Simulations of gravitational collapse in null coordinates.
I. Formulation and weak-field tests in generalized Bondi gauges

Carsten Gundlach⊙

Mathematical Sciences, University of Southampton, Southampton SO17 1BJ, United Kingdom

David Hilditch⊙

*CENTRA, Departamento de Física, Instituto Superior Técnico IST, Universidade de Lisboa UL,
Avenida Rovisco Pais 1, 1049 Lisboa, Portugal*

Thomas W. Baumgarte⊙

Department of Physics and Astronomy, Bowdoin College, Brunswick, Maine 04011, USA

We present a code for numerical simulations of the collapse of regular initial data to a black hole in null coordinates. We restrict to twist-free axisymmetry with scalar field matter. Our coordinates are (u, x, θ, φ), where the retarded time u labels outgoing null cones emerging from a regular central worldline, the angles (θ, φ) label the null generators of each null cone, and the radial coordinate x labels points along these generators. We focus on a class of generalized Bondi radial coordinates x with the twin properties that $x = 0$ is the central worldline and that the numerical domain ($u \geq 0, 0 \leq x \leq x_{\max}$) is a subset of the domain of dependence of the initial data on ($u = 0, 0 \leq x \leq x_{\max}$). In critical collapse, an appropriate choice of these coordinates can be made to zoom in on the accumulation point of scale echoes of the critical solution, without the need for explicit mesh refinement. We introduce a novel numerical scheme that in effect reduces the angular resolution at small radius, such that the time step Δu for an explicit numerical scheme is limited by the radial resolution Δx, rather than $\Delta x (\Delta \theta)^2$. We present convergence tests in the weak-field regime, where we have exact solutions to the linearized scalar and gravitational-wave equations.

I. INTRODUCTION

A. Motivation

Formulations of the Einstein equations on null surfaces are attractive in both mathematical and numerical relativity for several reasons.

As is generally done in the literature, we will here consider null coordinates where the surfaces of constant coordinate u are null hypersurfaces, and the lines of constant (u, θ, φ) are their null generators [1,2]. Such coordinates are sometimes called "Bondi-like," not to be confused with Bondi coordinates, where in addition the radial coordinate x is chosen to be the area radius.

As we will review, the Einstein equations in Bondi-like coordinates can be formulated in a way that makes them maximally constrained, in the following sense. In affine gauge or Bondi gauge, one solves two evolution equations, representing two polarizations of gravitational wave (or one in twist-free axisymmetry). On each null slice of constant u, the metric is completely determined by solving partial differential equations (PDEs) with derivatives only in the slice. Moreover these can be solved by explicit integration along the null cone generators. In double-null coordinates there is one additional evolution equation for the area radius R.

This is useful in numerical relativity in two ways. First, the evolution cannot drift away from a consistent state on each time slice through numerical error, in the sense that the hypersurface equations are solved on each time slice, not just the initial one, and so only free data are evolved.

Secondly, the equations can easily be discretized in a way that is compatible with causality. This makes it straightforward to evolve on the domain of dependence of the initial data, or to impose boundary conditions on timelike inner and outer boundaries, or to extend the numerical domain to future null infinity.

A third reason for using Bondi-like coordinates in numerical relativity is that any results obtained in them have immediate geometric significance, in contrast to, say, harmonic coordinates or "puncture" coordinates. For all these reasons, null coordinates are often the method of choice in spherical symmetry (see Sec. I B for references).

Beyond spherical symmetry, there is an obvious problem with null coordinates: null surfaces generically form caustics. However, we expect the expansion of outgoing null cones to prevent caustics in spacetimes that are

sufficiently close to being either flat or spherically symmetric, with the origin of the null cones near the center of approximate spherical symmetry. In a companion paper [3] (from now, Paper II) we will consider axisymmetric spacetimes with an additional reflection symmetry through the equatorial plane, so that there is a preferred worldline fixed by this symmetry.

This motivates us to investigate *when* null coordinates can be used to simulate nonspherical gravitational collapse, and how best to do this. We are not, in fact, aware of *any* use of null coordinates in the numerical evolution of regular initial data to a black hole (gravitational collapse) beyond spherical symmetry. Our ultimate motivation for this is vacuum critical collapse, see [4] for a general review of critical collapse and [5] for what we consider to be the state of the art, at the time of writing, in vacuum critical collapse.

The present paper is concerned with general considerations, the derivation of a number of possible radial gauges adapted to critical collapse, spacetime diagnostics, the presentation of our numerical methods, and weak-field convergence tests. In Paper II we apply these methods to axisymmetric scalar field critical collapse. In another companion paper [6], we consider issues of hyperbolicity and well-posedness.

B. Previous numerical work in null coordinates

A spacetime coordinate u is called null if the surfaces of constant u are null or, in terms of the spacetime metric, $g^{uu} = 0$. We speak of null coordinates when one of the coordinates is null, and of double null coordinates when two of them (usually called u and v) are null. In our terminology, u will always be an outgoing null coordinate, also called retarded time. For general review papers on the use of such (single) null coordinates in numerical relativity see [7,8].

One natural choice of null surfaces is the set of null cones emerging from a regular central worldline. The cones are labeled by the retarded time u and their null generators by (u, θ, φ). The fourth coordinate, which we generically call x, then labels points on each generator.

This formulation has been implemented in twist-free axisymmetry, in Bondi gauge, where x is the area radius R, compactified at future null infinity, both in vacuum [9] and with perfect fluid matter [10]. Here, the null cones emanate from a regular center. This brings about a severe limitation of the time step to $\Delta u \sim \Delta x (\Delta \theta)^2$, see also Appendix A.

Without symmetry restrictions, Bondi gauge in vacuum has been implemented in [11], and with perfect fluid matter in [12], both using stereographic coordinates on the 2-spheres of constant (u, x). The vacuum case has also been implemented using angular coordinates on the 2-spheres in [13]. These papers are focused on gravitational wave extraction, and so their null cones emanate from a regular timelike world cylinder, on which boundary data

must be given. This also avoids the time step problem at the origin.

A formulation where the radial coordinate x is the affine parameter λ along the null generators has been implemented in spherical symmetry in a cosmological setting with fluid matter in [14], and in an application to spherical scalar field critical collapse in [15], and to vacuum in spherical symmetry with initial data on two intersecting null cones in [16]. Affine gauge in vacuum without symmetries was formulated in [17], but not implemented in a code.

Affine gauge has also been used in a number of papers in the context of asymptotically anti-de Sitter spacetimes, on *ingoing* null surfaces emanating from the timelike infinity and terminating inside a black hole apparent horizon [18]. Mentioning only the two applications most relevant for us, in [19] this was done in $4 + 1$ dimensions with two commuting translation symmetries (and therefore mathematically similar to the twist-free axisymmetric case in $3 + 1$ dimensions), and in $3 + 1$ dimensions without symmetries in [20,21].

As far as we know, no attempt has been made to simulate the collapse of regular initial data to a black hole on null cones emanating from a regular center, except in spherical symmetry. From among the many successful applications in spherical symmetry, we review here only the application to the gravitational collapse of a spherically symmetric massless scalar field. An early study of scalar field collapse in Bondi coordinates was [22]. Essentially the same algorithm was used in [23] for the study of power-law tails and quasinormal modes in scalar field collapse. Double-null coordinates were used for the study of spherical scalar field critical collapse in [24], Bondi coordinates compactified at future null infinity in [25], and (as already mentioned) affine coordinates in [15].

"Type II" critical collapse is characterized by an arbitrarily large range of spacetime scales, and hence typically requires adaptive mesh refinement in numerical simulations. In fact, the pioneering paper [26] was made possible only by the first use of adaptive mesh refinement in numerical relativity. However, with the benefit of hindsight, the required mesh refinement zooms in on a single spacetime point. Once that point has been identified, a much simpler refinement scheme is possible, such as nested boxes in Cartesian coordinates. A fixed grid in polar-radial coordinates centered on this point and spaced logarithmically in radius also provides the required spatial resolution, but at the cost of a time step everywhere set by the smallest radial grid spacing.

A fixed grid providing the required mesh refinement in space and time for spherically symmetric critical collapse was implemented by Garfinkle [27] using double-null coordinates. The numerical domain is $(u \geq 0, x \leq x_0)$, with x an ingoing null coordinate. Hence the numerical domain is exactly the domain of dependence of the initial data.

Its outer boundary is the ingoing null cone $x = x_0$, which converges to a point. In the evolution of near-critical initial data, an appropriate choice of x_0 then puts the apex of the numerical domain near the accumulation point of scale echoes in near-critical solutions. Whenever half the x-grid points have fallen into the center, the resolution is doubled by regridding, with the time step adjusted accordingly. The numerical grid thus "zooms in" on the (approximate) critical solution as efficiently as possible, without the considerable complications of standard adaptive mesh refinement schemes.

Beyond spherical symmetry, ingoing null cones generically develop caustics, rather than refocus on a central world line. (This is not a concern for a sufficiently short time in a setup with initial data on two null cones $u = 0$ and $v = 0$ that intersect in a spacelike two-sphere, one often used in mathematical relavity for the study of black-hole spacetimes, see for example [28]).

However, we can rescue the key idea of [27] if we choose x in such a way that $x = 0$ is the regular center while $x = x_0$ is ingoing null (possible in spherical symmetry only) or future spacelike (ingoing faster than light). We first implemented this in spherical symmetry [29], and found that it provides mesh refinement for critical collapse as efficiently as the algorithm of [27].

In hindsight this is similar to using an ingoing radial shift with spacelike time slices to make the outer boundary ingoing null or future spacelike. This had already been used for spherical critical collapse on spacelike time slices in [30].

C. Plan of this paper

A fresh approach to critical collapse using null coordinates looks promising for the following three reasons, already mentioned above. First, because of their geometric rigidity, the use of null cones give us any approximate critical solution we find in coordinates already adapted to discrete self-similarity. Second, we do not expect problems with constraint violations. Finally, in a suitable discretization the outer boundary (assumed future spacelike) can be treated exactly like the interior points.

A natural stepping stone from spherical symmetry to vacuum collapse is nonspherical scalar field collapse, which can be examined with an arbitrary degree of nonsphericity, whereas vacuum collapse is necessarily very nonspherical. We shall present our results for scalar field critical collapse in twist-free axisymmetry in Paper II.

The structure of the present paper is as follows. In order to highlight the general mathematical structure of the Einstein equations in null coordinates, in Sec. II we review them in $n + 2$ spacetime dimensions, without symmetry assumptions.

Starting from Sec. III, we restrict to twist-free axisymmetry in the usual $3 + 1$ dimensions, and add a massless scalar field as matter. We review standard gauge choices, and propose several new ones for critical collapse. We also

discuss diagnostic quantities such as the Hawking mass, and how one can hope to identify apparent and event horizons.

Section IV describes our numerical methods, and in particular a novel method for completely overcoming the time step problem mentioned above, such that $\Delta u \sim \Delta x$.

Section V describes convergence tests of the full nonlinear equations in a small data regime where the linearization of the equations about Minkowski is a good approximation.

We conclude with a summary and outlook in Sec. VI.

A number of appendixes give details of (Appendix A) the time step problem, null coordinates in (Appendix B) Minkowski spacetime and (Appendix C) spherical symmetry, (Appendix D) exact solutions of the linearized equations that we use as test beds in the small data regime, (Appendix E) the residual gauge freedom in our coordinate choice, and (Appendix F) regularity conditions for the metric at the origin and on the symmetry axis.

II. NULL COORDINATES ON GENERIC SPACETIMES OF ARBITRARY DIMENSION

A. Metric ansatz

Throughout this paper, $a, b, \ldots$ are abstract tensor indices on the full spacetime ($n + 2$-dimensional in this section, and $2 + 2$-dimensional in the rest of the paper), and ∇_a is the covariant derivative with respect to the spacetime metric g_{ab}. $i, j, k = 1 \ldots n$ are angular coordinate indices, and $\mu, \nu, \ldots = 1 \ldots n + 2$ are spacetime coordinate indices, with $n \geq 2$ in this section only, and $n = 2$ in the remainder of the paper.

In $n + 2$ spacetime dimensions, we define null coordinates (u, x, θ^i), with $i = 1 \ldots n$, by demanding that the hypersurfaces of constant u are null, in the sense that $g^{ab} \nabla_a u \nabla_b u = g^{uu} = 0$. The vector field

$$U^a := -\nabla^a u \tag{1}$$

is obviously null, and obeys

$$U^a \nabla_a U_b = U^a \nabla_b U_a = \frac{1}{2} \nabla_b (U_a U^a) = 0. \tag{2}$$

Hence U^a is the tangent vector to the affinely parametrized null geodesics ruling the null surfaces of constant u. It is easy to verify that the most general metric obeying $g^{uu} = 0$ can be written in $n + 2$ form as

$$ds^2 = -2G du dx - H du^2 + \tilde{\gamma}_{kl}(d\tilde{\theta}^k + \hat{\beta}^k du + \tilde{\beta}^k dx)$$
$$\times (d\tilde{\theta}^l + \hat{\beta}^l du + \tilde{\beta}^l dx). \tag{3}$$

Note there are

$$2 + \frac{n(n + 1)}{2} + 2n = \frac{(n + 3)(n + 2)}{2} - 1 \tag{4}$$

metric coefficients $(G, H, \tilde{\gamma}_{kl}, \tilde{\beta}^k, \hat{\beta}^k)$, which is the full number of independent metric coefficients in $n + 2$ spacetime dimensions, minus the one coordinate condition $g^{uu} = 0$. At this point we still have two independent n-dimensional angular "shift vectors" $\tilde{\beta}^k$ and $\hat{\beta}^k$.

We now make the gauge transformation from $(u, x, \tilde{\theta}^k)$ to (u, x, θ^i) where $\theta^i(u, x, \tilde{\theta}^k)$ is given implicitly by a solution of the system

$$\tilde{\theta}^k_{,x}(u, x, \theta^i) = -\tilde{\beta}^k[u, x, \tilde{\theta}^l(u, x, \theta^j)] \quad (5)$$

of n coupled first-order ordinary differential equations (ODEs) in x for the functions $\tilde{\theta}^k(u, x, \theta^i)$. In the new coordinates the metric then takes the form

$$ds^2 = -2G\,du\,dx - H\,du^2 + \gamma_{ij}(d\theta^i + \beta^i du)(d\theta^j + \beta^j du),$$

$$(6)$$

where

$$\gamma_{ij} = \tilde{\gamma}_{kl}\tilde{\theta}^k_{,i}\tilde{\theta}^l_{,j}, \quad (7)$$

$$\beta^i = \theta^i_{,k}(\tilde{\theta}^k_{,u} + \hat{\beta}^k), \quad (8)$$

and $\theta^i_{,k}$ is the matrix inverse $\tilde{\theta}^k_{,i}$.

With the coordinates in the order $x^\mu := (u, x, \theta^i)$, the metric tensor can be written in matrix form as

$$g_{\mu\nu} = \begin{pmatrix} -H + \gamma_{ij}\beta^i\beta^j & -G & \gamma_{jk}\beta^k \\ -G & 0 & 0 \\ \gamma_{ik}\beta^k & 0 & \gamma_{ij} \end{pmatrix}. \quad (9)$$

Note that the induced metric on the surfaces of constant u, given by the bottom right submatrix of (9), is degenerate with signature $0 + + \cdots +$, as one would expect.

The inverse metric is

$$g^{\mu\nu} = \begin{pmatrix} 0 & -\frac{1}{G} & 0 \\ -\frac{1}{G} & \frac{H}{G^2} & \frac{\beta^j}{G} \\ 0 & \frac{\beta^i}{G} & \gamma^{ij} \end{pmatrix}, \quad (10)$$

where γ^{ij} is the inverse of γ_{ij}. We see that $G = 0$ would be a coordinate singularity where the metric has no inverse. Hence we assume that $G > 0$. There is no such restriction on H.

We see from (10) that in our coordinates the vector field U^a takes the simple form

$$U = \frac{1}{G}\partial_x, \quad (11)$$

and so the outgoing null geodesics that rule each surface of constant u (its generators) are simply the lines of constant (u, θ^i). (We have already mentioned that such coordinates are sometimes called "Bondi-like"). With $G > 0$, x is
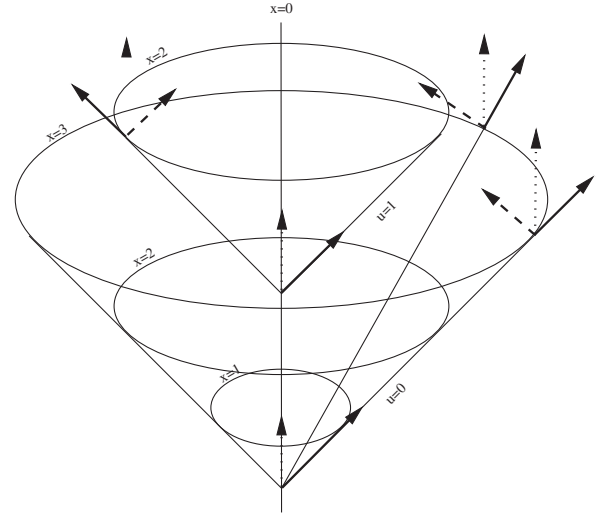

FIG. 1. Schematic spacetime picture showing our coordinates and basis vectors. Shown are two null cones of constant u, four spacelike closed 2-surfaces of constant (u, x), the central worldline $x = 0$, and outgoing null rays of constant (u, y, φ) (the angular coordinate $y = -\cos\theta$ is suppressed here). Solid arrows represent the outgoing null vector $U \propto \partial_x$, dashed arrows the ingoing null vector Ξ, and dotted arrows the timelike vector ∂_u (which are timelike near the origin but may tip inward to become spacelike further out).

strictly increasing with the affine parameter of the null cone generators.

For the following discussions, we denote by $\mathcal{N}_u^+$ the outgoing $n + 1$-dimensional null hypersurfaces of constant u, and by $\mathcal{L}_{u,\theta^i}^+$ the outgoing affinely parametrized null geodesics that rule each $\mathcal{N}_u^+$ [lines of constant (u, θ^i)]. We denote by $\mathcal{S}_{u,x}$ the n-dimensional spacelike surfaces of constant u and x, by $\mathcal{N}_{u,x}^-$ the ingoing null surface that emerges from each $\mathcal{S}_{u,x}$, and by $\mathcal{L}_{u,x,\theta^i}^-$ the affinely parametrized null geodesics that rule it. Note that on $\mathcal{N}_{u,x}^-$ and $\mathcal{L}_{u,x,\theta^i}^-$ none of the coordinates are constant: the coordinate values that label them are only starting values. Our coordinates and basis vectors are sketched in Fig. 1.

B. Standard radial gauge choices

We see from (9) that $g_{xi} = 0$ for $i = 1 \ldots n$, and from (10) that $g^{uu} = 0$, both by construction. We have used up $n + 1$ of the possible $n + 2$ coordinate conditions, with one remaining to be imposed.

From an $n + 2$ perspective, this final gauge condition should not single out any spatial coordinate θ^i, and so should involve only G, H and $\det \gamma_{ij}$. If we think of this condition as fixing, for example, the metric coefficient H, we have

$$\frac{(n+3)(n+2)}{2} - (n+2) = 1 + \frac{n(n+1)}{2} + n \quad (12)$$

independent metric coefficients: on the left-hand side the number of algebraically independent metric coefficients $g_{\mu\nu}$ in $n+2$ spacetime dimensions without symmetry, minus $n+2$ coordinate conditions, and on the right-hand side the number of components in $(G, \gamma_{ij}, \beta^k)$. We are aware of three such conditions in the literature.

1. Bondi

One may be able to assume that the surfaces $\mathcal{S}_{u,x}$ are n-spheres and that their volume increases monotonically with x. Then Bondi coordinates are defined by the coordinate condition $\det \gamma_{ij} = x^2 \det \bar{\gamma}_{ij}$, where $\bar{\gamma}_{ij}$ is the unit round metric on S^n in the coordinates θ^i. x is then called the area radius and is usually, and in this paper, denoted by r.

2. Double-null

Double-null coordinates are defined by $g^{xx} = 0$, which, from (10), is equivalent to $H = 0$. x is then a second null coordinate, and is usually, and in this paper, denoted by v. The $\mathcal{N}_{u,v}^-$ are now surfaces of constant v. The affinely parametrized $\mathcal{L}_{u,v,\theta^i}^-$ have the tangent vector

$$V^a := -\nabla^a v, \tag{13}$$

which in coordinates takes the form

$$V = \frac{1}{G}(\partial_u - \beta^i \partial_i). \tag{14}$$

As already mentioned, we are not aware of a numerical application of double null coordinates beyond spherical symmetry. This may be because we expect ingoing null cones to develop caustics generically, rather than converge to a point.

3. Affine

Finally, one sees from (11) that x is an affine parameter along the outgoing null geodesics if and only if the coordinate condition $G_{,x} = 0$ holds. x is then often called λ. λ on each outgoing null geodesic is fixed up to an additive constant by fixing the function $G(u, \theta^i)$. The most common choice is $G = 1$.

C. The hierarchy of Einstein equations

In order to have nontrivial spacetimes with a regular center even in the limit of spherical symmetry, we add a massless minimally coupled scalar field ψ that obeys the wave equation

$$\nabla^a \nabla_a \psi = 0. \tag{15}$$

The Einstein equations with scalar field matter in any spacetime dimension can be written, in trace-reversed form, as

$$E_{ab} := R_{ab} - 8\pi \nabla_a \psi \nabla_b \psi = 0. \tag{16}$$

R_{ab} is the spacetime Ricci tensor, and we use gravitational units where $c = G = 1$.

We define the ingoing null vector field

$$\Xi := \partial_u - \frac{H}{2G}\partial_x - \beta^i \partial_i. \tag{17}$$

Together Ξ^a and U^a span the normal space to each $\mathcal{S}_{u,x}$. They are normalized so that $\Xi^a U_a = -1$. U^a, given in (11), is tangent to the affinely parametrized generators of $\mathcal{N}_u^+$, while Ξ^a is tangent to the generators of $\mathcal{N}_{u,x}^-$ where they emerge from $\mathcal{S}_{u,x}$.

Given only the metric components γ_{ij} on a surface of constant u, and boundary values for G, β^i and $\beta^i_{,x}$ on any past boundary $x = x_{\min}(u, \theta^i)$ of that surface, we can solve the Einstein equation $E_{xx} = 0$ for G, $E_x{}^i = 0$ for β^i, and $E_{ij} = 0$ for the null derivatives $\Xi \gamma_{ij}$, in this order, by explicit integration in x. In the terminology of [2] these are the "main equations." We will call them the "hierarchy equations." They contain H only in the combination Ξ. Explicit expressions in twist-free axisymmetry in $3+1$ spacetime dimensions will be given in Sec. III below. Given the $\Xi \gamma_{ij}$, one gauge condition is required to fix H and so find the "time" derivatives $\gamma_{ij,u}$. We can then advance γ_{ij} in u and repeat.

The remaining $n+2$ Einstein equations $E_{ux} = 0$ (the "trivial equation" [2]), $E_{uu} = 0$ and $E_u{}^i = 0$ (the "supplementary conditions" [2]) contain higher u-derivatives than the hierarchy equations and are redundant modulo the $n+2$ contracted Bianchi identities. The supplementary conditions also act as constraints on the boundary data imposed at $x = x_{\min}$. We do not discuss them here because $x = x_{\min} = 0$ will always be a regular central world line here and in Paper II, so no free data can be imposed there. Finally, the trivial equation is an algebraic consequence of imposing all other equations.

With the shorthand

$$B := \frac{H}{2G}, \tag{18}$$

we can write (17) as

$$\partial_u = \Xi + B\partial_x + \beta^i \partial_i. \tag{19}$$

This suggests that we consider B and β^i as the x and θ^i components of a "shift vector" representing the difference between the coordinate time direction ∂_u and the null vector Ξ. However, while the future-pointing unit vector n^a normal to a spacelike hypersurface Σ is unique, the null vector Ξ^a depends not only on the null hypersurface $\mathcal{N}_u^+$, but also on its foliation by n-surfaces $\mathcal{S}_{u,x}$.

III. TWIST-FREE AXISYMMETRY IN 3 + 1 WITH A SCALAR FIELD

A. Metric ansatz and matter field

We now restrict to $3 + 1$ spacetime dimensions in spherical polar null coordinates (u, x, θ, φ). We assume axisymmetry with Killing vector $K := \partial_\varphi$, meaning that $g_{\mu\nu,\varphi} = 0$ in those coordinates. In addition, we assume a reflection symmetry $\varphi \to -\varphi$. This is a consistent truncation, in the sense that if we impose $\gamma_{\theta\varphi} = 0$ on the initial data, and the boundary conditions $\beta^\varphi = \beta^\varphi_{,x} = 0$ at $x = x_{\min}$ to start up the integration, the hierarchy equations give us $\beta^\varphi = 0$ and $\gamma_{\theta\varphi,u} = 0$. Geometrically, the twist vector $\epsilon^{abcd} K_a \nabla_b K_c$ vanishes, hence the name twist-free axisymmetry. Physically, this symmetry removes one of the two gravitational wave degrees of freedom, hence the alternative name polarized axisymmetry.

We identify a worldline on the symmetry axis world sheet, calling it the central worldline, or center for short. (There is a preferred choice for this if the spacetime has a reflection symmetry $z \to -z$, or $\theta \to \pi - \theta$, but in general the choice is arbitrary. For now we stay in the general case.) The null cones of constant u are assumed to have a regular vertex on the central worldline.

We parametrize the metric under these conditions as

$$ds^2 = -2Gdudx - Hdu^2$$
$$+ R^2\left[e^{2F}(d\theta + \beta du)^2 + e^{-2F}\sin^2\theta d\varphi^2\right], \quad (20)$$

where the metric coefficients (G, H, R, F, β) depend on the coordinates (u, x, θ) only. R is the area radius, in the sense that $\det \gamma_{ij} = R^2 \sin^2\theta$, and so the area of $\mathcal{S}_{u,x}$ is $4\pi R^2$. The central worldline is at $R = 0$.

B. Regularization of the axis

The field equations can be regularized on the symmetry axis by reparametrizing [9]

$$\beta =: \sin\theta b, \qquad F =: \sin^2\theta f \quad (21)$$

and replacing the coordinate θ by

$$y := -\cos\theta, \qquad -1 \le y \le 1. \quad (22)$$

Intuitively, b is the z-component of the shift vector $\beta\partial_\theta$, and therefore regular on the symmetry axis. The metric becomes

$$ds^2 = -2Gdudx - Hdu^2$$
$$+ R^2\left[e^{2Sf}S^{-1}(dy + Sbdu)^2 + e^{-2Sf}Sd\varphi^2\right], \quad (23)$$

where we have defined the shorthand

$$S := 1 - y^2 = \sin^2\theta. \quad (24)$$

We also have

$$\Xi = \partial_u - B\partial_x - bS\partial_y. \quad (25)$$

Even though the metric now has a division by S, the hierarchy equations for $(G, b, H, \Xi R, \Xi f, \Xi \psi)$ are regular on the axis in the sense that they contain neither square roots nor divisions by y or S, see Eqs. (29)–(38) below.

We note in passing the identities

$$\psi_{,\theta} = \sqrt{1 - y^2}\psi_{,y}, \quad (26)$$

$$\psi_{,\theta\theta} = (1 - y^2)\psi_{,yy} - y\psi_{,y}. \quad (27)$$

From (26) we see that the usual regularity condition for scalars on the symmetry axis, $\psi_{,\theta} = 0$ at $\theta = 0$ and π, corresponding to $y = -1$ and $y = 1$, does not impose a condition on $\psi_{,y}$ there. Rather, we see from (27) that $\psi_{,y} = \pm\psi_{,\theta\theta}$ there, which is unconstrained by regularity.

C. The equation hierarchy

In twist-free axisymmetry, the Einstein equations have seven algebraically independent components $E_{\mu\nu}$. It is convenient to define the two linear combinations

$$E_\pm := e^{-2F}SE_{yy} \pm e^{2F}S^{-1}E_{\varphi\varphi}. \quad (28)$$

Geometric free data on an outgoing null cone of constant u consist of f and ψ as functions of R and y. In double-null gauge, we also need to specify R as a function of x and y there, considering this as fixing a gauge freedom in the initial data.

The two Einstein equations $E_{xx} = 0$ and $E_{xy} = 0$ do not contain any u-derivatives. They can, respectively, be written as

$$\left(\ln\frac{G}{R_{,x}}\right)_{,x} = S_G[R, f, \psi] \quad (29)$$

$$\left(\frac{R^4 e^{2Sf} b_{,x}}{G}\right)_{,x} = S_b[R, f, \psi, G]. \quad (30)$$

The Einstein equations $E_+ = 0$ and $E_- = 0$ and the scalar wave equation all contain u-derivatives. They can, respectively, be written as

$$(R\Xi R)_{,x} = S_R[R, f, \psi, G, b], \quad (31)$$

$$(R\Xi f)_{,x} = S_f[R, f, \psi, G, b] - (\Xi R)f_{,x}, \quad (32)$$

$$(R\Xi\psi)_{,x} = S_\psi[R, f, \psi, G, b] - (\Xi R)\psi_{,x}, \quad (33)$$

where Ξ is the derivative operator defined in (25), and H only appears as part of Ξ. The right-hand sides $S[f, \ldots]$ contain the derivatives $f_{,x}, f_{,y}, f_{,xy}$ and $f_{,yy}$ (but not $f_{,xx}$),

and the same derivatives of R, G, b and ψ, with the exceptions of $\psi_{,xy}$ and $b_{,yy}$.

The remaining algebraically independent Einstein equations are E_{uu}, E_{ux} and E_{yy}, and contain u-derivatives other than those already appearing in the hierarchy equations. In particular, E_{uu} (only) contains $R_{,uu}$; E_{uu} and E_{uy} contain $R_{,uy}$, $f_{,uy}$ and $G_{,uy}$; E_{uu} and E_{ux} contain $G_{,ux}$; and all three

contain $G_{,u}$. We do not investigate here what relevance these have as constraints on the data at an inner boundary $x = x_{\min}$.

The full right-hand sides of the hierarchy equations are as follows, beginning with the principal terms on a separate line (S_G is all nonprincipal), and the scalar field stress-energy terms at the end

$$S_G = \frac{R}{R_{,x}}\left(S^2 f_{,x}^2 + 4\pi\psi_{,x}^2\right), \tag{34}$$

$$S_b = 2R^2 S f_{,xy} - \frac{R^2 G_{,xy}}{G} - 2RR_{,xy} + f_{,x}(-4R^2 S^2 f_{,y} + 8R^2 y(fS-1) + 4RSR_{,y}) + \frac{R^2 G_{,x}G_{,y}}{G^2} + \frac{2RG_{,y}R_{,x}}{G}$$
$$+ 2R_{,x}R_{,y} - 16\pi R^2 \psi_{,x}\psi_{,y}, \tag{35}$$

$$S_R = \frac{1}{4}R^2 S b_{,xy} + \frac{GSe^{-2fS}R_{,yy}}{2R} - \frac{1}{2}GS^2 e^{-2fS}f_{,yy} + \frac{1}{4}Se^{-2fS}G_{,yy} + \frac{R^4 Sb_{,x}^2 e^{2fS}}{8G} + b_{,x}\left(-\frac{R^2 y}{2} - \frac{1}{2}RSR_{,y}\right)$$
$$+ RSb_{,y}R_{,x} - 2bRyR_{,x} + e^{-2fS}\left[f_{,y}\left(-4GSy(fS-1) - \frac{GS^2 R_{,y}}{R}\right) + \frac{Gy(2fS-1)R_{,y}}{R} + GS^3 f_{,y}^2 - \frac{SG_{,y}^2}{8G}\right.$$
$$+ G_{,y}\left(-\frac{1}{2}S^2 f_{,y} + fSy - \frac{y}{2}\right) + \frac{1}{2}G\left(2f(S(5 - 4f(S-1)) - 4) - 1\right) - \frac{GSR_{,y}^2}{2R^2} + 2\pi GS\psi_{,y}^2\right], \tag{36}$$

$$S_f = \frac{Rb_{,xy}}{4} + \frac{e^{-2fS}G_{,yy}}{4R} - 3bRyf_{,x} + \frac{R^3 b_{,x}^2 e^{2fS}}{8G} + b_{,x}\left(-\frac{1}{2}RSf_{,y} - fRy\right) + b_{,y}\left(\frac{1}{2}RSf_{,x} + \frac{R_{,x}}{2}\right)$$
$$- 2bfyR_{,x} + e^{-2fS}\left(-\frac{G_{,y}R_{,y}}{2R^2} - \frac{G_{,y}^2}{8GR} + \frac{2\pi G\psi_{,y}^2}{R}\right), \tag{37}$$

$$S_\psi = \frac{GSe^{-2fS}\psi_{,yy}}{2R} + \frac{1}{2}RSb_{,y}\psi_{,x} - \frac{1}{2}RSb_{,x}\psi_{,y} - bRy\psi_{,x} + e^{-2fS}\left(-\frac{GS^2 f_{,y}}{R} + \frac{Gy(2fS-1)}{R} + \frac{SG_{,y}}{2R}\right)\psi_{,y}. \tag{38}$$

D. Formulation in terms of $\gamma := \ln(G/R_{,x})$ and R-derivatives

Note that S_G given in (34) has a division by $R_{,x}$. If we assume that $R_{,x} > 0$ everywhere, we can reparametrize the metric variable G by the new variable

$$g := \frac{G}{R_{,x}} \Rightarrow G = R_{,x}g. \tag{39}$$

Then all x-derivatives in the hierarchy equations can be eliminated in favor of the derivative

$$\mathcal{D} := \frac{1}{R_{,x}}\partial_x, \tag{40}$$

that is, the derivative with respect to R at constant u and y. Moreover, g is invariant under reparametrizing x, and in particular is simply 1 in flat spacetime. Obviously, $\mathcal{D}$ does not commute with ∂_u and ∂_y, so we have to specify the order of mixed derivatives. As a convention, we apply $\mathcal{D}$ last.

If we further replace g by

$$\gamma := \ln g, \tag{41}$$

the equations simplify a little further, and γ appears undifferentiated only in the two combinations $\exp\pm(2Sf - \gamma)$. Hence in our final numerical formulation we treat γ as the primary variable.

The hierarchy equations now take the form

$$\mathcal{D}\gamma = \bar{S}_\gamma[R, f, \psi], \tag{42}$$

$$\mathcal{D}(R^4 e^{2Sf-\gamma}\mathcal{D}b) = \bar{S}_b[R, f, \psi, \gamma], \tag{43}$$

$$\mathcal{D}(R\Xi R) = \bar{S}_R[R, f, \psi, \gamma, b], \tag{44}$$

$$\mathcal{D}(R\Xi f) = \bar{S}_f[R, f, \psi, \gamma, b] - (\Xi R)\mathcal{D}f, \tag{45}$$

$$\mathcal{D}(R\Xi\psi) = \bar{S}_\psi[R, f, \psi, \gamma, b] - (\Xi R)\mathcal{D}\psi, \tag{46}$$

where $\bar{S} = S/R_{,x}$. With $G = gR_{,x}$, third derivatives of R appear in the Einstein equations when we write them in terms of g or γ. In the hierarchy equations, only $R_{,xxy}$ and $R_{,xyy}$ appear. The former, which appears only in the equation for b, is problematic numerically, but it can be eliminated by writing

$$\bar{S}_b[R, f, \psi, \gamma] = -\mathcal{D}(R^2\mathcal{D}(R_{,y})) + \tilde{S}_b[R, f, \psi, \gamma]. \quad (47)$$

The modified source $\tilde{S}_b$ no longer contains third derivatives, and we can explicitly integrate the first term on the right.

The source terms are now

$$\bar{S}_\gamma = R\big(S^2\mathcal{D}(f)^2 + 4\pi\mathcal{D}(\psi)^2\big), \quad (48)$$

$$\tilde{S}_b = R^2\Big(-\big(4\mathcal{D}(f)(S^2f_{,y} - 2fSy + 2y) - 2S\mathcal{D}(f_{,y}) + \mathcal{D}(\gamma_{,y}) + 16\pi\mathcal{D}(\psi)\psi_{,y}\big)\Big)$$
$$+ 2R(2S\mathcal{D}(f)R_{,y} + \mathcal{D}(R_{,y}) + \gamma_{,y}) + 2R_{,y}, \quad (49)$$

$$\bar{S}_R = \frac{1}{8}e^{2fS-\gamma}R^4S\mathcal{D}(b)^2 - \frac{1}{4}R\big(-S(R\mathcal{D}(b_{,y}) + 4b_{,y}) + 2\mathcal{D}(b)(SR_{,y} + Ry) + 8by\big) + \frac{1}{8}e^{\gamma-2fS}\Big(32f^2Sy^2$$
$$- 4\mathcal{D}(R_{,y})(S^2f_{,y} - 2fSy + y) - \frac{8R_{,y}}{R}(S^2f_{,y} - 2fSy + y) + 8S^3f_{,y}^2 - 8f(4S^2yf_{,y} + 5y^2 - 1)$$
$$+ 2S\gamma_{,y}\mathcal{D}(R_{,y}) - S\mathcal{D}(R_{,y})^2 + 2S\mathcal{D}(R_{,yy}) + \frac{4SR_{,yy}}{R} - S\gamma_{,y}^2 + 2S(\gamma_{,y}^2 + \gamma_{,yy}) - 4$$
$$- 4\gamma_{,y}(S^2f_{,y} - 2fSy + y) + 32Syf_{,y} - 4S^2f_{,yy} - \frac{4SR_{,y}^2}{R^2} + 16\pi S\psi_{,y}^2\Big), \quad (50)$$

$$\bar{S}_f = \frac{1}{8}e^{2fS-\gamma}R^3\mathcal{D}(b)^2 + \frac{1}{4}\Big(R\big(2Sb_{,y}\mathcal{D}(f) + 2\mathcal{D}(b)(-Sf_{,y} - 2fy) + \mathcal{D}(b_{,y})\big) - 4by\big(3R\mathcal{D}(f) + 2f\big) + 2b_{,y}\Big)$$
$$+ \frac{e^{\gamma-2fS}}{8R^2}\big(-4R_{,y}\gamma_{,y} + (2R\gamma_{,y} - 4R_{,y})\mathcal{D}(R_{,y}) - R\mathcal{D}(R_{,y})^2 + R\big(2\mathcal{D}(R_{,yy}) + \gamma_{,y}^2 + 2\gamma_{,yy} + 16\pi\psi_{,y}^2\big)\big), \quad (51)$$

$$\bar{S}_\psi = \frac{1}{2}\big(\mathcal{D}(\psi)(R(Sb_{,y} - 2by)) - RS\mathcal{D}(b)\psi_{,y}\big) + \frac{e^{\gamma-2fS}}{2R}\big(\psi_{,y}(-2S^2f_{,y} + 4fSy + S\mathcal{D}(R_{,y}) + S\gamma_{,y} - 2y) + S\psi_{,yy}\big). \quad (52)$$

E. Regular center

When the null surfaces are cones emanating from a regular central world line, null geodesics leaving the central worldline at the same time must carry the same u, and those leaving at different times are naturally identified as setting off in the same direction (y, φ) via parallel transport along the central wordline. The only remaining gauge freedom is to relabel x by $\tilde{x}(u, x, y)$, and u by $\tilde{u}(u)$.

Obviously, all metric coefficients must be single-valued on the central worldline, that is, at $x = 0$ they must be independent of y for all u. We stress this by writing $G(u, 0, y) = G(u, \mathbf{0})$, and so for all other quantities that are single-valued at the center.

We show in Appendix B that when the center $R = 0$ is regular, geodesic, and has coordinate location $x = 0$, in our gauge we have $b(u, \mathbf{0}) = f(u, \mathbf{0}) = 0$ and

$$g(u, \mathbf{0}) = U'(u), \quad (53)$$

$$H(u, \mathbf{0}) = U'(u)^2, \quad (54)$$

where $U(u)$ is proper time along the central worldline. Choosing u itself to be proper time, we have $g(u, \mathbf{0}) = H(u, \mathbf{0}) = 1$. Because g and G are single-valued at the origin, $R_{,x}$ at the origin must also be single-valued, and we denote it by $R_{,x}(u, \mathbf{0})$.

F. Standard radial gauge choices

We review here the standard radial gauge choices already outlined for arbitrary spacetime dimension above.

1. Bondi

The condition $R = x =: r$ defines Bondi coordinates (u, r, y, φ). In particular, the definition

$$\Xi R := R_{,u} - BR_{,x} - bSR_{,y}, \quad (55)$$

reduces in Bondi gauge, where $R_{,u} = R_{,y} = 0$, to

$$B = B_{\text{Bondi}} := -\frac{\Xi R}{R_{,x}} \quad (56)$$

Given ψ and f as functions of (x, y) on a surface of constant u, such as $u = 0$, and the gauge initial data $R(0, x, y) = x$, the hierarchy equations are solved for $(\gamma, b, \Xi R, \Xi f, \Xi \Psi)$ by integration. We then find B from ΞR and (56).

2. Double-null

The condition $H = 0$ and hence $B = 0$ defines double-null coordinates (u, v, y, φ), where $x =: v$ becomes the second null coordinate. Given ψ, f and R as functions of (x, y) on a surface of constant u, the hierarchy equations are again solved for $(\gamma, b, \Xi R, \Xi f, \Xi \psi)$ by integration. Here $R(0, x, y)$ can be thought of as fixing a gauge freedom. As already mentioned, this formulation is only almost-maximally constrained because the evolution equation for R does not relate to a physical degree of freedom but propagates this initial gauge choice.

3. Affine

The coordinate condition $G_{,x} = 0$ defines $x =: \lambda$ to be an affine parameter along the null geodesic generators of our time slices. In this case, the first hierarchy equation (29), (34) becomes

$$R_{,xx} + (S^2 f_{,x}^2 + 4\pi\psi_{,x}^2)R = 0, \qquad (57)$$

and so does not contain G at all, but instead becomes an ODE in x (at constant u and y) for R, given f and ψ. ΞR is now used for finding H, rather than for evolving R. By taking an x-derivative of the hierarchy equation (31) for ΞR and a u-derivative of (57), and eliminating $R_{,uxx}$ between them, we find an equation of the form

$$H_{,xx} = S_H[R, b, f, \psi, \Xi R, \Xi f, \Xi\psi], \qquad (58)$$

which can be integrated twice to find H.

We stress that any null gauge in which we solve the Raychaudhuri equation (29) for G or γ, such as Bondi or double null gauge, breaks down where $R_{,x} = 0$. But, as we will see later, the divergence ρ_+ of the null generators of our $\mathcal{N}_u^+$ is proportional to $R_{,x}$, so this will generically happen in strong gravity. The only alternative is to solve (29) for R, for example in affine gauge. We will consider this elsewhere.

G. Choices of radial gauge for critical collapse

We now present possible choices of radial gauge adapted to critical collapse that generalize the ideas of [27] beyond spherical symmetry. In these coordinates, we want $x = 0$ to be the regular center $R = 0$ and the outer boundary $x = x_{\max}$ to be future spacelike or null. We therefore evolve on the domain of dependence of the initial data, without the need for an explicit outer boundary condition, and the numerical domain shrinks with time. We can then hope to

control this shrinking in such a way that resolution of the critical solution is maintained without the need for adaptive mesh refinement.

The domain of dependence of the data on any $u = u_0$, $0 \leq x \leq x_{\max}$ is bounded by the ingoing null surface $\mathcal{N}_{u_0, x_{\max}}^-$, whose null tangent vector at $S_{u_0, x_{\max}}$ is Ξ. Hence at $x = x_{\max}$, x should not decrease along Ξ, or

$$\Xi x = -\frac{H}{2G} \geq 0 \quad \text{at } x = x_{\max}, \qquad (59)$$

and this must hold for all u. An equivalent requirement is that the surface $x = x_{\max}$ is null or spacelike, that is

$$|\nabla x|^2 = \frac{H}{G^2} \leq 0 \quad \text{at } x = x_{\max}. \qquad (60)$$

In short, H and therefore B must be nonpositive at $x = x_{\max}$. This also means that numerically we can consistently upwind the advection term $B\partial_x$ in the time evolution equations at the outer boundary, with the one-sided stencil pointing into the numerical domain.

At the inner boundary $x = 0$, the condition that $R = 0$ remains at $x = 0$ fixes B to be given by (56). This is positive, and so again we can upwind consistently with the one-sided stencil pointing into the numerical domain.

For applications to critical collapse we impose a condition at some $0 < x_0 \leq x_{\max}$ that makes $x = x_0$ approximately null: approximately in the sense that $B = 0$ there in some average sense, or that $B \leq 0$ with equality at one or more values of y. If a spacetime approaches a self-similar critical solution, and x_0 is chosen appropriately, the past light cone of the critical solution will then be at $x \simeq x_0$. This gauge condition should then also imply $B \leq 0$ at the outer boundary.

We now present a few possible gauges that obey these two boundary conditions at $x = x_0$ and $x = 0$.

1. Shifted double null gauge

Consider first the choice

$$B = B_{\text{sdn}} := \left(1 - \frac{x}{x_0}\right) B_{\text{Bondi}}(\mathbf{0}), \qquad (61)$$

for some constant parameter $x_0 \leq x_{\max}$. The Bondi radial shift was defined in (56). In other words, B decreases linearly in x from its value in Bondi gauge at the center to zero at the some x_0. Then $x = 0$ remains at $R = 0$, and $x = x_0$ is an ingoing null surface.

Our convention that u is proper time along the central worldline implies that $\Xi R = -1/2$ there, and so (56) gives

$$B_{\text{sdn}} = \left(1 - \frac{x}{x_0}\right)\frac{1}{2R_{,x}(u, \mathbf{0})}. \qquad (62)$$

We shall call this gauge choice shifted double null (from now on, sdn) gauge because it simply rescales the ingoing null coordinate v. More precisely, if we choose $x = v$ at $u = 0$, then at fixed u the new coordinate x is a linear function of v defined by $v = v_0$ at $x = 0$ and $R = 0$ at $x = 0$ [29].

Therefore, sdn gauge is a continuous version of the repeated regridding of the double-null coordinate v in [27]. We have previously implemented it in spherical symmetry in [29], and found that it works exactly as well as our earlier implementation, in [31], of the original Garfinkle regridding algorithm. The resulting numerical domains are identical for $x_0 = x_{\max}$, but we found in [29,31] that choosing $x_{\max} > x_0$, which gives us a spacelike buffer zone, has the advantage of revealing more of the critical solution spacetime.

On a regular spacetime beyond spherical symmetry, we expect sdn gauge to fail because ingoing null cones, and in particular $x = x_0$, do not reconverge on the central worldline. It is, however, extremely useful in spherical symmetry. We will now discuss alternatives that work better beyond spherical symmetry, but reduce to sdn gauge in spherical symmetry.

We have also not been able to find a stable discretization of the equations in sdn gauge beyond spherical symmetry, even in the limit of weak fields. The instability *looks* like a purely numerical problem at the origin, but we cannot exclude formation of caustics or ill-posedness as problems in the continuum. We have not explored this further.

2. Global shifted Bondi gauge

An alternative starting point is to demand that

$$R(u, x, y) = s(u)x, \tag{63}$$

where $s(u)$ is a function to be specified. This simplifies the hierarchy equations in the same way as Bondi gauge does, and in particular $G = g/s(u)$. Substituting this into the equation defining ΞR, we have

$$B = B_{\text{gsB}} := \frac{s'(u)x - \Xi R}{s(u)}. \tag{64}$$

We shall this class of gauges global shifted Bondi (from now on, gsB) gauge.

To avoid potential numerical instabilities from evolving R as a dynamical variable, we update it directly with $R_{,u} = s'(u)x$, rather than the generic expression based on ΞR plus shift terms, and also use (63) in simplifying other derivatives. This gauge does not suffer from the same numerical instabilities as sdn gauge.

To fix $s(u)$, we demand that the surface $x = x_0$ is ingoing null or future spacelike, in the sense that $H(u, x_0, y) \le 0$. This gives

$$s'(u) = \frac{1}{x_0} \min_y \Xi R(u, x_0, y). \tag{65}$$

As we shall see in Paper II, gsB gauge behaves very different from sdn gauge in strong fields, already in spherical symmetry, and does not seem to be a good choice for critical collapse.

3. Local shifted Bondi gauge

Yet another starting point is the observation that the instability we observe in our implementation of sdn gauge seems to be connected to the y-dependence of R, while the choice (63) is too restricted. Hence we can attempt the more general gauge

$$R(u, x, y) = \bar{R}(u, x), \tag{66}$$

by setting

$$B = B_{\text{lsB}} := \frac{\bar{R}_{,u} - \Xi R}{\bar{R}_{,x}}, \tag{67}$$

where $\bar{R}_{,u}(u, x)$ and hence $\bar{R}(u, x)$ is yet to be specified. We call this local shifted Bondi (from now on, lsB) gauge. We require

$$\bar{R}(u, 0) = 0 \tag{68}$$

to keep the origin regular at $x = 0$ and

$$\bar{R}_{,u}(u, x_{\max}) \le \min_y \Xi R(u, x_{\max}, y) \tag{69}$$

to make $B \le 0$ at the outer boundary.

We restrict the possible choices of $\bar{R}_{,u}$ by demanding that in spherical symmetry we revert to sdn gauge. An obvious possibility is to start from the Bondi shift (56) with its spherical part subtracted ("nonspherical Bondi", from now on nsB),

$$B_{\text{nsB}} := -\frac{\Xi R - (\Xi R)_{l=0}}{\bar{R}_{,x}}, \tag{70}$$

where the suffix $l = 0$ denotes the spherical part, and add the (purely spherical) sdn shift:

$$B = B_{\text{lsB1}} := B_{\text{nsB}} + B_{\text{sdn}}. \tag{71}$$

In this gauge x_0 is a null surface "on average."

We can add a further term that makes the shift nonpositive everywhere at x_0, giving

$$B_{\text{lsB2}} := B_{\text{lsB1}} + \frac{x}{x_0} \frac{(\min_y \Xi R) - (\Xi R)_{l=0}}{R_{,x}}(x_0). \tag{72}$$

In this gauge, $B_{,u}$ is discontinuous at such values of u where the location in y of $\min_y \Xi R(u, x_0, y)$ changes discontinuously. In either lsB1 or lsB2, there is no guarantee that $H < 0$ at the outer boundary to make it future spacelike.

Another possibility is to subtract, at *every* (u, x), the global maximum over y of the Bondi radial shift instead of its spherical part ("non-negative Bondi")

$$B_{\text{nnB}} := -\frac{\Xi R - \min_y(\Xi R)}{\bar{R}_{,x}}, \tag{73}$$

and again add the sdn shift,

$$B_{\text{lsB4}} := B_{\text{nnB}} + B_{\text{sdn}}. \tag{74}$$

This version has the property that every surface of constant $x \geq x_0$, and so the outer boundary in particular, is null or future spacelike. It has the disadvantage that $B_{,x}$ and $B_{,u}$ are discontinuous at all values of (u, x) where the location in y of $\min_y \Xi R(u, x, y)$ changes discontinuously.

We can also make a transition from lsb gauge, with the spherical part given by sdn, near the center, to full sdn at the outer boundary, so that the outer boundary is null everywhere. In other words, we consider

$$B_{\text{lsBtosdn}} := B_{\text{sdn}} + \left[1 - K_{01}\left(\frac{x - \lambda x_0}{x_0 - \lambda x_0}\right)\right]$$
$$\times \left[\left(\frac{\Xi R}{R_{,x}}\right)_{l=0} - \frac{\Xi R}{R_{,x}}\right], \tag{75}$$

where $K_{01}(x)$ is a sufficiently smooth switching function with $K_{01}(x \leq 0) = 0$ and $K_{01}(x \geq 1) = 1$, and $0 < \lambda \leq 1$ is a parameter. We then have pure lsb gauge for $0 \leq x \leq \lambda x_0$, and a transition to pure sdn over the interval $\lambda x_0 \leq x \leq x_0$. For K_{01} on the transition range $0 \leq x \leq 1$ we have settled on the 9th order polynomial defined by the first four derivatives vanishing at $x = 0$ and $x = 1$, and the symmetry condition $K_{01}(1/2) = 1/2$.

H. Spacetime diagnostics

1. Affine parameter

In this section we look at a number of ways of extracting geometric information from our simulations. After fixing the central worldline, a completely geometric coordinate system is given by (u, λ, y, φ), where u labels the null cones emanating from the central worldline, and (y, φ) label the generators of these null cones. u is fixed to be the proper time along the central worldline. The same (y, φ) at different u are identified by parallel transport of the vector ∂_u along the central worldline. λ is the affine parameter along the generators, with origin $\lambda = 0$ at $R = 0$ and normalization $\lambda \simeq R$ near the origin.

The tangent vector to the affinely parametrized generators of our null cones is U given by (11). We have $dx/d\lambda := U^a \nabla_a x = G^{-1}$ and $\lambda = 0$ on the central worldline $R = x = 0$. Integration then gives us

$$\lambda(u, x, y) = \int_0^x G(u, x', y) dx' = \int_0^R g(u, x', y) dR'. \tag{76}$$

for the affine parameter. Recall that in our convention $g = 1$ at the origin, so we have the required normalization.

2. Redshift

Let τ be the proper time measured by a timelike observer at coordinate location (u, x, y, φ) and with 4-velocity u^a (normalized to $u^a u_a = -1$). The redshift of photons emitted from the central world line at $(u, \mathbf{0})$, measured by this observer, is

$$Z := \frac{d\tau}{du} = \frac{1}{u^a \nabla_a u} \tag{77}$$

where we have used our convention that $d\tau/du = 1$ along the central world line.

One natural choice of a family of timelike observers puts them at constant R, y and φ. The ansatz

$$u^a = Z^{-1}\left((\partial_u)^a - \frac{R_{,u}}{R_{,x}}(\partial_x)^a\right) \tag{78}$$

for the corresponding u^a gives $u^a \nabla_a R = u^a \nabla_a y = u^a \nabla_a \varphi = 0$ and $u^a \nabla_a u = Z^{-1}$ as required. From $u_a u^a = -1$ we then find that Z is given by

$$Z = \left(-2g(\Xi R + SbR_{,y}) - R^2 e^{2Sf}Sb^2\right)^{\frac{1}{2}}. \tag{79}$$

3. Hawking mass and Hawking compactness

Following [32], let $\mathcal{S}$ be a smooth closed spacelike 2-surface. Let l^a and n^a be a pair of future-pointing null vectors normal to $\mathcal{S}$, normalized such that $l^a n_a = -1$. Let n^a be outgoing and l_a be ingoing. This is unique up to multiplying n^a by a positive function e^Λ on $\mathcal{S}$ and multiplying l_a by $e^{-\Lambda}$. The projection operator onto the tangent space of $\mathcal{S}$ is

$$\pi_{ab} := g_{ab} + l_a n_b + n_a l_b, \tag{80}$$

and is unique. We then define the null congruence expansions

$$\rho_+ := \frac{1}{2}\pi^{ab}\nabla_a n_b, \qquad \rho_- := \frac{1}{2}\pi^{ab}\nabla_a l_b, \tag{81}$$

It is easy to see that

$$\pi^{ab}\nabla_a(e^\Lambda n_b) = e^\Lambda(\nabla_a n^a + n^b l^a \nabla_a l_b) = e^\Lambda \nabla_a n^a, \tag{82}$$

where the first equality holds when n^a is null and the second when l^a is an affinely parametrized geodesic. A similar result then holds for $e^{-\Lambda}l_a$. Without loss of generality we now define n^a and l^a to be continued off $\mathcal{S}$ as affinely parametrized geodesics. Then the product

$$\rho_+\rho_- = \frac{1}{4}(\nabla_a l^a)(\nabla_b n^b) \tag{83}$$

is independent of the normalization e^Λ, and therefore uniquely determined by the spacetime geometry and $\mathcal{S}$.

From $\rho_+\rho_-$, the Hawking mass M of $\mathcal{S}$ is now defined by

$$M(\mathcal{S}) := \frac{1}{2}\sqrt{\frac{A(\mathcal{S})}{4\pi}}C(\mathcal{S}), \tag{84}$$

$$C(\mathcal{S}) := 1 + \frac{1}{2\pi}\int_{\mathcal{S}}\rho_+\rho_-dS, \tag{85}$$

where $A := \int_{\mathcal{S}}dS$ is the area of $\mathcal{S}$. We have defined the "Hawking compactness" C as an intermediate quantity that is of interest in its own right. In particular, a marginally outer-trapped surface, defined by $\rho_+ = 0$ and $\rho_- < 0$ at each point, has Hawking compactness $C = 1$, and an outer-trapped surface, defined by $\rho_+ < 0$ and $\rho_- < 0$ at each point, has Hawking compactness $C > 1$, but the converses are not true.

We now restrict attention to spacelike 2-surfaces $\mathcal{S}_0$ that lie within a single coordinate null cone. We define such surfaces by

$$u - u_0 = 0, \qquad x - x_0(y) = 0. \tag{86}$$

A basis of tangent vectors to $\mathcal{S}_0$ is given by

$$t_{(y)} := \partial_y + x_0'\partial_x, \qquad t_{(\varphi)} := \partial_\varphi. \tag{87}$$

The outgoing and ingoing null vectors orthogonal to $\mathcal{S}_0$, normalized so that $n^a l_a = -1$ and $l^a\nabla_a u = 1$, are

$$n = U, \tag{88}$$

$$l = \Xi + \frac{e^{-2Sf}SgR_{,x}}{R^2}\left(x_0'\partial_y + \frac{x_0'^2}{2}\partial_x\right). \tag{89}$$

$n^a = U^a$ holds because the outgoing null surface that emanates from $\mathcal{S}_0$ is simply the part of $\mathcal{N}_u^+$ that lies to the future of $\mathcal{S}_0$. If and only if $x_0(y)$ is constant, we also have $l^a = \Xi^a$.

The corresponding null geodesic expansions are

$$R\rho_+ = \frac{1}{g}, \tag{90}$$

$$2R\rho_- = K_0 + K_1 x_0' + K_2 x_0'^2 + K_3 x_0''. \tag{91}$$

The K_i are evaluated at $(u_0, x_0(y), y)$, but do not contain derivatives of $x_0(y)$. Here we only give

$$K_0 = 2\Xi R - R(Sb)_{,y} \tag{92}$$

for later reference.

The induced metric on $\mathcal{S}_0$ is

$$h_{ij} = x_{,i}^\mu x_{,j}^\nu h_{\mu\nu} = x_{,i}^\mu x_{,j}^\nu g_{\mu\nu} = g_{ij}, \tag{93}$$

where $x^i := (y, \varphi)$ are the coordinates on $\mathcal{S}_0$. Its determinant is therefore simply R^4, independent of $x_0(y)$, so the volume element on $\mathcal{S}_0$ is

$$dS = R^2 dy d\varphi. \tag{94}$$

We therefore have

$$\sqrt{\frac{A(\mathcal{S}_0)}{4\pi}} = \left(\frac{1}{2}\int_{-1}^{1}R^2 dy\right)^{1/2} \tag{95}$$

and

$$C(\mathcal{S}_0) = 1 + \frac{1}{2}\int_{-1}^{1}2\rho_+\rho_-R^2 dy. \tag{96}$$

Note that the integral $\frac{1}{2}\int_{-1}^{1}\ldots dy$ is equal to the $l = 0$ component of its integrand, and that $2R^2\rho_+\rho_- = -1$ for light cones in flat spacetime.

After an integration by parts to eliminate the x_0'' term, C becomes

$$C(\mathcal{S}_0) = 1 + \frac{1}{2}\int_{-1}^{1}L(x_0, x_0')dy, \tag{97}$$

where

$$L = L_0 + L_1 x_0' + L_2 x_0'^2, \tag{98}$$

$$L_0 := \frac{K_0}{g} = \frac{2\Xi R - R(Sb)_{,y}}{g}, \tag{99}$$

$$L_1 := \frac{S(e^{-2Sf}R_{,x}(Rg)_{,y} - R^3 b_{,x})}{gR^2}, \tag{100}$$

$$L_2 := \frac{e^{-2Sf}Sg_{,x}R_{,x}}{gR}. \tag{101}$$

If we now further restrict to surfaces $\mathcal{S}_{u,x}$ of constant u and x, (91) simplifies to

$$2R\rho_- = K_0, \tag{102}$$

and (97) simplifies to

$$C(u,x) := C(\mathcal{S}_{u,x}) = 1 + \frac{1}{2}\int_{-1}^{1} L_0 \, dy. \qquad (103)$$

Beyond spherical symmetry, $M(\mathcal{S}_{u,x})$ need not be monotonic in x or non-negative. However, in the lsB class of gauges, one can show that it is nondecreasing with x, and non-negative. To prove this, note that for $R = R(u,x)$, the integral for A is trivial, giving $\sqrt{A/4\pi} = R$. We can pull this factor of R into the integral for $C(u,x)$ to obtain

$$M(u,x) := M(\mathcal{S}_{u,x}) = \frac{1}{4}\int_{-1}^{1} R(1 + 2\rho_+\rho_- R^2)\,dy. \qquad (104)$$

To evaluate $M_{,x}(\mathcal{S}_{u,x})$, one can pull ∂_x into the integral and use integration by parts in y to express it as

$$M(u,x)_{,x} = \frac{R_{,x}}{4}\int_{-1}^{1}\left\{\frac{e^{-2Sf}S}{4g^2}(g_{,y} - R^2 e^{2Sf}\mathcal{D}b)^2\right.$$
$$- 2R^4\rho_+\rho_-\left[S^2(\mathcal{D}f)^2 + 4\pi(\mathcal{D}\psi)^2\right]$$
$$\left. + 4\pi e^{-2Sf}S\psi_{,y}^2\right\}dy. \qquad (105)$$

Hence a sufficient condition for $\mathcal{D}M := M_{,x}/R_{,x}$ to be nonnegative is that $\rho_+\rho_- < 0$. (This result is a special case of Eardley's observation [33] that the Hawking mass increases on a foliation of an outgoing null surface by luminosity distance, as long as $\rho_+\rho_- < 0$ and the matter obeys the dominant energy condition.) From $\mathcal{D}M \geq 0$ and $M = 0$ along the central worldline we then obtain $M \geq 0$.

In lsB gauge, where

$$M(u,x) = \frac{R(u,x)}{2}C(u,x) \qquad (106)$$

we can therefore either compute M from (106) with C given by (97), or by integrating (105) as

$$\tilde{M}(u,x,) := \int_0^x M_{,x}(u,x')\,dx', \qquad (107)$$

with

$$\tilde{C}(u,x) := \frac{2\tilde{M}(u,x)}{R(u,x)} \qquad (108)$$

then defined from $\tilde{M}$. In the continuum, $C = \tilde{C}$ and $M = \tilde{M}$, but their discretizations are very different, to the extent that their agreement is a highly nontrivial test of the accuracy of our discretization.

4. No MOTS on coordinate light cones $u = \mathrm{const}$

A marginally outer-trapped surface (MOTS) in 4-dimensional spacetime is a smooth closed 2-dimensional spacelike surface in the spacetime, such that the generators

of its outgoing null cone have zero divergence. We now ask if a MOTS $\mathcal{S}_0$ can exist that lies in a single null time slice, defined by $\rho_+ = 0$ on the surface $u = u_0$, $x = x_0(y)$.

A first problem with this is that in any gauge where R is specified as free data and evolved, including sdn, gsB and lsB gauge, the Raychaudhuri equation will necessarily involve division by $R_{,x}$. To see this, write it as

$$\left(\ln\frac{G}{R_{,x}}\right)_{,x} = \frac{R}{R_{,x}}\tilde{S}_G, \qquad \tilde{S}_G := S^2 f_{,x}^2 + 4\pi\psi_{,x}^2, \quad (109)$$

where $\tilde{S}_G$ is now regular even when $R_{,x} = 0$. It is now easy to check that if we write this as a first-order ODE in x for G, $g := G/R_{,x}$ or $\rho_+ = R_{,x}/(RG)$, or as a second-order ODE in x for the affine parameter λ with $\lambda_{,x} = G$, these ODEs involve a division by $R_{,x}$, and so become singular where the null expansion ρ_+ changes sign. Where $R_{,x} > 0$, we can absorb it into $\mathcal{D} := d/dR$ (as we have done), but $\mathcal{D}$ is not defined where $R_{,x} = 0$. This division by $R_{,x}$ is purely a gauge problem: in any gauge where we specify G a priori and solve the Raychaudhuri equation for R, for example in affine gauge, $R_{,x}$ and therefore ρ_+ can change sign without a problem.

A second, purely geometrical, problem is that when we trace the future-outgoing null rays that emerge from the MOTS backward in time we generically do not expect them to converge to a point, but rather to form caustics, except in spherical symmetry. In particular this means that this backward null surface cannot in general be a *coordinate* null cone with regular vertex.

We now ask if one can find a *nonmarginally* outer-trapped surface (OTS) on $u = u_0$, one where $\rho_+ \leq 0$ everywhere. The geometric nongenericity argument then does not apply, so we expect to be able to find OTS (but not MOTS) in affine gauge.

But in twist-free vacuum axisymmetry even this is not possible because of a third, again purely geometrical, problem: the shear along the symmetry axis is zero, (as we see from $S = 0$ there), so in vacuum the Raychaudhuri equation on the symmetry axis on the symmetry axis reduces to $g_{,x} = 0$ or $R_{,xx} = 0$, and so $\rho_+ = 1/(gR) = R_{,x}/G$ cannot become zero or change sign. There may of course be OTS and MOTS in such a spacetime, but they cannot be embedded in an outgoing null cone with regular vertex.

5. Surfaces of maximal Hawking compactness

In spite of the obstacles set out in the last subsection, in spherical symmetry we and other authors have happily identified MOTS, and used their Hawking mass as an estimate of the initial black hole mass while working in double-null, Bondi or sdn gauge. How did this work?

Quite naively, we were led by what one does in spherical symmetry on Cauchy surfaces, for example in polar-radial

coordinates (t, r) (see Appendix C): these coordinates become singular on a MOTS, but one simply identifies the first appearance of a local maximum in r of $C(t, r)$ with $C \geq 0.99$, say, as an approximate MOTS, and estimates $M \simeq r/2$ at its location.

Similarly, in null coordinates we called a surface $\mathcal{S}_{u,x}$ an approximate MOTS when it was a local maximum in x of $C(u, x)$ with $C \geq 0.99$, say [29,31]. Unlike $\rho_+(u, x)$, which is often a decreasing function of x (while remaining strictly positive), $C(u, x)$ typically does have one or more local maxima in x. This singled out a specific $\mathcal{S}_{u,x}$ as our MOTS candidate. The same approach was also taken by the authors of [25,27].

How then can we generalize this procedure in spherical symmetry to the nonspherical case? The obvious difference is that we no longer have a preferred foliation of our coordinate null cones into 2-surfaces.

Ideally, we would look for surfaces $x = x_0(y)$ in $u = u_0$ that maximize $C[u_0, x_0(y)]$ given by (97). The resulting Euler-Lagrange equation is a quasilinear second-order ODE for $x_0(y)$. The boundary term usually obtained in deriving the Euler-Lagrange equation vanishes, and so the variation is unconstrained, consistent with our earlier observation following Eq. (27) that a regular scalar need not have vanishing y-derivative on the symmetry axis.

In Paper II, we shall for simplicity limit ourselves to finding the local maxima in x of $C(u, x)$ given by (97) for the compactness of the *coordinate* 2-spheres $\mathcal{S}_{u,x}$, as we did in spherical symmetry. It turns out that with a smaller threshold value, say $C \geq 0.8$, our heuristic criterion consistently distinguishes collapsing and dispersing solutions. Recall, however, our earlier observation following Eq. (85) that beyond spherical symmetry $C > 1$ is necessary but not sufficient for a spacelike 2-surface to be outer-trapped.

6. Event horizons and coordinate null cones

We now show that if a spacetime admits an event horizon, this has at least one generator in common with each of at least two coordinate null cones.

The intersection $\mathcal{H} \cap \Sigma$ of an event horizon $\mathcal{H}$ with a Cauchy surface Σ with topology $\mathbb{R}^3$ (which excludes external black holes with two spatial infinities) at sufficiently late time is a 2-sphere (or consists of disconnected 2-spheres). Hence on $\mathcal{H} \cap \Sigma$ the coordinate u attains a global minimum and global maximum. If the intersection is at least C^2 in our coordinate system, they are also local extrema. At any such local extremum u_*, $\mathcal{H} \cap \Sigma$ and $\{u = u_*\} \cap \Sigma$ have the same (2-dimensional, spacelike) tangent space V. It follows that the unique future outgoing null geodesic through that point and normal to V is a generator of both $\mathcal{H}$ and $u = u_*$.

Independently, in axisymmetry there will be (at least) two local extrema u_* of u on $\mathcal{H} \cap \Sigma$ located at the poles $y = \pm 1$. If the spacetime has an additional $y \to -y$ symmetry (reflection through the equatorial plane), and

$\mathcal{H} \cap \Sigma$ (or one of its components) straddles the equatorial plane, there is a third local extremum on the equator $y = 0$, while the other two are related by the reflection symmetry.

It seems unlikely that we can identify any horizon generators, even if they are also coordinate null cone generators.

IV. NUMERICAL METHODS

A. Legendre pseudospectral method in y

1. Choice of basis functions, and synthesis matrices

In the physical scenarios we are interested in, it is natural to maintain constant angular resolution, and to expect the solution to be smooth. We therefore use a pseudospectral method in the angle θ, or y. In this subsection, we write N for N_y, the number of grid points in y.

We represent ψ as

$$\psi = \sum_{l=0}^{N-1} \psi_l(u, x) P_l(y), \tag{110}$$

where P_l is the Legendre polynomial of order l, proportional to the spherical harmonic Y_{l0}. We represent any other quantity that transforms as a scalar under coordinate changes on the 2-spheres $\mathcal{S}_{u,x}$, including the metric components R, G and H, in the same way.

By contrast, b and f are not scalars but components of a vector and symmetric 2-tensor on S^2, respectively. We show in Appendix D that if we represent them as

$$b = \sum_{l=1}^{N} b_l(u, x) P'_l(y), \tag{111}$$

$$f = \sum_{l=2}^{N+1} f_l(u, x) P''_l(y), \tag{112}$$

then in the linearization of the Einstein and wave equations about spherical symmetry the different l decouple. We choose this spectral representation for b, f and the scalars also in the nonlinear case.

In any pseudospectral method, one goes backward and forward between a finite number of coefficients in Fourier space and an equal number of carefully chosen collocation points in real space, carrying out differentiation in Fourier space and nonlinear algebraic operations in real space.

Here we transform between collocation points y_i and spherical harmonic components l by full matrix multiplication. This is clearly inefficient for large N, in contrast to the fast Fourier transform available for a Fourier series or Chebyshev polynomials.

We call the matrices that take variables from Fourier space (with index l) to real space (with index i) synthesis matrices. For all variables that transform as a scalar under

coordinate changes on the coordinate 2-spheres, we use the synthesis matrix $S_{il}^{(0)} := P_l(y_i)$, where y_i are the collocation points (to be determined below) and l is the spectral index. In other words, each *column* of $S^{(0)}$ represents one P_l. For b we use $S_{il}^{(1)} := P'_l(y_i)$, and for f we use $S_{il}^{(2)} := P''_l(y_i)$. Note this means that with $i = 1, …N$, we can represent $l = 0, …N − 1$ for scalars, $l = 1, …N$ for b and $l = 2, …N + 1$ for f.

2. Differentiation in y and choice of collocation points

Rather than transforming back to Fourier space in order to differentiate in y, we differentiate directly in real space. Either method requires full-matrix multiplication. Making differentiation exact for certain functions of y then fixes the collocation points.

We implement first y-derivatives through the Legendre-Gauss-Lobatto differentiation matrix [34]

$$D_{ij} = \begin{cases} \frac{P_{N-1}(y_i)}{(y_i-y_j)P_{N-1}(y_j)}, & i \neq j, \\ -\frac{N(N-1)}{4}, & i = j = 1, \\ \frac{N(N-1)}{4}, & i = j = N, \\ 0, & \text{otherwise}, \end{cases} \quad (113)$$

where the grid points $y_2, …, y_{N-1}$ are the zeros of $P'_{N-1}(y)$ in increasing order, $y_1 = -1$ and $y_N = 1$. For second derivatives, we use the matrix square of D. This discretization is known to be exact for polynomials up to order $2N − 3$. We choose Legendre-Gauss-Lobatto because we want the north and south poles $y = \pm 1$ to be on the grid. In order to also have the equator on the grid, we then choose N to be odd, typically $2^K + 1$.

3. Analysis matrices

An analysis matrix takes grid functions from physical to Fourier space. One might compute the analysis matrix $A_{li}^{(0)}$ using the formulas for Legendre-Gauss-Lobatto quadrature and the fact that $\int P_m P_n dy = (2n + 1)\delta_{mn}/2$. However, with $A^{(0)}$ defined this way, the product $A^{(0)} S^{(0)}$ differs from the expected unit matrix in that its bottom right element is $2 + 1/(N − 1)$. This is due to the fact that Legendre-Gauss-Lobatto quadrature is exact for polynomials in y up to order $2N − 3$, whereas this bottom right element requires integration of $P_{N-1}P_{N-1}$, which is a polynomial of order $2N − 2$.

Therefore we use as our analysis matrix $A_{li}^{(0)}$ the matrix inverse of $S_{il}^{(0)}$, which differs from the naive analysis matrix only in its last row. (Having to make this choice could have been avoided by using Gauss-Legendre quadrature, which is exact for polynomials up to order $2N − 1$.) We similarly define $A_{li}^{(1)}$ and $A_{li}^{(2)}$ as the matrix inverses of $S_{il}^{(1)}$ and $S_{il}^{(2)}$.

4. Discrete versions of continuum identities and consistent truncation

We define the $N \times N$ matrices

$$\Delta^{(s)} := (1 − Y^2)D^2 − 2(s + 1)YD, \quad (114)$$

where Y denotes the diagonal matrix

$$Y := \text{diag}\{y_i\}_{i=1}^N, \quad (115)$$

and we have defined the diagonal matrices

$$\Lambda^{(s)} := \text{diag}\{\lambda_l^{(s)}\}_{l=s}^{N-1+s}, \quad (116)$$

where

$$\lambda_l^{(s)} := -(l + s + 1)(l − s), \quad l \geq s, \quad (117)$$

are the eigenvalues of the Laplace operator on S^2 for the "spins" $s = 0, 1, 2$.

Our synthesis and differentiation matrices obey discrete equivalents of the continuum identities (D15), (D33) and (D38) between Legendre polynomials given in Appendix D, which in this notation can be written concisely as

$$\Delta^{(s)} S^{(s)} = S^{(s)}\Lambda^{(s)}, \quad (118)$$

for $s = 0, 1, 2$. We also have

$$DS^{(0)} = S^{(1)}I_+, \quad (119)$$

$$DS^{(1)} = S^{(2)}I_+, \quad (120)$$

$$\Rightarrow D^2 S^{(0)} = S^{(2)}I_+^2, \quad (121)$$

relating the different spins. Here I_+ denotes the matrix with ones in the super-diagonal. We also define I_- as the matrix with ones in the sub-diagonal, and I_0 as the matrix with ones in the diagonal, except for a zero in the bottom right element. These obey $I_+I_- = I_0$. In particular, Eq. (120) is a discrete version of the identity (D29), which we need for the linearized hierarchy equation for f, Eq. (D26), to be discretized exactly in y.

With the definition

$$\not{\Delta}^{(1)} := (1 − Y^2)D − 4Y \Rightarrow \not{\Delta}^{(1)}D = \Delta^{(1)}, \quad (122)$$

we can derive the identity

$$\not{\Delta}^{(1)} S^{(2)}I_0 = S^{(1)}\Lambda^{(1)}I_-. \quad (123)$$

Except for the presence of the factor I_0 on the left, (123) is the discrete version of the identity (D30), which is needed for the linearized hierarchy equation for b, Eq. (D25). Without the right factor of I_0, the left-hand side of (123) would have nonzero entries for $l = N − 1$ and $N − 3$ in its

last column. With N grid points we can represent f_{N+1} but not b_{N+1}, so one can think of this as an aliasing error arising from the spectral truncation. To avoid this error, which would lead to an instability, we need to suppress the unpartnered f_{N+1}.

In the linearized field equations on a general spherically symmetric background, or in a gauge other than Bondi gauge, f_l and b_l couple not only to each other but also to ψ_l, G_l, R_l and H_l. Therefore we must suppress f_{N+1}, f_N and b_N, as all three have no scalar partners.

5. Half-range spectral method

Most papers on axisymmetric gravitational collapse assume an additional reflection symmetry, $z \to -z$ in cylindrical coordinates, or $y \to -y$ in our spherical coordinates. In such situations, we can save computing time by only representing the even l in Fourier space, or the values $-1 \leq y \leq 0$ in real space. Let $\bar{N} := (N+1)/2$ denote the number of grid points on the half-range equivalent to N grid points on the full range. We define reduced analysis and synthesis matrices as

$$\bar{A}^{(0,2)} := Q_+ A^{(0,2)} V_+^T, \tag{124}$$

$$\bar{S}^{(0,2)} := X S^{(0,2)} Q_+^T, \tag{125}$$

$$\bar{A}^{(1)} := Q_- A^{(1)} V_-^T, \tag{126}$$

$$\bar{S}^{(1)} := X S^{(1)} Q_-^T, \tag{127}$$

where we have defined (with $\bar{N} = 3 \Leftrightarrow N = 5$ serving as a prototype)

$$Q_+ := \begin{pmatrix} 1 & 0 & & \\ & 0 & 1 & 0 \\ & & & 0 & 1 \end{pmatrix}, \tag{128}$$

$$Q_- := \begin{pmatrix} 0 & 1 & 0 & & \\ & & 0 & 1 & 0 \\ & & & & 0 \end{pmatrix}, \tag{129}$$

$$X := \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & 0 & 0 \end{pmatrix}, \tag{130}$$

$$V_+ := \begin{pmatrix} 1 & & & 1 \\ & 1 & & 1 \\ & & 1 & \end{pmatrix}, \tag{131}$$

$$V_- := \begin{pmatrix} 1 & & & -1 \\ & 1 & & -1 \\ & & 0 & \end{pmatrix}. \tag{132}$$

Note that $\bar{S}$ with $N = 5$ or $\bar{N} = 3$ is a 3×3 matrix, etc. These obey

$$\bar{A}^{(0,2)} \bar{S}^{(0,2)} = I, \tag{133}$$

$$\bar{A}^{(1)} \bar{S}^{(1)} = I_0. \tag{134}$$

To understand this, consider again the case $\bar{N} = 3$: we can represent P_l for $l = 0, 2, 4$, and $P_l^{(2)}$ for $l = 2, 4, 6$, but $P_l^{(1)}$ only for $l = 2, 4$. Q_- has row rank $\bar{N} - 1$, and therefore $S^{(1)}$ and $A^{(1)}$ have rank $\bar{N} - 1$.

We also define separate derivative matrices acting on even and odd functions,

$$D_\pm := X D V_\pm^T, \tag{135}$$

$$D_\pm^2 := D_\mp D_\pm. \tag{136}$$

We have then the same identities as already discussed, again with the proviso that $P_{2\bar{N}}^{(2)} = P_{N+1}^{(2)}$ can be represented but must be suppressed for consistency.

6. Tests at finite numerical precision

We numerically calculate the synthesis, analysis and differentiation matrices for selected values of N in Mathematica. The collocation points (zeros of the Legendre polynomials) need to be determined numerically, and we do this with precision 10^{-50}. The matrix calculations are then carried out with the same precision. We then save the resulting expressions for the collocation points y_i and the synthesis, analysis and differentiation matrices result into ascii files in slightly more than double precision, and read this into the F90 code at runtime.

As as test of the numerical error of the pseudospectral method, we have explicitly evaluated the following matrices, which all vanish in the continuum, numerically in the F90 code (in double precision).

$$T_0 := A^{(0)} S^{(0)} - I, \tag{137}$$

$$T_1 := A^{(1)} S^{(1)} - \begin{cases} I \\ I_0 \end{cases}, \tag{138}$$

$$T_2 := A^{(2)} S^{(2)} - I, \tag{139}$$

$$T_3 := \Delta_+^{(0)} S^{(0)} - S^{(0)} \Lambda^{(0)}, \tag{140}$$

$$T_4 := \Delta_-^{(1)} S^{(1)} - S^{(1)} \Lambda^{(1)}, \tag{141}$$

$$T_5 := \Delta_+^{(2)} S^{(2)} - S^{(2)} \Lambda^{(2)}, \tag{142}$$

$$T_6 := D_+ S^{(0)} - S^{(1)} I_+, \tag{143}$$

$$T_7 := D_- S^{(1)} - S^{(2)} \begin{cases} I_+, \\ I_0 \end{cases} \tag{144}$$

$$T_8 := D_+^2 S^{(0)} - S^{(2)} \begin{cases} I_+^2, \\ I_+ \end{cases} \tag{145}$$

$$T_9 := \slashed{\Delta}_+^{(1)} S^{(2)} \begin{cases} I_+ \\ I_0 \end{cases} - S^{(1)} \Lambda^{(1)}. \tag{146}$$

The matrices A, S, Δ and Λ are understood as either half-range or full-range. Where there is a case distinction, the upper case applies to the full-range setup and the lower case to the half-range setup. In the full-range case $D_+ = D_- := D$, and similarly for Δ and $\slashed{\Delta}$.

To avoid duplication of code, in the numerical code we use the half-range notation throughout, and simply set both D_+ and D_- to D in the full-range case, and similarly D_+^2 and D_-^2 to D^2. By contrast, the combinations Δ and $\slashed{\Delta}$ do not explicitly appear in the nonlinear equations and so are not stored as matrices, but are used here as shorthand for their definitions.

We also check that each row of D or D_+ adds up to zero, as the differencing of a constant grid function should give zero. However, no such test applies to D_-, which acts only on odd functions of y. Hence we define another test

$$T_{10} := D_+ \mathbf{1}, \tag{147}$$

where $\mathbf{1}$ is the column vector with 1 in every row.

At $N = 3$ or $\bar{N} = 2$, the error in all tests is zero or at machine precision. At all higher resolutions, the largest error occurs in T_5. All errors at the two equivalent resolutions N and $\bar{N} = (N + 1)/2$ are approximately the same, as we would expect. At $N = 5$, $\bar{N} = 3$, this error is $\simeq 10^{-13}$. At each doubling of resolution, it increases by a factor of ~ 100, up to 4×10^{-5} at $N = 65$, $\bar{N} = 33$, and 8×10^{-3} at $N = 129$, $\bar{N} = 65$.

We believe that the observed error in these tests is essentially round-off error in double precision computations in the F90 code, and that it increases so rapidly with N because the synthesis and analysis matrices become increasingly ill-conditioned with both resolution N and spin s.

We note that while the error in T_0 at $N = 65$, evaluated within *Mathematica*, is 10^{-49} as expected, in T_1 and T_2 it is already 10^{-23}. We can avoid this by noting that P'_l is a linear combination with integer coefficients of P_{l-1}, P_{l-3} and so on. Hence we can write $S^{(1)} = S^{(0)} T^{(01)}$ and $S^{(2)} = S^{(0)} T^{(02)}$, where $T^{(01)}$ and $T^{(02)}$ are lower diagonal matrices with integer coefficients, and so can be inverted exactly. T_1 and T_2 then have a much smaller internal error of 10^{-44}. The maximum difference between $A^{(1)}$ or $A^{(2)}$ obtained in the two ways is still only 10^{-24} at $N = 65$,

so they are identical when reduced to double precision in Fortran code. However, for $N = 129$, *Mathematica* cannot find $A^{(1)}$ and $A^{(2)}$ by direct matrix inversion, but we can still find them from $A^{(0)}$ and the inverses of $T^{(01)}$ and $T^{(02)}$.

7. High-frequency filtering

As in any pseudospectral code, nonlinear terms spuriously excite high-l modes through aliasing, and without care this eventually leads to numerical instabilities, even if the linearized code is stable. For high-frequency filtering of a grid function in real space, we therefore multiply by $F := S \hat{F} A$, where $\hat{F}$ is a diagonal matrix where only the entries corresponding to $l \leq l_{\max}$ diagonal entries are one, and those for $l > l_{\max}$ are zero. (A smooth transition from one to zero would also be possible, but we have not tried this).

When the Einstein equations are linearized around any spherically symmetric solution, ψ_l, f_l, b_l, R_l and γ_l couple only for the same l. This suggests that for consistency we truncate at the same $l_{\max}$ for all variables, rather than the same degree of polynomial in y, using the appropriate analysis and synthesis matrices. Define $\hat{F}(k)$ to be the diagonal matrix with 1 in its first $0 \leq k \leq N$ entries, and 0 in the remaining ones. We then filter the scalars with $\bar{F}^{(0)} := S^{(0)} \hat{F}(l_{\max} + 1) A^{(0)}$, the variable b with $\bar{F}^{(1)} := S^{(1)} \hat{F}(l_{\max}) A^{(1)}$, and the variable f with $\bar{F}^{(2)} := S^{(2)} \hat{F}(l_{\max} - 1) A^{(2)}$. These matrices have therefore different rank.

B. Finite differencing in x of the hierarchy equations

In this subsection, we write N for N_x. As we have seen, the hierarchy equations take the form

$$\phi_{\mathrm{int},x} = S(y, \phi, \phi_{,x}, \phi_{,y}, \phi_{,xy}, \phi_{,yy}) \tag{148}$$

in terms of the intermediate quantities

$$\phi_{\mathrm{int}} := \left(\gamma, R^4 e^{2Sf - \gamma} \mathcal{D}b, b, R\Xi R, R\Xi f, R\Xi\psi \right), \tag{149}$$

and the basic variables (metric coefficients and matter field)

$$\phi := (\gamma, R, b, f, \psi). \tag{150}$$

These equations can be solved by integration as

$$\phi_{\mathrm{int}}(u, x, y) = \phi_{\mathrm{int}}(u, 0, y) + \int_0^x S(u, x', y) dx', \tag{151}$$

for the intermediate quantities, and hence the constrained variables

$$\phi_{\mathrm{cons}} := (\gamma, b, \Xi R, \Xi f, \Xi \psi), \tag{152}$$

in this order. The evolved variables

$$\phi_{\mathrm{evol}} := (f, R, \psi) \tag{153}$$

are specified freely at $u = 0$ and evolved in u using their Ξ-derivatives.

In double-null gauge no influence propagates from larger to smaller x. In a general gauge, this happens only through the shift term $B\partial_x$. This suggests to us that the hierarchy equations should be solved by an integration scheme in x that does not evaluate the integrand to the right of the point where the integral is being approximated. Then for $H \leq 0$, strictly no information flows to larger x. In the following, we present a simple second-order accurate scheme that has these properties.

We use a grid equally spaced in x, $x_i = i\Delta x$, with $x_0 = 0$ and $x_N = x_{\mathrm{max}}$, so $\Delta x = x_{\mathrm{max}}/N$. In the code, array storage for fields at the mid-point $x_{i+1/2}$ is labeled by the array index i, that is by the grid point to its left. We do not store field values at the first grid point $x_0 = 0$ and first mid-point $x_{1/2} = \Delta x/2$. Therefore arrays representing grid points have array index ranging from 1 to N, and arrays representing mid-points from 1 to $N - 1$, corresponding to $i = 3/2$ to $N - 1/2$. With this convention in mind, we define the shorthand

$$\Delta_i\phi := \phi_{i+1} - \phi_i, \tag{154}$$

$$\bar{\phi}_i := \frac{\phi_{i+1} + \phi_i}{2}. \tag{155}$$

Here we write only the index i corresponding to the coordinate x, but not the index j corresponding to the coordinate y.

For sufficiently smooth functions ϕ we then have

$$(\phi_{,x})_{i+1/2} = \frac{\Delta_i\phi}{\Delta x} + O(\Delta x^2), \tag{156}$$

$$\phi_{i+1/2} = \bar{\phi}_i + O(\Delta x^2). \tag{157}$$

We then discretize the integrations (151) with the midpoint rule, that is

$$\phi_{\mathrm{int},i+1} = \phi_{\mathrm{int},i} + S_{i+\frac{1}{2}}\Delta x, \tag{158}$$

using the discretizations (156), (157), the discretizations of $\phi_{,y}$ and $\phi_{,yy}$ using D and D^2, and the discretization of $\phi_{,xy}$ and $\phi_{,xyy}$ resulting from combining (156) with D. The resulting $\phi_{\mathrm{int},i+1}$ is second-order accurate but depends only on ϕ_{i+1} and ϕ_i, thus respecting causality. By contrast, the trapezoid rule would require left derivatives at the right-hand gridpoint to respect causality.

In the formulation where we use $\mathcal{D}$ defined in (40) rather than ∂_x, we differentiate

$$(\mathcal{D}\phi)_{i+1/2} \simeq \frac{(\phi_{,x})_{i+1/2}}{(R_{,x})_{i+1/2}} = \frac{\Delta_i\phi}{\Delta_i R} \tag{159}$$

and integrate

$$\phi_{\mathrm{int},i+1} = \phi_{\mathrm{int},i} + \bar{S}_{i+\frac{1}{2}}\Delta_i R. \tag{160}$$

C. Shift terms

For the evolved variables (153), the definition of Ξ gives

$$\phi_{\mathrm{evol},u} = \Xi\phi_{\mathrm{evol}} + B\phi_{\mathrm{evol},x} + Sb\phi_{\mathrm{evol},y} \tag{161}$$

The analogy of B with the x-component of a shift vector suggests that, in contrast to the x-derivatives in the right-hand side of (148), $\phi_{\mathrm{evol},x}$ in (161) should be upwinded: as right derivatives $\phi_{,x}^+$ for $B > 0$, and left derivatives $\phi_{,x}^-$ for $B < 0$. In particular, no numerical boundary condition is then required at an outer boundary $x = x_{\mathrm{max}}$ as long as $B \leq 0$ there, or at the inner boundary where $B = B_{\mathrm{Bondi}} > 0$. We do not upwind the shift term in the y direction, as we take all y-derivatives spectrally.

The upwind x-derivatives are evaluated with the second-order accurate three-point formulas on grid points, namely

$$(\phi_{,x})_i^+ := \frac{2\phi_{i+1} - \frac{3}{2}\phi_i - \frac{1}{2}\phi_{i+2}}{\Delta x}, \tag{162}$$

$$(\phi_{,x})_i^- := -\frac{2\phi_{i-1} - \frac{3}{2}\phi_i - \frac{1}{2}\phi_{i-2}}{\Delta x}. \tag{163}$$

Where $R_{,x}(u, \mathbf{0})$ is needed, we use $R_{0,j} = 0$ in (162) to evaluate the right derivative as

$$(R_{,x})_{0,j}^+ = \frac{2R_{1,j} - \frac{1}{2}R_{2,j}}{\Delta x}, \tag{164}$$

and then average over y to obtain $R_{,x}(u, \mathbf{0})$.

D. The time step problem, and a resolution

An important, if somewhat vaguely defined, necessary condition for numerical stability of any explicit time-evolution scheme is the Courant-Friedrichs-Lewy (from now, CFL) condition. This says that the outermost characteristic cone, in our case the light cone, is contained in the spacetime numerical stencil. We consider this as a heuristic guide to a stability limit on the time step, without carrying out an actual discrete stability analysis. Unusually, in spherical symmetry, in double-null coordinates (u, v), this causality condition does not impose any restriction on the time step Δu. However, and not surprisingly, we found in [31] that even then a limit $\Delta u \sim \Delta x$ is required for stability. Beyond spherical symmetry, however, the combination of spherical polar coordinates and null coordinates imposes a

severe restriction, and we need to address this problem in order to make our code efficient.

Explicit methods for hyperbolic problems using spacelike time slices and Cartesian coordinates typically have a time step condition $\Delta t \sim \Delta x$, where Δx is the grid spacing of the Cartesian spatial coordinates. By contrast, polar spatial coordinates on spacelike slices give rise to a time step condition of $\Delta t \sim \Delta r \Delta \theta$, which comes from evaluating the CFL condition in the tangential direction, $\Delta t \sim r \Delta \theta$, near the center, that is at $r \sim \Delta r$. In Appendix A, we show that in polar coordinates on null cones, the stability criterion is an even worse $\Delta u \sim \Delta r \Delta \theta^2$.

This is a problem not only when we use finite differencing in the angle θ. When we split the linearized wave equation into spherical harmonics as in (110) and finite-difference in r for given l, we find empirically that the time step limit for evolving $\psi_l(u, r)$ is $\Delta u \sim l^{-2} \Delta r$. There is no differentiation in y to give rise to a CFL condition in the tangential direction, but instead the wave equation decomposed into spherical harmonics now contains an $l(l+1)/r^2$ potential. It turns out that this requires the same restriction on the time stesp as if we had used finite differencing in θ, with $\Delta \theta \sim 1/l_{\max}$.

In the context of the wave equation on Minkowski spacetime, the $l(l+1)/r^2$ barrier can be transformed into a $2(l+1)/r$ barrier for a first-order reduction of the wave equation, and this was made stable for $\Delta t \sim \Delta r$ by the introduction of a suitable summation by parts (SBP) differencing scheme in r [35]. We could not see how to do this in null coordinates, even for the flat-space scalar wave equation.

However, the spectral method in y that we use suggests a possible remedy to the time step restrictions in polar coordinates, both on null slices and on the usual spacelike slices: we simply filter out all spatial frequencies above $l \sim i$, where l is the spherical harmonic index, and i the grid index in the radial coordinate x. In other words, at radius x only spherical harmonics up to $l \sim x/(\Delta x)$ are represented. We discuss how this is done in the code in the next subsection.

An intuitive understanding of why this boundary condition removes the extra restrictions on the time step due to spherical polar coordinates and null coordinates is that, with $\Delta \theta \sim 1/l_{\max} \sim 1/i \sim \Delta x/x$ near the center, we now have an effective angular resolution such that $x \Delta \theta \simeq \Delta x$, so that the effective "grid cells" have roughly equal sides, giving us the stability benefit of a Cartesian grid.

Outside the central region, for $i \gtrsim l_{\max}$, the angular resolution is constant, giving us the physical benefits of a spherical grid, namely efficient resolution of ingoing and outgoing waves of finite l and an outer boundary with spherical topology.

A similar filtering approach to overcoming the stricter CFL limit in spherical polar coordinates has been presented on spacelike time slices in [36]. Here the filtering uses fast Fourier transforms in θ and φ.

These adjustments allow us to run the axisymmetric code with the time step

$$\Delta u = \min_{i,j} \min(C_1 \Delta u_1, C_2 \Delta u_2), \qquad (165)$$

where we have defined the local time step criteria

$$\Delta u_1(u, x, y) := \Delta x \left| \frac{R_{,x}}{\Xi R} \right|, \qquad (166)$$

$$\Delta u_2(u, x, y) := \Delta x |B|^{-1}. \qquad (167)$$

The parameters C_1 and C_2 are independent of Δx and Δy (or $l_{\max}$), and are of order one. As B is the shift in the x-direction, $\Delta u \leq \Delta u_2$ is the standard limit on the time step for any explicit finite differencing of an advection equation. In double-null gauge, $B = 0$, and so this criterion is empty, but empirically the limit $\Delta u \leq \Delta u_1$ on the time step is required even then. The quantity $|R_{,x}/\Xi R|$ can be understood in two ways: as $|R_{,x}|/|R_{,u}|$ in double-null gauge, implying that R always changes less per time step than per grid point, or as the value taken by B in Bondi gauge. Empirically, we find that this term guarantees stability in double null and Bondi gauges and their variants. Because of our gradual suppression of high angular frequencies near the center this holds independently of $l_{\max}$.

E. Treatment of the central region

1. High-frequency filtering after each time step

In order to allow for a time step $\Delta u \sim \Delta x$ in the way just discussed, in the initial data and after each full time step we apply a filter to the evolved variables R, f and ψ as discussed in Sec. IV A 7, but with a local $l_{\max}$. This sets

$$\phi_{\mathrm{evol},l} = 0 \quad \text{for } l > l_{\max,\mathrm{local}}(i) \qquad (168)$$

where

$$l_{\max,\mathrm{local}}(i) := \min\big(\max(2, 2i - 2), l_{\max,\mathrm{global}}\big), \qquad (169)$$

which equals 2, 2, 4, 6,… for $i = 1, 2, 3, 4…$ This filtering means that at $i = 1, 2$, only $l = 0, 1, 2$ are present. At $i = 3$, $l = 0, 1, 2, 3, 4$ are present, at $i = 4$, $l = 0…6$, and so on. $l_{\max,\mathrm{global}}$ is even to treat the full and half-range discretization equally. We use the relevant analysis and synthesis matrices for the scalars, b and f, respectively.

The filter always removes at least the top two l-modes of f on the full-range grid, or the top mode on the half-range grid, as these do not have a counterpart in γ, H, R, ψ and b. This means that $l_{\max,\mathrm{global}} \leq N_y - 1$ on the full y-range and $2(\bar{N}_y - 1)$ on the half-range. (We always choose N_y and $\bar{N}_y$ to be odd.)

2. Expansion of the field equations about the origin

The filtering near the origin that gets round the bad CFL condition de facto imposes unphysical numerical boundary conditions, for example $\psi_l(u, x_l) = 0$, at some $x_l \sim l\Delta x$. This is compatible with second (or higher) order numerical accuracy in Δx for large l, but not for the smallest l. For example, in a regular continuum solution $\psi_l(u, x) \sim x^l$ near the center, so imposing $\psi_l(u, x_l) = 0$ imposes an error of $O(\Delta x^l)$. If we want the code to be second-order accurate, this is acceptable for $l \geq 2$, but for $l = 0$ and $l = 1$ we need to impose a more accurate, nonzero, boundary condition.

In the code, we impose boundary conditions at an inner boundary (including an unphysical one) as integration constants when we solve the hierarchy equations by integration in x, for example the constant c_l in $(R\Xi\psi)_l(u, x) = c_l + \int_{x_l}^{x} \ldots dx'$. We now obtain the values of those integration constants that correspond to a regular center (to a given order of accuracy) by expanding the full hierarchy equations in powers of x.

We shall see that for second-order accuracy we need nonzero integration constants only for the spherical harmonic components $b_{1,2}$, $(\Xi R)_{0,1,2}$, $(\Xi f)_2$ and $(\Xi\psi)_{0,1,2}$ (where the suffix denotes the value of l). The general argument above applies also to $\gamma_{0,1,2}$ but their integration constants turn out to be zero.

We assume that in a regular axisymmetric solution of the Einstein-scalar equations, the scalar field $\psi(u, x, y)$ admits a convergent expansion around the origin $x = 0$ of the form

$$\psi = \sum_{l=0}^{\infty} \sum_{k=l}^{\infty} \psi_{l(k)}(u) x^k P_l(y) \tag{170}$$

$$= \sum_{k=0}^{\infty} \sum_{l=0}^{k} \ldots, \tag{171}$$

where the re-ordering obviously requires convergence. The first suffix in $\psi_{l(k)}$ denotes the spherical harmonic and the second one the power of x. Appendix F shows that this [together with analyticity of the $\psi_{l(k)}(u)$] corresponds to analyticity in suitable Cartesian coordinates (t, ξ, η, z).

Expanding any hierarchy equation $F = 0$ as (171), and truncating this expansion as

$$F \simeq \sum_{k=0}^{k_{\max}} \sum_{l=0}^{k} F_{l(k)}(u) x^k P_l(y) \tag{172}$$

the result is a polynomial of finite order $k_{\max}$ in both x and y. This observation guarantees that when, order by order in x, we set the coefficients of all nonvanishing powers of y to zero separately to obtain a system of algebraic equations for the $\psi_{l(k)}$, the expansion remains exact in y.

A priori, we expand γ and R in the same way as ψ. However, we impose $\gamma(u, \mathbf{0}) = \gamma_{0(0)} = 0$ in order to make u proper time at the origin, and so

$$\gamma = \sum_{k=1}^{\infty} \sum_{l=0}^{k} \gamma_{l(k)}(u) x^k P_l(y). \tag{173}$$

We additionally impose $R(u, \mathbf{0}) = R_{0(0)} = 0$ to locate the origin $R = 0$ at $x = 0$ and $R_{1(1)} = 0$ to make $R_{,x}$ single-valued at the origin. Moreover, we show in Appendix F that regularity requires $R_{l(l)} = 0$ for all l, so that

$$R = \sum_{k=1}^{\infty} \sum_{l=0}^{k-1} R_{l(k)}(u) x^k P_l(y). \tag{174}$$

Generalizing from the behavior of analytic solutions to the Einstein equations linearized about Minkowski spacetime (see Appendix D 5), and consistent with the regularity requirements in Appendix F we expand

$$f = \sum_{k=2}^{\infty} \sum_{l=2}^{k} f_{l(k)}(u) x^k P_l''(y), \tag{175}$$

$$b = \sum_{k=0}^{\infty} \sum_{l=1}^{k+1} b_{l(k)}(u) x^k P_l'(y). \tag{176}$$

Here $b_{1(0)}(u)$ is free, corresponding to a gauge choice, see Appendix E. Finally, from $\Xi = \partial_u - B\partial_x + \ldots$ we expect that the expansions of Ξ-derivatives start at one power of x lower, that is

$$\Xi R = \sum_{k=0}^{\infty} \sum_{l=0}^{k+1} (\Xi R)_{l(k)}(u) x^k P_l(y), \tag{177}$$

$$\Xi f = \sum_{k=1}^{\infty} \sum_{l=2}^{k+1} (\Xi f)_{l(k)}(u) x^k P_l''(y), \tag{178}$$

$$\Xi\psi = \sum_{k=0}^{\infty} \sum_{l=0}^{k+1} (\Xi\psi)_{l(k)}(u) x^k P_l(y). \tag{179}$$

As a test of consistency, we have explicitly expanded all fields to $O(x^5)$. This allows us to consistently expand the hierarchy equation for γ to $O(x^2)$, for b to $O(x^5)$ and for ΞR, Ξf and $\Xi\psi$ to $O(x^3)$, and the resulting coefficient equations can be solved for $(\gamma, b, \Xi R, \Xi f, \Xi\psi)$ to $O(x^3)$.

In the code, we only need the expansions to $O(x)$, as the error of $O(x^2)$ corresponds to $O(\Delta x)^2$ for the innermost few grid points. The nonvanishing terms then involve only spherical harmonics up to $l = 2$, and are

$$\gamma = O(x^2), \tag{180}$$

$$b = \left[b_{1(0)} + \left(-4\pi R_{0(1)}^{-1} \psi_{0(1)} \psi_{1(1)} - R_{0(1)}^{-2} R_{1(3)} \right. \right.$$
$$\left. + 2 R_{0(1)}^{-3} R_{0(2)} R_{1(2)} \right) x \right] P_1'(y)$$
$$+ \left[-4 R_{0(1)}^{-1} \left(f_{2(2)} + \frac{\pi}{3} \psi_{1(1)}^2 \right) - R_{0(1)}^{-2} R_{2(3)} \right.$$
$$\left. + \frac{2}{3} R_{0(1)}^{-3} R_{1(2)}^2 \right] x P_2'(y) + O(x^2), \tag{181}$$

$$\Xi R = -\frac{1}{2} P_0(y) + \left(-R_{0(1)} b_{1(0)} - R_{0(1)}^{-1} R_{1(2)} \right) x P_1(y)$$
$$+ O(x^2), \tag{182}$$

$$\Xi f = -R_{0(1)}^{-1} f_{2(2)} x P_2''(y) + O(x^2), \tag{183}$$

$$\Xi \psi = \left[\frac{1}{2} R_{0(1)}^{-1} \psi_{0(1)} + \frac{1}{2} \left(R_{0(1)}^{-1} \psi_{0(2)} + R_{0(1)}^{-2} \left(R_{1(2)} \psi_{1(1)} \right. \right. \right.$$
$$\left. \left. \left. - R_{0(2)} \psi_{0(1)} \right) \right) x \right] P_0(y) - \frac{1}{2} R_{0(1)}^{-1} \psi_{1(1)} P_1(y)$$
$$- R_{0(1)}^{-1} \psi_{2(2)} x P_2(y) + O(x^2). \tag{184}$$

Note that $P_0(y) = 1$, $P_1(y) = y$, $P_2(y) = (3y^2 - 1)/2$, $P_1'(y) = 1$, $P_2'(y) = 3y$, $P_2''(y) = 3$ but we have not substituted these values for clarity of exposition.

Assuming both $b_{10} = 0$ (the origin is geodesic) and $R_{,y} = 0$ (lsB gauge), as we do here and in Paper II, these equations simplify to

$$\gamma = O(x^2), \tag{185}$$

$$b = -4\pi R_{0(1)}^{-1} \psi_{0(1)} \psi_{1(1)} x P_1'(y)$$
$$- 4 R_{0(1)}^{-1} \left(f_{2(2)} + \frac{\pi}{3} \psi_{1(1)}^2 \right) x P_2'(y) + O(x^2), \tag{186}$$

$$\Xi R = -\frac{1}{2} P_0(y) + O(x^2), \tag{187}$$

$$\Xi f = -R_{0(1)}^{-1} f_{2(2)} x P_2''(y) + O(x^2), \tag{188}$$

$$\Xi \psi = \left[\frac{1}{2} R_{0(1)}^{-1} \psi_{0(1)} + \frac{1}{2} \left(R_{0(1)}^{-1} \psi_{0(2)} \right. \right.$$
$$\left. \left. - R_{0(1)}^{-2} R_{0(2)} \psi_{0(1)} \right) x \right] P_0(y) - \frac{1}{2} R_{0(1)}^{-1} \psi_{1(1)} P_1(y)$$
$$- R_{0(1)}^{-1} \psi_{2(2)} x P_2(y) + O(x^2). \tag{189}$$

This means that we need to fit only the following coefficients from the evolved variables R, f and ψ: $R_{0(1)}$, $R_{0(2)}$, $f_{2(2)}$, $\psi_{0(0)}$, $\psi_{0(1)}$, $\psi_{0(2)}$, $\psi_{1(1)}$ and $\psi_{2(2)}$. $\psi_{0(0)}$ must be fitted for consistency but is not used.

We have implemented both a least-squares fit of, say, ψ_l to $a x^l + b x^{l+1}$ (direct fit), and a least-squares fit of ψ_l / x^l to $a + bx$ (linear fit), and similarly for f_l and R_l. The direct fit weights grid points by x^l relative to the linear fit. In each case we fit to the first n_{fit} points. We obtain $R_{0(1)}$ as the value of the left difference at $x = 0$, as this term will have to cancel the equivalent transport term, and we then fit to $R - R_{0(1)}(u)x$ with the general method. We choose to also fit $R_{0(3)}$, $f_{2(3)}$, $\psi_{1(2)}$ and $\psi_{2(3)}$, which are not used, in order to fit two powers of x to each function. The exception is that we fit three powers to ψ_0, which then requires $n_{fit} \geq 3$.

If we restrict to linear perturbations about Minkowski spacetime in Bondi gauge, with $R = R_{0(1)}(u)x$ exactly in the background solution, we are dropping all other $R_l(k)$ and all products of expansion coefficients. In an early version of the code, we did that, and also truncated the expansions at different orders from the above, resulting in the following expansion:

$$\gamma_{0,1,2} = O(x^2), \tag{190}$$

$$b_1 = O(x^2), \tag{191}$$

$$\frac{R^4 b_{2,x}}{G} = -4 R_{0(1)}^2 \left(f_{2(2)} x^4 + \frac{6}{5} f_{2(3)} x^5 \right), \tag{192}$$

$$b_2 = -4 R_{0(1)}^{-1} \left(f_{2(2)} x + \frac{3}{5} f_{2(3)} x^2 \right), \tag{193}$$

$$(R \Xi R)_0 = -\frac{1}{2} R, \tag{194}$$

$$(R \Xi f)_2 = -f_{2(2)} x^2 - \frac{3}{10} f_{2(3)} x^3, \tag{195}$$

$$(R \Xi \psi)_0 = \frac{1}{2} \psi_{0(1)} x, \tag{196}$$

$$(R \Xi \psi)_1 = -\frac{1}{2} \psi_{1(1)} x, \tag{197}$$

$$(R \Xi \psi)_2 = -\psi_{2(2)} x^2, \tag{198}$$

with all other components initialized to zero. The almost-linear simulations presented here were carried out with this expansion, but we have checked since that the fully nonlinear expansion makes only a very small difference to the error we have measured in convergence tests. In particular, the magnitude and qualitative behavior of the error is unchanged.

3. Integration of the hierarchy equations

We initialize the integrals for all hierarchy equations up to $i = i_{expand} \geq 1$, and integrate from there. We truncate the integrand to $l_{max,local}(i)$ when integrating from x_{i-1} to x_i.

We then use ΞR to find the x-shift B. This allows us to calculate the upwinded x-derivatives of the evolved variables R, f and ψ, as these depend on the local sign of B.

In the integrals for $R\Xi f$ and $R\Xi\psi$ we truncate not the integrand but the whole integral to $l \leq l_{\text{max,local}}(i)$. We also set

$$(R\Xi\psi)_l = -(RB\psi_{,x})_{\text{upwind},l}, \qquad l > l_{\text{max,local}}(i). \quad (199)$$

With (161), this gives $(R\psi_{,u})_l = 0$ for $l > l_{\text{max,local}}(i)$. As R does not depend on y in lsB and gsB gauges, this then also sets $(\psi_{,u})_l = 0$, consistent with the boundary condition $\psi_l = 0$ that we impose on $l > l_{\text{max,local}}(i)$. We treat $R\Xi f$ similarly.

F. Time evolution

We set initial data for f and ψ on $u = 0$, $0 \leq x \leq x_{\text{max}}$. In gauges other than affine gauge we must also initialize R, and in these gauges we think of this initialization as pure gauge. We choose

$$R(0, x, y) = x/2, \quad (200)$$

in analogy with $R = (v + u)/2$ in the standard double null coordinates on Minkowski spacetime.

After solving all hierarchy equations, and applying the shift terms in Ξ, the resulting "time" derivatives $\phi_{\text{evol},u}$ are discretized using the second-order Runge-Kutta method, with all hierarchy equations and gauge conditions evaluated at each Runge-Kutta sub-step, so that our time update can be characterized as the "method of lines."

In double-null or sdn gauges, $R(u, x, y)$ is genuinely evolved, but in Bondi gauge $R = x$, and only f and ψ are evolved. In gsB gauge we have $R(u, x, y) = s(u)x$. Numerically, we evolve only $s(u)$ (as an auxiliary variable). In lsB gauge we have $R(u, x, y) = \bar{R}(u, x)$. Numerically, we evolve $R(u, x, y)$ but filter out the $l > 0$ components that are created by numerical error after each full time step.

V. TESTS IN THE ALMOST-LINEAR REGIME

We test convergence of the full nonlinear code in a regime of small deviations from Minkowski spacetime (with $\psi = 0$). We evolve in several of the nonlinear gauge choices we have discussed above. Specifically, we compare sdn gauge (62), gsB gauge (64) with $s(u)$ given by (65), lsB2 gauge (72), and lsBtosdn gauge (75), each using the (G, x) formulation and the (γ, R) formulation, the latter with and without the integration by parts (47). All tests use direct fits near the origin with $n_{\text{fit}} = 2$ and $n_{\text{expand}} = 1$ and the expansion (190)–(198).

A. Linearized solutions as test beds

In the small-data regime we can use exact solutions of the *linearized* field equations as test beds. Small data here

means in practice that the difference between the solutions of the linearized and nonlinear field equations can be neglected in comparison with the numerical error in the nonlinear evolution.

For clarity, in this subsection we denote the exact solutions of the linearized equations by $\delta\psi$, δf, δb, δR. These will then be good approximations to nonlinear but small ψ, f and b, and a small nonspherical part of R.

In the linearized equations, $\delta\psi$ on the one hand, and δf, δb and δR on the other, evolve independently, while different spherical harmonic components l also decouple from each other.

We write the Minkowski background $f = b = \psi = 0$ in a gauge which agrees with all of our gauge choices. The spherical background metric coefficients G, R and H in this gauge are given by Eqs. (B15)–(B16) in Appendix B.

The quantities δb and δR for the same physical solution disagree in different gauges. By contrast $\delta\psi$ is linearly gauge-invariant, and δf is linearly gauge-invariant within the class of lsB and gsB gauges; see also Appendix D 3. This means that the linearized solution in Bondi gauge also gives us $\delta\psi$ in any other gauge, and δf in any lsB or gsB gauge. In these gauges, $\delta R = 0$ (in the linearized equations). By contrast, in sdn gauge δR develops dynamically even if it is set to zero in the initial data.

B. Convergence test method

To look for second-order self-convergence of a variable ϕ with respect to Δx, we assume that the Richardson expansion

$$\phi_{\Delta x}(x) = \phi_0(x) + \phi_2(x)\Delta x^2 + O(\Delta x^3) \quad (201)$$

holds, where we have suppressed the other arguments of ϕ. $\phi_0(x)$ is the solution in the continuum limit in x, and $\phi_2(x)$ is the second-order error, assumed to be leading.

We now distinguish two cases. If $\phi(u, x, y)$ obeys a hypersurface equation (PDE in x and y only), then u is just a parameter for the purposes of the convergence test. At fixed finite resolution in y we can then think of the PDE as a (large) system of ODEs in x.

If, on the other hand, ϕ obeys an evolution equation (PDE in u, x and y), then at fixed finite resolution in y we can consider it as a (large) system of PDEs in u and x. Our time step criterion (165) scales Δu in proportion to Δx. If the discretization in u is also at least second-order accurate, then we expect that the discretization errors in both u and x are proportional to Δx^2, so we are testing convergence in u and x together.

In halving Δx exactly, Δu is halved approximately, but not exactly, by the application of the time step condition (165). To compensate for this, we align the output times at both resolutions exactly by adjusting the last time step coming up to the scheduled output time.

From pairs of numerical evolutions, we calculate the self-convergence error estimate

$$\mathcal{E}_{\phi,\Delta x} := \frac{\phi_{\Delta x} - \phi_{\Delta x/2}}{\Delta x^2 - (\Delta x/2)^2} \Delta x_{\text{ref}}^2 \qquad (202)$$

$$= \frac{3}{4} \left(\frac{\Delta x_{\text{ref}}}{\Delta x} \right)^2 [\phi_{\Delta x} - \phi_{\Delta x/2}] \qquad (203)$$

$$\simeq \Delta x_{\text{ref}}^2 \phi_2. \qquad (204)$$

Any pair of resolutions could be used to estimate the error, but for simplicity we use Δx and $\Delta x/2$, as the coarse grid is then aligned with the fine grid, so we can evaluate the difference on the coarse grid without interpolation. If we know the continuum solution ϕ_0, we can also compute the alternative error estimate

$$\mathcal{E}_{\phi,\Delta x} := \left(\frac{\Delta x_{\text{ref}}}{\Delta x} \right)^2 [\phi_{\Delta x} - \phi_0] \qquad (205)$$

$$\simeq \Delta x_{\text{ref}}^2 \phi_2. \qquad (206)$$

Note that $\mathcal{E}_{\phi,\Delta x}$ depends on the reference resolution Δx_{ref}, the fixed resolution Δy (or $l_{\max}$), and (u, x, y), but for brevity we do not write these arguments. If $\mathcal{E}_{\phi,\Delta x}$ calculated for two or more pairs of resolutions is similar at all Δx below some threshold, we have pointwise second-order convergence in Δx, and $\mathcal{E}_{\phi,\Delta x}$ itself is approximately equal, pointwise, to the discretization error in x, or in u and x, at the reference resolution Δx_{ref} and the fixed resolution Δy. The error at any other (smaller) Δx can be estimated by scaling with $(\Delta x/\Delta x_{\text{ref}})^2$.

In this paper we do not yet carry out systematic convergence testing in y, as we expect different spherical harmonics to decouple in almost-linear evolutions. Hence the error in $\phi_l(u, x)$ becomes negligible once $l_{\max} > l$, and otherwise $\phi_l(u, x)$ cannot be represented at all. However, for completeness, looking ahead to Paper II, we discuss testing convergence in y already here.

On a smooth nonlinear solution, a spectral method should converge exponentially, but we shall see in Paper II that our code converges only to second order in Δy. Hence, to look for second-order convergence with respect to $\Delta y \propto 1/l_{\max} \propto 1/N_y$, we assume that the Richardson expansion

$$\phi_{l_{\max}}(y) = \phi_0(y) + \phi_2(y) l_{\max}^{-2} + O(l_{\max}^{-3}) \qquad (207)$$

holds, where we have again suppressed the other arguments of ϕ and are keeping the resolution in them fixed. We can then consider the PDE as a large system of ODEs in y. $\phi_0(y)$ is the solution in the continuum limit in y (but at fixed finite resolution in x and u), and $\phi_2(y)$ is the second-order error, assumed to be leading. From pairs of numerical

evolutions, we then calculate the quantity

$$\mathcal{E}_{\phi,l_{\max 1}}(x, y) := \frac{\phi_{l_{\max 1}} - \phi_{l_{\max 2}}}{l_{\max 1}^{-2} - l_{\max 2}^{-2}} l_{\max\text{ref}}^{-2} \qquad (208)$$

$$\simeq l_{\max\text{ref}}^{-2} \phi_2. \qquad (209)$$

We can compare $\phi_l(u, x)$ at different $l_{\max}$, and so do not need to align y-grids at different resolutions.

For code checks, it is often useful to plot single Fourier components $\mathcal{E}_{\phi_l,\Delta x}(x, u)$ of the error against x, and animate these plots with time u. In Figs. 2–4, for data dominated by a single spherical harmonic, we show such single-l errors against x, at a representative moment of time u. In Fig. 5, for data containing all spherical harmonics, we instead take the root-mean-square (from now, rms) norm of $\mathcal{E}_{\phi,\Delta x}(x, u, y)$ over $0 \leq x \leq x_{\max}$ and $-1 \leq y \leq 1$, and plot this against u. We also evaluate the maximum norm of $\mathcal{E}_{\phi_l,\Delta x}(x, u)$ over x, and the maximum norm of $\mathcal{E}_{\phi,\Delta x}(x, u, y)$ over x and y, but we do not present plots here.
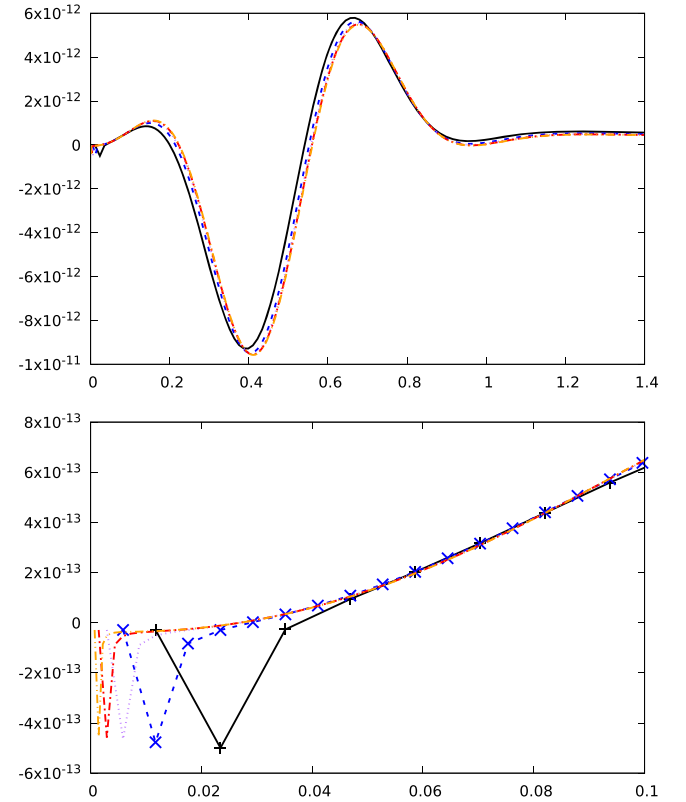

FIG. 2. D'Alembert test in lsB gauge: A snapshot of the scaled errors $\mathcal{E}_{\psi_2,\Delta x}(x)$ for the quadrupole component ψ_2 of the scalar field ψ, at five resolutions from $N_x = 256$ to $N_x = 4096$. We show a moment of time $u = 0.48$ where the wave passes through the center. In the upper plot we truncate the numerical domain $0 \leq x \leq 3$ to $0 \leq x \leq 1.4$, as nothing interesting happens at larger x. The lower plot, restricted to $0 \leq x \leq 0.1$, focuses on the center, and we show grid points for the two coarsest grids.
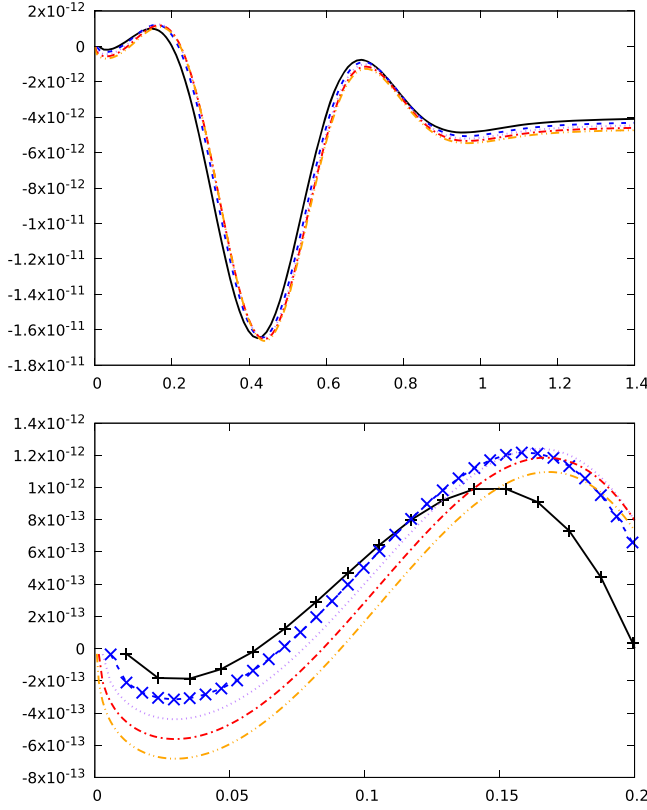
FIG. 3. D'Alembert test: As in Fig. 2, but now for the scaled error $\mathcal{E}_{f_2,\Delta x}(x)$. Again the lower plot is a detail of the upper one near the center, restricted here to $0 \leq x \leq 0.2$, and with grid points for the two lowest resolutions.

FIG. 4. D'Alembert test: As in Fig. 2, but now for the scaled error $\mathcal{E}_{b_2,\Delta x}(x)$. Again the lower plot is a detail of the upper one near the center, but now restricted to $0 \leq x \leq 0.02$, on a log-log scale, and showing grid points at all resolutions.

C. Single-l tests

1. Generalized d'Alembert exact solutions

We have tested two kinds of exact solutions. The first are the generalized d'Alembert solutions for a single spherical harmonic derived in Appendix D and given in (D23) for the scalar field $\delta\psi_l$, and in (D48)–(D49) for the coupled metric perturbations $(\delta b_l, \delta f_l)$, both in terms of an arbitrary function χ of one variable. These were derived, and used as test beds, previously in [9].

We choose the free function χ to be a Gaussian with center at 0.8 and width 0.2. As the amplitude for χ we choose 10^{-11-l}, which results in a maximum value $\sim 10^{-11}$ for ψ and f in the initial data. The numerical domain is $0 \leq x \leq 3$ with $x_0 = 2$ and $0 \leq u \leq 1.9$. We set $N_y = 5$, and $N_x = 64 \ldots 8192$, increasing by factors of two. For these tests, and in contrast to critical collapse applications, the fact that the grid shrinks to a point is irrelevant, and this is why we stop at $u = 1.9$, when the grid has shrunk to 0.05 of its original size, and the waves have essentially left the grid.

At high resolution, our numerical evaluation of the d'Alembert exact solution at the innermost grid points suffers from large round-off error, resulting from the
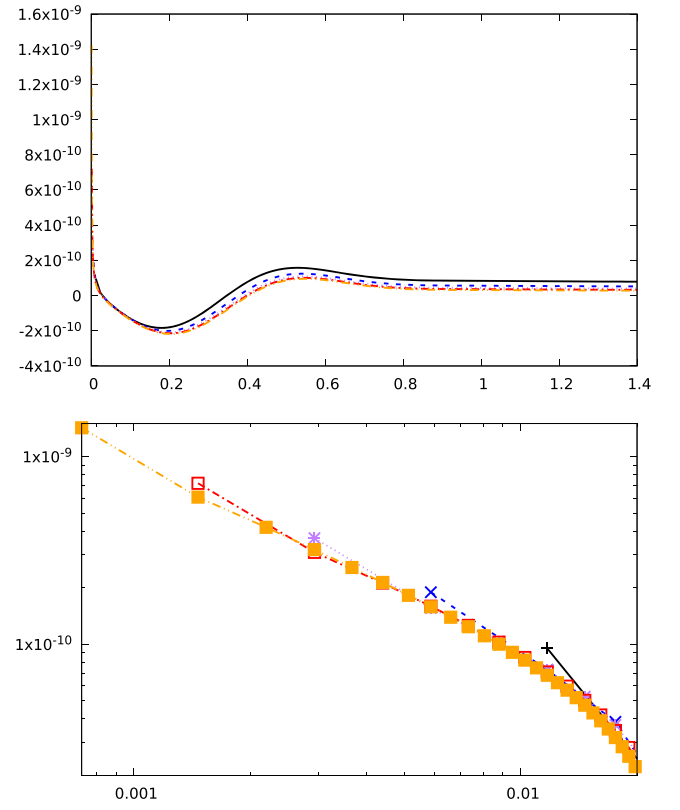
division by powers of r up to r^{l+1}. To get round this, at small r we implement a truncated power-series expansion of the d'Alembert solution.

In initial data in lsB and gsB gauge, R is set to the spherical background value given in Eq. (B14), that is $R_0(0, x) = x/2$. In sdn gauge, we add to this a Gaussian δR, so that we do not have to wait for δR to develop dynamically.

To start with a summary of the numerical results that follow, sdn and lsBtosdn gauge are unstable, our flavor of gsb is unstable in the (G, x) formulation but stable in the (γ, R) formulation (with and without integration by parts), and the lsB2 flavor of lsB gauge is stable in all three formulations. We then find pointwise second-order convergence (both self-convergence and convergence against the exact solution) for all l-components of all variables.

2. Results in sdn gauge

In sdn gauge, for $l = 3$ d'Alembert data and $l = 3$ d'Alembert plus Gaussian-in-R data, we have only gone to 1024 grid points to see that in all three formulations the error (against the exact solution) has a constant in x, oscillating in u, part that does not decrease with resolution.
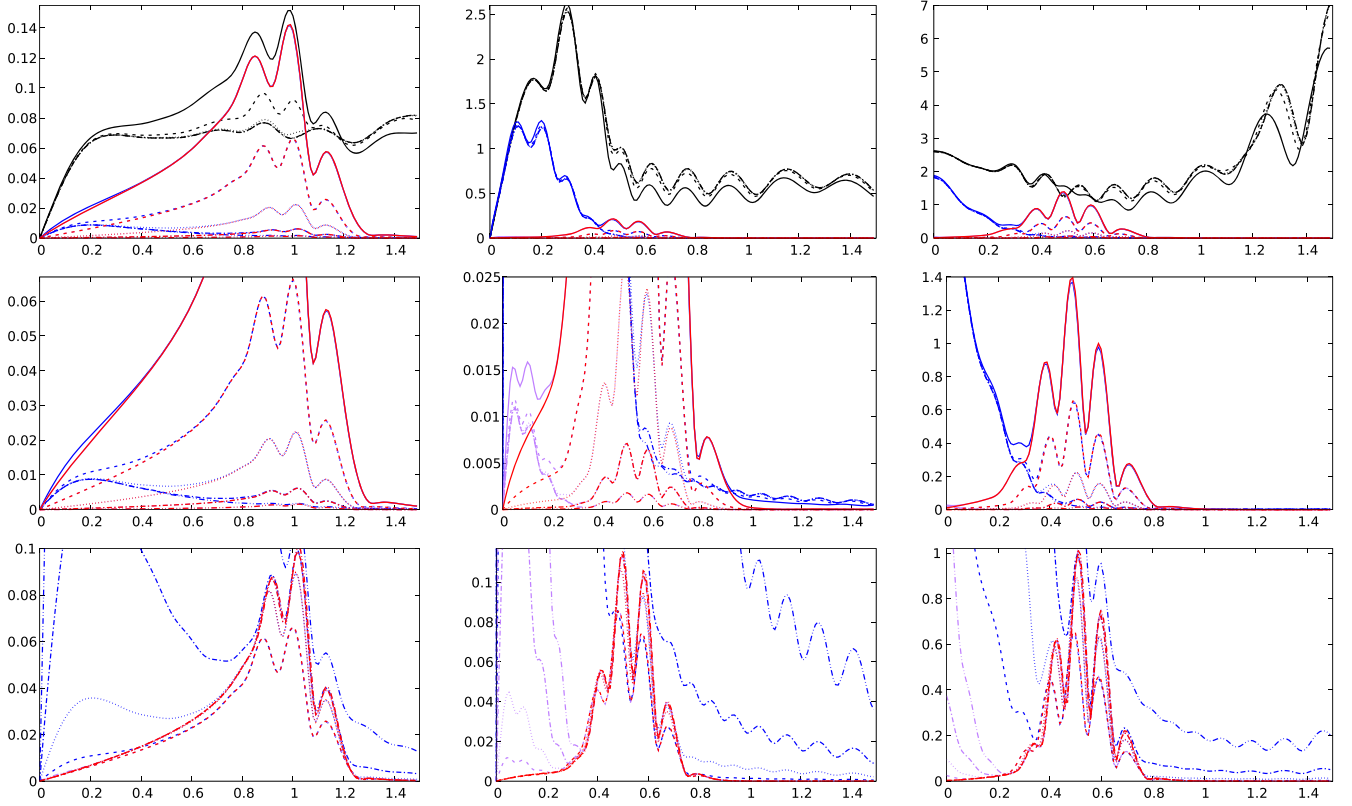
FIG. 5. (Single) plane-wave tests: the rms (over x and y) norm of the error in ψ, f and b (scaled by 10^{-11} to give an indication of relative error), against u. ψ is in the left column, f in the middle column, and b in the right column. Angular resolution is indicated by line color: $N_y = 17$ black, $N_y = 33$ blue, $N_y = 65$ purple and $N_y = 129$ red. Radial resolution is indicated by line type: $N_x = 64$ solid, $N_x = 128$ dashed, $N_x = 256$ dotted, $N_x = 512$ dot-dashed, and $N_x = 1024$ dot-dot-dashed. The middle plot in each column enlarges the lower part of the upper plot, and for clarity omits $N_y = 17$. The lower plot additionally omits $N_x = 64$, and is scaled with $(N_x/128)^2$, to test for (local-in-u) second-order convergence with N_x. The horizontal range is always $0 \leq u \leq 1.5$, but the vertical range (rms error) has been chosen differently in each of the nine plots.

In other words, in sdn gauge there is an instability at the origin.

3. Results in lsB tosdn gauge

In lsBtosdn gauge, f is unstable at the center and not converging from the start for $l = 2, 3$. (ψ still converges, presumably because it sees essentially flat spacetime). It is sufficient to go to 1024 grid points to see this. We have only tried the (γ, R) formulation with integration by parts.

A mild instability at the center is still present when the blending is between $0.3x_0$ and x_0, with pure lsB gauge for $x \leq 0.3x_0$. This could be because we are then not projecting R down to $l = 0$ even at those x. We also see something like a linear gauge shock at $x \sim 1.45$, in the middle of the blending zone.

4. Results in gsB gauge

In the (G, x) formulation, $l = 2$ is unstable in our flavor of gsb gauge, with again an oscillating instability at the center giving rise to an error constant in x that increases

with better resolution. We needed to go to 4096 points to see this. In the (γ, R) formulation with and without integration by parts, it is stable and perfectly second order convergent to 8192 points.

By contrast, $l = 3$ in gsb gauge is stable up to 8192 points already in the (G, x) formulation.

5. Results in lsB2 gauge

We have tested lsB2 gauge in the (G, x) formulation only for $l = 3$ and only up to 1024 gridpoints. The error at 1024 grid points is visually identical with that in the (γ, R) formulation with integration by parts. We have tested lsB2 gauge in the (γ, R) formulation with integration by parts in more detail. All error plots in the following are in this gauge and formulation.

In particular, we have tested the d'Alembert solution for $l = 0, 1$ (ψ only), $l = 2, 3, 4$ (ψ, f and b) and $l = 5$ (f and b only, not suppressing this highest frequency for once). We find second-order pointwise convergence of ψ_0, ψ_1 and, for $l \geq 2$, of ψ_l, f_l and b_f, from $N_x = 64$ to $N_x = 8092$ radial grid points.

For $l = 2$, the above statement needs to be qualified. The error in ψ_2 is not smooth near the origin, but has a blip at the second grid point (rather than a specific value of x). This is demonstrated in Fig. 2. However, we get pointwise second-order convergence at fixed x for all resolutions $\Delta x < x/2$.

Figure 3 shows the scaled error in f_2. Here, it is not clear if we have reached convergence even at $N_x = 4096$, and it appears again that the error is not smooth at the origin, although in a different manner to that in ψ_2.

Figure 4 shows the scaled error in b_2. Near the origin, the scaled error is $\mathcal{E}_{b_2, \Delta x}(x) \sim (\Delta x)^2/x$. At the innermost grid point this evaluates to $\sim \Delta x$. Hence b_2 converges pointwise to second order at all points, including near the origin, but converges only to first order in the maximum norm, which at high resolution is dominated by that innermost grid point. In other norms, such as the root-mean-square norms, the convergence would be at an intermediate order.

For $l > 2$, we suspect that the error is still technically unsmooth near the origin, but this is hard to see as the errors, like the variables themselves, are suppressed at the origin by powers of x^l for f, ψ and the other scalars, and x^{l-1} for b. Hence the second-order pointwise convergence looks fine. It is possible that our methods can be improved to make the error better behaved at the origin, but we leave this to future work.

We note finally that, where we have the exact solution, the self-convergence estimate of the error is pointwise approximately equal to the true error (difference from the exact solution).

To test convergence at $l > 5$, we change from the d'Alembert solution (which becomes increasingly complicated) to simple Gaussian data for a single l. As we need larger N_y to represent larger l, we go to the half-range, with $\bar{N}_y = 17$. We now set initial data directly for f_l and ψ_l, namely a Gaussian again with center at $x = 0.8$, width 0.2, and amplitude 10^{-11}. The numerical domain is again $0 \leq x \leq 3$ with $x_0 = 2$ and $0 \leq u \leq 1.9$, with $\bar{N}_y = 17$, and $N_x = 64\ldots8192$, increasing by factors of two. At $l = 2$, we find the same as with the d'Alembert $l = 2$ data. At $l = 4$, 8, 16, 32 we find apparent perfect second-order pointwise convergence (as for the $l = 4$ d'Alembert data).

D. Plane-wave tests

1. Plane-wave exact solutions

We show in Appendix D 6 that any solution of the scalar wave equation gives rise to a solution of the linearized Einstein equations, without the need for a spherical harmonic decomposition. In particular, the scalar wave equation admits plane-wave exact solutions, and in Appendix D 6 b we give these in Eq. (D61) for $\delta\psi$ and in (D62)–(D63) for the related $(\delta b_l, \delta f_l)$. Of course, these are just the translation into our coordinates of the linear limit of the

well-known plane-symmetric gravitational waves. This is our second exact solution test bed.

A scalar field plane wave moving in the negative z-direction on flat spacetime takes the form

$$\psi = \chi(t + z) = \chi(u + r + z) = \chi(u + r(1 + y)). \quad (210)$$

Its contours at constant z are parabolas in the (ρ, z) plane (where $\rho^2 := r^2 + z^2$), as one expects of the intersection of a null plane with a null cone. When this intersection reaches the vertex of the cone it closes up and disappears.

We choose χ to be a Gaussian with center 1.0 and width 0.1, and with amplitude 10^{-11} for ψ, which means that the maximum of ψ is exactly 10^{-11}, and amplitude 10^{-13} for f, which means that the maximum of f is approximately 4×10^{-11}. The numerical domain is again $0 \leq x \leq 3$ with $x_0 = 2$, now with $0 \leq u \leq 1.5$. This means that the plane wave crosses the origin and then disappears out of the numerical domain during the evolution.

The (single) plane wave solutions contain both even and odd spherical harmonics. In order to test plane waves with our half-range formulation, we add a copy of the same wave moving in the opposite direction to make ψ and f even functions of y (at constant u and x), and b an odd function.

2. Results in lsB2 gauge

We evolve with all combinations of $N_y = 17$, 33, 65, 129, $\bar{N}_y = 9$, 17, 33, 65 (for the double plane wave only), and $N_x = 64\ldots1024$.

This double plane wave solution contains only even spherical harmonics, and so can be run on the full or half range in y. We have verified that the errors in the double plane wave test are identical in the equivalent resolution pairs. We remove the top two frequencies in f and top frequency in b in the exact solution before the comparison with the numerical solution, as the numerical solution is similarly truncated.

Beginning with the single plane wave, we evaluate the maximum and rms norms of the error against u. The norms are taken over all grid values of x and y, with $y = \pm -1$ weighted half in the rms norm in order to make it the same on the full and half-grid. The rms error in ψ, f and b for the single plane wave at the different resolutions is shown in Fig. 5.

The figure consists of nine plots laid out in a square. The three columns from left to right show the variables ψ, f and b. Focus initially on the middle column, showing f. Each plot shows four differences of angular resolution and five differences of radial resolution. Different radial resolutions are distinguished by line type, and different angular resolution by line color.

In the top plots, the rms error is *not* scaled. We see that the error decreases quickly with angular resolution N_y, but is almost independent of radial resolution N_x. This

indicates that for $N_y = 17$, 33 and $N_x = 64...1024$, the total error budget is dominated by angular discretization error.

The middle plots zoom in on the smallest errors, which arise from the highest angular resolutions. They demonstrate that at $N_y = 65$, 129 the error does decrease with increasing radial resolution N_x. In the lowest plot these errors are scaled with $(N_x/128)^2$. The scaled curves with $N_y = 129$ (omitting $N_x = 64$) lie on top of each other, indicating second-order convergence with N_x. This indicates that at $N_y = 129$, for $N_x = 128...1024$ the error budget is dominated by the radial discretization error.

Comparing now the three variables ψ, f and b, we note that as b is computed from f at each time step, it shows a numerical error already in the initial data, whereas ψ and f are evolved from analytic initial data, and so their error is zero at $u = 0$. For $N_y = 129$ only, the initial error in b is negligible, and the subsequent error in b is similar to the error in f or ψ.

We see perfect second-order convergence with N_x in ψ (but not f and b) already at $N_y = 65$. Moreover, the errors in ψ with $N_y = 65$ and 129 (purple and red curves) are indistinguishable in our plots. We conclude that a plane scalar wave requires less angular resolution than a plane gravitational wave of the same shape χ (where χ is a solution of the scalar wave equation). This may be because the exact solution for f and b involves derivatives of χ.

A closer look at the middle and bottom plots shows a transition from an error dominated by y-discretization at early times and small N_y and by x-discretization at late times and large N_y. Plotting single-l components of the error against u and x further shows that the early discretization error in y arises mostly at $x \gtrsim 1.5$ and the late error discretization error in x mostly at $x \lesssim 1.5$.

The rms errors in the double plane wave test are very similar as functions of u, but larger by roughly $\sqrt{2}$, as there are now two identical waves of error instead of one in the same numerical domain. We therefore do not present plots here.

E. Tests of the computation of the Hawking mass

As a different indication of numerical error, we have implemented the expression for M given by (84) with (103), a centered finite differencing of this to compute the resulting $M_{,x}$, and, independently, the direct expression (105) for $M_{,x}$. The two expressions for $M_{,x}$ are affected in different ways by finite differencing error in x and spectral error in y, and so their agreement in the continuum limit is a nontrivial test of the correctness of (84) with (103), (105), the hierarchy equations, and their discretizations. The cleanest test is one where the solution is close to Minkowski, so that the hierarchy equations are approximately linear in ψ, b and f, and the expressions for γ and M are approximately quadratic.

We have tested single-l Gaussian initial data in f or ψ, and have varied the amplitude, N_x, N_y, and the numerical method. We have not evolved these data.

We first take a Gaussian in f only with center at $x = 1$, width 0.25, $x_{\max} = 5$, $l = 2$, and amplitude 10^{-4}. Then $M \sim 9 \times 10^{-8}$ at the outer boundary, and $M_{,x}$ has a maximum of $\sim 2.5 \times 10^{-7}$. Our baseline resolution is $N_x = 1000$ and $N_y = 5$. At this baseline, the difference between the two versions of $M_{,x}$, an estimate of the error in either, has maximum $\sim 1.1 \times 10^{-10}$, a relative error of $\sim 4 \times 10^{-4}$.

Decreasing the amplitude by a factor of 10 reduces M and the error in $M_{,x}$ by a factor of 100, as expected. If we decrease the amplitude much more, the error is dominated by noise. This would be expected as round-off-error in the cancellation between 1 and $2R^2 \rho_+ \rho_-$ in C.

The error with $N_y = 5$, 9, 17, 33, 65, 129 is the same, and so for $\bar{N}_y = 3$, 5, 9, 17, 33, 65 on the half-range. Doubling N_x from the baseline reduces the error by a factor of 4. Hence the error in M in almost-linear evolutions is dominated by finite differencing in x, while spectral error in y is negligible.

Initial data in which only ψ is nonvanishing test other parts of the two expressions for $M_{,x}$. Note that $b \neq 0$ when hierarchy equations are solved for these data even with $f = 0$. We use a Gaussian with the same center and width as for f above. An amplitude of 10^{-4} gives $M \sim 5.5 \times 10^{-8}$. $M_{,x}$ has a maximum of $\sim 1.5 \times 10^{-7}$, and the error at baseline has a maximum of $\sim 4.5 \times 10^{-11}$, a relative error of $\sim 3 \times 10^{-4}$. All other comments for pure f data just above also apply here.

VI. CONCLUSIONS

We have begun an investigation of the use of null coordinates in numerical relativity, applied to gravitational collapse. In this first paper we have both made progress and identified new difficulties.

A. Progress

On the purely mathematical side, we have shown that the ingoing null derivative Ξ normal to the surfaces of constant (u, x) plays a role similar to the Lie derivative $\mathcal{L}_n$ normal to spacelike time slices. Specifically, and oversimplifying a bit, the Einstein equations give us the geometric time derivative $\mathcal{L}_n \mathcal{L}_n g_{ij}$ on the usual spacelike time slices, but Ξg_{ij} on null slices. Writing the hypersurface equations in terms of Ξ removes any explicit appearance of the radial shift B, just as writing the $3 + 1$ equations in terms of $\mathcal{L}_n$ removes any explicit appearance of the lapse and shift.

On the numerical side, we have removed a practical obstacle to using null cones with a regular center, already identified in [9], namely that the time step $\Delta u \sim \Delta x (\Delta\theta)^2$ is unreasonably small at high angular resolution. We have been able to replace this by $\Delta u \sim \Delta x$, similar to Cartesian

coordinates, by reducing the angular resolution at small radius.

For applications to critical collapse in spherical polar coordinates, we have found an equivalent of the method of repeated radial regridding of [27] that works beyond spherical symmetry. Essentially this is done by adding to a standard gauge choice, such as Bondi gauge, an ingoing radial shift that shrinks the grid continuously, with the outer boundary becoming future spacelike [30].

In the present paper, we have demonstrated convergence of these numerical methods in the evolution of weak data, and in Paper II we successfully apply them axisymmmetric scalar field critical collapse.

B. Problems

As we have discussed, one cannot find marginally outer-trapped surfaces embedded in null cones with a regular center, and in vacuum axisymmetry not even any (non-marginally) outer-trapped surfaces. We propose using closed 2-surfaces of large Hawking compactness (Hawking mass/area) as an alternative diagnostic of black hole formation. In Paper II, this allows us to find the threshold of black-hole formation by bisection, and demonstrate critical scaling of the black hole mass.

It becomes clear in Paper II that in sufficiently non-spherical spacetimes the divergence of the congruence of the generators of our null cones becomes negative in some directions even when no black hole is formed. Bondi coordinates and double-null coordinates, and the generalized Bondi coordinatees of Paper II, break down when this happens.

C. Outlook

Two key questions remain open. The first is whether outgoing null cones emanating from a regular center remain regular in strong gravity further away from spherical symmetry, and in particular in electromagnetic or vacuum critical collapse. This is a purely geometric question, independent of gauge or numerical methods. Note also that the potential problem is nonsphericity, rather than black hole formation. One reason to be optimistic is that in [5] we have constructed outgoing null cones emanating from a regular vertex in postprocessing of $3 + 1$ near-critical vacuum evolutions, as a way of comparing evolutions in different coordinate systems, without finding caustics.

A second question is how much useful information about any newly formed black hole we can find, given that outgoing null coordinates cannot penetrate into the horizon, or at least not very deeply, and that we cannot find MOTS on a single coordinate null cone.

As already discussed, going further will probably require a change to a generalized affine parameter gauge. We leave this to future work.

APPENDIX A: THE TIME STEP PROBLEM IN NULL COORDINATES

We show here that the time step for explicit finite differencing schemes is even more severely limited in the combination of polar spatial coordinates with a regular center and a null "time" coordinate u than for polar coordinates and the usual (spacelike) time coordinate t. For simplicity we restrict to axisymmetry in $3 + 1$ dimensions. We focus on the causal geometry, rather than giving a rigorous argument. Our presentation closely follows [9].

Consider at first the Minkowski metric in standard spherical polar coordinates,

$$ds^2 = -dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2). \quad \text{(A1)}$$

In an abuse of notation, we now let dx^μ stand for the small finite coordinate distance from the grid point that is being updated to any one of the other grid points from which the update is calculated. These will be integer multiples of the grid spacings Δx^μ. For an explicit numerical method, the update is found from only a few neighboring grid points to the past. The set of these points is called the "stencil" of the point being updated. Causality requires that the numerical stencil be wider than the physical light cone, or, in our notation, $ds^2 > 0$ for the outermost grid points in the stencil. Empirically, this is also a necessary criterion for the stability of any explicit finite difference scheme, often referred to as the Courant-Friedrichs-Lewy (CFL) condition.

For simplicity we assume that the stencil involves, besides the point being updated, only grid points on the previous time level. We parametrize such a stencil by $dt = -\Delta t$, $dr = s\Delta r$, $d\theta = q\Delta\theta$, with $\Delta t, \Delta r, \Delta\theta > 0$ by definition, and $s, q = -1, 0, 1$ as a simple example of a 3×3 point stencil in r and θ. We assume axisymmetry, so nothing depends on φ. We also have $r = p\Delta r$, where, for example, $p = 1, 2, 3, \ldots$ on a centered equally spaced grid. What matters here is only that s and q are of order unity and so is the smallest possible value of p, which occurs next to the center.

The CFL condition $ds^2 > 0$ translates into

$$\Delta t < \Delta r \sqrt{s^2 + p^2 q^2 \Delta\theta^2}. \tag{A2}$$

The choice $s = 0$, $q = \pm 1$ of points in the stencil, with the choice $p = 1$ to locate the stencil next to the center, gives

$$\Delta t \lesssim \Delta r \Delta\theta, \tag{A3}$$

where we have replaced $<$ by $\lesssim$ to allow for more general stencils and for other $O(1)$ factors in the argument. As is well-known, this is worse than the time step restriction $\Delta t \lesssim \min(\Delta x^i)$ in Cartesian coordinates by the factor $\Delta\theta \ll 1$. Put differently, the right-hand side of (A3) is quadratic in small quantities. Halving the grid spacing in all spatial directions halves Δt in Cartesian coordinates, but reduces it to a quarter in polar coordinates.

Consider now the Minkowski metric in the null coordinate form

$$ds^2 = -2G du dx - H du^2 + R^2(d\theta^2 + \sin^2\theta d\varphi^2). \tag{A4}$$

We let $du = -\Delta u$, $dx = -s\Delta x$, $d\theta = q\Delta\theta$, and $R = p\Delta x$. At the center $R = 0$, any radial gauge must approach Bondi gauge to keep it at $x = 0$, and setting $H = 1$ and $R = x$ there without loss of generality, we also have $G = 1$. The CFL condition (A2) is now

$$\Delta u \lesssim \Delta x \left(\sqrt{s^2 + p^2 q^2 (\Delta\theta)^2} - s \right). \tag{A5}$$

With $s = 1$, $p = 1$, $q = \pm 1$, and assuming $\Delta\theta \ll 1$, the equivalent of (A3) is now

$$\Delta u \lesssim \frac{1}{2} \Delta x (\Delta\theta)^2, \tag{A6}$$

worse by a second factor of $\Delta\theta \ll 1$ than for polar coordinates on spacelike time slices. Except in spherical symmetry, this is a problem of any explicit numerical time evolution scheme on null cones with a regular vertex.

At large R, the sharpest CFL condition arises from stencil points with $q = 0$ (and p then drops out). We now consider arbitrary G and H. The CFL condition becomes

$$\Delta u < -\frac{2sG}{H} \Delta x. \tag{A7}$$

Recall that $G > 0$. Choosing $s = \pm 1$ with sign opposite to that of H we obtain

$$\Delta u < \frac{\Delta x}{B}, \tag{A8}$$

where $B := H/(2G)$ as defined previously. This is just the CFL condition for the x-advection term in $\partial_u = \Xi + B\partial_x + Sb\partial_y$.

APPENDIX B: MINKOWSKI SPACETIME

In Minkowski spacetime, we can choose coordinates where $f = b = 0$ and the remaining metric coefficients R, G and H depend only on u and x. We denote them by R_0, G_0 and H_0. With these assumptions, there are only two nontrivial hierarchy equations, namely

$$\left(\ln \frac{G_0}{R_{0,x}} \right)_{,x} = 0, \tag{B1}$$

$$(R_0 \Xi_0 R_0)_{,x} = -\frac{1}{2} G_0. \tag{B2}$$

Clearly, these can be integrated in terms of two arbitrary functions of u.

To clarify what these two functions are in a regular spacetime, we note that in the standard double null coordinates (U, V), the Minkowski metric is

$$ds_0^2 = -dU dV + \left(\frac{V - U}{2} \right)^2 d\Omega^2. \tag{B3}$$

We now change to the most general null coordinates u and x adapted to the spherical symmetry, defined by

$$U = U(u), \tag{B4}$$

$$V = 2R_0(u, x) + U(u), \tag{B5}$$

and obtain the metric

$$ds^2 = -2G_0 du dx - H_0 du^2 + R_0^2 d\Omega^2, \tag{B6}$$

where

$$G_0 = U' R_{0,x}, \tag{B7}$$

$$H_0 = U'(U' + 2R_{0,u}), \tag{B8}$$

and hence

$$\Xi_0 R_0 = -\frac{1}{2} U'. \tag{B9}$$

Therefore the general solution of (B1), (B2) with a regular center is (B7) and (B9): note this has only one free function $U(u)$. (B7) can be written as

$$g_0 = U'(u). \tag{B10}$$

In Bondi coordinates, where $R_{,u} = 0$, we have

$$H_{0,\text{Bondi}} = U'(u)^2. \tag{B11}$$

If we choose $x = 0$ to be a geodesic and the coordinate basis vectors to be parallely transported along it, the metric

at $x = 0$ is given by the Minkowski metric (B6) at $x = 0$ for all times.

In flat spacetime, shifted double null coordinates, shifted global Bondi coordinates and a natural choice of local shifted Bondi coordinates are all identical. To find the metric in these coordinates, we start again from the metric (B3) and make the specific coordinate transformation

$$U = u - x_0, \tag{B12}$$

$$V = (u - x_0)\left(1 - \frac{x}{x_0}\right), \tag{B13}$$

where $x_0 > 0$ is a parameter, to obtain (B6) with

$$R_0 = \frac{1}{2}\left(1 - \frac{u}{x_0}\right)x, \tag{B14}$$

$$G_0 = \frac{1}{2}\left(1 - \frac{u}{x_0}\right), \tag{B15}$$

$$H_0 = 1 - \frac{x}{x_0}. \tag{B16}$$

It follows that

$$g_0 = 1, \tag{B17}$$

$$B_0 := \frac{H_0}{2G_0} = \frac{x_0 - x}{x_0 - u}, \tag{B18}$$

$$\Xi_0 R_0 = -\frac{1}{2}. \tag{B19}$$

We call this the shifted Minkowski (from now on, sM) background gauge.

APPENDIX C: SPHERICAL SYMMETRY

We restrict to spherical symmetry, but bring in a spherical scalar field as matter, by setting $f = b = 0$ and making R, G, H and ψ functions of u and x only. The metric is

$$ds^2 = -2G\,du\,dx - H\,du^2 + R^2 d\Omega^2. \tag{C1}$$

The hierarchy equations become

$$\mathcal{D}\ln g = 4\pi R(\mathcal{D}\psi)^2 \tag{C2}$$

$$\mathcal{D}(R\Xi R) = -\frac{g}{2}, \tag{C3}$$

$$\mathcal{D}(R\Xi\psi) = -(\Xi R)\mathcal{D}\psi, \tag{C4}$$

where

$$\Xi = \partial_u - \frac{H}{g}\mathcal{D}. \tag{C5}$$

In spherical symmetry, diagnosing collapse and estimating the horizon mass is straightforward. The Hawking compactness C and mass M are given by

$$M = \frac{R}{2}C, \qquad C = 1 + \frac{2\Xi R}{g}, \tag{C6}$$

and, with $|\nabla R|^2 = -2\Xi R/g$, the Hawking mass in spherical symmetry is equal to the well-known Misner-Sharp mass

$$M = \frac{R}{2}(1 - |\nabla R|^2). \tag{C7}$$

(Unlike the Hawking mass in general, the Misner-Sharp mass in spherical symmetry derives from a conserved stress-energy current [37].) The mass aspect (105) is

$$M_{,x} = -4\pi R^2 R_{,x}\frac{\Xi R}{g}(\mathcal{D}\psi)^2, \tag{C8}$$

and the redshift defined in (77) is

$$Z = \sqrt{-2g\Xi R}, \tag{C9}$$

We see that, as long as ΞR remains finite, $g \to \infty$ gives both $Z \to \infty$ and $C \to 1$, so this is the obvious criterion for black hole formation, both from the compactness reaching one and the red shift from the center to infinity diverging.

Assuming that the time step is limited by the Bondi shift (56), as in Eq. (165), we have

$$\frac{\Delta u}{\Delta x} \lesssim \left|\frac{R_{,x}}{\Xi R}\right| = \frac{G}{|g||\Xi R|}. \tag{C10}$$

At finite G and ΞR, this goes to zero again as $g \to \infty$, or equivalently $R_{,x} \to 0$.

A commonly used non-null coordinate system in spherical symmetry is the polar-radial one, defined by

$$ds^2 = -\alpha^2 dt^2 + a^2 dR^2 + R^2 d\Omega^2, \tag{C11}$$

where R is now a coordinate. It is instructive to relate this to our null coordinates.

Expressing (t, R) in terms of (u, x), and comparing coefficients of (C11) and (C1), we can write the resulting three equations as

$$t_{,x} = \frac{a}{\alpha}R_{,x}, \tag{C12}$$

$$g = -a^2\left(R_{,u} - \frac{\alpha}{a}t_{,u}\right), \tag{C13}$$

$$H = g\left(R_{,u} + \frac{\alpha}{a}t_{,u}\right), \tag{C14}$$

and from this we can derive

$$\Xi R = -\frac{g}{2a^2} = \frac{1}{2}\left(R_{,u} - \frac{\alpha}{a}t_{,u}\right), \tag{C15}$$

$$C = 1 - \frac{1}{a^2} = 1 - |\nabla R^2|, \tag{C16}$$

$$Z = \frac{g}{a} = \alpha t_{,u} - aR_{,u}. \tag{C17}$$

From (C12) and (C17) we find

$$Z du = \alpha dt - a dR, \tag{C18}$$

and so the redshift of the center with respect to other observers at constant R is

$$Z = \alpha \frac{dt}{du}\bigg|_R = \frac{dt_0}{du}\bigg|_R, \tag{C19}$$

where t_0 is proper time along worldlines of constant R. In a static spacetime only, we can set $t_{,u} = 1$ without loss of generality and $R_{,u} = 0$, and we obtain $Z = \alpha$.

APPENDIX D: AXISYMMETRIC LINEAR PERTURBATIONS OF MINKOWSKI SPACETIME

1. Linearized field equations in any radial gauge

As a test of our formulation and numerical methods beyond spherical symmetry, we linearize about flat spacetime. We denote the background values of all fields by a subscript 0, and so $\psi_0 = 0$. We adapt the background coordinates to spherical symmetry by making G_0, H_0 and R_0 functions of u and x only. Although $G_0 = R_{0,x}$ holds in the background, for clarity we write either G_0 or $R_{0,x}$ in the linearized equations, as they appear in the linearization process.

The linearized hierarchy equations are

$$\left(\frac{\delta G}{G_0} - \frac{\delta R_{,x}}{R_{0,x}}\right)_{,x} = 0 \tag{D1}$$

for δG, then

$$\left(\frac{R_0^4 \delta b_{,x}}{G_0}\right)_{,x} = R_0^2(2S\delta f_{,xy} - 8y\delta f_{,x}) - \frac{R_0^2}{G_0}\delta G_{,xy}$$
$$- 2R_0\delta R_{,xy} + 2R_{0,x}\delta R_{,y}$$
$$+ \left(\frac{R_0^2 G_{0,x}}{G_0^2} + 2\frac{R_0 R_{0,x}}{G_0}\right)\delta G_{,y}, \tag{D2}$$

$$(R_0\Xi_0\delta f)_{,x} = \frac{R_0}{4x}\delta b_{,xy} + \frac{R_{0,x}}{2}\delta b_{,y} - (\Xi_0 R_0)\delta f_{,x}$$
$$+ \frac{1}{4R_0}\delta G_{,yy} \tag{D3}$$

for δb and δf,

$$(R_0\Xi_0\delta R + (\Xi_0 R_0)\delta R - R_0 R_{0,x}\delta B)_{,x} = \delta S_R \tag{D4}$$

for either δR or δH (we have not written out δS_R as it is long), and

$$(R_0\Xi_0\delta\psi)_{,x} = \frac{G_0 S}{2R_0}\delta\psi_{,yy} - \frac{G_0 y}{R_0}\delta\psi_{,y} - (\Xi_0 R_0)\delta\psi_{,x} \tag{D5}$$

for $\delta\psi$. We see that $\delta\psi$ decouples from the metric equations in any radial gauge, and is independent of the choice of linearized radial gauge.

With the background solution $G_0 = R_{0,x}$, and the boundary condition $\delta G = \delta R_{,x}$ at the origin, which follows from linearizing the gauge condition $G = R_{,x}$ that u is proper time at the origin, the solution of (D1) is

$$\delta G = \delta R_{,x} \tag{D6}$$

everywhere. This is far as we can go without choosing a (linearized) radial gauge.

2. Linearized field equations in linearized Bondi gauge

The perturbation equations take their simplest form in Bondi gauge. This is defined by $R = x =: r$ in the full equations, and hence by $R_0 = x$, $G_0 = H_0 = 1$ in the background, and $\delta R = \delta G = 0$ for the perturbations.

However, the choices of background gauge and linear perturbation gauge are in principle independent. In particular, we can choose to use sM gauge in the background, but linearized Bondi gauge $\delta R = \delta G = 0$ for the perturbations. Under a change of background gauge, for example from Bondi to sM, the linear perturbations δb, δf and $\delta\psi$ change only their argument, as if they were scalars.

In linearized Bondi gauge, defined by $\delta R = \delta G = 0$, δf and δb obey the coupled equations

$$\left(\frac{R_0^4 b_{,x}}{G_0}\right)_{,x} = R_0^2(2S\delta f_{,xy} - 8y\delta f_{,x}), \tag{D7}$$

$$(R_0\Xi_0\delta f)_{,x} = \frac{R_0}{4}\delta b_{,xy} + \frac{R_{0,x}}{2}\delta b_{,y} - (\Xi_0 R_0)\delta f_{,x}. \tag{D8}$$

These are just the first lines of (D2), (D3) above. The linearized wave equation (D5) does not simplify further. Using $G_0 = R_{0,x}$, as well as $\delta G = \delta R = 0$, the linearized hierarchy equation (D4) becomes

$$\left(-\frac{R_0}{2}\delta H\right)_{,x} = \frac{R_0^2}{4}\left(S\delta b_{,xy} - 2y\delta b_{,x}\right)$$
$$+ R_{0,x}\bigg[(6S - 4)\delta f - 2yR_0\delta b$$
$$+ \frac{S}{2}\left(2R_0\delta b_{,y} + 8y\delta f_{,y} - S\delta f_{,yy}\right)\bigg]. \tag{D9}$$

Given a solution $(\delta f, \delta b)$ of (D7), (D8), (D9) can be solved for δH by integration, but as δH does not couple back into the equations for δb and δf, we can ignore it.

The perturbation equations in linearized Bondi gauge (D8), (D7) are given in Bondi background coordinates as Eqs. (D25), (D26) below.

3. Other linearized radial gauges

Consider now the equations linearized about Minkowski in any linear gauge obtained by linearizing a nonlinear gauge. The different spherical harmonics l still evolve independently.

If $R_{,y} = 0$ in the nonlinear gauge, such as in gsB and lsB gauge, the $l \neq 0$ components of δR are absent, and hence, from (D6), also the $l \neq 0$ components of δG. (D3), (D2) then still reduce to (D8), (D7), and exact solutions derived in linear Bondi gauge are still relevant, after changing only their argument.

By contrast, in any gauge where $R_{,y} \neq 0$, such as sdn gauge, δR and δG couple back to δf and δb. However, δf transforms as a scalar under changes of radial gauge, even nonlinearly, and because f is already a perturbation the change due to the change of argument is quadratically small. By contrast, δb changes already to linear order. δR does transform as a scalar, but $R_{,x} \neq 0$ in the background solution, so the change of argument changes δR to linear order. In summary, we can still use the exact solution for δf as a test bed, but not δb.

4. Solution of the scalar wave equation in Bondi gauge

We now find explicit solutions of the scalar wave equation for $\delta \psi$, and of the coupled equations for δf and δb in linear Bondi gauge, using separation of variables. We use Bondi gauge $R = x$ in the background, and to indicate this we write r for x. At the end we translate the results back into sM background gauge. We begin in this subsection with the wave equation.

The linear wave equation on flat spacetime in Bondi coordinates is

$$-2\delta\psi_{,ur} + \delta\psi_{,rr} + \frac{2}{r}(\delta\psi_{,r} - \delta\psi_{,u})$$
$$+ \frac{1}{r^2}[(1-y^2)\delta\psi_{,y}]_{,y} = 0. \quad \text{(D10)}$$

Replacing the retarded time coordinate u with the Minkowski time coordinate

$$t := u + r, \quad \text{(D11)}$$

and writing

$$\delta\psi(u, r, y) =: \delta\phi(t, r, y) = \delta\phi(u+r, r, y), \quad \text{(D12)}$$

this takes the more familiar form

$$-\delta\phi_{,tt} + \delta\phi_{,rr} + \frac{2}{r}\delta\phi_{,r} + \frac{1}{r^2}[(1-y^2)\delta\phi_{,y}]_{,y} = 0. \quad \text{(D13)}$$

We initially work in coordinates (t, r, y) and switch back to (u, r, y) later.

a. Separating off the y-dependence.

Spherical symmetry of the background allows us to separate the angle y with the ansatz

$$\delta\phi(t, r, y) =: \sum_{l=0}^{\infty} \delta\phi_l(t, r) P_l(y), \quad \text{(D14)}$$

where $P_l(y)$ is the Legendre polynomial of order l, obeying

$$L^2 P_l := (1-y^2)P_l'' - 2yP_l' = -l(l+1)P_l. \quad \text{(D15)}$$

We then have the one-dimensional wave equation

$$-\delta\phi_{l,tt} + \delta\phi_{l,rr} + \frac{2}{r}\delta\phi_{l,r} - \frac{l(l+1)}{r^2}\delta\phi_l = 0, \quad \text{(D16)}$$

for the l-th partial wave, or equivalently

$$-2\delta\psi_{l,ur} + \delta\psi_{l,rr} + \frac{2}{r}(\delta\psi_{l,r} - \delta\psi_{l,u}) - \frac{l(l+1)}{r^2}\delta\psi_l = 0, \quad \text{(D17)}$$

where

$$\delta\psi(u, r, y) =: \sum_{l=0}^{\infty} \delta\psi_l(u, r) P_l(y), \quad \text{(D18)}$$

b. General solution of lth partial wave equation

We can find the general solution of (D16) that is regular at $r = 0$ in terms of a single free function χ_l of one variable, in the form of a generalized d'Alembert solution [38], as

$$\delta\phi_l(t, r) = \sum_{p=0}^{l} A_l^p r^{-p-1} \Big[\chi_l^{(l-p)}(t-r) - (-1)^{l-p}\chi_l^{(l-p)}(t+r) \Big], \quad \text{(D19)}$$

where

$$A_l^p := \frac{(l+p)!}{2^p p!(l-p)!}, \quad \text{(D20)}$$

and $\chi^{(n)}$ denotes the nth derivative of χ.

Expanding $\chi(t \pm r)$ into its Taylor series about $r = 0$, we obtain the double sum

$$\delta\phi_l(t,r) = 2\sum_{p=0}^{l}\sum_{\substack{q=0 \\ l-p+q \text{ odd}}}^{\infty}\frac{(-1)^q A_l^p}{q!}r^{q-p-1}\chi_l^{(l-p+q)}(t). \quad (D21)$$

Noting that $l-p+q \geq 0$ and odd in all terms, we parametrize q as $q = 2k+1-l+p$. We then have

$$\delta\phi_l(t,r) = 2\sum_{k=0}^{\infty}\left(\sum_{\substack{p=\max \\ (0,l-2k-1)}}^{l}\frac{(-1)^{l-p}A_l^p}{(2k+1-l+p)!}\right)r^{2k-l}\chi_l^{(k)}(t).$$
$$(D22)$$

The inner sum (in round brackets) vanishes for $k < l$, and so the outer sum starts up only at $k = l$. Hence, the expression (D19) admits a formal expansion in even (odd) powers of r at constant t, for l even (odd), starting at r^l. In particular, $\delta\psi_l(t,-r) = (-1)^l\delta\psi_l(t,r)$, and $\delta\psi_l \sim r^l$ at the origin.

Switching back from (t,r,y) to (u,r,y), we trivially find (D18) with

$$\delta\psi_l(t,r) = \sum_{p=0}^{l}A_l^p r^{-p-1}\Big[\chi_l^{(l-p)}(u)$$
$$- (-1)^{l-p}\chi_l^{(l-p)}(u+2r)\Big], \quad (D23)$$

Similarly, by setting $t = u+r$, in (D22), we trivially obtain

$$\delta\psi_l(t,r) = 2\sum_{k=0}^{\infty}\left(\sum_{\substack{p=\max \\ (0,l-2k-1)}}^{l}\frac{(-1)^{l-p}A_l^p}{(2k+1-l+p)!}\right)$$
$$\times r^{2k-l}\chi_l^{(k)}(u+r). \quad (D24)$$

5. Relating the perturbed Einstein equations in Bondi gauge to the scalar wave equation, in spherical harmonics

In the vacuum Einstein equations in Bondi gauge, linearized about flat spacetime, the coupled equations for δb and δf (D8), (D7) become

$$r^2\delta b_{,rr} + 4r\delta b_{,r} - 2(1-y^2)\delta f_{,ry} + 8y\delta f_{,r} = 0,$$
$$(D25)$$

$$r(4\delta f_{,ur} - 2\delta f_{,rr}) + 4(\delta f_{,u} - \delta f_{,r}) - r\delta b_{,ry} - 2\delta b_{,y} = 0,$$
$$(D26)$$

where we have used $x = R_0 =: r$ and $\Xi_0 R_0 = -1/2$. These equations are equivalent to Eqs. (6), (7) of [9]. In that paper solutions of these coupled equations were related to solutions of the scalar wave equation. Here we give a self-contained derivation of that relation.

a. Separating off the y-dependence

Spherical symmetry of the background means we can make the separation of variables ansatz

$$\delta b = \delta b_l(u,r)P_l^{(1)}(y), \quad (D27)$$

$$\delta f = \delta f_l(u,r)P_l^{(2)}(y), \quad (D28)$$

with the $P_l^{(1,2)}(y)$ to be found below. We obtain the four separated equations

$$P_l^{(1)\prime} = P_l^{(2)}, \quad (D29)$$

$$(1-y^2)P_l^{(2)\prime} - 4yP_l^{(2)} = \lambda P_l^{(1)}, \quad (D30)$$

$$(r^4\delta b_{l,r})_{,r} = 2\lambda r^2\delta f_{l,r}, \quad (D31)$$

$$2r(r\delta f_{l,u})_{,r} - (r^2\delta f_{l,r})_{,r} = \frac{1}{2}(r^2\delta b_l)_{,r}, \quad (D32)$$

where, without loss of generality, we have fixed a separation constant in (D29) to one. Clearly (D29) and (D30) together give

$$(1-y^2)P_l^{(1)\prime\prime} - 4yP_l^{(1)\prime} = \lambda P_l^{(1)}. \quad (D33)$$

b. Tensor spherical harmonics

We would like to relate the unfamiliar spherical harmonic-like functions $P_l^{(1)}$ and $P_l^{(2)}$ to the scalar spherical harmonics P_l. To motivate this, we note, following [39], that a geometrically natural ansatz for a perturbation δg_{ij} of the round unit metric g_{ij}^0 on S^2 with covariant derivative $\overline{\nabla}_i$ is

$$\delta g_{ij} = g_1 Y_{lm}g_{ij}^0 + g_2\overline{\nabla}_i\overline{\nabla}_j Y_{lm} \quad (D34)$$

where we have gone beyond axisymmetry but have restricted to polar perturbations, and where Y_{lm} is a scalar spherical harmonic. Similarly, the perturbation of any vector on S^2 such as the shift must take the form

$$\delta\beta^i = g_3 g_0^{ij}\overline{\nabla}_j Y_{lm}. \quad (D35)$$

Here $g_{1,2,3}$ are functions of u and r. In axisymmetry, the spherical harmonic Y_{lm} reduces to the Legendre polynomial $Y_{l0} = P_l$. Comparing (D34), (D35) with our metric ansatz (23) and separation of variables ansatz (D27), (D28) suggests

$$P_l^{(1)} = P_l', \qquad P_l^{(2)} = P_l'', \quad (D36)$$

consistent with (D29). Substituting (D36) into (D33), and using (D15) and its y-derivative, we find that (D33) is indeed obeyed with the separation constant given by

$$\lambda = -l(l+1) + 2 = -(l+2)(l-1). \quad (D37)$$

We note in passing that $P_l^{(2)}$ obeys the ODE

$$(1-y^2)P_l^{(2)\prime\prime} - 6P_l^{(2)\prime} = \tilde{\lambda}P_l^{(2)}, \quad (D38)$$

where

$$\tilde{\lambda} := -l(l+1) + 6 = -(l+3)(l-2). \quad (D39)$$

We are now left with the two coupled PDEs (D31), (D32) for δb_l and δf_l, with λ given by (D37).

c. Potential ansatz

We can solve (D31) identically by introducing the potential Ψ_l, in terms of which

$$\delta b_l(u, r) = 2\lambda \int_{r_l(u)}^{r} \frac{\Psi_l(u, \bar{r})}{\bar{r}^4} d\bar{r}, \quad (D40)$$

$$\delta f_l(u, r) = \int_{r_l(u)}^{r} \frac{\Psi_{l,r}(u, \bar{r})}{\bar{r}^2} d\bar{r}. \quad (D41)$$

$r_l(u)$ is arbitrary, but will later be set to zero. Substituting this into (D32), dividing by r and differentiating with respect to r in order to eliminate the integrals, we obtain the third-order PDE $E_3 = 0$, where

$$E_3 := -2r^3\Psi_{l,urr} + r^3\Psi_{l,rrr} - r^2\Psi_{l,rr} + \lambda(r\Psi_{l,r} - \Psi). \quad (D42)$$

We can write

$$E_3 = r^2 \left(\frac{E_2}{r}\right)_{,r}, \quad (D43)$$

where we have defined

$$E_2 := -2r^2\Psi_{l,ur} + r^2\Psi_{l,rr} + 2r\Psi_{l,u} - 2r\Psi_{l,r} + \lambda\Psi_l. \quad (D44)$$

Hence the general solution of $E_3 = 0$ is

$$E_2 = Cr, \quad (D45)$$

where C is an arbitrary constant, and the general solution of this is in turn

$$\Psi_l^{(3)}(u, r) = \Psi_l^{(2)}(u, r) + \frac{Cr}{l(l+1)}, \quad (D46)$$

where $\Psi_l^{(2)}$ is the general solution of the second-order ODE $E_2 = 0$. With the substitution

$$\Psi_l =: r^2 \delta\psi_l, \quad (D47)$$

$E_2 = 0$ for Ψ_l becomes the scalar wave equation (D17) for $\delta\psi_l$.

Substituting (D47) into (D40), (D41), we have

$$\delta b_l(u, r) = 2\lambda \int_{r_l(u)}^{r} \frac{\delta\psi_l(u, \bar{r})}{\bar{r}^2} d\bar{r}, \quad (D48)$$

$$\delta f_l(u, r) = \delta\psi_l(u, r) - \delta\psi_l(u, r_l(u))$$
$$+ 2\int_{r_l(u)}^{r} \frac{\delta\psi_l(u, \bar{r})}{\bar{r}} d\bar{r}. \quad (D49)$$

Note that, in addition to an arbitrary solution $\delta\psi_l$ of the scalar wave equation, the gravitational wave $(\delta f_l, \delta b_l)$ is parametrized also by the constant C and the function $r_l(u)$. The particular integral parametrized by the constant C is both nondynamical and gives rise to δb and δf that are singular at $r = 0$. We therefore set $C = 0$, and so obtain $E_2 = 0$ as the master equation governing (axisymmetric, polar) linear gravitational waves. We believe that $r_l(u)$ parametrizes some kind of gauge freedom, but have not tried to show this. We now set it to zero.

Using the fact that $\delta\psi_l \sim r^l$ at constant t, we see from (D48), (D49) that near the center, each spherical harmonic component of the linear gravitational wave has the spatial dependence $\delta b \sim P_l' r^{l-1}$ and $\delta f \sim P_l'' r^l$, for $l \geq 2$. In particular, near the origin $r = 0$ we therefore have $\delta b \sim ry$ and $\delta f \sim r^2$ for generic regular solutions with nonvanishing $l = 2$ components.

Substituting the general regular solution (D23) for $\delta\psi_l(u, r)$ with a sufficiently simple function $\chi_l(u)$ into (D48), (D49), and with $r_l(u) = 0$, we can carry out the integrations and so obtain $\delta b_l(u, r)$ and $\delta f_l(u, r)$ in closed form. In particular, this works when $\chi_l(u)$ is a Gaussian, and gives us a class of exact solutions for testing our code.

The general solution of (D25), (D26) is obtained by summing

$$\delta f(u, r, y) = \sum_{l=2}^{\infty} \delta f_l(u, r) P_l''(y), \quad (D50)$$

$$\delta b(u, r, y) = \sum_{l=2}^{\infty} \delta b_l(u, r) P_l'(y). \quad (D51)$$

If we now substitute (D48) and (D49), and then (D24), we see that at constant u, δf_l admits a formal expansion in even (odd) powers of r when l is even (odd), starting at r^l. This is just as for $\delta\psi_l$. Similarly, δb_l admits a formal expansion in odd (even) powers of r when l is even (odd), starting at r^{l-1}.

Obviously, the scalar field $\delta\psi$ governing the linearized gravitational waves is completely independent from the matter scalar field $\delta\psi$. We have used the same notation merely to stress that they obey identical wave equations.

d. The cases $l = 0$ and $l = 1$

Our separation of variables ansatz gives $b_0 = f_0 = f_1 = 0$ because $P'_0 = P''_0 = P''_1 = 0$. In addition the potential ansatz (D48) gives $\delta b_1 = 0$ because $\lambda = 0$. However, as $P_1 = y$ and $P'_1 = 1$, there can be nonvanishing perturbations δb_1, δH_1, even though there is no f_1.

We therefore look at the case $l = 1$ of the linearized equations in Bondi gauge without making the potential ansatz (D48), (D49). With $\delta b = \delta b_1(u, r)$ and $\delta f = 0$, (D25) reduces to

$$(r^4 \delta b_{1,r})_{,r} = 0, \tag{D52}$$

while (D26) is obeyed trivially. The solution of (D52) that is finite at the center is constant in r. The resulting regular perturbation of H is given by (D9). Putting both together, we have

$$\delta b_1(u, r) = \delta b_1(u), \qquad \delta H_1 = 2r\delta b_1(u). \tag{D53}$$

We show in Appendix E that this represents a linear gauge transformation, which physically corresponds to an acceleration $-\delta b_1(u)$ of the origin of our coordinate system along the symmetry axis.

e. Transformation to shifted Minkowski background gauge

Having found $\delta b(u, r, y)$, $\delta f(u, r, y)$ and $\delta\psi(u, r, y)$ in Bondi background gauge, we transform them to shifted Minkowski background gauge, with the coordinate change from (u, r, y) to (u, x, y) given by

$$r = \frac{x}{2}\left(1 - \frac{u}{x_0}\right), \tag{D54}$$

compare (B14). Under this coordinate transformation, only δH changes nontrivially. The metric coefficients δb, δf and R and the scalar field $\delta\psi$ change only as if they were scalar fields, and δG remains zero. We are only interested in δb, δf and $\delta\psi$, and we only need to transform their argument r to x using (D54).

f. Consistent boundary conditions for the linearized Einstein equations decomposed into spherical harmonics

We now return to the case of a nontrivial $r_l(u) > 0$. From (D48) and (D49) we read off that

$$\delta b_l(u, r_l(u)) = 0, \tag{D55}$$

$$\delta f_l(u, r_l(u)) = 0, \tag{D56}$$

$$\delta b_{l,r}(u, r_l(u)) = \frac{2\lambda}{r_l(u)^2}\delta\psi_l(u, r_l(u)), \tag{D57}$$

$$\delta f_{l,r}(u, r_l(u)) = \delta\psi_{l,r}(u, r_l(u)) + \frac{2}{r_l(u)}\delta\psi_l(u, r_l(u)). \tag{D58}$$

The first three equations gives us a class of startup conditions for integrating the linearized hierarchy equations for δb_l and δf_l that have a consistent continuum limit, in terms of a boundary condition at $r = r_l(u)$ for $\delta b_{l,r}$, which is equivalent to a boundary condition on the underlying scalar wave $\delta\psi_l$.

The simplest choice is to set $\delta b_{l,r} = 0$ at $r = r_l(u)$. As this corresponds to the homogeneous Dirichlet boundary condition $\delta\psi_l = 0$ at $r = r_l(u)$ on the underlying scalar field, this boundary condition gives rise to a well-posed PDE problem for the spherical harmonic component $\psi_l(u, r)$. We have therefore shown that, for the equations linearized about flat spacetime in linear Bondi gauge, it is consistent to impose the three boundary conditions $\delta f_l = \delta b_l = \delta b_{l,r} = 0$ at $r = r_l(u)$.

This is of course what we do, as a numerical trick, for the full nonlinear Einstein equations, with $r_l \sim l\Delta x$. What we have shown here is that this has a continuum limit for the Einstein equations linearized about Minkowski and split into spherical harmonics.

6. Solution without spherical harmonic decomposition

a. Relation between solutions for $\delta\psi$ and for $(\delta f, \delta b)$

We can remove the spherical harmonic decomposition, and obtain expressions for $\delta b(u, r, y)$ and $\delta f(u, r, y)$ in terms of a solution $\delta\psi(u, r, y)$ of the scalar wave equation. Combining (D18), (D49) and (D50), we obtain the unseparated expression

$$\delta f(u, r, y) = \delta\psi_{,yy}(u, r, y) - \delta\psi_{,yy}(u, 0, y)$$
$$+ 2\int_0^r \frac{\delta\psi_{,yy}(u, \bar{r}, y)}{\bar{r}} d\bar{r} \tag{D59}$$

for δf in terms of $\delta\psi$. Substituting (D51) into (D25) and integrating twice, we obtain the expression

$$\delta b(u,r,y) = \int_0^r \frac{1}{\bar{r}^4} \int_0^{\bar{r}} \tilde{r}^2 \left(2(1-y^2)\delta f_{,ry}(u,\tilde{r},y) \right.$$
$$\left. - 8y\delta f_{,r}(u,\tilde{r},y) \right) d\tilde{r} \tag{D60}$$

for δb in terms of δf.

b. Plane-wave solutions

The scalar wave equation admits the plane-wave solution $\delta\psi = \chi(t \pm z)$, that is

$$\delta\psi(u,r,y) = \chi(u_\pm), \qquad u_\pm := u + r(1 \pm y), \tag{D61}$$

for arbitrary functions χ. Substituting this into (D59), we can carry out the integration explicitly to obtain the plane wave solution

$$\delta f(u,r,y) = r^2\chi''(u_\pm) + 2\frac{r\chi'(u_\pm)}{1 \pm y} - 2\frac{\chi(u_\pm) - \chi(u)}{(1 \pm y)^2}, \tag{D62}$$

and substituting this into (D60), we can again carry out the integrations in closed form to obtain

$$\delta b(u,r,y) = \pm 2 \left[r(1 \mp y)\chi''(u_\pm) \right.$$
$$\left. - \frac{1 \pm 3y}{1 \pm y} \left(\chi'(u_\pm) - \chi'(u) \right) \right]. \tag{D63}$$

Note that $\psi(u,r,-y) = \psi(u,r,y)$, $f(u,r,-y) = f(u,r,y)$ and $b(u,r,-y) = -b(u,r,y)$.

To see that these expressions are regular as $y \to \mp 1$, we define the auxiliary variable $\epsilon := r(1 \pm y)$. We can then write

$$\delta\psi(u,r,y) = \chi(u + \epsilon), \tag{D64}$$

$$\delta f(u,r,y) = r^2\chi''(u+\epsilon) + 2r^2 \frac{d}{d\epsilon}\left(\frac{\chi(u+\epsilon) - \chi(u)}{\epsilon}\right),$$

$$\delta b(u,r,y) = \pm 2r \left[(1 \mp y)\chi''(u+\epsilon) \right.$$
$$\left. - (1 \pm 3y)\left(\frac{\chi(u+\epsilon) - \chi(u)}{\epsilon}\right) \right], \tag{D65}$$

where we can now see that the fraction in large round brackets is regular as $\epsilon \to 0$ if $\chi(u)$ is regular.

APPENDIX E: RESIDUAL GAUGE FREEDOM

A redefinition of the coordinates (u,x,y) leaves the form (6) of the metric invariant if $g_{xx} = g_{xy} = 0$ in the new, as in the old, coordinates. A nonlinear gauge transformation that does this and that can be written explicitly in terms of free functions is

$$u = u(\tilde{u}), \tag{E1}$$

$$y = y(\tilde{u},\tilde{y}), \tag{E2}$$

$$x = x(\tilde{u},\tilde{x},\tilde{y}). \tag{E3}$$

This means we separately relabel the outgoing null cones $\mathcal{N}_u^+$, the outgoing null rays $\mathcal{L}_{u,y,\varphi}^+$ on each outgoing cone, and the coordinate x along each outgoing ray. However, we cannot express the most general gauge freedom, which also moves the central worldline, in explicit form.

Instead, we look for the most general *infinitesimal* gauge transformation that leaves the form of the metric invariant. We make the ansatz $\delta x^\mu =: \xi^\mu$, and hence

$$\delta g_{\mu\nu} = 2\nabla_{(\mu}\xi_{\nu)}, \tag{E4}$$

where

$$\xi^\mu := (\xi^u, \xi^x, \xi^y, 0), \tag{E5}$$

are functions of (u,x,y). The general solution of $\delta g_{xx} = \delta g_{xy} = 0$ is

$$\xi^u = A(u,y), \tag{E6}$$

$$\xi^x = \xi^x(u,x,y), \tag{E7}$$

$$\xi^y = B(u,y) + A_{,y}(u,y)S\int_0^x \frac{G}{e^{2Sf}R^2}dx'. \tag{E8}$$

For $A = A(u)$ this reduces to the infinitesimal version of (E1)–(E3).

For simplicity, we now restrict the background metric to flat spacetime in Bondi gauge, with $R = x$, $G = H = 1$ and $b = f = 0$. (E8) then simplifies to

$$\xi^y = B(u,y) - \frac{SA_{,y}(u,y)}{x}. \tag{E9}$$

The resulting metric perturbation is

$$\delta G = A_{,u} + \xi^x_{,x}, \tag{E10}$$

$$\delta H = 2(A_{,u} + \xi^x_{,u}), \tag{E11}$$

$$\delta R = \xi^x + \frac{xB_{,y}}{2} - \frac{SA_{,yy} - 2yA_{,y}}{2}, \tag{E12}$$

$$\delta f = -\frac{A_{,yy}}{2x} + \frac{B_{,y} + 2yB}{2S}, \tag{E13}$$

$$\delta b = -\frac{A_{,y}}{x^2} + \frac{B_{,u}}{S} - \frac{xA_{,uy} + \xi^x_{,y}}{x^2}. \tag{E14}$$

For analysis, we relate the coordinates (u, x, y) to the cylindrical Minkowski coordinates defined by

$$t := u + x, \qquad z := -yx, \qquad \rho := \sqrt{S}x, \qquad \text{(E15)}$$

with φ completing both coordinate systems. The inverse transformation is

$$u = t - \sqrt{\rho^2 + z^2}, \qquad \text{(E16)}$$

$$x = \sqrt{\rho^2 + z^2}, \qquad \text{(E17)}$$

$$y = -\frac{z}{\sqrt{\rho^2 + z^2}}, \qquad \text{(E18)}$$

and hence the coordinate basis vectors transform as

$$\partial_t = \partial_u, \qquad \text{(E19)}$$

$$\partial_z = y\partial_u - y\partial_x - \frac{S}{x}\partial_y, \qquad \text{(E20)}$$

$$\partial_\rho = -\sqrt{S}\partial_u + \sqrt{S}\partial_x - \frac{y\sqrt{S}}{x}\partial_y, \qquad \text{(E21)}$$

with ∂_φ in both coordinate systems.

From (E6)–(E7) and (E19)–(E21), the residual gauge transformations expressed in this basis are

$$\xi = (A + \xi^x)\partial_t + \left(\sqrt{S}(\xi^x + yA_{,y}) - \frac{xyB}{\sqrt{S}} \right)\partial_\rho + \left(SA_{,y} - xB - y\xi^x \right)\partial_z. \qquad \text{(E22)}$$

ξ is a regular vector field if and only if the coefficients of $\partial_t, \partial_z, \partial_\rho$ are regular functions of (t, z, ρ). In particular, they must be finite and single-valued at the origin $x = 0$ and on the axis $S = 0$.

We now focus on gauge transformations that map the tz-plane to itself, as these include all that move the central worldline along the symmetry axis. Setting the component ξ^ρ to zero is equivalent to

$$\xi^x = \frac{xyB}{S} - yA_{,y}, \qquad \text{(E23)}$$

and substituting this back into (E22), we have

$$\xi = \left(A - yA_{,y} + \frac{xyB}{S} \right)\partial_t + \left(A_{,y} - \frac{xB}{S} \right)\partial_z. \qquad \text{(E24)}$$

The resulting metric perturbations are regular at the origin and on the axis if and only if

$$A = T(u) + Z(u)y, \qquad B = -SZ'(u), \qquad \text{(E25)}$$

where T and Z are arbitrary functions. The corresponding regular gauge vector field is

$$\xi = [T(u) - zZ'(u)]\partial_t + [Z(u) + xZ'(u)]\partial_z, \qquad \text{(E26)}$$

which explains why we have named the free coefficients T and Z. The corresponding pure gauge metric perturbations are

$$\delta H = 2(T'(u) + zZ''(u)), \qquad \text{(E27)}$$

$$\delta G = T'(u), \qquad \text{(E28)}$$

$$\delta b = -Z''(u), \qquad \text{(E29)}$$

with $\delta R = \delta f = 0$. This family includes the three Killing vector fields ∂_t, ∂_z and $z\partial_t + t\partial_z$ of Minkowski spacetime.

Restricting to $T(u) = 0$ leaves us with

$$\xi = Z'(u)z\partial_t + [Z(u) + xZ'(u)]\partial_z. \qquad \text{(E30)}$$

This is equal to $Z(u)\partial_z$ at the origin, and so is the gauge transformation that moves the origin along the symmetry axis. The resulting metric perturbation is

$$\delta H = -2Z''(u)xy, \qquad \delta b = -Z''(u), \qquad \text{(E31)}$$

with the other metric coefficients unchanged. Hence an infinitesimal noninertial motion of the origin along the symmetry axis gives exactly the $l = 1$ regular metric perturbation (D53) that we already derived from the linearized Einstein equations.

From elementary flatness at the origin, the same statement must be true for an arbitrary regular, twistfree axisymmetric spacetime. Hence we have shown that $b(u, \mathbf{0}) = 0$ is precisely the gauge condition that forces the origin of our coordinate system to move on a geodesic along the z-axis. We are imposing this as a boundary condition on (30) or (43). Conversely, by imposing an arbitrary value of $b(u, \mathbf{0})$, we could accelerate the origin during the evolution.

APPENDIX F: REGULARITY OF THE METRIC AT THE ORIGIN

Following common practice, Rinne [40] defines a scalar field as regular if it is analytic in Cartesian coordinates (t, ξ, η, z) (which in turn we assume to give rise to a chart without coordinate singularities). Transforming to the cylindrical coordinates (t, ρ, φ, z) defined by

$$\xi := \rho\cos\varphi, \qquad \eta := \rho\sin\varphi, \qquad \text{(F1)}$$

with t and z unchanged, Rinne shows that a scalar field is regular and axisymmetric if and only if it is an analytic function of (t, ρ^2, z), that is, analytic in (t, ρ, z), with only even powers of ρ.

Rinne then shows that an axisymmetric metric is regular, in the sense that its components in the Cartesian coordinate basis are analytic functions of the Cartesian coordinates, if and only if in the cylindrical coordinate basis it takes the form

$$
\begin{pmatrix}
\mathcal{A} & \mathcal{B} & \rho\mathcal{D} & \rho^2\mathcal{F} \\
\mathcal{B} & \mathcal{C} & \rho\mathcal{E} & \rho^2\mathcal{G} \\
\rho\mathcal{D} & \rho\mathcal{E} & \mathcal{H}+\rho^2\mathcal{J} & \rho^3\mathcal{K} \\
\rho^2\mathcal{F} & \rho^2\mathcal{G} & \rho^3\mathcal{K} & \rho^2(\mathcal{H}-\rho^2\mathcal{J})
\end{pmatrix}, \quad \text{(F2)}
$$

with the coordinates in order (t,z,ρ,φ), and the "Rinne coefficients" $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{E}, \mathcal{F}, \mathcal{G}, \mathcal{K}, \mathcal{H}, \mathcal{J}$ regular, that is analytic functions of (t,z,ρ^2). Clearly, the metric is twist-free if and only if $\mathcal{F}=\mathcal{G}=\mathcal{K}=0$, and we now restrict to this case.

In the following, we define a scalar field and metric as "Rinne-regular" if they obey these definitions. We shall now ask what a Rinne-regular metric and scalar field look like, first in spherical polar coordinates retaining the regular time slicing t, and secondly replacing t by a retarded time u.

In the first step, we change from the cylindrical coordinates (t,ρ,z,φ) to the spherical coordinates (t,x,y,φ) defined by

$$
\rho := x\sqrt{1-y^2}, \qquad z := -xy, \quad \text{(F3)}
$$

or equivalently

$$
x = \sqrt{\rho^2+z^2}, \qquad y = -z/\sqrt{\rho^2+z^2}, \quad \text{(F4)}
$$

with t and φ unchanged. Note that x and y are our radial and angular coordinates, not the Cartesian coordinates (ξ,η,z) defined above.

With $\rho^2 = x^2 - z^2$, a function is axisymmetric and regular if and only if it is an analytic function of (t,x^2-z^2,z), but this is the case if and only if it is an analytic function of (t,x^2,z) or, expressed entirely in spherical coordinates, of $(t,x^2,-xy)$. In particular, this means it is even in x at constant z, not constant y: this will be important below.

By transforming to the coordinate basis with respect to (t,x,y,φ), and reparametrizing the Rinne coefficients as

$$
\mathcal{B} = z\mathcal{D} - \mathcal{X}_1, \quad \text{(F5)}
$$

$$
\mathcal{E} = 2z\mathcal{J} - \mathcal{X}_2 - z\mathcal{X}_3, \quad \text{(F6)}
$$

$$
\mathcal{C} = \mathcal{H} + (2z^2-\rho^2)\mathcal{J} - 2z\mathcal{X}_2 + (\rho^2-z^2)\mathcal{X}_3, \quad \text{(F7)}
$$

we can show that the Rinne-regular twist-free axisymmetric metric is spherically symmetric if and only if $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}_3 = 0$.

Moreover, because the coefficients of the redefinitions (F5)–(F7) themselves are Rinne-regular functions, and because we cannot absorb them into redefinitions of $\mathcal{X}_{1,2,3}$ without making those singular, the reparametrized twist-free axisymmetric is Rinne-regular if *and only if* the "modified Rinne coefficients" $\mathcal{A}, \mathcal{D}, \mathcal{H}, \mathcal{J}, \mathcal{X}_{1,2,3}$ are Rinne-regular.

In the second step, we change from the spherical coordinates (t,x,y,φ) to the retarded coordinates (u,x,y,φ), where the retarded time u is defined as

$$
u(t,x,y) := t - h(t,x,y). \quad \text{(F8)}
$$

Without loss of generality, we set $h(t,0,y)=0$, and we split h into its odd and even in x (at constant z) parts as

$$
h(t,x,y) := xh_o(t,x^2,-xy) + x^2h_e(t,x^2,-xy). \quad \text{(F9)}
$$

For u to be a nontrivial retarded time, we assume that $h_o \neq 0$. Hence h is not Rinne-regular even if h_o and h_e are. The simplest choice would be $u = t - x$, the standard retarded null coordinate on Minkowski spacetime. In this case, we can invert the definition of u to give $t = u + x$, and so a Rinne-regular axisymmetric function would be an analytic function of $(u+x,x^2,-xy)$.

We now impose that the lines of constant (u,y,φ) are null, which is equivalent to $g_{xx} = g_{xy} = 0$. These are linear equations relating the modified Rinne coefficients, but their coefficients are neither even nor odd in x (at constant z), so they have Rinne-regular solutions only if we impose their even and odd parts separately, giving us four equations. Hence we see that beyond spherical symmetry we need the two generic Rinne-regular functions h_e and h_o in order to be left with five regular functions to match to our metric coefficients (G,H,R,f,b).

Which four of $\mathcal{A}, \mathcal{D}, \mathcal{H}, \mathcal{J}, \mathcal{X}_{1,2,3}, h_e$ and h_o should we solve for? In spherical symmetry the simple height function $h = x$ is already sufficiently general. This suggests that we should always solve for four of the seven modified Rinne coefficients, rather than the height function, and that this solution should be regular also in the special case $h = x$. These requirements force us to solve for $\mathcal{D}, \mathcal{H}, \mathcal{X}_1$ and $\mathcal{X}_2$.

We can now express our five metric coefficients in terms of the five remaining arbitrary regular functions $\mathcal{A}, \mathcal{J}, \mathcal{X}_3, h_e$ and h_o. To avoid square roots and logarithms in the solution, we solve not for R and f but for R^4 and e^{4Sf}. The resulting expressions are not immediately helpful, as they are explicit functions of (t,x,y), not (u,x,y), and we do not have an explicit expression for $t(u,x,y)$.

However, we can expand (G,H,R,f,b) in powers of x at constant (u,y) by using the derivative operator

$$
\left.\frac{\partial}{\partial x}\right|_{u,y} = \left.\frac{\partial}{\partial x}\right|_{t,y} - \frac{u_{,x}(t,x,y)}{u_{,t}(t,x,y)}\left.\frac{\partial}{\partial t}\right|_{x,y} \quad \text{(F10)}
$$

to find the coefficients of the Taylor series. We are then in effect inverting the height function order by order.

The Taylor coefficients are given in terms of $\mathcal{A}$, $\mathcal{J}$, $\mathcal{X}_3$, h_e and h_o and their partial derivatives with respect to (t, x^2, z), evaluated at $x = z = 0$, which are finite and independent of each other.

We do not give the series expressions here, but it is clear from their construction that G, H, R/x, f/x^2 and b are analytic functions of $(x^2, -xy)$, at constant u, with the coefficients of the Taylor series given explicitly in terms of the coefficients of the Taylor series in $(x^2, -xy)$, at constant $t = u$, of $\mathcal{A}$, $\mathcal{J}$, $\mathcal{X}_3$, h_e. This confirms the ansatz that we found to be consistent for expanding the field equations in powers of x at constant u in Sec. IV E 2.

[1] H. Bondi, M. G. J. van der Burg, and A. W. K. Metzner, Gravitational waves in general relativity. VII. Waves from axisymmetric isolated systems, Proc. R. Soc. A **269**, 21 (1962).

[2] R. K. Sachs, Gravitational waves in general relativity VIII. Waves in asymptotically flat spacetime, Proc. R. Soc. A **270**, 103 (1962).

[3] C. Gundlach, T. W. Baumgarte, and D. Hilditch, following paper, Simulations of gravitational collapse in null coordinates: II. Critical collapse of an axisymmetric scalar field, Phys. Rev. D **110**, 024019 (2024).

[4] C. Gundlach and J. M. Martín-García, Critical phenomena in gravitational collapse, Living Rev. Relativity **10**, 5 (2007).

[5] T. W. Baumgarte, B. Brügmann, D. Cors Agulló, C. Gundlach, D. Hilditch, A. Khirnov, T. Ledvinka, S. Renkhoff, and I. Suárez Fernández, Critical phenomena in the collapse of gravitational waves, Phys. Rev. Lett. **131**, 181401 (2023).

[6] C. Gundlach, this issue, Simulations of gravitational collapse in null coordinates: III. Hyperbolicity, Phys. Rev. D **110**, 024020 (2024).

[7] J. Winicour, Characteristic evolution and matching, Living Rev. Relativity **12**, 3 (2009).

[8] N. T. Bishop and L. Rezzolla, Extraction of gravitational waves in numerical relativity, Living Rev. Relativity **19**, 2 (2016).

[9] R. Gómez, P. Papadopoulos, and J. Winicour, Null cone evolution of axisymmetric vacuum spacetimes, J. Math. Phys. (N.Y.) **35**, 4184 (1994).

[10] F. Siebel, J. A. Font, E. Müller, and P. Papadopoulos, Simulating the dynamics of relativistic stars via a light-cone approach, Phys. Rev. D **65**, 064038 (2002).

[11] N. T. Bishop, R. Gómez, L. Lehner, M. Maharaj, and J. Winicour, High-powered gravitational news, Phys. Rev. D **56**, 6298 (1997).

[12] N. T. Bishop, R. Gómez, L. Lehner, M. Maharaj, and J. Winicour, Incorporation of matter into characteristic numerical relativity, Phys. Rev. D **60**, 024005 (1999).

[13] C. Reisswig, N. T. Bishop, and D. Pollney, General relativistic null-cone evolutions with a high-order scheme, Gen. Relativ. Gravit. **45**, 1069 (2013).

[14] P. J. van der Walt and N. T. Bishop, Observational cosmology using characteristic numerical relativity: Characteristic formalism on null geodesics, Phys. Rev. D **85**, 044016 (2012).

[15] J. A. Crespo, H. P. de Oliveira, and J. Winicour, The affine-null formulation of the gravitational equations: Spherical case, Phys. Rev. D **100**, 104017 (2019).

[16] T. Mädler, Affine-null metric formulation of general relativity at two intersecting null hypersurfaces, Phys. Rev. D **99**, 104048 (2019).

[17] J. Winicour, The affine-null metric formulation of the gravitational equations, Phys. Rev. D **87**, 124027 (2013).

[18] P. M. Chesler and L. G. Yaffe, Numerical solution of gravitational dynamics in asymptotically anti-de Sitter spacetimes, J. High Energy Phys. 07 (2014) 086.

[19] P. M. Chesler and L. G. Yaffe, Holography and colliding gravitational shock saves in asymptotically AdS5 spacetime, Phys. Rev. Lett. **106**, 021601 (2011).

[20] P. M. Chesler and D. A. Lowe, Nonlinear evolution of the AdS$_4$ black hole bomb, Phys. Rev. Lett. **122**, 181101 (2019).

[21] P. M. Chesler, Hairy black resonators and the AdS$_4$ super-radiant instability, Phys. Rev. D **105**, 024026 (2022).

[22] D. S. Goldwirth and T. Piran, Gravitational collapse of massless scalar field and cosmic censorship, Phys. Rev. D **36**, 3575 (1987).

[23] C. Gundlach, R. Price, and J. Pullin, Late-time behavior of stellar collapse and explosions: II. Nonlinear evolution, Phys. Rev. D **49**, 890 (1994).

[24] R. S. Hamade and J. M. Stewart, The spherically symmetric collapse of a massless scalar field, Classical Quantum Gravity **13**, 497 (1996).

[25] M. Pürrer, S. Husa, and P. C. Aichelburg, News from critical collapse: Bondi mass, tails, and quasinormal modes, Phys. Rev. D **71**, 104005 (2005).

[26] M. W. Choptuik, Universality and scaling in gravitational collapse of a massless scalar field, Phys. Rev. Lett. **70**, 9 (1993).

[27] D. Garfinkle, Choptuik scaling in null coordinates, Phys. Rev. D **51**, 5558 (1995).

[28] M. Dafermos, G. Holzegel, I. Rodnianski, and M. Taylor, The non-linear stability of the Schwarzschild family of black holes, arXiv:2104.08222.

[29] B. Porto-Veronese and C. Gundlach, Critical phenomena in gravitational collapse with competing scalar field and gravitational waves in 4 + 1 dimensions, Phys. Rev. D **106**, 104044 (2022).

[30] O. Rinne, Type II critical collapse on a single fixed grid: A gauge-driven ingoing boundary method, Gen. Relativ. Gravit. **52**, 117 (2020).

[31] C. Gundlach, T. W. Baumgarte, and D. Hilditch, Critical phenomena in gravitational collapse with two competing massless matter fields, Phys. Rev. D **100**, 104010 (2019).

[32] L. B. Szabados, Quasi-local energy-momentum and angular momentum in GR, Living Rev. Relativity **12**, 4 (2009).

[33] D. M. Eardley, Global problems in numerical relativity, in *Sources of Gravitational Radiation (Proceedings of the Battelle Seattle Workshop, July, 1978)*, edited by L. L. Smarr (Cambridge University Press, Cambridge, England, 1979).

[34] J. P. Boyd, *Chebyshev and Fourier Spectral Methods*, 2nd ed. (Dover Publications, New York 2000).

[35] C. Gundlach, J. M. Martín-García, and D. Garfinkle, Summation by parts methods for the spherical harmonic decomposition of the wave equation in arbitrary dimensions, Classical Quantum Gravity **30**, 145003 (2013).

[36] L. Ji, V. Mewes, Y. Zlochower, L. Ennoggi, F. G. Lopez Armengol, M. Campanelli, F. Cipolletta, and Z. B. Etienne, Ameliorating the Courant-Friedrichs-Lewy condition in spherical coordinates: A double FFT filter method for general relativistic MHD in dynamical spacetimes, Phys. Rev. D **108**, 104005 (2023).

[37] S. Kinoshita, On geometrical origin of Kodama vector, arXiv:2402.16484.

[38] R. H. Price, Nonspherical perturbations of relativistic gravitational collapse. I. Scalar and gravitational perturbations, Phys. Rev. D **5**, 2419 (1972).

[39] U. H. Gerlach and U. K. Sengupta, Relativistic equations for aspherical gravitational collapse, Phys. Rev. D **18**, 1789 (1978).

[40] O. Rinne, Axisymmetric general relativity, arXiv:gr-qc/0601064.