REVISTA MEXICANA DE FÍSICA S **53** (2) 5–30 FEBRERO 2007

Introducción a la relatividad numérica

M. Alcubierre
Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México,
Apartado Postal 70-543, México D.F. 04510, México

Recibido el 18 de julio de 2005; aceptado el 14 de marzo de 2005

Se presenta una introducción a los conceptos básicos de la relatividad numérica. Partiendo de una breve discusión de la relatividad general, se presenta la formulación 3+1, que es la más utilizada en cálculos numéricos. Se introducen los conceptos de foliación, métrica espacial, funciones de norma y curvatura extrínseca y se discute la separación de las ecuaciones de Einstein que dan lugar a las ecuaciones de Arnowitt-Deser-Misner (ADM). Se menciona también la existencia de formulaciones alternativas de las ecuaciones de evolución y se discute el concepto de hiperbolicidad. Finalmente se presentan las ideas principales de las aproximaciones en diferencias finitas, que son el método más comúnmente utilizado en relatividad numérica.

Descriptores: Relatividad numérica.

I present an introduction to the basic concepts of numerical relativity. Starting from a brief discussion of general relativity, I present the 3+1 formulation, which is the most widely used for numerical calculations in relativity. I introduce the concepts of foliation, spatial metric, gauge functions and extrinsic curvature, and discuss the splitting of Einstein's equations that result in the Arnowitt-Deser-Misner (ADM) equations. I also mention the existence of alternative formulations of the evolution equations and discuss the concept of hyperbolicity. Finally, I discuss the main ideas of finite difference approximations, which are the most common method used in numerical relativity.

Keywords: Numerical relativity.

PACS: 04.20.Ex; 04.25.Dm; 95.30.Sf

1. Introducción

La teoría de la relatividad general es una teoría altamente exitosa. No sólo ha modificado de manera radical nuestra visión de la gravitación, el espacio y el tiempo, sino que también posee un enorme poder predictivo. A la fecha, ha pasado con extraordinaria precisión todas las pruebas experimentales y observacionales a que se le ha sometido. Entre sus logros se encuentran la predicción de las ondas gravitacionales, la predicción de objetos exóticos como las estrellas de neutrones y los agujeros negros y el modelo cosmológico de la Gran Explosión.

La relatividad general, aunque conceptualmente sencilla y elegante, resulta ser en la práctica una teoría extremadamente compleja. Las ecuaciones de Einstein forman un sistema de diez ecuaciones diferenciales parciales en 4 dimensiones, acopladas y no lineales. Escritas en su forma más general estas ecuaciones poseen miles de términos. Debido a esta complejidad, soluciones exactas se conocen sólo en situaciones con alto grado de simetría, ya sea en el espacio o en el tiempo: soluciones con simetría esférica o axial, soluciones estáticas o estacionarias, soluciones homogéneas, isótropos etc. Si uno está interesado en estudiar situaciones con relevancia astrofísica, que involucren campos gravitacionales intensos, dinámicos y con poca o ninguna simetría, resulta imposible resolver las ecuaciones de manera exacta. De la necesidad de estudiar estos sistemas ha surgido el área de la relatividad numérica, que intenta resolver las ecuaciones de Einstein utilizando aproximaciones numéricas.

La relatividad numérica se desarrolló como un campo de investigación independiente a mediados de la década de los sesentas, comenzando con los esfuerzos pioneros de Hahn y Lindquist [1], pero no fue sino hasta finales de los setentas cuando las primeras simulaciones realmente existosas fueron realizadas por Smarr y Eppley en el contexto de la colisión de frente de dos agujeros negros [2–4]. En esa época, sin embargo, el poder de las computadoras era aún muy modesto, y las simulaciones que podían realizarse se limitaban a problemas con simetría esférica, o a lo más simetría axial a muy baja resolución. Esta situación ha cambiado, durante las décadas de los ochentas y noventas una verdadera revolución tuvo lugar en el campo de la relatividad numérica. Se han estudiado problemas cada vez más complejos en muy diversos aspectos de la teoría de la gravitación, desde la simulación de estrellas en rotación, al estudio de defectos topológicos, el colapso gravitacional y las colisiones de objetos compactos. Quizá el resultado de mayor importancia que ha obtenido la relatividad numérica sea el descubrimiento por parte de Choptuik de los fenómenos críticos en el colapso gravitacional [5, 6]. Un resumen de la historia de la relatividad numérica se pueden encontrar en la Ref. 7.

La relatividad numérica está alcanzando un estado de madurez. El desarrollo de poderosas súper computadoras, junto con el desarrollo de técnicas numéricas robustas, han permitido que se realicen simulaciones de sistemas tridimensionales con campos gravitacionales intensos y dinámicos. Toda esta actividad está ocurriendo en el momento justo. En efecto, una nueva generación de detectores de ondas gravitacionales está en estado avanzado de construcción, o incluso haciendo ya las primeras pruebas de calibración [8–11]. Las señales esperadas, sin embargo, son tan débiles que incluso con la enorme sensitividad de los nuevos detectores resultará necesario extraerlas del ruido de fondo. Es mucho más sencillo

extraer una señal enterrada en ruido si se sabe qué buscar. Es aquí donde la relatividad numérica es requerida urgentemente, proporcionando a los observadores predicciones precisas de las ondas generadas por las fuentes astrofísicas más comunes. Vivimos en verdad un momento emocionante en el desarrollo de esta disciplina.

2. La relatividad general

La teoría moderna de la gravitación es la teoría de la relatividad general postulada por Albert Einstein a fines de 1915 [12, 13]. De acuerdo a esta teoría la gravitación no es una fuerza, sino una manifestación de la "curvatura" del espacio-tiempo. Un objeto masivo produce una distorsión en la geometría del espacio-tiempo, y a su vez esta distorsión controla o altera el movimento de los objetos. Utilizando el lenguaje de John A. Wheeler, la materia le dice al espacio cómo curvarse, y el espacio le dice a la materia cómo moverse [14].

Ya al postular la teoría especial de la relatividad en 1905, Einstein sabía que la gravitación de Newton debería ser modificada. La principal razón para esto era que en la teoría de Newton la fuerza de gravedad se propagaba entre distintos objetos a velocidad infinita, lo que contradecía un principio fundamental de la relatividad: ninguna interacción física puede viajar más rápido que la luz. Es importante notar que al mismo Newton nunca le pareció convincente la existencia de esta "acción a distancia", pero consideró que era una hipótesis necesaria hasta que se encontrara una mejor explicación de la naturaleza de la gravedad. En la década de 1905 a 1915, Einstein se dedicó a buscar esa explicación. Las ideas principales que guiaron a Einstein en su camino hacia la relatividad general fueron el llamado "principio de equivalencia", que dice que todos los objetos caen exactamente de la misma forma en un campo gravitacional, y el "principio de Mach", llamado así en honor a Ernst Mach, quien lo postuló a fines del siglo XIX, y que dice que la inercia local de un objeto debe ser producida por la distribución total de la materia en el Universo. El principio de equivalencia llevó a Einstein a concluir que la gravedad debía identificarse con la geometría del espacio-tiempo, y el principio de Mach lo llevó a concluir que dicha geometría debería ser alterada por la distribución de materia y energía.

En las siguientes secciones veremos cómo se expresan estas ideas en lenguaje matemático. Antes de seguir adelante, mencionaré primero la convención de unidades que seguiré en estas notas. Utilizaré siempre las llamadas "unidades geométricas", en las que la velocidad de la luz c y la constante de la gravitación universal de Newton G son ambas iguales a la unidad. Cuando se trabaja en este sistema, la distancia, el tiempo, la masa y la energía tienen las mismas unidades (de distancia). Las unidades convencionales del sistema internacional siempre pueden recuperarse añadiendo los factores de c y G que sean necesarios en cada caso (por ejemplo, $t \to ct$ y $M \to GM/c^2$). Existen muchos libros introductorios a la relatividad general. La presentación que se da aquí está basada principalmente en las Ref. 15 a 17.

2.1. Métrica y geodésicas

Como ya hemos mencionado, el principio de equivalencia llevó a Einstein a pensar que la gravitación podía identificarse con la curvatura del espacio-tiempo. Matemáticamente esto quiere decir que la teoría de la gravedad debería ser lo que se conoce como una "teoría métrica", en la cual la gravedad se manifiesta única y exclusivamente a través de una distorsión en la geometría del espacio-tiempo.

Consideremos un espacio-tiempo de cuatro dimensiones (tres de espacio y una de tiempo). Sean x^α las coordenadas de un evento en este espacio-tiempo, donde el índice α toma los valores $\{0, 1, 2, 3\}$: cuatro números que indican en qué momento del tiempo, y en qué lugar en el espacio ocurre ese evento (en estas notas siempre tomaremos la componente 0 como aquella que se refiere al tiempo, y las componentes $\{1, 2, 3\}$ como las que se refieren al espacio; también utilizaremos la convención de que índices griegos toman valores de 0 a 3, mientras que índices latinos toman valores de 1 a 3). Entre dos eventos infinitesimalmente cercanos con coordenadas x^α y $x^\alpha + dx^\alpha$ es posible definir una "distancia invariante" ds^2 de la siguiente forma:

$$ds^2 = \sum_{\alpha,\beta=1}^{4} g_{\alpha\beta}\, dx^\alpha dx^\beta \equiv g_{\alpha\beta}\, dx^\alpha dx^\beta \,, \qquad (1)$$

donde $g_{\alpha\beta}$ se conoce como el "tensor métrico", o simplemente "la métrica", y donde la última igualdad define la "convención de suma de Einstein": índices que aparecen repetidos se suman. La distancia invariante es, como su nombre lo indica, una cantidad absoluta que no depende del sistema de coordenadas que se utilice para describir al espacio tiempo. En cada punto del espacio-tiempo, el tensor métrico es una matriz simétrica de 4×4 elementos, con eigenvalores cuyos signos son $(-,+,+,+)$, es decir, un eigenvalor negativo asociado al tiempo y tres eigenvalores positivos asociados al espacio. En la relatividad especial, el tensor métrico corresponde al de un espacio-tiempo plano, y está dado por la llamada "métrica de Minkowski":

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2 \equiv \eta_{\alpha\beta}\, dx^\alpha dx^\beta \,. \qquad (2)$$

Las transformaciones de Lorentz garantizan que el intervalo ds^2 tiene el mismo valor para cualquier observador. Esto es consecuencia directa del postulado (o mejor dicho, del hecho empírico) de la invariancia de la velocidad de la luz. Nótese que debido a la presencia de un eigenvalor negativo, la "distancia invariante" no es positiva definida. Uno puede distinguir entre eventos relacionados entre sí de 3 formas distintas:

$$ds^2 > 0 \qquad \text{intervalo espacialoide}\,, \qquad (3)$$

$$ds^2 < 0 \qquad \text{intervalo temporaloide}\,, \qquad (4)$$

$$ds^2 = 0 \qquad \text{intervalo nulo}\,. \qquad (5)$$

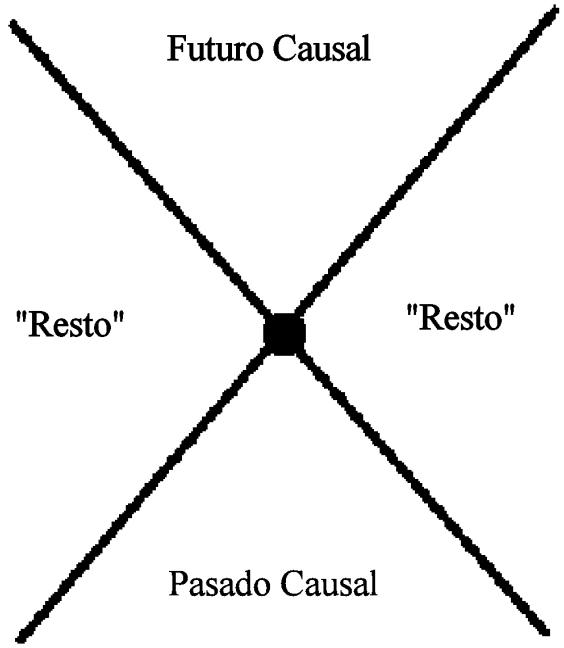

FIGURA 1. El cono de luz de un evento define las relaciones causales con otros eventos, y divide al espacio-tiempo en tres regiones: el pasado causal, el futuro causal y el "resto".

Los intervalos espacialoides corresponden a eventos separados de tal forma que un objeto tendría que moverse mas rápido que la luz para llegar de uno a otro (están separados principalmente en el "espacio"), los temporaloides corresponden a eventos donde un objeto tiene que moverse más lento que la luz para llegar de uno a otro (están separados principalmente en el "tiempo"), y los nulos a eventos que pueden alcanzarse viajando precisamente a la velocidad de la luz (la frontera entre separación espacial y temporal). Todo objeto material se mueve siguiendo trayectorias de tipo temporaloide, y la luz se mueve siguiendo trayectorias nulas. Las trayectorias nulas definen lo que se conoce como el "cono de luz" (véase Fig. 1). El cono de luz indica los eventos que pueden tener influencia física entre sí, y por lo tanto define la causalidad.

En relatividad especial, los objetos se mueven en líneas rectas en ausencia de fuerzas externas, la línea recta corresponde a una trayectoria de longitud extrema de acuerdo a la métrica de Minkowski (no necesariamente de longitud mínima debido al signo negativo de uno de los eigenvalores). Einstein postuló que en la presencia de un campo gravitacional, los objetos aún se mueven siguiendo las trayectorias más rectas posibles, es decir, trayectorias extremas, pero ahora en un espacio-tiempo curvo. A esa trayectoria extrema se le llama "geodésica". De esta forma, la gravedad no se ve como una fuerza externa, sino como una distorsión de la geometría. Dada esta distorsión, los objetos se mueven simplemente siguiendo una geodésica.

Se acostumbra parametrizar la trayectoria de un objeto utilizando el llamado "tiempo propio" $d\tau^2 := -ds^2$, que corresponde al tiempo medido por el objeto mismo (el tiempo que mide un reloj ideal atado al objeto). Nótese que este tiempo no tiene por que ser igual a dt, pues t es sólo una coordenada, es decir, una etiqueta arbitraria que asignamos a diferentes eventos. La ecuación para una geodésica está dada en general por

$$\frac{d^2x^\alpha}{d\tau^2} + \Gamma^\alpha_{\beta\gamma}\frac{dx^\beta}{d\tau}\frac{dx^\gamma}{d\tau} = 0\,, \tag{6}$$

donde las cantidades $\Gamma^\alpha_{\beta\gamma}$ se conocen como "símbolos de Christoffel" y están definidos por

$$\Gamma^\alpha_{\beta\gamma} := \frac{g^{\alpha\mu}}{2}\left[\frac{\partial g_{\beta\mu}}{\partial x^\gamma} + \frac{\partial g_{\gamma\mu}}{\partial x^\beta} - \frac{\partial g_{\beta\gamma}}{\partial x^\mu}\right]. \tag{7}$$

En la ecuación anterior introdujimos los coeficientes $g^{\alpha\beta}$ que se definen como los coeficientes de la matriz inversa a $g_{\alpha\beta}$: $g^{\alpha\mu}g_{\beta\mu} = \delta^\alpha_\beta$. La ecuación geodésica puede escribirse también como

$$\frac{du^\alpha}{d\tau} + \Gamma^\alpha_{\beta\gamma}u^\beta u^\gamma = 0\,, \tag{8}$$

donde hemos definido el vector $u^\alpha := dx^\alpha/d\tau$. Este vector se conoce como la "cuatro-velocidad", y es una generalización del concepto de velocidad ordinario.

Dado un campo gravitacional, es decir, dada la métrica del espacio-tiempo, la ecuación de las geodésicas (6) nos da la trayectoria de los objetos: el espacio-tiempo le dice a los objetos cómo moverse.

2.2. Curvatura

Como hemos visto, la métrica del espacio-tiempo nos permite obtener la trayectoria de los objetos. Sin embargo, el tensor métrico no es la forma más conveniente de describir la presencia de un campo gravitacional. Para ver esto, basta con notar que incluso en un espacio plano, uno puede cambiar la forma del tensor métrico mediante una simple transformación de coordenadas. Por ejemplo, la métrica de un espacio plano de tres dimensiones en coordenadas cartesianas $\{x, y, z\}$ esta dada simplemente por el teorema de Pitagoras:

$$dl^2 = dx^2 + dy^2 + dz^2\,, \tag{9}$$

mientras que la métrica del mismo espacio plano en coordenadas esféricas $\{r, \theta, \phi\}$ resulta ser

$$dl^2 = dr^2 + r^2 d\theta^2 + r^2\sin^2\theta d\phi^2\,, \tag{10}$$

lo que puede verse fácilmente de la transformación de coordenadas

$$x = r\sin\theta\cos\phi, \quad y = r\sin\theta\sin\phi, \quad z = r\cos\theta, \tag{11}$$

que implica

$$dx = dr\,\sin\theta\,\cos\phi - r\,\sin\theta\,\sin\phi\,d\phi + r\,\cos\theta\,\cos\phi\,d\theta\,, \tag{12}$$

$$dy = dr\,\sin\theta\,\sin\phi + r\,\sin\theta\,\cos\phi\,d\phi + r\,\cos\theta\,\sin\phi\,d\theta\,, \tag{13}$$

$$dz = dr\,\cos\theta - r\,\sin\theta\,d\theta\,. \tag{14}$$

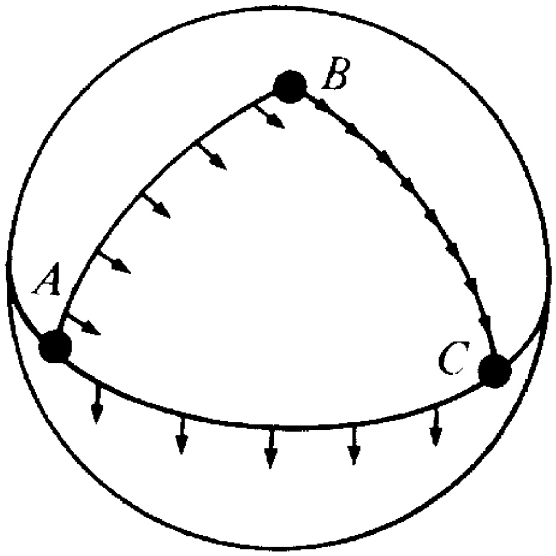

FIGURA 2. Transporte paralelo en la superficie terrestre.

A partir de la métrica (10) es posible calcular los símbolos de Christoffel para las coordenadas esféricas y ver que la ecuación de una linea recta escrita en estas coordenadas ya no es trivial. Llevando a cabo tranformaciones de coordenadas aún más elaboradas es posible terminar con una métrica mucho más compleja.

Debemos encontrar entonces una forma de distinguir con certeza entre un espacio plano y uno que no lo es. La manera de hacer esto es a través del llamado "tensor de curvatura de Riemann". Este tensor mide el cambio de un vector al transportarlo alrededor de un circuito manteniéndolo siempre paralelo a sí mismo ("transporte paralelo"). En un espacio plano, el vector no cambia al hacer esto, mientras que en un espacio curvo sí lo hace. Esto puede verse si uno mueve un vector en la supericie terrestre. Si comenzamos en el ecuador con un vector apuntando al este, nos movemos hasta el polo norte siguiendo un meridiano, bajamos siguiendo otro meridiano que haga un ángulo recto hacia el este con el primero hasta llegar de nuevo al ecuador, y luego seguimos el ecuador hasta volver al punto original, descubrimos que el vector ahora apunta al sur (véase la Fig. 2).

En estas notas no derivaremos el tensor de Riemann de primeros principios, nos limitaremos sólo a escribirlo:

$$R^{\sigma}{}_{\mu\nu\rho} := \partial_\nu \Gamma^\sigma_{\mu\rho} - \partial_\mu \Gamma^\sigma_{\nu\rho} + \left(\Gamma^\alpha_{\mu\rho}\Gamma^\sigma_{\alpha\nu} - \Gamma^\alpha_{\nu\rho}\Gamma^\sigma_{\alpha\mu}\right), \quad (15)$$

donde ∂_μ es una abreviación de $\partial/\partial x^\mu$, y donde $\Gamma^\alpha_{\mu\nu}$ son los símbolos de Christoffel definidos anteriormente. Nótese que el tensor de Riemann tiene 4 índices, es decir, $4{\times}4{\times}4{\times}4{=}256$ componentes. Sin embargo, tiene muchas simetrías, por lo que solamente tiene 20 componentes independientes. Es posible demostrar que el tensor de Riemann es igual a cero si y sólo si el espacio-tiempo es plano. A partir del tensor de Riemann podemos definir el llamado "tensor de Ricci" como

$$R_{\mu\nu} := \sum_\lambda R^\lambda{}_{\mu\lambda\nu} = R^\lambda{}_{\mu\lambda\nu}\,. \quad (16)$$

Nótese que el hecho de que el tensor de Ricci sea cero *no* significa que el espacio sea plano.

Es importante hacer notar que en algunas ocasiones hemos escrito tensores con los índices arriba y en otras con los índices abajo. Esto no es un error tipográfico, los tensores con índices arriba o abajo no son iguales, pero están relacionados entre sí. La regla es la siguiente: los índices de objetos geométricos se suben y bajan contrayendo con el tensor métrico $g_{\mu\nu}$ o su inversa $g^{\mu\nu}$. Por ejemplo:

$$v_\mu = g_{\mu\nu}\, v^\nu\,, \qquad v^\mu = g^{\mu\nu}\, v_\nu\,,$$
$$T^{\mu\nu} = g^{\mu\alpha} g^{\nu\beta}\, T_{\alpha\beta}\,, \qquad R_{\sigma\mu\nu\rho} = g_{\sigma\lambda} R^\lambda{}_{\mu\nu\rho}.$$

2.3. Base coordenada y derivadas covariantes

Cuando se considera el cambio de un campo vectorial (objetos con un índice) o tensorial (objetos con más de un índice) al moverse en el espacio-tiempo, debemos tomar en cuenta que al movernos de un punto a otro no solo las componentes de dichos objetos pueden cambiar, sino que también la base en las que dichas componentes se miden cambia de un punto a otro. En efecto, cuando consideramos las componentes de un objeto geométrico (vector o tensor), dichas componentes siempre están dadas con respecto a una base específica.

En un espacio plano tridimensional en coordenadas cartesianas, la base es en general la llamada "base canónica": $\vec{i} = (1,0,0), \vec{j} = (0,1,0), \vec{k} = (0,0,1)$. En un espacio curvo, o un espacio plano en coordenadas no triviales, es necesario elegir una base antes de poder hablar de las componentes de un objeto geométrico. Una base común (aunque no la única) es la llamada "base coordenada", en la que se toman como elementos de la base a aquellos vectores que tienen una componente igual a 1 a lo largo de la dirección de una coordenada dada, y componentes iguales a 0 en las otras direcciones.

Por ejemplo, en un espacio plano en coordenadas esféricas $\{r,\theta,\phi\}$ la base coordenada es

$$\vec{e}_r = (1,0,0)\,, \quad \vec{e}_\theta = (0,1,0)\,, \quad \vec{e}_\phi = (0,0,1)\,. \quad (17)$$

Escritos en coordenadas cartesianas, estos vectores resultan ser

$$\vec{e}_r = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)\,, \quad (18)$$
$$\vec{e}_\theta = (r\cos\theta\cos\phi, r\cos\theta\sin\phi, -r\sin\phi)\,, \quad (19)$$
$$\vec{e}_\phi = (-r\sin\theta\sin\phi, r\sin\theta\cos\phi, 0)\,. \quad (20)$$

Nótese que estos vectores no son todos unitarios, sus magnitudes son

$$|\vec{e}_r|^2 = (\sin\theta\cos\phi)^2 + (\sin\theta\sin\phi)^2 + (\cos\theta)^2{=}1 \quad (21)$$
$$|\vec{e}_\theta|^2 = (r\cos\theta\cos\phi)^2 + (r\cos\theta\sin\phi)^2 + (r\sin\phi)^2 = r^2, \quad (22)$$
$$|\vec{e}_\phi|^2 = (r\sin\theta\sin\phi)^+(r\sin\theta\cos\phi)^2 = r^2\sin^2\theta, \quad (23)$$

donde para obtener la magnitud sumamos los cuadrados de las componentes cartesianas.

Las magnitud de un vector puede calcularse directamente a partir de las componentes esféricas utilizando el tensor métrico. Para cualquier vector $\vec{v}$, la magnitud se define en términos del tensor métrico como

$$|\vec{v}|^2 = g_{\alpha\beta}\, v^\alpha v^\beta \,. \tag{24}$$

En general, el producto escalar de dos vectores $\vec{v}$ y $\vec{u}$ se define como

$$\vec{v}\cdot\vec{u} = g_{\alpha\beta}\, v^\alpha u^\beta \,. \tag{25}$$

Debido a que los vectores de la base coordenada tienen por definición componentes 1 para la coordenada correspondiente y 0 para las demas, la definición anterior implica

$$\vec{e}_\alpha \cdot \vec{e}_\beta = g_{\alpha\beta} \,. \tag{26}$$

No es difícil ver que si utilizamos la expresión (24) y la métrica del espacio plano en coordenadas esféricas, obtenemos las mismas magnitudes para los vectores de la base coordenada esférica que escribimos arriba.

Es importante notar que en las expresiones anteriores, v^α se refiere a la componente del vector $\vec{v}$ respecto a la coordenada x^α, mientras que $\vec{e}_\alpha$ se refiere al vector de la base coordenada que apunta en la dirección x^α. Esto implica que, en general,

$$\vec{v} = v^\alpha\, \vec{e}_\alpha; \tag{27}$$

ecuación que expresa a $\vec{v}$ como una combinación lineal de los elementos de la base $\{\vec{e}_\alpha\}$ con coeficientes v^α.

Consideremos ahora el cambio de un vector a lo largo de una coordenada:

$$\frac{\partial\vec{v}}{\partial x^\alpha} = \frac{\partial}{\partial x^\alpha}\,\left(v^\beta \vec{e}_\beta\right) = \frac{\partial v^\beta}{\partial x^\alpha}\,\vec{e}_\beta + v^\beta\,\frac{\partial\vec{e}_\beta}{\partial x^\alpha} \,. \tag{28}$$

Esta ecuación muestra cómo la derivada de un vector es más que la simple derivada de sus componentes, debemos también tomar en cuenta el cambio en los vectores de la base.

La derivada $\partial\vec{e}_\beta/\partial x^\alpha$ es a su vez un vector, por lo que puede expresarse como combinación lineal de los vectores de la base. Introducimos los símbolos $\Gamma^\mu_{\alpha\beta}$ para denotar los coeficientes de dicha combinación lineal:

$$\frac{\partial\vec{e}_\beta}{\partial x^\alpha} = \Gamma^\mu_{\alpha\beta}\,\vec{e}_\mu \,. \tag{29}$$

Los coeficientes $\Gamma^\mu_{\alpha\beta}$ son precisamente los símbolos de Christoffel que introdujimos anteriormente. Utilizando estos coeficientes tendremos

$$\frac{\partial\vec{v}}{\partial x^\alpha} = \frac{\partial v^\beta}{\partial x^\alpha}\,\vec{e}_\beta + v^\beta\,\Gamma^\mu_{\alpha\beta}\,\vec{e}_\mu \,. \tag{30}$$

Cambiando el nombre de los índices sumados, podemos reescribir esto como

$$\frac{\partial\vec{v}}{\partial x^\alpha} = \left(\frac{\partial v^\beta}{\partial x^\alpha} + v^\mu\,\Gamma^\beta_{\alpha\mu}\right)\vec{e}_\beta \,. \tag{31}$$

Esta ecuación nos da las componentes del vector $\partial\vec{v}/\partial x^\alpha$. Definimos la "derivada covariante" del vector $\vec{v}$ como

$$v^\alpha{}_{;\beta} \equiv \nabla_\beta\, v^\alpha := \frac{\partial v^\alpha}{\partial x^\beta} + v^\mu\,\Gamma^\alpha_{\beta\mu} \,. \tag{32}$$

Nótese que hemos introducido dos notaciones distintas para la derivada covariante: En un caso con un punto y coma, y en otro con el símbolo de gradiente (nabla). Ambas notaciones son comunes y se utilizan indistintamente.

La derivada covariante nos dice como cambian las componentes de un vector al movernos de un punto a otro, e incluye el cambio de los elementos de la base. La derivada covariante se reduce a la derivada parcial cuando los símbolos de Christoffel son cero, lo que ocurre en un espacio plano en coordenadas cartesianas, pero no en coordenadas esféricas. Sin embargo, siempre es posible encontrar una tranformación de coordenadas para la cual los símbolos de Christoffel son iguales a cero en un punto dado (pero no en otros puntos excepto si el espacio es plano). Esto se debe a que todo espacio curvo es "localmente plano": en una región infinitesimalmente cercana a todo punto la geometría se aproxima a la plana (es por eso que los mapas de una ciudad en un plano son muy buenos, mientras que los mapas del mundo entero introducen distorsiones serias).

La derivada covariante de un vector con los índices abajo v_α resulta ser

$$v_{\alpha;\beta} = \frac{\partial v_\alpha}{\partial x^\beta} - \Gamma^\mu_{\alpha\beta} v_\mu \,. \tag{33}$$

El mismo concepto de derivada covariante puede extenderse a tensores de muchas componentes, la regla es añadir un término con símbolos de Christoffel por cada índice libre, con el signo adecuado dependiendo de si el índice esta arriba o abajo. Por ejemplo:

$$\begin{aligned} T^{\mu\nu}{}_{;\alpha} &= \partial_\alpha T^{\mu\nu} + \Gamma^\mu_{\alpha\beta}\, T^{\beta\nu} + \Gamma^\nu_{\alpha\beta}\, T^{\mu\beta} \,, \\ T_{\mu\nu;\alpha} &= \partial_\alpha T_{\mu\nu} - \Gamma^\beta_{\alpha\mu}\, T_{\beta\nu} - \Gamma^\beta_{\alpha\nu}\, T_{\mu\beta} \,, \\ T^\mu{}_{\nu;\alpha} &= \partial_\alpha T^\mu{}_\nu + \Gamma^\mu_{\alpha\beta}\, T^\beta{}_\nu - \Gamma^\beta_{\alpha\nu}\, T^\mu{}_\beta \,. \end{aligned}$$

Utilizando esta regla es posible mostrar que la derivada covariante del tensor métrico es cero:

$$g_{\mu\nu;\alpha} = 0 \,, \qquad g^{\mu\nu}{}_{;\alpha} = 0 \,, \tag{34}$$

lo que implica que la operación de subir y bajar índices conmuta con las derivadas covariantes:

$$v^\mu{}_{;\alpha} = \left(g^{\mu\nu}\, v_\nu\right)_{;\alpha} = g^{\mu\nu}\,\left(v_{\nu;\alpha}\right) \,. \tag{35}$$

2.4. Las ecuaciones de Einstein

El elemento que aún nos falta en la teoría de la gravitación de Einstein es aquel que nos dice cómo se relaciona la geometría del espacio-tiempo con la distribución de materia

y energía. Este elemento final está contenido en las "ecuaciones de campo de Einstein", o simplemente las "ecuaciones de Einstein". Estas ecuaciones pueden derivarse de diversas maneras, ya sea buscando una generalización relativista y consistente de la ley de la gravitación de Newton (el camino que siguió Einstein), o derivándolas de manera formal a partir de un lagrangiano (el camino que siguió Hilbert casi simultáneamente). Aquí nos limitaremos a escribirlas sin derivación. En su forma más compacta, las ecuaciones de Einstein tienen la forma

$$G_{\mu\nu} = 8\pi T_{\mu\nu}\,, \tag{36}$$

donde $G_{\mu\nu}$ es el "tensor de Einstein" que está relacionado con el tensor de curvatura de Ricci, y $T_{\mu\nu}$ es el "tensor de energía-momento" de la materia. Es decir, el lado izquierdo representa la geometría del espacio-tiempo y el lado derecho la distribución de materia y energía. El factor de 8π es simplemente una normalización necesaria para obtener el límite newtoniano correcto. Nótese que hay 10 ecuaciones independientes en la expresión anterior, pues los índices μ y ν toman valores de 0 a 3, lo que da $4 \times 4 = 16$ ecuaciones, pero los tensores de Einstein y de energía-momento son simétricos, lo que reduce a 10 el número de ecuaciones independientes.

Las ecuaciones de Einstein que acabamos de escribir no podrían parecer más simples. Esta simplicidad, sin embargo, es sólo aparente, pues cada término es en realidad una abreviación que representa objetos muy complejos. Escritas en su manera más extensa, en un sistema de coordenadas arbitrario, las ecuaciones de Einstein son 10 ecuaciones diferenciales parciales acopladas en 4 coordenadas, y tienen miles de términos.

Consideremos ahora cada término en las ecuaciones de Einstein por separado. El tensor de Einstein se define en términos del tensor de Ricci como

$$G_{\mu\nu} := R_{\mu\nu} - \frac{1}{2}\, g_{\mu\nu} R\,, \tag{37}$$

con $R := g^{\mu\nu}\, R_{\mu\nu}$ la traza del tensor de Ricci, tambien llamada el "escalar de curvatura".

La segunda parte de las ecuaciones de Einstein, el tensor de energía-momento, describe la densidad de energía, la densidad de momento y el flujo de momento de un campo de materia ($i, j = 1, 2, 3$):

$$T^{00} = \text{densidad de energía}\,, \tag{38}$$

$$T^{0i} = \text{densidad de momento}\,, \tag{39}$$

$$T^{ij} = \text{flujo de momento } i \text{ a través de la superficie } j. \tag{40}$$

Por ejemplo, para un fluido perfecto sin presión (llamado "polvo") en un espacio plano tenemos

$$T^{00} = \rho/(1 - v^2)\,, \tag{41}$$

$$T^{0i} = \rho\, v^i/(1 - v^2)\,, \tag{42}$$

$$T^{ij} = \rho\, v^i v^j/(1 - v^2)\,, \tag{43}$$

donde ρ es la densidad de energía en el marco de referencia de un elemento de fluido, v^i es el campo de velocidad del fluido y $v^2 = v_x^2 + v_y^2 + v_z^2$.

Antes de terminar con nuestra discusión de las ecuaciones de Einstein, hace falta un último comentario. En el caso del espacio vacío, el tensor de energía-momento es cero, y las ecuaciones de Einstein se reducen a

$$G_{\mu\nu} = 0\,, \tag{44}$$

o de forma equivalente

$$R_{\mu\nu} = 0\,. \tag{45}$$

Nótese que, como ya hemos mencionado, el hecho de que el tensor de Ricci sea cero no significa que el espacio sea plano. Esto es como debe ser, pues sabemos que el campo gravitacional de un objeto se extiende más alla del objeto mismo, por lo que la curvatura del espacio en una región del vacío cercana a un objeto masivo no puede ser cero. Las ecuaciones de Einstein en el vacío tienen otra aplicación importante, describen la forma en la que el campo gravitacional se propaga en el vacío y, de manera análoga a las ecuaciones de Maxwell, predicen la existencia de las ondas gravitacionales: perturbaciones del campo gravitacional que viajan a la velocidad de la luz. La predicción de la existencia de las ondas gravitacionales nos dice que en la teoría de Einstein las interacciones gravitacionales no se propagan a velocidad infinita, sino que lo hacen a la velocidad de la luz.

2.5. Identidades de Bianchi y leyes de conservación

El tensor de curvatura de Riemann tiene la siguiente propiedad:

$$R_{\alpha\beta\mu\nu;\lambda} + R_{\alpha\beta\lambda\mu;\nu} + R_{\alpha\beta\nu\lambda;\mu} = 0\,. \tag{46}$$

A la ecuación anterior se le conoce como las "identidades de Bianchi". Estas identidades resultan muy importantes en la relatividad general. Una de sus consecuencias es el hecho de que la divergencia covariante del tensor de Einstein es igual a cero:

$$G^{\mu\nu}{}_{;\nu} = 0\,. \tag{47}$$

El tensor de Einstein es la única combinación que puede obtenerse a partir del tensor de Ricci que tiene esta propiedad, y es precisamente por esto que las ecuaciones de Einstein involucran a este tensor y no al tensor de Ricci directamente. Si utilizamos las ecuaciones de Einstein vemos que la propiedad anterior implica

$$T^{\mu\nu}{}_{;\nu} = 0\,. \tag{48}$$

Esta ecuación (4 ecuaciones en realidad) es de fundamental importancia, pues representa las leyes locales de conservación de energía y momento, y garantiza que la perdida de energía y momento en una región está compensada por el flujo de energía y momento fuera de dicha región. Podemos ver cómo las ecuaciones de Einstein implican automáticamente a las leyes de conservación de energía y momento.

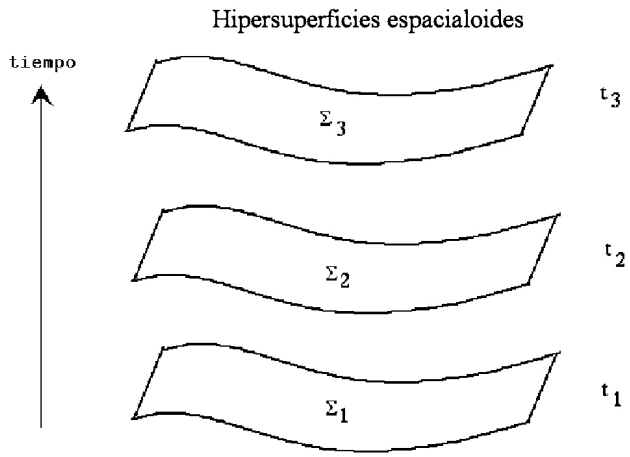

FIGURA 3. Foliación del espacio-tiempo en hipersuperficies tridimensionales de tipo espacialoide.

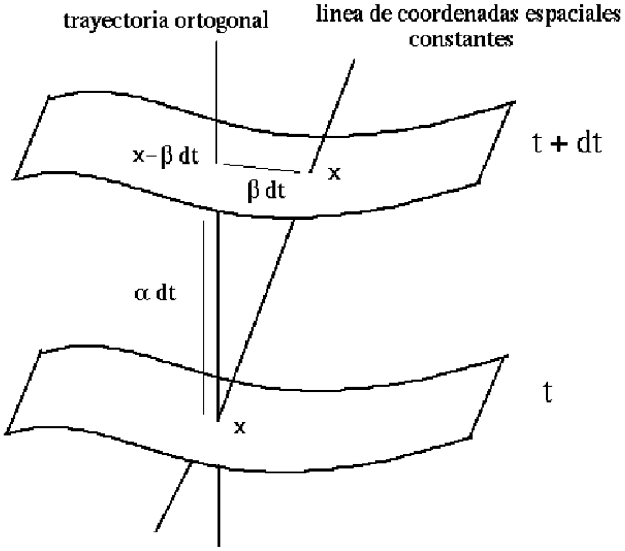

FIGURA 4. Dos hipersuperficies espacialoides infinitesimalmente cercanas.

3. El formalismo de la relatividad numérica

Existen diferentes formalismos utilizados en la relatividad numérica. En cada caso, lo que debe hacerse es separar las ecuaciones de campo de Einstein de forma tal que podamos dar ciertas condiciones iniciales, y a partir de ellas obtener la evolución del campo gravitacional. Los diferentes formalismos difieren en la forma específica en que se hace esta separación. En estas notas nos limitaremos a estudiar el llamado "formalismo 3+1", en donde se hace una separación entre las tres dimensiones del espacio, por un lado, y el tiempo, por otro. El formalismo 3+1 es el más comúnmente utilizado en la relatividad numérica, pero no es el único. Formalismos alternativos son el llamado "formalismo característico" [18], donde el espacio-tiempo se separa en conos de luz centrados en una cierta trayectoria temporaloide; y el "formalismo conforme" [19], donde se utiliza una transformación que permite tener la frontera del espacio-tiempo, es decir, el "infinito asintótico", en una región finita del sistema de coordenadas. Los diferentes formalismos tienen ventajas y desventajas dependiendo del tipo de sistema físico que se quiera estudiar.

En las siguientes secciones veremos una introducción al formalismo 3+1 de la relatividad general. La discusión que aquí se presenta puede verse con mayor detalle en las Ref. 15 y 20.

3.1. El formalismo 3+1

Para estudiar la evolución en el tiempo de cualquier sistema físico, lo primero que debe hacerse es formular dicha evolución como un "problema de valores iniciales" o "problema de Cauchy": dadas condiciones iniciales adecuadas, las ecuaciones fundamentales deben poder predecir la evolución futura (o pasada) del sistema.

Al intentar escribir las ecuaciones de Einstein como un problema de Cauchy nos enfrentamos inmediatamente a un problema: las ecuaciones de Einstein están escritas en forma tal que el espacio y el tiempo son simétricos y juegan papeles equivalentes. Esta "covariancia" de las ecuaciones es importante (y elegante) desde el punto de vista teórico, pero no permite pensar claramente en la evolución en el tiempo del campo gravitacional. Lo primero que debemos hacer entonces para reescribir las ecuaciones de Einstein como un problema de Cauchy es separar los papeles del espacio y el tiempo de forma clara. A la formulación de la relatividad general que resulta de esta separación se le conoce como el "formalismo 3+1".

Consideremos un espacio-tiempo con una métrica $g_{\alpha\beta}$. Si dicho espacio-tiempo puede ser totalmente "foliado" (es decir, separado en cortes tridimensionales) de tal forma que cada hoja tridimensional sea de tipo espacialoide, entonces se dice que dicho espacio-tiempo es "globalmente hiperbólico". Un espacio-tiempo globalmente hiperbólico no posee curvas temporaloides cerradas, lo que significa que no permite viajes hacia atrás en el tiempo. No todos los espacio-tiempos posibles tienen esta propiedad, pero en el formalismo 3+1 se asume que los espacio-tiempos físicamente razonables son de este tipo.

Una vez que se tiene un espacio-tiempo globalmente hiperbólico, por definición podemos foliarlo en una serie de hipersuperficies de tipo espacialoide (ver Fig. 3). Esta foliación en general no es única. Definimos el parámetro t como aquel que identifica a las distintas hojas de la foliación; t puede considerarse entonces como un "tiempo universal" (pero cuidado, t no tiene por qué coincidir con el tiempo propio de nadie). Consideremos ahora una cierta foliación, y tomemos dos hipersuperficies infinitesimalmente cercanas Σ_t y Σ_{t+dt}. La geometría de la región del espacio-tiempo contenida entre ambas hipersuperficies puede determinarse a partir de los siguientes 3 ingredientes (véase Fig. 4):

- La métrica tridimensional γ_{ij} $(i,j=1,2,3)$ que mide las distancias dentro de la hipersuperficie misma:

$$dl^2 = \gamma_{ij}\, dx^i dx^j \,. \tag{49}$$

- El "lapso" de tiempo propio α entre ambas hipersuperficies que mide un observador que se mueve en la dirección normal a ellas (observador de Euler):

$$d\tau = \alpha(t, x^i)\, dt\,. \tag{50}$$

- La velocidad relativa β^i entre los observadores de Euler y las líneas con coordenadas espaciales constantes:

$$x^i_{t+dt} = x^i_t - \beta^i(t, x^j)\, dt, \qquad \text{para observadores de Euler} \tag{51}$$

  Al vector β^i se le llama el "vector de corrimiento".

Nótese que la manera en la que se hace la foliación no es única, y de la misma forma, la manera en la que se propaga el sistema de coordenadas espacial de una superficie a otra tampoco lo es. Esto significa que tanto la función de lapso α como el vector de corrimiento β^i son funciones que pueden especificarse libremente. Estas funciones determinan nuestro sistema de coordenadas y se conocen como "funciones de norma".

En términos de las funciones $\{\alpha, \beta^i, \gamma_{ij}\}$, la métrica del espacio-tiempo toma la siguiente forma:

$$ds^2 = \left(-\alpha^2 + \beta_i\beta^i\right) dt^2 + 2\,\beta_i\, dt dx^i + \gamma_{ij}\, dx^i dx^j\,, \tag{52}$$

donde hemos definido $\beta_i := \gamma_{ij}\,\beta^j$.

En forma explícita tenemos

$$g_{\mu\nu} = \begin{pmatrix} -\alpha^2 + \beta_k\beta^k & \beta_i \\ \beta_j & \gamma_{ij} \end{pmatrix},$$
$$g^{\mu\nu} = \begin{pmatrix} -1/\alpha^2 & \beta^i/\alpha^2 \\ \beta^j/\alpha^2 & \gamma^{ij} - \beta^i\beta^j/\alpha^2 \end{pmatrix}. \tag{53}$$

De la misma forma, las componentes del vector normal unitario n^μ a las hipersuperficies resultan ser

$$n^\mu = \left(\frac{1}{\alpha}, \frac{-\beta^i}{\alpha}\right), \quad n_\mu = (-\alpha, 0)\,, \quad n^\mu n_\mu = -1. \tag{54}$$

Este vector corresponde por definición a la cuatro-velocidad de los observadores de Euler.

3.2. Curvatura intrínseca y curvatura extrínseca

Al hablar de las hipersuperficies espaciales que forman la foliación del espacio-tiempo, debemos distinguir entre la curvatura "intrínseca" de dichas hipersuperficies proveniente de su geometría interna, y su curvatura "extrínseca" relacionada con la forma en que éstas se encuentran inmersas en el espacio-tiempo de 4 dimensiones.

La curvatura intrínseca estará dada por el tensor de Ricci tridimensional que se define en términos de la métrica espacial γ_{ij}. La curvatura extrínseca, por otro lado, se define en términos de lo que le ocurre al vector normal n^α al transportarlo paralelamente de un sitio a otro de la superficie. En general encontraremos que al tranportar paralelamente este vector a un punto cercano, el nuevo vector ya no será normal a la superficie. El "tensor de curvatura extrínseca" $K_{\alpha\beta}$ es una medida del cambio en el vector normal bajo transporte paralelo. Para definir el tensor de curvatura extrínseca es necesario primero introducir el operador de "proyección" P^α_β a las hipersuperficies espaciales:

$$P^\alpha_\beta := \delta^\alpha_\beta + n^\alpha n_\beta\,, \tag{55}$$

Es fácil mostrar que para cualquier vector v^α tenemos (recuerdese que $n^\alpha n_\alpha = -1$)

$$\left(P^\alpha_\beta\, v^\beta\right) n_\alpha = 0\,, \tag{56}$$

es decir, todo vector proyectado a la hipersuperficie es ortogonal a n^α. Es posible también proyectar tensores con varios índices, basta contraer todos los índices libres con el operador de proyección:

$$P\,T_{\alpha\beta} \equiv P^\mu_\alpha P^\nu_\beta\, T_{\mu\nu}\,. \tag{57}$$

Usando el operador de proyección, el tensor de curvatura extrínseca se define como

$$K_{\alpha\beta} := -P\,\nabla_\alpha n_\beta\,, \tag{58}$$

donde el símbolo P significa proyectar todos los índices libres a la hipersuperficie espacial, y donde ∇_α es la derivada covariante, que como hemos visto está dada en términos de los símbolos de Christoffel como

$$\nabla_\alpha n_\beta = \partial_\alpha n_\beta - \Gamma^\lambda_{\alpha\beta}\, n_\lambda\,, \tag{59}$$

A partir de la definición anterior es posible mostrar que el tensor de curvatura extrínseca es simétrico ($K_{\alpha\beta} = K_{\beta\alpha}$), y que tiene la siguiente propiedad:

$$n^\alpha K_{\alpha\beta} = 0\,, \tag{60}$$

lo que significa que el tensor $K_{\alpha\beta}$ es "puramente espacial". Debido a esto consideraremos de ahora en adelante sólo sus componentes espaciales K_{ij}.

Substituyendo la forma explícita del vector normal (54) en la definición de la curvatura extrínseca es posible demostrar que K_{ij} está dado en términos de la métrica espacial como

$$K_{ij} = \frac{1}{2\alpha}\left[\partial_t\gamma_{ij} + D_i\beta_j + D_j\beta_i\right], \tag{61}$$

donde D_i se refiene a la derivada convariante con respecto a la métrica espacial γ_{ij} (que difiere de la derivada covariante con respecto a la métrica del espacio-tiempo $g_{\mu\nu}$ representada por ∇_μ). La ecuación anterior puede reescribirse como

$$\partial_t\gamma_{ij} = -2\alpha K_{ij} + D_i\beta_j + D_j\beta_i\,. \tag{62}$$

La curvatura extrínseca K_{ij} está entonces relacionada con el cambio en el tiempo de la métrica espacial.

Esto nos permite llegar a la mitad del camino necesario para reescribir la relatividad general como un problema de Cauchy: ya tenemos una ecuación de evolución para γ_{ij}. Pero para cerrar el sistema aún nos falta una ecuación de evolución para K_{ij}. Es importante notar que hasta ahora hemos trabajado sólo con conceptos geométricos y no hemos utilizado las ecuaciones de campo de Einstein. Es precisamente a partir de las ecuaciones de Einstein de donde obtendremos la ecuación de evolución para K_{ij}. Dicho de otra forma, la ecuación de evolución (62) para γ_{ij} es meramente cinemática, mientras que la ecuación de evolución para K_{ij} contendrá la información dinámica.

3.3. Las ecuaciones de Einstein en el formalismo 3+1

En la sección anterior definimos el operador de proyección P^α_β a las hipersuperficies espaciales. Utilizando este operador podemos separar las ecuaciones de Einstein en 3 grupos:

- Proyección normal (1 ecuación):

$$n^\alpha n^\beta \left(G_{\alpha\beta} - 8\pi T_{\alpha\beta}\right) = 0\,. \tag{63}$$

- Proyección a la hipersuperficie (6 ecuaciones):

$$P\ \left(G_{\alpha\beta} - 8\pi T_{\alpha\beta}\right) = 0\,. \tag{64}$$

- Proyección mixta (3 ecuaciones):

$$P\ \left[n^\alpha \left(G_{\alpha\beta} - 8\pi T_{\alpha\beta}\right)\right] = 0\,. \tag{65}$$

Para expresar estos conjuntos de ecuaciones en el lenguaje 3+1, es necesaria un álgebra larga que no haremos en estas notas. Aquí nos limitaremos a señalar los resultados finales.

De la proyección normal obtenemos la siguiente ecuación:

$$R^{(3)} + \left(\mathrm{tr\,K}\right)^2 - K_{ij}K^{ij} = 16\pi\rho, \tag{66}$$

donde $R^{(3)}$ es el escalar de curvatura de la métrica espacial, $\mathrm{tr\,K} \equiv \gamma^{\mathrm{ij}}\mathrm{K_{ij}}$ es la traza del tensor de curvatura extrínseca y ρ es la densidad de energía de la materia medida por los observadores de Euler:

$$\rho := n_\alpha n_\beta\, T^{\alpha\beta}\,. \tag{67}$$

La Ec. (66) no tiene derivadas temporales (aunque sí tiene derivadas espaciales de γ_{ij} dentro del escalar de Ricci). Debido a esto no es una ecuación de evolución sino una "constricción" o "ligadura" (a veces también llamada "vínculo") del sistema. Como está relacionada con la densidad de energía ρ, a esta constricción se le conoce como la "constricción de energía" o la "constricción hamiltoniana".

De la proyección mixta de las ecuaciones de Einstein obtenemos

$$D_j\left[K^{ij} - \gamma^{ij}\mathrm{tr\,K}\right] = 8\pi j^i\,, \tag{68}$$

donde j^i es el "flujo de momento" medido por los observadores de Euler:

$$j^i := P^i_\beta\ \left(n_\alpha T^{\alpha\beta}\right)\,. \tag{69}$$

La Ec. (68) tampoco tiene derivadas temporales, por lo que es otra constricción (o, mejor dicho, 3 constricciones más). A estas ecuaciones se les llama las "constricciones de momento".

La existencia de las contricciones implica que en la relatividad general no es posible especificar de manera arbitraria las 12 cantidades dinámicas $\{\gamma_{ij}, K_{ij}\}$ como condiciones iniciales. Las constricciones deben satisfacerse ya desde el inicio, o no estaremos resolviendo las ecuaciones de Einstein.

Las últimas 6 ecuaciones se obtienen a partir de la proyección a la hipersuperficie de las ecuaciones de Einstein y contienen la verdadera dinámica del sistema. Estas ecuaciones toman la forma

$$\begin{aligned}\partial_t K_{ij} = {} & \beta^a D_a K_{ij} + K_{ia} D_j \beta^a + K_{ja} D_i \beta^a - D_i D_j \alpha \\ & + \alpha\left[R^{(3)}_{ij} - 2K_{ia}K^a_j + K_{ij}\,\mathrm{tr\,K}\right] \\ & + 4\pi\alpha\left[\gamma_{ij}\left(\mathrm{tr\,S} - \rho\right) - 2S_{ij}\right]\,,\end{aligned} \tag{70}$$

donde S_{ij} es el "tensor de esfuerzos" de la materia definido como

$$S_{ij} := P\,T_{ij}\,. \tag{71}$$

Las Ecs. (62) y (70) forman un sistema cerrado de ecuaciones de evolución. A estas ecuaciones se les conoce como las ecuaciones de Arnowitt-Deser-Misner, o simplemente las ecuaciones ADM [20, 21]. Es importante notar que no tenemos ecuaciones de evolución para las variables de norma $\{\alpha, \beta^i\}$. Como hemos dicho antes, las variables de norma se pueden elegir libremente.

Es posible demostrar también, usando las identidades de Bianchi, que las ecuaciones de evolución garantizan que si las contricciones se satisfacen al tiempo inicial, entonces se satisfarán a todo tiempo posterior: Las ecuaciones de evolución propagan las constricciones.

3.4. El problema de los valores iniciales

La existencia de las contricciones en la relatividad general implica que no es posible elegir de manera arbitraria las 12 cantidades dinámicas $\{\gamma_{ij}, K_{ij}\}$ como condiciones iniciales. Los datos iniciales deben elegirse de tal modo que las constricciones se satisfagan desde el principio. Esto significa que antes de comenzar cualquier evolución, es necesario primero resolver el problema de los valores iniciales para obtener valores apropiados de $\{\gamma_{ij}, K_{ij}\}$ que representen la situación física de interés.

Las contricciones forman un conjunto de 4 ecuaciones diferenciales de tipo elíptico, y en general son difíciles de resolver. Existen, sin embargo, varios procedimientos conocidos para resolver estas ecuaciones en situaciones específicas. Hasta hace poco, el procedimiento más común era la llamada "descomposición conforme" de York-Lichnerowicz [20, 22]. Más recientemente, el llamado "formalismo del sandwich delgado" [23] también se ha vuelto muy popular para construir datos iniciales. En estas notas nos concentraremos en el primero de estos procedimientos debido a que es el mejor conocido y por lo mismo mejor entendido. Una reseña reciente sobre los diferentes procedimientos puede encontrase en la Ref. 24.

La descomposición de York-Lichnerowicz parte de considerar una trasformación de la métrica del tipo

$$\gamma_{ij} = \phi^4 \tilde{\gamma}_{ij} \,, \tag{72}$$

donde la métrica $\tilde{\gamma}_{ij}$ se considera como dada. Una transformación de este tipo se conoce como "trasformación conforme" debido a que preserva los ángulos. A la métrica $\tilde{\gamma}_{ij}$ se le llama la "métrica conforme".

Por otro lado, la curvatura extrínseca se separa primero en su traza y su parte de traza cero A_{ij}:

$$\mathrm{trK} = \mathrm{g^{ij}\ K_{ij}} \,, \qquad \mathrm{A_{ij}} := \mathrm{K_{ij}} - \frac{1}{3}\gamma_{\mathrm{ij}}\ \mathrm{trK} \,. \tag{73}$$

También se lleva a cabo una transformación conforme de A_{ij} de la forma

$$A_{ij} = \phi^{-10} \tilde{A}_{ij} \,, \tag{74}$$

donde la décima potencia se elige por conveniencia, ya que simplifica las ecuaciones que se obtienen después.

Ahora podemos utilizar el hecho de que cualquier matriz simétrica con traza cero puede separase en dos partes de la siguiente manera:

$$\tilde{A}_{ij} = \tilde{A}^*_{ij} + (\hat{L}W)_{ij} \,, \tag{75}$$

donde $\tilde{A}^*_{ij}$ es un tensor de divergencia cero $\tilde{D}^i \tilde{A}^*_{ij} = 0$ (y traza cero también), W_i es un vector y $\hat{L}$ es un operador diferencial definido como

$$(\hat{L}W)_{ij} = \tilde{D}_i W_j + \tilde{D}_j W_i - \frac{2}{3}\,\tilde{\gamma}_{ij} \tilde{D}_k W^k \,. \tag{76}$$

En las ecuaciones anteriores, $\tilde{D}_i$ representa la derivada covariante con respecto a la métrica conforme $\tilde{\gamma}_{ij}$.

Para encontrar nuestros datos inciales, asumimos ahora que las cantidades $\tilde{\gamma}_{ij}$, trK y $\tilde{A}^*_{ij}$ están dadas, y utilizamos las cuatro ecuaciones de constricción para determinar las cuatro cantidades $\{\phi, W_i\}$.

No es difícil mostrar que la constricción hamiltoniana toma la forma

$$8\tilde{D}^2\phi - \tilde{R}\,\phi + \phi^{-7}\left(\tilde{A}_{ij}\tilde{A}^{ij}\right) - \frac{2}{3}\,\phi^5(\mathrm{trK})^2 + 16\pi\tilde{\phi}^{-3}\rho = 0 \,, \tag{77}$$

lo que nos da una ecuación elíptica para ϕ, esencialmente la ecuación de Poisson. Es importante recordar que el operador laplaciano $\tilde{D}^2$ que aparece en esta ecuación debe ser el correspondiente a la métrica conforme.

Por otro lado, las contricciones de momento se convierten en

$$\tilde{D}^2\, W^i - \frac{2}{3}\phi^6 \tilde{D}^i \mathrm{trK} - 8\pi\tilde{\mathrm{j}}^{\mathrm{i}} = 0 \,, \tag{78}$$

que son tres ecuaciones elípticas (acopladas) para W^i.

Las ecuaciones anteriores son la forma más general de escribir las constricciones en la descomposición de York-Lichnerowicz. Nótese que las cuatro ecuaciones están acompladas entre sí.

Una manera de simplificar notablemente el problema y desacoplar las ecuaciones es simplemente elegir $\mathrm{trK} = 0$. Si, además, escogemos la métrica conforme como la correspondiente a un espacio plano $\tilde{\gamma}_{ij} = \delta_{ij}$, las constricciones se reducen (en el vacío) a

$$8\, D^2_{\mathrm{plano}}\phi + \phi^{-7}\left(\tilde{A}_{ij}\tilde{A}^{ij}\right) = 0 \,, \tag{79}$$

$$D^2_{\mathrm{plano}} W^i = 0 \,, \tag{80}$$

donde D^2_{plano} es simplemente el operador laplaciano estándar. La segunda ecuación es lineal y puede resolverse de forma analítica en muchos casos. Una vez teniendo la solución para W^i, se reconstruye A_{ij} y se resuelve le ecuación de Poisson para ϕ.

3.5. Datos iniciales para agujeros negros múltiples

Como ejemplo de la solución del problema de valores inciales, consideremos el siguiente caso: una métrica conformemente plana $\tilde{\gamma}_{ij} = \delta_{ij}$ y un momento de simetría temporal, es decir, $K_{ij} = 0$. En ese caso, las constricciones de momento se satisfacen trivialmente y la condición hamiltoniana toma la forma sencilla

$$D^2\phi = 0 \,, \tag{81}$$

que no es otra cosa que la ecuación de Laplace. Las condiciones de frontera corresponden a un espacio asintóticamente plano (muy lejos el campo gravitacional tiende a cero), por lo que en infinito debemos tener $\phi = 1$. La solución más sencilla de esta ecuación que satisface la condición de frontera es entonces

$$\phi = 1 \,, \tag{82}$$

lo que implica que la métrica espacial es simplemente

$$dl^2 = dx^2 + dy^2 + dz^2 \,. \tag{83}$$

Es decir, hemos recuperado los datos iniciales para el espacio de Minkowski.

La siguiente solución de interés es

$$\phi = 1 + k/r\,, \tag{84}$$

con k una constante arbitraria, por lo que la métrica espacial es ahora (en coordenadas esféricas)

$$dl^2 = (1 + k/r)^4 \left[dr^2 + r^2 d\Omega^2\right]\,. \tag{85}$$

No es difícil mostrar que esto no es otra cosa que la métrica de un agujero negro de Schwarzschild en las llamadas "coordenadas isótropas", donde la masa del agujero negro está dada por $M = 2k$. Hemos encontrado la solución para el problema de valores iniciales correspondiente a un agujero negro de Schwarzschild.

Como la ecuación de Laplace es lineal, podemos superponer soluciones para obtener nuevas soluciones. Por ejemplo, la solución

$$\phi = 1 + \sum_{i=1}^{N} \frac{M_i}{2|r - r_i|} \tag{86}$$

representa a N agujeros negros en los puntos r_i incialmente en reposo y se conoce como los datos inciales de Brill-Lindquist. Es posible generalizar dicha solución al caso en que los agujeros negros no estén en reposo, construyendo así, por ejemplo, datos inciales para dos agujeros negros en órbita. Sin embargo, en ese caso la solución de la constricción hamiltoniana ya no es tan sencilla y debe obtenerse por métodos numéricos.

Es importante notar que este tipo de datos iniciales para agujeros negros son modelos de tipo topológico: los agujeros negros están representados por agujeros de gusano (túneles de Einstein-Rosen) que unen nuestro Universo con otros. Esto ocurre en la solución de Schwarzschild y corresponde a un agujero negro "eterno" (no formado por colapso). En el mundo real, uno espera que los agujeros negros se formen por un colapso y el túnel de Einstein-Rosen no estará ahi.

3.6. Condiciones de norma

Como hemos mencionado, en el formalismo 3+1 la elección del sistema de coordenadas está dada en términos de las "variables de norma": la función de lapso α y el vector de corrimiento β^i. Estas funciones aparecen en las ecuaciones de evolución (62) y (70) para la métrica y la curvatura extrínseca. Sin embargo, las ecuaciones de Einstein no nos dicen qué valores deben tomar las variables de norma. Esto es lo que deberíamos esperar, pues es claro que las coordenadas deben poder elegirse libremente.

La libertad en la elección de las variables es una bendición a medias. Por un lado, nos permite elegir las cosas de manera que las ecuaciones se simplifiquen, o de manera que la solución sea mejor comportada. Por otro lado, surge inmediatamente la pregunta: ¿Cuál es una buena elección de las funciones α y β^i? Recuérdese que esta decisión debe tomarse incluso antes de comenzar la evolución. Existen, desde luego, infinitas elecciones posibles. En estas notas nos limitaremos a mencionar algunas de las más conocidas.

3.6.1. Condiciones de foliación

Consideremos en primer lugar la elección de la función de lapso α. La elección de esta función determina cómo avanza el tiempo propio entre distintas hipersuperficies espaciales, es decir, determina cómo está rebanado el espacio-tiempo en superficies de coordenada t constante. Debido a esto, a la elección del lapso se le llama comúnmente una "condición de foliación".

Antes de considerar diferentes condiciones de foliación, debemos mencionar un par de resultados que son muy importantes al hablar sobre este tema. En primer lugar, consideremos el movimiento de los observadores de Euler, es decir, aquellos que se mueven siguiendo las líneas normales a las hipersuperficies que forman la foliación. Nótese que no hay razón para suponer que estos observadores están en caída libre, por lo que en general pueden tener lo que se conoce como una "aceleración propia" que mide la fuerza requerida para mantenerlos en una trayectoria diferente a la caida libre. Esta aceleración está dada por la derivada direccional de la cuatrovelocidad de los observadores de Euler (el vector normal n^μ) a lo largo de sí misma:

$$a^\mu = n^\nu \nabla_\nu n^\mu\,. \tag{87}$$

Nótese que esto involucra a la derivada covariante en 4 dimensiones ∇_μ. No es difícil mostrar que esta ecuación implica que la aceleración está dada en términos de la función de lapso por

$$a_i = \partial_i \ln \alpha\,. \tag{88}$$

Otra relación importante está relacionada con la evolución de los elementos de volumen asociados a los obervadores de Euler. El cambio de estos elementos de volumen en el tiempo está dado por la divergencia de las velocidades de estos observadores $\nabla_\mu n^\mu$. Usando la definición de la curvatura extrínseca encontramos ahora

$$\nabla_\mu n^\mu = -\mathrm{trK}\,, \tag{89}$$

es decir, el cambio de los elementos de volumen en el tiempo es menos la traza de la curvatura extrínseca.

Regresemos ahora al problema de cómo elegir la función de lapso. La manera más obvia de elegir esta función es simplemente decidir que queremos que el tiempo coordenado t coincida en todo punto con el tiempo propio de los observadores de Euler. Después de todo, ¿qué podría ser más natural? Esta elección corresponde a tomar $\alpha = 1$ en todo el espacio. De la Ec.88 vemos que en este caso la aceleración de los observadores de Euler es cero, es decir, están en caída libre y se mueven a lo largo de geodésicas. Debido a esto, a esta elección se le conoce como "foliación geodésica".

La foliación geodésica, pese a parecer muy natural, es en la práctica una muy mala elección. La razón de esto es fácil

de ver si pensamos que les ocurre a observadores en caída libre en un campo gravitacional no uniforme. Al ser el campo no uniforme, diferentes observadores caen en distintas direcciones y nada impide que choquen entre ellos. Cuando esto sucede, el sistema de coordenadas que está atado a estos observadores deja de ser uno a uno: un mismo punto tiene ahora más de un valor de las coordenadas asociado a él. Esto significa que el sistema de coordenadas se ha vuelto singular. Debido a esto, la foliación geodésica no se utiliza.

Para buscar una mejor condición de lapso, debemos pensar cuál fue el problema con la foliación geodésica. El problema básico es que observadores en caída libre son "enfocados" por el campo gravitacional. Esto significa que se acercan unos a otros, por lo que los elementos de volumen disminuyen hasta hacerse cero. Podemos entonces construir una foliación donde se exija que los elementos de volumen se mantengan constantes. De la Ec. (89) vemos que esto es equivalente a pedir

$$\mathrm{trK} = \partial_t \mathrm{trK} = 0 \; . \tag{90}$$

Como se ve en esta expresión, es claro que debemos exigir no sólo que trK sea cero inicialmente, sino que se mantenga cero durante la evolución. Por otro lado, a partir de las Ecs. (62) y (70) encontramos

$$\partial_t \mathrm{trK} = \beta^{\mathrm{i}} \partial_{\mathrm{i}} \mathrm{trK} - \mathrm{D}^2\alpha + \alpha \left[\mathrm{K_{ij}K^{ij}} + \frac{1}{2}\left(\rho + \mathrm{trS}\right) \right], \tag{91}$$

donde se utilizó la constricción hamiltoniana para eliminar el escalar de curvatura. Imponiendo ahora la condición (90) encontramos que la función de lapso debe satisfacer la siguiente ecuación elíptica:

$$D^2\alpha = \alpha \left[K_{ij}K^{ij} + \frac{1}{2}\left(\rho + \mathrm{trS}\right) \right] \; . \tag{92}$$

Esta condición se conoce como "foliación maximal" [25]. Este nombre se debe a que puede demostrarse que ésta es la condición que garantiza que el volumen de una región cualquiera de la hipersuperficie espacial se maximiza ante pequeñas variaciones.

La condición de foliación maximal se ha utilizado en muchas simulaciones numéricas de diferentes sistemas, incluyendo agujeros negros, y sigue siendo popular a la fecha. Sus ventajas están en que es una ecuación sencilla, cuya solución es suave (por ser producto de una ecuación elíptica), y que garantiza que los observadores de Euler no pueden enfocarse. En particular, la foliación maximal tiene la propiedad de evadir singularidades, tanto singularidades de coordenadas debidas al enfoque de observadores, como singularidades físicas como aquellas que se encuentran en el interior de un agujero negro. En el caso de un agujero negro, la solución de la condición maximal hace que la función de lapso se haga exponencialmente cero en la región interior al agujero negro, fenómeno que se conoce como el "colapso del lapso" (ver Fig. 5). El colapso del lapso resulta muy útil, ya que de otra forma las hipersuperficies marcharían directo a la singularidad física en un tiempo muy corto, donde se encontrarían valores infinitos del campo gravitacional, lo que causaría que el código computacional se detenga.

Una de las desventajas de utilizar una foliación maximal es el hecho de que en tres dimensiones la solución de la ecuación elíptica (92) resulta muy costosa en tiempo de máquina. Debido a esto se han buscado condiciones de foliación alternativas que tengan propiedades similares a la foliación maximal, pero que a la vez sean más fáciles de resolver. Una familia de condiciones de lapso que tiene estas propiedades es la llamada condición de Bona-Masso [26]. Esta familia especifica una ecuación de evolución para la función de lapso de la forma

$$\partial_t \alpha - \beta^i \partial_i \alpha = -\alpha^2 f(\alpha)\, K \; , \tag{93}$$

donde $f(\alpha)$ es una función positiva pero por lo demás arbitraria de α (la razón por la que debe ser positiva tiene que ver con el concepto de hiperbolicidad que se dicute en la Sec. 4). El término con el vector de corrimiento β^i se incluye para formar la "derivada total" respecto al tiempo, es decir, la derivada a lo largo de la dirección normal. La condición de Bona-Masso incluye a subfamilias muy conocidas. Por ejemplo, si se toma $f = 1$ se obtiene la llamada "foliación armónica", para la cual la coordenada t resulta obedecer la ecuación de onda $g^{\mu\nu}\nabla_\mu \nabla_\nu t = 0$. En este caso la ecuación de evolución del lapso se puede integrar para obtener

$$\alpha = h(x^i)\, \gamma^{1/2}, \tag{94}$$

donde $h(x^i)$ es una función independiente del tiempo y γ es el determinante de la métrica espacial, es decir, $\gamma^{1/2}$ es el elemento de volumen.

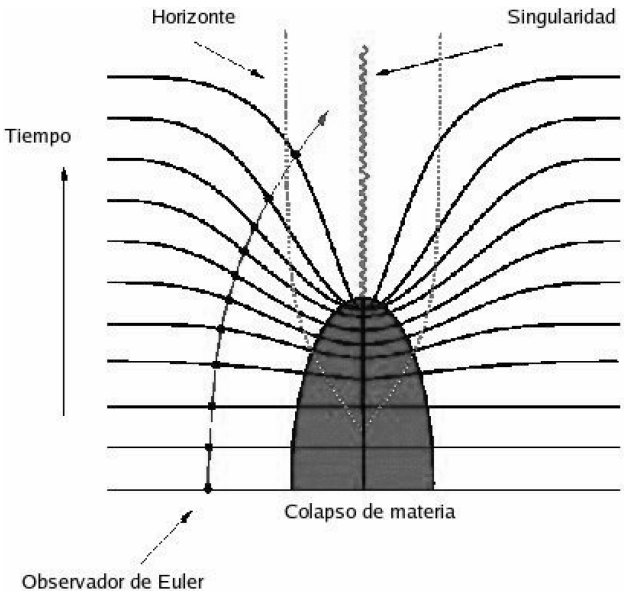

FIGURA 5. Diagrama esquemático del colapso del lapso.

Otra familia interesante se obtiene tomando $f(\alpha) = N/\alpha$ con N una constante positiva. Esto corresponde a la familia "1+log" [27, 28], cuyo nombre proviene de que en este caso la función del lapso resulta tener la forma

$$\alpha = h(x^i) + \ln\left(\gamma^{N/2}\right) . \tag{95}$$

La familia 1+log conduce a foliaciones muy similares al lapso maximal y también tiene la propiedad de evadir singularidades. En la práctica se ha encontrado que el caso $N = 2$ es particularmente bien comportado [29].

3.6.2. Condiciones de corrimiento

Es menos lo que se conoce sobre elecciones del vector de corrimiento que sobre condiciones de foliación. La razón es que tomar el vector de corrimiento igual a cero funciona bien en muchos casos. Sin embargo, sí existen propuestas de cómo elegir el vector de corrimiento en diversas situaciones.

La propuesta mejor conocida para elegir el vector de corrimiento se conoce como "distorsión mínima" [30]. La idea básica es utilizar al vector de corrimiento para eliminar lo más posible la distorsión de los elementos de volumen. Para ello se define el llamado "tensor de distorsión" Σ_{ij} como

$$\Sigma_{ij} = \frac{1}{2}\gamma^{1/2}\partial_t\tilde{\gamma}_{ij} , \tag{96}$$

donde γ es el determinante de la métrica espacial y $\tilde{\gamma}_{ij} = \gamma^{-1/3}\gamma_{ij}$. Es decir, $\tilde{\gamma}_{ij}$ es una transformación conforme de la métrica en la que se ha factorizado el elemento de volumen ($\tilde{\gamma} = 1$). El tensor de distorsión puede entenderse entonces como el cambio en la *forma* de los elementos de volumen durante la evolución (y no en su volumen interno).

A partir de la ecuación de evolución para la métrica (62), puede demostrarse que

$$\Sigma_{ij} = -\alpha\left(K_{ij} - \frac{1}{3}\gamma_{ij}\mathbf{tr}K\right) + \frac{1}{2}\left(L\beta\right)_{ij} , \tag{97}$$

donde L es el mismo operador que definimos cuando discutimos cómo encontrar datos iniciales (Ec. (76).

La condición de distorsión mínima se obtiene de pedir que la integral sobre todo el espacio del "cuadrado" no-negativo del tensor de distorsión, $\Sigma_{ij}\Sigma^{ij}$, sea lo más pequeña posible. En este sentido, esta condición minimiza la distorsión de manera global. La condición que acabamos de describir se expresa matemáticamente como un principio variacional, que tiene como solución

$$D^j\Sigma_{ij} = 0 , \tag{98}$$

o de manera equivalente

$$D^j\left(L\beta\right)_{ij} = 2D^j\left[\alpha\left(K_{ij} - \frac{1}{3}\gamma_{ij}\mathbf{tr}K\right)\right] . \tag{99}$$

Ésta es la forma final de la condición de distorsión mínima.

Hay varias cosas que es importante notar de esta condición. En primer lugar, se trata de tres ecuaciones diferenciales (el índice i está libre), que en principio nos permiten calcular las tres componentes del vector β^i. En segundo lugar, es un sistema de ecuaciones de tipo elíptico (generalizaciones de la ecuación de Poisson), acopladas entre sí. Debido a esto, la condición de distorsión mínima es muy difícil de resolver en el caso tri-dimensional (aunque no imposible), lo que ha hecho que sea poco utilizada en la práctica.

Condiciones de corrimiento relacionadas con la distorsión mínima, pero mucho más fáciles de resolver, se han propuesto en los últimos años. En particular, propuestas del tipo

$$\partial_t^2\beta^i \propto D^j\Sigma_{ij} , \tag{100}$$

que transforman el sistema de ecuaciones elípticas en un sistema hiperbólico se han propuesto en la Ref 31, pero en estas notas no hablaremos sobre ellas.

Para terminar esta sección es importante comentar que, si bien es cierto que en muchos casos tomar el vector de corrimiento igual a cero funciona muy bien, hay casos en los que es claro que esto no puede ser así. En particular, sistemas con momento angular (agujeros negros o estrellas de neutrones en rotación) producen un efecto físico llamado "arrastre de sistemas inerciales", que en pocas palabras significa que observadores cercanos al objeto que rota se ven arrastrados en la rotación de éste. En este tipo de situaciones, la única forma de impedir que el sistema de coordenadas sea arrastrado, y termine completamente enredado alrededor del objeto, es utilizando un vector de corrimiento que se oponga al arraste. Estas situaciones también se dan en el caso de la órbita de objetos compactos, por lo que el estudio de condiciones adecuadas de corrimiento se ha vuelto un tema de gran actualidad.

4. Formulaciones alternativas e hiperbolicidad

En la sección anterior se introdujeron las ecuaciones de evolución de ADM, Ecs. (62) y (70). De hecho, estas ecuaciones no están escritas en la forma original de Arnowitt, Deser y Misner [21], sino que son una re-escritura no trivial por parte de York [20]. Aun así, en la comunidad numérica las Ecs. (62) y (70) se conocen como ecuaciones ADM.

Es importante señalar precisamente en qué difieren las ecuaciones de ADM originales de las ecuaciones ADM *a la* York (que llamaremos de ahora en adelante "ADM estándar"). Ambos grupos de ecuaciones difieren en dos aspectos distintos. En primer lugar, las variables ADM originales son la métrica espacial γ_{ij} y su momento canónico conjugado π_{ij} que se obtiene a partir de la formulación lagrangiana de la relatividad general, y que está relacionado con la curvatura extrínseca como

$$K_{ij} = -\gamma^{-1/2}\left(\pi_{ij} - \frac{1}{2}\,\gamma_{ij}\pi\right) , \tag{101}$$

con γ el determinante de la métrica espacial y $\pi = \mathrm{tr}\;\pi_{ij}$.

Sin embargo, incluso si se reescribe en términos de la curvatura extrínseca, la ecuación de evolución que ADM obtienen para K_{ij} difiere de (70), y tiene la forma

$$\partial_t K_{ij} = \beta^a D_a K_{ij} + K_{ia} D_j \beta^a + K_{ja} D_i \beta^a - D_i D_j \alpha + \alpha \left[R^{(3)}_{ij} - 2 K_{ia} K^a_j + K_{ij} \, \mathrm{tr\,K} \right] + 4\pi\alpha \left[\gamma_{ij} \left(\mathrm{tr\,S} - \rho \right) - 2 S_{ij} \right] - \frac{\alpha}{4} \, \gamma_{ij} H \,, \qquad (102)$$

donde H es la constricción hamiltoniana (66):

$$H := R^{(3)} + \left(\mathrm{tr\,K} \right)^2 - K_{ij} K^{ij} - 16\pi\rho = 0 \,. \qquad (103)$$

Es claro que ambas ecuaciones de evolución son físicamente equivalentes, ya que difieren sólo en la adición de un término proporcional a la constricción hamiltoniana que debe ser cero para cualquier solución física. Sin embargo, las diferentes ecuaciones de evolución para K_{ij} no son *matemáticamente* equivalentes. Hay dos razones por las que ambas ecuaciones son distintas matemáticamente:

1. En primer lugar, el espacio de soluciones de las ecuaciones de evolución es distinto, y sólo coincide para soluciones físicas, es decir, aquellas que satisfacen las contricciones. Dicho de otro modo, ambos sistemas de ecuaciones sólo son equivalentes en un subconjunto del espacio total de soluciones. A este subconjunto se le conoce como la "hipersuperficie de constricción".

   Desde luego, uno podría argumentar que dado que sólo estamos interesados en soluciones físicas, la distinción que hemos hecho es irrelevante. Esto es estrictamente cierto si uno puede resolver las ecuaciones de forma exacta. Pero en el caso de soluciones numéricas siempre habrá un elemento de error que nos llevará fuera de la hipersuperficie de constricción, y la pregunta de qué ocurre en ese caso se vuelve no sólo relevante sino de fundamental importancia: Si nos salimos ligeramente de la hipersuperficie de constricción, ¿nos mantenemos a una distancia pequeña, o nos alejamos de ella cada vez más rápido?

2. La segunda razón por la que ambos sistemas difieren matemáticamente está relacionada con la anterior y es de mayor importancia. Debido a que la constricción hamiltoniana tiene derivadas de la métrica (escondidas en el escalar de curvatura R), al añadir un múltiplo de ésta estamos cambiando la *estructura misma* de la ecuación diferencial, es decir, podemos pasar de un sistema de tipo hiperbólico a otro de tipo elíptico, e incluso de un sistema bien comportado ("bien planteado") a otro que no lo es.

Estas consideraciones nos llevan a una observación que es hoy en día una de las áreas de investigación más importantes en la relatividad numérica: las ecuaciones de evolución son altamente no únicas, pues siempre es posible añadirles múltiplos arbitrarios de las contricciones. Diferentes ecuaciones coinciden en las soluciones físicas, pero pueden diferir de manera dramática en sus propiedades matemáticas y en particular en la manera en la que responden a pequeñas violaciones de las constricciones (inevitables numéricamente). Cuál es la mejor manera de escribir las ecuaciones de evolución es un problema abierto de gran actualidad.

4.1. La formulación BSSN

Aun cuando las ecuaciones de ADM son el punto de partida del formalismo 3+1, en la práctica no han resultado ser muy bien comportadas frente a pequeñas violaciones de las constricciones. En la siguiente sección discutiremos una posible explicación de este hecho, pero de momento lo tomaremos como una observación empírica: las ecuaciones ADM no son muy estables ante violaciones de las contricciones.

Desde principios de la década de los 90, se han propuesto una gran cantidad de formulaciones alternativas de las ecuaciones de evolución con diversas motivaciones. En este curso no podemos verlas todas, ni siquiera una porción representativa, así que nos limitaremos a discutir una formulación particular que se ha vuelto muy popular en los últimos años debido a que en la práctica ha resultado ser muy bien comportada. Ésta es la formulación conocida como BSSN (Baumgarte y Shapiro [32], Shibata y Nakamura [33]).

La formulación BSSN difiere de la formulación ADM en varios aspectos. En primer lugar, BSSN introduce nuevas variables similares a las utilizadas en el problema de valores iniciales que discutimos en la Sec. 3.4. La métrica espacial γ_{ij} se reescribe en términos de una métrica conforme $\tilde{\gamma}_{ij}$ mediante la transformación:

$$\gamma_{ij} = \psi^4 \tilde{\gamma}_{ij}, \qquad (104)$$

donde el factor conforme ψ se escoge de manera que el determinante de $\tilde{\gamma}_{ij}$ sea igual a 1, es decir,

$$\psi = \gamma^{1/12}, \quad \tilde{\gamma}_{ij} = \psi^{-4} \gamma_{ij} = \gamma^{-1/3} \gamma_{ij}, \quad \tilde{\gamma} = 1, \quad (105)$$

donde γ es el determinante de γ_{ij} y $\tilde{\gamma}$ el determinante de$\tilde{\gamma}_{ij}$. Por otro lado, la curvatura extrínseca se separa en su traza $K := \mathrm{tr}K$ y su parte de traza cero:

$$A_{ij} = K_{ij} - \frac{1}{3} \gamma_{ij} K. \qquad (106)$$

En vez de las variables ADM γ_{ij} y K_{ij}, se utilizan entonces las variables

$$\phi = \ln \psi = \frac{1}{12} \ln \gamma, \quad K = \gamma_{ij} K^{ij}, \qquad \tilde{\gamma}_{ij} = e^{-4\phi} \gamma_{ij}, \qquad \tilde{A}_{ij} = e^{-4\phi} A_{ij}, \qquad (107)$$

Además, se introducen tres variables auxiliares llamadas "funciones de conexión conformes" definidas como

$$\tilde{\Gamma}^i = \tilde{\gamma}^{jk} \tilde{\Gamma}^i{}_{jk} = -\partial_j \tilde{\gamma}^{ij}, \qquad (108)$$

donde $\tilde{\Gamma}^i{}_{jk}$ son los símbolos de Christoffel de la métrica conforme, y donde la segunda igualdad se deriva de la definición de los símbolos de Christoffel en el caso en que el determinante de $\tilde{\gamma}$ sea igual a 1 (lo que debe ser cierto por construcción). Llamamos a las 17 variables ϕ, K, $\tilde{\gamma}_{ij}$, $\tilde{A}_{ij}$ y $\tilde{\Gamma}^i$ las "variables BSSN".

Hasta ahora no hemos hecho otra cosa que redefinir variables e introducir tres variables auxiliares adicionales. Aún falta dar las ecuaciones de evolución de las nuevas variables. Por simplicidad, de aquí en adelante asumiremos que estamos en vacío y que el vector de corrimiento β^i es cero (el caso general puede obtenerse fácilmente). De la ecuación de evolución para la métrica espacial (62) encontramos

$$\partial_t \tilde{\gamma}_{ij} = -2\alpha \tilde{A}_{ij}, \tag{109}$$

$$\partial_t \phi = -\frac{1}{6}\alpha K, \tag{110}$$

mientras que de la ecuación de evolución para la curvatura extrínseca (70) obtenemos

$$\partial_t \tilde{A}_{ij} = e^{-4\phi}[-D_i D_j \alpha + \alpha R_{ij}]^{TF} + \alpha(K\tilde{A}_{ij} - 2\tilde{A}_{ik}\tilde{A}^k{}_j), \tag{111}$$

$$\partial_t K = -D^i D_i \alpha + \alpha(\tilde{A}_{ij}\tilde{A}^{ij} + \frac{1}{3}K^2), \tag{112}$$

donde TF denota la parte de traza cero. Es importante notar que en la ecuación de evolución para K se ha utilizado ya la constricción hamiltoniana para eliminar al escalar de curvatura:

$$R = K_{ij}K^{ij} - K^2 = \tilde{A}_{ij}\tilde{A}^{ij} - \frac{2}{3}K^2. \tag{113}$$

Vemos entonces que ya hemos comenzado a jugar el juego de sumar múltiplos de las constricciones a las ecuaciones de evolución. Falta aún la ecuación de evolución para las $\tilde{\Gamma}^i$. Esta ecuación puede obtenerse directamente de (108) y (62). Encontramos que

$$\partial_t \tilde{\Gamma}^i = -2\left(\alpha \partial_j \tilde{A}^{ij} + \tilde{A}^{ij}\partial_j \alpha\right). \tag{114}$$

Nótese que en las ecuaciones de evolución para $\tilde{A}_{ij}$ y K aparecen derivadas covariantes de la función de lapso con respecto a la métrica física γ_{ij} que están dadas por

$$D^i D_i \alpha = e^{-4\phi}(\tilde{\gamma}^{ij}\partial_i\partial_j\alpha - \tilde{\Gamma}^k\partial_k\alpha + 2\tilde{\gamma}^{ij}\partial_i\phi\partial_j\alpha). \tag{115}$$

Además, en la ecuación para $\tilde{A}_{ij}$ aparece el tensor de Ricci asociado a γ_{ij} que se separa de la siguiente forma:

$$R_{ij} = \tilde{R}_{ij} + R^\phi_{ij}, \tag{116}$$

donde $\tilde{R}_{ij}$ es el tensor de Ricci asociado a la métrica conforme $\tilde{\gamma}_{ij}$:

$$\tilde{R}_{ij} = -\frac{1}{2}\tilde{\gamma}^{lm}\partial_l\partial_m\tilde{\gamma}_{ij} + \tilde{\gamma}_{k(i}\partial_{j)}\tilde{\Gamma}^k + \tilde{\Gamma}^k\tilde{\Gamma}_{(ij)k} + \tilde{\gamma}^{lm}\left(2\tilde{\Gamma}^k{}_{l(i}\tilde{\Gamma}_{j)km} + \tilde{\Gamma}^k{}_{im}\tilde{\Gamma}_{klj}\right), \tag{117}$$

y donde R^ϕ_{ij} denota términos adicionales que dependen de ϕ:

$$R^\phi_{ij} = -2\tilde{D}_i\tilde{D}_j\phi - 2\tilde{\gamma}_{ij}\tilde{D}^k\tilde{D}_k\phi + 4\tilde{D}_i\phi\,\tilde{D}_j\phi - 4\tilde{\gamma}_{ij}\tilde{D}^k\phi\,\tilde{D}_k\phi, \tag{118}$$

donde $\tilde{D}_i$ es ahora la derivada covariante con respecto a la métrica conforme.

La motivación para el cambio de variables realizado arriba es múltiple. La transformación conforme y la separación de la traza de la curvatura extrínseca se lleva a cabo para tener mejor control sobre las condiciones de lapso, que como se mencionó en la secc. 3.6.1, en general están relacionadas con la traza de K_{ij} y se escogen de manera que se pueda controlar mejor la evolución de los elementos de volumen. Por otro lado, la introducción de las variables auxiliares $\tilde{\Gamma}^i$ se debe a que, cuando éstas se consideran variables independientes, las segundas derivadas de la métrica conforme que aparecen en el lado derecho de la ecuación de evolución (111) (contenidas en el tensor de Ricci (117)) se reducen simplemente al operador de Laplace $\tilde{\gamma}^{lm}\partial_l\partial_m\tilde{\gamma}_{ij}$. Todos los demás términos que tienen segundas derivadas de $\tilde{\gamma}_{ij}$ pueden reescribirse en términos de primeras derivadas de $\tilde{\Gamma}^i$. Esto significa que las Ecs. (109) y (111) toman la estructura de una ecuación de onda.

Falta aún un elemento clave en la formulación BSSN que no hemos mencionado. En la práctica resulta que, pese a las motivaciones arriba mencionadas, si uno utiliza las ecuaciones de evolución dadas por (109), (110), (111), (112) y (114), el sistema resulta ser *violentamente inestable* en simulaciones numéricas.

Para corregir este problema escribimos primero las contricciones de momento, que en las variables BSSN toman la forma

$$\partial_j \tilde{A}^{ij} = -\tilde{\Gamma}^i{}_{jk}\tilde{A}^{jk} - 6\tilde{A}^{ij}\partial_j\phi + \frac{2}{3}\tilde{\gamma}^{ij}\partial_j K. \tag{119}$$

Podemos ahora utilizar esta ecuación para substituir a la divergencia de $\tilde{A}^{ij}$ que aparece en la ecuación de evolución (114) para $\tilde{\Gamma}^i$, obteniendo

$$\partial_t\tilde{\Gamma}^i = -2\tilde{A}^{ij}\partial_j\alpha + 2\alpha\left(\tilde{\Gamma}^i{}_{jk}\tilde{A}^{jk} + 6\tilde{A}^{ij}\partial_j\phi - \frac{2}{3}\tilde{\gamma}^{ij}\partial_j K\right). \tag{120}$$

El sistema de ecuaciones de evolución es ahora: (109), (110), (111), (112), y (120). Este nuevo sistema no sólo ya no presenta la violenta inestabilidad antes mencionada, sino que además resulta ser mucho mejor comportado que el sistema ADM en todos los casos estudiados a la fecha. Que esto es así fue mostrado por primera vez de forma empírica (es decir, mediante comparaciones directas de simulaciones numéricas) por Baumgarte y Shapiro en la Ref. 32, y explicado por Alcubierre y colaboradores en la Ref. 34. La explicación de porqué BSSN es mejor que ADM está relacionada con el concepto de hiperbolicidad que discutiremos en la siguiente sección.

4.2. El concepto de hiperbolicidad

En esta sección daremos una breve introducción al concepto de hiperbolicidad de sistemas de ecuaciones de evolución, que resulta clave en el análisis de las propiedades matemáticas de los sistemas de ecuaciones utilizados en diversas áreas de la física, y en particular en la relatividad numérica. Una descripción más detallada de la teoría de sistemas hiperbólicos puede encontrarse en la Ref. 35.

Consideremos un sistema de ecuaciones de evolución en una dimensión de la forma

$$\partial_t u_i \, + \, \partial_x F_i \, = \, q_i \qquad i \in \{1, \ldots, N_u\} \, , \tag{121}$$

donde F_i y q_i son funciones arbitrarias, posiblemente no lineales, de las u's pero no de sus derivadas. Este sistema también puede escribirse como

$$\partial_t \, u_i \, + \, \sum_j M_{ij} \partial_x u_j \, = \, q_i \qquad i \in \{1, \ldots, N_u\} \, , \tag{122}$$

con

$$M_{ij} = \frac{\partial F_i}{\partial u_j}$$

la llamada "matriz jacobiana".

Es importante mencionar que la mayor parte de las ecuaciones diferenciales de la física se pueden escribir de esta forma. En el caso en el que haya derivadas de orden mayor, se pueden simplemente definir variables auxiliares para obtener un sistema de primer orden.

Sean ahora λ_i los eigenvalores de la matriz jacobiana M. El sistema de ecuaciones de evolución se denomina "hiperbólico" si todas las λ_i resultan ser funciones reales. Más aún, el sistema se denomina "fuertemente hiperbólico" si la matriz M tiene un conjunto completo de eigenvectores. En caso que todos los eigenvalores sean reales, pero no exista un conjunto completo de eigenvectores, el sistema se denomina "débilmente hiperbólico".

Asumamos ahora que el sistema es fuertemente hiperbólico. Definimos las "eigenfunciones" w_i como

$$\mathbf{u} \, = \, R \, \mathbf{w} \quad \Rightarrow \quad \mathbf{w} \, = \, R^{-1} \, \mathbf{u} \ , \tag{123}$$

donde R es la matriz de eigenvectores columna $\mathbf{e}_i$. Es posible mostrar que la matriz R es tal que

$$R M R^{-1} = \Lambda \, , \tag{124}$$

con $\Lambda = \mathrm{diag}(\lambda_i)$. Las ecuaciones de evolución para los eigencampos w_i resultan ser entonces

$$\partial_t w_i \, + \, \lambda_i \, \partial_x w_i \, = \, q'_i \, , \tag{125}$$

con q'_i funciones de las w's pero no sus derivadas. El sistema se ha transformado entonces en una serie de ecuaciones de advección con velocidades características dadas por los eigenvalores λ_i. Dicho de otro modo, tenemos una serie de perturbaciones propagándose con velocidades λ_i.

La hiperbolicidad es de fundamental importancia en el estudio de las ecuaciones de evolución asociadas con un problema de Cauchy. En un sentido físico, la hiperbolicidad implica que el sistema de ecuaciones es causal y local, es decir, que la solución en un punto dado del espacio-tiempo depende sólo de información contenida en una región compacta al pasado de ese punto, el llamado "cono característico" (o cono de luz en el caso de la relatividad). Matemáticamente, se puede mostrar que un sistema fuertemente hiperbólico está "bien planteado", es decir, sus soluciones existen y son únicas (al menos localmente), y además las soluciones son estables en el sentido de que cambios pequeños en los datos iniciales corresponden a cambios pequeños en la solución.

El concepto de hiperbolicidad se puede extender a tres dimensiones considerando ecuaciones del tipo

$$\partial_t u_i \, + \, \partial_x F^x_i + \partial_y F^y_i + \partial_z F^z_i \, = \, q_i \, , \qquad i \in \{1, \ldots, N_u\} \, , \tag{126}$$

y analizando las tres matrices jacobianas

$$M^k_{ij} = \partial F^k_i / \partial u_j \qquad (k = x, y, z).$$

Para ver un ejemplo sencillo de hiperbolicidad, consideremos la ecuación de onda en una dimensión:

$$\frac{\partial^2 \phi}{\partial t^2} - c^2 \, \frac{\partial^2 \phi}{\partial x^2} = 0 \, , \tag{127}$$

donde ϕ es la función de onda y c la velocidad de onda. Esta ecuación es de segundo orden, pero podemos pasarla a primer orden si introducimos las variables auxiliares:

$$\Pi := \partial \phi / \partial t \, , \qquad \Psi := \partial \phi / \partial x \, . \tag{128}$$

La ecuación de onda puede reescribirse entonces como el sistema

$$\frac{\partial \phi}{\partial t} = \Pi \, , \quad \frac{\partial \Pi}{\partial t} - c^2 \frac{\partial \Psi}{\partial x} = 0 \, , \quad \frac{\partial \Psi}{\partial t} - \frac{\partial \Pi}{\partial x} = 0 \, , \tag{129}$$

donde la primera ecuación es la definición de Π, la segunda es la ecuación de onda propiamente dicha, y la tercera es el requerimiento de que las derivadas de ϕ conmuten. El sistema anterior claramente tiene ya la forma (121). Como la función ϕ no aparece en ninguna de las dos últimas ecuaciones, podemos analizar únicamente el sistema $\{\Pi, \Psi\}$. La matriz jacobiana en este caso es de 2×2 y resulta ser

$$\mathbf{M} = \begin{pmatrix} 0 & -c^2 \\ -1 & 0 \end{pmatrix} \, , \tag{130}$$

cuyos correspondientes eigenvalores son reales y están dados por $\lambda_\pm = \pm c$, es decir, las ondas pueden moverse a la izquierda y a la derecha con velocidad c. La martriz M tiene dos eigenvectores independientes, es decir un conjunto completo, por lo que el sistema es fuertemente hiperbólico. Los eigenvectores resultan ser $v_\pm = (\mp c, 1)$, de donde podemos encontrar fácilmente las eigenfunciones $\omega_\pm = \Pi \mp c \Psi$ (nótese que a la velocidad $+c$ corresponde la eigenfunción $\Pi - c\Psi$

y viceversa). Las ecuaciones de evolución para las eigenfunciones son

$$\frac{\partial \omega_\pm}{\partial t} \pm c \frac{\partial \omega_\pm}{\partial x} = 0 \ . \tag{131}$$

Vemos que las eigenfunciones se propagan en direcciones opuestas de manera independiente.

4.3. Hiperbolicidad de las ecuaciones 3+1

Vamos ahora a estudiar la hiperbolicidad de las ecuaciones de evolución de ADM, Ecs.(62) y (70). Como sólo estamos interesados en la idea básica, vamos a analizar un caso muy simple (las conclusiones no se modifican fundamentalmente en el caso general). En primer lugar, asumiremos que estamos en el vacío y que el vector de corrimiento β^i es cero. Las ecuaciones que vamos a utilizar son entonces

$$\partial_t \gamma_{ij} = -2\alpha K_{ij} \ , \tag{132}$$

$$\partial_t K_{ij} = - D_i D_j \alpha + \alpha \left[R^{(3)}_{ij} - 2K_{ia}K^a_j + K_{ij}\,\mathrm{tr\,K} \right] \ , \tag{133}$$

Para simplificar aún más las cosas, consideraremos perturbaciones lineales de un espacio plano. En este caso, la métrica espacial y la función de lapso se pueden escribir como

$$\alpha = 1 + a, \qquad \gamma_{ij} = \delta_{ij} + h_{ij}, \tag{134}$$

con a y h_{ij} mucho menores que 1. A primer orden en cantidades pequeñas, las ecuaciones de evolución toman la forma simplificada (nótese que en estas circumstancias K_{ij} resulta ser una cantidad pequeña):

$$\partial_t h_{ij} = -2K_{ij}, \tag{135}$$

$$\partial_t K_{ij} = -\partial_i \partial_j a + R^{(1)}_{ij}, \tag{136}$$

donde el tensor de Ricci linealizado está dado por

$$R^{(1)}_{ij} = -1/2 \ \left(\nabla^2_{\rm plano} h_{ij} - \partial_i \Gamma_j - \partial_j \Gamma_i \right). \tag{137}$$

y donde hemos definido

$$\Gamma_i := \sum_k \partial_k h_{ik} - 1/2 \ \partial_i h, \tag{138}$$

con

$$h = \sum_i h_{ii}.$$

Para continuar el análisis debemos ahora elegir nuestra condición de foliación. La opción más simple es tomar $a = 0$, pero eso equivale a la foliación geodésica y ya hemos mencionado que ésa no es una buena elección. Tomaremos mejor el lapso armónico, que en esta aproximación corresponde a ($K = \sum_i K_{ii}$)

$$\partial_t a = -K \ . \tag{139}$$

El siguiente paso es reescribir estas ecuaciones como un sistema de primer orden. Para ello debemos introducir las variables auxiliares

$$A_i := \partial_i a \ , \qquad D_{ijk} := \frac{1}{2} \partial_i h_{jk} \ . \tag{140}$$

El sistema de ecuaciones toma ahora la forma

$$\partial_t a = -K \ , \tag{141}$$

$$\partial_t h_{ij} = -2K_{ij} \ , \tag{142}$$

$$\partial_t A_i = -\partial_i K \ , \tag{143}$$

$$\partial_t D_{ijk} = -\partial_i K_{jk} \ , \tag{144}$$

$$\partial_t K_{ij} = -\partial_{(i} A_{j)} + \sum_k \left(2\partial_{(i} D_{kkj)} - \partial_{(i} D_{j)kk} - \partial_k D_{kij} \right). \tag{145}$$

Para hacer el análisis de hiperbolicidad notamos primero que las variables a y h_{ij} evolucionan sólo con términos de fuente (no hay derivadas del lado derecho). Además, no aparecen en las siguientes ecuaciones, por lo que pueden ignorase al estudiar la hiperbolicidad. Nos concentraremos entonces en el subsistema $\{A_i, D_{ijk}, K_{ij}\}$. Como el sistema es tridimensional, en principio debemos construir las tres matrices jacobianas. Analizaremos únicamente la dirección x, las otras dos direcciones son enteramente análogas.

Si consideramos sólo derivadas en la dirección x, notamos que únicamente debemos considerar las componentes A_x y D_{xjk}, pues las componentes A_q y D_{qjk}, para q distinta de x, se pueden tomar como fijas en este caso (podemos asumir que son cero sin afectar el análisis). Haciendo un poco de álgebra encontramos que el sistema a analizar toma la forma

$$\partial_t A_x + \partial_x \left(K_{xx} + K_{yy} + K_{zz} \right) = 0 \ , \tag{146}$$

$$\partial_t K_{xx} + \partial_x A_x + \partial_x \left(D_{xyy} + D_{xzz} \right) = 0 \ , \tag{147}$$

$$\partial_t K_{yy} + \partial_x D_{xyy} = 0 \ , \tag{148}$$

$$\partial_t K_{zz} + \partial_x D_{xzz} = 0 \ , \tag{149}$$

$$\partial_t K_{xy} = 0 \ , \tag{150}$$

$$\partial_t K_{xz} = 0 \ , \tag{151}$$

$$\partial_t K_{yz} + \partial_x D_{xyz} = 0 \ , \tag{152}$$

$$\partial_t D_{xjk} + \partial_x K_{jk} = 0. \tag{153}$$

Éste es un sistema de trece variables:

$$u := \{A_x, K_{xx}, K_{yy}, K_{zz}, K_{xy}, K_{xz}, K_{yz}, D_{xxx}, D_{xyy}, D_{xzz}, D_{xxy}, D_{xxz}, D_{xyz}, \}$$

La matriz jacobiana resulta ser

$$M=\begin{pmatrix}
0&1&1&1&0&0&0&0&0&0&0&0&0\\
1&0&0&0&0&0&0&0&1&1&0&0&0\\
0&0&0&0&0&0&0&0&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&1&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1\\
0&1&0&0&0&0&0&0&0&0&0&0&0\\
0&0&1&0&0&0&0&0&0&0&0&0&0\\
0&0&0&1&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&1&0&0&0&0&0&0&0\\
0&0&0&0&0&0&1&0&0&0&0&0&0
\end{pmatrix}. \quad (154)$$

La matriz anterior, de 13×13 elementos, tal vez parezca muy grande, pero la mayor parte de sus entradas son ceros, por lo que no es difícil encontrar sus eigenvalores. Éstos resultan ser: $+1$ con multiplicidad 4, -1 con multiplicidad 4, y 0 con multiplicidad 5, para un total de 13. Como todos estos eigenvalores son reales, el sistema es hiperbólico. Más interesantes son los eigenvectores. Asociados al eigenvalor $\lambda=-1$ tenemos los siguientes tres eigenvectores:

$$v_1^- = (1,-1,0,0,0,0,0,1,0,0,0,0,0), \quad (155)$$

$$v_2^- = (0,0,1,-1,0,0,0,0,-1,1,0,0,0), \quad (156)$$

$$v_3^- = (0,0,0,0,0,0,1,0,0,0,0,0,-1); \quad (157)$$

asociados al eigenvalor $\lambda=+1$ los tres eigenvectores:

$$v_1^+ = (1,1,0,0,0,0,0,1,0,0,0,0,0), \quad (158)$$

$$v_2^+ = (0,0,1,-1,0,0,0,0,1,-1,0,0,0), \quad (159)$$

$$v_3^+ = (0,0,0,0,0,0,1,0,0,0,0,0,1); \quad (160)$$

y asociados al eigenvalor $\lambda=0$ los tres eigenvalores:

$$v_1^0 = (0,0,0,0,0,0,0,1,0,0,0,0,0), \quad (161)$$

$$v_2^0 = (0,0,0,0,0,0,0,0,0,0,1,0,0), \quad (162)$$

$$v_3^0 = (0,0,0,0,0,0,0,0,0,0,0,1,0); \quad (163)$$

para un total de solo nueve. Es decir, no hay un conjunto completo de eigenvectores y el sistema es sólo débilmente hiperbólico. Como hemos mencionado, un sistema débilmente hiperbólico no es bien comportado, por lo que deberíamos esperar problemas cuando utilizamos ADM para simulaciones numéricas.

Es posible resolver el problema de la hiperbolicidad débil de ADM añadiendo múltiplos de las constricciones en lugares adecuados. Para ver esto, escribimos la aproximación lineal a las constricciones que resulta ser

$$\sum_j \partial_j f_j = 0, \qquad (\text{hamiltoniana}) \quad (164)$$

$$\sum_j \partial_j K_{ij} - \partial_i K = 0, \qquad (\text{momento},) \quad (165)$$

donde hemos definido

$$f_i := \sum_k \partial_k h_{ik} - \partial_i h = 2\sum_k \left(D_{kki} - D_{ikk}\right).$$

Si consideramos sólo derivadas en la dirección x (e ignoramos las componentes D_{qjk} con q distinta de x), las constricciones toman la forma

$$\partial_x \left(D_{xyy} + D_{xzz}\right) = 0, \quad (166)$$

$$\partial_x \left(K_{yy} + K_{zz}\right) = 0, \quad \partial_x K_{xy} = 0, \quad \partial_x K_{xz} = 0. \quad (167)$$

Substituyendo estas constricciones en las ecuaciones de evolución, la matriz jacobiana se transforma en

$$M=\begin{pmatrix}
0&1&0&0&0&0&0&0&0&0&0&0&0\\
1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&1&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1\\
0&1&0&0&0&0&0&0&0&0&0&0&0\\
0&0&1&0&0&0&0&0&0&0&0&0&0\\
0&0&0&1&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&1&0&0&0&0&0&0
\end{pmatrix}. \quad (168)$$

Los eigenvalores resultan ser los mismos que antes, pero ahora sí se encuentra un conjunto completo de eigenvectores. La lección aprendida es la siguiente: substituyendo las constricciones en las ecuaciones de evolución es posible transformar a las ecuaciones linearizadas de ADM en un sistema fuertemente hiperbólico.

En el caso general, cuando estamos en un campo no linealizado y con un vector de corrimiento distinto de cero, se puede hacer un análisis similar. El problema es que en ese caso no es nada evidente cómo se deben substituir las constricciones en las ecuaciones de evolución para obtener sistemas hiperbólicos. De hecho hay muchas formas distintas de hacerlo, inclusive se han construido familias multiparamétricas de formulaciones hiperbólicas, entre las que se pueden mencionar las formulaciones de Bona-Masso [26, 36], Frittelli-Reula [37], Kidder-Scheel-Teukolsy [38], entre otras. Incluso es posible mostrar que la formulación BSSN que se introdujo en la sección anterior es fuertemente hiperbólica bajo ciertas suposiciones [39].

Hoy en día, el problema de encontrar criterios que digan qué formulación hiperbólica es mejor que otra se ha convertido en un área de investigación de gran actualidad, y varios grupos trabajan activamente en ella en diversos lugares del mundo.

5. Aproximaciones en diferencias finitas

Las teorías de campo juegan un papel fundamental en la física moderna. Desde la electrodinámica clásica de Maxwell, hasta las teorías cuánticas de campo, pasando por la ecuación de Schrödinger, la hidrodinámica y, desde luego, la relatividad general, la noción de campo como una entidad física en sí misma ha tenido implicaciones profundas en nuestra concepción del Universo. Los campos son funciones continuas del espacio y el tiempo, y la descripción matemática de sus leyes dinámicas se realiza en el contexto de las ecuaciones diferenciales parciales.

Las ecuaciones diferenciales parciales asociadas a teorías físicas son en general imposibles de resolver analíticamente excepto en casos muy idealizados. Esta dificultad puede tener diversos orígenes, de la presencia de fronteras irregulares a la existencia de términos no lineales en las ecuaciones mismas. Para resolver este tipo de ecuaciones en situaciones dinámicas generales resulta inevitable utilizar aproximaciones numéricas.

Existen muchas formas distintas de resolver ecuaciones diferenciales parciales de forma numérica. Los métodos más populares son tres: las "diferencias finitas" [40], los "elementos finitos" [41] y los "métodos espectrales". En este curso nos limitaremos a estudiar las diferencias finitas por ser el método conceptualmente más simple y también por ser el más común en la relatividad numérica (aunque no el único, en particular los métodos espectrales han comenzado a ganar popularidad en los últimos años [42–44]).

5.1. Ideas fundamentales en las diferencias finitas

Cuando uno estudia un campo en un espacio-tiempo continuo, se ve en la necesidad de considerar un número infinito (y no contable) de variables desconocidas: el valor del campo en todo punto del espacio y a todo tiempo. Para encontrar el valor del campo utilizando aproximaciones numéricas primero se debe reducir el número de variables a una cantidad finita. Hay muchas formas de hacer esto. Los métodos espectrales, por ejemplo, expanden la solución como combinación lineal finita de una base adecuada de funciones. Las variables a resolver son entonces los coeficientes de dicha serie. Un enfoque distinto es tomado por las diferencias y los elementos finitos. En ambos casos se reduce el número de variables discretizando el dominio de dependencia de las funciones, aunque con distintas estrategias.

La idea básica de las diferencias finitas es substituir al espacio-tiempo continuo por un conjunto discreto de puntos. Este conjunto de puntos se conoce como la "malla" o "red" computacional. Las distancias en el espacio entre los puntos de esta red no tienen porqué ser uniformes, pero en estas notas asumiremos que sí lo son. El paso de tiempo entre dos niveles consecutivos se denomina Δt, y la distancia entre dos puntos adyacentes en el espacio Δx. La Fig. 6 es una representación gráfica de la red computacional.

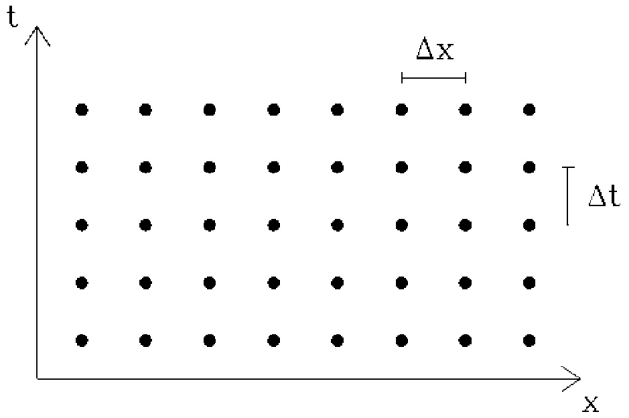

FIGURA 6. Discretización del espacio-tiempo utilizada en diferencias finitas.

Una vez establecida la malla computacional el siguiente paso es subsitituir las ecuaciones diferenciales por un sistema de ecuaciones algebraicas. Esto se logra aproximando los operadores diferenciales por diferencias finitas entre los valores de las funciones en puntos adyacentes de la malla. De esta forma se obtiene una ecuación algebraica en cada punto de la malla por cada ecuación diferencial. Estas ecuaciones algebraicas involucran los valores de las funciones en el punto considerado y en sus vecinos más cercanos. El sistema de ecuaciones algebraicas se puede resolver de manera sencilla, el precio que hemos pagado es que ahora tenemos muchísimas ecuaciones algebraicas, por lo que se requiere utilizar una computadora.

Para ver cómo se hace esto en la práctica, consideremos como módelo la ecuación de onda en una dimensión. Esta ecuación presenta muchas ventajas. En primer lugar se puede resolver de manera exacta y la solución exacta se puede utilizar para comparar con la solución numérica. Además, la mayoría de las ecuaciones fundamentales en las teorías de campo modernas se pueden ver como generalizaciones de diversos tipos de la ecuación de onda.

5.2. La ecuación de onda en una dimensión

La ecuación de onda en una dimensión (en un espacio plano) tiene la forma

$$\frac{\partial^2 \phi}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 \phi}{\partial t^2} = 0 \,, \tag{169}$$

donde ϕ es la función de onda y c la velocidad de onda.

Para encontrar una aproximación en diferencias finitas a esta ecuación comenzamos por introducir la siguiente notación para los valores de ϕ en la red computacional:

$$\phi^n_m := \phi\left(n\Delta t,\, m\Delta x\right) \,. \tag{170}$$

Podemos aproximar los operadores diferenciales que aparecen en la Ec. 169 utilizando expansiones en Taylor de ϕ alrededor del punto (n, m). Consideremos el valor de ϕ en los

puntos $(n, m+1)$ y $(n, m-1)$:

$$\phi^n_{m+1} = \phi^n_m + \left(\frac{\partial\phi}{\partial x}\right)\Delta x + \frac{1}{2}\left(\frac{\partial^2\phi}{\partial x^2}\right)(\Delta x)^2 + \frac{1}{6}\left(\frac{\partial^3\phi}{\partial x^3}\right)(\Delta x)^3 + \cdots , \quad (171)$$

$$\phi^n_{m-1} = \phi^n_m - \left(\frac{\partial\phi}{\partial x}\right)\Delta x + \frac{1}{2}\left(\frac{\partial^2\phi}{\partial x^2}\right)(\Delta x)^2 - \frac{1}{6}\left(\frac{\partial^3\phi}{\partial x^3}\right)(\Delta x)^3 + \cdots , \quad (172)$$

donde las derivadas son evaluadas en el punto $(t = n\,\Delta t,\, x = m\,\Delta x)$. A partir de estas expresiones vemos que

$$\left(\frac{\partial^2\phi}{\partial x^2}\right) = \frac{\phi^n_{m+1} - 2\,\phi^n_m + \phi^n_{m-1}}{(\Delta x)^2} + \frac{(\Delta x)^2}{12}\left(\frac{\partial^4\phi}{\partial x^4}\right) + \cdots . \quad (173)$$

Podemos entonces aproximar la segunda derivada como

$$\left(\frac{\partial^2\phi}{\partial x^2}\right) \simeq \frac{\phi^n_{m+1} - 2\,\phi^n_m + \phi^n_{m-1}}{(\Delta x)^2} . \quad (174)$$

Qué tan buena es esta aproximación depende, desde luego, del tamaño de la malla Δx. Si Δx es pequeño en el sentido de que la función ϕ varía muy poco en una región de ese tamano, entonces la aproximación puede ser muy buena. El error involucrado en esta aproximación es llamado "error de truncación" y, como vemos, su parte dominante resulta ser proporcional a Δx^2, por lo que se dice que esta aproximación es de segundo orden.

La segunda derivada de ϕ con respecto a t se puede aproximar exactamente de la misma manera. De esta forma obtenemos la siguiente aproximación a segundo orden de la ecuación de onda:

$$\frac{\phi^n_{m+1} - 2\,\phi^n_m + \phi^n_{m-1}}{(\Delta x)^2} - \frac{1}{c^2}\,\frac{\phi^{n+1}_m - 2\,\phi^n_m + \phi^{n-1}_m}{(\Delta t)^2} = 0 . \quad (175)$$

Podemos reescribir esta ecuación de forma más compacta si introducimos el llamado "parámetro de Courant": $\rho := c\Delta t/\Delta x$. La aproximación toma la forma final

$$\rho^2\,\left(\phi^n_{m+1} - 2\,\phi^n_m + \phi^n_{m-1}\right) - \left(\phi^{n+1}_m - 2\,\phi^n_m + \phi^{n-1}_m\right) = 0 . \quad (176)$$

Esta ecuación tiene una propiedad muy importante: involucra sólo un valor de la función de onda en el último nivel de tiempo, el valor ϕ^{n+1}_m. Podemos entonces despejar este valor en términos de valores en tiempos anteriores para obtener

$$\phi^{n+1}_m = 2\,\phi^n_m - \phi^{n-1}_m + \rho^2\,\left(\phi^n_{m+1} - 2\,\phi^n_m + \phi^n_{m-1}\right) . \quad (177)$$

Por esta propiedad la aproximación anterior se conoce como "aproximación explícita". Si conocemos los valores de la función ϕ en los niveles n y $n-1$, podemos usar la ecuación anterior para calcular directamente los valores de la función en el nuevo paso de tiempo $n+1$. El proceso puede luego iterarse tantas veces como se quiera. Es evidente que todo lo que se necesita para comenzar la evolución es el conocimiento del valor de la función de onda en los primeros dos pasos de tiempo. Pero obtener estos datos es muy fácil de hacer. Como se trata de una ecuación de segundo orden, los datos iniciales incluyen

$$f\,(x) := \phi\,(0, x) , \qquad g\,(x) := \left.\frac{\partial\phi}{\partial t}\right|_{t=0} . \quad (178)$$

El conocimiento de $f(x)$ evidentemente nos da el primer nivel de tiempo:

$$\phi^0_m \;=\; f\,(m\,\Delta x) . \quad (179)$$

Para el segundo nivel basta con aproximar la primera derivada en el tiempo en diferencias finitas. Una posible aproximación es

$$g\,(m\,\Delta x) \;=\; \frac{\phi^1_m - \phi^0_m}{\Delta t} , \quad (180)$$

de donde obtenemos

$$\phi^1_m \;=\; g\,(m\,\Delta x)\;\Delta t \;+\; \phi^0_m . \quad (181)$$

La expresión anterior tiene un inconveniente importante. De la expansión en Taylor podemos ver fácilmente que el error de truncación para esta expresión es de orden Δt, por lo que la aproximación es sólo de primer orden. Sin embargo, es fácil resolver este problema. Una aproximación a segundo orden resulta ser

$$g\,(m\,\Delta x) \;=\; \frac{\phi^1_m - \phi^{-1}_m}{2\,\Delta t} . \quad (182)$$

El problema ahora es que esta expresión hace referencia al valor de la función ϕ^{-1}_m que también es desconocido. Pero ya tenemos otra ecuación que hace referencia a ϕ^1_m y ϕ^{-1}_m: la aproximación a la ecuación de onda (176) evaluada en $n = 0$. Podemos entonces utilizar estas dos ecuaciones para eliminar ϕ^{-1}_m y resolver para ϕ^1_m. De esta forma obtenemos la siguiente aproximación a segundo orden para el segundo nivel de tiempo:

$$\phi^1_m \;=\; \phi^0_m \;+\; \frac{\rho^2}{2}\,\left(\phi^0_{m+1} - 2\,\phi^0_m + \phi^0_{m-1}\right) \;+\; \Delta t\; g\,(m\,\Delta x) . \quad (183)$$

Las Ecs. (179) y (183) nos dan ahora toda la información necesaria para comenzar la evolución.

Hay un punto importante que debe mencionarse aqui. Para reducir el número total de variables a un número finito, también es necesario tomar una región finita del espacio, conocida como el "dominio computacional", con un número finito de puntos N. Resulta entonces crucial especificar las condiciones de frontera que deberán aplicarse a las orillas de la red computacional. Es claro que la aproximación a la ecuación de onda (176) no puede ser utilizada en las fronteras debido a que hace referencia a puntos fuera del dominio computacional. Existen diversas maneras conocidas de imponer condiciones de frontera para la ecuación de onda. Sin embargo, para sistemas más complejos, como las ecuaciones de Einstein, la elección de condiciones de frontera adecuadas y consistentes es un problema no resuelto de gran actualidad [45, 46]. Para los objetivos de estas notas podemos olvidarnos del problema de las condiciones de frontera y asumir que tenemos un espacio periódico, de manera que la elección de las condiciones de frontera será simplemente

$$\phi_0^n \equiv \phi_N^n \, . \tag{184}$$

Esta elección, además de ser sumamente simple, es equivalente a utilizar la aproximación interna en todos lados, por lo que permite concentrarnos en las propiedades del método interior únicamente, sin preocuparnos por los efectos que puedan introducir las fronteras.

5.3. Aproximaciones implícitas y moléculas computacionales

En la sección anterior se introdujeron las ideas básicas de las aproximaciones en diferencias finitas utilizando la ecuación de onda en una dimensión como ejemplo. La aproximación que encontramos, sin embargo, está lejos de ser única. En principio, existe un número infinito de formas distintas de aproximar una misma ecuación diferencial utilizando diferencias finitas. Diferentes aproximaciones tienen distintas propiedades. En la siguiente sección se mencionarán cuáles de estas propiedades pueden hacer una aproximación más útil que otra. En esta sección me limitaré a introducir una generalización de la aproximación explícita que vimos antes.

Para hacer las cosas más fáciles, introduciremos una notación compacta para las diferencia finitas. Definimos los operadores de "primeras diferencias centradas" como

$$\delta_x \, \phi_m^n := \frac{1}{2} \left(\phi_{m+1}^n - \phi_{m-1}^n \right) \, , \tag{185}$$

$$\delta_t \, \phi_m^n := \frac{1}{2} \left(\phi_m^{n+1} - \phi_m^{n-1} \right) \, , \tag{186}$$

y los operadores de "segundas diferencias centradas" como

$$\delta_x^2 \, \phi_m^n := \phi_{m+1}^n - 2 \, \phi_m^n + \phi_{m-1}^n \, , \tag{187}$$

$$\delta_t^2 \, \phi_m^n := \phi_m^{n+1} - 2 \, \phi_m^n + \phi_m^{n-1} \, . \tag{188}$$

(Cuidado: Con esta notación $(\delta_x)^2 \neq \delta_x^2$).

Habiendo definido estos operadores, podemos regresar a las aproximaciones que hicimos a los operadores diferenciales que aparecen en la ecuación de onda. Partiendo de nuevo de las series de Taylor, es posible mostrar que la segunda derivada espacial puede aproximarse de manera más general como

$$\left(\frac{\partial^2 \phi}{\partial x^2} \right) \simeq \frac{1}{(\Delta x)^2} \, \delta_x^2 \left[\frac{\theta}{2} \left(\phi_m^{n+1} + \phi_m^{n-1} \right) + (1-\theta) \, \phi_m^n \right] , \tag{189}$$

con θ un parámetro arbitrario. La expresión que teníamos antes, Ec. (174), puede recuperarse tomando $\theta = 0$. Esta nueva aproximación corresponde a tomar promedios, con un cierto peso, de operadores en diferencias finitas actuando en diferentes pasos de tiempo. En el caso particular en que $\theta = 1$, la contribución del paso de tiempo intermedio de hecho desaparece por completo.

Si utilizamos esta aproximación para la segunda derivada espacial, pero mantenemos la aproximación anterior para la derivada temporal, obtenemos la siguiente aproximación en diferencias finitas de la ecuación de onda:

$$\rho^2 \, \delta_x^2 \left[\frac{\theta}{2} \left(\phi_m^{n+1} + \phi_m^{n-1} \right) + (1-\theta) \, \phi_m^n \right] - \delta_t^2 \, \phi_m^n = 0 \, . \tag{190}$$

Ésta es una posible generalización de (176), pero claramente no la única. Es claro que podemos jugar este juego de muchas maneras y obtener aproximaciones aún más generales, todas ellas válidas, y todas ellas a segundo orden (incluso es posible ingeniárselas para encontrar aproximaciones a cuarto o mayor orden). La aproximación dada por (190) tiene una nueva propiedad muy importante: hace referencia no a uno, sino a tres valores diferentes de ϕ en el último paso de tiempo. Esto significa que ahora no podemos resolver para ϕ en el último paso de manera explícita en términos de los valores en los dos pasos anteriores. Debido a esto, a la aproximación (190) se le conoce como "aproximación implícita".

Cuando las ecuaciones para todos los puntos de la malla, incluyendo las fronteras, se consideran a la vez, es posible resolver el sistema completo invirtiendo una matriz no trivial, lo que desde luego toma más tiempo que el necesario para una aproximación explícita. Parecería entonces que no tiene ningún sentido considerar aproximaciones de tipo implícito. Sin embargo, en muchas ocasiones las aproximaciones implícitas resultan tener mejores propiedades que las explícitas, en particular relacionadas con la estabilidad del esquema numérico, concepto que se discutirá en la siguiente sección.

La diferencia entre una aproximación implícita y otra explícita puede verse gráficamente utilizando el concepto de "molécula computacional", que no es otra cosa que un diagrama que muestra las relaciones entre los distintos puntos utilizados en una aproximación en diferencias finitas. La Fig.7 muestra las moléculas computacionales para los casos implícito y explícito que hemos considerado.

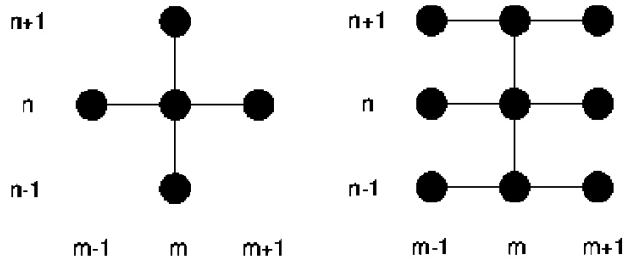

FIGURA 7. Moléculas computacionales.

5.4. Consistencia, convergencia y estabilidad

En la última sección nos topamos con quizá uno de los conceptos más importantes de las diferencias finitas: existe, en general, un número infinito de maneras posibles de aproximar una misma ecuación diferencial. Esto sigue siendo cierto incluso si uno se restringe a considerar aproximaciones con un mismo orden de error.

La multiplicidad de posibles aproximaciones nos lleva inmediatemente a la siguiente pregunta: ¿Cómo saber en qué caso usar una cierta aproximación y no otra? Desgraciadamente, no existe una respuesta general a esta pregunta, por lo que muchas veces se dice que las diferencias finitas son un arte más que una ciencia. Sin embargo, sí existen algunas guías que nos permiten escoger aproximaciones en ciertos casos. Estas guías tienen que ver con los conceptos de consistencia, convergencia y estabilidad.

Consideremos una cierta aproximación en diferencias finitas a una ecuación diferencial. Cuando la malla se refina (es decir, cuando Δt y Δx se hacen cada vez más pequeños), uno esperaría que la aproximación fuera cada vez mejor en el sentido de que los errores de truncación se hacen cada vez más pequeños. Decimos entonces que en el límite continuo nuestra aproximación se acerca a la ecuación diferencial original y no a otra. Cuando esto ocurre localmente se dice que nuestra aproximación es "consistente". En general, esta propiedad es muy fácil de ver de la estructura de las aproximaciones en diferencias finitas, y puede comprobarse casi a ojo. Las excepciones importantes son situaciones en las que el sistema de coordenadas es singular, donde mostrar consistencia en el punto singular puede no ser trivial. Por ejemplo, es común que aproximaciones en diferencias finitas "estándar" fallen en el punto $r = 0$ cuando se utilizan coordenadas esféricas. La consistencia es claramente fundamental en una aproximación en diferencias finitas. Si falla, aunque sea en un solo punto, implica que no recuperaremos la solución correcta de la ecuación diferencial.

La consistencia es sólo una propiedad local: una aproximación consistente se reduce *localmente* a la ecuación diferencial en el límite continuo. En la práctica, estamos realmente interesados en una propiedad más global. Lo que realmente buscamos es que la aproximación mejore *a un tiempo finito* T cuando refinamos la malla. Es decir, la diferencia entre la solución exacta y la solución numérica a un tiempo fijo T debe tender a cero en el límite continuo. Esta condición se conoce como "convergencia".

La convergencia es claramente diferente a la consistencia: esquemas consistentes pueden fácilmente no ser convergentes. Esto no es difícil de entender si pensamos que en el límite cuando Δt se hace cero, un tiempo finito T se puede alcanzar solo después de un número infinito de pasos. Esto implica que incluso si el error en cada paso es infinitesimal, su integral total puede muy fácilmente ser finita. La solución numérica puede incluso diverger y el error de hecho resultar infinito.

En general es muy difícil verificar analíticamente si un esquema de aproximación es convergente o no lo es. Numéricamente, por otro lado, es muy fácil ver si la solución aproximada converge a algo (es decir, no diverge). Lo difícil es saber si la solución numérica coverge hacia la solución exacta y no a otra cosa.

Hay otra propiedad importante de las aproximaciones en diferencias finitas. Independientemente del comportamiento de la solución a la ecuación diferencial, debemos pedir que las *soluciones exactas de las ecuaciones en diferencias finitas* permanezcan acotadas para cualquier tiempo finito T y cualquier intervalo de tiempo Δt. Este requisito se conoce como "estabilidad", e implica que ninguna componente de los datos iniciales debe amplificarse arbitrariamente. La estabilidad es una propiedad del sistema de ecuaciones en diferencias finitas, y no tiene nada que ver con la ecuación diferencial que estamos aproximando.

Un resultado fundamental de la teoría de las aproximaciones en diferencias finitas es el teorema de Lax (para su demostración ver, por ejemplo, la Ref. 40):

TEOREMA: *Dado un problema de valores iniciales bien planteado matemáticamente y una aproximación en diferencias finitas a él que satisface la condición de consistencia, entonces la estabilidad es condición necesaria y suficiente para la covergencia*.

Este teorema relaciona el objetivo final de toda aproximación en diferencias finitas, es decir, la convergencia a la solución exacta, con una propiedad que es mucho más fácil de probar: la estabilidad.

5.5. Estabilidad de von Newmann

Un método general para probar la estabilidad de los sistemas de ecuaciones en diferencias finitas se obtiene de la definición de estabilidad directamente. Comenzamos por escribir las ecuaciones en diferencias finitas como

$$\mathbf{v}^{n+1} = \mathbf{B}\,\mathbf{v}^{n}\,, \tag{191}$$

donde $\mathbf{v}^n$ es el vector solución en el nivel de tiempo n y $\mathbf{B}$ es una matriz (en general con muchos ceros). Es importante notar que toda aproximación en diferencias finitas, incluso aquellas que involucran más de dos niveles de tiempo (como las que introdujimos para la ecuación de onda en las secciones anteriores), puede escribirse de la forma (191) simplemente introduciendo variables auxiliares.

Como el vector $\mathbf{v}^n$ se puede escribir como una combinación lineal de los eigenvectores de $\mathbf{B}$, el requerimiento de

estabilidad se reduce a pedir que la matriz $\mathbf{B}$ no amplifique ninguno de sus eigenvectores, es decir, que la magnitud del mayor de sus eigenvalores sea menor o igual que 1. Dicho de otra forma, el "radio espectral" de $\mathbf{B}$ debe ser menor o igual a 1.

El análisis de estabilidad basado en la idea que hemos expuesto es muy general, pero requiere del conocimiento de los coeficientes de $\mathbf{B}$ en todo el espacio, incluida la frontera. Existe, sin embargo, un método muy popular de análisis de estabilidad que, aunque en principio sólo nos proporciona condiciones necesarias para la estabilidad, en muchos casos resulta también dar condiciones suficientes. Este método, introducido originalmente por von Newmann, está basado en una descomposición en Fourier de la solución.

Para introducir este método, empezamos entonces por expandir la solución de (191) en una serie de Fourier:

$$\mathbf{v}^n\left(\mathbf{x}\right) \,=\, \sum_{\mathbf{k}} \tilde{\mathbf{v}}^n\left(\mathbf{k}\right)\, e^{i\,\mathbf{k}\cdot\mathbf{x}}\ , \tag{192}$$

donde la suma es sobre todos los vectores de onda $\mathbf{k}$ que pueden representarse en la malla[i]. Si ahora substitumos esto en la ecuación original (191) obtenemos

$$\tilde{\mathbf{v}}^{n+1} \,=\, \mathbf{G}\left(\Delta x,\, \Delta t,\, \mathbf{k}\right)\, \tilde{\mathbf{v}}^n\ . \tag{193}$$

La matriz $\mathbf{G}$ se conoce como "matriz de amplificación". La condición de estabilidad ahora corresponde a pedir que ningun modo de Fourier se amplifique, es decir, el radio espectral de $\mathbf{G}$ debe ser menor o igual que uno. Ésta es la condición de estabilidad de von Newmann. Es importante recalcar que para poder utilizar este criterio de estabilidad se han supuesto dos cosas: 1) Las condiciones de frontera son períodicas, y 2) Los elementos de la matriz $\mathbf{B}$ son constantes.

Como ejemplo del análisis de estabilidad de von Newmann vamos a utilizar la aproximación implícita a la ecuación de onda que derivamos antes [Ec. (190)]:

$$\rho^2\,\delta_x^2\,\left[\left(\theta/2\right)\,\left(\phi_m^{n+1}\,+\,\phi_m^{n-1}\right)+\left(1-\theta\right)\,\phi_m^n\right] \\ -\delta_t^2\,\phi_m^n\,=\,0\ . \tag{194}$$

Consideraremos ahora un modo de Fourier de la forma

$$\phi_m^n \,=\, \xi^n\, e^{i\,m\,k\,\Delta x}\ . \tag{195}$$

Si substituimos esto en la ecuación en diferencias finitas encontramos, después de un poco de álgebra, la siguiente ecuación cuadrática para ξ:

$$A\,\xi^2\,+\,B\,\xi\,+\,C\,=\,0\ , \tag{196}$$

con coeficientes dados por

$$A = \rho^2\,\theta\,\Big[\cos\left(k\,\Delta x\right)-1\Big]\,-\,1\ , \tag{197}$$

$$B = 2\,\rho^2\,\left(1-\theta\right)\,\Big[\cos\left(k\,\Delta x\right)-1\Big]\,+\,2\ , \tag{198}$$

$$C = \rho^2\,\theta\,\Big[\cos\left(k\,\Delta x\right)-1\Big]\,-\,1\ . \tag{199}$$

La dos raices de la ecuación cuadrática son, claramente,

$$\xi_\pm \,=\, \frac{-B\,\pm\,\left(B^2\,-\,4\,A\,C\right)^{1/2}}{2\,A}\ , \tag{200}$$

y la solución general a la ecuación en diferencias finitas resulta ser

$$\phi_m^n = \sum_k \Big[Z_k^+\,\Big(\xi_+\left(k\right)\Big)^n \\ +Z_k^-\,\Big(\xi_-\left(k\right)\Big)^n\Big]\,e^{imk\Delta x}, \tag{201}$$

donde Z_k^+ y Z_k^- son constantes arbitrarias.

Por otro lado, del hecho de que $A = C$ es fácil mostrar que

$$\left|\xi_+\,\xi_-\right| \,=\, \left|\frac{C}{A}\right| \,=\, 1\ . \tag{202}$$

Ésta es una propiedad muy importante, implica que si el sistema es estable para toda k, es decir, si $|\xi_\pm(k)| \,\leq\, 1$, entonces necesariamente será también no disipativo (los modos de Fourier no sólo no crecen, sino que tampoco se disipan). Para que el sistema sea estable debemos ahora pedir que

$$\xi_+(k) \,=\, \xi_-(k) \,=\, 1\ . \tag{203}$$

Es fácil ver que esto ocurrirá siempre que

$$B^2\,-\,4\,A\,C\,\leq\,0\ . \tag{204}$$

Substituyendo los valores de A, B y C en esta expresión obtenemos la siguiente condición de estabilidad:

$$\rho^2\,\left(1-2\,\theta\right)\,\Big[1\,-\,\cos\left(k\,\Delta x\right)\Big]\,-\,2\,\leq\,0\ . \tag{205}$$

Como esto se debe cumplir para toda k, debemos considerar el caso cuando el lado izquierdo es máximo. Para $\theta < 1/2$ esto ocurrirá para $k = \pi/\Delta x$, en cuyo caso la condición de estabilidad se reduce a

$$\rho^2\,\leq\,1/\left(1-2\,\theta\right)\ . \tag{206}$$

Para el esquema explícito se tiene $\theta = 0$, y la condición se reduce a la bien conocida condición de estabilidad de Courant-Friedrich-Lewy (CFL):

$$c\Delta t \leq \Delta x\ . \tag{207}$$

La condición CFL tiene una clara interpretación geométrica: el dominio de dependencia numérico debe ser mayor que el dominio de dependencia físico y no al revés (ver Fig. 8). Si esto no fuera así, resultaría imposible para la solución numérica converger a la solución exacta, pues al refinar la malla siempre habría información física relevante que quedaría fuera del dominio de dependencia numérico. Y, como hemos visto, el teorema de Lax implica que si no hay convergencia el sistema es inestable.

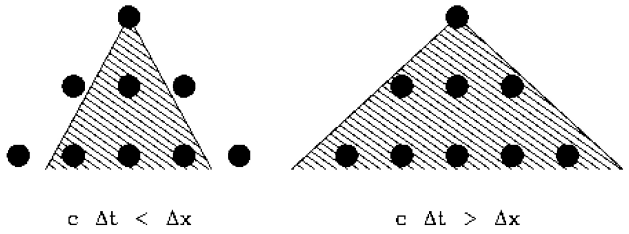

FIGURA 8. Condición de estabilidad CFL. Para $c\,\Delta t < \Delta x$, el dominio de dependencia numérico es mayor que el dominio de dependencia físico (región sombreada), y el sistema es estable. Para $c\,\Delta t > \Delta x$, la situación es la opuesta, y el sistema es inestable.

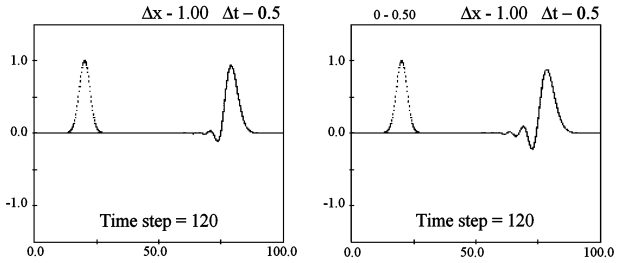

FIGURA 9. Simulación numérica para el caso $\rho = 1/2$, utilizando un sistema explícito (izquierda), y uno implícito con $\theta = 1/2$ (derecha).

El argumento que acabamos de dar claramente sólo se aplica a esquema explícitos. Esto se debe a que para un sistema implícito, el dominio de dependencia numérico es toda la malla. En este caso no hay un argumento físico simple que nos diga cúal debe ser la condición de estabilidad.

Para llegar a la condición de estabilidad (206), supusimos que $\theta < 1/2$. Si, por otro lado, tomamos $\theta \geq 1/2$, debemos regresar a la condición general (205). Sin embargo, en este caso es fácil ver que la condición se satisface siempre. Esto significa que un sistema implícito con $\theta \geq 1/2$ es estable para todo valor de ρ, es decir, es "incondicionalmente estable".

Esto nos lleva a quizá una de las lecciones más importantes de la teoría de las diferencias finitas: los esquemas más simples no siempre tienen las mejores propiedades de estabilidad.

5.6. Ejemplos

Como última parte de esta introducción a las diferencias finitas, vamos a considerar los resultados de experimentos numéricos utilizando la ecuación de onda y nuestra aproximación (190). En todas las simulaciones mostradas aquí se toma como región de integración $x \in [0, 100]$ y se utilizan condiciones de frontera periódicas. También tomamos la velocidad de onda c igual a 1.

Primero consideremos el caso cuando Δx y Δt son tales que

$$\Delta x = 1 \ , \qquad \Delta t = 1/2 \ . \qquad (208)$$

El parámetro de Courant resulta ser $\rho = 1/2$. La Fig. 9 muestra resultados de dos simulaciones usando estos parámetros de malla. La gráfica de la izquierda corresponde a un esquema explícito y la de la derecha a un sistema implícito con $\theta = 1/2$. En ambos casos los datos inciales (línea punteada) corresponden a un paquete gaussiano moviéndose inicialmente a la derecha. El resultado de la simulación numérica se muestra después de 120 pasos de tiempo.

De la figura vemos que ambas simulaciones son muy similares. En ambos casos se nota que el paquete inicial se ha dispersado, en contraste con la solución exacta para la cual el paquete se propaga manteniendo su forma original. La dispersión es un efecto numérico, ocasionado por el hecho de que, en la apoximación numérica, diferentes modos de Fourier viajan a diferentes velocidades.

La siguiente simulación corresponde a la misma situación, pero ahora tomando $\Delta x = \Delta t = 1$, correspondiente a un parámetro de Courant $\rho = 1$. La Fig. 10 muestra los resultados de esta simulación. Nótese que como Δt es ahora el doble de grande, se requieren sólo 60 pasos de tiempo para llegar a la misma etapa.

Lo primero que hay que notar es que para el esquema explícito la dispersión ha desaparecido. Esto es un resultado soprendente, y sólo ocurre para la ecuación de onda en una dimensión: en este caso es posible mostrar que para $\rho = 1$ todos los errores numéricos se cancelan y la solución numérica es de hecho exacta. El esquema implícito, por otro lado sigue siendo dispersivo (incluso más que antes).

Finalmente, en la Fig. 11 se muestran resultados de una simulación con $\Delta x = 1$ y Δt =1.02 , para un parámetro de Courant de ρ =1.02. Para el esquema explícito es evidente que una perturbación de alta frecuencia ha aparecido después de sólo 55 pasos de tiempo. Si la simulación continúa, esta perturbación crece exponencialmente y rápidamente domina la solución completa. Ésta es una clásica inestabilidad numérica: una perturbación localizada, generalmente de alta frecuencia (aunque no siempre), que crece exponencialmente sin moverse. El esquema implícito, por otro lado, no muestra ninguna señal de una inestabilidad. De hecho, podemos continuar la integración en este caso y nunca encontrar ningún problema. La dispersión, por otro lado, sí está presente y eventualmente causará que la solución numérica difiera mucho de la solución analítica.

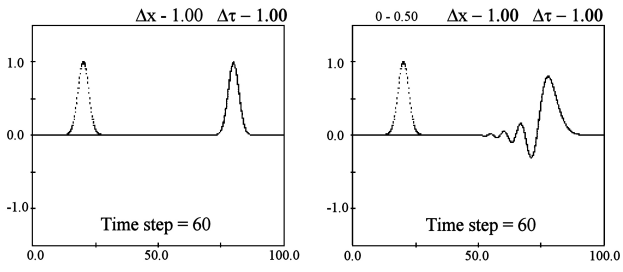

FIGURA 10. Simulación numérica para el caso $\rho = 1$.

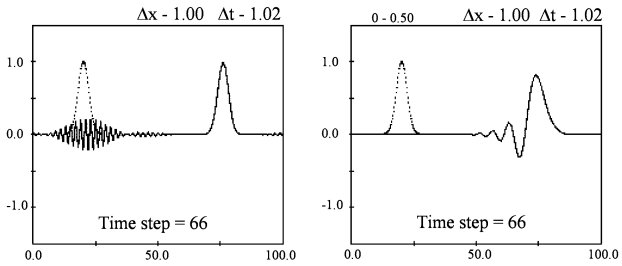

FIGURA 11. Simulación numérica para el caso $\rho = 1.02$.

6. Conclusiones

En estas notas se ha presentado una introducción a la relatividad numérica. Partiendo de una breve discusión de los principios fundamentales de la relatividad general, se ha introducido el formalismo 3+1 que es el más comúnmente utilizado en la relatividad numérica. El formalismo 3+1 lleva directamente a las ecuaciones de evolución de Arnowitt-Deser-Misner (ADM), que son el punto de partida de prácticamente todas las simulaciones numéricas de sistemas gravitacionales. También se han discutido las formulaciones alternativas de las ecuaciones de evolución que difieren de las ecuaciones ADM en la manera en que se utilizan las constricciones y tienen por tanto las mismas soluciones físicas. Sin embargo, diferentes formulaciones pueden tener distintas propiedades matemáticas. En este contexto se ha discutido el concepto de hiperbolicidad y su relación con el buen comportamiento de las soluciones. Finalmente, se ha presentado una breve introducción a las diferencias finitas, que son uno de los métodos más comunes de aproximar ecuaciones diferenciales numéricamente.

La relatividad numérica es un área de investigación que finalmente está alcanzando un estado de madurez. Esta disciplina tiene un futuro prometedor, pues representa la única manera de estudiar de manera precisa el comportamiento de sistemas astrofísicos realistas, con campos gravitacionales intensos y dinámicos. En particular, el estudio de fuentes astrofísicas de ondas gravitacionales será fundamental para poder analizar los datos provenientes de los grandes detectores que se encuentran actualmente en las fases finales de construcción y calibración.

---

i. La menor longitud de onda que se puede representar en la malla es claramente $2\,\Delta x$, también conocida como la "longitud de onda de Niquist". Esto implica que el valor máximo de cualquier componente del vector de onda es $\pi/\Delta x$.

1. S.G. Hahn y R.W. Lindquist, *Ann. Phys.*, **29** (1964) 304.
2. L. Smarr, A. Čadež, B. DeWitt, y K. Eppley, *Phys. Rev. D*, **14** N. 10 (1976) 2443.
3. K. Eppley, *The Numerical Evolution of the Collision of Two Black Holes*. PhD thesis, Princeton University, Princeton, New Jersey, (1975).
4. K. Eppley, *Phys. Rev. D*, **16** N 6 (1977) 1609.
5. M. W. Choptuik, "Universality and scaling in gravitational collapse of massless scalar field," *Phys. Rev. Lett*, **70** (1993) 9.
6. C. Gundlach, "Critical phenomena in gravitational collapse," *Living Reviews in Relativity*, **2** (1999) 4.
7. L. Lehner, "Numerical relativity: A review," *Class. Quantum Grav.* **18** (2001) R25-R86.
8. LIGO - http://www.ligo.caltech.edu/.
9. http://www.virgo.infn.it/.
10. GEO600 - http://www.geo600.uni-hannover.de/.
11. TAMA - http://tamago.mtk.nao.ac.jp/.
12. A. Einstein, "Die feldgleichungen der gravitation," *Preuss. Akad. Wiss. Berlin, Sitzber*, (1915) 844-847.
13. A. Einstein, "Zur algemeinen relativitätstheorie," *Preuss. Akad. Wiss. Berlin, Sitzber*, (1915) 778-786.
14. J. A. Wheeler, *A journey into gravity and spacetime*. distributed by W. H. Freeman, (New York, U.S.A.: Scientific American Library, 1990).
15. C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation*. (San Francisco: W. H. Freeman, 1973).
16. B. Schutz, *A First Course in General Relativity*. (Cambridge University Press, 1985).
17. R. M. Wald, *General Relativity*. (Chicago: The University of Chicago Press, 1984).
18. J. Winicour, *Living Reviews in Relativity*, **1** (1998).
19. H. Friedrich, *Proc. Roy. Soc. London*, **378** (1981) 401.
20. J. W. York, Jr., Kinematics and dynamics of general relativity, in *Sources of Gravitational Radiation* L. L. Smarr, ed., (Cambridge, UK: Cambridge University Press, 1979) p-83.
21. R. Arnowitt, S. Deser, and C. W. Misner, The dynamics of general relativity, in *Gravitation: An Introduction to Current Research* L. Witten, ed., (New York: John Wiley, 1962) p-227.
22. A. Lichnerowicz, *J. Math Pures et Appl.*, **23** (1944) 37.
23. J. W. York, *Phys. Rev. Lett.*, **82** (1999) 1350.
24. G. B. Cook, *Living Reviews in Relativity*, **2000** (200) 5.
25. L. Smarr and J. York, '*Phys. Rev. D*, **17** (1978) 2529.
26. C. Bona, J. Massó, E. Seidel, and J. Stela, *Phys. Rev. Lett.*, **75** (1995) 600.
27. A. Abrahams, D. Bernstein, D. Hobill, E. Seidel, and L. Smarr, *Phys. Rev. D*, **45** (1992) 3544.
28. P. Anninos, D. Hobill, E. Seidel, L. Smarr, and W.-M. Suen, *Phys. Rev. Lett.*, **71** N. 18 (1993) 2851.
29. M. Alcubierre, *Class. Quantum Grav.*, **20** N. 4 (2003) 607.
30. L. Smarr and J. York, *Phys. Rev.D*, **17** (1978) 1945.
31. M. Alcubierre, B. Brügmann, P. Diener, M. Koppitz, D. Pollney, E. Seidel, and R. Takahashi, *Phys. Rev. D*, **67** (2003) 084023.

32. T. W. Baumgarte and S. L. Shapiro, *Physical Review D*, **59** (1999) 024007.

33. M. Shibata and T.Ñakamura, *Phys. Rev. D*, **52** (1995) 5428.

34. M. Alcubierre, G. Allen, B. Brügmann, E. Seidel, and W.M. Suen, *Phys. Rev. D*, **62** (2000) 124011.

35. H.-O. Kreiss and J. Lorenz, *Initial-Boundary Value Problems and the Navier-Stokes Equations*. (New York: Academic Press, 1989).

36. C. Bona, J. Massó, E. Seidel, and J. Stela, *Phys. Rev. D*, **56** (1997) 3405.

37. S. Frittelli and O. Reula, *Phys. Rev. Lett.*, **76** (1996) 4667. gr-qc/9605005.

38. L. E. Kidder, M. A. Scheel, and S. A. Teukolsky, *Phys. Rev. D*, **64** (2001) 064017. gr-qc/0105031.

39. O. Sarbach, G. Calabrese, J. Pullin, and M. Tiglio, *Phys. Rev. D*, **66** (2002) 064002.

40. R. D. Richtmyer and K. Morton, *Difference Methods for Initial Value Problems*. (New York: Interscience Publishers, 1967).

41. A. R. Mitchell, *The finite element method in partial differential equations*. (U.S.A.: J. Wiley and Sons, 1977).

42. S. Bonazzola and J.-A. Marck, in *Frontiers in Numerical Relativity* C. Evans, L. Finn, and D. Hobill, eds., (Cambridge, England: Cambridge University Press, 1989) p. 239.

43. L. E. Kidder and L. S. Finn, *Phys. Rev. D*, **62** (2000) 084026.gr-qc/9911014.

44. L. E. Kidder, M. A. Scheel, S. A. Teukolsky, E. D. Carlson, and G. B. Cook, *Phys. Rev. D*, **62** (2000) 084032. gr-qc/0005056.

45. B. Szilagyi, R. Gómez, N. T. Bishop, and J. Winicour, *Phys. Rev. D*, **62** (2000). gr-qc/9912030.

46. B. Szilagyi, B. Schmidt, and J. Winicour, *Phys. Rev. D*, **65** (2002). gr-qc/0106026.