Geometría en la Relatividad General

por
José María Martín Senovilla, Universidad del País Vasco

1. Introducción

La Relatividad General es la teoría física de la gravitación. Creada por la mente lúcida de Einstein en 1907-16, ha cambiado radicalmente nuestra forma de entender el mundo prediciendo, o dando explicación científica, a fenómenos fundamentales revolucionarios tales como la expansión universal, la existencia de agujeros negros, o las lentes gravitatorias —que permiten deducir la existencia de materia oscura. Notablemente, la Relatividad General es pura y simple geometría: el campo gravitatorio es, nada más y nada menos, que la geometría del espacio y del tiempo. En esta contribución se presentan las ideas básicas subyacentes a la geometría espacio-temporal, y sus consecuencias. con algunos ejemplos relevantes.

2. El principio de equivalencia

La piedra angular sobre la que se edificó toda la teoría de la Relatividad General es el *Principio de Equivalencia*, una idea genial pero sencillísima que asaltó a Einstein alrededor de 1907. Es un principio básico que *interpreta*—o sea, explica el sentido de— la observada igualdad de masas inercial y gravitatoria[1], o equivalentemente, la propiedad puesta de manifiesto por Galileo de que todos los cuerpos caen con la misma aceleración en un campo gravitatorio.

El principio se basa en la propiedad fundamental de los campos gravitatorios, ya comprobada por Galileo, de que todos los cuerpos en caída libre alcanzan la

[1] La masa gravitatoria de un cuerpo es la que mide la intensidad con que reacciona a la atracción de la gravedad. La masa inercial mide la inercia de un cuerpo, su resistencia a cambiar de movimiento, y es la que aparece en la segunda ley de Newton: $\vec{F} = m\vec{a}$, donde $\vec{a}$ es la aceleración del cuerpo respecto del espacio absoluto.

misma aceleración independientemente de su masa y constitución (en el vacío). Es evidente que el principio de equivalencia requiere la igualdad de las masas gravitatoria e inercial, porque de la segunda ley de Newton y la fórmula de la gravitación universal se obtiene

$$m\vec{a} = \frac{GMm}{r^2}\frac{\vec{r}}{r} \quad \Longrightarrow \quad \vec{a} = \frac{GM}{r^2}\frac{\vec{r}}{r}\,.$$

Como se ve, la cancelación de las masas (inercial y gravitatoria) a ambos lados de la ecuación conlleva que el campo vectorial de aceleración $\vec{a}$ es independiente de la masa m del cuerpo sometido a la fuerza de la gravedad creado por la masa M.

Este hecho es conocido desde mediados del siglo XVII, por ejemplo era una propiedad que Newton ya conocía. No obstante, a pesar de que grandes científicos avisaron de la relevancia y de la necesidad de explicar este hecho, nadie supo extraer la información ni las consecuencias que de él se derivan hasta la llegada de Einstein y su enunciado del principio de equivalencia. Por citar mi caso predilecto, el gran Hertz, en su obra *Über der Constitution der Materie* (1884), escribió: “ciertamente tenemos frente a nosotros dos propiedades totalmente fundamentales de la materia que deben considerarse como completamente independientes entre sí, pero en nuestra experiencia, y sólo experimentalmente, se nos aparecen como iguales. Esta correspondencia debe significar mucho más que ser un simple milagro (…) Debemos percatarnos claramente de que la correspondencia entre masa e inercia tiene que tener una explicación más profunda, y no puede despacharse como de algo de poca importancia …”. ¡Qué profunda intuición!

Einstein comprendió que esa misma propiedad es obviamente cierta también en los sistemas de referencia acelerados o no inerciales[2], o sea, en ellos todos los cuerpos adquieren la misma aceleración independientemente de su masa y constitución. Para comprenderlo en toda su extensión, Einstein mismo propuso un “Gedanken experiment” en el que se usa una caja aislada y lejana de todas las demás masas del Universo: el ascensor de Einstein. En ella podemos suponer que hay una científica exprimentada, digamos Eva, que conoce las leyes de la física. Si dotamos a esa caja de una aceleración constante (pongamos por caso que “superman” se dedica a transmitirle esta aceleración durante el rato que dure el experimento, y que esa aceleración vale g, la misma que en la superficie terrestre), Eva puede dedicarse a colgar pesas del techo —al haber una dirección de movimiento y aceleración,

[2]Un sistema de referencia inercial se define como aquel en el que los cuerpos sin agentes externos obrando sobre ellos están en reposo o en movimiento rectílineo y uniforme; la realización material, o incluso la existencia real, de tales sistemas ideales sólo se puede conseguir o asegurar de forma aproximada, pero su valor teórico es enorme y su papel fundamental en toda la física. Un sistema no inercial es cualquiera que esté acelerado con respecto a uno inercial, por ejemplo el de un tiovivo, o de un cohete en su lanzamiento, o el de un tren frenando, arrancando o dando una curva.

los conceptos de arriba y abajo quedan inmediatamente definidos—, a subirse en una báscula, etcétera. Dado que, como es evidente, todos los cuerpos adquieren la misma aceleración g transmitida por el superhéroe, y puesto que Eva conoce la propiedad fundamental de los campos de gravedad, ella puede llegar a la conclusión de que se encuentra inmersa en un campo gravitatorio homogéneo de intensidad g.

Así fue como llegó Einstein a, en sus propias palabras, "glücklichste Gedanke meines Lebens" ("la idea más lúcida de toda mi vida"; lo cual, tratándose de Albert Einstein, ¡es mucho!). Esa idea feliz es el reconocido Principio de equivalencia local entre sistemas de referencia acelerados y campos de gravedad:

> [Principio de Equivalencia] *Cualquier sistema de referencia no inercial es* localmente *equivalente a un sistema inercial con un cierto campo de gravitación.*

Dándole la vuelta al Principio, y notando que siempre se puede cambiar de sistema de referencia para pasar de uno no inercial a uno inercial, se obtiene la sorprendente versión

> [Principio de Equivalencia] *Todo campo gravitatorio puede hacerse desaparecer* localmente *escogiendo un sistema de referencia adecuado.*

Un ejemplo es poniéndose en caída libre. Obsérvese que si uno se tira por la ventana de un rascacielos, y hace abstracción de la resistencia del aire, se encontraría en un sistema libre de gravedad localmente, inercial: si soltamos un objeto cuidadosamente, tal como el reloj, quedaría en reposo con respecto a nosotros —hasta que nos estrellemos contra el suelo, eso sí—. En palabras del propio Einstein:

> El campo gravitatorio sólo tiene una existencia relativa (...) Porque, para un observador en caída libre (...) no hay campo gravitatorio —al menos en su entorno inmediato—.

El principio de equivalencia ha sido comprobado experimentalmente a día de hoy hasta la saciedad y con precisiones cada vez mayores.

3. La geometrización de la gravedad

El principio de equivalencia conduce inexorablemente a la *geometrización* de la gravedad: dado que todos los cuerpos se mueven igual independientemente de su masa, podemos imaginar que de hecho siguen trayectorias determinadas por "la forma" del espacio, que simplemente van por los caminos menos esforzados de un espacio que no es plano. Hay que observar que esta geometrización es imposible para el electromagnetismo, porque las partículas cargadas siguen trayectorias que dependen de la relación entre su carga y su masa. Diferentes cargas viajan por diferentes caminos.

Claro está que los cuerpos libres no siguen en general rectas cuando viven bajo la influencia de un campo gravitatorio, piénsese en un satélite en órbita terrestre, en la Luna, o en la propia Tierra en torno al Sol. Esto es análogo a lo que sucede, por poner un ejemplo, cuando se quiere ir óptimamente de un punto a otro de la superficie terrestre: para ello hay que seguir las geodésicas, que como se sabe son trozos de circunferencias de radio máximo. La geometría subyacente a un campo gravitatorio debe ser, por consiguiente, más general que la euclídea tradicional. Como es sabido, a principios del siglo XX ya se había descubierto la independencia del axioma de las paralelas, y se conocían las geometrías no euclídeas de Riemann en cualquier dimensión.

Repasemos brevemente las ideas de Riemann. El teorema de Pitágoras (tridimensional) local se puede expresar escribiendo

$$l^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 \tag{4.1}$$

donde $\{x, y, z\}$ son coordenadas cartesianas en $\mathbb{R}^3$, y l mide la distancia entre dos puntos con coordenadas (a, b, c) y $(a + \Delta x, b + \Delta y, c + \Delta z)$, respectivamente. Esto es elemental. Riemann propuso mantener la propiedad cuadrática de la relación anterior, pero sólo localmente (o infinitesimalmente) de manera que la distancia entre dos puntos puede depender de los propios puntos. Así, Riemann definió lo que se denomina el *elemento de línea o de longitud*

$$dl^2 = A(x, y, z)dx^2 + B(x, y, z)dy^2 + C(x, y, z)dz^2$$

o incluso con mayor generalidad, añadiendo términos cruzados pero cuadráticos

$$dl^2 = \sum_{i,j=1}^{3} g_{ij}(x)dx^i dx^j \tag{4.2}$$

donde las $g_{ij} = g_{ji}$ son funciones del punto y ahora, además, las coordenadas $\{x^i\}$ *no son necesariamente cartesianas*. La interpretación sigue manteniéndose: dl^2 mide la distancia infinitesimal entre dos puntos cualesquiera del espacio con coordenadas (x^i) y $(x^i + dx^i)$. Por ello, a este nivel, la única restricción es que la matriz (g_{ij}) sea, en cada punto, definida positiva.

Por otro lado, en 1907 Hermann Minkowski se percató de que las fórmulas de la Relatividad Especial de Einstein tenían una expresión más natural y simple en una generalización cuadridimensional del espacio euclídeo, pero con una "pequeña" sutileza: la cuarta coordenada era el tiempo, y tenía un carácter ligeramente diferente a las demás coordenadas. En todo caso, el espacio y el tiempo por separado dejaban de tener sentido absoluto y quedaban indisolublemente unidos para siempre. De la más famosa cita de Minkowski:

A partir de ahora, el espacio por sí solo, y el tiempo por sí mismo, se han desvanecido en meras sombras y solamente una especie de mezcla de los dos existe por derecho propio.

Lo que Minkowski propuso fue entender la Relatividad Especial y la electrodinámica usando un continuo espacio-temporal con coordenadas cartesianas $\{x, y, z\}$ para describir el espacio tridimensional y t para denotar el tiempo en un sistema de referencia inercial, y con un *elemento de línea* dado por

$$ds^2 = -c^2dt^2 + dx^2 + dy^2 + dz^2 \tag{4.3}$$

donde c es la constante universal que corresponde al valor de la velocidad de la luz en el vacío en un sistema de referencia inercial.

Hoy en día se suele denominar *intervalo* al ds^2, y da una medida de la separación *espacio-temporal* entre dos puntos —llamados eventos o sucesos— cercanos: ds^2 mide el *intervalo* infinitesimal entre cualesquiera dos sucesos con coordenadas (t, x, y, z) y $(t + dt, x + dx, y + dy, z + dz)$. Lo primero que se debe entender es el significado del signo menos delante de $d(ct)^2$. Por ejemplo, si mido el intervalo entre los puntos $(0, 0, 0, 0)$ y $(cdt, 0, 0, 0)$ se obtiene una cantidad negativa. Por ello, "el cuadrado" ds^2 no es necesariamente positivo en general. En otras palabras, debido al signo menos delante del primer término del miembro derecho de (4.3), ds^2 puede ser negativo o nulo.

Lo verdaderamente relevante aquí es la interpretación física. Si el intervalo entre dos sucesos es negativo, entonces se dice que la separación entre ellos es de tipo *temporal*, y la cantidad $+\sqrt{-ds^2}/c$ mide el tiempo propio transcurrido entre ambos sucesos en el sistema de referencia inercial dado (por ejemplo, entre Ud. mismo al leer *esto* y *esto otro*). El caso tradicional donde el ds^2 es positivo corresponde a una separación *espacial* y $\sqrt{ds^2}$ mide la distancia tradicional entre dos puntos en el sistema de referencia dado. Finalmente, queda el caso en el que el intervalo $ds^2 = 0$ se anula, que corresponde a puntos separados de manera *luminosa*. Pongamos por caso los eventos con componentes $(0, 0, 0, 0)$ y $(cdt, dx, 0, 0)$. Claramente en este caso $ds^2 = 0$. Notemos que uno puede considerar el suceso $(cdt, dx, 0, 0)$ como combinación de los sucesos temporal $(cdt, 0, 0, 0)$ y espacial $(0, dx, 0, 0)$, de manera que la distancia que separa los dos eventos originales (medida simultáneamente) es la longitud que recorre la luz en un tiempo propio dt. O sea, esos dos eventos pueden conectarse por un rayo de luz, de manera que ningún otro agente físico puede conectarlos (la velocidad de la luz es la máxima posible).

3.1. Variedades pseudo-riemannianas o lorentzianas

Las dos vertientes anteriores, referidas a las geometrías no euclídeas, y al espacio-temporal con intervalo no definido positivo, se pueden combinar dando lugar a las llamadas geometrías pseudo-riemannianas o lorentzianas [2,8]. Estas geometrías permiten un intervalo de tipo general

$$ds^2 = \sum_{\alpha,\beta=0}^{3} g_{\alpha\beta}(x) dx^\alpha dx^\beta \tag{4.4}$$

donde las diez funciones $g_{\alpha\beta} = g_{\beta\alpha}$ pueden depender de las cuatro coordenadas $\{x^\alpha\}$ con la única restricción de que la matriz $(g_{\alpha\beta})$ tenga en cada punto *signatura lorentziana*. Esto significa simplemente que se puede escoger, en cada punto, una base ortonormada (con un vector de longitud -1 y tres vectores de longitud +1, todos ellos mutuamente ortogonales) de forma que $(g_{\alpha\beta})$ =diag$(-1,1,1,1)$. O hablando en plata, que exista un único tiempo y tres dimensiones espaciales. Naturalmente, estas geometrías son, por un lado, no euclídeas, o sea, con curvatura, y por otro lado matienen la clasificación en intervalos espaciales, temporales y luminosos de acuerdo con el signo de ds^2.

Veamos ahora cómo puede deducirse en primera aproximación el marco matemático de la nueva teoría de la relatividad general basándonos exclusivamente en el principio de equivalencia. Supongamos que deseamos describir un sistema de referencia no inerical, como por ejemplo el que está asociado a un tiovivo que se mueve con velocidad angular constante ω. Dado que los caballitos giran con respecto a un sistema de referencia inercial tradicional cuyas coordenadas cartesianas son $\{t,x,y,z\}$, es suficiente realizar el siguiente cambio elemental de coordenadas

$$T = t, \qquad \rho = \sqrt{x^2+y^2}, \qquad \phi = \arctan\frac{y}{x} + \omega t, \qquad Z = z$$

válido en el dominio $c > \omega\rho$ para no sobrepasar la velocidad de la luz. Entonces, el intervalo (4.3) se transforma en

$$ds^2 = -(c^2 - \omega^2\rho^2)dT^2 - 2\omega\rho^2 dT d\phi + d\rho^2 + \rho^2 d\phi^2 + dZ^2. \tag{4.5}$$

Como se ve, el intervalo deja de ser diagonal y sus componentes no son constantes. Usando el principio de equivalencia, se sigue por lo tanto que este tipo de propiedades serán generales para intervalos que describan regiones con un campo gravitatorio. Obsérvese que (4.5) es un caso particular de los intervalos lorentzianos (4.4). En consecuencia, se postula que un campo gravitatorio vendrá determinado por un cierto intervalo con signatura lorentziana —o sea, hay un tiempo y tres dimensiones espaciales—. Así pues, el tiempo alcanza la misma categoría que el resto de coordenadas espaciales, pero con un *status* especial, el que le da el signo menos en la signatura del intervalo.

El cuadro matemático en el que se desenvuelve la Relatividad General es el anterior, con una geometría lorentziana como base. Un ejemplo sencillo de espacio-tiempo es

$$ds^2 = -c^2 dt^2 + bt\left(dx^2 + dy^2 + dz^2\right)$$

donde b es una constante. Como se intuye, la geometría es análoga a la tradicional, pero el "espacio" va creciendo con el tiempo t, las distancias espaciales van aumentando a medida que pasa el tiempo. Aunque parezca un ejemplo demasiado simple, de hecho esta geometría sencilla ¡describe las primeras épocas del Universo adecuadamente!

De hecho, del principio de equivalencia y la geometrización de la gravitación, que pasa a ser la curvatura del espacio-tiempo, se pueden deducir, directamente, algunos de los resultados más sorprendentes predichos por Einstein, tales como el cambio de la frecuencia de la luz al subir/bajar por la vertical (y en general el efecto Doppler gravitatorio), la desviación de la luz al pasar rasante al Sol, y en general la influencia del campo gravitatorio sobre la luz y sus trayectorias. Todo esto lo dedujo Einstein en los años 1907-13.

3.2. Distancias mínimas, tiempos máximos. Puntos focales

Cuestiones básicas en geometría euclídea, y también en sus generalizaciones riemannianas, son: ¿cuál es la distancia mínima entre dos puntos? ¿Cuáles son las trayectorias óptimas para ir de un punto a otro? Aquí óptimo significa recorriendo la mínima distancia posible. La respuesta es clásica y fácil de entender. Dado un arco continuo, diferenciable a trozos, de curva γ en forma paramétrica

$$\gamma \, : \, x^i = X^i(\lambda)$$

donde λ es el parámetro, se calcula la longitud L entre dos cualesquiera de sus puntos $p = \gamma(\lambda_0)$ y $q = \gamma(\lambda_1)$ integrando el elemento de línea dl (4.2) a lo largo de la curva:

$$L(p \xrightarrow{\gamma} q) \equiv \int_{\lambda_0}^{\lambda_1} \sqrt{\sum_{i,j=1}^{3} g_{ij}(X(\lambda)) \frac{dX^i}{d\lambda} \frac{dX^j}{d\lambda}} \; d\lambda \tag{4.6}$$

Se sobreentiende que la integral se computa sumando las integrales parciales correspondientes a los concatenados trozos diferenciables. Obviamente, en el caso riemanniano $L > 0$ y basta entonces con usar el cálculo variacional para obtener qué tipo arcos de curva minimizan, localmente, el funcional L. El resultado que se obtiene define las curvas llamadas *geodésicas*, nombre que originalmente denotaba las curvas de longitud mínima en la superficie de la esfera bidimensional —en este caso son las circunferencias de radio máximo, como por ejemplo los meridianos—.

En el caso de la Relatividad General, la gravedad se geometriza y se postula que las partículas libres siguen las trayectorias extremales naturales de la geometría lorentziana (4.4) que define el campo gravitatorio. Estas curvas son las que hacen extremo variacionalmente "el intervalo", de forma que la directa y natural generalización de (4.6) es

$$S(p \xrightarrow{\gamma} q) \equiv \int_{\lambda_0}^{\lambda_1} \sqrt{\left| \sum_{\alpha,\beta=0}^{3} g_{\alpha\beta}(X(\lambda)) \frac{dX^\alpha}{d\lambda} \frac{dX^\beta}{d\lambda} \right|} \, d\lambda \tag{4.7}$$

para arcos de curva γ en el espacio-tiempo. Los arcos que extremizan (4.7) se siguen llamando *geodésicas* y es elemental comprobar que satisfacen localmente

$$\frac{d^2 X^\alpha}{d\lambda^2} + \sum_{\mu,\nu=0}^{3} \Gamma^\alpha_{\mu\nu}(X(\lambda)) \frac{dX^\mu}{d\lambda} \frac{dX^\nu}{d\lambda} \propto \frac{dX^\alpha}{d\lambda} \tag{4.8}$$

donde

$$\Gamma^\alpha_{\mu\nu} = \frac{1}{2} \sum_{\rho=0}^{3} g^{\alpha\rho} \left(\frac{\partial g_{\nu\rho}}{\partial x^\mu} + \frac{\partial g_{\rho\mu}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\rho} \right)$$

son los símbolos de Christoffel de segunda especie, y $(g^{\mu\nu})$ es la matriz inversa de $(g_{\mu\nu})$. Escogiendo la parametrización adecuadamente siempre puede conseguirse que el segundo miembro de (4.8) se anule. A esta parametrización se le denomina afín y la retomaremos más tarde para definir la completitud de las geodésicas. Los arcos geodésicos se pueden por lo tanto calcular localmente, dado el intervalo, sin más que resolver las ecuaciones (4.8).

Físicamente, estas geodésicas son las trayectorias que describen los cuerpos de prueba, o sea, los no sometidos a influencia externa alguna aparte de la gravedad. No obstante, hay que resaltar varias diferencias con el caso euclídeo o propiamente riemanniano

- si la curva γ es tal que $ds^2|_\gamma < 0$, o sea, la curva es *temporal* en el sentido de que todos sus puntos tienen separación de intervalo temporal, entonces el valor de S calculado mediante (4.7) proporciona el *tiempo propio* a lo largo de esa curva entre los puntos inicial y final.

- Por otra parte, si $ds^2|_\gamma > 0$, de manera que γ es una curva *espacial* lo que se obtiene es la distancia espacial a lo largo de γ.

- en general, $S \geq 0$ y de hecho existen curvas extensas tales que $S = 0$, por ejemplo las curvas *luminosas* en las que $ds^2|_\gamma = 0$.

De este último punto se deduce que S no puede minimizarse curvas que unen puntos con separación temporal, ya que siempre se puede conseguir ir de uno a otro en zig-zag mediante curvas luminosas —y por lo tanto $S = 0$ a lo largo de este camino tortuoso—. En este caso dos puntos unibles por al menos una curva temporal, es elemental comprobar que la segunda variación de S a lo largo de una geodésica —o sea, de una curva extremal— es siempre no negativa y, por consiguiente, las geodésicas temporales *maximizan* el intervalo. Esto significa que hacen máximo, localmente, el tiempo propio entre esos dos puntos.[3]

Las geodésicas minimizan la distancia espacial y maximizan el tiempo transcurrido entre eventos espacio-temporales *localmente*, pero no necesariamente globalmente. Para entender esto en detalle recordemos las geodésicas clásicas, o sea, las circunferencias de radio máximo en una esfera. Pensando por ejemplo en la superficie de la Tierra, la distancia mínima del polo sur S al polo norte N es la longitud de un meridiano. Nótese, empero, que hay infinitos meridianos, por lo que obtenemos la primera enseñanza:

> *La longitud mínima se puede alcanzar con varias, incluso con infinitas, geodésicas.*

Considerénse ahora dos puntos situados sobre el Ecuador, digamos P y Q. Para ir de uno a otro geodésicamente, se puede seguir la circunferencia del ecuador, o también ir de P a S por el meridiano que los une, y luego de S a Q por el correspondiente meridiano que pasa por ellos. En ambos trozos la curva seguida es geodésica, al ser un meridiano, pero es evidente que la distancia recorrida yendo por la circunferencia del Ecuador —salvo que P y Q sean antípodas mutuas— es menor que la que se recorre yendo primero al polo sur. Esto es de hecho una propiedad general, la segunda lección:

> *Para que una curva minimize la distancia, ha de estar exenta de* esquinas*, o sea, de puntos donde la curva no sea diferenciable.*

Aún así, esto no caracteriza las curvas *globalmente* mínimas. Resta conocer lo que probablemente es lo más importante. Para ir de S a un punto cualquiera R óptimamente, hay que escoger el meridiano que pasa por R y S, pero éste contiene dos trozos conexos que unen S y R: el que pasa por el polo norte N, y el que no. Ambos arcos son geodésicos, y no tienen esquina alguna. Los dos son, en consecuencia, buenos candidatos a minimar la distancia. Resulta obvio comprender, no obstante, que el trozo de meridiano que pasa por N tiene mayor longitud. Pasa saber por qué, la cuestión básica reside en el concepto de *punto focal o conjugado*. Antes se hizo notar que, para ir de S a N, se puede seguir cualquier meridiano; eso implica que *todas* las geodéscias que emanan de S van a parar a N. Se dice entonces que N es un punto conjugado, o focal, de S.[4] De esta digresión

[3]Por cierto, una vez conocido este hecho la famosa “paradoja” de los gemelos —que no es tal, claro está— queda fatalmente desmontada. Obviamente, el gemelo que está en movimiento libre, sin aceleraciones, es el que envejece más, ya que sigue una geodésica: la curva que maximiza el tiempo.

[4]La definición de puntos focales y conjugados es un poco más sutil, pero valga aquí la idea intuitiva anterior. En general, se dice que un punto C es conjugado, o focal, de otro P si existen geodésicas distintas “infinitamente próximas” que pasan por P y C.

se deduce que hay las geodésicas sin puntos conjugados son las que realmente minimizan la distancia. En el caso concreto considerado, el trozo de meridiano que no pasa por el polo norte N es de menor longitud, al ser N punto focal de S. Esta regla es de nuevo general y obtenemos la tercera moraleja:

Las curvas que minimizan la distancia en geometría riemanniana son las geodésicas sin esquinas ni puntos focales o conjugados.

La discusión anterior se traslada, mutatis mutandis, al caso lorentziano sin más que tener en cuenta la diferencia entre geodéscias temporales o espaciales y se obtiene el siguiente teorema fundamental [4,10,11]:

Teorema 1. *Dados dos sucesos p y q con separación temporal (respectivamente espacial), una curva γ transcurre mediante el tiempo máximo (resp. recorre la distancia mínima) de p a q si y sólo si γ es una geodésica temporal (resp. espacial), sin esquinas y sin puntos focales o conjugados entre p y q.*

3.3. Superfices mínimas y superficies atrapadas

Supongamos que en vez de querer saber cuáles son las trayectorias más cortas entre dos puntos del espacio, lo que deseamos es conocer las superficies de área mínima que unen dos curvas dadas. Por ejemplo, se pueden fijar dos circunferencias concretas del mismo radio centradas en el eje z a diferentes alturas en $\mathbb{R}^3$ y considerar todas las superficies que las unen (una obvia es el cilindro del mismo radio centrado en el mismo eje, por ejemplo). De entre todas ellas queremos discernir la (o las) que tenga(n) el área mínima. Este problema tradicional de las matemáticas se resuelve, como no podía ser de otra manera, minimizando el funcional de área. éste se obtiene directamente del elemento de longitud (4.2), para cualquier geometría riemanniana dada, de manera totalmente análoga a la que se usó para definir el funcional de longitud (4.6). Dada una superficie diferenciable Σ, se calcula el área A entre dos curvas γ_1, γ_2 cualesquiera contenidas en S integrando el elemento de línea dl (4.2) a lo largo de la superficie entre γ_1 y γ_2:

$$A(\gamma_1 \xrightarrow{S} \gamma_2) \equiv \int_{\gamma_1}^{\gamma_2} dl|_S \tag{4.9}$$

de manera que $A > 0$ y se puede miniminizar como es habitual:

$$\frac{\delta A}{\delta S}(\gamma_1 \xrightarrow{S} \gamma_2) = 0\,.$$

El resultado resolver esta variación nula son las superficies denominadas *mínimas* —o a veces, minimales—. Por ejemplo, para el caso mencionado anteriormente de dos circunferencias centradas en el eje z, y en la geometría euclídea clásica con longitud (4.1), Meusnier obtuvo en 1776 su famosa *catenoide*.

Una caracterización local de las superficies mínimas, de manera similar a la de las geodésicas (4.8), es que su *curvatura media* se anule. En geometría euclídea la curvatura media es la media de las curvaturas principales, y por lo tanto es un escalar. Resulta paradójico, a primera vista, que ahora se obtenga un escalar en contraposición al caso de las curvas geodésicas, cuya caracterización está dada por el vector (4.8). La paradoja se desmantela si se hace el mismo cálculo en dimensiones arbitrarias. La respuesta a la hora de minimizar el área, volumen, longitud, hipervolumen, etcétera, en dimensiones cualesquiera resulta ser *siempre* un vector, ortogonal a la subvariedad que se desea optimizar (la curva, la superficie, etcétera), y por lo tanto con tantas componentes independientes como el valor de la codimensión: la dimensión de la variedad menos la de la subvariedad. En el caso de superficies en $\mathbb{R}^3$ este vector de curvatura media sólo tiene una componente independiente, en la dirección normal a la superficie, y esa componente es el escalar tradicionalente llamado curvatura media. Pero en el caso ya visto de las geodésicas, lo que se obtiene es el vector de curvatura de la curva (o en física la aceleración), que es lo que se anula en la ecuación (4.8) sin segundo miembro. Por lo tanto, en dimensiones mayores, por ejemplo en $\mathbb{R}^4$, las superficies mínimas están carectirazadas localmente por la anulación del vector de curvatura media $\vec{H}$.

¿Qué ocurre si, además de tener 4 dimensiones, la geometría es de tipo lorentziano (4.4)? Nada fuera de lo común, las superficies mínimas siguen caracterizándose por tener vector de curvatura media nulo: $\vec{H} = \vec{0}$. Hay que hacer notar, no obstante, lo siguiente: en el caso de una geometría riemanniana, con elemento de longitud definido positivo, todos los vectores tiene módulo positivo — o nulo, lo que ocurre sólo cuando el propio vector se anula—. Consecuentemente, en el caso de la geometría riemanniana en cualquier dimensión, las superficies (y todas las subvariedades de cualquier dimensión) se pueden dividir en dos clases y sólo dos: las que tienen un vector de curvatura media nulo, y las que no. En otra palabras, las minimales y el resto, siendo aquellas las únicas con una característica destacada en lo que concierne al vector curvatura media. Por el contrario, en geometría lorentziana hay cuatro tipos de vectores, atediendo al signo de su módulo, que puede ser positivo (vector espacial), negativo (vector temporal) o nulo (en cuyo caso el vector es luminoso si no se anula, o puede ser el propio vector cero). Por consiguiente, las superficies también se pueden clasificar en cuatro tipos: las mínimas con $\vec{H} = \vec{0}$, las “normales” con $\vec{H}$ espacial, y dos tipos nuevos correspondientes a los casos con $\vec{H}$ luminoso o temporal. Pues bien

Definición. *Una superficie en un espacio-tiempo lorentziano cuadridimensional se dice que es* atrapada *si su vector de curvatura media $\vec{H}$ es temporal en todas partes, y se dice que es* marginalmente atrapada *si $\vec{H}$ es luminoso por doquier.*

Usualmente, sólo se consideran superficies cerradas espaciales, o sea, tales que son compactas y sin borde (cerradas), como por ejemplo una esfera, y con al menos una normal de tipo temporal (espaciales, viven en un “instante de tiempo”). El concepto de *superficie cerrada atrapada* es bello, intuitivo y muy poderoso, a la par que produjo un punto de no retorno en la ciencia relativista, a la que cambió de manera irreversible. Fue introducido por el matemático Roger Penrose en [9], quien demostró una extraordinaria inspiración.

Como acabamos de ver, la idea de superficie cerrada atrapada es estrictamente geométrica, lo que no obsta para poseer también un alto contendio físico. Físicamentre se suele introducir el concepto de atrapamiento de una superficie considerando flashes de luz emitidos desde ella hacia dentro o fuera, e imaginando las nuevas, imaginarias, superficies que los rayos de luz alcanzan en las direcciones interna y externa un instante fijo posterior. Se comparan entonces las áreas de la superficie inicial con las de las dos superficies imaginarias posteriores. Parece intuitivamente claro que estas superficies tendrán, respectivamente, menor y mayor área que la inicial (al estar una dentro y la otra fuera de la inicial). Curiosamente, esto sólo es así en situaciones *estacionarias*, independientes del tiempo. En situaciones dinámicas, variables con el tiempo —tales como las que se obtienen en las geometrías asociadas a un universo en expansión, o a una estrella en colapso—, puede suceder que las dos tengan mayor, o las dos menor, área que la inicial. éstas son las superficies atrapadas, en un caso hacia el futuro, en otro al pasado. Hay que entender que aquí queremos decir “la inicial *en el instante inicial*”, por supuesto, de manera que se está comparando el área de la superficie inicial *en el instante inicial* con el área de las superficies imaginarias finales *en un instante posterior*.

En fin, si existe una superficie cerrada atrapada en el espacio-tiempo, y dadas sus propiedades, se puede constatar que las geodéscias que emanan de ella tienen una longitud del intervalo máxima *acotada* por un valor relacionado con el inverso del módulo de $\vec{H}$. Esto será de la máxima importancia para la demostración de los teoremas de singularidades que se tratarán en breve. En particular, se da por hecho que la existencia de esferas atrapadas al futuro son una señal inconfundible de la existencia de un agujero negro, y tambiés se sabe que los modelos corrientes del Universo en expansión contienen esferas atrapadas al pasado.

4. Curvatura y las ecuaciones de Einstein

La existencia de superficies cerradas atrapadas será por tanto determinante en casos de expansión o contracción, o sea, e algunas situaciones extremas pero reales, tales como las acaecidas al principio de la etapa expansiva de nuestro Universo (o sea, en el “big-bang”) o las que pueden acontecer en el interior de estrellas con mucha masa que gastan todo su combustible de fusión nuclear y por ello implosionan catastróficamente originando, en ciertas ocasiones, agujeros negros. ¿Cómo discernimos si el espacio está en contracción, o en expansión? Es fácil de entender que para ello hace falta saber cómo es la geometría que describe el espacio-tiempo en cada caso, o sea, el intervalo. Pero, ¿cómo se sabe cuál es el intervalo ds^2 de (4.4) para un campo gravitatorio dado?

Para eso están las ecuaciones de campo de Einstein, que relacionan la curvatura del espacio-tiempo con su contenido material. Por ello, hay que hacer algunas puntualizaciones y una pequeña digresión acerca del concepto de curvatura. Recordemos, para empezar, el uso de la palabra *local* en los enunciados del principio de equivalencia. Eso está relacionado con el hecho, ya discutido, de que si se quiere ir del polo sur al polo norte de una esfera con desgaste mínimo, se puede recorrer cualquiera de sus meridianos. Por tanto, dos personas viajando simultáneamente por dos meridianos diferentes, y por tanto en esfuerzo mínimo, verán como su distancia mutua varía, creciendo inicialmente y disminuyendo después de cruzar el Ecuador, hasta que se topen de nuevo en el polo norte. Esto ocurre a pesar de que ambas siguen curvas geodésicas, y no es nada más, ni nada menos, que la manifestación de que curvas paralelas en la esfera acaban cruzándose. La *aceleración relativa* que sufren los viajeros que se mueven *geodésicamente* es una consecuencia directa de la *curvatura* de la superficie de la esfera. De manera análoga, si dos cuerpos se dejan caer libremente desde una misma altura en el campo gravitatorio terrestre, es evidente que a medida que pasa el tiempo los dos cuerpos se acercan, dado que sus trayectorias apuntan hacia el centro de la Tierra. O sea, sufren una *aceleración relativa.*

Por otro lado, que en un sistema de referencia no inercial, sin gravitación, dos partículas de prueba libres jamás sufren tal aceleración relativa. No hay por tanto curvatura. En resumidas cuentas, hay una manera de distinguir entre los campos gravitatorios *reales* y los sistemas de referencia no inerciales, *siempre y cuando se usen dos o más masas de prueba para sondear los efectos no locales, y estos se midan en regiones suficientemente extensas.*

El objeto matemático que mide la curvatura es el campo tensorial de Riemann, $R_{\alpha\beta\mu\nu}$. Si dicho tensor se anula idénticamente en un dominio, la geometría es plana en ese dominio. En caso contrario, la curvatura, que pasa a ser el propio campo de gravitación, tiene efectos físicos relevantes. El tensor de Riemann depende de los $g_{\alpha\beta}$ de (4.4) y de sus derivadas primeas y segundas en forma no lineal.

La cuestión de cómo determinar en cada caso concreto el intervalo, o lo que es lo mismo, los $g_{\alpha\beta}$ de (4.4), y a fortiori poder calcular la curvatura del espacio-tiempo, le costó a Einstein al menos 7 años de su vida y, de hecho, propuso diversas ecuaciones para la nueva teoría. En un realmente ajetreado mes de Noviembre de 1915, en el que semanalmente presentó diversas propuestas para las ecuaciones, finalmente se convenció de cuáles eran las correctas y las presentó un día 25. La cuestión era, repitamos: ¿Cómo cuantificar la curvatura del espacio-tiempo, la deformación producida por una distribución de masas? ¿Cómo determinar los $g_{\alpha\beta}$ en casos concretos de interés físico?

Dado que la teoría era relativista, las ecuaciones debían contener la constante fundamental c, y como se desea describir la gravitación, también debe aparecer G, la constante de la gravitación universal de Newton. Por otro lado, la relatividad especial había relacionado la masa con la energía, y había unificado a ésta con la cantidad de movimiento en un 4-vector momento. En consecuencia, no sólo la densidad de masa sino todas las densidades de energía y momento deben ser fuentes del campo gravitatorio y han de aparecer a la derecha de las ecuaciones. ¿Qué poner a la izquierda? Para descifrar este misterio Einstein usó como inspiración la aproximación newtoniana de su teoría para un campo gravitatorio aislado y débil, que él ya había deducido basándose en su principio de equivalencia. El intervalo en este caso se puede escribir [7,12]

$$ds^2 = -(c^2 - 2\Phi)\, dt^2 + dx^2 + dy^2 + dz^2 \tag{4.10}$$

donde $\Phi(t, x, y, z)$ es el clásico potencial newtoniano. Se intuye que las funciones g_{ab} serán, en general, lo que generalice al potencial Φ.

Se sigue de ello que las derivadas primeras (los símbolos de Christoffel) son análogos a las fuerzas gravitatorias newtoniannas, lo que concuerda con la expresión en coordenadas de la ecuación de las geodésicas (4.8), y con la interpretación de éstas como trayectorias de las partículas sometidas a la acción de la gravedad. Teniendo entonces en cuenta que el potencial Φ se determina en la teoría newtoniana mediante la ecuación de Poisson $\Delta\Phi = 4\pi G\varrho$, donde ϱ es la densidad de materia, y que esta ecuación es lineal y contiene las derivadas segundas de Φ, Einstein supuso que las buscadas ecuaciones deben tener en el miembro izquierdo *algo* que contenga las derivadas segundas de las $g_{\alpha\beta}$, a ser posible que sea lineal en ellas. Claro está, este objeto tiene que estar matemáticamente bien definido — o sea, tener carácter tensorial—.

Una posibilidad sería poner un campo vectorial proveniente de las derivadas segundas de las $g_{\alpha\beta}$ para igualarlo a la densidad de 4-momento de la distribución material y energética. Sin embargo, es imposible construir un campo vectorial de las citadas características. Se sabe de hecho que el único objeto tensorial que se puede fabricar con las derivadas segundas de g, lineal en éstas, es ni más ni menos que *el tensor de Riemann.* Desafortunadamente, el tensor de curvatura es demasiado complicado, y tiene un excesivo número de componentes independientes (20 en dimensión 4). A lo mejor que se podía aspirar, por lo tanto, era a usar una de sus trazas. La traza natural es el conocido tensor de Ricci

$$R_{\mu\nu} \equiv \sum_{\rho=0}^{3} R^{\rho}{}_{\mu\rho\nu} = \sum_{\rho,\sigma=0}^{3} g^{\rho\sigma} R_{\rho\mu\sigma\nu}\,.$$

El tensor de Ricci $R_{\mu\nu}$ es un campo tensorial 2-covariante simétrico. A primera vista se puede pensar en igualar este tensor de Ricci al tensor energía-momento de la materia $T_{\mu\nu}$, que es del mismo tipo tensorial y ya había aparecido en electromagnetismo y en dinámica de fluidos. No obstante, esa posibilidad tenía problemas irresolubles, uno de ellos es que no se deduce el límite newtoniano correcto. Otro grave problema es que estas ecuaciones no incorporan las necesarias relaciones de conservación de la energía y el momento de la materia.

Las ecuaciones del campo gravitatorio que Einstein finalmente propuso siguen hoy plenamente vigentes y se escriben

$$R_{\mu\nu} - g_{\mu\nu}\left(\frac{1}{2}\sum_{\rho,\sigma=0}^{3} g^{\rho\sigma}R_{\rho\sigma} - \Lambda\right) = \frac{8\pi G}{c^4}T_{\mu\nu} \tag{4.11}$$

donde Λ es una constante, hoy en día famosísima, denominada la *constante cosmológica*, protagonista de incesantes controversias que quedan fuera del alcance de esta reseña. De hecho, el miembro izquierdo de (4.11) es la *única* combinación tensorial posible lineal en las segundas derivadas de las $g_{\alpha\beta}$ que satisface las propiedades físicas necesarias. Desde este punto de vista —y dejando aparte el valor concreto de Λ—las ecuaciones de Einstein (4.11) son las únicas posibles.

Como se aprecia, las ecuaciones de Einstein relacionan las propiedades métricas del espacio-tiempo junto con su deformación en términos de la curvatura con la cantidad y distribución energético-material que crea esa deformación. La fórmula (4.11) representa un sistema de diez ecuaciones diferenciales en derivadas parciales no lineales y acopladas de segundo orden, por lo que son matemáticamente muy complicadas. Desde un punto de vista físico no se conocía nada igual, ya que las fuentes crean un campo gravitatorio que es una deformación de la geometría subyacente, pero la propia materia se tiene que mover según dicta dicha geometría, y este movimiento afecta a la geometría de nuevo que a su vez modifica la dinámica de las fuentes…, y así sucesivamente en un baile sin fin.

5. Una aplicación teórica: los teoremas de singularidades

Una de las cimas de las matemáticas aplicadas se consiguió probablemente con los teoremas de singularidades desarrollados por R. Penrose [9], S.W. Hawking [4,5] y otros en la década de los 60 del siglo pasado, dentro de la teoría general de la relatividad. Se pueden consultar compendios originales excelentes en [4] y [10], y más recientes con diversos desarrollos posteriores en [11,12]. Los teoremas de singularidades combinan geometría diferencial —en su vertiente lorentziana—, técnicas de topología diferencial, métodos variacionales, una novedosa teoría de la causalidad (que se puede considerar de la conectividad), geometría de subvariedades y algunos otros ingredientes. Son, por tanto, teoremas interdisciplinares en matemáticas. Además, tienen aplicación directa al mundo real, ya que —siquiera parcialmente— creemos que dan indicaciones de propiedades de objetos reales tales como el Universo entero o los estados finales de estrellas con mucha masa. Son, por todo ello, uno de los mayores logros de la física matemática teórica.

Las conclusiones de los teoremas de singularidades (la existencia de al menos una singularidad en espacio-tiempos generales) proporcionan pruebas insoslayables del comienzo *clásicamente* singular del Universo, y del destino fatal *clásico* de algunas estrellas. Hasta qué punto esto es cierto o no se puede debatir [11], pero no cabe duda de que en muchas situaciones, y bajo ciertas circunstancias, eso es cierto. brevemente discutiremos en lo que sigue esas situaciones y la razonabilidad de las circunstancias necesarias, desentrañando, de paso, los entresijos geométricos sobre los que se sustentan los teoremas. Naturalmente, hay magníficos libros dedicados al tema. Para los lectores interesados en la parte más física de la Relatividad General es recomendable el libro clásico [7], mientras que los interesados en la vertiente matemática pueden consultar [1,6,8]. Para los indecisos, quizás [12] sea recomendable.

5.1. ¿Qué es una singularidad espacio-temporal?

Pero, ¿de qué hablan los teoremas de singularidades mecionados? ¿Qué es lo que predicen?

Las nociones intuitivas de las singularidades, tales como concentración infinita de materia o energía, no son estrictamente hablando las singularidades que se deducen de los teoremas –aunque puede ocurrir que, ocasionalmente,lo sean–. Como se aprecia en (4.11), las ecuaciones de Einstein relacionan propiedades geométricas del espacio-tiempo, y su deformación en términos de la curvatura, con la cantidad y distribución energético-material que crea esa deformación. Por ello, la propia geometría pasa a ser una variable dinámica, lo que conduce a una de las muchas dificultades prácticas de la Relatividad General: a diferencia de otras teorías físicas, ésta eleva el espacio y el tiempo a la categoría de entes dinámicos, de manera que no son meramente el escenario en el que se desarrolla la obra, sino que forman parte del argumento, incluso con papeles protagonistas. Las singularidades son hechos catastróficos, destruyen en particular la arena espacio-temporal, la cual es esencial para simplemente *hablar de las cosas*. Así que la situación que se afronta es la de no poder hablar de algo que se quiere definir.... Ardua tarea. Para entender las dificultades que afloran se puede consultar [3], un entretenido artículo donde se desmenuzan estas cuestiones de manera pedagógica.

De esta manera, si por ejemplo nos topamos con una evolución catastrófica —digamos de una estrella, de manera que colapsa y su densidad de masa parece concentrarse sin medida, por lo que $T_{\mu\nu}$ diverge—, resulta que *también* $R_{\mu\nu}$ diverge debido a las ecuaciones de Einstein (4.11). Pero $R_{\mu\nu}$ es parte de la curvatura, es una propiedad puramente geométrica del espacio-tiempo. Por lo tanto, la propia geometría, el mismísimo espacio-tiempo, queda destruído. Estas catástrofes son, claramente, mucho peores que las previamente conocidas, en las que muy bien pudiera ocurrir que algo divergiese, pero siempre dejando intacta la arena básica: el espacio y el tiempo. En la relaticidad general, en contraste, el sustrato de cualquier proceso físico, el espacio y el tiempo, queda irremisiblemente roto. Notemos que en essas situaciones catastróficas no tienen sentido frases como "tal o cual cosa se hace infinito *en el punto x*, o *en el instante t* ", porque esos supuestos puntos, esos eventos, no forman parte del espacio-tiempo, que ha desaparecido precisamente allí — si se me permite el abuso—. No se puede hablar de "cuando" ocurrirá la catástrofe, ni de "donde". En rsumidas cuentas, *las singularidades no forman parte del espacio-tiempo*.[5]

La estrategia que ha conseguido cierto consenso en la comunidad relativista para solventar estas dificultades consiste en tratar de *señalar* las singularidades mediante objetos que, ellos sí, estén bien definidos en el espacio-tiempo. Una manera relativamente sencilla de hacerlo es imaginando qué ocurre en las situaciones que queremos describir, tales como por ejemplo el colapso catastrófico de una estrella. ¿Qué le ocurriría a una sonda que mandásemos a explorar lo sucedido? Pues simplemente que el monstruo se la tragaría al cabo de un cierto tiempo *finito*. Igualmente, si el Universo comenzó en una gran explosión catastrófica, resulta que las cosas, todas, aparecieron de pronto hace un cierto tiempo *finito*.

Por todo ello parece lógico tratar de señalizar las singularidades, o de diagnosticarlas, siempre que haya partículas hipotéticas que vayan hacia, o vengan desde, ellas de forma que en el primer caso desaparezcan o en el segundo surjan de la nada. Y esta es la idea principal: las singularidades existen si aparecen *repentinamente*, o desaparecen súbitamente, hipotéticas partículas físicas. Estas partículas —una sonda, un cohete, una pelota, o un rayo de luz, por ejemplo— alcanzan, o provienen de, el "filo" del espacio-tiempo, de su "margen o frontera", de algo que no pertenece al espacio-tiempo pero al que se puede acceder desde él. La manera matemática de definir esto es mediante curvas de longitud, o tiempo propio, finito. En concreto, concentrándonos en las curvas geodésicas:

> Se dice que una curva que sale desde el punto x es *completa* si existe, o se puede extender diferenciablemente, a todos los valores positivos de su parámetro afín.

Por ejemplo, la curva que va desde el origen hasta el evento $(1,0,0,0)$ en el espacio-tiempo de Minkowski (4.3) tiene una duración finita, pero se puede alargar sin más

[5] Además de lo dicho, hay problemas de otro tipo. Ni siquiera permitiendo una definición de divergencias/infinitos del tensor de curvatura la cosa se arregla, porque esas caracterizaciones pueden depender de la base escogida. Y si esto se evita, calculando por ejemplo los invariantres escalares de la curvatura, resulta que dejamos fuera muchos casos singulares, porque existen por ejemplo geometrías con todos estos escalares nulos, pero con curvatura no nula (e incluso singular). Esto es reminiscente de la existencia de curvas extensas con $L = 0$, o de vectores luminosos —que no son cero pero tienen norma nula—. Cuestiones genuinas de geometría lorentziana.

que extenderla a $(t, 0, 0, 0)$ para todos los valores de t. Esto define una geodéscia completa. Recordemos que el parámetro afín es el que permite anular el término de la derecha en (4.8).

Según todo lo anterior:

> Si existen geodésicas temporales incompletas entonces el espacio-tiempo contiene singularidades.

5.2. Un teorema patrón

Estamos en disposición de enunciar lo que podemos denominar teorema patrón, en el que se pone de manifiesto el esqueleto de un teorema de singularidades típico [11]. Prácticamente todos los teoremas concretos de singularidades concuerdan con este patrón, únicamente escogiendo entre diversas posibilidades para sus diferentes hipótesis.

Teorema 2 (Teorema patrón de singularidades). *Si un espacio-tiempo es suficientemente diferenciable y satisface*

1. *una cierta condición en la curvatura*
2. *una condición adecuada de causalidad*
3. *una condición inicial o de contorno apropiada*

entonces necesariamente contiene geodésicas temporales o luminosas incompletas, ergo singularidades.

La cuestión interesante es saber y explicar cuales son los efectos, y las consecuencias, de cada una de las hipótesis anteriores. En otras palabras, para qué se necesitan y qué implican, por separado y conjuntamente. Esto nos va a permitir entender como surge la incompletitud de algunas geodésicas, es decir, las singularidades. La sorprendente es que, aparte sutilezas técnicas, se puede comprender todo eso perfectamente con geometría elemental y lo que ya hemos dicho.

5.2.1. La hipótesis sobre la curvatura

El cometido de esta suposición es simplemente *asegurar la focalización de las geodésicas.*

Para entender esto, volvamos al ejemplo de la esfera terrestre. Como se recordará, se puede viajar desde el polo sur S al polo norte N, por el camino más corto, de muchas maneras: por todos los meridianos. Así que hay muchos "caminos más cortos". En particular decíamos que el polo norte N es un punto focal de S, ya que todas las geodésicas que salen de S se cruzan en N. Esto no ocurre en el plano $\mathbb{R}^2$, ya que si emitimos desde cualquier punto de $\mathbb{R}^2$ todas las rectas posibles, éstas nunca se encuentran de nuevo. No hay puntos focales. Intuitivamente se comprende que la razón por la que en la esfera aparecen puntos focales tales como N y S es la *curvatura* de la esfera. En definitiva, los puntos focales aparecen cuando existe (al menos un tipo de) curvatura.

Dicha existencia de curvatura se puede medir como se explicó al introducir la curvatura, que es observable mediante la *aceleración relativa* que sufren cuerpos que se mueven *geodésicamente*. Las hipótesis concretas que se usan en los teoremas propiamente dichos, sin tener aplicación universal, al menos son realistas en casos relevantes. Por lo tanto, supuesta esta condición, tenemos asegurada la existencia de puntos focales para cualquier conjunto de geodésicas emanando de cualquier punto.

5.2.2. La condición de causalidad

Pasemos pues a la segunda condición en el teorema patrón de singularidades, la condición de causalidad. Esta suposición se necesita por dos razones fundamentales: primero, para hacer obligatorio el paso del tiempo, o sea, para hacer inevitable la llegada del futuro; y segundo, para asegurar la existencia de geodésicas temporales *máximas* entre cualesquiera dos sucesos con separación temporal—o sea, para asegurar la existencia de la curva donde se alcanza el mayor tiempo propio—.

Por lo que se refiere al primero de estos objetivos, quizás pueda resultar extraño tener que asegurar semejante obviedad: todos sabemos del irremediable paso del tiempo. Desafortunadamente, el paso inexorable del tiempo no es una propiedad de todos los espacio-tiempos de la Relatividad General, como puso de manifiesto el gran matemático Kurt Gödel. Así se suele suponer esta hipótesis, harto razonable, "a mano".

En lo que respecta al segundo objetivo, la existencia de geodésicas máximas, resuta que la condición de causalidad más fuerte conocida —llamada *hiperbolicidad global*—, se define exigiendo que dados dos eventos x, y cualesquiera del espacio-tiempo, el cierre del conjunto de los puntos que están en todas las curvas temporales que van de x a y sea *compacto*. Claro está, esta caracterización implica que dados dos puntos cualesquiera siempre existe la curva que maximiza el tiempo propio entre esos dos sucesos. Pero de lo que ya sabemos, esta curva tiene que ser necesariamente una geodésica, sin esquinas, y sin *puntos focales*.

Recapitulando: si se combinan las condiciones de curvatura y de causalidad, 1 y 2, en el teorema patrón de singularidades, se sigue que

- de la hipótesis sobre la curvatura se deduce la existencia de puntos focales a lo largo de cualquier geodésica temporal inextensible;

- de la suposicón de causalidad obtenemos la existencia de curvas máximas, ergo geodésicas sin puntos focales, entre cualesquiera dos sucesos, por alejados que estén.

Se perfila una contradicción, que todavía hay que rematar. El problema que resta por sortear es que no se tiene, a este nivel, una cota máxima para la L (el tiempo propio) de las geodésicas temporales. Para entender el porqué, escojamos una geodésica temporal que parte de x y, dadas las hipótesis anteriores, alcanza un cierto punto focal y al cabo de un cierto tiempo propio; se sigue entonces por lo anterior que, a partir de y debe existir *otra* geodésica que sale de x sin puntos conjugados hasta y y más allá, pero claro, tiene que existir otro y', posterior a y, que es el focal de x a lo largo de la nueva curva, ya que todas las geodésicas tienen puntos focales. De nuevo esto significa que hay una *nueva* geodéscia temporal sin puntos conjugados desde x hasta y', y vuelta a empezar. No obstante, siempre que L a lo largo de las geodésicas no esté acotada superiormente, este razonamiento se puede seguir indefinidamente sin contradicción.

5.2.3. La condición inicial o de contorno

Y es precisamente para asegurar esta contradicción —o sea, la existencia de ciertas cotas máximas para L a lo largo de conjuntos adecuados de geodésicas temporales— por lo que se ha de incluir la tercera condición en el teorema patrón. La condición inicial o de contorno se revela así como la auténtica puntilla, la hipótesis irrenunciable. Esta condición puede adoptar muy diversas formas, cada una adaptada a un caso concreto de interés físico. El ejemplo de mayor relevancia es la existencia de una *superfice cerrada atrapada*. Como se mencionó en su momento, se sabe que las geodéscias que emanan de una tal superficie tienen una L máxima acotada por un valor relacionado con el inverso del mdulo de su vector de curvatura media, que no es cero en ningún punto. Por lo tanto, se cierra el círculo y se llega a la contradicción buscada: no todas las geodésicas pueden ser completas, porque entonces tendrían puntos focales, pero esto impediría que algunas de ellas fueran máximas como requiere la condición de causalidad.

Agradecimientos: Agradezco a los organizadores del Paseo por la Geometría en la UPV/EHU la amable invitación que me cursaron, que me ha permitido tratar de acercar a los estudiantes de matemáticas algunas ideas que espero sean de su interés.

Bibliografía

[1] J. K. Beem, P. E. Ehrlich, K. L. Easley, *Global Lorentzian Geometry*, Pure and Applied Math. 202, 1996.

[2] L. P. Eisenhart, *Riemannian Geometry*, Princeton University Press, 1966.

[3] R. Geroch, *What is a singularity in General Relativity?*, Ann. Phys. (NY) 48, 526-540, 1968.
[4] S. W. Hawking, G. F. R. Ellis, *The large scale structure of spacetime,* Cambridge University Press, 1973.
[5] S. W. Hawking, R. Penrose, *The singularities of gravitational collapse and cosmology*, Proc. Roy. Soc. London A, 314, 529-548, 1970.
[6] M. Kriele, *Spacetime*, Springer-Verlag, 1999.
[7] C. W. Misner, K. S. Thorne, J. A. Wheeler, *Gravitation*, W H Freeman and Company, 1973.
[8] B. O'Neill, *Semi-Riemannian geometry*, Academic Press, 1983.
[9] R. Penrose, *Gravitational Collapse and Space-Time Singularites*, Phys. Rev. Letters 14, 57-59, 1965.
[10] R. Penrose, *Techniques of Differential Topology in Relativity*, Regional Conference Series in Applied Math. 7, SIAM, 1972.
[11] J. M. M. Senovilla, *Singularity Theorems and Their Consequences*, Gen. Rel. Grav. 30, 701-848, 1998.
[12] R. M. Wald, *General Relativity*, The University of Chicago Press, 1984.

José María Martín Senovilla
Universidad del País Vasco
Facultad de Ciencia y Tecnología
Departamento de Física Teórica e Historia de la Ciencia
Barrio Sarriena s/n. 48940 (Leioa)
e-mail: ***josemm.senovilla@ehu.es***