



Universidad de Sevilla

FACULTAD DE MATEMÁTICAS
DPTO. DE GEOMETRÍA Y TOPOLOGÍA

CAUSALIDAD EN VARIETADES DE
LORENTZ
TEOREMAS DE SINGULARIDAD

Manuel Francisco Ruiz-Herrera Bernal

Doble Grado en Física y Matemáticas

Tutorizado y dirigido por

Dr. Alfonso Carriazo Rubio

2020 - 2021

Resumen

En este trabajo se realizará un primer acercamiento a la teoría de Causalidad en variedades de Lorentz enfocada al estudio de los teoremas de singularidad en un marco físico-matemático. Se comenzará introduciendo el concepto de variedad diferenciable y, a partir de este, se irán generando nuevas estructuras cada vez más complejas con las ideas que se desarrollarán a lo largo de los capítulos. En primer lugar, las variedades diferenciables se equiparán con un tensor métrico que permitirá definir el concepto de variedad semi-Riemanniana generándose un modo general de medir longitudes, ángulos, curvaturas, etc. Posteriormente, se destacará el estudio de un tipo particular de estas variedades: las variedades de Lorentz; se dotará a estas de una orientación temporal construyéndose la idea general de espacio-tiempo. En este punto, se estudiará brevemente la Teoría General de la Relatividad y algunos de sus modelos más simples sirviendo de nexo natural entre las ideas matemáticas previas — sustentándose en ellas — y las posteriores, introduciendo el concepto de singularidad y motivando el estudio de la teoría de Causalidad. En este sentido, el último capítulo estará enfocado en los conceptos más básicos de la teoría de la Causalidad — relaciones causales, hipersuperficies de Cauchy, etc. — generándose un formalismo que culminará con la construcción de los teoremas de singularidad que permitirán abordar de un modo más general la idea de singularidad en cualquier espacio-tiempo bajo determinadas condiciones.

Abstract

This project consists in a first approach to the theory of Causality in Lorentz manifolds aimed at the study of the singularity theorems from a physical-mathematical perspective. New increasingly complex structures will be generated from a very first introduction to the concept of smooth manifold and the different ideas developed throughout the chapters. First of all, smooth manifolds will be equipped with a metric tensor which will make the definition of semi-Riemannian manifold possible, generating a general way of measuring lengths, angles, curvature, etc. Afterward, a particular type of semi-Riemannian manifolds will be outlined: Lorentz manifolds; the general idea of space-time will be built equipping them with the concept of time orientation. At this point, General Relativity Theory and some of its simplest models will be studied briefly providing a natural connection between the previous mathematical ideas — as this theory relies on them — and the subsequent ones, introducing the concept of singularity and motivating the study of Causality theory. In this sense, the last chapter of the project will focus on the most basic concepts of Causality theory — causal relations, Cauchy hypersurfaces, etc. — generating a formalism that will culminate with the construction of the singularity theorems which will generalize the idea of singularity in any space-time under certain conditions.

Índice general

Introducción	3
1. Preámbulo: Variedades diferenciables y tensores	7
1.1. Variedades diferenciables	7
1.2. Aplicaciones diferenciables	10
1.3. Espacio tangente. Aplicación Diferencial	11
1.4. Campos vectoriales. Curvas integrales y flujos	14
1.5. 1-Formas	17
1.6. Campos de tensores. Formas bilineales simétricas y producto escalar	19
2. Elementos principales de la teoría de variedades semi-Riemannianas	27
2.1. Variedades semi-Riemannianas	27
2.2. La conexión de Levi-Civita	32
2.3. Transporte paralelo. Geodésicas y aplicación exponencial	37
2.4. Tensor de curvatura Riemanniano. Curvatura seccional	42
2.5. Tensor de Ricci y curvatura escalar	47
3. Variedades de Lorentz: El espacio-tiempo	51
3.1. Espacio-tiempo. Orientabilidad temporal	52
3.2. Geometría extrínseca. Segunda forma fundamental.	57
3.3. Variaciones de longitud. Campos de Jacobi y puntos focales	61
4. Introducción a la Teoría General de la Relatividad	71
4.1. Fundamentos físicos. Materia, gravedad y curvatura	72
4.2. Ecuación de campo de Einstein	78
4.3. Motivación para el estudio de la Causalidad: El Modelo FLRW y las singularidades	82
5. Causalidad en Variedades de Lorentz	95
5.1. Relaciones y condiciones de causalidad. Separación temporal	96
5.2. Conjuntos acronales. Hipersuperficies de Cauchy	106
5.3. Dominios de dependencia y horizontes de Cauchy	111
5.4. Teoremas de singularidad	118
Bibliografía	127

Introducción

Los conceptos de espacio y tiempo han sido siempre intrínsecos a la forma de entender el mundo que tienen los seres humanos. Es por ello que el entendimiento de la naturaleza de ambos ha generado gran controversia, debate y, consecuentemente, producción científica a lo largo de la historia, constituyendo a través de la geometría una de las mayores intersecciones de las disciplinas físicas y matemáticas.

Aunque el concepto de tiempo siempre ha sido más complicado de entender debido a sus propiedades, en concreto su irreversibilidad, el espacio siempre fue una entidad bien conocida tanto desde el punto de vista físico a través del espacio absoluto de Newton — interpretándose como un simple contenedor de objetos, es decir, como aquello que resta en el mundo si desapareciera todo lo demás — como desde el punto de vista matemático a través de la geometría clásica de Euclides. En este contexto, el tiempo en ambos marcos era considerado un simple número, un parámetro capaz de encapsular nuestra idea de movimiento y cambio dentro de este concepto de espacio.

Este mundo clásico que, a priori, podría parecer sólido y verdadero ya que está basado en nuestra intuición geométrica más básica de la realidad, ya presentaba problemas en el siglo XVII cuando el propio Newton se enfrentaba a Leibniz por su interpretación del espacio como una abstracción de las relaciones existentes entre objetos en lugar de algo absoluto existente independientemente de los objetos que contenga. A pesar de que la balanza histórica cayó hacia el lado de Newton, la brecha conceptual debida a la naturaleza del espacio siguió abierta y, en el siglo XVIII, en esta línea de espacio absoluto, sólido e inmutable apareció la figura del filósofo alemán Immanuel Kant (1724 - 1804) quien en su obra *Crítica a la razón pura* estableció que el espacio y el tiempo solo son construcciones mentales que nos permiten estructurar los elementos de nuestra experiencia. En palabras del propio Kant: “El espacio y el tiempo son el marco dentro del cual la mente se ve obligada a construir su experiencia de la realidad”.

Paralelamente a esta revolución del pensamiento en el ámbito filosófico, el mundo de las matemáticas también se desarrolló y la geometría fue evolucionando con las ideas del matemático alemán Carl Friedrich Gauss (1777 - 1855). Gauss, con su desarrollo del análisis intrínseco de superficies — esto es, independientemente del espacio ambiente donde se encuentre la superficie — construyó conceptos como la curvatura o la medida de longitudes, ángulos y áreas en superficies, alejando al mundo de la idea de espacio

inmutable, sólido y absoluto, planteando con fuerza por primera vez si el espacio físico podría ser curvo. En esta misma época, el surgimiento de las geometrías no euclídeas — hipérbolica y elíptica — fueron difuminando cada vez más las ideas clásicas y permitieron superar el jaque mate conceptual que había supuesto la *Crítica a la razón pura* de Kant en el siglo anterior.

El culmen de esta revolución llegó con Bernhard Riemann (1826 - 1866), discípulo de Gauss, que en su tesis de habilitación *Sobre las hipótesis en que se funda la geometría* elevaba las ideas de su maestro a dimensiones superiores, generalizando toda la geometría previa y dando un estocazo final al mundo geométrico clásico o, mejor dicho, incluyéndolo en un nuevo mundo geométrico inmenso lleno de posibilidades. Realmente, este estocazo nunca llegó de parte de Riemann, quien a pesar de construir una obra que venía como anillo al dedo para el momento histórico en el que vivió, colapsando ideas filosóficas profundas con el desarrollo de una nueva matemática, fue completamente incomprendido debido a la oscuridad con la que transmitía sus ideas y la dificultad de estas. Los conceptos Riemannianos no fueron accesibles hasta final de siglo gracias al trabajo de los matemáticos italianos Tullio Levi-Civita (1873-1941) y Gregorio Ricci-Curbastro (1853-1925) quienes usando de hilo conductor el trabajo del matemático Elwin Cristoffel formalizaron la matemática iniciada por Riemann y Gauss dando lugar al cálculo diferencial general y a la geometría diferencial tal y como la conocemos hoy.

La revolución más tardía llegó de manos de la física; al no haberse realizado experimentos que rompieran con las ideas clásicas de espacio y de tiempo, estas se seguían manteniendo intactas. La historia cambió con la aparición de las ecuaciones de Maxwell (1865) — ecuaciones que regían la teoría electromagnética — que no se mostraban invariantes bajo las transformaciones relativistas de Galileo lo que significaba que la teoría electromagnética era incompatible con la idea de espacio de Newton. Tras diversos experimentos, la balanza cayó del lado del electromagnetismo ya que se demostró que experimentalmente la velocidad de la luz tenía que ser constante y además, tenía que ser una cota máxima de velocidad. La mecánica clásica de Newton fue generalizada a la Teoría de la Relatividad Especial de Albert Einstein en 1905, teoría que insertaba la idea de espacio-tiempo de cuatro dimensiones equipado con un nuevo tipo de transformaciones: las transformaciones de Lorentz, que sí mantenían la velocidad de la luz constante y que eran compatibles con las ecuaciones de Maxwell.

De este modo, las ideas clásicas de espacio y tiempo se intercambiaron con el concepto de espacio-tiempo en el que el tiempo era una nueva dimensión más, solo que con características especiales. El carácter clásico y absoluto se perdía, pues las transformaciones de Lorentz implicaban que el espacio se contraía y que el tiempo se dilataba pasando más lento para aquellos observadores que se movieran más rápido. Aunque la mayor parte de estas ideas se le deben a Einstein, otros muchos matemáticos y físicos trabajaron en la formalización geométrica de estas ideas como fueron el propio Lorentz, Minkowski o Poincaré. Este último mantenía gran debate con Einstein acerca de la interpretación de estas nuevas ideas; Einstein, interesado en la filosofía del espacio e incluso en las ideas de Kant, pensaba que con la Teoría de la Relatividad Especial

se había demostrado la falsedad de estas ideas puesto que la teoría probaba cuál era la geometría y la naturaleza “real” del espacio. Sin embargo, Poincaré mantenía una idea completamente opuesta y defendía que la geometría “real” del espacio no podía determinarse experimentalmente.

En este punto, cuestiones filosóficas aparte, parecía que la física volvía a ser consistente. Sin embargo, faltaba una cuestión por resolver, puesto que si la velocidad de la luz era máxima, la gravitación tal y como la había explicado Newton era incompatible con la nueva mecánica. Hacía falta una nueva teoría de la gravedad. Inspirado por su amigo matemático Marcel Grossman, Einstein comenzó a leer las matemáticas de Levi-Civita y Ricci-Curbastro de las que hablamos previamente, proporcionándole las ideas matemáticas perfectas para el desarrollo de sus ideas físicas. En 1915, Einstein publicó la Teoría General de la Relatividad en la que unificaba las ideas de la Teoría Especial con una nueva forma de entender la gravitación. La gravedad dejaba de ser una fuerza para ser simplemente la manifestación de una propiedad del espacio-tiempo, en concreto, su geometría o curvatura en cierto sentido que ya estableceremos: el espacio-tiempo se convirtió en una variedad semi-Riemanniana curvada por la presencia de energía haciendo que las trayectorias sobre él de los cuerpos libres fueran las geodésicas de esta variedad, claramente curvas como lo son las órbitas reales.

La Teoría de la Relatividad General fue probada experimentalmente en múltiples ocasiones, afianzando esta nueva idea de espacio-tiempo mutable y moldeable. El éxito de la teoría llevó a gran parte de la comunidad científica a querer resolver las ecuaciones de Einstein — ecuaciones por las que se rige la teoría — siendo especialmente relevantes las soluciones encontradas por Schwarzschild sobre la gravedad alrededor de un cuerpo esférico como una estrella y las debidas a Friedmann, Lemaître, Robertson y Walker que trataban el universo como un todo dinámico. Ambas soluciones, a pesar de ser claramente diferentes, tenían en común la aparición de puntos que desquiciarían a Einstein, puntos donde la teoría fallaba y donde el espacio-tiempo dejaba de tener sentido: las singularidades. En el caso de Schwarzschild su solución llevó a lo que hoy conocemos como agujeros negros y en el caso de Friedmann, Lemaître, Robertson y Walker a conceptos como el Big Bang y el estudio del final cosmológico del universo; puntos de densidad y curvatura infinita donde el cosmos se rompe.

Durante muchos años los físicos, y en particular Einstein, se mantuvieron reticentes a que estas soluciones fueran físicamente plausibles, y argumentaban que eran matemáticamente posibles debido a la gran simetría supuesta en los modelos para resolver las ecuaciones de campo. Con el paso del tiempo, la experimentación y ciertos desarrollos teóricos, empezaron a sostener la idea de que la formación de agujeros negros como los que predecía la solución de Schwarzschild eran posibles en el colapso de ciertas estrellas como argumentaban Oppenheimer y Snyder en su artículo de 1939. Sin embargo, el hecho de que existieran las singularidades de forma física y solo bajo condiciones de alta simetría seguía siendo un enigma. La mayor parte de la comunidad física siguió abordando el problema buscando soluciones detalladas de las ecuaciones de campo de Einstein, y en 1963 los científicos rusos Lifshitz y Khalatniikov afirmaron haber demos-

trado que este tipo de soluciones solo podían obtenerse bajo distribuciones de energía y velocidades muy concretas.

El paradigma no cambió hasta que posteriormente (1965) el matemático británico Roger Penrose (1931-Presente) y el físico británico Stephen Hawking (1942-2018) abordaron el problema de una forma mucho más general y matemática. Penrose y Hawking tuvieron la perspicacia de percatarse de que las singularidades estaban estrechamente relacionadas con el hecho de que las geodésicas de la variedad estén incompletas en cierto sentido que ya precisaremos y que, por ello, en estos puntos no tiene sentido hablar de espacio-tiempo. Ambos aplicaron las herramientas de la teoría matemática de Causalidad para estudiar las relaciones causa-efecto entre diferentes puntos de la variedad espacio-temporal con la finalidad de encontrar qué condiciones suficientes debe de cumplir esta para que se produzcan singularidades. Como veremos, estas condiciones son simples e intuitivas, y es muy factible que nuestro universo las cumpla.

Hemos viajado históricamente desde un espacio tridimensional inmutable y rígido en que el tiempo no era más que un parámetro, a un mundo en cuatro dimensiones en el que una de esas dimensiones es el tiempo y que es completamente moldeable y cambiante, llegando incluso a puntos donde este no tiene sentido y donde sería necesaria una nueva definición de espacio-tiempo o una nueva forma de entender la gravedad; preguntas que aún a día de hoy siguen abiertas.

En este documento se abordarán las teorías geométricas desarrolladas por Gauss, Riemann, Levi-Civita, Cristoffel y Ricci en los dos primeros capítulos, haciendo una introducción a la idea de variedad diferenciable y posteriormente de variedad semi-Riemanniana al equipar esta con un tensor métrico. En el tercer capítulo se introducirá la idea de orientación temporal definiendo el concepto de espacio-tiempo a través de un tipo especial de variedad semi-Riemanniana, las variedades de Lorentz, estudiándose además con más profundidad la geometría en ellas centrándonos en particular en las geodésicas. En el cuarto capítulo se introducirá la Teoría de la Relatividad General y los modelos de Schwarzschild y Friedmann, Lemaître, Robertson y Walker, cuyo formalismo se basa en toda esta geometría previa sirviéndonos de nexo perfecto entre estas ideas matemáticas y la teoría de la Causalidad a través del concepto de singularidad. Finalmente, en el quinto capítulo desarrollaremos la teoría de la Causalidad construyendo los teoremas de singularidad debidos a Hawking y Penrose, estableciendo unas condiciones causales y energéticas sobre el espacio-tiempo suficientes para la aparición de singularidades.

Destacamos que el carácter de este texto será principalmente matemático aunque se traten conceptos con finalidad física buscando un equilibrio entre ambas disciplinas. Es por ello que el capítulo cuatro, en el que principalmente se tratará una teoría física, se desarrollará desde un punto de vista formal y físico-matemático, y aquellos capítulos matemáticos, irán acompañados de intuiciones físicas e ideas que se relacionarán con conceptos que se mostrarán más adelante. Destacamos además en este punto que, al principio de cada capítulo, aparecerán introducciones conectando con la historia aquí

presentada para dotar de contexto a las ideas planteadas y, además, se mostrará la literatura usada para el capítulo en cuestión con el fin de dejar claro que la mayor parte de las demostraciones pertenecen a dichos textos y para que el lector tenga una bibliografía propia de cada capítulo si necesitara profundizar en los conceptos tratados. A pesar de esto último, recalamos que el desarrollo general del documento está basado en [18] y [24].

Recalamos además la excesiva extensión del documento para un trabajo de este tipo. Debido a la necesidad de recopilar todos los resultados de *Variedades Diferenciables* y de *Geometría semi-Riemanniana* como preámbulo para poder realizar el desarrollo posterior, las dimensiones del documento son excesivas. Sin embargo, no existía otra manera de agrupar todo este contenido haciendo un documento autocontenido de menor extensión.

En relación a esto último, en el desarrollo del trabajo se ha dedicado una mayor profundidad a aquellos conceptos más cercanos a los tratados en las asignaturas del Grado en Matemáticas de la Universidad de Sevilla, y se han desarrollado de forma más escueta aquellas ideas más complejas, que simplemente son necesarias para alcanzar los resultados deseados. Por ello, se podrá apreciar quizás un mayor detalle a la hora de desarrollar las ideas en los Capítulos 1,2 y 4, mientras que los Capítulos 3 y 5 — más cercanos a asignaturas de Máster — aparecen también con profundidad, pero de forma menos detallada.

Para acabar esta introducción, me gustaría agradecer a todas las personas que han estado a mi lado estos cinco años de carrera, que en algún momento despertaron en mí curiosidad o que simplemente estuvieron ahí cuando lo necesité. En primer lugar, quiero mostrar mis más sinceros agradecimientos a mi tutor Alfonso Carriazo por haberme apoyado desde el segundo año de carrera, por la motivación y curiosidad que ha despertado en mí y por haberme dado un hueco en su tiempo para poder aprender todas las matemáticas que siempre había querido, echando un primer vistazo formal a una de mis teorías físicas favoritas como es la Relatividad General.

Me gustaría agradecer también en el ámbito académico al profesor José Oñorbe por haberme dedicado tanto tiempo este último año de carrera permitiéndome aplicar todos los conocimientos que he adquirido estos años a la disciplina que siempre me había apasionado: la cosmología. Aprovecho también para agradecer a mis compañeros tanto del Doble Grado en Física y Matemáticas como de los grados separados por el apoyo que nos hemos dado y el grupo que hemos creado durante estos cinco años.

Finalmente, en un plano más personal, me gustaría dedicar este trabajo a mis profesores del instituto Arroyo Hondo en Rota, en especial a Carlos Vázquez, Regli Niño y Marta Cirilo, por despertar en mí una curiosidad implacable por la física, la filosofía y las matemáticas; a mis amigos de siempre (BdD) y a mi familia, en especial a mi madre y a mis abuelos.

Para acabar, y aunque es algo cliché, me gustaría dejar aquí esta frase que, además

de otras muchas experiencias, creo que condensa muy bien el sentimiento de pequeñez que podemos tener al acabar el grado ante el mundo de conocimiento que se nos ha planteado, pero, a su vez, la fuerza, el saber y la curiosidad con la que nos hemos cargado estudiando estos cinco años:

El aleteo de una mariposa puede ser la causa de un huracán en la otra punta del mundo

Con afecto, el autor.

Capítulo 1

Preámbulo: Variedades diferenciables y tensores

La primera parte de este documento, constituida por los tres primeros capítulos, consistirá en una escalera de complejidad conceptual hacia la idea de espacio-tiempo y hacia aquellos términos matemáticos necesarios para establecer el formalismo de la Teoría de la Relatividad General y de la Teoría de Causalidad. En este primer capítulo presentamos el objeto “espacial” más simple en complejidad: la variedad diferenciable, concepto que nos servirá de punto de partida y que iremos equipando con estructuras cada vez más complejas hasta llegar al espacio-tiempo.

Dedicaremos este capítulo a hacer un repaso general de aquellos conceptos fundamentales relacionados con variedades diferenciables y también introduciremos las ideas de tensor y forma bilineal que serán de vital importancia en el desarrollo que presentaremos en el Capítulo 2. Dado que, en general, estas ideas han sido estudiadas en la asignatura de Variedades Diferenciables del Grado en Matemáticas de la Universidad de Sevilla, presentaremos los resultados sin demostración. Haremos esto también en la parte dedicada a formas bilineales, ya que la mayor parte de los resultados tienen una demostración sencilla basada en el *Álgebra Lineal*. Para el lector interesado en las demostraciones omitidas o que desee profundizar en los conceptos que se presentan, pueden consultarse los textos en los que está basado este capítulo [16], [18], [21].

1.1. Variedades diferenciables

La forma más naif de pensar en una variedad diferenciable es considerar un espacio topológico que al ser observado desde cerca — localmente — se parece al espacio Euclídeo usual \mathbb{R}^n pero que al ser observado de lejos — globalmente — puede tener una forma o estructura más compleja. En este mismo sentido, podemos pensar de forma simple en la construcción matemática de una variedad diferenciable como un conjunto

de parches locales que se parecerán a \mathbb{R}^n — cartas — que coseremos entre sí con el fin de tener un tejido — atlas — de forma que dichos parches se cosan de manera que el paso de uno a otro sea suave — diferenciabilidad de la variedad. Nótese que, al fin y al cabo, lo que buscamos es que este parecido entre el espacio topológico y el espacio usual sea lo suficientemente fino como para poder generalizar todo el cálculo diferencial que conocemos a la variedad diferenciable.

A continuación, estableceremos formalmente las definiciones de los conceptos esbozados. En lo que sigue, consideraremos que una función es diferenciable si es C^∞ y siempre que nos refiramos a un espacio topológico (N, \mathcal{T}) lo consideraremos 2º numerable con la propiedad de Hausdorff y abusaremos de la notación llamándolo simplemente N .

Comencemos de este modo definiendo nuestros parches, esto es, las cartas.

Definición 1.1.1 (Carta) *Dado un espacio topológico N diremos que el par (U, φ) es una carta de dimensión n de N si U es un abierto de N y $\varphi : U \rightarrow \mathbb{R}^n$ es un homeomorfismo sobre su imagen $\varphi(U)$ que es abierto de \mathbb{R}^n . En particular, llamaremos a U dominio de la carta y a φ función coordenada.*

Definición 1.1.2 (Función componente y coordenadas locales) *Dados una carta $(U, \varphi = (x^1, \dots, x^n))$ y un punto $p \in U$, se podrá escribir $\varphi(p) = (x^1(p), \dots, x^n(p))$ y llamaremos a cada una de las funciones $x^i : U \rightarrow \mathbb{R}$ con $i = 1, \dots, n$ funciones componentes. En particular, llamaremos al conjunto $(x^1(p), \dots, x^n(p))$ coordenadas locales de p respecto de la carta (U, φ) .*

Nótese que si en la definición anterior tomáramos como carta (\mathbb{R}^n, id) podríamos escribir la identidad como $id = (u^1, \dots, u^n)$ que serían funciones que proyectarían de forma natural un punto de \mathbb{R}^n a su componente en \mathbb{R} , esto es, $u^i : \mathbb{R}^n \rightarrow \mathbb{R}$ para $i = 1, \dots, n$. Estas funciones vendrían a ser las funciones componentes de la identidad. Sin embargo, se suelen llamar *funciones coordenadas naturales* debido a su carácter canónico ya que dada una carta (U, φ) podríamos escribir sus funciones componentes como $x^i = u^i \circ \varphi$ para $i = 1, \dots, n$.

Retomando la idea inicial, una vez hemos formalizado nuestros parches, tenemos que establecer matemáticamente qué significa que se cosan de forma suave. Para ello introducimos el concepto de compatibilidad.

Definición 1.1.3 (Cartas compatibles) *Dadas dos cartas n -dimensionales (U, φ) y (V, ψ) de N , diremos que son compatibles si la intersección de sus dominios es vacía o si las composiciones $\varphi \circ \psi^{-1} : \psi(U \cap V) \rightarrow \varphi(U \cap V)$ y $\psi \circ \varphi^{-1} : \varphi(U \cap V) \rightarrow \psi(U \cap V)$ son diferenciables en el sentido anteriormente establecido.*

Finalmente, podemos definir nuestro atlas y a través de él el concepto de variedad diferenciable.

Definición 1.1.4 (Atlas) *Dado un espacio topológico N diremos que \mathcal{A} es un atlas de dimensión n de N si es una colección de cartas $\{(U_i, \varphi_i)\}_{i \in \Lambda}$ de dimensión n compatibles dos a dos tales que $N = \cup_{i \in \Lambda} U_i$.*

Definición 1.1.5 (Variedad diferenciable) *Una variedad diferenciable de dimensión n es un par (N, \mathcal{A}) donde N es un espacio topológico con la propiedad de Hausdorff y \mathcal{A} numerable y \mathcal{A} un atlas de dimensión n .*

Definición 1.1.6 (Carta de una variedad) *Dada una variedad diferenciable (N, \mathcal{A}) diremos que (U, φ) es una carta de N si es una carta de \mathcal{A} .*

Nótese que la dimensión de la variedad diferenciable siempre será la dimensión de su atlas. Además, recalco que esta no es la única forma de definir variedad diferenciable: se podrían haber exigido las condiciones de numerabilidad y de propiedad Hausdorff al propio atlas en lugar de al espacio topológico. Ambas descripciones son completamente equivalentes. De aquí en adelante, nos referiremos a una variedad diferenciable simplemente como variedad y la notaremos como N o N^n donde n es su dimensión y omitiremos su atlas.

En este punto, como adelantábamos antes, es el atlas el que permite hablar de cálculo y análisis sobre N . Podríamos preguntarnos por tanto qué ocurre si diferentes atlas generan la misma forma de hacer análisis sobre N . Resolvemos esta tecnicidad del siguiente modo:

Definición 1.1.7 (Atlas compatibles) *Diremos que dos atlas \mathcal{A} y \mathcal{B} de N son compatibles si todas las cartas de \mathcal{A} son compatibles con todas las cartas de \mathcal{B} y recíprocamente. En particular, $\mathcal{A} \cup \mathcal{B}$ es también un atlas sobre N .*

Proposición 1.1.1 *La relación de compatibilidad anterior es una relación de equivalencia.*

Definición 1.1.8 (Atlas maximal o estructura diferenciable) *Dado un atlas \mathcal{A} sobre N llamaremos atlas maximal o estructura diferenciable a la clase de equivalencia de \mathcal{A} a través de la relación de equivalencia de compatibilidad.*

Nótese que el atlas de la definición anterior se denomina maximal puesto que constituye la familia de todas las cartas que son compatibles con \mathcal{A} , entendiendo la compatibilidad entre una carta y un atlas como la compatibilidad de la carta con todas las

cartas de dicho atlas. Del mismo modo, se denomina estructura diferenciable puesto que en el sentido que comentábamos antes este objeto engloba todos los atlas que generan el mismo análisis o cálculo sobre la variedad. Así pues, hemos solucionado nuestro problema pues todo atlas de una variedad estará contenido en un atlas maximal, esto es, la variedad tendrá asignada una estructura diferenciable.

Para acabar este apartado introduciremos el siguiente resultado que nos permite generar un tipo de variedad que será necesario en apartados posteriores.

Lema 1.1.1 *Dadas dos variedades N^n y M^m , $M \times N$ será una variedad de dimensión $m + n$ a través de las cartas $(U \times V, \varphi \times \psi)$ con (U, φ) carta de N , (V, ψ) carta de M y $\varphi \times \psi : U \times V \rightarrow \mathbb{R}^{n+m}$ definida como $(p, q) \mapsto (\varphi(p), \psi(q))$ para todo $p \in U$ y $q \in V$.*

Llamaremos a este tipo de cartas *carta producto* y a la variedad generada *variedad producto*.

1.2. Aplicaciones diferenciables

Habiendo definido ya el concepto de variedad diferenciable es natural preguntarse ahora cómo funcionan las aplicaciones definidas sobre ellas. En particular, nos interesará saber qué entendemos ahora por función diferenciable. Si comenzamos pensando en el caso más simple, por ejemplo, una función escalar f sobre una variedad N , $f : N \rightarrow \mathbb{R}$ la forma natural para determinar si f es diferenciable sería pasar a su expresión coordenada en el mismo sentido que en la Definición 1.1.2. Así pues, si dada cualquier carta (U, φ) de M se tiene que $f \circ \varphi^{-1} : \varphi(U) \rightarrow \mathbb{R}$ es diferenciable en el sentido usual, entonces diremos que f es diferenciable. Nuestro objetivo es generalizar esta construcción para cualquier aplicación:

Definición 1.2.1 (Aplicación diferenciable) *Dadas dos variedades M y N diremos que una aplicación continua $f : M \rightarrow N$ es diferenciable en un punto $p \in M$ si para toda carta (V, ψ) de N con $f(p) \in V$ existe una carta (U, φ) de M tal que $p \in U$ y se tiene que la aplicación $\bar{f} = \psi \circ f \circ \varphi^{-1} : \varphi(U \cap f^{-1}(V)) \rightarrow \psi(V)$ es diferenciable en $\varphi(p)$. En particular, diremos que f es diferenciable si lo es en todo $p \in M$.*

Se puede demostrar que la definición anterior no depende de las cartas elegidas.

Notaremos el conjunto de todas las aplicaciones diferenciables entre dos variedades M y N como $\mathcal{F}(M, N)$ y en el caso de que la función sea escalar, esto es, que N sea el conjunto real simplemente escribiremos $\mathcal{F}(M)$. En particular, $\mathcal{F}(M)$ es un anillo con las operaciones habituales de suma y producto de funciones.

Para acabar la sección simplemente introduciremos el concepto de difeomorfismo que será esencial en los capítulos posteriores.

Definición 1.2.2 (Difeomorfismo) *Dada una aplicación f entre dos variedades M y N se dirá que f es difeomorfa o que es un difeomorfismo si f es biyectiva y tanto ella como su inversa son diferenciables en el sentido anterior.*

Proposición 1.2.1 *Las cartas son difeomorfismos sobre sus imágenes.*

1.3. Espacio tangente. Aplicación Diferencial

En la sección previa hemos visto cómo la forma natural de definir objetos relacionados con la diferenciabilidad en una variedad era traducir estos a través de las cartas a las ideas usuales que tratamos en \mathbb{R}^n . El siguiente paso para poder generalizar todo el cálculo de \mathbb{R}^n a las variedades tendrá la misma esencia: construiremos el espacio tangente a una variedad aproximándola localmente, lo que nos permitirá traducir todos los conceptos locales ligados a \mathbb{R}^n a cualquier tipo de espacio.

Para la construcción del espacio tangente existen varias estrategias en función de cómo se traten los vectores tangentes. En la asignatura de Variedades Diferenciables definimos el vector tangente a una variedad en un punto como la clase de equivalencia de las curvas que pasan por dicho punto, estando relacionadas dos curvas con origen común si sus vectores tangentes en el origen — vistos en \mathbb{R}^n a través de las cartas — eran iguales. En esta sección seremos aún más generales y definiremos los vectores tangentes axiomatizando el carácter de derivada direccional que tienen los vectores Euclídeos.

Definición 1.3.1 (Vector tangente) *Dada una variedad N y un punto $p \in N$, diremos que una función $v : \mathcal{F}(N) \rightarrow \mathbb{R}$ es un vector tangente si:*

- *Es \mathbb{R} -lineal, esto es: para cualesquiera $f, g \in \mathcal{F}(N)$ y $a, b \in \mathbb{R}$ se tiene que $v(af + bg) = av(f) + bv(g)$.*
- *Cumple la condición de Leibniz, esto es: para cualesquiera $f, g \in \mathcal{F}(N)$ se tiene que $v(fg) = v(f)g(p) + f(p)v(g)$.*

De este modo, definimos el espacio tangente en un punto como el conjunto de todos los vectores tangentes en ese punto.

Definición 1.3.2 (Espacio tangente) *Dada una variedad N y un punto $p \in N$, llamaremos espacio tangente de N en p al conjunto:*

$$T_p(N) = \{v : \mathcal{F}(N) \rightarrow \mathbb{R} \mid v \text{ es vector tangente a } N \text{ en } p\}.$$

En la introducción a la sección ya adelantábamos que queríamos aproximar localmente la variedad por su espacio tangente que, como es lógico, será isomorfo al espacio Euclídeo, permitiéndonos generalizar el cálculo que era nuestro objetivo esencial. Así pues, el espacio tangente en cada punto tendrá que ser un espacio vectorial — ya podíamos intuirlo de la propia axiomatización de vector tangente — y además será isomorfo a \mathbb{R}^n siendo n la dimensión de la variedad — ya que tendríamos que ser capaces de asociar a cada vector tangente el correspondiente vector tangente visto a través de las cartas en \mathbb{R}^n .

Lema 1.3.1 *Dada una variedad N^n el espacio tangente $T_p(N)$ tiene estructura de espacio vectorial de dimensión n para todo $p \in N$. En particular, existe un isomorfismo de espacios vectoriales, que llamaremos isomorfismo lineal canónico, entre $T_p(N)$ y \mathbb{R}^n , que lleva cada vector a él mismo en el punto p .*

En estas circunstancias podemos definir además una nueva estructura uniendo los espacios tangentes en todos los puntos de la variedad.

Definición 1.3.3 (Fibrado tangente) *Dada una variedad N llamaremos fibrado tangente de N al conjunto $T(N) = \cup_{p \in N} T_p(N)$.*

Nótese que la idea geométrico-física detrás del fibrado tangente es caracterizar un sistema a través de sus posiciones y sus velocidades, esto es, los puntos de la variedad y los vectores tangentes en cada punto. En la línea de esta idea tenemos el siguiente resultado:

Lema 1.3.2 *Dada una variedad N^n su fibrado tangente $T(N^n)$ tiene estructura de variedad diferenciable de dimensión $2n$.*

El fibrado tangente y la variedad estarán conectados de manera natural a través de la proyección canónica $\pi : T(N) \rightarrow N$ caracterizada como $\pi(v_p) = p$ para cualquier $p \in N$ y $v_p \in T(N)$.

Una vez axiomatizados los vectores tangentes y construidas las estructuras necesarias estamos en condiciones de generalizar el análisis definiendo las derivadas del siguiente modo:

Definición 1.3.4 (Derivada parcial) *Dada una carta $(U, \varphi = (x^1, \dots, x^n))$ de una variedad N^n y un punto $p \in U$ llamaremos derivada parcial en p a la función:*

$$\partial_i|_p := \frac{\partial}{\partial x^i} \Big|_p : \mathcal{F}(N) \rightarrow \mathbb{R}$$

En particular, dada una función $f \in \mathcal{F}(N)$ su derivada parcial en p será:

$$\frac{\partial f}{\partial x^i}(p) := \frac{\partial(f \circ \varphi^{-1})}{\partial u^i}(\varphi(p))$$

para todo $i = 1, \dots, n$ siendo u^i las funciones coordenadas naturales.

Lema 1.3.3 *En las condiciones de la definición anterior las derivadas parciales $\partial_i|_p$ son vectores tangentes a la variedad N^n en p .*

Nótese que hemos conseguido lo que queríamos, pues basta ver que, a través de la Definición 1.3.1 de vector tangente, el objeto que acabamos de construir tiene las características de una derivada parcial en el sentido usual de forma local. Además, no solo tenemos que las derivadas parciales sean vectores tangentes, sino que tenemos un resultado mucho más fuerte pues constituirán una base de la estructura de espacio vectorial que tiene el espacio tangente en cada punto.

Teorema 1.3.1 (Teorema de la Base) *Dados una variedad N^n , una carta $(U, \varphi = (x^1, \dots, x^n))$ y un punto $p \in U$, el conjunto $\{\partial_i|_p\}_i$ es una base de $T_p(M)$ pudiéndose escribir cualquier $v \in T_p(M)$ como*

$$v = v(x^i)\partial_i|_p,$$

donde se ha usado el convenio de sumación de Einstein e $i = 1, \dots, n$.

Recordamos que el convenio de sumación de Einstein consiste en omitir un sumatorio sobre la dimensión tratada que se considera implícito cuando se repiten índices arriba y abajo — contravariante y covariante como ya estableceremos más adelante en el documento. Recalamos además en todo el desarrollo anterior la importancia de diferenciar objetos y esos mismos objetos *en un punto*. Intentaremos mantener esta notación a lo largo del documento por evitar confusión aunque cuando la notación se haga pesada y se sobrentienda el sentido por contexto la omitiremos.

Para acabar este apartado, en el mismo sentido que nos preguntábamos tras la Sección 1.1 cómo son las aplicaciones entre variedades, ahora nos podemos preguntar cómo se relacionan los espacios tangentes a través de aplicaciones.

Definición 1.3.5 (Aplicación diferencial) *Sea $f : N \rightarrow M$ una aplicación diferenciable. Definimos la aplicación diferencial de f en p para cada $p \in M$ como:*

$$\begin{aligned} df_p : T_p(N) &\rightarrow T_p(M) \\ v &\mapsto df_p(v) \end{aligned}$$

definiéndose $df_p(v)$ como $df_p(v)(g) = v(g \circ f)$ para toda $g \in \mathcal{F}(M)$.

Nótese que la aplicación diferencial es lineal y podemos calcular su matriz jacobiana del siguiente modo:

Lema 1.3.4 *Dados un punto $p \in N^n$, $f : N^n \rightarrow M^m$ una aplicación diferenciable, $(U, \varphi = (x^1, \dots, x^n))$, $(V, \psi = (y^1, \dots, y^m))$ cartas de N^n y M^m respectivamente con $p \in U$ y $f(p) \in V$, se tiene que:*

$$df_p \left(\frac{\partial}{\partial x^j} \Big|_p \right) = \frac{\partial (y^i \circ f)}{\partial x^j} (p) \frac{\partial}{\partial y^i} \Big|_{f(p)}$$

con $i = 1, \dots, m$ y $j = 1, \dots, n$. En particular, estamos estableciendo que la matriz de df_p respecto a las bases de $T_p(U)$ y $T_{f(p)}(V)$ que nos da el Teorema de la Base es la matriz de la aplicación $\bar{f} = \psi \circ f \circ \varphi^{-1}$. Esto es, la matriz jacobiana de f en p relativa a las cartas previas es $J_p(f) := \frac{\partial (y^i \circ f)}{\partial x^j} (p) = df_p$.

El lema previo nos permite definir qué significa que una variedad sea orientable, concepto que será necesario diferenciar del concepto de orientación temporal que aparecerá en el capítulo tres.

Definición 1.3.6 (Variedad orientable) *Diremos que una variedad N es orientable si existe un conjunto \mathcal{R} de cartas cuyos dominios recubran N y tal que para cada par de cartas el determinante de la matriz jacobiana introducida en el Lema 1.3.4 sea positivo.*

Acabemos la sección introduciendo el Teorema de la Función Inversa que será necesario en resultados posteriores:

Teorema 1.3.2 (Teorema de la función inversa) *Dadas dos variedades N y M y una función $f \in \mathcal{F}(N)$, la aplicación diferencial df_p con $p \in N$ es un isomorfismo lineal si y solo si existe un entorno U de p en N tal que f restringida a dicho entorno sea un difeomorfismo sobre su imagen $f(U)$ entorno de $f(p)$ en M .*

1.4. Campos vectoriales. Curvas integrales y flujos

Continuando con nuestro desarrollo de conceptos, en esta sección vamos a dotar de mayor complejidad al concepto de vector tangente previo introduciendo la idea de campo vectorial. Los campos vectoriales se pueden interpretar de tres maneras diferentes como veremos a continuación.

Definición 1.4.1 (Campo vectorial (1)) *Un campo vectorial sobre una variedad N es una aplicación $V : N \rightarrow T(N)$ que a cada punto $p \in N$ le asocia un vector tangente $V_p \in T_p(N)$ verificando que $\pi \circ V = id$ con π la proyección canónica.*

Definición 1.4.2 (Campo vectorial (2)) *Un campo vectorial sobre una variedad N es una aplicación $V : \mathcal{F}(N) \rightarrow \mathcal{F}(N)$ definida como $(Vf)(p) = V_p(f)$ para todo $p \in N$, con $f \in \mathcal{F}(N)$ y con $V_p(f) \in T_p(N)$.*

Las dos definiciones anteriores son equivalentes. Las introducimos por separado pues en ocasiones nos será útil pensar en un campo vectorial como un objeto que a cada punto de la variedad le asocia un vector pero, en otras ocasiones, será mucho más intuitivo pensar en un campo vectorial en el sentido de la segunda definición.

Nótese que la segunda definición es una generalización mucho más directa del concepto de vector tangente ya que cumple exactamente las condiciones de la Definición 1.3.1 solo que ahora tenemos una aplicación de $\mathcal{F}(N)$ en sí mismo. A este tipo de objetos se les conoce como derivaciones. Así, podemos pensar — sin ser formales — que cuando “aplicamos” un campo vectorial en cierto sentido estamos derivando; de hecho, en el sentido de la segunda definición Vf sería la derivada de f en la dirección del campo V . Recalamos además que aunque el concepto de derivación es una abstracción mayor, se puede probar que toda derivación proviene de un campo vectorial [18].

Podemos extender el concepto de diferenciabilidad a los campos tal y como sigue:

Definición 1.4.3 (Campo vectorial diferenciable) *Diremos que un campo vectorial V sobre N es diferenciable si para toda $f \in \mathcal{F}(N)$ se tiene que Vf es diferenciable.*

Notaremos el conjunto de campos vectoriales diferenciables sobre N por $\mathcal{X}(N)$. Este conjunto tiene la siguiente estructura:

Lema 1.4.1 *Dada una variedad N , el conjunto de los campos vectoriales diferenciables sobre ella $\mathcal{X}(N)$ es un espacio vectorial sobre \mathbb{R} y un módulo sobre $\mathcal{F}(N)$.*

Al igual que ocurría antes con los vectores y el Teorema de la Base, el hecho de que $\mathcal{X}(N)$ sea un espacio vectorial nos permite dar una forma general de escribir los campos vectoriales diferenciables:

Lema 1.4.2 *Dada una carta $(U, \varphi = (x^1, \dots, x^n))$ de N y $V \in \mathcal{X}(U)$ se tiene que el conjunto $\{\partial_i\}_i$ es una base de $\mathcal{X}(U)$. Además, podemos expresar cualquier $V \in \mathcal{X}(U)$ como:*

$$V = (Vx^i)\partial_i$$

para $i = 1, \dots, n$.

Nótese que en el lema anterior estamos viendo ∂_i como campos vectoriales, esto es: en las condiciones del lema anterior $\partial_i : U \rightarrow T(U)$ con $\partial_i(p) = \partial_i|_p$.

Además de las formas anteriores, como anunciábamos al inicio de la sección, podemos pensar en campos vectoriales de un tercer modo. Podemos introducirlos como ecuaciones diferenciales que tienen una solución particular. Esta solución la introducimos con la idea de curva integral.

De aquí en adelante I denotará un intervalo abierto de \mathbb{R} y la definición de curva es la misma que para superficies, pero en nuestro formalismo. Una curva en una variedad N es una aplicación diferenciable $\alpha : I \rightarrow M$ que tendrá un vector velocidad en cada punto p dado por $\alpha'(t) = d\alpha\left(\frac{d}{du}\Big|_p\right)$ — nótese que t es el parámetro y u la coordenada asociada en \mathbb{R} .

Definición 1.4.4 (Curva integral) *Una curva α en una variedad N será una curva integral de $V \in \mathcal{X}(N)$ si $\alpha'(t) = V_{\alpha(t)}$ para todo $t \in I$.*

Podemos pensar en cierto modo que el campo le proporciona a la curva la velocidad de una especie de flujo a través de la variedad.

Al estar tratando, en cierta manera, con ecuaciones diferenciales sobre N , tenemos los siguientes resultados de unicidad para curvas integrales:

Proposición 1.4.1 *Dado $V \in \mathcal{X}(N)$ para todo $p \in N$ existe un intervalo I alrededor del 0 y una única curva integral α con $\alpha(0) = p$.*

Corolario 1.4.1 *Sean α y β curvas integrales de V tales que existe $t^* \in I$ para el que $\alpha(t^*) = \beta(t^*)$ entonces $\alpha = \beta$ para todo $t \in I$.*

Las curvas integrales podrán estar definidas en todo \mathbb{R} o no, esto nos lleva a la siguiente definición:

Definición 1.4.5 (Campo vectorial completo) *Diremos que un campo vectorial $V \in \mathcal{X}(N)$ es completo si todas sus curvas integrales están definidas en todo \mathbb{R} .*

En este caso concreto, como esbozamos antes, podemos colapsar todas las curvas integrales de un campo vectorial en una sola aplicación que puede interpretarse como un flujo moviéndose en el tiempo.

Definición 1.4.6 (Flujo de un campo vectorial) *Dado un campo vectorial completo $V \in \mathcal{X}(N)$ llamaremos flujo a la aplicación $\psi : N \times \mathbb{R} \rightarrow N$ caracterizada por $\psi(p, t) = \alpha_p(t)$ con α_p la curva integral de V que empieza en p para todo $p \in N$ y $t \in I$.*

Para acabar la sección introducimos el corchete de Lie que será una operación que usemos con regularidad en el siguiente capítulo.

Definición 1.4.7 (Corchete de Lie) *Llamaremos corchete de Lie en una variedad diferenciable N a la aplicación:*

$$\begin{aligned} [\cdot, \cdot] : \mathcal{X}(N) \times \mathcal{X}(N) &\rightarrow \mathcal{X}(N) \\ (X, Y) &\mapsto [X, Y] = XY - YX \end{aligned}$$

Nótese que el corchete de Lie cumple las propiedades usuales de un álgebra de Lie [18].

1.5. 1-Formas

Antes de centrarnos en la última sección del capítulo, que estará enfocada en tensores, tenemos que introducir el concepto de 1-forma. Las 1-formas son los objetos duales a los campos vectoriales. Por ello, para introducirlos tendremos que repetir el desarrollo previo pero de forma dual.

Definición 1.5.1 (Espacio cotangente) *Dada una variedad N y un punto $p \in N$, llamaremos espacio cotangente de N en p al dual del espacio tangente en dicho punto:*

$$T_p(N)^* = \{\omega : T_p(N) \rightarrow \mathbb{R} \mid \omega \text{ es lineal}\}$$

También podemos construir el fibrado, pero de forma dual.

Definición 1.5.2 (Fibrado cotangente) *Dada una variedad N llamaremos fibrado cotangente de N al conjunto $T(N)^* = \cup_{p \in N} T_p(N)^*$.*

Al igual que ocurría antes, el fibrado cotangente y la variedad estarán conectados de manera natural a través de la proyección canónica $\pi^* : T(N)^* \rightarrow N$ caracterizada como $\pi^*(\omega_p) = p$ para cualquier $p \in N$ y $\omega_p \in T(N)^*$.

Una vez construidas las estructuras duales podemos establecer el objeto dual al campo vectorial. Como ocurría con los campos vectoriales, podemos definir de dos maneras equivalentes las 1-formas solo que sobre los espacios duales.

Definición 1.5.3 (1-forma (1)) *Una 1-forma sobre una variedad N es una aplicación $\theta : N \rightarrow T(N)^*$ que a cada punto $p \in N$ le asocia un elemento del espacio cotangente $\omega_p \in T_p(N)^*$ verificando que $\pi^* \circ \theta = id$ con π^* la proyección canónica.*

Definición 1.5.4 (1-forma (2)) Una 1-forma sobre una variedad N es una aplicación $\omega : \mathcal{X}(N) \rightarrow \mathcal{F}(N)$ definida como $(\omega V)(p) = \omega_p(V_p)$ para todo $p \in N$, con $V \in \mathcal{X}(N)$ y con $\omega_p \in T_p(N)^*$.

De nuevo, nos centramos en la segunda definición para establecer la diferenciabilidad.

Definición 1.5.5 (1-forma diferenciable) Diremos que una 1-forma ω sobre N es diferenciable si para todo $V \in \mathcal{X}(N)$ se tiene que ωV es diferenciable.

Notaremos el conjunto de 1-formas diferenciables sobre N por $\mathcal{X}(N)^*$. Este conjunto tiene la siguiente estructura:

Lema 1.5.1 Dada una variedad N , el conjunto de las 1-formas diferenciables sobre ella $\mathcal{X}(N)^*$ es un espacio vectorial sobre \mathbb{R} y un módulo sobre $\mathcal{F}(N)$.

Finalmente, el hecho de que $\mathcal{X}(N)^*$ sea un espacio vectorial nos permite dar una manera general de escribir las 1-formas diferenciables. A diferencia de antes, la base no vendrá dada por las ∂_i , sino que tenemos que encontrar la base dual. Para ello, definimos la diferencial — no debe confundirse con la aplicación diferencial definida previamente.

Definición 1.5.6 (Diferencial) Llamaremos diferencial a la aplicación d que lleva cada función diferenciable f a la 1-forma df verificándose que $(df)(V) = Vf$ para todo campo vectorial diferenciable V .

Podemos ahora acabar la sección con el resultado dual que buscábamos:

Lema 1.5.2 Dados una carta $(U, \varphi = (x^1, \dots, x^n))$ de N y $V \in \mathcal{X}(N)$ se tiene que el conjunto $\{dx^i\}_i$ es una base de $\mathcal{X}(N)^*$, es decir, es la base dual de $\{\partial_i\}_i$ cumpliendo que $dx^i(\partial_j) = \delta_{ij}$ siendo δ_{ij} la delta de Kroenecker para $i, j = 1, \dots, n$. Además, podemos expresar cualquier $\omega \in \mathcal{X}(N)^*$ como

$$\omega = (\omega \partial_i) dx^i$$

para $i = 1, \dots, n$.

Nótese que en el sentido del lema anterior podemos expresar la diferencial de cualquier función $f \in \mathcal{F}(N)$ como: $df = \frac{\partial f}{\partial x^i} dx^i$.

1.6. Campos de tensores. Formas bilineales simétricas y producto escalar

Para acabar el capítulo generalizaremos los objetos previamente definidos — las 1-formas y los campos vectoriales — a través del concepto de tensor. Los tensores jugarán un papel esencial a lo largo del documento ya que serán los que nos permitan dotar de mayor estructura a la variedad diferenciable definiendo nuevos objetos con determinadas características como veremos en el capítulo siguiente.

Esta sección estará dividida en dos partes bien diferenciadas. Primero desarrollaremos la idea de tensor de un modo general tal y como hemos hecho antes con las 1-formas y los campos vectoriales: estudiaremos su estructura, las operaciones que podemos realizar con ellos y cómo generalizan toda la teoría ya desarrollada. En la segunda parte de la sección, nos centraremos en un tipo concreto de tensor: las formas bilineales simétricas; en concreto, generalizaremos la idea que tenemos de producto escalar. Esto será esencial en el siguiente capítulo a la hora de definir las variedades semi-Riemannianas ya que necesitaremos un objeto que genere cierta noción de medida dentro de la variedad.

Adelantamos además que, durante toda la sección, se seguirá un formalismo puramente matemático bastante alejado de la idea de tensor que se tiene en física. En física se suele definir un tensor como la generalización de un vector en el sentido de que toma determinada forma en una base y la mantiene “bien” bajo transformaciones (cambios de base o sistema de referencia). Aunque esta idea es completamente compatible con el desarrollo que aquí realizamos, no nos detendremos en estos aspectos ni partiremos de ellos para nuestras definiciones pues se aleja de nuestros objetivos principales.

Comencemos pues definiendo qué es un tensor sobre un K -módulo V .

Definición 1.6.1 (Tensor) *Dado un K -módulo V con K un anillo, llamaremos tensor de tipo (r,s) a toda aplicación K -multilineal $T : (V^*)^r \times V^s \rightarrow K$ con $(r, s) \in \mathbb{N} \times \mathbb{N}$ no simultáneamente nulos.*

Nótese que si r y s son cero a la vez simplemente tendríamos un objeto de K .

Seguidamente generalizamos esta definición a la de campo tensorial sobre una variedad.

Definición 1.6.2 (Campo tensorial) *Un campo tensorial T de tipo (r, s) sobre una variedad N es un tensor de tipo (r, s) sobre el $\mathcal{F}(N)$ -módulo $\mathcal{X}(N)$.*

De este modo, un campo tensorial no es más que una “máquina” a la que damos r 1-formas θ^i y s campos vectoriales V_j , y que nos devuelve una función diferenciable que depende de ellos: $T(\theta^1, \dots, \theta^r, V_1, \dots, V_s) \in \mathcal{F}(N)$.

Sin embargo, aún no es claro el sentido que tenemos que darle a esta función, puesto que no sabemos cómo se comporta en cada punto. Solucionemos esto con el siguiente resultado.

Lema 1.6.1 *Todo campo tensorial T de tipo (r, s) sobre una variedad N devuelve en cada punto $p \in N$ un tensor τ_p de tipo (r, s) sobre el módulo $T_p(N)$, esto es, $\tau_p : (T_p(N)^*)^r \times T_p(N)^s \rightarrow \mathbb{R}$.*

De esta manera, nuestra función anterior $T(\theta^1, \dots, \theta^r, V_1, \dots, V_s) \in \mathcal{F}(N)$, nos devolverá en cada punto $p \in N$ un tensor que nos llevará a un elemento de \mathbb{R} . Este tensor será simplemente la función evaluada en los vectores y elementos del cotangente que devuelven las 1-formas y los campos vectoriales en dicho punto, esto es: $T(\theta^1, \dots, \theta^r, V_1, \dots, V_s)(p) = \tau_p(\omega^1, \dots, \omega^r, v_1, \dots, v_s)$ con $\omega^i = \theta^i|_p$ y $v_j = V_j|_p$ para todo $i = 1, \dots, r$ y $j = 1, \dots, s$.

Podemos, por tanto, interpretar un campo tensorial en cierto modo como una generalización del campo vectorial — como un campo propiamente dicho — ya que en cada punto de la variedad nos devuelve un tensor tal y como hemos establecido.

De aquí en adelante, notaremos el conjunto de los tensores de tipo (r, s) sobre N como Λ_s^r y por simplicidad hablaremos indistintamente de tensores o campos tensoriales. Además, se puede probar (es directo por todo lo anterior) que Λ_s^r es un $\mathcal{F}(N)$ -módulo.

A continuación, caracterizaremos los tensores en función de “sobre qué objeto actúa”, es decir, los caracterizaremos en función de si actúan sobre el módulo en sí — índices abajo — o sobre su dual — índices arriba —.

Definición 1.6.3 (Contravariante y covariante) *Diremos que un tensor es contravariante si es de tipo $(r, 0)$ y diremos que es covariante si es de tipo $(0, s)$; en cualquier otro caso diremos que es un tensor mixto.*

Una vez caracterizados los tensores y fijada la notación, veamos cómo recuperamos los conceptos de 1-forma y campos vectoriales a través del concepto de tensor.

- Las 1-formas son tensores covariantes de tipo $(0, 1)$, es decir, $\Lambda_1^0 = \mathcal{X}(N)^*$. Recuérdese que las 1-formas llevan campos vectoriales a funciones diferenciables asociadas que en cada punto estarán evaluadas sobre un vector y nos darán un escalar.
- Los campos vectoriales son tensores contravariantes de tipo $(1, 0)$, es decir, $\Lambda_0^1 = \mathcal{X}(N)$. Nótese que los campos vectoriales llevan 1-formas a funciones diferenciables asociadas que en cada punto estarán evaluadas sobre un elemento del cotangente y nos darán de nuevo un escalar.

Por uniformidad con lo que acabamos de mencionar, llamaremos a los vectores usuales contravariantes y a los elementos del espacio cotangente vectores covariantes.

Otro tipo de “tensor” que será de gran utilidad más adelante es aquel que surge de la siguiente construcción. Si consideramos una aplicación $\mathcal{F}(N)$ -multilineal $T : \mathcal{X}(N)^s \rightarrow \mathcal{X}(N)$ podemos construir un tensor $(1, s)$ $\tau : \mathcal{X}(N)^* \times \mathcal{X}(N)^s \rightarrow \mathcal{F}(M)$ caracterizado del siguiente modo: $\tau(\theta, V_1, \dots, V_s) = \theta(T(V_1, \dots, V_s))$.

Tras haber visto cómo los tensores engloban todos los objetos previos, la idea ahora es desarrollar una notación en componentes para los tensores de modo que se puedan escribir en bases como ocurre en los casos particulares de 1-forma y campo vectorial. Para ello, simplemente estudiemos cómo vienen dadas las componentes de un tensor y definamos la operación de *producto tensorial* que nos permitirán extrapolar las bases que ya conocemos a bases más complejas y generales.

Definición 1.6.4 (Componentes de un tensor) *Dada una carta $(U, \varphi = (x^1, \dots, x^n))$ de una variedad N^n las componentes de un tensor $T \in \Lambda_s^r(N)$ relativas a dicha carta serán las funciones:*

$$T_{j_1, \dots, j_s}^{i_1, \dots, i_r} = T(dx^{i_1}, \dots, dx^{i_r}, \partial x_{j_1}, \dots, \partial x_{j_s}) \quad i_1, \dots, i_r, j_1, \dots, j_s \in \{1, \dots, n\}$$

Nótese como hemos dicho antes que estas expresiones son consistentes con los Lemas 1.4.2 y 1.5.2.

Pasemos a continuación a definir las operaciones. La suma de tensores se realizará en un sentido natural siendo el tensor resultante el formado por la suma de las componentes de la definición anterior. Centrémonos en el producto de tensores.

Definición 1.6.5 (Producto tensorial) *Dados dos tensores $A \in \Lambda_s^r(N)$ y $B \in \Lambda_{s'}^{r'}(N)$ definimos el producto tensorial de A y B como el tensor $T_1 \otimes T_2 : (\mathcal{X}(N)^*)^{r+r'} \times \mathcal{X}(N)^{s+s'} \rightarrow \mathcal{F}(N)$ caracterizado como:*

$$(T_1 \otimes T_2)(\theta^1, \dots, \theta^{r+r'}, V_1, \dots, V_{s+s'}) = T_1(\theta^1, \dots, \theta^r, V_1, \dots, V_s) T_2(\theta^{r+1}, \dots, \theta^{r+r'}, V_{s+1}, \dots, V_{s+s'}).$$

En particular, la relación en la expresión coordenada es:

$$(T_1 \otimes T_2)_{j_1, \dots, j_{s+s'}}^{i_1, \dots, i_{r+r'}} = T_{j_1, \dots, j_s}^{i_1, \dots, i_r} T_{j_{s+1}, \dots, j_{s+s'}}^{i_{r+1}, \dots, i_{r+r'}}$$

Podemos finalmente escribir los tensores de forma general:

Lema 1.6.2 *Dada una carta $(U, \varphi = (x^1, \dots, x^n))$ de N cualquier tensor $T \in \Lambda_s^r$ se puede escribir como:*

$$T = T_{j_1, \dots, j_s}^{i_1, \dots, i_r} \partial_{i_1} \otimes \dots \otimes \partial_{i_r} \otimes dx^{j_1} \otimes \dots \otimes dx^{j_s}$$

Para acabar con la primera parte de la sección presentamos dos operaciones tensoriales más que nos serán de utilidad: la contracción y el pullback.

Definición 1.6.6 (Contracción) *Dada una carta $(U, \varphi = (x^1, \dots, x^n))$ de N y un tensor $T \in \Lambda_s^r(U)$, llamaremos contracción (i, j) del tensor T al tensor $C_j^i(T) \in \Lambda_{s-1}^{r-1}(U)$ que cumple que*

$$C_j^i(T)(\theta^1, \dots, \theta^{r-1}, V_1, \dots, V_{s-1}) = T(\theta^1, \dots, \theta^{i-1}, dx^k, \theta^i, \dots, \theta^{r-1}, V_1, \dots, V_{j-1}, \partial_k, V_j, \dots, V_{s-1}),$$

donde nótese que se está usando el convenio de sumación de Einstein sobre k .

Definición 1.6.7 (Pullback) *Dados dos variedades N y M , $f \in \mathcal{F}(N, M)$, un punto $p \in N$ y $T \in \Lambda_s^0(M)$ llamamos pullback de T a través de f en p al siguiente tensor:*

$$(f^*T)_p(v_1, \dots, v_s) = T_{f(p)}(df_p v_1, \dots, df_p v_s)$$

para todos $v_i \in T_p(N)$ con $i = 1, \dots, s$. Además, el pullback también es de tipo $(0, s)$.

A continuación, como dijimos al principio de la sección, nos dedicaremos a estudiar un tipo particular de tensor, las formas bilineales simétricas que nos permitirán llegar al concepto de producto escalar y con él al de medida en una variedad como veremos en el Capítulo 2.

Comencemos definiendo qué es una forma bilineal simétrica. La idea de tensor como generalización del concepto de 1-forma nos permite definir las formas bilineales simétricas de forma trivial como tensores de tipo $(0, 2)$ que son simétricos en el siguiente sentido:

Definición 1.6.8 (Forma bilineal simétrica) *Dados un \mathbb{R} -espacio vectorial V llamaremos formas bilineales simétricas a los tensores $b \in \Lambda_2^0$ cumpliendo que $b(u, v) = b(v, u)$ para todo $u, v \in V$.*

Nótese que en la definición hemos concretado que hablamos de un \mathbb{R} -espacio vectorial debido a nuestros propósitos. La definición podría darse para un cuerpo K en general de acuerdo con la idea previa de tensor.

Caractericemos ahora las formas bilineales simétricas.

Definición 1.6.9 (Formas bilineales definidas y semidefinidas) *Dado un espacio vectorial V y una forma bilineal simétrica b diremos que:*

- b es definida positiva (resp. negativa) si $b(u, u) > 0$ (resp. $b(u, u) < 0$) para todo $u \in V - \{0\}$. En ambos casos diremos que la forma bilineal simétrica es definida.

- b es semidefinida positiva (resp. negativa) si $b(u, u) \geq 0$ (resp. $b(u, u) \leq 0$) para todo $u \in V - \{0\}$. En ambos casos diremos que la forma bilineal simétrica es semidefinida.
- b es degenerada si fijado $u \in V$ con $b(u, v) = 0$ para todo $v \in V$ no se cumple necesariamente que u sea nulo. En caso contrario diremos que b es no degenerada.

Toda forma bilineal simétrica está relacionada con una forma cuadrática, esto es, una aplicación que toma elementos del espacio vectorial y los lleva al cuerpo:

Definición 1.6.10 (Forma cuadrática) *Dado un espacio vectorial b y una forma bilineal simétrica b llamaremos forma cuadrática asociada a b a la aplicación $q_b : V \rightarrow \mathbb{R}$ definida como $q_b(v) = b(v, v)$ para todo $v \in V$. Además, la forma cuadrática nos permite reconstruir la forma bilineal simétrica sin perder información a través de la identidad de polarización*

$$b(u, v) = \frac{1}{2}(q_b(u + v) - q_b(u) - q_b(v))$$

para todos $u, v \in V$.

Una vez caracterizadas las formas bilineales simétricas, nos centraremos en dos resultados principales. Para poder definir el producto escalar nos interesa saber cuándo la forma bilineal simétrica será no degenerada y, además, queremos definir el concepto de índice de una forma bilineal simétrica.

Comenzando con lo segundo, la definición de índice es natural a través de la caracterización de las formas bilineales simétricas teniendo en cuenta su carácter local, es decir, si tenemos un subespacio vectorial de V , sea W , la forma bilineal simétrica restringida al subespacio se comportará igual que la forma sin restringir heredando las propiedades de semidefinida o definida.

Definición 1.6.11 (Índice de una forma bilineal simétrica) *Dados un espacio vectorial V y una forma bilineal simétrica b , llamaremos índice de la forma bilineal simétrica al mayor entero ν que coincida con la dimensión de un subespacio vectorial $W \subset V$ donde se cumpla que $b|_W$ sea definida negativa.*

Nótese que en la definición hemos introducido la notación de la restricción especificando solo el espacio al que nos restringimos, aunque realmente la forma bilineal simétrica actúe sobre el producto de este espacio por sí mismo.

Para establecer condiciones que aseguren que la forma bilineal simétrica sea no degenerada será más fácil trabajar con su representación matricial.

Definición 1.6.12 (Representación matricial de una forma bilineal simétrica)

La representación matricial de una forma bilineal simétrica b sobre un espacio vectorial V de dimensión n con base $\{e_i\}_i$ vendrá dada por la matriz simétrica con componentes $(b_{ij}) = b(e_i, e_j)$ con $i, j = 1, \dots, n$.

Trabajando con esta definición tenemos el siguiente resultado sobre la degeneración de b :

Lema 1.6.3 *Una forma bilineal simétrica b sobre un espacio vectorial V es no degenerada si y solo si su representación matricial es invertible en cualquier base de V .*

En estas condiciones podemos definir el concepto de producto escalar.

Definición 1.6.13 (Producto escalar) *Un producto escalar es una forma bilineal simétrica g sobre un espacio vectorial V no degenerada.*

Nótese que hemos relajado la noción de producto escalar al que solemos estar acostumbrados en la mayor parte de las ramas de las matemáticas. El producto escalar usual (Euclídeo) es definido positivo. Sin embargo, aquí estamos permitiendo que el producto escalar pueda ser definido o semidefinido negativo lo que nos da una mayor riqueza y flexibilidad geométrica como veremos a continuación. Es bastante común en lengua castellana confundir estos conceptos debido al abuso del lenguaje. En inglés la diferenciación es más obvia ya que al producto escalar usual se le suele llamar *inner product* — en el sentido del Análisis Funcional — mientras que el producto escalar que usaremos en este texto se suele llamar *scalar product* — en el sentido de la Geometría semi-Riemanniana.

Veamos qué peculiaridades y propiedades nuevas surgen de esta definición. Esencialmente rompemos la noción clásica de ortogonalidad.

Definición 1.6.14 (Ortogonalidad) *Dado un espacio vectorial V equipado con un producto escalar g diremos que dos vectores $u, v \in V$ son ortogonales, y lo notaremos $u \perp v$, si $g(u, v) = 0$. Esta noción se extenderá naturalmente a conjuntos siendo dos conjuntos ortogonales si lo son entre sí todos sus vectores.*

Definición 1.6.15 (Conjunto ortogonal) *Dado un espacio vectorial V equipado con un producto escalar g llamaremos conjunto ortogonal de $W \subset V$ al conjunto:*

$$W^\perp = \{v \in V \mid v \text{ es ortogonal a } W\}$$

Detengámonos en las diferencias de estas definiciones con las usuales. En el sentido clásico, dos vectores eran ortogonales si forman entre sí un ángulo recto. Sin embargo, ahora esta definición se extiende pudiendo ser ortogonales vectores que no cumplan esto, vectores incluso colineales en un sentido clásico. Si consideramos (ver Figura 1.1) en \mathbb{R}^2 por ejemplo el producto escalar $g = -dx^1 \otimes dx^1 + dx^2 \otimes dx^2$ — producto escalar Lorentziano que será esencial más adelante — podemos ver cómo todos los vectores de las diagonales del plano real son ortogonales entre sí e incluso ortogonales a sí mismos; es más, todos los vectores simétricos a través de esas diagonales son ortogonales entre sí.

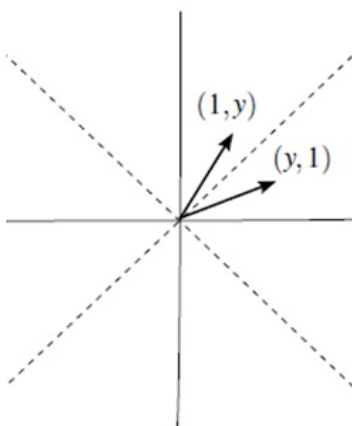


Figura 1.1: Ortogonalidad en \mathbb{R}^2 con el producto escalar Lorentziano.

En este mismo sentido, podemos ver cómo el conjunto ortogonal ya no tiene el significado del complemento ortogonal en el sentido clásico. Debido a la nueva noción de ortogonalidad, la idea de independencia y de poder representar cualquier elemento del espacio como combinación de vectores y sus ortogonales se pierde — basta ver el ejemplo anterior. Nótese que todo esto se debe a que el producto escalar restringido sí puede estar degenerado.

Veamos algunas características del conjunto ortogonal que nos permitan formalizar esta última idea:

Lema 1.6.4 *Dado un espacio vectorial V equipado con un producto escalar g , se cumple para $W \subset V$ que:*

- $\dim(W) + \dim(W^\perp) = \dim(V)$,
- $(W^\perp)^\perp = W$.

Definición 1.6.16 (Subespacio degenerado) *Dado un espacio vectorial V equipado con un producto escalar g , diremos que un subespacio $W \subset V$ es (no) degenerado si $g|_W$ es (no) degenerado.*

Para caracterizar la degeneración tenemos el siguiente resultado que formaliza el esbozo conceptual que hicimos antes.

Lema 1.6.5 *Dado un espacio vectorial V equipado con un producto escalar g , un subespacio $W \subset V$ es no degenerado si y solo si $V = W \oplus W^\perp$. En particular W es degenerado si y solo si su conjunto ortogonal también lo es.*

En este sentido, vemos que cuando un subespacio es no degenerado recuperamos la intuición clásica. Es más, de los resultados anteriores es directo, por la definición de producto escalar, que si equipamos con un producto escalar un espacio vectorial el conjunto ortogonal de este tiene que reducirse al cero — ya que el producto escalar en el espacio vectorial es no degenerado aunque pueda degenerarse como ocurría con el ejemplo del plano real con el producto Lorentziano.

Para acabar la sección introducimos los conceptos de norma — en paralelo con la idea clásica — y de signatura, que será esencial en la geometría semi-Riemanniana del capítulo siguiente.

Definición 1.6.17 (Norma) *Dado un espacio vectorial V equipado con un producto escalar g , llamamos norma de un vector $u \in V$ a $\|u\| = \sqrt{|g(u, u)|}$. Diremos que un vector es unitario o que está normalizado si su norma es 1.*

Recordemos aquí que llamaremos base ortonormal, al igual que en el sentido usual, a una base de un espacio vectorial constituida por vectores unitarios y ortogonales entre sí. Además, cualquier espacio vectorial posee una base ortonormal — quitando el caso trivial en que se reduzca al cero. Esto nos permite introducir la signatura:

Definición 1.6.18 (Signatura) *Dado un espacio vectorial V con base ortonormal $\{e_i\}_i$ equipado con un producto escalar g , llamaremos signatura al conjunto $(\epsilon_1, \dots, \epsilon_n)$ donde $\epsilon_i = g(e_i, e_i) = \pm 1$ con $i = 1, \dots, n$.*

Finalicemos con el siguiente resultado que nos permite relacionar la signatura con el índice. Nótese que por comodidad ordenaremos la signatura dando preferencia al signo negativo.

Lema 1.6.6 *Dado un espacio vectorial V con base ortonormal $\{e_i\}_{i=1, \dots, n}$ equipado con un producto escalar g , si llamamos k al número de signos negativos en la signatura entonces $k = \nu$ con ν el índice del producto escalar en V .*

En consecuencia, el resultado anterior nos permite concluir que la signatura y el índice son independientes de la base que tomemos en V y, por tanto, solo dependen del producto escalar.

Capítulo 2

Elementos principales de la teoría de variedades semi-Riemannianas

El segundo peldaño en nuestra escalera conceptual hacia la idea de espacio-tiempo serán las variedades semi-Riemannianas. Hasta ahora, a través del concepto de variedad diferenciable hemos generalizado la idea clásica de espacio, sin embargo, en ningún momento se han mencionado longitudes, ángulos o curvaturas; en definitiva, no se ha especificado cómo se *mide* dentro de la variedad. Así, nuestro objetivo central será generalizar la noción clásica de estos objetos. Para ello, dotaremos a las variedades de una estructura más compleja a través del tensor métrico — generalización de la primera forma fundamental en superficies — que formalizará la idea de medida para las variedades diferenciables dando lugar a las variedades semi-Riemannianas. Las demostraciones y desarrollos de este capítulo están basados esencialmente en el contenido de [18] que ha sido apoyado con otros textos como [2] y [16].

2.1. Variedades semi-Riemannianas

La estrategia para generar nuestra nueva estructura, esto es, las variedades semi-Riemannianas, será, una vez más, fijarnos en el caso Euclídeo y generalizar las ideas. En este sentido, es bien conocido que la noción de medida en \mathbb{R}^n viene generada por el producto escalar (en el sentido usual) a través del cual podemos establecer formalmente cómo medir distancias, ángulos, etc. Aumentando un poco la complejidad, sabemos que la noción de medida en una superficie también proviene del producto escalar, pero, en este caso, a través de la primera forma fundamental, es decir, a través del producto escalar inducido por ella en el espacio tangente a la superficie. Esta idea será la que generalicemos para las variedades, puesto que, como vimos en el capítulo anterior, una variedad, hablando burdamente, no es más que la generalización de una superficie a n dimensiones, siendo natural entonces generalizar los conceptos ya estudiados en

superficies.

Si recordamos la idea de primera forma fundamental, podemos determinar de forma directa que el formalismo tensorial desarrollado al final de la sección anterior encaja perfectamente con ella, es más, la primera forma fundamental es un tensor $(0, 2)$ — 2-covariante — y simétrico, es decir, una forma bilineal simétrica — recuérdese que abusamos del lenguaje y aquí sería más correcto hablar de “campo” tensorial. En este sentido, el objeto que generalice la primera forma fundamental será un tensor 2-covariante simétrico que induzca un producto escalar en el espacio tangente a la variedad, pero ahora en el sentido de la Definición 1.6.13.

Definición 2.1.1 (Tensor métrico) *Llamaremos tensor métrico o métrica sobre una variedad diferenciable N a un campo tensorial $g \in \Lambda_2^0$ simétrico, no degenerado y de índice constante.*

Podemos ver, como veníamos diciendo, que, en el sentido de los campos tensoriales, g tomará campos vectoriales $X, Y \in \mathcal{X}(N)$ dándonos una función diferenciable que notaremos $g(X, Y) \in \mathcal{F}(N)$, y en particular, en cada punto $p \in N$ generará un tensor g_p en el sentido del Lema 1.6.1 que, en este caso, será — por serlo g — de tipo $(0, 2)$, simétrico y no degenerado, es decir, un producto escalar sobre el espacio tangente. Nótese además, que debido a la relajación de la definición de producto escalar exigimos que el índice de g y, en consecuencia, de g_p sea el mismo independientemente del punto de la variedad por uniformidad — el porqué de esta idea en nuestro desarrollo se intuirá mejor más adelante cuando veamos cómo podemos interpretar de un modo más físico el índice.

Trabajemos un poco más con la noción de tensor métrico antes de introducir la noción de variedad semi-Riemanniana. Tal y como hemos adelantado en el párrafo anterior escribiremos la actuación de g sobre campos vectoriales $X, Y \in \mathcal{X}(N)$ como $g(X, Y)$ y la actuación de g_p sobre vectores $u, v \in T_p(N)$ como $g_p(u, v)$. En general, no distinguiremos entre la métrica y el producto escalar inducido y notaremos ambos por g .

En el capítulo anterior, vimos cómo representar en general un tensor por sus componentes. Fijada una carta $(U, \varphi = (x^1, \dots, x^n))$ en N las componentes de g tal y como vimos en la Definición 1.6.4 son $g_{ij} = g(\partial_i, \partial_j)$ con $i, j = 1, \dots, n$ en todo el desarrollo que vamos a realizar. Así, a través del Lema 1.6.2, podemos expresar g como

$$g = g_{ij} dx^i \otimes dx^j$$

donde dx^k denotan, como venimos haciendo en todo el texto, las 1-formas coordenadas asociadas a la carta tomada.

Además, fijada una base, podemos representar g de forma matricial tal y como vimos en la Definición 1.6.12, siendo sus entradas las componentes anteriores $g_{ij} = g(\partial_i, \partial_j)$

— es usual en la literatura llamar a g_{ij} *coeficientes métricos*. En este sentido, gracias al Lema 1.4.2 de representación de campos vectoriales podemos representar la actuación de g sobre campos $X = X^i \partial_i, Y = Y^j \partial_j \in \mathcal{X}(N)$ como:

$$g(X, Y) = g(X^i \partial_i, Y^j \partial_j) = X^i Y^j g(\partial_i, \partial_j) = g_{ij} X^i Y^j$$

Nótese que todo lo anterior es válido tanto para g tensor métrico como para g_p producto escalar inducido. En esta última expresión simplemente introduciríamos vectores del espacio tangente usando el Teorema de la Base y el resultado sería análogo — pudiéndose interpretar, en este caso, esta última expresión como un producto de matrices y vectores en el sentido usual.

Siguiendo con el recorrido de resultados de la última sección del Capítulo 1 en el caso concreto del tensor métrico, al igual que a toda forma bilineal simétrica al producto escalar inducido le podemos asociar una forma cuadrática como en la Definición 1.6.10.

Definición 2.1.2 (Elemento de línea) *Dada una variedad N con un tensor métrico g llamaremos elemento de línea $q_g := ds^2$ a la forma cuadrática asociada al producto escalar inducido por g .*

Nótese que el nombre de elemento de línea se debe a que este objeto en cierto modo representa un infinitésimo de longitud tal y como se puede notar en su representación fijada una carta $(U, \varphi = (x^1, \dots, x^n))$:

$$ds^2 = g_{ij} dx^i dx^j.$$

Para acabar con nuestro estudio del tensor métrico, vemos que, por ser no degenerado, tenemos a través del Lema 1.6.3 que:

Corolario 2.1.1 *Dada una variedad N con un tensor métrico g y dada una carta (U, φ) la representación matricial de g es invertible en todo punto de U , es decir, tiene una inversa que notaremos $g^{-1} \equiv (g^{ij})$.*

En este punto, una vez tenemos bien caracterizado el tensor métrico podemos introducir el concepto clave de todo el documento: la variedad semi-Riemanniana.

Definición 2.1.3 (Variedad semi-Riemanniana) *Una variedad semi-Riemanniana es un par (N, g) donde N es una variedad diferenciable y g un tensor métrico sobre N*

De aquí en adelante abreviaremos el par (N, g) y nos referiremos a una variedad semi-Riemanniana como N , omitiendo la métrica. Ha de tenerse en cuenta sin embargo, que la métrica es esencial en la variedad semi-Riemanniana ya que es el objeto

que eleva la estructura de variedad diferenciable como venimos diciendo. Así pues, a partir de una misma variedad diferenciable se pueden generar diferentes variedades semi-Riemannianas al equiparla con diferentes métricas. En la práctica, sobre todo en el Capítulo 4 cuando introduzcamos las ecuaciones de campo, estaremos interesados en variedades semi-Riemannianas que sean isométricas, esto es, que haya una función entre ambas que no las deforme — difeomorfismo — y que preserve el tensor métrico en el sentido que presentamos a continuación. El motivo de esta tecnicidad es que, aunque no sea determinante, cuando resolvamos las ecuaciones de campo de Einstein las soluciones no serán únicas, en el sentido de que habrá varias variedades semi-Riemannianas “diferentes” con la misma geometría, es decir, el espacio-tiempo — cuya definición precisaremos en el Capítulo 3 — no es único. La solución a este problema es tomar como espacio-tiempo — ya que físicamente sí es único — la clase de equivalencia a través de la relación inducida por la isometría, esto es, “colapsamos” todos los espacios-tiempo posibles a uno único que represente la geometría general.

Definición 2.1.4 (Isometría) *Diremos que dos variedades semi-Riemannianas (N, g) y (M, h) son isométricas si existe una isometría entre ellas, esto es, si existe un difeomorfismo $f : N \rightarrow M$ tal que el pullback del tensor métrico en M sea el tensor métrico de N : $f^*h = g$.*

A continuación, caracterizamos dos tipos de variedades semi-Riemannianas en función de su índice — nótese que, una vez más, abusamos del lenguaje y llamamos índice de una variedad semi-Riemanniana al de su tensor métrico que, como sabemos, por definición es constante en toda la variedad.

Definición 2.1.5 (Variedades Riemannianas y Lorentzianas) *Sea N^n una variedad semi-Riemanniana. Diremos que:*

- N^n es Riemanniana si su índice $\nu = 0$.
- N^n es Lorentziana si $n > 1$ y tiene índice $\nu = 1$.

Así pues, las variedades Riemannianas son aquellas que heredan la geometría Euclídea generalizada puesto que al tener índice nulo, el producto escalar inducido será definido positivo, como ocurre usualmente. En el documento nos centraremos en las variedades Lorentzianas, pues el hecho de que el índice sea uno será nuestra manera de introducir el tiempo como una dimensión más, generando así la idea primigenia de espacio-tiempo. En este sentido, la dimensión de la variedad a la que vaya asociado el signo negativo de la signatura será por convenio el tiempo. Aunque en el Capítulo 3 estudiaremos todo esto con más detalle, puede entenderse ya por qué el índice debe ser constante, ya que de no serlo, en cierta manera estaríamos diciendo que, en función del punto de la variedad donde nos encontremos, habría más o menos dimensiones con carácter temporal, hecho que — al menos actualmente — no es cierto.

Para acabar la sección, introducimos dos variedades semi-Riemannianas que usaremos a lo largo del documento. En primer lugar presentamos el espacio semi-Euclídeo.

Definición 2.1.6 (Espacio semi-Euclídeo) *Llamaremos espacio semi-Euclídeo a la variedad semi-Riemanniana $\mathbb{R}_\nu^n = (\mathbb{R}^n, g)$ con*

$$g = -\delta_{ij}dx^i \otimes dx^j + \delta_{kj}dx^k \otimes dx^j$$

con $j = 1, \dots, n$, $i = 1, \dots, \nu$, $k = \nu + 1, \dots, n$ y $\nu \in [1, n]$. A la métrica anterior la denominaremos métrica Lorentziana generalizada.

Es claro que el espacio semi-Euclídeo no es más que el espacio Euclídeo equipado con el producto escalar usual, solo que cambiamos los signos positivos por negativos en las primera ν componentes.

Finalmente, retomemos la idea de variedad producto introducida en el Lema 1.1.1. En el Capítulo 4 veremos que en el modelo cosmológico actual, debido a varias razones experimentales, el espacio-tiempo se modela como una variedad semi-Riemanniana cuya variedad diferenciable es una variedad producto y cuyo tensor métrico precisamos en la siguiente definición:

Definición 2.1.7 (Producto deformado) *Dadas dos variedades semi-Riemannianas (B, g) y (F, h) , y una función $a \in \mathcal{F}(B)$ positiva, llamamos producto deformado a la variedad semi-Riemanniana $B \times_a F = (B \times F, g_a)$ con*

$$g_a = \pi^*(g) + (a \circ \pi)^2 \sigma^*(h),$$

donde π es la proyección de la variedad producto a B y σ la proyección de la variedad producto a F .

Hemos decidido usar el nombre de producto deformado al traducir directamente del inglés *warped product*. Sin embargo, en la literatura es común llamar a la estructura anterior *producto alabeado* o *producto warped*. Además, aunque pueda resultar extraño, usaremos usualmente la letra a para denotar la función de los productos deformados como es usual en la literatura relativista (ver Capítulo 4).

En la definición anterior se suele llamar a B *base* y a F *fibra* puesto que la idea detrás del producto deformado es “pegar” en cada punto de la base una copia de la fibra que se verá deformada a través de la función a . Así pues, la geometría en cada copia de la fibra se verá alterada a través de dicha deformación de acuerdo con el factor $(a \circ \pi)^2$ de la métrica. Aquellos lectores que estén familiarizados con ideas cosmológicas como la expansión del espacio-tiempo, podrán intuir ya que nuestra idea en el Capítulo 4 será coger la línea del tiempo como base y el espacio tridimensional como fibra, teniéndose una rodaja de espacio en cada instante de tiempo que se verá expandida o contraída a través de la función a que dependerá del tiempo.

2.2. La conexión de Levi-Civita

Una vez definidas las variedades semi-Riemannianas nuestro objetivo es construir en ellas objetos que nos caractericen la curvatura en un cierto sentido que ya estableceremos más adelante. Para hablar de curvatura, necesitamos generalizar la idea de derivada en una variedad semi-Riemanniana en analogía con el caso en superficies. De este modo, como vimos antes, la geometría de una variedad viene dada esencialmente por su espacio tangente, que irá cambiando punto a punto. Necesitaremos, por tanto, un objeto que relacione las geometrías locales de diferentes puntos de la variedad dándonos una idea de cómo cambia el espacio tangente, o lo que es lo mismo, ver cómo varían — idea de derivada — los campos vectoriales tangentes en direcciones determinadas por otros campos tangentes — transportar vectores de punto a punto. El objeto que englobará estas ideas será la *conexión* que generalizará la idea de derivada covariante en superficies.

Definición 2.2.1 (Conexión) *Una conexión D sobre una variedad N es una aplicación $D : \mathcal{X}(N) \times \mathcal{X}(N) \rightarrow \mathcal{X}(N)$ tal que para $X, Y \in \mathcal{X}(N)$ cumple las siguientes propiedades:*

1. $D_Y(X)$ es $\mathcal{F}(N)$ -lineal en Y .
2. $D_Y(X)$ es \mathbb{R} -lineal en X .
3. $D_Y(fX) = (Yf)X + fD_Y(X)$ para toda $f \in \mathcal{F}(N)$.

En particular a cada campo $D_Y(X) \in \mathcal{X}(M)$ lo llamaremos derivada covariante de X con respecto a Y para la conexión D .

Nótese que las conexiones no son tensores debido a la propiedad 3 — no son $\mathcal{F}(N)$ -lineales en X .

De todas las conexiones sobre una variedad nos interesará una que cumplirá unas propiedades particulares: *la conexión de Levi-Civita*. Llegaremos a ella a través del Teorema Fundamental de la Geometría semi-Riemanniana. Para probar este, necesitamos el siguiente resultado, análogo al teorema de representación de Riesz del Análisis Funcional pero en el ámbito de la geometría semi-Riemanniana, estableciendo que el espacio de las 1-formas es isomorfo al de los campos vectoriales.

Teorema 2.2.1 *Dados una variedad semi-Riemanniana N y $X \in \mathcal{X}(N)$, si llamamos X^* a la 1-forma que cumple que $X^* = g(X, Y)$ para todo $Y \in \mathcal{X}(N)$, entonces la aplicación f que lleva X a X^* es un isomorfismo $\mathcal{F}(N)$ -lineal entre $\mathcal{X}(N)$ y $\mathcal{X}(N)^*$. Los campos vectoriales y 1-formas relacionadas por este isomorfismo se dirán métricamente equivalentes.*

Demostración: La $\mathcal{F}(N)$ -linealidad es directa pues g es $\mathcal{F}(N)$ -lineal en sus dos componentes y consecuentemente tenemos que X^* es una 1-forma pues también es $\mathcal{F}(N)$ -lineal. Así, solo tenemos que probar que es isomorfismo. Para ello, tenemos que ver que dada una 1-forma $\theta \in \mathcal{X}(N)$ existe un único campo vectorial $X \in \mathcal{X}(N)$ tal que $\theta(X) = g(X, Y)$ para todo $Y \in \mathcal{X}(N)$.

Para probar la existencia, tomemos una carta $(U, \varphi = (x^1, \dots, x^n))$ de N que nos permite escribir localmente la 1-forma θ como $\theta = \theta_i dx^i$. Veamos que el campo vectorial $X = g^{ij} \theta_j \partial_j$ cumple lo que queremos. Lo vemos aplicando g sobre él y sobre ∂_k teniéndose el resultado inicial por $\mathcal{F}(N)$ -linealidad para todo $Y \in \mathcal{X}(N)$:

$$g(X, \partial_k) = g^{ij} \theta_j g(\partial_j, \partial_k) = \theta_j g^{ij} g_{jk} = \theta_j \delta_k^j = \theta_k = \theta(\partial_k).$$

La unicidad del campo anterior es directa pues, supongamos que existe otro campo Z que cumple lo mismo que X . Entonces, tendríamos que $g(X, Y) = g(Z, Y)$ para todo $Y \in \mathcal{X}(N)$ pero, por linealidad, esto es equivalente a que $g(X - Z, Y) = 0$. Basta recordar ahora la no degeneración de la métrica que implica que, en cada punto, si $g(X_p - Z_p, Y_p) = 0$ entonces $X_p = Z_p$ y, consecuentemente, $X = Z$ concluyendo la demostración. ■

En estas condiciones podemos introducir el Teorema Fundamental de la Geometría semi-Riemanniana y la conexión de Levi-Civita.

Teorema 2.2.2 (Teorema Fundamental de la Geometría semi-Riemanniana)

Dada una variedad semi-Riemanniana N existe una única conexión ∇ tal que para cada $X, Y, Z \in \mathcal{X}(N)$ cumple que:

1. *Es libre de torsión, esto es: $\nabla_Z X - \nabla_X Z - [Z, X] = 0$.*
2. *Es compatible con la métrica, esto es: $Zg(X, Y) = g(\nabla_Z X, Y) + g(X, \nabla_Z Y)$.*

Denominaremos conexión de Levi-Civita a dicha conexión ∇ que queda caracterizada por la siguiente fórmula llamada fórmula de Koszul:

$$2g(\nabla_Z X, Y) = Zg(X, Y) + Xg(Y, Z) - Yg(Z, X) - g(Z, [X, Y]) + g(X, [Y, Z]) + g(Y, [Z, X]).$$

Demostración: La existencia y la unicidad de la conexión de Levi-Civita se demuestran de forma análoga a la demostración del teorema previo. Llamando $F(X, Y, Z)$ al lado derecho de la fórmula de Koszul, fijamos $X, Z \in \mathcal{X}(N)$ y consideramos la función que lleva cada $Y \in \mathcal{X}(N)$ a $F(X, Y, Z)$, es directo ver que dicha función es $\mathcal{F}(N)$ -lineal y por tanto, como ocurría antes, es 1-forma. Aplicando ahora directamente el Teorema 2.2.1, deducimos que existe un único campo que denotaremos $\nabla_Z X$ que cumpla que $2g(\nabla_Z X, Y) = F(X, Y, Z)$ teniéndose la fórmula de Koszul y la existencia y unicidad de la conexión de Levi-Civita.

A continuación, tendremos que probar que el objeto antes caracterizado es conexión, es decir, tenemos que ver que se cumplen las propiedades de la Definición 2.2.1. Además, tendremos que ver que dicha conexión cumple las propiedades del Teorema. Todo ello lo podemos probar trabajando con la fórmula de Koszul. En toda la demostración consideramos $X, Y, Z \in \mathcal{X}(N)$ y $f \in \mathcal{F}(N)$. La $\mathcal{F}(N)$ -linealidad (la \mathbb{R} -linealidad es análoga) se puede probar del siguiente modo:

$$\begin{aligned}
2g(\nabla_{fZ}X, Y) &= fZg(X, Y) + Xg(Y, fZ) - Yg(fZ, X) - \\
&\quad - g(fZ, [X, Y]) + g(X, [Y, fZ]) + g(Y, [fZ, X]) = \\
&= fZg(X, Y) + (Xf)g(Y, Z) + fXg(Y, Z) - (Yf)g(Z, X) - fYg(Z, X) - \\
&\quad - fg(Z, [X, Y]) + g(X, (Yf)Z + f[Y, Z]) + g(Y, -(Xf)Z + f[Z, X]) = \\
&= 2g(f\nabla_ZX, Y).
\end{aligned}$$

La tercera propiedad de las conexiones se prueba como sigue:

$$\begin{aligned}
2g(\nabla_ZfX, Y) &= Zg(fX, Y) + fXg(Y, Z) - Yg(Z, fX) - \\
&\quad - g(Z, [fX, Y]) + g(fX, [Y, Z]) + g(Y, [Z, fX]) = \\
&= (Zf)g(X, Y) + fZg(X, Y) + fXg(Y, Z) - (Yf)g(Z, X) - fYg(Z, X) - \\
&\quad - g(Z, (Xf)Y + f[X, Y]) + fg(X, [Y, Z]) + g(Y, (Zf)X + f[Z, X]) = \\
&= 2g(f\nabla_ZX + (Zf)X, Y).
\end{aligned}$$

Finalmente, una vez que hemos probado que es conexión, veamos que cumple las condiciones del teorema. Comencemos viendo que es libre de torsión:

$$\begin{aligned}
2g(\nabla_ZX - \nabla_XZ, Y) &= Zg(X, Y) + Xg(Y, Z) - Yg(Z, X) - \\
&\quad - Xg(Z, Y) - Zg(Y, X) + Yg(X, Z) - \\
&\quad - g(Z, [X, Y]) + g(X, [Y, Z]) + g(Y, [Z, X]) + \\
&\quad + g(X, [Z, Y]) - g(Z, [Y, X]) - g(Y, [X, Z]) = \\
&= 2g([Z, X], Y).
\end{aligned}$$

Acabemos la demostración viendo la compatibilidad con la métrica:

$$\begin{aligned}
2(g(\nabla_ZX, Y) + g(X, \nabla_ZY)) &= Zg(X, Y) + Xg(Y, Z) - Yg(Z, X) + \\
&\quad + Zg(Y, X) + Yg(X, Z) - Xg(Z, Y) - \\
&\quad - g(Z, [X, Y]) + g(X, [Y, Z]) + g(Y, [Z, X]) - \\
&\quad - g(Z, [Y, X]) + g(Y, [X, Z]) + g(X, [Z, Y]) = \\
&= 2Zg(X, Y).
\end{aligned}$$

Concluyendo así la demostración. ■

A partir de este punto, todo nuestro desarrollo se enfocará en la conexión de Levi-Civita y, siempre que consideremos una conexión, será esta cumpliendo las propiedades del Teorema Fundamental.

En este punto, tal y como adelantábamos al principio de la sección, hemos generalizado la noción de derivada covariante. De hecho, podemos introducir la derivada covariante de un tensor que será esencial en las secciones posteriores.

Definición 2.2.2 (Derivada Covariante tensorial) *Dados un tensor $T \in \Lambda_s^1(N)$ y un campo vectorial $X \in \mathcal{X}(N)$, definimos la derivada covariante de T en la dirección de X como el tensor $\nabla_X T \in \Lambda_s^1$ caracterizado por:*

$$(\nabla_X T)(V_1, \dots, V_s) = \nabla_X(T(V_1, \dots, V_s)) - \sum_i T(V_1, \dots, V_{i-1}, \nabla_X V_i, V_{i+1}, \dots, V_s)$$

para cada $V_i \in \mathcal{X}(N)$ con $i = 1, \dots, s$.

Además, podemos crear un objeto que condense todas las derivadas covariantes.

Definición 2.2.3 (Diferencial Covariante) *Dado un tensor $T \in \Lambda_s^1(N)$ definimos la diferencial covariante de T como el tensor $\nabla T \in \Lambda_s^1$ caracterizado por:*

$$(\nabla T)(V_1, \dots, V_s, X) = (\nabla_X T)((V_1, \dots, V_s))$$

para todos $V_i, X \in \mathcal{X}(N)$ con $i = 1, \dots, s$.

Nótese que en la definición hemos hecho uso de la interpretación exótica que hicimos en el Capítulo 1 de los tensores tipo $(1, s)$, pero todo lo anterior es también válido para tensores puramente covariantes. En este sentido, la definición anterior no es más que una generalización de la diferencial que vimos en la Definición 1.5.6 — si $s = 0$ recuperamos su definición para funciones escalares. Recalamos además que, al igual que la conexión, la diferencial y la derivada covariante tienen carácter local; basta ver que solo dependen del valor en un punto de los campos. Así, fijada una carta podremos representar de nuevo estos tensores por sus componentes siguiendo por convenio la notación: $\nabla T \equiv T_{i_1, \dots, i_s; k}$ indicando con “;” que se está “derivando” respecto a esa componente.

Destacamos además que, a través de las definiciones previas, podemos darle un sentido más claro a la segunda condición del Teorema Fundamental de la geometría semi-Riemanniana, pues que la conexión es compatible con la métrica es completamente equivalente a que $\nabla g = 0$. Se suele decir que g es paralelo respecto a la conexión de Levi-Civita.

Para acabar la sección, a pesar de que es claro, a través del Teorema Fundamental, que no podríamos haber caracterizado la conexión de Levi-Civita sin introducir las variedades semi-Riemannianas, esto es, sin haber introducido la noción de métrica, podemos preguntar cómo determina la métrica la conexión o, siendo más claros, qué dependencia explícita con la métrica tiene la expresión de la conexión de Levi-Civita.

Para resolver esta cuestión vamos a expresar la conexión localmente por componentes a través de los llamados símbolos de Christoffel — nótese que estos objetos son componentes pero no de un tensor ya que, como dijimos antes, la conexión no es un tensor.

Definición 2.2.4 (Símbolos de Christoffel) *Dada una carta $(U, \varphi = (x^1, \dots, x^n))$ de una variedad N llamaremos símbolos de Christoffel a las funciones componentes $\Gamma_{ij}^k \in \mathcal{F}(U)$ de $\nabla_{\partial_i} \partial_j$ en la base de $\mathcal{X}(N)$:*

$$\nabla_{\partial_i} \partial_j = \Gamma_{ij}^k \partial_k.$$

Corolario 2.2.1 *Los símbolos de Christoffel son simétricos respecto a sus subíndices: $\Gamma_{ij}^k = \Gamma_{ji}^k$.*

Demostración: Es directo de la definición sin más que tener en cuenta que la conexión de Levi-Civita es libre de torsión y que el corchete de Lie de los elementos de la base de $\mathcal{X}(N)$ es nulo ■

Nótese cómo la simetría de los símbolos de Christoffel nos da una intuición de qué significa que la conexión sea libre de torsión — en particular por qué usamos la palabra torsión — ya que esta simetría nos indica que en nuestra variedad la variación de un campo de la base ∂_i en la dirección de cualquier otro campo de la base ∂_j es equivalente a la variación del segundo campo en la dirección del primero, dándonos cierta noción de uniformidad en el sentido de que la variedad no está torcida — pues si lo estuviera y hubiera torsión habría alguna dirección de la base privilegiada siendo las variaciones diferentes.

Veamos a continuación cómo están relacionados los símbolos de Christoffel con la métrica:

Proposición 2.2.1 *Dada una carta $(U, \varphi = (x^1, \dots, x^n))$ de una variedad N los símbolos de Christoffel vienen dados por:*

$$\Gamma_{ij}^k = \frac{1}{2} g^{km} \left(\frac{\partial g_{jm}}{\partial x^i} + \frac{\partial g_{im}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^m} \right).$$

Demostración: La demostración es por computación directa; primero a través de la fórmula de Koszul:

$$2g(\nabla_{\partial_i} \partial_j, \partial_m) = \frac{\partial g_{jm}}{\partial x^i} + \frac{\partial g_{im}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^m}.$$

Aplicando ahora la definición de los símbolos al primer término de la fórmula de Koszul:

$$2g(\nabla_{\partial_i} \partial_j, \partial_m) = 2g(\Gamma_{ij}^l \partial_l, \partial_m) = 2\Gamma_{ij}^l g_{lm}.$$

Finalmente, sin más que igualar ambas expresiones y multiplicar por la inversa del tensor métrico g^{km} tenemos lo que queríamos. ■

Adelantamos, en relación a lo que comentamos al inicio de esta sección, la gran importancia que tendrán los símbolos de Christoffel a la hora de hablar de curvatura en una variedad. Del lema anterior podemos ver ya la directa relación que tienen estos objetos con la variación de la métrica que, como decíamos, estará intrínsecamente relacionado con la noción de curvatura, ya que si, por ejemplo, la métrica — la geometría — no varía de punto a punto, los símbolos de Christoffel son nulos, pudiéndose intuir que la variedad será en cierto sentido llana — nótese por ejemplo que es exactamente esto lo que ocurre para \mathbb{R}^n que sabemos de antemano que es llano.

Finalicemos pues la sección viendo explícitamente cómo expresamos en general la conexión en componentes determinando, como queríamos, la forma en que la métrica caracteriza la conexión.

Proposición 2.2.2 *La métrica caracteriza la conexión de Levi-Civita a través de los símbolos de Christoffel, esto es, dada una carta $(U, \varphi = (x^1, \dots, x^n))$ de N y un campo vectorial $V = V^i \partial_i \in \mathcal{X}(N)$ se tiene:*

$$\nabla_{\partial_j} V = \left(\frac{\partial V^k}{\partial x^j} + \Gamma_{ij}^k V^i \right) \partial_k.$$

Demostración: Es directo que

$$\nabla_{\partial_j} (V^i \partial_i) = (\partial_j V^i) \partial_i + V^i \Gamma_{ij}^k \partial_k = \left(\frac{\partial V^k}{\partial x^j} + \Gamma_{ij}^k V^i \right) \partial_k,$$

donde simplemente hemos aplicado la tercera propiedad de las conexiones. ■

2.3. Transporte paralelo. Geodésicas y aplicación exponencial

Como venimos diciendo, nuestro propósito esencial dentro de las variedades semi-Riemannianas es llegar a la noción de curvatura. La manera de introducir la curvatura, una vez vista la conexión de Levi-Civita, no es única, pues sus interpretaciones son varias. En nuestro desarrollo usaremos la idea de *transporte paralelo* que, tal y como su propio nombre indica, será un objeto que nos permita transportar vectores de forma paralela a la variedad en un sentido que ya precisaremos. El motivo principal por el que elegimos esta formulación y no otra es que nos va a permitir generalizar la idea clásica de línea recta a través del concepto de *geodésica* que será clave en el formalismo de la Teoría General de la Relatividad.

Para empezar a formalizar estas ideas, lo primero será determinar cómo varían los vectores, o mejor dicho, campos de vectores a lo largo de una curva. Dada una curva α dentro de una variedad N — tal y como introducimos en el Capítulo 1 — denotaremos el conjunto de todos los campos vectoriales tangentes a α como $\mathcal{X}(\alpha)$.

Proposición 2.3.1 *Sea α una curva en una variedad semi-Riemanniana N . Entonces existe una única función de $\mathcal{X}(\alpha)$ en $\mathcal{X}(\alpha)$ que lleva cada V en $V' := \frac{DV}{dt}$ llamada derivada covariante inducida y que cumple las siguientes propiedades para todos $V, W \in \mathcal{X}(\alpha)$:*

1. $(aV + bW)' = aV' + bW'$ para todo $a, b \in \mathbb{R}$.
2. $(fV)' = \frac{df}{dt}V + fV'$ para toda $f \in \mathcal{F}(N)$.
3. $(V|_\alpha)' = \nabla_{\alpha'(t)}V$ para todo $t \in I$.
4. $\frac{d}{dt}g(V, W) = g(V', W) + g(V, W')$.

Demostración: Comencemos con la unicidad. Para ello consideremos una aplicación que cumpla 1, 2 y 3, y una carta $(U, \varphi = (x^1, \dots, x^n))$ o conjunto de cartas (si es necesario) tal que su dominio contenga a α . Podemos ver cómo V' es único, ya que queda completamente determinada por la conexión de Levi-Civita ∇ , pues, a través de las propiedades y expresando V en una base $V = V^i \partial_i := V^i \partial_i$, podemos escribir:

$$V' = \frac{dV^i}{dt} \partial_i|_\alpha + V^i (\partial_i|_\alpha)' = \frac{dV^i}{dt} \partial_i + V^i \nabla_{\alpha'} \partial_i. \quad (2.1)$$

Para la existencia, no tenemos más que definir localmente en subintervalos J del dominio de α el campo V' con la expresión de arriba — recuérdese que dichos subintervalos deben quedar dentro de las cartas. Probar que cada uno de estos campos V' cumple las propiedades de la proposición es una computación directa. Finalmente, tendremos el campo V' deseado ya que, por la unicidad, cada uno de estos campos locales será el único V' que buscamos. ■

Nótese que los objetos previos no son más que generalizaciones de los conceptos de velocidad y aceleración, aunque en cualquier dirección, no solo en aquella en que se mueve la curva. Es trivial ver que si en la proposición anterior tomamos concretamente $V = \alpha'$ — la velocidad de la curva — entonces $V' = \alpha''$ — su aceleración.

A través de los símbolos de Christoffel podemos encontrar una expresión coordinada para la derivada covariante inducida:

Corolario 2.3.1 *En las condiciones de la proposición anterior, dada una carta $(U, \varphi = (x^1, \dots, x^n))$ podemos expresar V' como:*

$$V' = \left(\frac{dV^k}{dt} + \Gamma_{ij}^k \frac{d(x^i \circ \alpha)}{dt} V^j \right) \partial_k.$$

Demostración: La demostración es directa sin más que tener en cuenta la Ecuación (2.1) de la demostración previa. ■

En el mismo sentido que antes, el corolario anterior nos proporciona una expresión generalizada para la aceleración en cualquier variedad. Introducimos a continuación la noción de paralelismo — generalización de aceleración nula — que nos permitirá definir el transporte paralelo.

Definición 2.3.1 (Paralelismo) *Dada una curva α en una variedad N diremos que un campo $V \in \mathcal{X}(\alpha)$ es paralelo si $V' = 0$.*

Nótese, de la definición previa, que la condición de paralelismo equivale a igualar la expresión del Corolario 2.3.1 a cero, y por tanto, a resolver un sistema de n ecuaciones diferenciales — ya que los ∂_i son base. Así pues, si fijamos unas condiciones iniciales, esto es, si fijamos el valor de V en un punto, gracias a los resultados de existencia y unicidad de los problemas de valores iniciales, podremos asegurar que existe un único campo paralelo con esa condición inicial. Es en este contexto en el que podemos introducir la noción de transporte paralelo:

Definición 2.3.2 (Transporte paralelo) *Dada una curva α en una variedad N , consideremos el campo paralelo $V \in \mathcal{X}(\alpha)$ que cumple que $V(t_0) = v$ para $t_0 \in I$ y $v \in T_{\alpha'(t_0)}(N)$ fijos. Llamaremos transporte paralelo a lo largo de α de p a q a la aplicación $P_{pq}^\alpha : T_p(N) \rightarrow T_q(N)$ que lleva cada v en $V(t)$, donde hemos llamado $p = \alpha(t_0)$ y $q = \alpha(t_1)$ con $t_1 \in I$. En particular, debido a la unicidad P_{pq}^α , es difeomorfismo con inversa P_{qp}^α .*

Antes de cerrar el círculo retomando con la idea expuesta al principio de la sección, introduzcamos la noción de geodésica que clarificará aún más las ideas.

Definición 2.3.3 (Geodésica) *Una geodésica es una curva ρ en una variedad N tal que su campo de velocidades ρ' es paralelo.*

Vemos cómo hemos llevado de forma natural el concepto de geodésicas que tenemos en superficies a variedades. Una geodésica será la generalización de una línea recta en cualquier variedad: aquellas curvas que tengan aceleración nula — como les ocurría a las rectas sobre el espacio Euclídeo — y que localmente son rectas en el sentido usual.

Nótese además que el paralelismo hace que el comportamiento de la geodésica sea muy uniforme. Esta, hablando burdamente, no puede decelerar y parar, puesto que si su $\rho' \neq 0$ en un instante ya luego no podrá anularse nunca (retomaremos esta noción en el Capítulo 3). Téngase en cuenta sin embargo, que existen curvas que no son geodésicas

pero que al reparametrizarlas — en el sentido usual — pueden convertirse en geodésicas; a estas curvas las llamaremos *pregeodésicas*.

Trabajemos un poco más con las geodésicas que serán esenciales tanto en el desarrollo del Capítulo 4 con la Teoría de la Relatividad, como en el Capítulo 5 con la teoría de Causalidad.

Comencemos viendo que el resultado anterior sobre la expresión de la derivada covariante inducida se puede aplicar directamente al campo de velocidades de la geodésica, obteniéndose una ecuación para ellas.

Corolario 2.3.2 *Dada una curva ρ en una variedad N y dada una carta $(U, \varphi = (x^1, \dots, x^n))$ que la contenga, ρ es geodésica si y solo si sus funciones coordenadas $x^i \circ \rho$ cumplen que*

$$\frac{(x^k \circ \rho)^2}{dt} + \Gamma_{ij}^k \frac{(x^i \circ \rho)}{dt} \frac{(x^j \circ \rho)}{dt} = 0,$$

con $k = 1, \dots, n$.

En el mismo sentido que ocurría antes en general para los campos, resultados de la teoría de ecuaciones diferenciables nos aseguran la existencia y unicidad de geodésicas fijada una condición inicial (que tengan una velocidad fija en el punto en que comiencen)¹. Así, presentamos los siguientes lemas sin demostración pues se pueden encontrar en cualquier texto de ecuaciones diferenciales o de geometría a través ellas [18]:

Lema 2.3.1 *Dados una variedad N y un vector $v \in T_p(N)$, existe un intervalo I en torno al 0 y una única geodésica ρ con $\rho(0) = p$ y $\rho'(0) = v$. Llamaremos a esta geodésica, geodésica que empieza en p con velocidad inicial v .*

Lema 2.3.2 *Dadas dos geodésicas ρ y γ en una variedad N , si existe un número real t tal que $\rho(t) = \gamma(t)$ entonces $\rho = \gamma$.*

Estos dos lemas nos permiten enunciar el siguiente enunciado, que será esencial conceptualmente para la teoría de Causalidad.

Proposición 2.3.2 *Dados una variedad N , un punto $p \in N$ y un vector $v \in T(N)$, existe una única geodésica que notaremos ρ_v que cumpla las siguientes propiedades:*

1. *La velocidad inicial de la geodésica es v , es decir, $\rho'_v(0) = v$ (es claro que empieza en p).*
2. *Es maximal o inextendible, es decir, su dominio de definición I_v es el máximo posible de modo que, si existe otra geodésica γ definida en un intervalo J , entonces $J \subset I_v$ y $\gamma = \rho_v|_J$.*

¹Nótese el gran paralelismo con los resultados que presentamos para curvas integrales en el Capítulo 1. De hecho, existe una correspondencia 1-1 entre curvas integrales y geodésicas [16].

Demostración: La demostración es bastante directa, pues si tomamos toda la colección de geodésicas que por el Lema 2.3.1 tienen velocidad inicial v , sabemos que dichas geodésicas coinciden en la intersección de sus dominios por el Lema 2.3.2. Así, es trivial que si consideramos toda la colección de geodésicas sobre el dominio que surge de unir todos los dominios particulares, estas constituyen una única geodésica con las propiedades del enunciado. ■

Este resultado nos lleva a la siguiente noción que será crucial cuando hablemos de singularidades en los últimos capítulos ya que introduce la idea de que una geodésica este incompleta o, lo que es lo mismo, que a la variedad le “falten” puntos en el siguiente sentido:

Definición 2.3.4 (Variedad geodésicamente completa) *Diremos que una variedad es geodésicamente completa o, simplemente, completa, si todas sus geodésicas maximales están definidas en todo \mathbb{R} .*

Para acabar la sección, introducimos el concepto de *aplicación exponencial*. Este concepto nos permite determinar de cierta manera “cómo de buena” es la aproximación que hacemos de una variedad a través de su espacio tangente localmente — tal y como vimos en el Capítulo 1 en la línea del isomorfismo lineal canónico.

Definición 2.3.5 (Aplicación exponencial) *Dado un punto p de una variedad N , si consideramos el conjunto \mathcal{D}_p de vectores de $T_p(N)$ tal que la geodésica maximal ρ_v esté definida al menos en el intervalo $[0, 1]$, definimos la aplicación exponencial de N en p como $\exp_p : \mathcal{D}_p \rightarrow N$ tal que $\exp_p(v) = \rho_v(1)$ para todo $v \in \mathcal{D}_p$.*

La idea detrás de la aplicación exponencial es que “envuelve” el espacio tangente sobre la variedad, determinando cómo se adhiere este. Hablando burdamente es como si determináramos cómo de bien podemos envolver algo curvo con un trozo de papel plano; lógicamente cuanto más pequeño, más local, mejor es la identificación. Podemos formalizar esto a través de la siguiente proposición.

Previamente, necesitaremos incluir este último resultado de ecuaciones diferenciales. Recuérdese que si un sistema de ecuaciones diferenciales de segundo orden está dado en términos de funciones diferenciables, entonces sus soluciones también lo son y no solo en el parámetro, también en los valores iniciales y sus derivadas. Esto nos lleva al siguiente resultado en el caso de las geodésicas:

Lema 2.3.3 *Dados una variedad N y un vector de su fibrado tangente $v \in T(N)$, existe un entorno \mathcal{N} de v en $T(N)$ y un intervalo I alrededor del 0 tal que la función que lleva cada par $(u, t) \in T(N)$ a $\rho_u(t) \in N$ está bien definida y es diferenciable.*

Proposición 2.3.3 *Para cada punto p en una variedad N existe un entorno \bar{U} del 0 en $T_p(N)$ en el que la aplicación exponencial \exp_p es un difeomorfismo sobreyectivo sobre un entorno de p en N .*

Demostración: Es directo, del lema previo que la exponencial en p está bien definida y es diferenciable en un entorno del 0 en $T_p(N)$. Además si consideramos la diferencial de la exponencial en p entre $T_p(T_p(N))$ y $T_p(N)$, esta coincide con el isomorfismo lineal canónico. Si ahora consideramos una curva $\rho = tv$ con t un parámetro y $v \in T_p(N)$, por definición $v_p = \rho'(0)$ ($\exp_p(tv) := \gamma_v(t)$). Finalmente:

$$d\exp_p(v_p) = d\exp_p(\rho'(0)) = (\exp_p \circ \rho)'(0) = \gamma'_v(0) = v.$$

El resultado sigue del Teorema 1.3.2 de la Función Inversa . ■

2.4. Tensor de curvatura Riemanniano. Curvatura seccional

Durante todo el capítulo hemos estado desarrollando un formalismo que nos permitiera definir de algún modo la curvatura en nuestra variedad. Llegados a este punto, somos capaces de introducir la noción de curvatura. De hecho, no solo introduciremos una idea de curvatura, sino cuatro, aunque todas ellas parten del mismo objeto: *el tensor de curvatura de Riemann*. Es bien conocido para las superficies que la curvatura es una noción intrínseca a estas, dependiendo solo de la métrica (*Theorema Egregium* de Gauss). Sin embargo, generalizar esta idea en variedades no es trivial, trabajo que le debemos a Riemann (también a Ricci, Levi-Civita y Christoffel) tal y como desarrollamos históricamente en la introducción al documento.

La idea actual detrás de esta generalización se puede llevar a cabo de diferentes maneras. Nosotros nos centraremos, como dijimos en el apartado anterior, en aquella relacionada con el transporte paralelo. En este sentido, tal y como explicamos en los ejemplos anteriores, la idea esencial es que al hacer el transporte paralelo de un vector en un ciclo — curva cerrada — en una variedad plana, el vector al final del transporte es el mismo que al principio. Sin embargo, si hacemos esto en una variedad curva, el vector final no será igual que el inicial (como vimos en el caso de la esfera). La diferencia entre estos vectores, que en teoría clásica de superficies se denomina ángulo de holonomía, está directamente relacionado con la curvatura de la superficie y, esencialmente, es esta noción la que se pretende determinar a través del tensor de curvatura de Riemann. Otra forma de interpretar el tensor de Riemann es considerarlo, en el sentido de la aplicación exponencial, la diferencia o separación de la métrica de la variedad con respecto a la métrica plana.

Definición 2.4.1 (Tensor de curvatura de Riemann) *Dada una variedad N con*

la conexión de Levi-Civita ∇ , llamaremos tensor de curvatura de Riemann a la aplicación $R_{XY}Z : \mathcal{X}(N)^3 \rightarrow \mathcal{X}(N)$ dada por $R_{XY}Z = \nabla_{[X,Y]}Z - [\nabla_X, \nabla_Y]Z$.

Nótese que el signo menos de la definición es cuestión de convenio y depende completamente del autor consultado. Aquí seguiremos este — se podría medir dicha diferencia justo del revés y todo el desarrollo quedaría igual. Además, lo hemos caracterizado como tensor pero no lo hemos probado. Veamos que es precisamente un campo tensorial:

Lema 2.4.1 *El tensor de curvatura de Riemann es un campo tensorial de tipo (1,3).*

Demostración: De nuestra caracterización de tensores, como $R : \mathcal{X}(N)^3 \rightarrow \mathcal{X}(N)$, si es $\mathcal{F}(N)$ -lineal (la \mathbb{R} -linealidad es trivial) podremos verlo como un tensor (1,3). Veamos pues la $\mathcal{F}(N)$ -linealidad en cada componente. Para las dos componentes inferiores es análogo, probémoslo para X por ejemplo:

$$\begin{aligned} R_{fXY}Z &= \nabla_{[fX,Y]}Z - [\nabla_{fX}, \nabla_Y]Z = \nabla_{[fX,Y]}Z - \nabla_{fX}\nabla_YZ + \nabla_Y\nabla_{fX}Z = \\ &= \nabla_{-(Yf)X+f[X,Y]}Z - f\nabla_X\nabla_YZ + \nabla_Y(f\nabla_X)Z = \\ &= -(Yf)\nabla_XZ + f\nabla_{[X,Y]}Z - f\nabla_X\nabla_YZ + f\nabla_Y\nabla_XZ + (Yf)\nabla_XZ = \\ &= fR_{XY}Z. \end{aligned}$$

Finalmente, veamos para la componente Z :

$$\begin{aligned} R_{XY}fZ &= \nabla_{[X,Y]}fZ - [\nabla_X, \nabla_Y]fZ = \nabla_{[X,Y]}fZ - \nabla_X\nabla_YfZ + \nabla_Y\nabla_XfZ = \\ &= ([X,Y]f)Z + f\nabla_{[X,Y]}Z - \nabla_X(Yf)Z - \nabla_X(f\nabla_YZ) + \nabla_Y(Xf)Z + \nabla_Y(f\nabla_XZ) = \\ &= (XYf)Z - (YXf)Z + f\nabla_{[X,Y]}Z - (XYf)Z - (Yf)\nabla_XZ - \\ &\quad - (Xf)\nabla_YZ - f\nabla_X\nabla_YZ + (YXf)Z + (Xf)\nabla_YZ + (Yf)\nabla_XZ + f\nabla_X\nabla_YZ = \\ &= fR_{XY}Z. \end{aligned}$$

Por tanto, R es un campo tensorial (1,3). ■

Veamos algunas propiedades del tensor de curvatura que serán útiles posteriormente. Estas propiedades suelen llamarse simetrías de la curvatura.

Proposición 2.4.1 *Dados una variedad N y $X, Y, Z, V, W \in \mathcal{X}(N)$ el tensor de curvatura de Riemann cumple las siguientes propiedades:*

1. *Es antisimétrico en su componentes subíndice: $R_{XY}Z = -R_{YX}Z$.*
2. *Es simétrico por pares: $g(R_{XY}V, W) = g(R_{VW}X, Y)$.*
3. *Es antiadjunto respecto de la métrica: $g(R_{XY}V, W) = -g(R_{XY}W, V)$.*
4. *Se da la primera identidad de Bianchi: $R_{XY}Z + R_{YZ}X + R_{ZX}Y = 0$.*
5. *Se da la segunda identidad de Bianchi: $(\nabla_X R)(Y, Z, V) + (\nabla_Y R)(Z, X, V) + (\nabla_Z R)(X, Y, V) = 0$.*

Demostración: Antes de comenzar la demostración destacamos una ventaja de la que nos aprovecharemos a lo largo de esta para simplificarla. Debido al caracter local de la derivada covariante y de los corchetes de Lie, podemos trabajar en sistemas de coordenadas que nos den determinadas facilidades sin perder generalidad. En nuestro caso, podemos hacer que todos los corchetes sean nulos (sin más que tomarlos constantes en el sistema de coordenadas concreto), resultando el tensor de Riemann: $R_{XY} = \nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z$. Probemos ahora las propiedades:

La antisimetría es directa:

$$R_{XY}Z = \nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z = -(\nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z) = -R_{YX}Z.$$

Para la simetría por pares necesitaremos las otras propiedades, y a partir de ellas, no será más que un cálculo combinatorio simple pero engorroso. Para una demostración consúltese [18]

Para la propiedad de antiadjunto con la métrica usemos la linealidad del tensor de Riemann y del tensor métrico:

$$g(R_{XY}V, W) = -g(R_{XY}W, V) \text{ si y solo si } g(R_{XY}(V+W), V+W) := g(R_{XY}U, U) = 0.$$

La demostración es ahora directa pues:

$$\begin{aligned} g(R_{XY}U, U) &= g(\nabla_Y \nabla_X U, U) - g(\nabla_X \nabla_Y U, U) = \\ &= Yg(\nabla_X U, U) - g(\nabla_X U, \nabla_Y U) - Xg(\nabla_Y U, U) + g(\nabla_Y U, \nabla_X U) = \\ &= \frac{1}{2}(YX - XY)g(U, U) = 0. \end{aligned}$$

Probemos la primera identidad de Bianchi:

$$\begin{aligned} R_{XY}Z + R_{YZ}X + R_{ZX}Y &= \nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z + \nabla_Z \nabla_Y X - \\ &\quad - \nabla_Y \nabla_Z X + \nabla_X \nabla_Z Y - \nabla_Z \nabla_X Y = 0. \end{aligned}$$

Para la segunda identidad de Bianchi desarrollemos cada término y simplemente tenemos el resultado sumando. Desarrollemos uno de los términos; los demás serán igual.

$$\begin{aligned} (\nabla_X R)(Y, Z, V) &= \nabla_X(\nabla_Z \nabla_Y V - \nabla_Y \nabla_Z V) - R_{(\nabla_X Y)Z}V - \\ &\quad - R_{Y\nabla_X Z}V - \nabla_Z \nabla_Y \nabla_X V + \nabla_Y \nabla_Z \nabla_X V. \end{aligned}$$

Concluyendo así la demostración. ■

Si seguimos analizando el caracter tensorial del tensor de Riemann, de forma parecida a cómo ocurría con el Lema 1.6.1, podemos deducir que en cada punto el tensor de curvatura nos genera un operador lineal del espacio tangente en sí mismo dado por $R_{uv} : T_p(N) \rightarrow T_p(N)$, que lleva cada $w \in T_p(N)$ a $R_{uv}w$ habiendo fijado $p \in N$ y $u, v \in T_p(N)$. A este operador se le conoce como operador de curvatura y hereda — ya que los vectores se pueden extender a campos vectoriales — todas las propiedades ya vistas del tensor de Riemann.

En la sección previa adelantábamos que los símbolos de Christoffel jugarían también un papel importante en la noción de curvatura. Estudiemos la expresión por componentes del tensor de Riemann para ver esta relación.

Lema 2.4.2 *Dada una carta $(U, \varphi = (x^1, \dots, x^n))$ de una variedad N , su tensor de curvatura de Riemann R se expresa localmente en coordenadas como*

$$R_{\partial_k \partial_l} \partial_j = R_{jkl}^i \partial_i,$$

donde las componentes del tensor de Riemann vienen dadas a través de los símbolos de Christoffel como

$$R_{jkl}^i = \frac{\partial}{\partial_l} \Gamma_{kj}^i - \frac{\partial}{\partial_k} \Gamma_{lj}^i + \Gamma_{lm}^i \Gamma_{kj}^m - \Gamma_{km}^i \Gamma_{lj}^m.$$

Demostración: La demostración consiste en una computación directa del tensor de Riemann $R_{\partial_k \partial_l} \partial_j = \nabla_{\partial_l} \nabla_{\partial_k} \partial_j - \nabla_{\partial_k} \nabla_{\partial_l} \partial_j$, ya que los corchetes de la base son nulos como sabemos. Desarrollemos uno de estos términos:

$$\begin{aligned} \nabla_{\partial_k} (\nabla_{\partial_l} \partial_j) &= \nabla_{\partial_k} (\Gamma_{lj}^r \partial_r) = \\ &= (\partial_k \Gamma_{lj}^r) \partial_r + \Gamma_{lj}^r \nabla_{\partial_k} \partial_r = \\ &= (\partial_k \Gamma_{lj}^i + \Gamma_{lj}^r \Gamma_{kr}^i) \partial_i. \end{aligned}$$

Del mismo modo: $\nabla_{\partial_l} (\nabla_{\partial_k} \partial_j) = (\partial_l \Gamma_{kj}^i + \Gamma_{kj}^r \Gamma_{lr}^i) \partial_i$. Sin más que restarlas tenemos la expresión del lema. ■

Vemos por tanto, como intuíamos en la sección anterior, que la curvatura viene determinada por los símbolos de Christoffel y por la métrica a través de ellos. Así pues, como ya adelantábamos, aquellas variedades con símbolos de Christoffel nulos serán llanas como los espacios Euclídeo y semi-Euclídeo — al tener métrica constante — ya que el tensor de Riemann se anula. Esto nos lleva a la siguiente noción:

Definición 2.4.2 (Riemann llana) *Diremos que una variedad N es Riemann llana o, simplemente, llana, si su tensor de curvatura de Riemann es idénticamente nulo.*

Para acabar la sección introduciremos una segunda noción de curvatura, *la curvatura seccional*, que será de utilidad cuando estudiemos las fibras del producto deformado del universo en el Capítulo 4. La curvatura seccional, como veremos, es un intento de generalización de la curvatura gaussiana de las superficies. La idea será considerar unas superficies concretas — los planos tangentes — y sobre ellas, estudiar el cociente del operador de curvatura, que hará las veces de segunda forma fundamental — recalco una vez más que esto solo es una idea intuitiva, formalmente no es así — sobre el determinante del tensor métrico, que como vimos se comporta igual que la primera forma fundamental.

Si siguiendo la línea del documento, pasemos ahora a formalizar estos conceptos. Comenzamos introduciendo la función $Q : T_p(N) \times T_p(N) \rightarrow \mathbb{R}$ dada por $Q(u, v) = g(u, u)g(v, v) - g(u, v)^2$ para todo $u, v \in T_p(N)$ — nótese, como se ha comentado antes, que esta función no es otra cosa que el determinante de la matriz que representa el tensor métrico.

A continuación definimos las superficies donde introduciremos la curvatura seccional.

Definición 2.4.3 (Sección plana) Dado $p \in N$ llamaremos sección plana a cualquier subespacio Π de dimensión 2 de $T_p(N)$.

Es común referirse a las secciones planas también como *planos tangentes*; usaremos ambas nomenclaturas indistintamente.

Como hemos esbozado antes, vamos a querer definir la curvatura seccional como un cociente, así que tendremos que asegurar que el denominador no se anule. Introducimos el siguiente resultado, que realmente es trivial puesto que, como Q es el determinante de la matriz asociada a g y g es no degenerado, Q no puede anularse por el Lema 1.6.3 y, más directamente, por el Corolario 2.1.1.

Corolario 2.4.1 Diremos que una sección plana Π es no degenerada si y solo si $Q(u, v) \neq 0$ para todo $u, v \in \Pi$.

Estamos ya en condiciones de introducir la curvatura seccional propiamente dicha.

Definición 2.4.4 (Curvatura seccional) Dada una sección plana no degenerada Π con base $\{u, v\}$ llamaremos curvatura seccional de Π a

$$\kappa(\Pi) = \frac{g(R_{uv}u, v)}{Q(u, v)}.$$

Nótese que la curvatura seccional es intrínseca a la sección plana:

Lema 2.4.3 La curvatura seccional de una sección plana Π no depende de la base elegida.

Demostración: Si tomamos otra base de Π aparte de $\{u, v\}$, llamémosla $\{u', v'\}$, deben existir escalares $a, b, c, d \in \mathbb{R}$ con $u = au' + bv'$ y $v = cu' + dv'$ con $ad - bc \neq 0$. Así pues, computando $Q(u, v)$ en la nueva base tenemos:

$$Q(u, v) = g(au' + bv', au' + bv')g(cu' + dv', cu' + dv') - g(au' + bv', cu' + dv')^2 = (ad - bc)^2 Q(u', v').$$

Del mismo modo, computando el operador curvatura en la nueva base y usando las propiedades del tensor de curvatura:

$$g(R_{uv}u, v) = (ad - bc)^2 g(R_{u'v'}u', v').$$

Sin más que hacer el cociente teniendo en cuenta que $ad - bc$ es distinto de cero tenemos el resultado. ■

En nuestro desarrollo nos interesarán aquellas variedades que tenga curvatura seccional constante.

Definición 2.4.5 (Curvatura seccional constante) *Diremos que una variedad N tiene curvatura seccional constante si $\kappa(\Pi)$ es independiente de la sección plana Π que tomemos en N . En este caso, notaremos la curvatura seccional de la variedad por κ .*

En particular nos interesarán tres variedades con curvatura seccional constante (que además son las simplemente conexas). El espacio Euclídeo con el producto escalar usual, que tendrá $\kappa = 0$. La n -esfera unidad, con el pullback de la métrica Euclídea de \mathbb{R}^{n+1} a través de la inclusión, que notaremos como S^n y que tendrá $\kappa = 1$. Y, por último, el n -espacio hiperbólico, con la métrica inducida del espacio semi-Euclídeo \mathbb{R}_1^{n+1} , que notaremos como \mathbb{H}^n y que tendrá $\kappa = -1$.

2.5. Tensor de Ricci y curvatura escalar

Aunque en la sección anterior hemos llegado a la noción de curvatura más general a través del tensor de curvatura de Riemann, para nuestros propósitos necesitamos introducir otras ideas de curvatura que son las que aparecerán en las ecuaciones de campo de la Teoría de la Relatividad General. Una pregunta obvia que surge es por qué esto es así y no pueden construirse a través del tensor de Riemann. Aunque al nivel del documento, y en el punto en que nos encontramos — aún no hemos hablado de la Teoría de la Relatividad — nuestra respuesta no pueda ser del todo satisfactoria, la idea clave es que buscamos objetos invariantes a través de isometrías, y aunque el tensor de Riemann lo es — recuérdese que viene determinado por la métrica a través de los símbolos de Christoffel — se pueden generar invariantes más simples y prácticos en el sentido que veremos a continuación.

Estos nuevos objetos serán *el tensor de Ricci y la curvatura escalar*. Como venimos diciendo, ambos se construirán como “simplificaciones” del tensor de Riemann; siendo más precisos, a través de contracciones de este. Ahora bien, dicha contracción no será la introducida en el Capítulo 1 para la curvatura escalar, sino que será una contracción inducida por la métrica de la variedad semi-Riemanniana. Para definir este tipo de contracción, primero tenemos que definir cómo podemos cambiar el tipo de un tensor:

Definición 2.5.1 (Subir índices) Dada una variedad N y dados $a, b, r, s \in \mathbb{Z}^+$ con $a \in [1, r]$ y $b \in [1, s]$, llamaremos operación subir índices o función sostenido a la aplicación $\uparrow_b^a: \Lambda_s^r(N) \rightarrow \Lambda_{s-1}^{r+1}(N)$ dada por:

$$(\uparrow_b^a T)(\theta^1, \dots, \theta^{r+1}, V_1, \dots, V_{s-1}) = T(\theta^1, \dots, \theta^{a-1}, \theta^{a+1}, \dots, \theta^{r+1}, V_1, \dots, \theta_a^*, \dots, V_{s-1})$$

donde θ_a^* es el campo vectorial métricamente equivalente a θ^a según el Teorema 2.2.1 y ocupa la entrada b que pasará a la a .

De hecho, la operación inversa viene dada por el propio Teorema 2.2.1:

Corolario 2.5.1 (Bajar índices) En las condiciones del apartado anterior, la función bajar índices es un isomorfismo $\mathcal{F}(N)$ -lineal cuya inversa es la función bajar índices o función bemol $\downarrow_b^a: \Lambda_s^r(N) \rightarrow \Lambda_{s+1}^{r-1}(N)$ dada por:

$$(\downarrow_b^a T)(\theta^1, \dots, \theta^{r-1}, V_1, \dots, V_{s+1}) = T(\theta^1, \dots, V_*^b, \dots, \theta^{r-1}, V_1, \dots, V_{b-1}, V_{b+1}, \dots, V_{s+1})$$

donde V_*^b es la 1-forma métricamente equivalente a V_b según el Teorema 2.2.1 y ocupa la entrada a que pasará a la b .

A través de la idea de contracción del Capítulo 1 y las operaciones que acabamos de introducir, podemos definir como queríamos la noción de contracción métrica que nos permitirá contraer índices arriba o abajo a través del propio tensor métrico.

Definición 2.5.2 (Contracción métrica) Dados una variedad N , un tensor $T \in \Lambda_s^r(N)$ y $a, b, r, s \in \mathbb{Z}^+$ con $a \in [1, r]$ y $b \in [1, s]$, definimos la contracción métrica covariante como $\mathbf{C}_{ab}T = C_b^a(\uparrow_a^b T) \in \Lambda_{s-2}^r$. Equivalentemente, definimos la contracción métrica contravariante como $\mathbf{C}^{ab}T = C_b^a(\downarrow_a^b T) \in \Lambda_s^{r-2}$.

En estas condiciones, antes de introducir el tensor de Ricci, es destacable que, a pesar de que introducimos las componentes del tensor de Riemann como un tensor (1,3) en el Lema 2.4.2, es más usual trabajar con sus componentes en forma (0,4) a través de la contracción métrica:

$$R_{ijkl} = g_{ir}R_{jkl}^r.$$

Pudiéndose escribir las simetrías de curvatura de forma compacta como $R_{ijkl} = -R_{jikl} = -R_{ijlk} = R_{klij}$.

Aprovechamos también esta introducción técnica para definir la divergencia de un tensor simétrico que usaremos más adelante en la sección.

Definición 2.5.3 (Divergencia) Dados una carta $(U, \varphi = (x^1, \dots, x^n))$ de una variedad N y un tensor simétrico $T \in \Lambda_2^0(N)$, definimos la divergencia de T como $\text{div}T = \mathbf{C}_{13}(\nabla T) = \mathbf{C}_{23}(\nabla T) \in \mathcal{X}(N)^*$ que en coordenadas será $(\text{div}T)_i = g^{jk}T_{ij;k} = T_{i;k}^k$.

Nótese que tanto la contracción métrica 13 como la 23 nos producen el mismo resultado por ser T simétrico — dicha contracción está relacionada con la traza de la representación matricial del tensor.

Tras esta introducción podemos introducir el tensor de Ricci y la curvatura escalar.

Definición 2.5.4 (Tensor de Ricci) *Dada una carta $(U, \varphi = x^1, \dots, x^n)$ de una variedad N con tensor de curvatura de Riemann R , llamaremos tensor de Ricci, Ric , a la contracción $C_3^1(R)$ que localmente podrá expresarse como:*

$$Ric_{ij} = R_{ijk}^k.$$

Definición 2.5.5 (Curvatura escalar) *Dada una carta $(U, \varphi = x^1, \dots, x^n)$ de una variedad N con tensor de Ricci Ric , llamaremos curvatura escalar R a la única contracción métrica del tensor de Ricci, $C_{13}(Ric) \in \mathcal{F}(N)$, que localmente podrá expresarse como:*

$$R = g^{ij} Ric_{ij}.$$

Recalcamos que notaremos a la curvatura escalar con la letra R al igual que al tensor de Riemann, como es común en textos sobre Teoría de la Relatividad. Ambos se diferenciarán fácilmente por el contexto, y porque, en general, usaremos notación indexada en la que aparecerá el tensor de Riemann con índices y la curvatura escalar sin ellos.

Nótese que es directo probar que el tensor de Ricci es simétrico — lo hereda de las propiedades del tensor de Riemann. Además, podemos dar una nueva noción de ser llano en el mismo sentido que con el tensor de Riemann.

Definición 2.5.6 (Ricci llana) *Diremos que una variedad N es Ricci llana si su tensor de Ricci, Ric , es idénticamente nulo.*

Para acabar la sección y el capítulo, presentamos el siguiente resultado, que será esencial a la hora de definir el tensor de Einstein en el Capítulo 4.

Proposición 2.5.1 *Dada una variedad N , su tensor de Ricci, Ric , y su curvatura escalar R cumplen la siguiente expresión:*

$$dR = 2div Ric.$$

Demostración: Recuérdese que $dR = \nabla R$ por ser R función escalar y, por tanto, es equivalente demostrar que $\nabla R = 2\text{divRic}$. Para demostrar la expresión, consideremos la segunda identidad de Bianchi aplicada sobre elementos de la base ∂_i :

$$R_{jkl;r}^i + R_{jlr;k}^i + R_{jrk;l}^i = 0.$$

Intercambiando r con k en el último término, generando un signo menos, y contrayendo en el sentido usual i con r :

$$R_{jkl;r}^r + R_{jlr;k}^r - R_{jrk;l}^r = R_{jkl;r}^r + R_{jl;k} - R_{jk;l} = 0.$$

Contrayendo métricamente j con k :

$$g^{jk} R_{jkl;r}^r + g^{jk} R_{jl;k} - R_{;l} = 0.$$

El segundo término es por definición divRic y el último término es ∇R . Por tanto, si probamos que el primer término es también divRic , tenemos la expresión que buscábamos. Trabajemos con el primer término usando las simetrías de la curvatura. Considerémoslo sin derivar:

$$g^{jk} R_{jkl}^r = g^{jk} g^{ir} R_{ijkl} = g^{jk} g^{ir} R_{jilk} = g^{ir} R_{il}.$$

Recuperando la diferencial covariante:

$$g^{jk} R_{jkl;r}^r = g^{ir} R_{il;r} = R_{l;r}^r = (\text{divRic})_l.$$

Por lo que concluye el resultado. ■

Capítulo 3

Variedades de Lorentz: El espacio-tiempo

El último paso para poder llegar a la definición de espacio-tiempo consistirá en tomar una variedad Lorentziana y equiparla con una última estructura: *la orientación temporal*.

En el capítulo previo explicamos que, dentro de las variedades semi-Riemannianas, aquellas que nos permitirían modelar un espacio-tiempo serían las variedades de Lorentz, puesto que el hecho de que tengan índice uno nos proporciona la condición idónea para identificar el tiempo con esta dimensión *anómala* en la que el producto escalar deja de tener sentido clásico y pasa a ser definido negativo.

A pesar de que, tal y como hemos dicho, dentro de estas variedades ya se pueda hablar de tiempo como una dimensión más, necesitamos caracterizar dicha dimensión con las propiedades intrínsecas del tiempo físico o real. En este sentido, dotaremos a la variedad de una orientación temporal, es decir, introduciremos un objeto que nos permita equipar la dimensión temporal con la irreversibilidad que la caracteriza o, equivalentemente, que nos permita formalizar las ideas de presente, pasado y futuro.

Dedicaremos la primera sección del capítulo a realizar el desarrollo que hemos descrito en el párrafo anterior. Las dos secciones posteriores constituirán desarrollos técnicos de conceptos y resultados que necesitaremos invocar en numerosas demostraciones cuando estudiemos la causalidad en las variedades de Lorentz en el último capítulo. Así pues, enfocaremos estas secciones en hacer una breve introducción a la geometría *extrínseca* de una variedad semi-Riemanniana y a la variación de longitudes de geodésicas. Los resultados de estas dos últimas secciones son generales, no se dan solo para espacios-tiempo. Sin embargo, los incluimos en este capítulo pues constituyen un paso previo y natural para pasar de los conceptos desarrollados en la primera sección y en los capítulos previos a la teoría que se desarrollará en los dos últimos capítulos.

Las ideas expuestas en este capítulo siguen directamente la línea argumental de [24] complementada con ideas y demostraciones de [18].

3.1. Espacio-tiempo. Orientabilidad temporal

Para la construcción del espacio-tiempo nos restringimos, como venimos diciendo, a variedades de Lorentz. En particular, para poder formalizar la idea de orientación temporal que hemos esbozado, tendremos que imponer condiciones sobre los vectores tangentes a la variedad, es decir, realizaremos nuestras construcciones sobre espacios vectoriales y, para nuestros propósitos, particularizaremos los resultados para el espacio tangente a nuestra variedad en cada punto.

Al tratar con variedades de Lorentz, notemos que los espacios tangentes y, en general, todos los espacios vectoriales de nuestras definiciones, serán de Lorentz en el sentido de que serán espacios vectoriales de dimensión mayor o igual que dos y equipados con un producto escalar de índice uno.

Antes de comenzar a formalizar la idea de orientación temporal, veamos primero cómo afecta a los vectores tangentes el hecho de que el tensor métrico tenga índice uno. Hasta ahora, habíamos comentado que la dimensión en que el producto escalar se vuelve definido negativo era perfecta para modelar nuestra idea de tiempo. Sin embargo, en ningún momento hemos formalizado esta idea. La siguiente clasificación de vectores pone de manifiesto esta noción de *tiempo* como una dimensión más.

Definición 3.1.1 (Carácter causal de vectores) *Dado un espacio vectorial de Lorentz (V, g) diremos que un vector $v \in V$ es*

- temporal si $q_g(v) < 0$,
- luminoso o nulo si $q_g(v) = 0$ pero $v \neq 0$,
- espacial si $q_g(v) > 0$,
- causal si es temporal o luminoso,

donde q_g es la forma cuadrática asociada a g .

Denominaremos *carácter causal* de un vector al carácter que tenga según la definición anterior. El hecho de que esta clasificación de vectores esté relacionada con la causalidad en la variedad — relaciones causa-efecto entre diferentes puntos — se hará más evidente al final de la sección cuando estudiemos esta para un tipo concreto de espacio-tiempo.

Notemos además, en relación a la definición previa, que esta no solo aplica a los vectores, es decir, dicha definición se extiende de forma natural a cualquier objeto que esté relacionado con vectores de un carácter causal concreto. En este sentido, diremos que un subespacio vectorial (por ejemplo, una hipersuperficie) tiene determinado carácter causal cuando todos sus vectores tienen dicho carácter, o lo que es lo mismo, que el producto escalar restringido a dicho espacio tenga las características asociadas a cada tipo causal — si es espacial (resp. temporal) definido positivo (resp. negativo), etc. Equivalentemente, si un espacio está generado por un vector, su carácter causal será el del vector.

En el mismo sentido, diremos que una geodésica tiene el carácter causal de su vector velocidad — recuérdese (ver capítulo previo) que el comportamiento del vector velocidad de las geodésicas estaba restringido y, por ello, su carácter causal no puede cambiar y está bien definido. Extendiendo esta idea, también podremos restringir propiedades ligadas a las geodésicas, como por ejemplo la completitud, a tipos de geodésicas diferentes en función de su carácter causal (podremos hablar de completitud temporal, causal, etc.).

Una vez hemos visto cómo el índice introduce el tiempo como una dimensión más en la variedad generando una estructura causal (que concretaremos más adelante), podemos comenzar a construir la idea de orientación temporal. Para ello, introducimos primero los conceptos de cono temporal y causal.

En el resto de la sección llamaremos $\mathbb{T}(V)$ al conjunto de todos los vectores temporales de un espacio vectorial de Lorentz V y $\mathcal{C}(V)$ al conjunto de todos sus vectores causales.

Definición 3.1.2 (Cono temporal) *Dados un espacio vectorial de Lorentz (V, g) y $v \in \mathbb{T}(V)$, llamaremos cono temporal de v al conjunto:*

$$C^+(v) = \{u \in \mathbb{T}(V) \mid g(v, u) < 0\}$$

Definición 3.1.3 (Cono causal) *Dados un espacio vectorial de Lorentz (V, g) y $v \in \mathcal{C}(V)$, llamaremos cono causal de v al conjunto:*

$$\bar{C}^+(v) = \{u \in \mathcal{C}(V) \mid g(v, u) < 0\}$$

De la propia definición es fácil ver que ambos conos son convexos y que dos vectores pertenecen al mismo cono temporal si y solo si su producto escalar es negativo. Nótese además, que para cada vector no hay un único cono temporal (causal) sino que hay dos. Es claro que podemos definir el cono temporal (causal) opuesto como $C^-(v) := C^+(-v) = -C^+(v)$ (idem para el causal) sin más que exigir ahora la condición $g(v, u) > 0$ sobre los vectores temporales (resp. causales). Es esta dualidad la que nos permitirá modelar las ideas de pasado y futuro.

En estas condiciones podemos introducir la orientación temporal.

Definición 3.1.4 (Orientación temporal) *Dada una variedad de Lorentz N , llamaremos orientación temporal a una aplicación τ^{+-} que a cada punto p de N le asigna un cono temporal de $T_p(N)$ y que es diferenciable, en el sentido de que, para cada punto p existe un campo vectorial diferenciable V que en cada punto q de un entorno de p pertenece al cono temporal asignado a dicho punto q . Diremos que N es temporalmente orientable si admite una orientación temporal.*

En vistas de la definición, podemos ver ahora que si exigimos que nuestra variedad sea temporalmente orientable, además de tener una dimensión temporal, vamos a tener bien definido en cada punto un par de conos temporales (causales) con orientaciones opuestas. Estos conos, dado que la orientabilidad es diferenciable, nos van a permitir definir de forma correcta una orientación que llamaremos futuro y otra que llamaremos pasado.

Recopilando las construcciones de estos tres primeros capítulos, hemos construido un objeto que generaliza nuestra noción más básica de espacio — variedad diferenciable — dentro de la cual podemos medir longitudes, ángulos y curvaturas — variedad semi-Riemanniana — que engloba nuestra idea de tiempo como una dimensión más — variedad Lorentziana — y que, si además, equipamos con una orientación temporal, nos permite asignarle al tiempo su carácter irreversible a través de una correcta definición de pasado y futuro. En resumen, hemos llegado a la idea de espacio-tiempo.

Definición 3.1.5 (Espacio-tiempo) *Dada una variedad de Lorentz (M, g) m -dimensional y una orientación temporal τ^{+-} llamaremos espacio-tiempo a la terna (M, g, τ^{+-}) . Arbitrariamente llamaremos futuro a la orientación temporal positiva, pasado a la negativa y denominaremos a los puntos de M sucesos o eventos.*

En el sentido de esta definición, una vez hayamos fijado una orientación temporal, diremos que una curva apunta al futuro (resp. pasado) si su vector velocidad está en el cono con orientación futuro (resp. pasado). Esta nomenclatura se extenderá de forma equivalente cuando hablemos de geodésicas, o extensión de geodésicas.

Recalcamos que en la definición previa hemos dicho que la variedad es m -dimensional. Sin embargo, en el siguiente capítulo nos restringiremos a un espacio-tiempo de 4 dimensiones, pues es el que tiene sentido físico (3 dimensiones espaciales + 1 temporal). En el Capítulo 5, aunque nuestro desarrollo se centrará en objetivos físicos y queremos particularizar para el caso 4-dimensional, lo presentaremos para el caso m -dimensional por generalidad matemática. De aquí en adelante reservaremos la letra M para los espacios-tiempo y seguiremos notando con N a las variedades que tratemos de un modo general.

Antes de introducir un ejemplo de espacio-tiempo para acabar la sección, introducimos el siguiente lema que, en relación a lo que acabamos de decir, nos permitirá estudiar siempre los espacio-tiempo de 4-dimensiones como un espacio usual de 3 dimensiones que va avanzando en el tiempo (llamaremos a esta idea descripción 3+1).

Lema 3.1.1 *Dados un espacio-tiempo de Lorentz (V, g) y $v \in \mathbb{T}(V)$, podemos descomponer V como: $V = \langle v \rangle_{\mathbb{R}} \oplus v^{\perp}$ donde hemos notado con $\langle \cdot \rangle_{\mathbb{R}}$ a la recta vectorial generada v .*

Demostración: La demostración es directa a través del Lema 1.6.5, pues como v es temporal $\langle v \rangle_{\mathbb{R}}$ no es degenerado y, por tanto, tampoco v^{\perp} , teniéndose la suma. ■

Nótese además en la demostración previa que, como el índice de v^{\perp} es cero, tendremos la descripción $3 + 1$, es decir, un espacio 3-dimensional que avanza en una línea temporal 1-dimensional (la recta vectorial generada por el vector temporal).

Para acabar la sección, introducimos como ejemplo uno de los espacio-tiempo más simples y sobre el que se basa la Teoría Especial de la Relatividad: el espacio-tiempo de Lorentz-Minkowski.

Definición 3.1.6 (Espacio-tiempo de Lorentz-Minkowski) *Llamaremos espacio-tiempo de Lorentz-Minkowski o simplemente espacio de Minkowski al espacio semi-Euclídeo $\mathbb{R}_1^n := \mathbb{L}^n$.*

Al igual que ocurría antes, aunque definimos el espacio-tiempo de forma n -dimensional, el espacio de Minkowski sobre el que se desarrolla la Teoría Especial de la Relatividad es \mathbb{L}^4 . Se recomienda, aunque en el siguiente capítulo se introducirán los conceptos básicos de Relatividad Especial para estudiar la Teoría de la Relatividad General, un estudio, al menos leve, de las propiedades de \mathbb{L}^4 para tener menos complicaciones en dicho capítulo [24], [8].

En esta sección, nos restringiremos al caso bidimensional con el fin de que el lector genere una intuición de todos los conceptos previos y para realizar un primer acercamiento a qué significa causalidad.

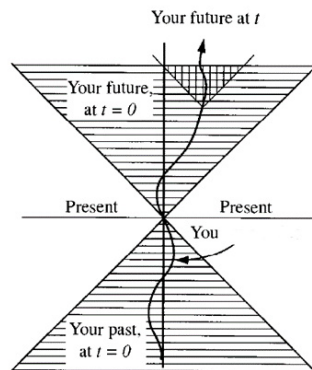


Figura 3.1: Causalidad en el espacio de Minkowski bidimensional.

De este modo, consideremos el plano \mathbb{R}^2 con el producto escalar tipo del espacio semi-Euclídeo para dimensión 2. Identificaremos el eje de ordenadas con el tiempo y el de abscisas con el espacio. Nótese que en este punto hay una simetría total en la elección de la signatura como ya comentamos, pudiéndose elegir el signo negativo o positivo indistintamente para la dimensión temporal. Para no crear confusión, en este ejemplo tomaremos el producto escalar mencionado sumando en las coordenadas espaciales y restando en las temporales.

Con esta elección, podemos ahora entender mejor el carácter causal de los vectores desde un punto de vista físico así como por qué los conos causales y temporales se llaman conos (ver Figura 3.1). Desde un punto de vista físico, como ya adelantamos en la introducción al documento y concretaremos en el capítulo 4, ningún sistema puede viajar más rápido que la velocidad de la luz. Este hecho limita la zona del espacio-tiempo que es accesible al observador (concepto que también se definirá), entendiendo accesible como aquella zona del espacio-tiempo que puede ser conectada por una curva que vaya desde la posición en el presente del observador hasta dicho evento.

Teniendo esto en mente, y considerando en nuestro ejemplo (aunque como veremos esto es general) la velocidad de la luz como uno, podemos interpretar las diagonales de los cuadrantes de \mathbb{R}^2 como rayos de luz que limitan el espacio-tiempo a un observador situado en el origen y cuyo presente es la recta $y = 0$. Es claro, según nuestra definición, que estas diagonales son precisamente los vectores luminosos o nulos (de ahí su nombre). Además, dichos rayos limitan un cono, que, de nuevo, con nuestra definición en mente estará formado por vectores temporales, tanto en el sentido positivo del eje y — futuro — como en el negativo — pasado, explicando el nombre de cono temporal. Es más, si consideramos junto con dicho cono los rayos en sí mismos, tendremos el cono causal.

Introducimos aquí por primera vez la noción de causalidad pues estamos viendo cómo desde un evento determinado en el presente, solo una serie de eventos son accesibles en el futuro, siendo estos sucesos aquellos que están dentro del cono causal (temporal) futuro o, equivalentemente, los puntos que se pueden unir con el origen a través de curvas causales (temporales). Nótese que aquellas zonas fuera de los conos (asociadas a zonas espaciales) no son accesibles desde el origen porque no pueden ser alcanzadas mediante curvas de este tipo, es decir, habría que viajar más rápido que la velocidad de la luz para alcanzarlos, lo cual es físicamente imposible.

Así pues, hemos formalizado matemáticamente las condiciones necesarias para que varios sucesos puedan relacionarse entre sí físicamente, esto es, mediante relaciones causa-efecto. El espacio de Minkowski es el caso más simple. Sin embargo, además de ayudarnos a intuir casos más generales, jugará un papel importante en el estudio general de la causalidad en espacios-tiempo.

3.2. Geometría extrínseca. Segunda forma fundamental.

En las condiciones de la sección previa, habiendo definido ya formalmente qué es un espacio-tiempo, podríamos comenzar a introducir ya la Teoría de la Relatividad General y la Teoría de la Causalidad en variedades de Lorentz. Sin embargo, necesitaremos varias herramientas previas para trabajar con esta última.

En primer lugar — y es a lo que dedicaremos esta sección — tendremos que generalizar la noción de geometría extrínseca de superficies a variedades. Hasta ahora, hemos estudiado las variedades desde un punto de vista intrínseco, esto es, desde el punto de vista de un observador que se encuentra inmerso en la variedad. Nuestro objetivo ahora será introducir la geometría de la variedad desde el punto de vista de un observador que la vea desde fuera, de forma extrínseca. De forma análoga a lo que ocurría en el caso de las superficies, en las que se introduce el concepto de vector normal para formalizar la idea de una visión externa de la geometría, vamos a considerar ahora nuestra variedad inmersa en una variedad más grande permitiendo el estudio extrínseco de su geometría.

Al igual que ocurría en el Capítulo 2, en el que generalizamos la geometría intrínseca, al generalizar la geometría extrínseca podríamos desarrollar una teoría completa y extensa generalizando todo tipo de conceptos a las variedades. En nuestro caso, haremos una breve introducción a la geometría extrínseca estudiando solo cómo cambia la conexión vista desde un punto de vista externo, permitiéndonos definir la segunda forma fundamental y la curvatura media que serán necesarias en desarrollos posteriores.

Para comenzar, como vamos a estudiar una variedad inmersa en otra, tendremos que formalizar la idea de subvariedad, primero sin métrica y luego con ella.

Para poder definir una subvariedad, primero tendremos que definir inmersiones y sumersiones:

Definición 3.2.1 (Sumersión e Inmersión) *Dadas dos variedades diferenciables N^n y M^m diremos que una aplicación diferenciable f entre ellas es inmersión (resp. sumersión) en un punto $p \in N$ si el rango de la matriz Jacobiana de $f(p)$ es n (resp. m). Diremos que f es inmersión (resp. sumersión) si lo es en todo punto de N .*

Así, una subvariedad se definirá del siguiente modo:

Definición 3.2.2 (Subvariedad) *Diremos que una variedad T es subvariedad de una variedad N si T es subespacio topológico de N y si la inclusión de T en N es una inmersión.*

La definición previa se extiende al caso semi-Riemanniano en el siguiente sentido:

Definición 3.2.3 (Subvariedad semi-Riemanniana) *Dada una subvariedad T de una variedad semi-Riemanniana N diremos que T es subvariedad semi-Riemanniana de N si el pullback de la métrica de N a través de la inclusión es una métrica en T .*

De este modo, de aquí en adelante consideraremos una variedad N que sea subvariedad semi-Riemanniana de una variedad \overline{N} . Además, notaremos en general todos los objetos de \overline{N} con una barra encima para diferenciarlos de los de N .

Comenzando con nuestro estudio para analizar cómo cambia la conexión, centrémonos en el espacio tangente de la variedad \overline{N} que llamaremos de forma usual variedad ambiente. Es claro que el espacio tangente en cada punto de N va a ser un subespacio no degenerado del espacio tangente de \overline{N} y el tensor métrico no será ambiguo en el sentido de que las visiones desde N y \overline{N} coincidirán en una visión puramente intrínseca. Ahora bien, tengamos en cuenta que el espacio tangente a \overline{N} estará constituido no solo por el espacio tangente de N sino por un conjunto más grande, de hecho, en visión del Lema 1.6.5 tendremos que:

$$T_p(\overline{N}) = T_p(N) \oplus T_p(N)^\perp$$

Así, tendremos vectores tangentes a N en el sentido usual, aquellos vectores de $T_p(N)$, y también vectores *normales*, a N aquellos vectores que pertenezcan a $T_p(N)^\perp$ (nótese que en relación a estructuras tendremos, en el mismo sentido que en el Capítulo 1, un fibrado normal, etc. y diremos que determinados objetos son normales si sus vectores tangentes son ortogonales). Además, en el sentido de la descomposición previa, la dimensión de $T_p(N)^\perp$ será la codimensión de N en \overline{N} , es decir, $\dim(\overline{N}) - \dim(N)$, y del mismo modo, su índice será el coíndice del tensor métrico: $\text{indice}(\overline{g}) - \text{indice}(g)$.

Notemos además que, al ser una suma directa, la descomposición previa nos permite — valga la redundancia — descomponer cualquier vector del espacio tangente de \overline{N} como la suma de su parte tangente a N y su parte normal a N . Esta descomposición induce en el sentido más clásico dos proyecciones ortogonales que llamaremos *tan* y *nor* — claramente \mathbb{R} -lineales — desde el espacio tangente de la variedad ambiente al espacio tangente de N y a su ortogonal, respectivamente.

Destacamos además que todo el desarrollo realizado para vectores tangentes se extiende de un modo natural a campos vectoriales, teniéndose las mismas propiedades: suma directa de los campos vectoriales tangentes y normales, y proyecciones — ahora $\mathcal{F}(N)$ -lineales — equivalentes a las previas.

En estas condiciones, ya podemos analizar cómo varía la conexión en esta nueva visión de la variedad:

Definición 3.2.4 (Conexión inducida) *Dados una variedad semi-Riemanniana \overline{N} , una subvariedad semi-Riemanniana $N \subset \overline{N}$ y campos vectoriales $V \in \mathcal{X}(N)$ y $X \in \overline{\mathcal{X}}(N)$, llamaremos conexión inducida a la conexión $\overline{\nabla} : \mathcal{X}(N) \times \overline{\mathcal{X}}(N) \rightarrow \overline{\mathcal{X}}(N)$ inducida en cada punto $p \in N$ a través de la conexión de Levi-Civita en \overline{N} (que también*

notaremos como $\overline{\nabla}$) quedando definida la derivada covariante $\overline{\nabla}_V X$ en cada $U \cap N$ como la restricción de $\overline{\nabla}_V \overline{X}$ a $U \cap N$ siendo U un entorno de $p \in \overline{N}$ y \overline{X} y \overline{V} las extensiones diferenciables a dicho entorno de X y V .

Una vez introducida la conexión inducida veamos que está bien definida y que de hecho es un campo vectorial diferenciable.

Lema 3.2.1 *La derivada covariante $\overline{\nabla}_V X$ está bien definida y es un campo vectorial diferenciable de \overline{N} en N .*

Demostación: Simplemente tendremos que demostrar que la definición no depende de las extensiones ya que la restricción de todo campo vectorial diferenciable también es campo vectorial diferenciable. Así pues, sin más que tomar un sistema coordenado (carta) como venimos haciendo, podemos escribir $X = f^i \partial_i$ y, por tanto, $\overline{\nabla}_V \overline{X} = \overline{V} f^i \partial_i + f^i \overline{\nabla}_V \partial_i$. Pero es claro que si, en las condiciones de la definición previa, tomamos un punto $p' \in U \cap N$ tendremos que $(\overline{V} f^i)(p') = V_{p'}(f^i|_{U \cap N})$ y $\overline{\nabla}_V(\partial_i)|_{p'} = \overline{\nabla}_{V_{p'}}(\partial_i)$ y consecuentemente $\overline{\nabla}_V \overline{X}$ solo depende de los campos. ■

La conexión inducida heredará todas las propiedades de la conexión de Levi-Civita de \overline{N} (basta considerar su restricción a N para demostrarlo).

Finalmente, ya estamos en condiciones de introducir la segunda forma fundamental. La idea esencial es que si ahora tomamos dos campos vectoriales tangentes a N , su derivada covariante no tendrá por qué ser tangente a N ya que ahora tendremos también una parte normal. Es concretamente esta parte normal de la derivada covariante la que constituye la segunda forma fundamental, también conocida como tensor de forma.

Las razones detrás de esta segunda denominación se deben a que, en el sentido más natural, la segunda forma fundamental nos permite en cierta manera darle forma a la variedad viéndola desde fuera y viendo cómo se aleja esta de la visión intrínseca desde dentro de la variedad. Así pues, la segunda forma fundamental *medirá* cuánto se aleja la geometría externa de la variedad de su visión desde su tangente introduciendo términos de corrección normal en la derivada covariante, en el tensor de curvatura, en la curvatura seccional, etc. (ver [18]). En este documento nos limitamos a definirla a través de la derivada covariante.

Comencemos viendo que la parte tangente de la derivada covariante inducida coincide con la derivada covariante en el sentido usual:

Lema 3.2.2 *Dados una variedad semi-Riemanniana \overline{N} , una subvariedad semi-Riemanniana $N \subset \overline{N}$ y V, W campos tangentes a N , se tiene que $\nabla_V W = \tan(\overline{\nabla}_V W)$.*

Demostración: Tomando otro campo tangente a N arbitrario, llamémoslo Z , la demostración es directa a través de la fórmula de Koszul aplicada sobre las extensiones de dichos campos $2g(\overline{\nabla}_{\overline{V}}\overline{W}, \overline{X}) = F(\overline{V}, \overline{W}, \overline{X})$, donde hemos seguido la misma notación que en la demostración del Teorema Fundamental de la Geometría semi-Riemanniana. Sin más que restringir sobre N , y teniendo en cuenta las propiedades heredadas de la conexión de Levi-Civita, podemos deducir que los términos de la igualdad anterior son iguales a $2g(\overline{\nabla}_V W, X)$ y $F(V, W, X)$ respectivamente. Por tanto, tenemos que $g(\overline{\nabla}_V W, X) = g(\nabla_V W, X)$ y como X es tangente podemos sustituir $\overline{\nabla}_V W$ por $\tan(\overline{\nabla}_V W)$ teniéndose el resultado. ■

Pasemos ahora a introducir la segunda forma fundamental propiamente dicha.

Definición 3.2.5 (Segunda forma fundamental) *Dadas una variedad semi-Riemanniana \overline{N} y una subvariedad semi-Riemanniana $N \subset \overline{N}$, llamaremos segunda forma fundamental o tensor de forma a la aplicación $II : \mathcal{X}(N) \times \mathcal{X}(N) \rightarrow \mathcal{X}(N)^\perp$ definida como $II(V, W) = \text{nor}(\overline{\nabla}_V W)$ para todos V, W campos vectoriales tangentes a N .*

No debe confundirse la segunda forma fundamental con el *endomorfismo de Weingarten* (en algunos textos también se le llama a este último tensor de forma).

Lema 3.2.3 *La segunda forma fundamental es $\mathcal{F}(N)$ -lineal y simétrica.*

Demostración: La linealidad la hereda directamente de la derivada covariante y de la proyección normal. Veamos que es simétrica para V, W tangentes:

$$II(V, W) - II(W, V) = \text{nor}(\overline{\nabla}_V W - \overline{\nabla}_W V) = \text{nor}[V, W] = 0.$$

Concluyéndose así la demostración. ■

Recalcamos que II es un tensor pero más general que los definidos previamente, al devolver campos vectoriales normales.

Finalmente introducimos el concepto de curvatura media que generaliza la curvatura media de superficies y que jugará un papel esencial a la hora de poner condiciones en los Teoremas de Singularidad ya que, como veremos, estará directamente relacionada con el hecho de que existan ciertas zonas en las que la gravedad es tan fuerte que atrapa regiones del espacio-tiempo.

Definición 3.2.6 (Curvatura media) *Dadas una variedad semi-Riemanniana \overline{N} y una subvariedad semi-Riemanniana N , llamaremos curvatura media \vec{H} a la traza (en el sentido tensorial) de la segunda forma fundamental.*

De hecho, siendo concisos, no será exactamente la curvatura media la que estará relacionada con dichas regiones sino su “proyección” sobre determinados vectores en el sentido de la siguiente definición (concretaremos todo esto desde un punto de vista físico en el Capítulo 5).

Definición 3.2.7 (Convergencia) *Dadas una variedad semi-Riemanniana \bar{N} y una subvariedad semi-Riemanniana N con curvatura media \vec{H} , llamaremos convergencia de N a la función real k definida sobre el fibrado normal como $k(v) = g(v, \vec{H}_p)$ para todo $v \in T_p(N)^\perp$.*

3.3. Variaciones de longitud. Campos de Jacobi y puntos focales

En esta última sección nos dedicaremos al estudio de las variaciones de longitud, esto es, medir el cambio de la longitud de arco de curvas dentro de la variedad para pequeñas variaciones o desplazamientos.

Esta disciplina concreta se aleja en cierta medida del desarrollo central de este texto aunque obviamente estén estrechamente relacionados como veremos. La razón por la que comentamos esto es, esencialmente, porque usaremos esta disciplina como una pura herramienta intermedia para poder demostrar los resultados que nos aparecerán en el Capítulo 5 y, por tanto, no entraremos en detalle en la teoría subyacente a los resultados de esta sección — esencialmente por autocontención del documento y extensión. De este modo, introduciremos la notación, los conceptos básicos y los resultados que necesitaremos para el desarrollo posterior, algunos de ellos con demostración con el fin de que el lector tenga una primera toma de contacto con esta teoría, y otros sin demostración debido, como decimos, a su extensión y a la falta de relación directa con el contenido central del documento.

Para esta sección seguiremos esencialmente la línea argumental seguida en [24], recomendándose este texto para las demostraciones que no añadimos. En toda la sección estaremos trabajando con una variedad N y curvas definidas sobre cierto intervalo de la recta real.

Como hemos comentado justo al iniciar, nuestra tarea será estudiar variaciones de longitud para pequeñas perturbaciones de una curva. En concreto, variaciones del *funcional longitud* definido para una curva γ en una variedad N como: $L(\gamma) = \int_a^b |\gamma'(s)| ds$ donde s es el parámetro del intervalo donde tengamos definida la curva y a y b los extremos del arco de longitud que consideremos.

La definición de este funcional puede hacerse de modo general sobre *curvas diferenciables a trozos* que, en el sentido usual, no son mas que curvas diferenciables en determinadas particiones de un intervalo pero no en los extremos de dichos intervalos

— que llamaremos saltos o roturas — donde solo admiten la derivada a derecha e izquierda y cuyo salto de valor se denotará como $\Delta\gamma'(t_i)$ siendo t_i el punto de salto. Esta idea nos permite definir las geodésicas rotas que serán de utilidad más adelante:

Definición 3.3.1 (Geodésica rota) *Diremos que una curva diferenciable a trozos es una geodésica rota cuando su restricción a cada uno de los intervalos donde es diferenciable sea geodésica.*

Una vez conocido el funcional que queremos variar, veamos qué es una variación:

Definición 3.3.2 (Variación) *Llamaremos variación de un segmento de curva $\gamma : [a, b] \rightarrow N$ o simplemente variación a la aplicación $\mathbf{x} : (-\delta, \delta) \times [a, b] \rightarrow N$ tal que $\mathbf{x}(0, t) = \gamma(t)$ para todo t en el intervalo $[a, b]$ con a, b, δ reales.*

La idea esencial es por tanto que variaremos de forma infinitesimal una curva moviendo sus extremos o dejándolos fijos. Nótese además que la variación también es diferenciable a trozos en el sentido de que será diferenciable en los intervalos producto del tipo $(-\delta, \delta) \times [t_{i-1}, t_i]$, siendo estos últimos los intervalos donde es diferenciable la curva inicial con t_i los saltos o roturas.

En todo lo que sigue llamaremos w al primer parámetro del que depende la variación y t al segundo. Tendremos también la siguiente definición:

Definición 3.3.3 (Curvas longitudinales y transversales) *Llamaremos curvas longitudinales a aquellas curvas con w constante y transversales a aquellas con t constante.*

Finalmente, nuestro último objeto previo será el campo vectorial de variación o variacional:

Definición 3.3.4 (Campo vectorial variacional) *Llamaremos campo vectorial variacional al campo diferenciable a trozos V dado por $V(t) = \mathbf{x}_w(0, t)$.*

Además, en todo lo que sigue usaremos prima para denotar cualquier derivada total (incluso si es covariante) y para las derivadas parciales de la variación usaremos subíndices para referirnos a la variable con respecto a la cual estamos derivando — nótese que las derivas cruzadas son intercambiables siempre que derivemos solo dos veces $\mathbf{x}_{tw} = \mathbf{x}_{wt}$ debido a la simetría de la conexión de Levi-Civita. Finalmente, en todo lo que sigue usaremos curvas γ de N diferenciables a trozos, definidas en un intervalo $[a, b]$ de velocidad $|\gamma'(t)| := c$ constante y de signo ϵ dado por el carácter causal de γ' salvo que se indique lo contrario.

Comenzaremos estudiando la *primera variación* del funcional:

$$L'_x(0) := \frac{\partial}{\partial w} L(\mathbf{x}(w, t))|_{w=0}.$$

y posteriormente estudiaremos la *segunda variación*

$$L''_x(0) := \frac{\partial^2}{\partial^2 w} L(\mathbf{x}(w, t))|_{w=0}.$$

Para la primer variación tendremos la siguiente fórmula que nos servirá tanto para variaciones con extremos fijos como sueltos:

Teorema 3.3.1 *Dada una curva γ con velocidad estrictamente positiva, para cualquier variación \mathbf{x} de un campo variacional V a lo largo de γ se tiene que:*

$$\begin{aligned} L'_x(0) &= \frac{\epsilon}{c} \int_a^b g(\gamma', V') ds \\ &= -\frac{\epsilon}{c} \left(\int_a^b g(\gamma'', V) ds + \sum_{i=1}^{k-1} g(\Delta\gamma'(t_i), V(t_i)) - g(\gamma', V)|_a^b \right). \end{aligned}$$

Demostración: Tendremos que probar ambas igualdades. Nótese que la segunda es directa probando la primera, pues bastaría tener en cuenta que por linealidad $g(\gamma', V') = \frac{d}{dt}g(\gamma', V) - g(\gamma'', V)$ e integrar esta primera igualdad.

Probemos la primera. Al ser la velocidad estrictamente positiva, podemos tomar nuestro δ lo pequeño que queramos con el fin de asegurar que $g(x_t, x_t) \neq 0$. Podemos escribir por tanto

$$L'_x(w) = \int_a^b \frac{\partial}{\partial w} \sqrt{(\epsilon g(x_t, x_t))} dt = \epsilon \int_a^b \frac{g(x_t, x_{wt})}{|x_t|} dt,$$

donde hemos usado que $x_{wt} = x_{tw}$ que, como dijimos, se tiene por la simetría de la conexión de Levi-Civita. Basta evaluar esta última expresión en $w = 0$ para tener el resultado que buscamos. ■

El resultado previo nos permite obtener los siguientes resultados sobre las geodésicas rotas que serán de gran utilidad en demostraciones de causalidad:

Corolario 3.3.1 *Una curva γ de velocidad constante estrictamente positiva es una geodésica que no está rota si y solo si la primera variación de la longitud es cero para toda variación con extremos fijos.*

Demostración: Centrémonos en la implicación a izquierda pues, a derecha es trivial usando la expresión obtenida para la primera variación. Para esta implicación busquemos una contradicción. Supongamos que la derivada segunda de la curva no se anula si tomamos $t_{k-1} < s < t_k$ — seguimos con la misma notación para las roturas — y tomemos un vector v del tangente a lo largo de la curva en el punto $T_{\gamma(s)}N$ que cumpla que $g(\gamma''(s), v) > 0$. Consideremos además el transporte paralelo V asociado a v tomando un δ positivo tal que $g(\gamma'', Y) > 0$ en todo $t \in (s - \delta, s + \delta)$ y un campo vectorial variacional $X = fV$ siendo f una función no negativa que se anula fuera del intervalo previo y que es estrictamente positiva en s .

En estas condiciones, si aplicamos la fórmula de primera variación llegamos a contradicción pues:

$$L'_x(0) = -\frac{\epsilon}{c} \int_{s-\delta}^{s+\delta} fg(\gamma'', Y)dt > 0.$$

Deducimos por tanto que γ es una geodésica rota teniéndose el resultado — nótese que la diferenciabilidad se prueba de modo análogo sin más que considerar los saltos y buscar un campo variacional que asegure que el producto a través de g de los saltos y el campo variacional sea distinto de cero dentro del intervalo en consideración pero se anule fuera de este. ■

Realmente, el resultado que usaremos no será este sino una consecuencia directa de él. Para enunciarlo primero consideramos el espacio de todas las curvas diferenciables que parten de una subvariedad no degenerada $Q \subset N$ hasta un punto $q \in N$ que notaremos como $\Omega(Q, q)$.

Aunque el espacio $\Omega(Q, q)$ no será considerado como variedad propiamente dicha, introducimos un modo de definir un espacio tangente para su estructura que nos ayudará más adelante a introducir la forma indexada de una geodésica. Definimos este espacio tangente de modo que cada uno de sus elementos sea un campo vectorial variacional asociado a cierta variación de una geodésica.

Definición 3.3.5 (Espacio tangente de $\Omega(Q, q)$) *Dados una geodésica γ definida sobre un intervalo $[a, b]$ y el conjunto $\Omega(Q, q)$, definimos el espacio tangente a este conjunto $T_\gamma\Omega$ como el conjunto de todos los campos vectoriales diferenciables a trozos a lo largo de γ que cumplan que $V(b) = 0$ y $V(a) \in T_{\gamma(a)}Q$.*

Tras este inciso, enunciemos el resultado que adelantábamos.

Corolario 3.3.2 *Sea $\gamma \in \Omega(Q, q)$ con velocidad constante y estrictamente positiva. Entonces su primera variación es nula para toda variación de γ si y solo si γ es una geodésica que no está rota y que es normal a Q .*

Demostración: La demostración a derecha es directa a través de la fórmula de la primera variación sin más que tener en cuenta que el campo variacional en a es tangente a Q . La demostración a izquierda también es directa pues, a través del Corolario 3.3.1 tenemos asegurado que γ es geodésica y, de nuevo, con la fórmula de la primera variación tenemos asegurada la ortogonalidad a Q ya que $L'_x(0) = \frac{\epsilon}{c}g(\gamma'(a), v) = 0$ habiendo considerado $v \in T_{\gamma(a)}Q$ y un campo variacional V con $V(a) = v$. ■

Una vez estudiada la primera variación, estudiemos la segunda e introduzcamos la forma indexada de una geodésica a través del tangente a $\Omega(Q, q)$ como comentábamos previamente.

Teorema 3.3.2 *Dada una geodésica γ con velocidad constante estrictamente positiva y x una variación de γ se tiene que:*

$$L''_x(0) = \frac{\epsilon}{c} \left(\int_a^b (g((V')^\perp, (V')^\perp) - g(R_{\gamma'V}V, \gamma')) ds + g(\gamma', \mathbf{x}_{ww}(0, t)) \Big|_a^b \right).$$

Demostración: Sabemos que la longitud de cada curva variada es $\int_a^b |x_t(w, t)| dt$ y tengamos en cuenta que $|x_t(w, t)|_w = \frac{\epsilon g(x_{tw}, x_t)}{|x_t(w, t)|}$. Mediante un cálculo directo tenemos que:

$$\begin{aligned} L''_x(0) &= \int_a^b |x_t(w, t)|_{ww}(0, t) dt = \\ &= \frac{\epsilon}{|x_t(w, t)|} \int_a^b \left(g(x_{tw}, x_{wt}) + g(x_t, x_{tw}) - \frac{\epsilon}{|x_t(w, t)|^2} g(x_t, x_{tw})^2 \right) dt = \\ &= \frac{\epsilon}{c} \int_a^b \left(g(V', V') - g(R_{\gamma'V}V, \gamma') + g(\gamma', x'_{ww}(0, t)) - \frac{\epsilon}{c^2} g(\gamma', V')^2 \right) dt, \end{aligned}$$

donde en la última igualdad se ha tenido en cuenta la conmutatividad de las derivadas cruzadas y que la pérdida de conmutatividad se cuantifica a través del tensor de curvatura R .

Finalmente, tengamos en cuenta que $g(\gamma', x'_{ww}(0, t)) = \frac{d}{dt}g(\gamma', x_{ww}(0, t))$ ya que γ es geodésica y que $V' = \frac{\epsilon}{c^2}g(V', \gamma)\gamma' + (V')^\perp$. Sustituyendo e integrando en la expresión previa tenemos el resultado. ■

Definición 3.3.6 (Forma indexada) *Llamaremos forma indexada de una geodésica no luminosa $\gamma \in \Omega(Q, q)$ a la única forma bilineal y simétrica $I_\gamma^Q : T_\gamma\Omega \times T_\gamma\Omega \rightarrow \mathbb{R}$ definida como $I_\gamma^Q(V, V) = L''_x(0)$ para todo $V \in T_\gamma\Omega$. Si la subvariedad Q se limita a un punto escribiremos simplemente I_γ .*

A continuación introducimos los campos de Jacobi, los puntos conjugados y los puntos focales, que serán los conceptos centrales de la sección en relación a la causalidad.

Estos conceptos nos permitirán encontrar resultados acerca de curvas con determinadas características que unan puntos en nuestra variedad — relaciones causa-efecto — estando estas características íntimamente relacionada con la longitud de las curvas y con su carácter causal como veremos en los últimos resultados de la sección. Además, a través de estos conceptos entenderemos mejor por qué hemos introducido la forma indexada; a través de esta construiremos un resultado que nos permita entender de forma muy geométrica qué significan los puntos focales.

Comencemos pues introduciendo los campos de Jacobi. Estos campos serán aquellos campos vectoriales diferenciables a lo largo de geodésicas generados por variaciones en las que toda curva longitudinal es una geodésica.

Definición 3.3.7 (Campo de Jacobi) *Dado un campo vectorial diferenciable J a lo largo de una geodésica γ definida en un intervalo $[a, b]$, decimos que es un campo de Jacobi si satisface la ecuación diferencial:*

$$J'' + R_{J\gamma'}\gamma' = 0.$$

En el mismo sentido que antes, podemos definir del mismo modo los Q -campos de Jacobi como aquellos campos de Jacobi cuyas geodésicas longitudinales parten ortogonalmente de Q .

Definición 3.3.8 (Q-campo de Jacobi) *En las condiciones de la definición previa, si tenemos además una subvariedad Q ortogonal a γ en a , diremos que J es un Q -campo de Jacobi si $J(a)$ es tangente a Q y si la parte tangente a $T_{\gamma(a)}Q$ de su derivada cumple que $\tan(J'(a)) = \nabla_{J(a)}\gamma'(a)$.*

Una vez definidos los campos de Jacobi podemos definir los puntos conjugados y los puntos focales. Ambos, en cierto sentido, estarán asociados a variaciones de geodésicas que unan *a primer orden* un punto q y un punto p , siendo la única diferencia entre ambos que los puntos conjugados estarán asociados a campos de Jacobi en general, mientras que los puntos focales están asociados a Q -campos de Jacobi y, en consecuencia, a geodésicas que partan normalmente de la subvariedad Q .

Definición 3.3.9 (Puntos conjugados y puntos focales) *Dada una geodésica γ definida en un intervalo $[a, b]$, diremos que un punto $r \in (a, b]$ es un punto conjugado si existe un campo de Jacobi $J \neq 0$ a lo largo de γ con $J(a) = J(r) = 0$. Además, si existe una subvariedad Q ortogonal a γ en a diremos que r es un Q -punto focal si existe un Q -campo de Jacobi J a lo largo de γ con $J(r) = 0$.*

Recalamos que se ha abusado de la notación en la definición previa pues suele llamarse en la literatura *valor conjugado (focal)* a r y a $\gamma(r)$ punto conjugado (focal).

En el documento utilizaremos ambos nombres para ambos conceptos pues no lleva a confusión.

Para introducir el resultado prometido que relacione los puntos focales y conjugados con la forma indexada necesitamos hacer la siguiente definición previa:

Definición 3.3.10 (Geodésica coespacial) *Dada una variedad semi-Riemanniana N diremos que una geodésica es coespacial si el subespacio ortogonal a la geodésica es definido positivo.*

Nótese que si N es Riemanniana toda geodésica será coespacial, mientras que si N es Lorentziana toda geodésica temporal será coespacial.

Enunciemos ahora el resultado (como comentábamos al principio del capítulo este teorema será uno de los que presentemos sin demostración).

Teorema 3.3.3 *Dada una geodésica coespacial $\gamma \in \Omega(Q, q)$ con signo ϵ , entonces:*

- *Si no existen Q -puntos focales de $\gamma(a)$ a lo largo de γ la forma indexada I_γ^Q restringida a los campo vectoriales ortogonales a γ es definida.*
- *Si existe Q -punto focal r a lo largo de γ con $r \in (a, b)$ entonces I_γ^Q no es semi-definida.*

Con el Teorema 3.3.3 en mano, es fácil ver ahora que el significado geométrico de los puntos focales es el siguiente: habiendo fijado un signo para la geodésica, por ejemplo, positivo — para el caso negativo simplemente habría que cambiar máximo por mínimo en lo que sigue — la inexistencia de puntos focales implica que la geodésica en consideración es la curva de longitud mínima de entre todas las curvas cercanas a ella — variaciones — que une nuestra subvariedad con el punto asociado al extremo derecho del intervalo. En caso de que no existan puntos focales tendremos que existen curvas de mayor y menor longitud que la geodésica y que también unan la subvariedad con el punto.

Para acabar la sección introducimos varias aplicaciones de los resultados previos que nos serán de utilidad como venimos comentando. Estos resultados ponen de manifiesto la estrecha relación que existe entre la curvatura de la variedad y la existencia de puntos focales que, a su vez, estará directamente relacionada con la existencia de geodésicas de máxima o mínima longitud que unan determinados puntos lo cual, añadiendo un carácter causal a dicha geodésica, nos será de gran utilidad a la hora de estudiar las relaciones de causalidad entre diferentes puntos del espacio.

Como veremos a continuación, tanto la curvatura media como la convergencia juegan un papel importante. Estos conceptos, como ya veníamos adelantando, y como volveremos a estudiar en los Capítulos 4 y 5, estarán ligados al hecho de que determinadas

zonas queden atrapadas debido a lo fuerte que es la gravedad dándonos una primera pista de cómo la causalidad y la gravedad deben estar relacionadas. Presentaremos el primer resultado con demostración y los dos posteriores, que están relacionados entre sí, sin ella, pudiéndose encontrar esta en [24] como dijimos al inicio de la sección.

Teorema 3.3.4 *Sean Q una hipersuperficie espacial en una variedad de Lorentz N y γ una geodésica coespacial normal a Q en el punto $p = \gamma(0) \in N$. En estas condiciones si suponemos que $k(\gamma'(0)) = g(\gamma'(0), \vec{H}_p) > 0$ y $Ric(\gamma', \gamma') > 0$ entonces existe un punto focal $\gamma(t)$ de γ con $t \in (0, 1/k(\gamma'(0))]$ siempre que la curva esté definida en dicho intervalo.*

Demostración: Sin pérdida de generalidad supongamos que la velocidad de la curva es uno y, para aliviar la notación, escribamos $k := k(\gamma'(0))$. Consideremos una base ortonormal e_i de $T_p Q$ y sus transportes paralelos E_i a lo largo de γ — en toda la demostración $i = 1, \dots, n - 1$.

Definamos $f(t) = 1 - kt$ en el intervalo $[0, 1/k]$ de modo que $fE_i \in T_\gamma \Omega$ para todo i y consecuentemente:

$$\epsilon I_\gamma^Q(fE_i, fE_i) = k - \int_0^{1/k} f^2 g(R_{E_i \gamma'} E_i, \gamma') dt - g(\gamma'(0), II(e_i, e_i)).$$

Teniendo en cuenta que los transportes paralelos son espaciales sumando para cada i la expresión previa tendremos:

$$\sum_i \epsilon I_\gamma^Q(fE_i, fE_i) = (n - 1)k - \int_0^{1/k} f^2 Ric(\gamma', \gamma') dt - g(\gamma'(0), (n - 1)\vec{H}_p).$$

De esta expresión deducimos que $\epsilon I_\gamma^Q \leq 0$ ya que el primer y el último término de la expresión anterior se cancelan y sabemos que el tensor de Ricci es positivo, por hipótesis. En virtud del Teorema 3.3.3, en concreto su primer punto, deducimos que debe existir un Q -punto focal en el intervalo $(0, 1/k]$. ■

El resultado previo puede extenderse de hipersuperficies a subvariedades con dimensión $n - 2$, siendo la demostración completamente análoga salvo detalles que pueden consultarse en la referencia mencionada anteriormente.

Proposición 3.3.1 *Dados una variedad de Lorentz N y dos puntos p y q que pueden unirse a través de una curva causal diferenciable a trozos γ que no es pregeodésica causal, entonces existe una variación \mathbf{x} de γ con extremos fijos tal que todas las curvas longitudinales son temporales menos la propia γ . En concreto existen curvas temporales uniendo ambos puntos arbitrariamente cerca de γ .*

Teorema 3.3.5 *Dada una subvariedad espacial Q de una variedad de Lorentz N , si $\gamma \in \Omega(Q, q)$ es causal entonces existe una curva temporal arbitrariamente cerca de γ en $\Omega(Q, q)$ a no ser que γ sea una geodésica nula Q -normal en la cual no hay puntos focales de Q antes que q .*

Capítulo 4

Introducción a la Teoría General de la Relatividad

Llegados a este punto del documento, habiendo ya presentado todo un formalismo sobre variedades diferenciables, geometría semi-Riemanniana y en concreto sobre variedades de Lorentz, el lector podría extrañarse de que ahora vayamos a presentar un capítulo puramente físico, aunque se introduzca con carácter matemático. Siendo claros en este sentido, para realizar un desarrollo puramente matemático de la materia que aquí se pretende abordar — la teoría de causalidad en variedades Lorentzianas — este capítulo es completamente innecesario y el lector con puros intereses matemáticos podría ignorarlo y centrarse directamente en el último capítulo del escrito. Sin embargo, considero clarificador y lógico, siguiendo la línea histórica en la que surgieron los resultados que se pretenden tratar, introducir aquí un primer acercamiento a la Teoría de la Relatividad General como motivación física para la búsqueda de los Teoremas de Singularidad que estudiaremos en la última sección; ya que, como veremos, dichos teoremas surgen de la necesidad de encontrar ciertas condiciones generales que aseguren la aparición de singularidades en el espacio-tiempo sobre el que se desarrolla esta teoría.

Esta última idea presentada nos guía hacia la esencia de la Teoría de la Relatividad General. Recuperando una frase típica de textos divulgativos, a diferencia del resto de teorías físicas en las que el “protagonista” son las partículas, en Relatividad General el actor principal será un espacio-tiempo tal y como los que hemos presentado en el capítulo anterior. En este sentido, la Relatividad General modelará fenómenos físicos a través de la geometría del espacio-tiempo haciéndose esencial todo el formalismo matemático sobre geometría semi-Riemanniana introducido en las secciones previas. En concreto, la Relatividad General es, a día de hoy, la única teoría física que consigue explicar la gravedad o, equivalentemente, la física a gran escala, de forma consistente.

Recalamos este último apunte porque, como es bien conocido, la teoría de la gravedad por antonomasia es la presentada por Isaac Newton en su Principia Mathematica (1687) basada en la Ley de Gravitación Universal. Sin embargo, esta teoría que, a día de

hoy sigue siendo usada y es buena aproximación en la mayor parte de problemas físicos observados, dejó de ser consistente con el resto de la física cuando, en 1905, Albert Einstein publicó su Teoría Especial de la Relatividad.

La Relatividad Especial surgió como solución a la inconsistencia de la mecánica clásica con la teoría electromagnética basada en las ecuaciones de Maxwell (1865), las cuales no se mostraban invariantes ante las transformaciones galileanas para cambiar de sistema de referencia. En este sentido, la Relatividad Especial, a partir de sus dos postulados, “las leyes de la física son las mismas en todos los sistemas de referencia inerciales” y “la velocidad de la luz en el vacío es una constante universal independiente del movimiento de la fuente de luz”, generó toda una nueva teoría de la mecánica en la que, se abandonaba la idea clásica de espacio, surgiendo el concepto de espacio-tiempo (en concreto, el espacio-tiempo de Minkowski) y en la que aparecían unas nuevas transformaciones que sustituían las de Galileo, las transformaciones de Lorentz [8]. En este contexto, la física no dispuso de una teoría de la gravedad consistente — nótese que los dos principios de la relatividad especial son completamente incompatibles con la teoría de Newton, por ejemplo, con la idea de que la gravedad sea una interacción instantánea — hasta que, en 1915, Albert Einstein publicó su Teoría de la Relatividad General cuyos principios estudiaremos en la primera sección del capítulo.

Para acabar con esta introducción al capítulo, remarcamos una cuestión técnica que no será esencial en las ideas principales del texto, pero que habrá que tener en cuenta a la hora de leer las ecuaciones correctamente. Como es usual en las ramas más teóricas de la física, en Relatividad General suele usarse un sistema de unidades especial con el fin de simplificar las constantes esenciales de la teoría en las ecuaciones. En el caso de la Relatividad General, este sistema de unidades se conoce como Sistema de Unidades Geométricas (o Geometrizadas) y esencialmente, consisten en hacer la velocidad de la luz (c) y la constante de gravitación universal (G) adimensionales e iguales a uno. Así, seguiremos este sistema de unidades, escribiendo las magnitudes con las unidades necesarias inducidas por esta consideración — el lector con mayor conocimiento en física reconocerá que estas unidades no son más que un caso menos restrictivo del Sistema de Unidades Naturales usadas en Teoría Cuántica de Campos.

Al igual que en el resto del documento, el desarrollo principal del capítulo se basa en [18], aunque en este caso ha sido combinado con textos como [8], [15], [22], [23] y [24]. Además, se acompaña el contenido con varios artículos de interés para el lector que desee profundizar en las ideas que se tratan.

4.1. Fundamentos físicos. Materia, gravedad y curvatura

Dedicaremos esta primera sección del capítulo a introducir los conceptos fundamentales y principios de la Teoría de la Relatividad General y a realizar una breve discusión

de los mismos con el fin de dar contexto a las secciones posteriores. Tal y como hemos esbozado en la introducción al capítulo, podemos entender la Relatividad General (RG) como un estudio de fenómenos físicos — en particular, de la gravedad — a través de la geometría de una variedad general que hemos definido y denominamos espacio-tiempo (Definición 3.1.5). Nuestro principal problema ahora, antes de preocuparnos por el modo en que modelaremos la gravedad a través de la geometría espaciotemporal, es cómo podemos traducir todos los conceptos de la mecánica clásica a esta nueva variedad genérica; esto es, cómo definimos las partículas, los observadores, cómo podemos hablar de masa, energía, momento, etc. En este sentido, los lectores que estén familiarizados con la Relatividad Especial (RE) tendrán cierta ventaja, puesto que, esta no es más que un caso particular de la RG aplicada a un espacio-tiempo concreto: el espacio-tiempo de Minkowski \mathbb{R}_1^4 . De este modo, todas aquellas características que no dependan de \mathbb{R}_1^4 podrán ser extrapoladas a la teoría general. Comenzaremos pues haciendo un breve repaso o introducción de estos conceptos de la RE que hereda la RG. De aquí en adelante, con el fin de hacer la notación más compacta, denotaremos al espacio-tiempo como M , aunque sepamos que realmente es una dupla (M, g) orientada temporalmente. Recuérdesse además que llamaremos eventos a los puntos de M .

Los primeros conceptos que trataremos son los de partícula y observador. Como veremos a continuación, ambas ideas están íntimamente relacionadas; sin embargo, introduciremos primero el objeto de observador pues, de algún modo, puede resultar más natural pensar primero en el objeto que permite la observación que en la observación en sí misma.

Definición 4.1.1 (Observador) *Un observador en un espacio-tiempo M es una curva temporal orientada hacia el futuro $\alpha : I \subset \mathbb{R} \rightarrow M$ tal que $g(\alpha'(s), \alpha'(s)) = -1$ para todo $s \in I$.*

De la definición de observador es claro que la curva está parametrizada naturalmente siendo s la longitud de arco. Sin embargo, sin entrar en demasiado detalle, llamaremos *tiempo propio* al parámetro s por motivos físicos. A continuación, perfilaremos la idea de observador caracterizando su velocidad en cada instante de tiempo propio a través del concepto de observador instantáneo. Este concepto nos permitirá utilizar la descomposición ortogonal descrita en Lema 3.1.1, pudiéndose realizar un estudio completo de las partículas como veremos seguidamente.

Definición 4.1.2 (Observador instantáneo) *Un observador instantáneo en un espacio-tiempo M es un par (p, V) con $p \in M$ y $V \in T_p(M)$ un vector unitario temporal que apunta al futuro.*

En este sentido, tal y como hemos dicho, considerando un observador instantáneo (p, V) podemos descomponer el espacio tangente como sigue

$$T_p(M) = \langle V \rangle_{\mathbb{R}} \oplus V^{\perp},$$

donde hemos denotado con $\langle V \rangle_{\mathbb{R}}$ al espacio generado por el vector V con coeficientes reales (la recta vectorial generada por V).

Con esta descomposición, podemos realizar una descripción $3 + 1$ dimensional de las partículas; esto es, un observador puede estudiar una partícula descomponiendo su movimiento en los dos espacios dados. Por un lado tendríamos la contribución espacial tridimensional de V^{\perp} que, básicamente, es la sección de espacio Euclídeo que ve el observador en un instante de tiempo propio — llamaremos a este espacio el espacio en reposo del observador. Y, por otro lado, tendremos la contribución de $\langle V \rangle_{\mathbb{R}}$, que nos dará el avance en el tiempo — llamaremos a este espacio el eje temporal del observador. Aunque para el resto de nuestro desarrollo bastará con las ideas ya expuestas, en vista de esta descomposición y dado que es común en la literatura, recalcaremos también la idea de observador a través del concepto de *sistema de referencia*. Así pues, en este contexto, se suele definir un sistema de referencia O como un observador instantáneo (p, V) junto con una base ortonormal de V^{\perp} . Podemos entender por tanto el observador O como la curva o trayectoria generada por el observador instantáneo (p, V) vista en el sistema de referencia O — nótese que se abusa de la notación, nos referimos a la base a través del propio sistema de referencia.

Una vez tratada la manera en que observaremos en nuestro espacio-tiempo, tratemos los propios observables: las partículas.

Definición 4.1.3 (Partícula) *Una partícula en un espacio-tiempo M es una curva temporal orientada hacia el futuro $\rho : I \subset \mathbb{R} \rightarrow M$ tal que $g(\rho'(t), \rho'(t)) = -m^2$ con $m \geq 0$ para todo $t \in I$. Denominaremos masa en reposo a la constante m , de modo que llamaremos partículas sin masa a las partículas con $m = 0$ y partículas materiales a aquellas con $m > 0$.*

Es ahora inmediato ver la relación tan estrecha entre partícula y observador que adelantábamos al principio de la sección. Así, partícula y observador son formalmente el mismo objeto solo que con parametrizaciones diferentes. Es claro, por tanto, que podríamos haber introducido ambos objetos en una sola definición. Podemos interpretar un observador como una partícula de masa unidad, esto es, con su parametrización natural o, del modo contrario, podemos interpretar una partícula como un observador al que se le asigna una masa m .

Esta separación explícita que se ha llevado a cabo en la exposición tiene como finalidad advertir del sentido que tiene la masa introducida de este modo. A pesar de que esta constante m nos permita distinguir entre los diferentes tipos de partículas y sea un reflejo de la realidad experimental, el hecho de añadirla en cierto sentido, a posteriori, no confiere de ningún sentido físico a la definición de partícula en relación al espacio-tiempo sobre la que esta vive. Este problema que, en principio, nos puede parecer insignificante, cobrará una gran importancia más adelante cuando queramos modelar las fuentes de gravedad a través de la materia.

Por otro lado, tras la definición de partícula, se hace mucho más claro el sentido físico de la descomposición ortogonal previa. Realizando la descomposición de un vector tangente de la curva que representa la partícula sobre un punto de la misma, podemos interpretar que en cada instante de tiempo la partícula vive en un espacio tridimensional diferente que va cambiando a medida que avanza en su eje temporal. Además, esta misma descomposición nos permite caracterizar la energía y el momento de una partícula. Usaremos negrita para denotar cantidades vectoriales y reservaremos la letra común para su norma.

Definición 4.1.4 (Energía-momento) *Dados una partícula ρ de masa m con $\rho(t) = p$ para algún t y un observador instantáneo (p, V) sobre un espacio-tiempo M , definimos la energía-momento de ρ como su vector velocidad $\rho'(t)$ identificándose respectivamente la energía y el momento con los únicos elementos $E \in \mathbb{R}$ y $\mathbf{P} \in V^\perp$ que permiten la descomposición ortogonal $\rho'(t) = EV + \mathbf{P}$. Además, la velocidad tridimensional de la partícula medida por dicho observador vendrá dada por $\mathbf{v} = \frac{\mathbf{P}}{E}$.*

Esta definición es consistente con las definiciones clásicas de energía y momento de RE

$$E = \frac{m}{\sqrt{1-v^2}}, \quad \mathbf{P} = \frac{m\mathbf{v}}{\sqrt{1-v^2}}, \quad (4.1)$$

donde recuérdese que no aparecen las típicas c^2 porque estamos tomando unidades geométricas. La demostración de la compatibilidad de estas expresiones con la Definición 4.1.4 surge de evaluar $g(\rho'(t), \rho'(t))$ que fácilmente, sin más que tener en cuenta que V y \mathbf{P} son ortogonales, nos lleva a la conocida expresión: $E^2 = m^2 + P^2$. A través de esta ecuación y la relación entre energía, momento y velocidad de la definición, podemos obtener de forma directa las ecuaciones (4.1).

Analizando las expresiones (4.1) junto con la relación $E^2 = m^2 + P^2$ podemos deducir algunas implicaciones físicas. En primer lugar, es directo de la expresión para la energía que, incluso una partícula en reposo ($v = 0$), tiene una energía asociada $E = m$ denominada energía de reposo — sin usar unidades geométricas surge la archiconocida ecuación $E = mc^2$. En un sentido parecido, usando la relación de las normas al cuadrado, podemos deducir que incluso una partícula sin masa posee energía $E = P$. Finalmente, despejando v de la expresión para la energía podemos deducir que $v \leq 1$ lo que implica que ninguna partícula puede viajar más rápido que la luz¹; en particular, las partículas con $v = 1$ son aquellas con $m = 0$ o lo que es lo mismo $E = P$. Además, interpretando en este contexto la expresión de la energía, podemos ver cómo ir aumentando la velocidad de una partícula para aproximarla a 1 — o lo que es lo mismo a c — requeriría cada vez más energía hasta que esta se indeterminase divergiendo a infinito. De este último límite para la velocidad corroboramos la invalidez de la teoría

¹Nótese que toda esta disquisición se ha llevado a cabo bajo la suposición de que $m \in \mathbb{R}$, condición que tiene sentido dadas las experiencias realizadas hasta el presente. En la actualidad, ya existen teorías cuánticas con partículas superlumínicas llamadas taquiones. Sin embargo, estas requerirían de una masa imaginaria.

de Newton que ya adelantamos en la introducción al capítulo (recuérdese que la teoría gravitatoria de Newton suponía que la interacción de la gravedad se “propagaba” de forma instantánea a distancias arbitrarias por el espacio).

Para acabar con esta primera parte de la sección y con las relaciones esenciales entre RE y RG, podemos observar a partir de todos los resultados anteriores cómo RE no es solo un caso particular de RG sino una muy buena aproximación de RG a nivel local. Nótese que, en el mismo sentido que durante toda la sección, fijado $p \in M$ RE tendrá sentido en $T_p(M) \approx \mathbb{R}_1^4$ a través del isomorfismo lineal canónico. Por tanto, tal y como ya vimos en el Capítulo 2, en entornos suficientemente pequeños M estará bien aproximado — siempre que la curvatura lo permita — por $T_p(M)$ pudiéndose comparar ambos a través de la aplicación exponencial en el punto (Definición 2.3.5). Podemos concluir por tanto que RG y RE coinciden en buena aproximación cuando la curvatura es despreciable, que, como veremos querrá decir que la gravedad es despreciable. En la misma línea, la teoría clásica de Newton será buena aproximación cuando la gravedad sea muy débil y las velocidades sean pequeñas comparadas con la de la luz.

Hasta ahora, hemos modelado todos los conceptos necesarios para hablar de movimiento y mecánica, pero en ningún momento hemos hablado de cómo la propia geometría del espacio-tiempo modela la gravedad, lo que, según habíamos adelantado, era la esencia de RG.

La RG se basa esencialmente en tres principios: los dos principios de relatividad enunciados en la introducción al capítulo y el principio de equivalencia fuerte. Nos centramos en este último pues nos da las claves esenciales de la teoría. Existen muchas formas de enunciar este principio; presentamos aquí una de las más formales y técnicas:

El movimiento gravitacional de un cuerpo de prueba depende únicamente de su posición inicial en el espacio-tiempo y no de su constitución, y el resultado de cualquier experimento local, gravitacional o no, en un laboratorio moviéndose en un sistema de referencia inercial es independiente de la velocidad del laboratorio y de su localización en el espacio-tiempo.

A priori, el enunciado nos puede parecer bastante difuso o abstracto. Sin embargo, su interpretación es simple y está directamente relacionada con la relación entre RE y RG que exponíamos antes: para un observador en caída libre — entendiéndose, por ahora, caída libre como el movimiento influido únicamente por la gravedad — en un entorno lo suficientemente pequeño del espacio-tiempo, todas las leyes de la física obedecen las leyes de RE. En particular, esto quiere decir que un sistema sometido a un campo gravitatorio es indistinguible de un sistema uniformemente acelerado o, equivalentemente, estar en reposo es indistinguible de estar cayendo libremente en el seno de un campo gravitatorio. Para entender mejor esta idea, desarrollamos un experimento mental típico en la literatura y propuesto por el propio Einstein:

Supóngase que estamos subidos en un ascensor en un rascacielos en su piso más alto y que en un instante determinado el ascensor comienza a caer. Si tomáramos cualquier

objeto de nuestro bolsillo y lo soltáramos en el ascensor, este empezaría a caer a la vez que nosotros, quedándose “estático” en el lugar donde lo soltamos. Para nosotros, al menos en nuestro entorno más cercano (localidad), no “existe” campo gravitatorio. Desde el punto de vista contrario, supóngase que ahora llevamos el ascensor a un punto suficientemente alejado de todo en el espacio de modo que no hay campo gravitatorio. Si ahora se colocase algún tipo de mecanismo en la parte superior del ascensor que tirara de este hacia arriba con una fuerza constante, es claro que si volvemos a soltar ahora el mismo objeto este sí cae. Ahora sí “existiría” campo gravitatorio para nosotros.

Del experimento mental podemos concluir nuestra incapacidad para distinguir entre un campo gravitatorio y estar uniformemente acelerados o en reposo. Por tanto, el campo gravitatorio es relativo, extendiéndose de este modo el principio de relatividad a la teoría gravitatoria tal y como necesitaba la física desde la ruptura entre la teoría de Newton y la RE. Ahora bien, la pregunta que surge naturalmente es cómo podemos relacionar esta implicación con el modelado de la gravedad a través de la geometría del espacio-tiempo. La respuesta es directa, pues, si según el principio de equivalencia somos incapaces de diferenciar entre los estados anteriormente expuestos, por ser estos experimentalmente indistinguibles, podemos afirmar que toda caída libre será una geodésica. Recordemos que, en la teoría clásica de Newton, al desarrollarse sobre el espacio Euclídeo, las geodésicas eran líneas rectas y correspondían a las curvas de las partículas que se movían libremente sin estar sujetas a una fuerza o aceleración — nótese que las geodésicas, tal y como recalcamos en su definición (Definición 2.3.3), son en cierto modo curvas con aceleración nula. La idea esencial es por tanto que, en RG, las partículas sometidas a un campo gravitatorio — en caída libre — se moverán también a través de las geodésicas — puesto que para un observador que cae con la partícula su aceleración es nula — solo que ahora, en lugar de ser rectas, serán curvas determinadas por cómo la materia — la fuente de gravedad — curve el espacio-tiempo en el que se encuentran.

Para resumir todas las ideas expuestas en el párrafo anterior volvamos a nuestro experimento mental. Supongamos ahora que tenemos nuestro ascensor con dos observadores A y B en las siguientes situaciones: A estará cayendo en ángulo hacia el Sol por su campo gravitatorio y, como ocurría antes, B estará en un punto del espacio alejado sin campo gravitatorio. Al igual que antes, ambas situaciones son indistinguibles por el principio de equivalencia. Y, además, podemos ver cómo ambos se moverán sobre geodésicas. Para B, el movimiento es recto — tanto para él como para un observador que lo mire desde fuera — pues no hay presencia de gravedad; sin embargo, en el caso de A, el movimiento es curvo visto desde fuera pues el Sol curva el espacio-tiempo haciendo que la geodésica por la que se mueve el ascensor sea curva, aunque, para el propio A, la sensación sea la de estar moviéndose sobre una recta al ser su situación indistinguible de la de B.

En vista de estas ideas planteamos la siguiente definición que sustituye la idea inicial de caída libre:

Definición 4.1.5 (Caída libre) *Una partícula $\rho(t)$ en un espacio-tiempo sufre una caída libre si $\rho(t)$ es geodésica, equivalentemente; si*

$$\frac{D\rho'}{dt} = 0.$$

Esta relación entre caída libre y geodésica se puede probar directamente a través del principio de equivalencia como hizo Weinberg en [26]; sin embargo, hemos preferido realizar esta discusión de forma cualitativa por su carácter ilustrativo.

Hemos llegado por tanto, como habíamos adelantado, a que en RG la gravedad es la curvatura del espacio-tiempo generada por la energía-momento (generalización de la masa). Nos queda pues un único problema en el aire; entender cómo la materia puede provocar la curvatura del espacio-tiempo. Nótese que esta dificultad surge, como ya destacamos, del modo en que definimos partícula. El hecho de añadir a posteriori la idea de masa — implícitamente la energía-momento — no define como tal la materia ni le da significado físico a la masa en relación al espacio-tiempo en el que esta se encuentra. Necesitamos por tanto definir un objeto matemático que modele la materia y que le permita perturbar de algún modo la geometría del espacio-tiempo: un tensor energía-momento. Tomando una vez más un observador instantáneo (p, V) , es lógico esperar que, en la parte espacial de su descomposición ortogonal, este tensor se comporte como el tensor tensión-energía clásico [13], [14] por lo que, en analogía, podemos plantear el tensor energía-momento como un tensor simétrico $(0,2)$ — nótese que esta idea también es consistente con el hecho de que este tensor esté relacionado con el tensor métrico (geometría de M) que también es $(0,2)$. En base a esta analogía clásica planteamos la siguiente definición:

Definición 4.1.6 (Tensor energía-momento) *Un tensor energía-momento T en un espacio-tiempo M es un campo tensorial $(0,2)$ simétrico cumpliendo que:*

1. $T(u, u) \geq 0$ para todo vector causal $u \in T_p(M)$ en todo $p \in M$.
2. $\text{div}T = 0$.

Las dos condiciones impuestas en la definición son mero producto de la observación experimental y el significado físico del tensor. Así, la condición 1 solo nos indica la positividad de toda forma de energía, de acuerdo con todas las experiencias realizadas hasta el presente, y la condición 2 es la generalización en forma tensorial de la conservación de la energía.

4.2. Ecuación de campo de Einstein

Hasta este momento hemos desarrollado todos los conceptos básicos y los principios generales de la RG llegando a la idea esencial de la teoría: la gravedad es la curvatura del

espacio-tiempo M producida por la energía-momento. En este contexto, hemos llegado también a la idea de tensor energía-momento T con la finalidad de modelar la materia a través de un objeto matemático que pueda perturbar la geometría del espacio-tiempo. Llegados a este punto, para cerrar la teoría, solo nos falta encontrar el modo en que este tensor se relaciona con la geometría de M , esto es, con su tensor métrico g . Es decir, al igual que ocurría en la teoría de Newton con la Ley de Gravitación Universal o la ecuación de Poisson, necesitamos una ecuación que nos permita describir la gravedad a través de la relación entre T y g — nótese que nos referimos a la ecuación de Poisson puesto que g , en cierto modo, actuaría como el potencial gravitatorio clásico.

La existencia de esta ecuación en RG no está asegurada. De hecho, no hay un método directo de derivación de esta ecuación como ocurre con sus análogas en la teoría clásica. Esto confiere de una gran dificultad al problema de encontrar qué ecuaciones son las correctas para la RG. Einstein, en su carrera con Hilbert para encontrar una ecuación de campo válida [4], propuso de forma heurística una serie de características principales que tenía que cumplir esta ecuación. Su idea principal era la existencia de un tensor que cumpliera, por simplicidad, una cierta relación de proporcionalidad con T . Este tensor, por motivos de predictibilidad, determinismo y por analogía con la ecuación de Poisson (antecedente clásico), no debería de contener derivadas del tensor métrico de orden superior a dos y debería de depender linealmente de las derivadas segundas del tensor métrico, lo cual nos puede llevar a la intuición de que pueda construirse como alguna variación del tensor de Ricci. Por otro lado, por estar relacionado con T , debido a la conservación de la energía, debemos de exigirle a G que tenga divergencia nula y además que sea simétrico y de tipo $(0,2)$.

En vista de estas ideas, y teniendo en cuenta todo el formalismo ya expuesto en los primeros capítulos, un buen candidato para G es el tensor que proponemos en la siguiente definición, que, de hecho fue el que propuso Einstein en su día:

Definición 4.2.1 (Tensor de Einstein) *Llamaremos tensor de Einstein en un espacio-tiempo M al campo tensorial:*

$$G = Ric - \frac{1}{2}Rg.$$

Nótese que este G es una buena propuesta porque, en vista de la Proposición 2.5.1, sabemos que $2 \operatorname{div} Ric = dR$ por lo que la combinación de la Definición 4.2.1 entre el tensor de Ricci Ric y la curvatura escalar R debería de proporcionar un tensor con las características adecuadas. En cualquier caso, en la siguiente proposición probamos que el tensor de Einstein cumple todas las características que habíamos exigido:

Proposición 4.2.1 *El tensor de Einstein es un campo tensorial simétrico, de tipo $(0,2)$ y con divergencia nula. Además, contiene la misma información que el tensor de Ricci pues podemos expresar este último como:*

$$Ric = G - \frac{1}{2}C(G)g.$$

En concreto, el tensor de Ricci y el tensor de Einstein tienen la misma expresión formal en términos del otro.

Demostración: El hecho de que G sea simétrico y (0,2) es consecuencia directa de que tanto g como Ric también lo sean. Para probar que tiene divergencia nula, la computamos directamente a partir de las divergencias de sus componentes por linealidad

$$div G = div Ric - \frac{1}{2}div(Rg) = \frac{1}{2}(dR - div(Rg)) = \frac{1}{2}(dR - dR) = 0,$$

donde se ha tenido en cuenta que $2 div Ric = dR$ por la Proposición 2.5.1 y que $div(Rg) = dR$. Probemos esto último realizando también un desarrollo directo de la divergencia de Rg :

$$\begin{aligned} div(Rg) &= \mathbf{C}_{23}(\nabla(Rg)) = \mathbf{C}_{23}((\nabla R)g + R(\nabla g)) = \mathbf{C}_{23}((dR)g) \equiv \\ &\equiv (\mathbf{C}_{23}((dR)g))_\alpha = g^{\mu\nu} g_{\alpha\mu} R_{;\nu} = \delta_\alpha^\nu R_{;\nu} = R_{;\alpha} \equiv dR. \end{aligned}$$

Finalmente, veamos la expresión para Ric . Trabajemos a la inversa y veamos qué es $\mathbf{C}(G)$. A partir de la Definición 4.2.1:

$$\mathbf{C}(G) = \mathbf{C}(Ric) - \frac{1}{2}R\mathbf{C}(g) \equiv R - \frac{1}{2}Rg^{\mu\nu}g_{\mu\nu} = R - \frac{4}{2}R = -R.$$

Nótese que, por definición, la contracción del tensor métrico será siempre la dimensión de la variedad donde este se define. Despejando directamente de la Definición 4.2.1 y sustituyendo el resultado:

$$Ric = G + \frac{1}{2}Rg = G - \frac{1}{2}\mathbf{C}(G)g.$$

Además, la equivalencia formal de las expresiones es directa pues, teniendo en cuenta que el valor de la contracción del tensor de Ricci es R :

$$G = Ric - \frac{1}{2}\mathbf{C}(Ric)g.$$

Concluyéndose así el resultado. ■

Una vez construido el tensor G deseado, solo requerimos encontrar su relación de proporcionalidad con T . Tal y como comentamos antes, no hay ninguna manera de demostrar formalmente o derivar esta relación; simplemente podemos basarnos en la experiencia para encontrar el valor adecuado de la constante de proporcionalidad. Esto es exactamente lo que hizo Einstein, realizar varias pruebas comparando con la realidad experimental y así determinar tanto la forma de G como la constante de proporcionalidad entre este y T , dando lugar a las ecuaciones de campo:

Definición 4.2.2 (Ecuación de campo de Einstein) *En un espacio-tiempo M con materia modelada por un tensor energía-momento T , la relación entre este tensor y el tensor de Einstein G viene dada por la ecuación de campo de Einstein:*

$$G = Ric - \frac{1}{2}Rg = 8\pi T.$$

O equivalentemente, en notación indexada

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi T_{\mu\nu},$$

con μ y ν variando de 0 a 3.

Estas ecuaciones — y decimos ecuaciones, puesto que a pesar de la simplicidad que puedan aparentar y parecer una sola ecuación, la notación indexada nos indica el número de ecuaciones — modelan exactamente lo que queríamos: la energía-momento a través del tensor T provoca una curvatura en el espacio-tiempo modelada a través de G . La gravedad, a través de todos los objetos infinitesimales e instantáneos que hemos desarrollado, ha pasado de ser una fuerza a convertirse en una consecuencia intrínseca de la geometría del espacio-tiempo donde se produce el movimiento. La constante de proporcionalidad propuesta, a pesar de no poder ser comprobada exactamente, ha sido validada a través de numerosas comprobaciones experimentales como el perihelio de la órbita de Mercurio o la desviación de rayos de luz por objetos masivos (lentes gravitacionales) — nótese que la luz es afectada por la gravedad ya que aunque no tenga masa tiene energía-momento. Además, esta constante se consolidó forzando a que la RG en su límite en gravedad débil y velocidades pequeñas respecto a c fuera equivalente a la teoría de Newton.

Recalamos además que, aunque como hemos dicho estas ecuaciones no puedan derivarse de ningún principio, la forma de introducirlas no es siempre la que aquí se ha elegido. Es común en la literatura que las ecuaciones de Einstein, en vez de presentarse como una definición, sean derivadas a través del Principio de Mínima Acción. Esto puede parecer contradictorio con lo que acabamos de comentar. Sin embargo, el Principio de Mínima Acción, como su propio nombre indica, se basa en buscar mínimos de un funcional llamado acción, de modo que para poder buscar dichos mínimos primero debemos de construir este funcional. Es aquí donde vuelve nuestro problema, puesto que, al igual que no hay un camino único y formal para llegar a las ecuaciones de Einstein tal y como las hemos presentado en este documento, no existe tampoco dicho camino para construir este funcional — funcional de Einstein-Hilbert [4] — de modo que tendremos que volver a recurrir a la experiencia y a técnicas heurísticas para determinar dicho objeto.

Para acabar con la sección, haremos una breve discusión de cómo funciona la ecuación de campo y algunas implicaciones que podemos deducir de la misma. A través de la ecuación de campo se genera una retroalimentación dinámica entre materia y espacio-tiempo. En este sentido, la materia le dice al espacio-tiempo cómo curvarse

y el espacio-tiempo le dice a la materia cómo moverse. Obsérvese que este es el sentido natural en que las ecuaciones de Einstein deben usarse; construimos un tensor energía-momento que introducimos en la ecuación de campo que nos devuelve cómo el espacio-tiempo se curva y a través de la condición $div T = 0$ deducimos cómo el espacio-tiempo le dice a la materia cómo moverse. Si se intentara usar la ecuación de campo en sentido contrario, práctica no tan poco común en artículos actuales (véase [1]), tenemos que percatarnos siempre de que el tensor energía-momento resultante tenga sentido físico, pues no tiene por qué tenerlo en general. Sin embargo, siempre que esto se tenga en cuenta, este segundo método de proponer una métrica y encontrar un tensor energía-momento compatible es plausible ².

Finalmente, recalcamos que las ecuaciones de campo nos permiten entender mejor el significado físico de T y nos proporcionan un modo de definir el vacío:

Definición 4.2.3 (Vacío) *Diremos que un espacio-tiempo M es un vacío (o que está vacío) cuando es Ricci llano; equivalentemente, cuando $T \equiv 0$.*

A partir de esta definición, podemos ver que aunque no consideremos materia el espacio-tiempo presentará curvatura, puesto que será Ricci llano — nótese que si T es cero lo será G y en consecuencia por la Proposición 4.2.1 lo será Ric — pero esto no implica que M sea llano. En ocasiones, el concepto de vacío se usa en la ecuación de campo para determinar la geometría del espacio-tiempo “fuera” del objeto que genera la gravedad, como por ejemplo en la solución de Schwarzschild para una estrella aislada que comentaremos muy brevemente en la siguiente sección. Podemos ver además cómo en T es despreciable la contribución de las partículas cuyo movimiento es determinado a través de las ecuaciones previas.

4.3. Motivación para el estudio de la Causalidad: El Modelo FLRW y las singularidades

Para acabar el capítulo, introducimos esta sección sobre modelos solución de las ecuaciones de campo que nos permitirán hacer un primer acercamiento a la idea de singularidad. Presentaremos esencialmente un desarrollo del modelo Friedmann-Lemaître-Robertson-Walker (FLRW) aunque haremos también una breve mención al modelo de Schwarzschild. Ambos modelos, a pesar de su gran importancia y uso en la física actual, son dos de las soluciones más simples de las ecuaciones de campo. Además, aunque son soluciones completamente diferentes — la solución FLRW modela el universo como un todo mientras que la solución de Schwarzschild modela el campo gravitatorio producido

²Nótese además que aunque no sea del todo relevante en nuestro desarrollo, tal y como adelantábamos en el Capítulo 2, las ecuaciones de campo no tienen solución única. Por ello en el fondo estamos trabajando con clases de equivalencia a través de isometrías como en la Definición 2.1.4. Véase [20]

por un cuerpo esférico masivo — ambas se encuentran directamente ligadas a través del concepto de singularidad.

Adelantándonos un poco en nuestro desarrollo, podemos entender, por ahora, una singularidad como un punto donde el espacio-tiempo deja de tener sentido y en cierto modo la RG se rompe, i.e. las singularidades constituyen los límites de la teoría. Así, como veremos, en estas primeras soluciones simples de las ecuaciones de campo aparecen este tipo de anomalías que limitan el uso de la RG a ciertas regiones del espacio-tiempo. Este hecho particular, que podríamos permitirnos catalogar incluso de sorprendente, será la motivación esencial del desarrollo de la Teoría de la Causalidad enfocado a los Teoremas de Singularidad y el nexo puro entre esta teoría de carácter puramente matemático y la propia física de la Relatividad General.

El modelo FLRW, tal y como hemos adelantando, es un modelo cosmológico, esto es, modela el universo como un todo. Además, es un modelo cosmológico dinámico, indicándonos que el espacio-tiempo no es estático sino que podría sufrir expansiones o contracciones. Esta idea iba completamente en contra de las intuiciones que tuvo Einstein al trabajar inicialmente con la ecuación de campo con un enfoque cosmológico. A pesar de la forma de las ecuaciones de campo que hemos introducido en la Definición 4.2.2, Einstein pensaba que el universo debía de ser completamente estático en el tiempo, lo que le llevó a incluir en sus ecuaciones la llamada constante cosmológica Λ , resultando

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi T_{\mu\nu},$$

con μ y ν de 0 a 3 como en la ecuación que presentábamos inicialmente.

Paralelamente, primero Friedmann entre 1922 y 1924, e independientemente, Lemaître en 1927, publicaban sus soluciones dinámicas a las ecuaciones de campo. Estas soluciones no fueron validadas hasta 1929, cuando el astrónomo estadounidense Edwin Hubble probó experimentalmente la expansión del universo a través del corrimiento al rojo de las galaxias. Fue en este año cuando Einstein, según la literatura, admitió haber cometido el mayor error de su vida y desechó la constante cosmológica³. Posteriormente, en los años 30, los matemáticos Robertson y Walker consolidaron teóricamente el modelo, dando lugar a lo que hoy conocemos como modelo Friedmann-Lemaître-Robertson-Walker debido a sus cuatro contribuidores.

Para construir estas soluciones dinámicas del universo, tendremos que usar uno de los procedimientos de resolución de las ecuaciones de campo que comentábamos en la Sección 4.2. Así pues, o construimos un tensor energía-momento y determinamos la geometría, o proponemos una geometría y determinamos si el tensor energía-momento es compatible físicamente. En el caso del modelo FLRW hay cierto paralelismo entre ambos métodos como veremos seguidamente.

³En cosmología actual la constante cosmológica ha sido recuperada en las ecuaciones de campo con el fin de modelar lo que se conoce como energía oscura, tipo de energía exótica a la que se le atribuye la expansión del universo. Este modelo se conoce actualmente como modelo Λ CMD puesto que es un modelo FLRW pero con constante cosmológica Λ y materia oscura fría (cold dark matter) [6].

La idea esencial del modelado en la solución FLRW se basa en el principio cosmológico, principio experimentalmente confirmado que nos indica que el universo a gran escala es homogéneo e isótropo y que, por tanto, la forma en que se distribuyen las galaxias es en cierto modo uniforme. Este principio, junto con otras experiencias, nos permite proponer modelos simples tanto para la forma de la variedad espacio-temporal como para la materia que haya en ella.

Así, la materia se puede modelar esencialmente como un fluido ideal cuyas “partículas” serían las galaxias y cuyo flujo (velocidad) vendría dado por un campo vectorial temporal U . En cuanto a la variedad, se propone lo que se conoce como un universo bloque, esto es, un producto $I \times S$ con I un intervalo abierto de \mathbb{R} que representa el tiempo y S una variedad tridimensional conexa que representa el espacio — recuérdese el análisis del movimiento en $3 + 1$ dimensiones que hicimos en la Sección 4.1.

De este modo, parametrizando las galaxias como curvas en el tiempo, es decir, en la forma $I \times p$ con $p \in S$, dado que U — nótese que $U = \partial_t$ puesto que es d/dt en I pero visto en $I \times S$ — representa su velocidad, es claro que las galaxias son las curvas integrales de U y que el tiempo definido a través de I representa el tiempo propio de las galaxias (observadores), de modo que U debe ser unitario. Además, por el principio cosmológico, podemos suponer que el espacio en reposo de todas las galaxias es el mismo, U es perpendicular a $S(t)$ para cada $t \in I$, puesto que su movimiento a gran escala es despreciable; además, buscaremos que S tenga curvatura k constante y que cada aplicación inyectiva de S a $S(t)$ con $t \in I$ fijo sea una homotecia con escala $a(t)$ — véase [18] para un análisis más detallado de estas últimas ideas.

Todas estas condiciones nos conducen a la idea de que el objeto matemático más adecuado para describir el espacio-tiempo bloque del modelo FLRW es un producto deformado (Definición 2.1.7). Esto nos lleva a la siguiente definición:

Definición 4.3.1 (Espacio-tiempo de Robertson-Walker) *Dada una variedad Riemanniana conexa tridimensional S con curvatura (seccional) constante $\kappa \in \{-1, 0, 1\}$ y dada una función diferenciable $a > 0$ definida en un intervalo abierto $I \subseteq \mathbb{R}_1^1$, llamaremos espacio-tiempo de Robertson-Walker al producto deformado:*

$$M(\kappa, a) = I \times_a S.$$

Nótese por tanto que la idea de universo bloque es exactamente la interpretación $3 + 1$ dimensional que hacíamos sobre $I \times S$ solo que tendremos que considerar esta variedad con elemento de línea $-dt^2 + a(t)^2 d\sigma^2$ donde hemos denotado por dt^2 y $d\sigma^2$ a los elementos de línea de cada variedad por separado. Entendemos por tanto el universo como una sucesión de secciones espaciales a través del tiempo solo que la geometría en cada sección espacial queda engarzada con el tiempo a través de la función $a(t)$. En este sentido, la función $a(t)$ modela la expansión o contracción del universo con el tiempo, es decir, su dinámica, por lo que la llamaremos factor de escala. Su interpretación es directa: $a(t)$ valdrá exactamente uno en el instante actual, un medio cuando el universo

era la mitad de grande, etc. Diremos además que I es maximal cuando a no pueda extenderse a una función diferenciable en un intervalo más grande.

En cuanto a la variedad S (la fibra), que en general se llamará espacio, suelen tomarse de forma estándar \mathbb{R}^3 , \mathbb{S}^3 y \mathbb{H}^3 debido a que son los simplemente conexos y a que tienen curvatura constante, 0, 1 y -1 respectivamente (Definición 2.4.5). Recalcamos finalmente que para que el espacio-tiempo de la Definición 4.3.1 sea en rigor un espacio-tiempo, en las condiciones del modelo FLRW que describíamos antes, el flujo de las galaxias U tendrá que estar dirigido al futuro.

Recapitulando, podemos ver que en nuestro proceso de construcción de la solución para las ecuaciones de campo, hemos propuesto el espacio-tiempo, en concreto el de Robertson-Walker, y además hemos propuesto que la materia tiene que ser un fluido perfecto.

Ahora bien, puede parecer que estamos siguiendo los dos procesos de resolución de la ecuación de campo a la vez. Sin embargo, no es esto lo que ocurre. Estamos siguiendo concretamente la ruta de proponer la geometría y deducir la compatibilidad física del tensor energía-momento, solo que estamos exigiendo que el tensor energía-momento tenga la “forma” de un fluido perfecto, pero no estamos construyendo o proponiendo explícitamente un tensor energía-momento concreto. Veamos exactamente qué entendemos en este contexto por fluido ideal para entender mejor lo que está ocurriendo.

Definición 4.3.2 (Fluido ideal o perfecto) *Un fluido ideal o perfecto en un espacio-tiempo M es una terna (U, ρ, \mathcal{P}) cumpliendo:*

- U es un campo vectorial temporal unitario en M que apunta al futuro y que denominaremos campo vectorial de flujo o simplemente flujo cuando no conduzca a confusión.
- ρ y \mathcal{P} son funciones diferenciables en M que denominaremos densidad de energía y presión, respectivamente.
- Denotando por U^* a la 1-forma métricamente equivalente a U , la terna de fluido ideal induce un tensor energía-momento de la siguiente forma:

$$T = (\rho + \mathcal{P})U^* \otimes U^* + \mathcal{P}g.$$

Es claro que el flujo de la definición anterior es exactamente igual al que proponíamos durante toda la sección. Por otro lado, podemos interpretar físicamente la forma del tensor energía-momento anterior. Nótese que la forma del tensor nos da tres propiedades esenciales que caracterizan un fluido ideal — si un tensor cumple estas propiedades tendrá la forma de la Definición 4.3.2. Si tomamos un observador que se mueva con el flujo, este medirá la densidad de energía

$$T(U, U) = (\rho + \mathcal{P})U^* \otimes U^*(U, U) + \mathcal{P}g(U, U) = (\rho + \mathcal{P}) - \mathcal{P} = \rho,$$

donde hemos usado la definición de U^* y que U es temporal y unitario.

Por otro lado, si tomamos campos vectoriales X e Y perpendiculares a U , $T(X, Y)$ nos dará la componente Y de la presión a través de la dirección X ,

$$T(X, Y) = (\rho + \mathcal{P})U^* \otimes U^*(X, Y) + \mathcal{P}g(X, Y) = \mathcal{P}g(X, Y),$$

donde hemos usado la ortogonalidad.

Finalmente, si consideramos uno de estos campos y el propio U , obtendremos con $T(U, X)$ la densidad de energía en dirección X y con $T(X, U)$ la densidad de la componente X del momento. Sin más que considerar la ortogonalidad:

$$T(U, X) = (\rho + \mathcal{P})U^* \otimes U^*(U, X) + \mathcal{P}g(U, X) = 0,$$

$$T(X, U) = (\rho + \mathcal{P})U^* \otimes U^*(X, U) + \mathcal{P}g(X, U) = 0.$$

Resultado completamente consistente con el hecho de que estemos considerando un fluido ideal — recordamos que en física clásica un fluido ideal se caracteriza por ser no viscoso, estacionario, incompresible y sin vorticidad.

Una vez caracterizado el fluido ideal podemos volver a la resolución de las ecuaciones de campo. Podemos ver ahora que, claramente, como decíamos, estamos introduciendo en la ecuación de campo una métrica concreta y a partir de ella vamos a determinar un tensor energía-momento que tenga sentido físico y, además, le vamos a exigir que tenga la forma del tensor asociado a un fluido ideal de la Definición 4.3.2. En este sentido, nuestra tarea será encontrar qué funciones de densidad de energía y presión son compatibles con la métrica de Robertson-Walker y una vez consigamos esto tendremos el modelo solución: espacio-tiempo y tensor energía-momento.

Para demostrar esto último usaremos el siguiente lema cuya demostración podemos encontrar en [2]:

Lema 4.3.1 *En un espacio-tiempo de Robertson-Walker $M(\kappa, a)$ con flujo $U = \partial_t$:*

- *Dados X e Y campos vectoriales ortogonales a U , el tensor de Ricci vendrá dado por:*
 - $Ric(U, U) = -3a''/a$,
 - $Ric(U, X) = 0$,
 - $Ric(X, Y) = [2(a'/a)^2 + 2\kappa/a^2 + a''/a]g(X, Y)$.
- *La curvatura escalar vendrá dada por:*

$$R = 6[(a'/a)^2 + \kappa/a^2 + a''/a].$$

Apoyándonos en el lema, determinamos las funciones densidad y presión que necesitamos:

Teorema 4.3.1 *Sea U el flujo de un espacio-tiempo de Robertson-Walker $M(\kappa, a)$. Entonces (U, ρ, \mathcal{P}) es un fluido ideal con densidad de energía y presión dadas por:*

$$8\pi\rho/3 = (a'/a)^2 + \kappa/a^2,$$

$$8\pi\mathcal{P} = -[2a''/a + (a'/a)^2 + \kappa/a^2].$$

Demostración: Para demostrar que la terna es fluido ideal, nos apoyamos en la definición 4.3.2. En primer lugar, es claro por construcción que U es un vector unitario temporal que apunta al futuro. Además, de las expresiones para ρ y \mathcal{P} podemos deducir que son funciones diferenciables puesto que son combinaciones de una función diferenciable, sus derivadas y escalares. Finalmente, solo tenemos que deducir que se genera un tensor energía-momento como el de la definición. Para ello, usaremos la caracterización de T que vimos justo después de la definición.

Como venimos diciendo, a través de la ecuación de campo y proponiendo la métrica, estamos caracterizando el tensor energía-momento que vendrá dado (despejando de las ecuaciones de campo) por:

$$T = \frac{1}{8\pi}(Ric - \frac{1}{2}Rg).$$

Consideremos una vez más campos vectoriales X e Y ortogonales a U y veamos qué nos devuelven $T(U, U)$, $T(X, Y)$ y $T(U, X)$, apoyándonos en el Lema 4.3.1:

$$\begin{aligned} T(U, U) &= \frac{1}{8\pi}[Ric(U, U) - \frac{1}{2}Rg(U, U)] = \frac{1}{8\pi}[-3a''/a + \frac{1}{2}[6((a'/a)^2 + \kappa/a^2 + a''/a)]] = \\ &= \frac{3}{8\pi}[(a'/a)^2 + a''/a] = \rho, \end{aligned}$$

$$T(U, X) = \frac{1}{8\pi}[Ric(U, X) - \frac{1}{2}Rg(U, X)] = 0,$$

$$\begin{aligned} T(X, Y) &= \frac{1}{8\pi}[Ric(X, Y) - \frac{1}{2}Rg(X, Y)] = \\ &= \frac{1}{8\pi}[[2(a'/a)^2 + 2\kappa/a^2 + a''/a] - \frac{1}{2}[6((a'/a)^2 + \kappa/a^2 + a''/a)]]g(X, Y) = \\ &= \frac{1}{8\pi}[-(a'/a)^2 - \kappa/a^2 - 2a''/a]g(X, Y) = \mathcal{P}g(X, Y). \end{aligned}$$

Se concluye que T es un tensor energía-momento de fluido ideal con la densidad de energía y la presión enunciadas en el teorema. ■

Hemos concluido por tanto la construcción del modelo FLRW puesto que hemos comprobado que la geometría del espacio-tiempo de Robertson-Walker junto con el tensor-energía de fluido ideal del Teorema 4.3.1 cumplen la ecuación de campo y que, en

consecuencia, son solución. Una vez hecho esto, aunque el modelo es rico en resultados, nuestro objetivo principal será estudiar cómo aparecen singularidades en este espacio-tiempo. Para ello, usaremos como herramienta el siguiente corolario y el siguiente lema:

Corolario 4.3.1 *Para un fluido ideal en un espacio-tiempo de Robertson-Walker se cumple que:*

$$3a''/a = 4\pi(\rho + 3\mathcal{P}).$$

Demostración: La demostración es directa sin más que considerar las expresiones del Teorema 4.3.1 y combinarlas para obtener las cantidades deseadas. ■

Lema 4.3.2 *Para un fluido ideal en un espacio-tiempo de Robertson-Walker se cumple que:*

$$\rho' = -3(\rho + \mathcal{P})a'/a.$$

Este lema también lo presentamos sin demostración pues, aunque es consecuencia directa de la condición $divT = 0$, para demostrarlo necesitaríamos diversos resultados y conceptos relacionados con la geometría extrínseca de la variedad, lo cual haría que el documento fuera difícilmente autocontenido. Una demostración directa se puede encontrar en [18] [2].

Para poder estudiar en detalle el modelo FLRW es claro que tenemos que analizar los parámetros ρ , \mathcal{P} , κ y a . Todas estas variables son medibles experimentalmente. De hecho, hoy en día sabemos que el universo está dominado por los efectos de la densidad de energía ρ que es mucho más grande que la contribución de la presión \mathcal{P} [6]. Además, diversos experimentos para la medición de κ muestran (siempre bajo cotas de error) que nuestro universo es prácticamente euclídeo o plano, con una κ prácticamente cero aunque pudiendo ser un poco positiva [6].

Sin embargo, quizás el parámetro más interesante es el factor de escala a , puesto que su valor, más bien, el de su derivada, nos indica si el universo se expande o se contrae. Como ya comentamos al inicio de la sección, la expansión del universo fue corroborada experimentalmente cuando Hubble observó que las galaxias se alejaban de nosotros. No solo observó esto, sino que vio cómo las galaxias se alejaban de nosotros más rápido cuanto más lejos se encontraban, esto es, existe una relación de proporcionalidad entre la velocidad v de las galaxias y su distancia d del tipo: $v = Hd$ donde H es una constante de proporcionalidad. Esta constante H que, realmente no es una constante, se denomina constante de Hubble y se define como $a'(t)/a(t)$, viéndose claramente cómo tiene dependencia con el tiempo.

El valor de la constante de Hubble en la actualidad — $H_0 = a'(t_0)/a(t_0)$ — nos dice si el universo se expande o se contrae; en nuestro caso particular, el universo se

expande, dado que H_0 es positiva, lo que nos indica que la derivada del factor de escala también es positiva.

Actualmente, existe gran controversia por la determinación de la constante de Hubble puesto que diferentes métodos experimentales proporcionan diferentes valores, cuyos errores, en un principio, intersecaban, pensándose que ambos métodos se pulirían acercándose ambos valores a una cantidad concreta bien definida. Sin embargo, con el paso de los años y el perfeccionamiento de estos métodos experimentales, las barras de error han ido disminuyendo, pero alejándose en cada caso del valor del método opuesto, generando dos valores completamente diferentes de la constante de Hubble. Esta discrepancia entre métodos — aunque no haya una diferencia abismal entre ambos valores — da indicios de un mal funcionamiento del modelo cosmológico actual, haciendo sonar la alarma de que podríamos estar en vistas de nueva física o cambios profundos en los modelos actuales [12].

Estos datos experimentales y los resultados anteriores nos permiten encontrar nuestra primera singularidad en el modelo FLRW:

Teorema 4.3.2 *Sea $M(\kappa, a)$ un espacio-tiempo de Robertson-Walker. Si $H_0 > 0$ en algún t_0 y $\rho > -3\mathcal{P}$, entonces el intervalo de tiempo I tendrá un extremo inicial t_* acotado del siguiente modo: $t_0 - H_0^{-1} < t_* < t_0$. Además, se tendrá que $a'(t) > 0$ o que $a(t)$ alcanzará un máximo tras pasar por t_0 y por tanto el intervalo temporal será finito, $I = (t_*, t^*)$.*

Demostración: Deducimos directamente que $a''(t) < 0$ a través del Corolario 4.3.1 puesto que $\rho > -3\mathcal{P}$. Teniendo esto en cuenta, si consideramos la recta tangente de $a(t)$ en t_0 $A(t) = a(t_0) + a'(t_0)(t - t_0) = a(t_0) + H_0 a(t_0)(t - t_0)$, podemos deducir que el grafo de $a(t)$ está por debajo de esta recta salvo en t_0 . Por tanto, como H_0 es positiva (y $a(t_0)$ también) si vamos reduciendo t por debajo de t_0 , a debe presentar una singularidad antes de que se alcance el cero de $A(t)$ puesto que $a(t)$ siempre es positiva. Deducimos por tanto que el intervalo temporal I debe tener un extremo inicial t_* y además, este debe estar entre el cero de $A(t)$, $t_0 - H_0^{-1}$ y el propio t_0 .

De nuestra deducción inicial sobre la segunda derivada de $a(t)$, también podemos inferir que, o $a'(t)$ es siempre positiva o que es positiva hasta que se hace cero, alcanzándose un máximo tras t_0 — nótese que en t_0 no se puede alcanzar puesto que ahí la derivada es positiva por serlo H_0 — y volviéndose luego negativa, decreciendo $a(t)$. En este segundo caso, haciendo un razonamiento análogo al que hicimos entre t_0 y el cero de la izquierda, pero considerando ahora un cero a la derecha ya que la función $a(t)$ se vuelve decreciente y no puede anularse, podemos deducir que I tiene que tener un extremo final t^* y que por tanto $I = (t_*, t^*)$. ■

Hemos probado por tanto que, inevitablemente, ligada a las características de nuestro modelo y a los valores experimentales de los parámetros que lo caracterizan, aparece

una singularidad. Esto es, el extremo inicial (y también final en el segundo caso) del intervalo de tiempo, es un punto que no pertenece al modelo y donde no tiene sentido hablar del espacio-tiempo puesto que la métrica no está definida al no estarlo $a(t)$. Podemos decir por tanto que la RG se rompe en t_* (también en t^*) puesto que es un punto donde no se puede hablar de ecuación de campo y, como venimos diciendo, tampoco de espacio-tiempo.

Aunque este resultado nos pueda hacer pensar directamente en conceptos populares como el Big Bang, un punto infinitamente pequeño y con una densidad de energía infinitamente grande donde comienza el universo, tenemos que notar que en nuestro resultado solo hemos probado que existe singularidad; no hemos probado que el factor de escala tienda a cero a medida que nos acercamos al extremo de I o que la densidad de energía se haga infinita. Sin embargo, podemos perfilar un poco más nuestro resultado y hablar de singularidades físicas — recalamos aquí que este concepto está dentro de la idea de singularidad que expondremos en el Capítulo 5 pero es mucho menos general.

Definición 4.3.3 (Singularidades Físicas) *Diremos que un espacio-tiempo de Robertson-Walker presenta una singularidad física en un extremo de su intervalo temporal si la densidad de energía se hace infinita en el límite del tiempo hacia dicho extremo. En concreto, diremos que el extremo inferior (superior) o singularidad inicial (final) es un Big Bang (Big Crunch) si el factor de escala tiende a cero y su derivada a más (menos) infinito en el límite temporal hacia el extremo inferior (superior).*

En vistas de esta definición podemos caracterizar las singularidades físicas de un espacio-tiempo de Robertson-Walker tal y como sigue.

Teorema 4.3.3 *Consideremos un espacio-tiempo de Robertson-Walker que solo tenga singularidades físicas y con un intervalo de tiempo maximal. Si $H_0 > 0$ para algún $t_0 \in I$, $\rho > 0$ y se dan las cotas $-1/3 < \alpha \leq \mathcal{P}/\rho \leq \beta$ para α y β constantes, entonces:*

1. *La singularidad inicial es un Big Bang.*
2. *Si $\kappa \in \{0, -1\}$ entonces $I = (t_*, \infty)$ y el factor de escala tenderá a infinito y la densidad energética a cero cuando el tiempo tienda a infinito. Diremos que se produce una muerte térmica.*
3. *Si $\kappa = 1$ entonces $I = (t_*, t^*)$ puesto que $a(t)$ alcanzará un máximo y posteriormente se producirá un Big Crunch.*

Demostración: Al igual que antes tendremos que $a''(t) < 0$ pues estamos en las mismas condiciones que en el Teorema 4.3.2. Basta considerar que dado $\epsilon > 0$ de las desigualdades del enunciado tenemos que $\rho + 3\mathcal{P} > \epsilon\rho > 0$. Para probar 1, tendremos que demostrar que el factor de escala tiende a cero y su derivada a más infinito en el

límite temporal al extremo inferior. Recuperemos también que $a(t)$ tiene derivada positiva entre t_* y t_0 . Además, de las desigualdades, dado que $\mathcal{P} \leq \beta\rho$, también podemos deducir a través del Lema 4.3.2 que:

$$\rho' = -3(\rho + \mathcal{P})a'/a \geq -3(\beta + 1)\rho a'/a.$$

Denotando $\delta = 3(\beta + 1) > 2$ y escribiendo $a = a^{\delta-1}/a^\delta$ tendremos que:

$$(\rho a^\delta)' \geq 0.$$

Deducimos por tanto que $\rho a^\delta \leq \rho(t_0)a(t_0)^\delta$ en (t_*, t_0) y, además, como por hipótesis las singularidades son físicas, esto es, ρ tiende a infinito cuando t tiende al extremo inferior, deducimos que el factor de escala tiende a cero.

Trabajando exactamente igual con la desigualdad $\rho - \epsilon\rho \geq -3\mathcal{P}$, deducimos que $\rho' \leq -(2 + \epsilon)\rho a'/a$ y consecuentemente que $(\rho a^{2+\epsilon})' \leq 0$. De este resultado, al igual que antes, tendremos que $\rho a^{2+\epsilon} \leq \rho(t_0)a(t_0)^{2+\epsilon}$ en (t_*, t_0) y, en consecuencia, cuando el tiempo tiende al extremo inferior ρa^2 tiende a más infinito. Apoyándonos ahora en la expresión para la densidad de energía del Teorema 4.3.1, tenemos que $a'^2 + \kappa$ tiende a más infinito cuando t tiende a t_* y, por tanto, tenemos que a' tiende a más infinito como queríamos. Hemos probado pues que la singularidad inicial es un Big Bang.

Para demostrar 2 y 3 diferenciamos dos casos, puesto que, como $a'' < 0$, $a(t)$ puede tener un máximo o no tenerlo.

Supongamos primero que $a(t)$ tiene un máximo. De este modo, existirá un $\tilde{t} \in I$ con $a'(\tilde{t}) = 0$. En consecuencia, apoyándonos de nuevo en la expresión de ρ del Teorema 4.3.1, deducimos que $3\kappa/(8\pi a^2(\tilde{t})) = \rho > 0$ infiriéndose que $\kappa = 1$. Además, tal y como ocurría en el Teorema 4.3.2, debido a la condición sobre la segunda derivada, a' será negativa después del máximo; podemos por tanto realizar argumentos equivalentes a los que hemos hecho en 1 para t_* , pero con t^* deduciéndose que esta singularidad final es un Big Crunch.

Supongamos ahora que $a(t)$ no tiene máximo. En este caso a' será positivo en todo el intervalo y tendremos todos los resultados previos pero extendidos al intervalo completo. De las cotas $\rho > 0$ y $\rho + 3\mathcal{P} > 0$ tenemos que $\rho + \mathcal{P} > 0$ y por el Lema 4.3.2 la derivada de ρ será negativa. En consecuencia, ρ no podrá diverger a infinito, evitando la existencia de singularidades físicas y pudiéndose extender el intervalo temporal hasta el infinito por la derecha.

En estas condiciones, se nos presenta otra disyuntiva en función de si $a(t)$ se va a infinito al tender el tiempo a infinito o está acotada. En el primer caso, como ya demostramos antes, se cumple la cota $(\rho a^{2+\epsilon})' \leq 0$ por lo que para tiempos suficientemente grandes tendremos que $\rho a^{2+\epsilon}$ está acotado. Deducimos entonces que, al avanzar en el tiempo, $\rho a^{2+\epsilon}$ tiende a cero y por tanto $a'^2 + \kappa$ tiende a cero concluyéndose que $\kappa \in \{0, -1\}$.

En el caso de que $a(t)$ esté acotada, demostremos que es imposible que esto pueda darse. Si hacemos tender el tiempo a infinito, como la segunda derivada de $a(t)$ es negativa y la primera tiende a cero, concluimos que $a(t)$ debe tender a cierta constante Q . Pero entonces, estudiando ρa^2 , deducimos que tiene que tender a $3\kappa/8\pi$ y por tanto $\kappa \in \{0, 1\}$.

Veamos si κ puede ser 0. Esto es imposible, puesto que, de deducciones previas, sabemos que ρa^δ es creciente con el tiempo y por tanto no puede tender a cero.

Finalmente, estudiemos si κ puede ser 1. En este caso ρa^2 tenderá a $3/8\pi$ y podemos tomar entonces un $\tilde{\epsilon} > 0$ tal que $\rho \geq \tilde{\epsilon}$. Pero, además, como la derivada del factor de escala tiende a cero, existirá una secuencia de tiempos $\{t_i\}$ tendiendo a infinito con $\{a''(t_i)\}$ tendiendo a cero. En estas circunstancias, es claro que a través del Corolario 4.3.1, $\{(\rho + 3\mathcal{P})(t_i)\}$ tiende a cero, llevándonos a contradicción con la primera cota que dedujimos, $\rho + \mathcal{P} \geq \epsilon\rho \geq \tilde{\epsilon}\rho > 0$, y concluyendo la demostración. ■

El Teorema anterior nos muestra cómo un universo modelado al estilo FLRW, con las condiciones de energía, presión y constante de Hubble presentadas, comienza en una singularidad de tipo Big Bang y en función de su curvatura puede acabar en muerte térmica o con un Big Crunch. En este sentido, en vistas de las propiedades de nuestro universo medidas experimentalmente, podemos decir que nuestro universo — al menos modelado con la solución FLRW — comenzó en un Big Bang, un punto singular externo al modelo, y que posteriormente ha ido expandiéndose estando destinado, probablemente, a una muerte térmica.

A pesar de la belleza del resultado previo, nuestro interés principal es estudiar estas singularidades, sean del tipo que sean y ver por qué aparecen. Es claro que, por ahora, estos puntos singulares están puramente ligadas al modelo que hemos elegido y a sus características. Sin embargo, nos podríamos preguntar si constituyen algo intrínseco e inevitable a la formulación de la RG bajo ciertas condiciones más generales.

Esta última pregunta cobra mucho sentido en vistas al otro modelo que mencionamos al principio de la sección: el modelo de Schwarzschild. Aunque no desarrollamos este modelo por extensión, al igual que el modelo FLRW, presenta singularidades. La solución de Schwarzschild, como ya adelantamos, modela el campo gravitatorio producido por un cuerpo masivo esférico, esto es, modela el campo gravitatorio de una estrella de masa M . Aunque no entremos en muchísimo detalle (se recomienda el desarrollo de [18]), el modelo se basa en una solución de vacío de las ecuaciones de campo (Definición 4.2.3), es decir, modela el campo gravitatorio en un espacio-tiempo donde “no está la estrella” o mejor dicho fuera del espacio que ocupa esta. Bajo varias suposiciones — que la métrica sea independiente del tiempo, simetría esférica y planitud asintótica, esto es, que en el infinito el espacio-tiempo considerado sea \mathbb{R}_1^4 — la solución de Schwarzschild genera la siguiente métrica o elemento de línea

$$ds^2 = -\left(1 - \frac{r_S}{r}\right)dt^2 + \left(1 - \frac{r_S}{r}\right)^{-1}dr^2 + r^2d\Omega^2,$$

donde r es la coordenada radial esférica, r_S es el llamado radio de Schwarzschild y

$d\Omega^2 = d\theta^2 + \sin^2(\theta)d\varphi^2$ el elemento de línea angular.

En esta expresión podemos observar dos singularidades, una en el cero y otra en el radio de Schwarzschild. La singularidad del radio de Schwarzschild, como se demuestra en [9], no es más que una singularidad coordinada, no es real, y está directamente relacionada con el hecho de que la solución de Schwarzschild divide el espacio-tiempo en dos regiones: la región exterior a la estrella y una zona interior que suele interpretarse como un primer modelo de agujero negro⁴. Siendo concretos, la solución de Schwarzschild proporciona un espacio-tiempo con una zona exterior $P_1 \times_r S^2$ llamada espacio-tiempo de Schwarzschild y, una zona interior, $P_2 \times S^2$ llamada agujero negro de Schwarzschild donde P_1 y P_2 denotan los semiplanos de $\mathbb{R} \times \mathbb{R}^+$ de $r > 2M$ y $0 < r < 2M$ respectivamente.

Así pues, la singularidad real es la que se encuentra en el cero, esto es, en el centro del agujero negro. De hecho, el significado físico del radio de Schwarzschild es la máxima distancia a la que un fotón — partícula de luz con masa cero y en caída libre — puede acercarse al agujero negro sin que este lo atrape; más allá de esta frontera, que también suele conocerse como horizonte de sucesos del agujero negro, los fotones, y en consecuencia cualquier partícula — recuérdese que la velocidad de la luz es un límite superior de velocidad — caerán dentro del agujero negro sin poder escapar debido al poder gravitatorio de este, hasta caer inevitablemente a la singularidad (de aquí el nombre de agujero negro). De hecho, esto no es más que una consecuencia de la perturbación de la causalidad en el entorno del agujero negro; si recordamos los conos de luz del espacio-tiempo de Minkowski, el poder gravitatorio del agujero negro es tan grande que los curva en dirección a la singularidad.

Al igual que ocurría en el modelo FLRW, la singularidad de la solución de Schwarzschild es, a priori, propia del modelo. Sin embargo, el hecho de que aparezcan singularidades en dos modelos tan diferentes y simples, nos puede dar un indicio de que las singularidades son algo mucho más general y común en la RG, y que no son una casualidad ligada a las múltiples suposiciones que se realizan en estos modelos. Además, el hecho de que las singularidades estén tan relacionadas con puntos donde convergen las trayectorias de las partículas — en el modelo FLRW a través del límite a cero del factor de escala, y en el modelo de Schwarzschild, a través del destino inevitable de todas las partículas de caer a la singularidad — nos indica que la solución a nuestra cuestión debe estar relacionada con un estudio de la forma en que se unen entre sí diferentes puntos del espacio-tiempo a través de estas trayectorias o curvas con carácter causal, esto es, con la causalidad. Como veremos en el siguiente capítulo, la Teoría de Causalidad resolverá este enigma a través de los Teoremas de Singularidad: el Teorema de Hawking, asociado a soluciones del tipo FLRW, y el Teorema de Penrose, asociado a soluciones del tipo Schwarzschild.

⁴En la actualidad el modelo de Schwarzschild está obsoleto y prácticamente no se usa puesto que la suposición de métrica estática es muy poco realista. Nótese que todos los cuerpos rotan. Por ello, hoy en día hay mejores soluciones relacionadas con este fenómeno como aquellas asociadas a los agujeros negros de Kerr o agujeros negros en rotación [11].

Capítulo 5

Causalidad en Variedades de Lorentz

En las condiciones del final del capítulo previo, tal y como comentamos al inicio del documento, se generó gran controversia en el mundo de la física ligada a la verdadera posibilidad física de las singularidades. En la época en la que aparecieron las soluciones FLRW y de Schwarzschild, parecía que nadie quería creer en la realidad física de estos objetos, incluso el padre de la teoría, Albert Einstein, se mostraba reticente a aceptar que en el mundo real se produjeran dichos fenómenos, y defendía, al igual que la mayor parte de la comunidad, que dichas singularidades eran aberraciones matemáticas debidas a las condiciones ideales impuestas en los modelos. En los años posteriores, gran parte de la comunidad abordó el problema de las singularidades con el fin de demostrar precisamente este hecho, que las singularidades solo se produjeran en sistemas con alta simetría o condiciones especiales.

En este contexto histórico, los físicos norteamericanos Oppenheimer y Snyder publicaron su famoso artículo *On Continued Gravitational Contraction* [19], en el que probaban que el colapso de determinadas estrellas podría provocar situaciones parecidas a las singularidades en el sentido de la solución de Schwarzschild. Este hecho, que parecía decantar la balanza hacia el lado contrario de lo que se creía hasta entonces, no hizo que los físicos dejaran de buscar soluciones de las ecuaciones de campo con el fin de probar lo opuesto. De hecho, como comentamos en la introducción, se dieron casos, como el de los físicos rusos Lifshitz y Khalatnikov, que incluso afirmaron haber probado que las singularidades solo se producían bajo distribuciones de energía y velocidades muy concretas.

La situación de crisis parecía haber concluido con este último resultado. Sin embargo, la desmotración dada por estos científicos era, cuanto menos, dudosa. A priori, ya era difícil de esperar que se obtuviesen resultados generales en la forma en que se estaba abordando el problema desde la comunidad física: buscando más y más soluciones concretas de las ecuaciones de campo. En este tira y afloja conceptual, aparecieron las

figuras de Roger Penrose y Stephen Hawking, quienes, haciendo una abstracción mucho mayor del problema y atacándolo con técnicas más generales, consiguieron enunciar una serie de teoremas puramente matemáticos que aseguraban la existencia de las singularidades bajo determinadas condiciones: los Teoremas de Singularidad.

La línea que siguieron Hawking y Penrose, como esbozábamos en la sección previa, está basada en la causalidad. Nótese, como dijimos, la intuición detrás de esta idea, pues es lógico abordar los problemas expuestos previamente estudiando cómo se afectan causalmente, esto es, a través de relaciones causa-efecto, diferentes eventos. Matemáticamente, la manera de modelar cómo se influyen entre sí diferentes puntos del espacio-tiempo es estudiar de forma general cómo se unen diferentes puntos de la variedad espacio-temporal a través de curvas con determinado carácter causal. En este sentido, podremos entender una singularidad como un punto al que no “llegan” este tipo de curvas, es decir, un espacio-tiempo con singularidades — como veremos — será una variedad geodésicamente incompleta, ya sea en el sentido luminoso o en el sentido temporal — observadores o rayos de luz que no acaban debido a estos “fallos” en la estructura del espacio-tiempo.

En este capítulo abordaremos la Teoría de la Causalidad de una forma puramente matemática — directamente relacionada con la geometría — pero destinada a resolver el problema físico de las singularidades llegando a los Teoremas de Singularidad. Así, aunque esta teoría, de un modo exclusivamente matemático, ya sea rica y profunda, nuestro desarrollo estará enfocado en esta línea uniendo todas las ideas físicas expuestas con estos resultados matemáticos. En este sentido, denotaremos de aquí en adelante por M a un espacio-tiempo m dimensional — dejamos de restringirnos a 4 dimensiones por hacer un desarrollo lo más general posible — que será la variedad a la que nos restringiremos en nuestro estudio de la causalidad. El capítulo se basará en los textos [5], [17], [18], [24].

5.1. Relaciones y condiciones de causalidad. Separación temporal

En esta primera sección desarrollaremos los elementos básicos de la Teoría de Causalidad, realizando una introducción lo más exhaustiva posible de todos los conceptos que necesitaremos posteriormente. Primero, estudiaremos las estructuras causales básicas que surgen de tomar las relaciones simples de conexión por curvas de diferentes carácter causal comentadas en la introducción. Posteriormente, introduciremos la separación temporal que hará las veces de distancia, permitiéndonos ver cómo están separados los puntos del espacio-tiempo. Finalmente, caracterizaremos diferentes tipos de espacio-tiempo en función de las condiciones causales que cumplan, permitiéndonos establecer diferentes “grados” de causalidad en función de lo fuertes o débiles que sean estas condiciones.

Comencemos introduciendo las relaciones de causalidad en función de cómo se conectan entre sí los diferentes puntos del espacio-tiempo. Recuérdese, tal y como vimos en el Capítulo 4, que los observadores en un espacio-tiempo se mueven a través de curvas temporales dirigidas hacia el futuro y los rayos de luz no son más que geodésicas luminosas. Por tanto, viendo cómo se unen diferentes puntos a través de curvas con diferentes carácter causal — curvas temporales y causales — estamos viendo exactamente cómo dichos puntos se pueden influir entre sí de un modo físico; esto es, estamos determinando su pasado y su futuro a través de qué eventos le han sido o le serán accesibles a través de relaciones causa-efecto. Tenemos por tanto la siguiente definición:

Definición 5.1.1 (Relaciones de causalidad) *Dados dos puntos p y q de un espacio-tiempo M , diremos que:*

- *p está en el pasado cronológico de q — de forma equivalente, q está en el futuro cronológico de p — si existe una curva temporal dirigida al futuro empezando en p y acabando en q ; denotaremos esta relación como $p \ll q$ y el pasado cronológico (resp. el futuro cronológico) de un punto q como $I^-(q)$ (resp. $I^+(q)$).*
- *p está en el pasado causal estricto de q — q está en el futuro causal estricto de p — si existe una curva causal dirigida al futuro empezando en p y acabando en q ; denotaremos esta relación como $p < q$.*
- *p está en el pasado causal de q — q está en el futuro causal de p — si $p \leq q$, es decir, o $p < q$, o bien $p = q$. Denotaremos al pasado causal (resp. futuro causal) de un punto q como $J^-(q)$ (resp. $J^+(q)$).*

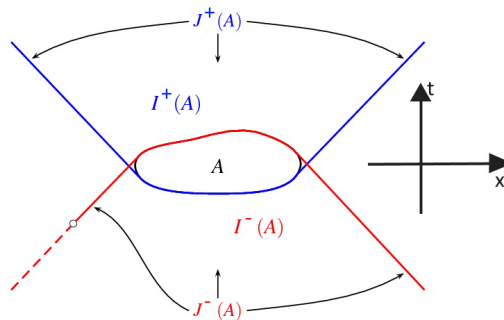


Figura 5.1: Ejemplo: Futuros y pasados causales y cronológicos de un conjunto A del espacio de Minkowski bidimensional.

Nótese que la definición anterior puede extenderse de forma trivial a un conjunto A sin más que considerar la unión de todos los pasados (futuros) cronológicos o causales de cada punto de A . En este sentido, el conjunto $I^+(A)$ ($I^-(A)$) será el subconjunto de M de los puntos que se relacionan temporalmente en el sentido de la definición anterior

con algún punto de A — con los conjuntos causales es exactamente igual. Recalcamos además que es directo de la definición que $I^+(A) \cup A \subset J^+(A)$.

Téngase en cuenta de aquí en adelante que, debido a la simetría entre pasado y futuro, las demostraciones se harán indistintamente para uno de los dos, teniéndose la demostración equivalente sin más que revertir el tiempo.

Recordando la Proposición 3.3.1 podemos establecer además una estrecha relación entre los conjuntos previos ya que:

Corolario 5.1.1 *Dados $p, q, r \in M$ si $p \ll q$, $q \leq p$ (o $p \leq q$) y $q \ll r$ entonces $p \ll r$. Equivalentemente, dado un subconjunto A de M entonces $I^+(A) = I^+(I^+(A)) = I^+(J^+(A)) = J^+(I^+(A)) \subset J^+(J^+(A)) = J^+(A)$.*

Haciendo referencia una vez más al Capítulo 3, si recordamos la estructura que se generaba en el espacio de Minkowski \mathbb{R}_1^4 debido al carácter causal de los vectores, podemos ver ahora que los conjuntos que allí interpretábamos como futuro y pasado son exactamente los futuros y pasados temporales y causales en el sentido de la definición anterior. La causalidad en el espacio de Minkowski, como vimos, es bastante sencilla y, además, sabemos como hemos recalcado varias veces en el documento que todo espacio-tiempo puede aproximarse localmente por el espacio de Minkowski (en la dimensión adecuada).

El siguiente resultado nos va a permitir sacar provecho de estas circunstancias pues, si consideramos un subconjunto abierto de M con sus relaciones de causalidad intrínsecas, estas relaciones determinarán la causalidad en M . En particular, si el subconjunto tomado es convexo (en el sentido usual), tendremos unas relaciones de causalidad tan simples como en el espacio de Minkowski.

De aquí en adelante llamaremos \vec{pq} a la única geodésica que une los puntos $p, q \in C \subset M$ con C abierto y convexo, y tal que dicha geodésica esté contenida en C . Además, cuando tratemos la causalidad de forma local será usual la notación $I^+(A, U) = I^+(A) \cap U$ donde $U \subset A \subset M$.

Lema 5.1.1 *Dado un subconjunto convexo abierto $C \subset M$, entonces:*

1. *Para $p \neq q \in C$, $q \in J^+(p, C)$ (también se da para el futuro temporal) si y solo si \vec{pq} es causal y apunta al futuro.*
2. *$I^+(p, C)$ es abierto en C y, por tanto, en M .*
3. *$J^+(p, C)$ es la clausura de $I^+(p, C)$ en C .*
4. *La relación \leq es cerrada en C si dadas $\{p_n\} \rightarrow p$ y $\{q_n\} \rightarrow q$ contenidas en C con $q_n \in J^+(p_n, C)$ para todo n , entonces $q \in J^+(p, C)$.*
5. *Una curva causal α en un compacto $K \subset C$ es extendible (de forma continua).*

Demostración: Las tres primeras propiedades no las probamos pues dependen de resultados acerca de la aplicación exponencial que se alejan bastante del contenido del documento y alargarían mucho su extensión. Estas propiedades y una demostración de estos puntos se pueden encontrar en [24].

Para la cuarta propiedad solo tenemos que considerar la aplicación que va de $C \times C \rightarrow TM$ que lleva cada par (p, q) al vector $\vec{p}\vec{q}'(0)$ que claramente (es una comprobación directa) es continua. Aplicando 1 tenemos el resultado.

Probemos la quinta propiedad. Para ello, supongamos que el dominio de α es de la forma $[0, t)$ con t pudiendo ser infinito. Para probar que es extendible, dada cualquier sucesión $\{s_n\}$ que tienda a t , toda sucesión $\{\alpha(s_n)\}$ tiene que converger al mismo punto $p \in K$. La existencia de estas sucesiones es clara, pues K es compacto. Veamos que todas tienden al mismo p . Supongamos que una de ellas tiende a un $q \neq p$. Esto significaría que, en cierto modo, α se movería de p a q yendo y volviendo al ser ambos límites. Ahora bien, si aplicamos el punto 4 tendría que cumplirse $q \in J^+(p, C)$ y que $p \in J^+(q, C)$, pero es claro, aplicando 1 que entonces $\vec{p}\vec{q}$ tendría que apuntar al futuro y al pasado lo cual es imposible. ■

En general, a diferencia de como ocurría en el espacio de Minkowski, $J^+(A)$ no es cerrado. Sin embargo, sí podemos generalizar el punto 2 del lema anterior viendo que $I^+(A)$ es siempre abierto. Además podemos asegurar que $J^+(A)$ es cerrado si A es compacto.

De hecho tenemos los siguientes resultados generales que nos permiten comparar la causalidad de Minkowski con una general:

Lema 5.1.2 *Dado un conjunto $A \subset M$ compacto en el que \leq es cerrada, entonces $J^+(A)$ es cerrado.*

Demostración: Sea $\{q_n\}_n$ una sucesión en $J^+(A)$ que converge a $q \in M$. Por definición, existirá una única sucesión $\{p_n\}_n$ en A tal que $p_n \leq q_n$ para todo n . Como A es compacto, cogiendo una subsucesión convergente es directo que la sucesión de p_n converge a un $p \in A$. Ahora bien, como ya hemos probado, la relación \leq es cerrada y, consecuentemente, $p \leq q$. Así, hemos probado que $J^+(A)$ es cerrado pues q está en él con lo que concluye la demostración. ■

Lema 5.1.3 *Dado un espacio-tiempo M :*

1. *La relación \ll es abierta. Dados dos puntos $p, q \in M$ con $p \ll q$ entonces para puntos p' y q' en entornos de p y q respectivamente se tiene que $p' \ll q'$.*
2. *$\text{int}(J^+(A)) = I^+(A)$ para cualquier $A \subset M$.*

3. $J^+(A) \subset \overline{I^+(A)}$ dándose la igualdad si y solo si $J^+(A)$ es cerrado para cualquier $A \subset M$.

Demostración: Comencemos probando el primer punto. Sea α una curva temporal de p a q . Si C es un entorno convexo de q , tomemos un punto q^- de C sobre α previo a q . Del mismo modo, dado un entorno convexo C' de p , tomemos un punto p^+ de C' sobre α entre p y q^- . Aplicando el Lema 5.1.1 tenemos que $I^+(q^-, C)$ y $I^-(p^+, C')$ son abiertos en M y por tanto son los entornos de p y q con las características que queremos.

Para la demostración de 2 usemos doble inclusión. Ya sabemos que $I^+(A)$ es abierto por 1 y ya sabíamos que está contenido en $J^+(A)$ y consecuentemente en $\text{int}(J^+(A))$. Tomemos ahora un $q \in \text{int}(J^+(A))$. Si consideramos un entorno convexo C de q entonces es claro que $I^-(q, C)$ contiene un punto de $J^+(A)$ y consecuentemente $q \in I^+(J^+(A)) \subset I^+(A)$.

Finalmente, probemos 3. Veamos la contención (nos basta considerar un único punto). Es claro que $p \in \overline{I^+(p)}$; por tanto tomemos $p < q$. Si consideramos ahora una curva α causal y apuntando al futuro de p a q y tenemos un entorno C de q convexo, tomemos un punto q^- de α en $J^-(q, C)$. Aplicando ahora de nuevo el Lema 5.1.1 la cadena de inclusiones es clara pues $q \in J^+(q^-, C) \subset \overline{I^+(q^-, C)}$ pero $I^+(q^-, C) \subset I^+(J^+(p)) \subset I^+(p)$ y por tanto $q \in \overline{I^+(p)}$. Tenemos así el resultado pues la condición de igualdad es directa ya que teníamos que $I^+(A) \subset J^+(A)$. ■

Podemos concluir junto con el siguiente resultado — consecuencia directa del Teorema 3.3.5 — que $J^+(A)$ es una unión de A con su futuro temporal y, en ocasiones, de curvas luminosas de A .

Corolario 5.1.2 *Si α es una curva causal que apunta al futuro de $A \subset M$ a un punto $q \in J^+(A) - I^+(A)$, entonces es una geodésica luminosa que no interseca a $I^+(A)$ y que no tiene puntos focales antes que q .*

Una vez vista la estructura causal general de un espacio-tiempo necesitamos un objeto que nos permita caracterizar de algún modo la separación entre puntos dentro de este, es decir, una especie de generalización de distancia.

Definición 5.1.2 (Separación temporal) *Dado un espacio-tiempo M llamaremos separación temporal o distancia Lorentziana entre dos puntos p y q a la aplicación $\tau : M \times M \rightarrow [0, +\infty]$ definida como:*

$$\tau(p, q) = \begin{cases} 0 & \text{si } C_{pq}^c = \emptyset, \\ \sup \{L(\alpha), \alpha \in C_{pq}^c\} & \text{si } C_{pq}^c \neq \emptyset, \end{cases}$$

donde C_{pq}^c denota el conjunto de todas las curvas causales diferenciables a trozos y dirigidas al futuro que unen p y q .

Si recordamos el tiempo propio de un observador que introdujimos en el Capítulo 4, podemos interpretar la separación temporal — cuando se toma el supremo — como el tiempo máximo que tarda un observador en un viaje de p a q . En este sentido, haciendo una comparación entre este objeto métrico o pseudodistancia con la distancia que solemos considerar de forma clásica (no hemos introducido una distancia Riemanniana pero la generalización es directa) podemos ver que, mientras la distancia usual se minimiza, la separación temporal se maximiza.

Nótese que la separación temporal no es simétrica debido al hecho de estar relacionada con una orientación temporal.

Proposición 5.1.1 *Dados tres puntos p , q y r de un espacio-tiempo M . Entonces:*

1. $\tau(p, q) > 0 \Leftrightarrow p \ll q$.
2. $\tau(p, p) = \infty$ si existe una curva temporal diferenciable a trozos que una a p consigo mismo y, en caso contrario, valdrá cero.
3. τ no es simétrica, es decir, si $0 < \tau(p, q) < \infty$ entonces $\tau(q, p) = 0$.
4. τ cumple la desigualdad triangular inversa: si $p \leq q \leq r$, entonces $\tau(p, q) + \tau(q, r) \leq \tau(p, r)$.

Demostración: El primer punto es directo de la propia definición de $I^+(p)$ (resp. $I^-(p)$).

Para el segundo, supongamos que existe una curva temporal diferenciable a trozos dirigida hacia el futuro α que una p con p . A partir de ella, dado que por ser temporal su longitud es siempre estrictamente positiva, podemos crear otras $k \in \mathbb{N}$ curvas α_k que unan p con p cumpliéndose que $\tau(p, p) \geq L(\alpha_k) = kL(\alpha) \rightarrow +\infty$ (ya que las estamos concatenando entre sí). En el caso de que no existiera dicha curva temporal pueden darse dos casos: o no existe ninguna curva causal dirigida al futuro que una p con p y por tanto la separación temporal es cero o, si existe dicha curva causal debido a la Proposición 3.3.1 no puede ser temporal en algún punto y por tanto la separación vuelve a ser cero.

Para ver que τ no es simétrica, basta usar el resultado previo ya que si se da la condición del punto 3 no puede existir ninguna curva causal dirigida al futuro que una q con p porque, de ser así, podríamos concatenar dicha curva con otra temporal que fuera de p a q y la separación temporal entre estos sería infinita.

Probemos finalmente la desigualdad triangular. Si existen curvas causales dirigidas al futuro α que una p con q y β que una q con r , podemos tomar un $\epsilon > 0$ tal que las longitudes de α y β estén entre $\epsilon/2\tau(p, q)$ y $\epsilon/2\tau(q, r)$. Tenemos entonces que

$$\tau(p, r) \geq L(\alpha + \beta) \geq \tau(p, q) + \tau(q, r) - \epsilon,$$

teniéndose el resultado. En el caso de que no existan dichas curvas la separación temporal es claramente cero y el resultado se da trivialmente. ■

Nótese cómo, aunque no hayamos introducido ya condiciones de causalidad sobre el espacio-tiempo, las propiedades de τ prácticamente nos lo exigen, pues van en contra de muchas intuiciones físicas. Por ejemplo, el hecho de que la separación temporal entre un punto y él mismo sea infinita es completamente ilógico desde un punto de vista físico, pero nótese que estamos diciendo que se puede unir consigo mismo a través de una curva temporal lo cual también es bastante problemático desde el punto de vista físico pues nos podría llevar a paradojas causales.

Otra propiedad importante de la separación temporal es que no es continua en general; solo podemos asegurar que es semi-continua por debajo (inferiormente).

Proposición 5.1.2 *La separación temporal τ es una función semi-continua inferiormente.*

Demostración: Nos centramos en el caso en que la separación temporal es positiva pues si es cero no hay nada que probar. Dada una curva causal dirigida al futuro α que una p con q tenemos que probar que $\liminf_{k \rightarrow \infty} \tau(p_k, q_k) \geq L(\alpha)$.

Sin pérdida de generalidad supongamos que $|\alpha'| < 1$, cojamos un $\epsilon \in (0, L/2)$ donde L denota el extremo superior del dominio de α y consideremos $p_\epsilon = \alpha(\epsilon/2)$ y $q_\epsilon = \alpha(L - \epsilon/2)$. Como $\{p_k\}$ tenderá a $p \in I^-(p_\epsilon)$ y $\{q_k\}$ a $q \in I^+(q_\epsilon)$, recordando que los futuros y pasados cronológicos son abiertos, podemos deducir que $p_k \in I^-(p_\epsilon)$ y $q_k \in I^+(q_\epsilon)$ para todo k mayor igual que cierto k_ϵ .

Por tanto, para $m \geq m_\epsilon$

$$\begin{aligned} L(\alpha) &\leq \tau(p_\epsilon, q_\epsilon) + \epsilon \leq \tau(p_k, p_\epsilon) + \tau(p_\epsilon, q_\epsilon) + \tau(q_k, q_\epsilon) + \epsilon \\ &\leq \epsilon + \tau(p_k, q_k), \end{aligned}$$

donde hemos usado la desigualdad triangular para deducir la última desigualdad. Tomando límite inferior tenemos el resultado. ■

Llegados a este punto, presentaremos los diferentes tipos de condiciones de causalidad que le podemos pedir al espacio-tiempo. La motivación de estas condiciones suele nacer de ideas puramente lógicas o intuiciones físicas como que no puedan existir curvas temporales o causales cerradas. Sin embargo, como veremos a continuación, estas condiciones pueden ser mucho más complicadas y antintuitivas.

Definición 5.1.3 (Condiciones de causalidad) *Dado un espacio-tiempo M diremos que es:*

- cronológico, si no existe ninguna curva temporal cerrada en M .

- causal, si no existe ninguna curva causal cerrada en M .
- distinguible, si cada punto puede ser caracterizado por su futuro y pasado cronológico, es decir, si $I^+(p) = I^+(q)$ y $I^-(p) = I^-(q)$ entonces $p = q$.
- fuertemente causal, si para cada punto p de M y dado cualquier entorno U de p existe otro entorno V de p contenido en U de modo que todo arco de curva causal con extremos en V esté completamente contenido en U .
- establemente causal, si existe una función temporal global, esto es, una función $t : M \rightarrow \mathbb{R}$ diferenciable tal que su gradiente sea temporal.
- causalmente continuo, si es distinguible y además reflexivo, esto es, si dados p y q de M entonces $p \in \overline{I^-(q)}$ si y solo si $q \in I^+(p)$.
- causalmente simple, si los conjuntos $J^+(p)$ y $J^-(p)$ son cerrados en M para todo punto $p \in M$.
- globalmente hiperbólico, si es causal y para todo par de puntos p y q con $p < q$ el conjunto $J(p) \cap J(q)$ es compacto.

La forma en que hemos presentado estos resultados no es arbitraria. De hecho, están ordenados de arriba a abajo de más débil a más fuerte, es decir, cada uno de estos tipos de causalidad implica el anterior en la lista previa. Nótese que además dichas definiciones pueden restringirse a subconjuntos de forma natural.

Cada uno de estos tipos de causalidad es importante por sí mismo, aunque en general, los espacio-tiempo más usados son los cronológicos, los fuertemente causales y los globalmente hiperbólicos. Las diferentes caracterizaciones e implicaciones de un tipo a otro son resultados excesivamente complicados y extensos para lo que se pretende abordar en este texto. Aunque estos no vayan a ser necesarios en lo que sigue — solo se usarán implicaciones de un tipo de causalidad más fuerte a otro más débil — se pueden consultar en textos como [7].

De aquí en adelante, nos centraremos en un espacio-tiempo que cumpla la condición de hiperbolicidad global pues, aparte de ser la condición más fuerte y con resultados matemáticos más interesantes, es también aquella que va a tener un mayor interés físico como iremos viendo a lo largo del capítulo — ya que es bastante lógico que nuestro universo sea globalmente hiperbólico por caracterizaciones que veremos más tarde.

A continuación, introducimos tres resultados que usaremos más adelante que combinan todos los elementos que hemos presentado: las relaciones y condiciones de causalidad y la separación temporal. Además, uno de ellos constituirá la primera justificación de por qué nuestro espacio-tiempo debe ser globalmente hiperbólico. Para poder llegar a estos resultados, necesitaremos una herramienta técnica previa que nos permita tener cierta noción de límite de una sucesión de curvas causales diferenciables a trozos. Introducimos así los cuasi-límites, geodésicas rotas que solo son límite de forma aproximada, midiéndose dicha aproximación a través de un recubrimiento convexo de M .

Definición 5.1.4 (Cuasi-límite) Sean $\{\alpha_n\}$ una sucesión infinita de curvas causales en M y \mathcal{R} un recubrimiento convexo de M . Una sucesión límite o cuasi-límite para $\{\alpha_n\}$ relativa a \mathcal{R} es una sucesión finita o infinita $p = p_0 < p_1, \dots$ que cumple que:

1. Para cada p_i existe una subsucesión $\{\alpha_k\}$ y para cada k existen números $s_{k0} < s_{k1} < \dots < s_{ki}$ tales que
 - $\lim_{k \rightarrow \infty} \alpha_k(s_{kj}) = p_j$ para todo $j \leq i$.
 - Para todo $j < i$, los puntos p_j, p_{j+1} y los segmentos $\alpha_k|_{[s_{kj}, s_{kj+1}]}$ están contenidos en un único conjunto de \mathcal{R} para todo k .
2. Si $\{p_j\}$ es infinito es no convergente mientras que, si es finito, debe tener más de un punto y no cumplirá la primera condición.

Los siguientes resultados sobre cuasi-límites los presentamos también sin demostración pues son meras tecnicidades que no aportan nada al contenido central del capítulo. Para un desarrollo muy detallado consúltese [18]

Proposición 5.1.3 Sea $\{\alpha_n\}$ una sucesión de curvas causales que apuntan al futuro tales que $\{\alpha_n(0)\}$ converge en el sentido previo a $p \in M$ pero $\{\alpha_n\}$ no. Entonces dicha sucesión tiene una sucesión límite empezando en p relativa a cualquier recubrimiento convexo de M .

Lema 5.1.4 Sea $\alpha_n : [0, 1] \rightarrow M$ con $n \in \mathbb{N}$ una sucesión de curvas causales dirigidas al futuro tal que $\{\alpha_n(0)\} \rightarrow p$ para algún $p \in M$ y con $\{\alpha_n(1)\}$ sin converger a p . Suponiendo que $\{\alpha_n\}$ tiene una sucesión límite finita $\{p_i\}_{i=1}^r$, entonces una subsucesión de $\{\alpha_n(1)\}$ converge al último punto de la sucesión límite $p_r (\neq p)$.

Lema 5.1.5 Sean M un espacio-tiempo, K un subconjunto compacto donde se cumple la condición de causalidad fuerte y, p y q dos puntos de K . Consideremos además una sucesión de curvas temporales dirigidas al futuro $\alpha_n : [0, 1] \rightarrow M$ con $n \in \mathbb{N}$ en K , tal que $\{\alpha_n(0)\} \rightarrow p$ y $\{\alpha_n(1)\} \rightarrow q$. Entonces existe un cuasilímite γ empezando en p y acabando en q y una subsucesión $\{\alpha_{n_k}\}$ de $\{\alpha_n\}$ cumpliendo que $\lim_{inf_k} L(\alpha_{n_k}) \leq L(\gamma)$.

En estas condiciones podemos presentar los dos resultados comentados previamente.

Proposición 5.1.4 Dados dos puntos p y q tales que $p < q$, si el conjunto $J(p, q) := J^+(p) \cap J^-(q)$ es compacto y cumple la condición de causalidad fuerte, entonces existe una geodésica causal dirigida al futuro de p a q de longitud máxima.

Demostración: Si $\tau(p, q) = 0$ entonces por la Proposición 3.3.1 toda curva causal que una p con q tiene que ser una pregeodésica luminosa. Si suponemos ahora que la separación temporal es positiva, podemos volver a usar la Proposición 3.3.1 y podemos suponer la existencia de una sucesión de curvas temporales dirigidas al futuro tales que su longitud converge a la separación temporal. En estas condiciones, basta usar el Lema 5.1.5 sobre el conjunto $J(p, q)$ para obtener un cuasi-límite γ de longitud $\tau(p, q)$. Finalmente, para determinar que γ no tiene saltos, basta usar una vez más la Proposición 3.3.1 ya que si los tuviera, podríamos construir una curva causal de mayor longitud. ■

Este primer resultado nos da una intuición de por qué nuestro universo debe ser globalmente hiperbólico puesto que, aparte de ser lógicamente causal, también parece bastante natural el hecho de que si podemos unir dos puntos mediante una curva causal también podamos unir esos dos puntos mediante una geodésica causal máxima. De hecho, esto es exactamente lo que ocurre en el espacio de Minkowski — que localmente aproxima bien cualquier espacio-tiempo — implicando también que lo sean los espacios de Robertson-Walker (siempre que S sea completo [18]), fortaleciendo la idea de que nuestro universo tenga que ser así.

Finalmente, presentamos dos resultados más sobre espacios globalmente hiperbólicos que serán de utilidad más tarde.

Proposición 5.1.5 *Si U es un conjunto abierto globalmente hiperbólico, entonces la separación temporal es continua en $U \times U$.*

Demostración: Tenemos que probar que la separación temporal es también semi-continua superiormente. Para ello, supongamos que no lo es en un par (p, q) de $U \times U$. En este caso, podemos tomar un número real $\epsilon > 0$ y sucesiones $\{p_n\}$ y $\{q_n\}$ convergiendo a p y q respectivamente con $\tau(p_n, q_n) \geq \tau(p, q) + \epsilon$. Además, como $\tau(p_n, q_n)$ es estrictamente positiva, existe una curva causal α_n tal que $L(\alpha_n) > \tau(p_n, q_n) - 1/n$. Como U es abierto, contendrá puntos $p^- \ll p$ y $q^+ \gg q$, pudiéndose suponer además que las sucesiones iniciales están contenidas en $I^+(p^-)$ y $I^-(q^+)$ respectivamente y por tanto están contenidas en $J^+(p^-) \cap J^-(q^+)$. Finalmente, basta usar el Lema 5.1.5 sobre el conjunto $J^+(p^-) \cap J^-(q^+)$ para deducir que existe una curva causal γ de p a q con $L(\gamma) \geq \tau(p, q) + \epsilon$ lo cual es imposible por la definición de τ . ■

Proposición 5.1.6 *Si un conjunto abierto $U \subset M$ es globalmente hiperbólico, entonces la relación \leq de M es cerrada en U .*

Demostración: Queremos probar exactamente lo mismo que en el tercer punto del Lema 5.1.1 solo que en el caso de este U concreto. Solo tenemos que fijarnos en el caso en que $p_n < q_n$ para todo n ya que el caso con $p_n = q_n$ para infinitos n es trivial. La demostración es igual que la de la proposición previa pues, si tomamos una curva

causal α_n que conecte p_n con q_n , esta, tal y como hemos visto, siempre estará en un $J^+(p^-) \cap J^-(q^+)$ adecuado y tendremos el resultado. Si estuvieramos en el caso en que $p \neq q$ la curva γ del Lema 5.1.5 nos da la relación $p < q$. ■

Así pues, en los espacios globalmente hiperbólicos los futuros y pasados causales de cada punto así como los conjuntos de la forma $J^+(p) \cap J^-(q)$ son cerrados. De aquí en adelante notaremos a estos últimos conjuntos como $J(p, q)$ para aliviar la notación.

5.2. Conjuntos acronales. Hipersuperficies de Cauchy

En esta segunda sección comenzaremos a introducir nuevas estructuras causales que serán de interés y que constituirán los cimientos de los objetos más complejos que introduciremos en la sección posterior. Para introducir estas nuevas estructuras, pondremos nuevas restricciones causales sobre las relaciones entre los puntos de los conjuntos. Esta será la línea principal tanto de esta sección como de la siguiente solo que con objetos diferentes; en esta primera sección nuestro foco central serán las hipersuperficies de Cauchy que nos permitirán caracterizar de un nuevo modo los espacios globalmente hiperbólicos además de ser esenciales en todo el desarrollo que sigue.

Las hipersuperficies de Cauchy son un caso concreto de conjuntos acronales. Introduzcamos estos primero.

Definición 5.2.1 (Conjunto acronal) *Un subconjunto A de un espacio-tiempo M es acronal si no se cumple que $p \ll q$ para ningún p y q de A .*

Nótese que un conjunto acronal es aquel que solo puede ser intersecado una vez (como mucho) por una curva temporal. La intuición más naif de conjunto acronal es considerar una rodaja de espacio a tiempo constante en el espacio de Minkowski que claramente podrá ser intersecada a lo más una vez por una curva temporal. Es claro además que cualquier subconjunto de un conjunto acronal es acronal y que su clausura también es acronal — recuérdese que \ll es abierta. Además, destacamos que paralelamente a los conjuntos acronales podemos definir los conjuntos acausales cambiando la relación cronológica por la causal haciendo la restricción más fuerte.

Dentro de los conjuntos acronales nos interesaremos especialmente en su borde. Este será el objeto con el que determinemos cuándo un conjunto acronal es hipersuperficie, permitiéndonos llegar a la idea de hipersuperficie de Cauchy como venimos diciendo.

Definición 5.2.2 (Borde de un conjunto acronal) *El borde de un conjunto acronal A es el conjunto formado por todos los puntos p de \bar{A} tales que todos sus entornos U contienen una curva temporal que va desde $I^-(p, U)$ a $I^+(p, U)$ que no interseca con A .*

En este punto ya podríamos intentar determinar a través del borde cuáles son las condiciones bajo las que un conjunto acronal es hipersuperficie. Sin embargo, antes de ello, vamos a debilitar nuestros requisitos sobre la hipersuperficie. Durante todo el documento hemos estado interesados en estructuras diferenciables pero, en este punto, vamos a relajar esta exigencia y vamos a considerar también hipersuperficies que no sean diferenciables. Así pues, diremos que un espacio topológico T es una *variedad topológica* si cada uno de sus puntos tiene un entorno homeomorfo a un abierto de \mathbb{R}^n .

La idea detrás de esta nueva condición viene motivada una vez más por el caso simple del espacio de Minkowski. En este, como hemos dicho antes, las rodajas espaciales a tiempo constante son conjuntos acronales; es más, son hipersuperficies y no tienen borde. En este sentido, parece intuitivo que vayamos a poder caracterizar las hipersuperficies como conjuntos acronales sin borde. Sin embargo, es claro que existen conjuntos acronales sin borde que no son diferenciables — por ejemplo los conos luminosos en el espacio de Minkowski — que, por consistencia, vamos a incluir también dentro de nuestra definición de hipersuperficie. Así pues, caracterizamos las hipersuperficies de forma topológica como sigue:

Definición 5.2.3 (Hipersuperficie topológica) *Un subespacio S de una variedad topológica T es una hipersuperficie topológica si para cada $p \in S$ existen un entorno U en T y un homeomorfismo sobreyectivo de U a un conjunto abierto de \mathbb{R}^n cumpliendo que $\phi(U \cap S) = \phi(U) \cap \Pi$ con Π un hiperplano de \mathbb{R}^n .*

Es directo, tal y como se adelantaba arriba, que las hipersuperficies diferenciables serán topológicas, siendo este conjunto más grande. Una vez definidas nuestras hipersuperficies formalizamos la idea de caracterización previa a través de la siguiente definición.

Proposición 5.2.1 *Un conjunto acronal A es una hipersuperficie topológica si y solo si A y su borde son disjuntos (en este contexto diremos que A no tiene borde).*

Demostración: Comencemos con la implicación a derecha. Como A es hipersuperficie topológica, dado un punto $p \in A$ podemos coger un entorno topológico tal y como el de la definición de arriba que supondremos conexo (en el sentido usual) pudiéndose suponer además que $U - A$ solo tiene dos componentes conexas. Veamos que A no tiene borde. Es directo pues, como A es acronal si consideramos los conjuntos (abiertos y conexos) $I^-(p, U)$ y $I^+(p, U)$, estos no pueden intersectar ni entre sí ni con A y, como cualquier curva temporal que pase por p interseca a ambos conjuntos, podemos concluir que cada uno está en una componente de $U - A$ y por tanto p no es un punto del borde. Como p era arbitrario concluimos esta implicación.

Para la implicación a izquierda supongamos que A no tiene borde, esto es, que A y su borde son disjuntos. Consideremos un punto $p \in A$ arbitrario y un sistema coordenado

φ en un entorno U (en definitiva una carta como en el Capítulo 1) con ∂^0 temporal y apuntando al futuro. A través de dicho sistema vamos a construir otro entorno V más pequeño tal que $\varphi(V)$ sea de la forma $(a - \epsilon, a + \epsilon) \times \mathcal{N} \subset \mathbb{R}^1 \times \mathbb{R}^{n-1}$ donde a, b y ϵ son reales y \mathcal{N} un subconjunto de \mathbb{R}^{n-1} . Además, exigimos también en este nuevo entorno que las rodajas constantes $x^0 = a$ y $x^0 = b$ estén contenidas en $I^-(p, U)$ y $I^+(p, U)$ respectivamente.

Como p no es un punto del borde, si cogemos el entorno inicial lo suficientemente pequeño y tomamos un $y \in \mathcal{N}$ la curva coordenada (también en el sentido usual) asociada a x^0 que lleva cada s a $\varphi^{-1}(s, y)$ con s entre a y b debe cortar al conjunto A . Además, este punto de corte debe ser único (también debido a la acronalidad); llamémoslo $h(y)$. Veamos que esta función h que lleva puntos de \mathcal{N} a su correspondiente (a, b) es continua. Tomemos una sucesión $\{y_n\}$ que converge a $y \in \mathcal{N}$, pero tal que $\{h(y_n)\}$ no converja a $h(y)$. Como la sucesión de h está acotada podemos tomar una subsucesión $\{h(y_k)\}$ que converja a un determinado t (que claramente no es $h(y)$). En estas condiciones es claro que llegamos a contradicción pues $\varphi^{-1}(t, y)$ estará contenida en la unión de abiertos $I^-(\varphi^{-1}(y, h(y)), V) \cup I^+(\varphi^{-1}(y, h(y)), V)$ y, consecuentemente, también lo estará $\varphi^{-1}(y_n, h(y_n))$ lo cual es totalmente incompatible con la acronalidad de A .

Una vez tenemos que h es continua tenemos el resultado pues cogiendo ahora el homeomorfismo $\psi = (x^0 - h(x^1, \dots, x^n), x^1, \dots, x^n)$ tenemos que A es hipersuperficie topológica. Nótese que ψ lleva $A \cap N$ a la rodaja constante $u^0 = 0$ de $\psi(N) \subset \mathbb{R}^n$. ■

Antes de introducir las hipersuperficies de Cauchy, definimos los conjuntos futuros y damos un corolario sobre su frontera — basado en la proposición previa — que usaremos en posteriores resultados de caracterización de hipersuperficies de Cauchy.

Definición 5.2.4 (Conjunto Futuro) *Dado un espacio-tiempo M diremos que un subconjunto F es un conjunto futuro si $I^+(F) \subset F$.*

Nótese que claramente el complementario de un conjunto futuro será un conjunto pasado definiendo conjunto pasado del mismo modo pero cambiado el futuro cronológico por el pasado (como a lo largo del capítulo, tenemos simetría en el tiempo).

Corolario 5.2.1 *La frontera de un conjunto futuro es una hipersuperficie topológica acronal cerrada.*

Demostración: Dado un conjunto futuro F , consideremos un punto de su frontera $p \in \partial F$. Veamos primero que es acronal. Si $q \in I^+(p)$ entonces $I^-(q)$ es un entorno de p y contiene al menos un punto de F , consecuentemente $q \in I^+(F) \subset F$. De este hecho, podemos ver que el futuro y el pasado cronológico de la frontera de F son

disjuntos y por tanto la frontera es acronal; nótese que hemos probado que $I^+(p) \subset F$ y equivalentemente $I^-(p) \subset M - F$.

Nos queda probar que es hipersuperficie topológica. De esta misma deducción, podemos ver cómo la frontera (que es cerrada) tiene borde vacío pues $I^+(p) \subset \text{int}(F)$ y $I^-(p) \subset \overline{\text{int}(F)}$ para todo punto de la frontera. Vamos a probar pues que un conjunto es hipersuperficie topológica cerrada si y solo si su borde es vacío.

La implicación a derecha es clara pues por la Proposición 5.2.1 nos dice que si A es hipersuperficie topológica entonces es disjunta de su borde pero, al ser cerrada, la clausura está contenida en el conjunto (y los bordes en la clausura) por tanto, el borde es vacío.

Para la implicación a izquierda, como la clausura también es acronal, si tomamos un punto que pertenezca a la clausura de A sin el propio A es claro que una curva que pase por él nunca cortará a A teniendo que pertenecer al borde de A . Por tanto $\overline{A} - A$ está contenido en el borde y A es cerrado. ■

El resultado previo vuelve a asegurarnos, como ya sabíamos, que los conos de luz en el espacio de Minkowski son hipersuperficies topológicas (no diferenciables) acronales y cerradas, ya que son la frontera del futuro (o del pasado) causal.

Llegados a este punto podemos introducir las hipersuperficies de Cauchy.

Definición 5.2.5 (Hipersuperficie de Cauchy) *Dado un espacio-tiempo M , una hipersuperficie de Cauchy es un subconjunto S que es intersecado exactamente una vez por toda curva temporal inextendible de M .*

Destacamos que una hipersuperficie de Cauchy viene a ser, desde un punto de vista físico, como la mayor abstracción de la idea de “instante de tiempo”; esto es, podemos estudiar la evolución temporal en un espacio-tiempo pasando de una superficie de Cauchy a otra. En este sentido, al igual que antes, las rodajas de espacio a tiempo constante en el espacio de Minkowski son hipersuperficies de Cauchy; sin embargo el cono de luz que también era acronal e hipersuperficie no lo es. De hecho, aunque parezca completamente ilógico, como veremos más adelante, no todo espacio-tiempo incluye una hipersuperficie de Cauchy.

Antes de continuar con el desarrollo introducimos el concepto de retracción a una hipersuperficie de Cauchy a través de un campo vectorial, que será de utilidad en la demostración del Teorema de Penrose.

Definición 5.2.6 (Retracción) *Dados una hipersuperficie de Cauchy S y un campo vectorial temporal X cuya curva integral máxima toca a S en un único punto, llamaremos retracción sobre la hipersuperficie de Cauchy S asociada al campo vectorial X a la aplicación $\rho : M \rightarrow S$ que es continua, deja S invariante punto a punto y que definimos*

como $\rho : \pi \circ \psi^{-1}$ siendo $\pi : \mathbb{R} \times S \rightarrow S$ la proyección natural y ψ^{-1} la inversa del flujo de X restringido a $S \times \mathbb{R}$.

Al principio de la sección adelantábamos que las hipersuperficies de Cauchy serían acronales. Sin embargo, no lo hemos explicitado en la definición. De hecho, las hipersuperficies de Cauchy serán hipersuperficies topológicas cerradas y reforzaremos su caracterización viendo que no solo son intersecadas por curvas temporales sino causales. Para probar esto nos valdremos del siguiente lema técnico, conocido como *Lema de la evasión*, cuya demostración puede encontrarse en [24], [18].

Lema 5.2.1 (Lema de la evasión) *Sea C un conjunto cerrado y α una curva causal inextendible hacia el pasado que empieza en un punto $p \in M$ y que no corta a C . Entonces:*

- *Dado $q \in M$, si $q \in I^+(p, M - C)$ entonces existe una curva temporal inextendible hacia el pasado que empieza en q y que no corta a C .*
- *Si α es una geodésica nula sin puntos conjugados, existe una curva temporal inextendible hacia el pasado que empieza en $\alpha(0)$ y que no corta a C .*

Proposición 5.2.2 *Una hipersuperficie de Cauchy S es una hipersuperficie topológica acronal y cerrada que es cortada por todas las curvas causales inextendibles.*

Demostración: La demostración es directa usando el Corolario 5.2.1 y dándonos cuenta de lo siguiente: de la propia definición de S podemos deducir trivialmente que M debe estar constituido por uniones disjuntas de hipersuperficies de Cauchy, sus futuros y sus pasados cronológicos. Así pues, como cualquier curva temporal que pase por S va a pasar también por su futuro y su pasado cronológico, podemos deducir que S es la frontera común a ambos concluyéndose que S es hipersuperficie topológica acronal cerrada.

Veamos finalmente que es intersecada exactamente una vez por toda curva causal. Trabajando por reducción a la absurdo, supongamos que existe una curva causal α que no interseca a S y supongamos, sin pérdida de generalidad, que empieza en un punto de $I^+(S)$. Por el lema de la evasión, debe existir una curva temporal inextendible hacia el pasado γ que empieza en $I^+(S)$ y evade S . La contradicción ahora es clara, pues basta coger cualquier curva temporal que apunte al futuro (que debe quedarse para siempre en $I^+(S)$) y concatenándola con γ tenemos una curva temporal inextendible que no corta a S , lo cual es totalmente incompatible con la definición de superficie de Cauchy con lo que se concluye la demostración. ■

A pesar de que al principio de la sección adelantábamos que las hipersuperficies de Cauchy nos permitirían dar una caracterización de los espacios globalmente hiperbólicos, damos por concluida la sección con el resultado previo, pues, para llegar a dicha

caracterización necesitamos construir nuevas estructuras causales que presentaremos en la siguiente sección.

5.3. Dominios de dependencia y horizontes de Cauchy

En la línea de la sección previa, en este apartado del documento seguiremos introduciendo nuevas estructuras causales que nos permitirán, en primer lugar, llegar a la caracterización de la hiperbolicidad global que prometíamos antes y, en segundo lugar, conseguir todos los resultados explícitamente necesarios en las demostraciones de los Teoremas de Singularidad. Podemos interpretar el desarrollo de esta penúltima sección como un prelude a dichos teoremas.

Comenzaremos definiendo los dominios de dependencia que constituirán el paso intermedio entre las hipersuperficies de Cauchy y la caracterización de la hiperbolicidad global a través de estas.

Definición 5.3.1 (Dominio de dependencia) *Dado un subconjunto acronal A de un espacio-tiempo M llamaremos dominio de dependencia futuro $D^+(A)$ (resp. pasado $D^-(A)$) de A al conjunto de los puntos p de M tales que toda curva causal inextendible hacia el pasado (resp. futuro) que pasa por ellos corta a A . Llamaremos dominio de dependencia del conjunto A a la unión de los dominios de dependencia futuro y pasado: $D(A) = D^-(A) \cup D^+(A)$.*

La interpretación física de los dominios de dependencia es bastante clara de su propia definición — y de su propio nombre — ; simplemente tenemos que percatarnos de que el propio A está contenido en su dominio de dependencia futuro (o pasado). De este modo, es claro que ninguna curva temporal o causal inextendible hacia el pasado puede llegar al dominio de dependencia sin antes haber pasado por A , es decir, ningún observador o rayo de luz puede llegar a este conjunto sin pasar por A estableciéndose una dependencia causal directa entre A y este conjunto. Así pues, el dominio de dependencia no es más que el conjunto de eventos que estarán en el futuro (causal) de A y que son predecibles físicamente desde A .

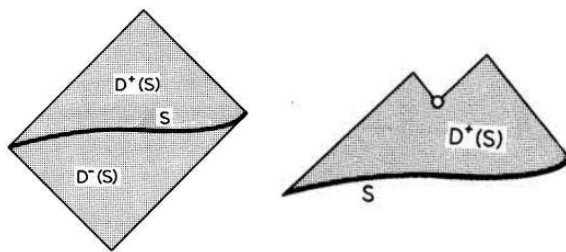


Figura 5.2: Ejemplos de dominios de dependencia.

Notemos además, después de esta definición, por qué las hipersuperficies de Cauchy son tan importantes. Recordando la caracterización previa de las hipersuperficies de Cauchy como hipersuperficies topológicas acronales cerradas que son intersecadas por toda curva causal, podemos deducir inmediatamente que, debido a esta última propiedad, el dominio de dependencia de una hipersuperficie de Cauchy es el espacio-tiempo completo. Este hecho nos permite formalizar la idea que adelantábamos cuando explicábamos la intuición física detrás de las hipersuperficies de Cauchy ya que, formalmente, podemos predecir cualquier evento en el espacio-tiempo colocándonos sobre una hipersuperficie de Cauchy (en un instante de tiempo).

Dualmente, podemos interpretar el dominio de dependencia de un conjunto A como el conjunto más grande donde A se comporta como una hipersuperficie de Cauchy.

En relación a todo lo que estamos comentando, podemos deducir — directamente de la definición — algunas igualdades conjuntistas para los dominios de dependencia que ponen de manifiesto estas propiedades. Así pues, si intersecamos los dominios de dependencia futuro y pasado, el resultado será el propio conjunto, ya que solo el mismo es predecible desde él hacia el futuro y el pasado: $D^+(A) \cap D^-(A) = A$ — esta igualdad se deduce de forma directa, tanto de la idea comentada, como del hecho de que $D^+(A) \subset A \cup I^+(A) \subset J^+(A)$ y que A es acronal. Otra propiedad interesante en el mismo sentido es que si al dominio de dependencia futuro le quitamos el propio conjunto A solo nos queda la parte del dominio de dependencia que interseca al futuro cronológico.

Los dominios de dependencia nos proporcionan el siguiente resultado que nos dará la caracterización de espacios globalmente hiperbólicos.

Teorema 5.3.1 *Si A es un conjunto acronal, entonces $\text{int } D(A)$ es globalmente hiperbólico (siempre que no sea vacío).*

La demostración del teorema, aunque interesante, ya que usa todos los conceptos que hemos desarrollado hasta ahora, es excesivamente larga para la extensión de un documento de este tipo y no aporta nada nuevo en relación a las demostraciones que estamos realizando. Puede consultarse en los textos indicados al comienzo del capítulo.

A partir del teorema, sin más que tener en cuenta la relación previa entre hipersuperficies de Cauchy y dominios de dependencia, nos da la siguiente caracterización — nótese que si S es hipersuperficie de Cauchy, $D(S) = M$, como ya hemos dicho y, trivialmente, $\text{int } D(S) = M$, pudiéndose aplicar el teorema.

Corolario 5.3.1 *Si un espacio-tiempo M tiene una hipersuperficie de Cauchy entonces es globalmente hiperbólico.*

El resultado previo puede extenderse a un si y solo si a través de técnicas bastantes más complejas que pueden consultarse en [7].

Podemos concluir por tanto que un espacio es globalmente hiperbólico si y solo si contiene una hipersuperficie de Cauchy. A través de esta caracterización podemos volver a ver cómo es bastante lógico, desde un punto de vista físico, que nuestro espacio-tiempo sea globalmente hiperbólico ya que, como es usual, suponemos que al fijar un instante de tiempo (una rodaja de espacio) podemos predecir todos los eventos en el futuro, es decir, existe una hipersuperficie de Cauchy. La discusión que aquí realizo no es en absoluto formal. Sin embargo, a través de resultados posteriores podremos asegurar que un espacio de Robertson-Walker — que suponemos que modela bastante bien nuestro universo — contiene una hipersuperficie de Cauchy y es, por tanto, globalmente hiperbólico.

Antes de adentrarnos en estos resultados, veamos un teorema sobre dominios de dependencia que será esencial en la demostración del Teorema de Singularidad de Hawking. Para la demostración del teorema necesitamos dos lemas previos. Presentamos uno con demostración y otro sin ella una vez más debido a su extensión — consúltese la bibliografía recomendada.

Lema 5.3.1 *Dado un conjunto acronal A , si p pertenece a $\text{int } D(A) - I^-(A)$, entonces $J^-(p) \cap D^+(A)$ es compacto.*

Demostración: Si p pertenece a A , el conjunto se reduce solo al propio p , así que supongamos que $p \in I^+(A) \cap \text{int } D(A)$. Dentro de $J^-(p) \cap D^+(A)$ consideremos una sucesión $\{x_n\}$ y, para cada n , α_n una curva causal apuntando al pasado que una x_n con p . En estas condiciones, por la Proposición 5.1.3 sabemos que existe una sucesión límite de la sucesión generada por dichas curvas. En particular, si dicha sucesión es finita podemos concluir que $\{x_n\}$ converge y hemos acabado en virtud del Lema 5.1.4 — nótese que es directo que el límite debe estar en el conjunto que queremos. Si la sucesión es infinita, el cuasi-límite asociado es inextendible y tocará a A en algún punto, pudiéndose deducir que x_n para algún n está en $I^-(A)$ por lo que se llega a contradicción — ya que podemos construir una subsucesión adecuada que corte a A más de una vez — con el hecho de que A es acronal. ■

Lema 5.3.2 *Dado un espacio-tiempo M , si S es una hipersuperficie topológica acausal, entonces $D(S)$ es abierto y, consecuentemente globalmente hiperbólico.*

Una vez vistos los lemas, enunciemos y probemos el teorema. Recalamos que el desarrollo aquí seguido no es único; de hecho, para la demostración del Teorema de Hawking pueden seguirse pasos parecidos a los que estamos realizando sin tener que entrar en estructuras causales más allá de la hipersuperficie de Cauchy — a través del *Teorema de Avez-Seifert* [24]. La línea argumental seguida en el documento se debe a que estos conjuntos tienen una mayor riqueza desde un punto de vista físico, sin perder el interés matemático, mientras que la otra línea tiene un difícil acercamiento desde la física.

Teorema 5.3.2 *Dada una hipersuperficie espacial acronal cerrada S en un espacio-tiempo M , si q pertenece a $D^+(S)$, entonces existe una geodésica γ de S a q de longitud $\tau(S, q)$ y consecuentemente γ es normal a S y no tiene puntos focales en S previos a q .*

Demostración: Antes de comenzar la demostración nótese que γ es temporal excepto en el caso trivial de que $q \in S$; además debido a la simetría temporal tendremos también el resultado en el caso dual extendiéndose a todo el dominio de dependencia.

Para la demostración, del lema anterior podemos deducir directamente que $D(S)$ es abierto y globalmente hiperbólico. Además del Lema 5.3.1 podemos deducir que $J^-(q) \cap S$ es compacto (ya que lo es $J^-(q) \cap D^+(S)$). Además, como tenemos que es globalmente hiperbólico, sabemos que la separación temporal es continua (Lema 5.1.5) y por el Teorema de Weierstrass alcanza un máximo que es exactamente $\tau(S, q)$. La existencia de la geodésica con esta longitud nos la proporciona la Proposición 5.1.4. La normalidad y la imposibilidad de que haya puntos focales nos la dan el Corolario 3.3.2 y el Teorema 3.3.4 respectivamente — téngase en cuenta, como aclaramos al principio, que si q no está en S entonces $p \ll q$ y tenemos que la curva es temporal. ■

Para acabar la sección, introducimos la última estructura causal que nos será necesaria: el horizonte de Cauchy. Relacionando este objeto con los dominios de dependencia obtendremos un resultado de vital importancia en la demostración del Teorema de Hawking. Además, esta estructura nos permitirá determinar que el espacio de Robertson-Walker es, como adelantábamos, globalmente hiperbólico.

El horizonte de Cauchy, hablando burdamente, no es más que el límite en el futuro del dominio de dependencia, es decir, si el horizonte de Cauchy (futuro) no es vacío entonces existe una zona a partir de la cual el futuro del conjunto no es predecible desde el propio conjunto. Formalmente, como veremos de la propia definición, el horizonte de Cauchy (futuro) no es más que la frontera futura del dominio de dependencia.

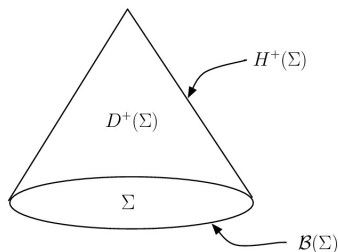


Figura 5.3: Ejemplo de una hipersuperficie espacial Σ con su dominio de dependencia y su horizonte de Cauchy.

Definición 5.3.2 (Horizonte de Cauchy) *Dado un conjunto acronal A , llamaremos horizonte de Cauchy futuro de A al conjunto*

$$H^+(A) = \overline{D^+(A)} - I^-(D^+(A)) = \left\{ p \in \overline{D^+(A)} \text{ tal que } I^+(p) \text{ interseca con } D^+(A) \right\}.$$

De forma dual, definimos exactamente igual el horizonte de Cauchy pasado de A , $H^-(A)$ y llamaremos horizonte de Cauchy de A a la unión de ambos $H(A) = H^-(A) \cup H^+(A)$.

Seguidamente presentamos los resultados necesarios anteriormente mencionados con sus demostraciones.

Lema 5.3.3 *Dado un conjunto acronal cerrado A , la clausura de su dominio de dependencia futuro $\overline{D^+(A)}$ coincide con el conjunto formado por todos los puntos p tales que toda curva temporal inextendible hacia el pasado que pasa por ellos corta al conjunto A .*

Demostración: Llamemos al conjunto del enunciado B y procedamos por doble inclusión. Veamos primero que la clausura está en B . Para ello tomemos un $p \in \overline{D^+(A)} - B$ y busquemos una contradicción. En estas condiciones, existe una curva temporal inextendible hacia el pasado que empieza en p y que evade a A por lo que p no pertenece a A y existirá un entorno convexo C de dicho punto que no interseque con A . Si nos movemos hacia atrás en la curva sin salirnos de C hasta un punto t es claro que $I^+(t, C)$ contiene a p y que por tanto $q \in D^+(A)$. Veamos que esto último es imposible pues si tomamos una geodésica que una q con t dentro del convexo y la concatenamos con el resto de la curva inicial tendremos una nueva curva temporal inextendible hacia el pasado que no toca a A .

Para la segunda inclusión veamos que si un punto p no pertenece a la clausura entonces no pertenece a B . Esto también es directo por resultados previos pues, si tomamos $t \in I^-(p, M - \overline{D^+(A)})$, existe una curva causal inextendible hacia el pasado que empieza en p y que no toca a A . El resultado se sigue trivialmente del Lema de la evasión. ■

Proposición 5.3.1 *Sea A un conjunto acronal cerrado. Entonces $\partial D^+(A) = A \cup H^+(A)$.*

Demostración: Demostramos de nuevo por doble inclusión. La inclusión a la izquierda es trivial por definición del horizonte de Cauchy sin más que considerar que A es acronal. Para la siguiente inclusión trabajaremos como en la demostración del lema previo suponiendo que $p \in \partial D^+(A) - A - H^+(A)$ y llegando a contradicción. Si p pertenece a este conjunto, en particular pertenece a $\partial D^+(A) - A$ y a $\partial D^+(A) - H^+(A)$, pudiéndose deducir de lo primero (por el Lema 5.3.3) que $p \in I^+(A)$, y de lo segundo que existe un $q \in I^+(p) \cap D^+(A)$. Sin más que tener en cuenta que $I^+(A) \cap I^-(q)$ es un entorno de p , tirando hacia detrás, llegamos a contradicción con el hecho de que p esté en $\partial D^+(A)$. Concluyéndose así la demostración. ■

A través de estos resultados intermedios llegamos a la proposición que nos interesa para el Teorema de Singularidad.

Proposición 5.3.2 *Sea S una hipersuperficie topológica acausal cerrada. Entonces: $H^+(S) = I^+(S) \cap \partial D^+(S) = \overline{D^+(S)} - D^+(S)$. En particular, $H^+(S)$ y S son disjuntos.*

Demostración: Abordemos el problema por doble inclusión. La inclusión hacia la izquierda es más sencilla; comencemos con la primera y dejemos esta para el final. En primer lugar, por el Lema 5.3.3 tenemos la inclusión $H^+(S) \subset \overline{D^+(S)} \subset S \cup I^+(S)$. Además $H^+(S)$ no toca a $D^+(S)$ pero $I^+(S)$ sí, ya que, por el Lema 5.3.2 el dominio de dependencia de S es abierto y si $p \in D^+(S) \subset S$ entonces $I^+(p)$ toca a $D(S)$ pero no su parte pasada. Finalmente, recordando que un conjunto siempre está en su dominio de dependencia, podemos deducir que $H^+(S) \subset I^+(S)$ e intersectando este último con los conjuntos de la Proposición 5.3.1 tenemos que $I^+(S) \cap \partial D^+(S) = H^+(S)$ deduciéndose lo que queríamos: $H^+(S) \subset \overline{D^+(S)} - D^+(S)$.

Para la inclusión contraria tomemos un punto p de $\overline{D^+(S)} - D^+(S)$. Veamos que dado un punto $q \in I^+(p)$ este no puede pertenecer al dominio de dependencia futuro de S y por tanto p está en el horizonte de Cauchy. Para ello, si tomamos este q sabemos que existe una curva temporal hacia el pasado que une p y q que no toca a S ya que p no pertenece ni a S ni a su pasado cronológico. Deducimos el resultado ya que también existe una curva causal inextendible hacia el pasado que empieza en p y que tampoco toca a S . ■

Finalmente, introduzcamos las ideas que nos permiten probar que el espacio de Robertson-Walker es globalmente hiperbólico.

Proposición 5.3.3 *Sea S una hipersuperficie topológica acausal cerrada. Entonces existen geodésicas luminosas inextendibles hacia el pasado sin puntos conjugados empezando en cada punto de $H^+(S)$ y que están completamente contenidas en él. Llamaremos a este tipo de geodésicas las generadoras del horizonte de Cauchy.*

Demostración: La idea detrás de que estas curvas sean generadoras es que se extienden lo máximo posible hacia el futuro en el horizonte. Para demostrar su existencia, tomemos un punto p del horizonte del cual puede partir (por la proposición anterior) una curva causal inextendible hacia el pasado que no toca a S y, que, de hecho, no puede ser temporal por el Lema 5.3.3 y por tanto, será, en virtud del Lema de evasión, una geodésica luminosa sin puntos conjugados.

Para acabar la demostración solo tenemos que demostrar que dicha curva no sale del horizonte. En primer lugar, es directo (ya que no toca a S) que la curva no toca a su dominio de dependencia futuro. Por tanto, para concluir la demostración solo tenemos que ver que la curva siempre pertenece a la clausura del dominio de dependencia futuro.

Si esto no fuera así, es decir, existe un punto de la curva que no pertenece a $D^+(S)$ entonces existiría otra curva temporal apuntando al pasado e intextensible hacia el pasado que empezaría en dicho punto y que tampoco tocaría a S . Este resultado, usando una vez más el Lema de evasión, en este caso sobre una concatenación adecuada de las dos curvas anteriores — basta tomar la inicial restringida a $[0, t]$ siendo t el parámetro del punto en que la curva se sale de $D^+(S)$ — nos lleva a contradicción con el hecho de que la curva empieza en $p \in D^+(S)$. ■

Lema 5.3.4 *Una hipersuperficie topológica acausal cerrada es una hipersuperficie de Cauchy si y solo si $H(S)$ es vacío.*

Demostración: El enunciado es equivalente a probar que $H(S)$ es vacío si y solo si $D(S) = M$. Por tanto, nos basta probar que $\partial D(S) = H(S)$. Probémoslo por doble inclusión. La inclusión a izquierda es trivial. Para ver la inclusión a derecha, nos basta usar la Proposición 5.3.1, a través de la cual deducimos que $\partial D(S) \subset S \cup H(S)$ pero, como $D(S)$ es abierto y contiene a S se tiene que $\partial D(S) \subset H(S)$. ■

Nótese que la conclusión previa es lógica en concordancia con lo que significan físicamente tanto la hipersuperficie de Cauchy como el horizonte. Si la hipersuperficie de Cauchy es aquella desde la que podemos predecir cualquier evento del espacio-tiempo completo es necesario que el horizonte sea vacío pues, en caso contrario, existiría un límite a partir del cual no podríamos predecir nada desde la hipersuperficie.

Corolario 5.3.2 *Una hipersuperficie topológica acausal cerrada S es una hipersuperficie de Cauchy si toda geodésica luminosa inextendible corta a S .*

Demostración: Para probarlo solo tenemos en cuenta el lema previo y demostramos que $H(S)$ es vacío. Trabajando por reducción al absurdo, supongamos que en las condiciones del corolario existe un punto p en el horizonte. Entonces, por la proposición anterior las curvas generadores del horizonte a través de p no tocan a S y, tampoco podrán tocarlo al extenderse ya que S es acronal y el punto inicial está en su futuro cronológico llegándose a contradicción. ■

En particular, el corolario fundamenta de forma sólida nuestra suposición de que nuestro universo es globalmente hiperbólico puesto que demuestra que el espacio de Robertson-Walker lo es (siempre que su fibra sea completa como ya adelantamos). En concreto, basta tomar cualquier rodaja del espacio-tiempo a tiempo constante, esto es, $t \times S$ con S la fibra del espacio de Robertson Walker. Es claro que $t \times S$ es una hipersuperficie cerrada y acausal y, además, por construcción — recuérdese el Capítulo 4 — toda geodésica luminosa corta dicha hipersuperficie aplicándose el lema previo. Por tanto, el espacio de Robertson-Walker contiene una hipersuperficie de Cauchy y, en virtud, del teorema de caracterización, es globalmente hiperbólico.

5.4. Teoremas de singularidad

Concluimos el documento llegando — al fin — a nuestros ansiados Teoremas de Singularidad. Como venimos explicando a lo largo de todo el texto, estos teoremas darán unas condiciones generales para asegurar la aparición de singularidades en el espacio-tiempo independientemente de las características del modelo que estemos usando sobre las ecuaciones de campo de Einstein. En este sentido, presentaremos dos teoremas, el Teorema de Hawking que generalizará el Teorema 4.3.2 sobre singularidades en el espacio-tiempo de Robertson-Walker (modelo FLRW), y el Teorema de Penrose que demostrará que la singularidad en el modelo de Schwarzschild no está ligada a la simetría de este.

Antes de comenzar a introducir los teoremas en sí mismos, tenemos que formalizar la idea de singularidad que introducimos en el Capítulo 4 tal y como adelantamos en la introducción a este capítulo.

Definición 5.4.1 (Espacio-tiempo singular) *Diremos que un espacio-tiempo M es singular o que contiene singularidades si M es geodésicamente incompleto. En particular, diremos que M es singular en el sentido temporal (resp. nulo) si es geodésicamente incompleto temporal (resp. nulo), es decir, si el carácter causal de las geodésicas asociadas a su incompletitud es temporal (resp. nulo).*

La definición anterior aún puede concretarse más indicando en qué dirección temporal es incompleto el espacio-tiempo. Así, diremos que un espacio-tiempo es *singular en el futuro* si es geodésicamente incompleto en el futuro (resp. pasado), es decir, si sus geodésicas no están definidas en todo su intervalo hacia el futuro (resp. pasado).

La motivación de la definición es clara en este punto del documento, pues las singularidades pueden interpretarse como puntos aislados a los que no llega información o puntos que no están en cierto sentido en el espacio-tiempo — puntos fuera del modelo relativístico — ya que impiden que las geodésicas maximales puedan definirse en la línea real completa teniendo que “detenerse” debido a la existencia de estas anomalías. Así, la idea es detectar las singularidades viendo que las geodésicas (rayos de luz, observadores) no pueden extenderse indefinidamente — considérese, por ejemplo, las singularidades del modelo de Robertson-Walker en que todas las curvas convergen a un punto determinado.

Con estas ideas en mente, la estructura de los Teoremas de Singularidad será siempre la misma, necesitaremos tres condiciones: una condición sobre la energía del espacio-tiempo, una condición sobre la causalidad global del espacio-tiempo y una condición que nos asegure que la gravedad es lo suficientemente fuerte como para colapsar, en un sentido que precisaremos a continuación, una región del espacio-tiempo.

La condición de causalidad global será la misma en ambos teoremas por todo lo

justificado en la sección previa: exigiremos que el espacio-tiempo sea globalmente hiperbólico.

La condición sobre la energía también será la misma y será la siguiente:

Definición 5.4.2 (Condición de energía fuerte) *Diremos que un espacio-tiempo M con tensor energía-momento T cumple la condición de energía fuerte si $T(u, u) \geq \frac{1}{2} \mathbf{C}(T)g(u, u)$ para todo u vector luminoso o temporal tangente a M .*

Nótese que la condición de energía fuerte, a través de las ecuaciones de campo de Einstein, es equivalente a exigir que $Ric(u, u) \geq 0$ para todo u vector luminoso o temporal tangente a M . Esta forma de escribir la condición es bastante más intuitiva ya que, a grosso modo, nos dice que la gravedad atrae, es decir, en cierto modo los vectores temporales convergen (al menos en media). De hecho, esta condición así escrita suele conocerse como condición de convergencia temporal. Sin embargo, hemos decidido escribirla en su forma energética pues es más común en la literatura.

Finalmente, tenemos que exigir la condición que nos asegure que una zona del espacio-tiempo vaya a colapsar debido a lo fuerte que es la gravedad. En cada Teorema de Singularidad esta condición será diferente pero, como venimos adelantando en varias partes del documento, dicha condición estará relacionada con la convergencia k . Por la propia definición de este concepto podemos intuir que la convergencia mide, valga la redundancia, cómo converge o cómo se focaliza una región del espacio-tiempo ya que esencialmente estamos estudiando la proyección de la curvatura media sobre un determinado vector normal a la región en cuestión, lo cual nos da una visión de cómo varía la forma — segunda forma fundamental — de esa región a medida que pasa el tiempo — si converge (colapsa) o si diverge (se estira).

Debido a que, como hemos dicho, esta última condición se interpreta de modo algo diferente en cada teorema, separaremos a partir de aquí el desarrollo. Comenzaremos con el Teorema de Penrose que generalizará la singularidad que aparece en el modelo de Schwarzschild dando condiciones generales para la aparición de singularidades en el colapso de estrellas.

La idea en la que nos basamos estará directamente relacionada con la característica principal que destacamos en un agujero negro de Schwarzschild. En estos agujeros negros, la gravedad es tan fuerte que incluso los rayos de luz que deberían salir de la superficie del agujero negro — siendo formales: esferas $S^2(r)$ a tiempo constante en el espacio-tiempo del agujero negro — caen hacia el interior del agujero negro. Si recordamos la causalidad en el espacio de Minkowski, esto quiere decir que la gravedad es tan fuerte que los conos de luz se giran apuntando hacia el interior del agujero negro. Desde un punto de vista físico, si los rayos de luz no pueden escapar, siendo estos los sistemas que viajan a velocidad máxima, nada podrá escapar de la región del agujero negro de ahí el nombre que venimos comentando de *colapso* o *superficie atrapada*.

Como veníamos diciendo, esta situación de colapso estará matemáticamente relacionada con la convergencia. Podría surgirnos la pregunta de por qué no caracterizamos este fenómeno a través del hecho directo de que los rayos de luz caen hacia el interior del agujero negro. La respuesta es clara, y está directamente relacionada con la discusión que venimos realizando en todo el documento: la solución de Schwarzschild goza de gran simetría y es por ello que podríamos escribir esta condición de forma sencilla a través de la coordenada r . Sin embargo este no es el caso general. Nuestra idea será caracterizar esta situación de colapso a través de una convergencia positiva.

De hecho, en una subvariedad espacial Q de codimensión mayor o igual que 2 (al igual que aquellas del modelo de Schwarzschild) es directo demostrar, a través de álgebra en el normal del espacio tangente a la subvariedad, que si la convergencia es positiva para todo vector causal normal a la subvariedad, entonces su curvatura media es temporal y apunta al pasado. Esto nos lleva a la siguiente definición:

Definición 5.4.3 (Superficie atrapada) *Una subvariedad espacial de M es una superficie atrapada si su curvatura media es temporal y apunta al pasado.*

Nuestro siguiente paso será transformar esta idea geométrico-diferencial de superficie atrapada a un concepto causal.

Definición 5.4.4 (Conjunto atrapado) *Diremos que un conjunto acronal cerrado A es un conjunto atrapado en el futuro (resp. pasado) si el conjunto $E^+ := J^+(A) - I^+(A)$ (resp. $E^- := J^-(A) - I^-(A)$) es compacto.*

A continuación vemos que estas ideas son equivalentes bajo determinadas condiciones.

Teorema 5.4.1 *Si se da la condición de energía fuerte para vectores nulos y M es geodésicamente completo nulo en el futuro, entonces si Q es una subvariedad espacial de M de dimensión $n - 2$ acronal y compacta que es superficie atrapada, también es conjunto atrapado.*

Demostración: Tenemos que probar que $E^+(Q)$ es compacto. Tomemos en cada punto de Q un par de vectores nulos que apunten al futuro y que pertenezcan a campos vectoriales normales independientes. Llamemos al conjunto de estos vectores Θ , que claramente pertenecerá al fibrado normal y será compacto por ser recubrimiento. Por la Proposición 3.3.1 para cada vector v de este conjunto existirá un punto focal de Q a lo largo de una geodésica γ_v restringida a $[0, 1/k(v)]$ y, al ser k continua en un compacto, existirá un número positivo b de modo que para cada $v \in Q$ habrá un punto focal de Q a lo largo de γ_v restringida a $[0, b)$.

Si tomamos un punto $q \in E^+(Q)$, como hemos visto, existe una geodésica γ nula de Q a q . Además, como q no pertenece a $I^+(Q)$ sabemos que la geodésica es normal a Q y que no tiene puntos focales de Q antes que q (Teorema 3.3.5). Por tanto, γ debe ser una reparametrización de alguna γ_v y podemos decir que $E^+(Q) \subset \exp(K)$ con K :

$$K = \{tv \mid v \in \Theta, t \text{ en el fibrado normal y } s \in [0, b]\}$$

y como Θ y K son compactos también lo será $\exp(K)$. Demostrar que $E^+(Q)$ es compacto es ahora trivial pues, tomando una sucesión en $E^+(Q)$, es claro que cualquier subsucesión convergerá a un punto de $\exp(K) \subset J^+(Q)$ que, como no puede estar en $I^-(Q)$ estará en $E^+(Q)$. Concluye así la demostración. ■

Notemos que las condiciones del Teorema previo son totalmente razonables como veremos a continuación al enunciar el Teorema de Penrose. En particular, la condición de energía fuerte, como dijimos antes, es completamente necesaria. La idea detrás del Teorema es que si existe una superficie atrapada que está colapsando y el tensor de Ricci es positivo para vectores nulos, es decir, si se da la condición de energía fuerte para vectores nulos, entonces dicho colapso continuará de modo que se producirá una concatenación de geodésicas nulas convergentes que provocará la existencia de la singularidad.

La forma en que enunciaremos el Teorema de Penrose no será como un Teorema de Singularidad propiamente dicho, sino que, enunciaremos el Teorema como una serie de propiedades que deben darse simultáneamente para que otra se cumpla y, a partir de este, presentaremos dos corolarios completamente directos que serán los resultados de singularidad en sí mismos.

Teorema 5.4.2 (Penrose) *Sea M un espacio-tiempo cumpliendo las siguientes condiciones:*

- M cumple la condición de energía fuerte para vectores nulos.
- M es globalmente hiperbólico.
- M contiene una subvariedad espacial Q de dimensión $n - 2$ acronal y compacta que es superficie atrapada.

Entonces $E^+(Q)$ es una hipersuperficie de Cauchy de M .

Demostración: Probemos que $E^+(Q)$ es hipersuperficie de Cauchy a través de su caracterización como hipersuperficie topológica acronal y cerrada que es cortada por todas las curvas causales inextendibles. Veamos que es una hipersuperficie topológica cerrada, ya que es claro que es acronal. De hecho, vamos a probar que es variedad topológica compacta ya que será la frontera de un conjunto futuro cerrado. Por la

Proposición 5.1.6 tenemos que \leq es cerrada y, consecuentemente, como Q es compacto, por el Lema 5.1.2 $J^+(Q)$ es cerrado. Finalmente, a través del Lema 5.1.3 tenemos que $\text{int}J^+(Q) = I^+(Q)$ y, por tanto, $E^+(Q) = \partial J^+(Q)$ como queríamos probar.

Acabemos de demostrar que es hipersuperficie de Cauchy aprovechando que sabemos que existe una hipersuperficie de Cauchy S en M por ser globalmente hiperbólico. Tomemos la restricción a $E^+(Q)$ de una retracción por un campo temporal X a S , $\rho : E^+(Q) \rightarrow S$ que será continua y biyectiva (a través de la unicidad de curvas integrales y la acronalidad de $E^+(Q)$). Además, por invariancia del dominio ρ es un homeomorfismo de $E^+(Q)$ a un abierto de S ; de hecho, será un homeomorfismo entre $E^+(Q)$ a S ya que $E^+(Q)$ es compacto y su imagen por ρ será cerrada en S que a su vez es conexo teniéndose que $\rho(E^+(Q)) = S$.

Para acabar la demostración veamos que toda curva temporal inextendible corta a $E^+(Q)$ (al ser acronal no hace falta probar nada más). Ahora bien, como M es globalmente hiperbólico en particular será establemente causal y, por tanto, toda curva temporal podrá ser vista como una curva integral de un campo vectorial temporal X . Pero acabamos de ver que la retracción a través de un campo de este tipo produce un homeomorfismo. En consecuencia, si la curva temporal corta a S también cortará a $E^+(Q)$ concluyéndose el resultado. ■

Recalcamos que la demostración aportada no es la más usual. En general suele probarse el Teorema de Penrose a través de la ecuación de Raychadhuri [17], [5]. Sin embargo, la demostración aquí realizada se limita a usar conceptos que hemos tratado previamente.

Introducimos ahora los Teoremas de Singularidad como Corolarios del Teorema de Penrose.

Corolario 5.4.1 (Teorema de Singularidad de Penrose (I)) *Sea M un espacio-tiempo cumpliendo las siguientes condiciones:*

- *M cumple la condición de energía fuerte para vectores nulos.*
- *M es globalmente hiperbólico.*
- *M contiene una subvariedad espacial Q de dimensión $n - 2$ acronal y compacta que es superficie atrapada.*

Si M contiene una hipersuperficie de Cauchy no compacta, entonces M es singular en el sentido nulo.

Corolario 5.4.2 (Teorema de Singularidad de Penrose (II)) *Sea M un espacio-tiempo cumpliendo las siguientes condiciones:*

- M cumple la condición de energía fuerte para vectores nulos.
- M es globalmente hiperbólico.
- M contiene una subvariedad espacial Q de dimensión $n - 2$ acronal y compacta que es superficie atrapada.

Si existe una curva causal inextendible que no corta a $E^+(Q)$, entonces M es singular en el sentido nulo.

Concluimos, como prometíamos desde el principio del documento, que a través del, o mejor dicho, los Teoremas de Singularidad de Penrose, hemos liberado a la singularidad del modelo de Schwarzschild de su simetría, generalizando la posible aparición de singularidades en forma de agujeros negros en estrellas que colapsen siempre que podamos asegurar las condiciones de los Teoremas. Dichas condiciones, como hemos justificado a lo largo del capítulo, son más que razonables para un espacio-tiempo como el nuestro — al menos en concordancia con la experimentación realizada hasta el presente — ya que sabemos que la gravedad atrae en media — condición de energía fuerte, — es más que esperable que existan hipersuperficies de Cauchy por razones de predictibilidad, por como es nuestro espacio-tiempo de forma local — hiperbolicidad global —, y es bien conocido que se producen superficies colapsantes en zonas de gravedad desmesuradamente alta como pueden ser las estrellas colapsantes — superficies atrapadas. Todo ello justifica la aparición de numerosos agujeros negros en nuestro universo ya que se podrán formar en condiciones de alta o baja simetría siempre que se cumpla lo anterior.

Pasemos a continuación a introducir el último resultado del documento: el Teorema de Singularidad de Hawking. Existen varios enunciados de este Teorema. Sin embargo, aquí simplemente enunciaremos aquel que prueba más aunque también exige más. El Teorema volverá a tener la estructura previamente mencionada y volveremos a tratar con la convergencia k para modelar que la gravedad es suficientemente fuerte, exigiendo que esta sea positiva y mayor que un determinado valor. Nótese que hemos dicho valor, pues recuérdese que la finalidad del Teorema de Hawking es generalizar el resultado que obtuvimos para el modelo de FLRW en el que aparecían singularidades iniciales, y consecuentemente estaremos tratando con el universo como un todo en este sentido. Así pues, nos interesará la convergencia de una hipersuperficie S de un espacio-tiempo sobre un vector normal unitario n a dicha hipersuperficie que apunte al futuro. En este caso, la convergencia se reduce a una función real: $k = g(n, \vec{H})$. Recalcamos además que al estar tomando el vector normal apuntando hacia el futuro llamaremos a esta convergencia k *convergencia futura*. Al igual que en todo el documento, a pesar de que el resultado que nos interesa se produce en el pasado — Big Bang — enunciaremos el resultado hacia el futuro siendo completamente equivalente debido a la simetría del tiempo — bastaría si no cambiar la convergencia futura por la convergencia pasada tomando un n que apunte al pasado y en la demostración del Teorema cambiar todos los conjuntos en el pasado en lugar de en el futuro.

La motivación detrás del Teorema es parecida a la de Penrose solo que desde un punto de vista cosmológico. Sabemos que las galaxias divergen. Sin embargo, como la gravedad atrae en media — condición de energía fuerte — podemos deducir que divergían mucho más lento en el pasado. Si extrapolamos esta idea a un pasado suficientemente lejano, en concreto, a épocas cercanas al nacimiento del cosmos — si es que este existe — es bastante lógico esperar que encontraremos un punto — singularidad — donde todas las galaxias colapsen.

Teorema 5.4.3 (Hawking) *Sea M un espacio-tiempo cumpliendo las siguientes condiciones:*

- *M cumple la condición de energía fuerte para vectores temporales.*
- *M es globalmente hiperbólico.*
- *Existe una hipersuperficie de Cauchy espacial S con un ínfimo de expansión b , es decir, su convergencia futura k cumple que: $k \geq b > 0$.*

Entonces, la longitud de cualquier curva temporal dirigida al futuro y que empiece en S es como máximo $1/b$.

Demostración: La demostración será consecuencia del Teorema 5.3.2 y el Teorema 3.3.4. En virtud de estos teoremas sabemos que existe una geodésica γ normal de S a $q \in D^+(S) - S$ sin puntos focales antes que q y con longitud $\tau(S, q) \leq 1/b$ (ya que si no habría un punto focal antes que q). En este sentido: $D^+(S) \subset \{p \in M \mid \tau(S, p) \leq 1/b\}$.

Tendremos el resultado a través de la definición de la separación temporal si demostramos que $I^+(S) \subset D^+(S)$. Pero esto es directo ya que, si una curva temporal que apunta al futuro desde S sale de $D^+(S)$, tocará a $\partial D^+(S)$ y, por la Proposición 5.3.2, tendríamos que $H^+(S)$ no sería vacío llegándose a contradicción, pues si S es de Cauchy su horizonte futuro tendría que ser vacío. Concluimos así la demostración. ■

Nótese que, obviamente, la implicación del teorema demuestra que M es singular. Hemos conseguido por tanto generalizar el resultado que obtuvimos para el caso del modelo FLRW, ya que la condición de isotropía que exigíamos ya no es necesaria. A pesar de que hemos liberado el resultado de este modelo concreto, es bastante lógico que nuestro universo se comporte como un espacio-tiempo de Robertson-Walker por todo lo que ya hemos argumentado a lo largo del capítulo. El hecho de que el Teorema de Hawking se pueda aplicar directamente al resultado obtenido para el modelo FLRW es fácil de probar. Si tomamos el espacio-tiempo de Robertson Walker con espacio completo y consideramos como hipersuperficie del teorema una rodaja de espacio a tiempo constante, es claro que dicha hipersuperficie es de Cauchy, se cumple la condición de energía fuerte — esta es equivalente a las condiciones que imponíamos en el modelo FLRW sobre la densidad y la presión — y además también se cumple la condición sobre

la convergencia, en este caso hacia el pasado, ya que es equivalente a que el universo se expanda debido a la constante de Hubble. Concluimos por tanto que, tanto si nuestro universo es del tipo FLRW como si no lo es, es bastante lógico que se cumplan las condiciones del Teorema de Hawking implicando que nuestro universo nació de una singularidad.

Concluimos el documento con los resultados que prometíamos. Hemos estudiado la naturaleza del espacio y el tiempo a través de la teoría de variedades diferenciables y variedades semi-Riemannianas, pasando de nuestra idea de espacio absoluto tridimensional con un parámetro que modela el cambio llamado tiempo, a la construcción de una entidad conjunta llamada espacio-tiempo, mutable y dinámica en la que el tiempo es solo una dimensión más con características peculiares. Además, hemos analizado cómo el espacio-tiempo es el objeto clave en la física actual a gran escala, constituyendo el corazón de la Teoría de la Relatividad General y modelando la gravedad y la física de nuestro cosmos a través de su geometría.

Finalmente, hemos concluido, a través de los Teoremas de Singularidad, que en espacios-tiempo de características similares al nuestro — desde el punto de vista físico y causal — se producen singularidades, puntos donde el espacio-tiempo se rompe y la Teoría de la Relatividad General no funciona. Estas singularidades — a grosso modo el Big Bang y los agujeros negros — son limitaciones actuales de la física a gran escala. Son puntos externos al espacio-tiempo que usamos y no forman parte de la teoría física en sí misma. Actualmente, desde el mundo físico dichas singularidades no son aceptadas como realidades físicas — hablando burdamente, en física no se aceptan los infinitos —; es por ello que gran parte de la comunidad trabaja arduamente hoy en día con el objetivo de extender la teoría a dichos puntos problemáticos y entender mejor tanto el inicio como el fin del tiempo. Las teorías de mayor renombre actualmente que consiguen esta extensión, al menos a nivel teórico, son *La teoría de cuerdas* y *La teoría cuántica de bucles*, que pretenden unir el mundo de la física cuántica — pequeñas escalas — con la cosmología — grandes escalas — dando lugar a una teoría unificada con gravedad cuántica. Sin embargo, ninguna de ellas cuenta con un apoyo experimental consistente.

Así pues, este problema sigue quedando abierto, esperando que la perspicacia de físicos y matemáticos en la intersección de sus disciplinas — Física Cuántica y Análisis Funcional, Relatividad General y Geometría Diferencial, Análisis Numérico y simulaciones, Geometría Discreta etc. — junto con un profundo conocimiento filosófico de la realidad sea capaz de resolverlo.

Bibliografía

- [1] Alcubierre, M. y Lobo, F. *Warp drive basics*. arXiv preprint arXiv:2103.05610 (2021).
- [2] Chen, B.-Y. *Pseudo-Riemannian geometry, δ -invariants and applications*. World Scientific, Singapur, 2011.
- [3] Clarke, C.J.S. *The Analysis of Space-Time singularities*. Cambridge University Press, Cambridge, 1994.
- [4] Ebner, D. *How Hilbert has found the Einstein Equations before Einstein and forgeries of Hilbert's page proofs*. arXiv preprint arXiv:physics/0610154 (2006).
- [5] Galloway J. G. *Notes on Lorentzian Causality*. ESI-EMS-IAMP Summer school on Mathematical Relativity. Department of Mathematics. University of Miami, 2014.
- [6] García, J. *Cosmology and Astrophysics*. arXiv preprint arXiv:astro-ph/0502139 (2005).
- [7] Geroch, R. *What is a singularity in General Relativity*. Ann. Phys., 48. 1968.
- [8] Griffiths, D. J. *Introduction to Electrodynamics*. Prentice-Hall, New Jersey, 1999.
- [9] Hawking, S. W. y Ellis, G. F. *The large scale structure of space-time*. Cambridge University Press, Cambridge, 1973.
- [10] Hawking, S.W. and Penrose, R. *La naturaleza del espacio y el tiempo*. Debolsillo, 2018.
- [11] Hooft, G. *Introduction to the theory of Black Holes*. Institute for Theoretical Physics Utrecht University and Spinoza Institute. Postbox 80.195. 3508 TD Utrecht, the Netherlands, 2009.
- [12] Jackson, N. *The Hubble Constant*. University of Manchester, Jodrell Bank Centre for Astrophysics. arXiv preprint arXiv:0709.3924. (2007).
- [13] Landau, L. D. y Lifshitz. E. M. *Fluid Mechanics*. Vol. 6 (1st ed.). Pergamon Press, 1959.

-
- [14] Landau, L. D. y Lifshitz. E. M. *Mechanics*. Vol. 1 (1st ed.). Pergamon Press, 1960.
- [15] Landau, L. D. y Lifshitz. E. M. *The Classical Theory of Fields*. Vol. 2 (2nd ed.). Pergamon Press, 1959.
- [16] Lee, J. M. *Introduction to Smooth Manifolds*. Springer, Nueva York, 2013.
- [17] Nátario, J. y Godinho, Leonor. *An Introduction to Riemannian Geometry with Applications to Mechanics and Relativity*. Springer, 2014.
- [18] O'Neill, B. *Semi-Riemannian geometry with applications to relativity*. Academic Press, Londres, 1983.
- [19] Oppenheimer, J.R. and Snyder, H. *On Continued Gravitational Contraction*. Physical Review Letters, 54, 455-459. 1939.
- [20] Pooley, O. *The Hole Argument*. arXiv preprint arXiv:2009.09982 (2020).
- [21] Quintero, A. Apuntes de *Varietades Diferenciables*. Departamento de Geometría y Topología. Facultad de Matemáticas. Universidad de Sevilla, 2020.
- [22] Romero, A. *Geometría de Lorentz: de lenguaje a herramienta básica en Relatividad General*. Facultat de Matemàtiques i Estadística. Universitat Politècnica de Catalunya.
- [23] Sachs, R.K. y Wu, H. *General Relativity for Mathematicians*. Graduate Texts in Mathematics (1st ed). Springer, 1977.
- [24] Sánchez, M. y Javaloyes, M. *An introduction to Lorentzian Geometry and its applications*. Departamento de Geometria y Topologia. Facultad de ciencias. Universidad de Granada.
- [25] Senovilla, J. y Garfinkle, D. *The 1965 Penrose singularity theorem*. arXiv preprint arXiv:1410.5226v1 (2014).
- [26] Weinberg, S. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. Wiley, 1972.

Las imágenes usadas para las figuras del Capítulo 1, 3 y 5 pertenecen a los textos [24], [8] y [5], respectivamente.