



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

ÓRBITAS ESTELARES COMO HERRAMIENTA PARA ESTUDIAR
LA NATURALEZA DEL ESPACIO-TIEMPO

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

FÍSICO

P R E S E N T A :

FERNANDO VÁZQUEZ CHÁVEZ

TUTOR

DR. NÉSTOR ENRIQUE ORTIZ MADRIGAL
INSTITUTO DE CIENCIAS NUCLEARES, UNAM



CIUDAD UNIVERSITARIA, Cd. Mx., 2024



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradezco a mis padres por siempre apoyarme y darme todas las herramientas necesarias para mejorar como estudiante y como persona.

Agradezco a mi hermano por compartir momentos valiosos juntos y por ser una gran fuente de inspiración y motivación en todos los aspectos.

Agradezco a mis amigos y compañeros por compartir esta etapa conmigo, llena de momentos inolvidables, y por siempre apoyarme cuando lo necesitaba.

Agradezco las becas otorgadas para realizar esta tesis a través del proyecto PAPIIT IA101123, “Estrellas de neutrones en relatividad general y en teorías alternativas de gravitación”.

Índice general

1. Introducción	7
1.1. Historia de la gravitación	7
1.2. Descubrimiento de Sgr A*	10
1.2.1. La naturaleza de Sgr A*	11
2. Variedades y curvatura	15
2.1. Variedades	15
2.1.1. Vectores y campos en la variedad	17
2.1.2. Formas lineales	20
2.1.3. Tensores	22
2.1.4. Tensor Métrico	24
2.2. Curvatura	27
2.2.1. Derivada covariante	27
2.2.2. Tensor de Riemann	30
2.2.3. Planitud local	33
2.2.4. Simetrías y propiedades del tensor de Riemann	34
2.2.5. Geodésicas	37
3. Relatividad General	41
3.1. Principios y construcción	41
3.1.1. Trayectorias a través del espacio-tiempo	43
3.2. Ecuaciones de campo	45

3.3. Solución de Schwarzschild	48
3.3.1. Cantidades conservadas	52
3.3.2. Trayectorias de partículas	53
3.4. Agujeros negros	59
3.4.1. Singularidades	59
3.4.2. Horizonte de eventos	62
4. Métodos MCMC	67
4.1. Estadística Bayesiana	67
4.2. Métodos de Monte Carlo	71
4.3. Cadenas de Markov	75
4.4. Métodos MCMC aplicados al ajuste de órbitas estelares	79
5. Implementación y resultados	83
5.1. Implementación del código	83
5.1.1. Integración y proyección de la órbita	83
5.1.2. Interpolación de la órbita	86
5.1.3. Implementación de las cadenas	89
5.2. Resultados	92
5.2.1. Distribuciones posteriores de los parámetros	92
5.2.2. Mejor modelo del sistema	93
6. Conclusiones	99

Capítulo 1

Introducción

La gravedad es, sin duda alguna, algo que está presente en nuestro día a día. Su estudio formal empezó en el S. XVII con el análisis de los datos astronómicos del movimiento de los planetas alrededor del Sol. A través de un análisis inductivo, Isaac Newton postuló una ley que describe esta fuerza [1]. Con el pasar de los años, han surgido nuevas formas de entenderla, y las técnicas implementadas para estudiarla han ido evolucionando. El gran avance de la tecnología nos permitió observar el movimiento de las estrellas en el centro galáctico. Los datos recabados de sus posiciones en el cielo abren las puertas a una forma de estudiar la gravedad en este régimen.

En este capítulo se habla del origen del estudio matemático de la gravitación en el S. XVII con Isaac Newton y su Ley Universal de la Gravedad. Además, se muestran las complicaciones que tuvo esta teoría para explicar la magnitud de precesión observada en la órbita de Mercurio alrededor del Sol. Más adelante veremos que a inicios del S. XX, Albert Einstein introdujo su teoría de la Relatividad General, la cual cambió drásticamente nuestro entendimiento de la gravedad; entre sus muchos éxitos está la obtención de la correcta magnitud de precesión de la órbita de Mercurio. Luego, se presenta la historia del descubrimiento de una de fuente intensa de ondas de radio en nuestro propio centro galáctico: Sagitario A* (Sgr A*). Finalmente, se muestran diversos métodos utilizados para estudiar su naturaleza; entre estos se habla del ajuste a modelos teóricos de órbitas estelares observadas alrededor de Sgr A*, siendo la más estudiada una estrella llamada S2.

1.1. Historia de la gravitación

El estudio matemáticamente formal de la gravedad como fuerza tiene sus orígenes en el S. XVII con los trabajos de Sir Isaac Newton [1]. Después de postular sus leyes de la mecánica, y gracias a los

trabajos previos de astrónomos como Johannes Kepler, Newton formuló una teoría capaz de explicar de manera matemática el movimiento de los planetas alrededor del Sol; ésta era una cuestión que había inquietado a grandes mentes como Robert Hooke, pero que no pudieron encontrar una forma de resolver [2]. De esta manera, Newton fue capaz de escribir su **Ley Universal de la Gravedad**, postulando la forma en la que se atraen diferentes cuerpos, como por ejemplo la Tierra y el Sol. Más aún, es *universal* porque explica tanto los movimientos celestes como los terrestres; esto es, la caída de una manzana en la Tierra tiene el mismo origen que la órbita de los planetas alrededor del Sol. Matemáticamente, la fuerza de atracción entre dos cuerpos se escribe como

$$\vec{F} = -\frac{Gm_1m_2}{r^2}\hat{r}, \quad (1.1)$$

donde m_1, m_2 representan las masas de los cuerpos, r y \hat{r} la distancia que los separa y el vector unitario que los une, respectivamente, y G es la **constante universal de la gravedad** y tiene un valor de $G = 6.673 \times 10^{-11} \text{m}^3\text{kg}^{-1}\text{s}^{-2}$ [3]. Un hecho importante a resaltar de la Ec. (1.1) es la dependencia de la atracción con las masas gravitacionales de los objetos. ¿Existe alguna relación entre las masas gravitacionales y las masas inerciales de estos? En el Cap. 3 exploraremos esta pregunta, la respuesta que se propuso y sus consecuencias.

Como se puede observar de la Ec. (1.1), la atracción entre cuerpos es proporcional al producto de sus masas pero inversamente proporcional al cuadrado de la distancia que los separa. Esta Ley nos dice entonces que las masas de los cuerpos dan origen a la gravedad y que cualesquiera cuerpos con masa se sienten atraídos entre sí, sin importar la distancia; en este sentido, se dice que la gravedad es una fuerza de alcance infinito.

Durante los siglos siguientes a Newton, su Ley Universal de la Gravedad fue la forma de entender esta interacción. De hecho, es esta misma ley la que se sigue utilizando para todo fin práctico al día de hoy, salvo en algunas situaciones muy particulares [4]. En términos generales, las leyes de Newton son suficientes para poder explicar el comportamiento de todos los cuerpos del Sistema Solar, incluso la precesión de las órbitas de *casi* todos los planetas. La precesión de una órbita quiere decir que el punto de máximo acercamiento, también llamado perihelio,¹ no es fijo en el tiempo y en cada ciclo orbital se mueve [3]; en este caso, avanza en el sentido de la órbita. En un principio, si se considera solamente la interacción de un planeta y el Sol, la órbita del primero alrededor del segundo no debería presentar precesión y el perihelio debería permanecer fijo en el espacio. No obstante, hay que recordar que esta situación no es realista, pues existen más cuerpos en el Sistema Solar que también ejercen una atracción con el planeta de acuerdo a la Ec. (1.1).

¹El sufijo *-helio* es porque orbita al Sol, pero en general el punto de máximo acercamiento se puede llamar *pericentro* o *periastro*.

Sin duda alguna, estas interacciones entre todos los planetas complican el problema ya que entonces cada planeta no solamente se siente atraído por el Sol, sino también por todos los demás cuerpos del Sistema Solar. Una manera de atacar esta situación es suponer que, dado un planeta, todos los demás son en realidad anillos concéntricos de densidad de masa constante y cuya masa total es igual a la de su respectivo planeta, y con radio igual al radio promedio de su órbita [3]. Este método permite promediar las perturbaciones gravitacionales que ejercerían los demás planetas sobre la órbita de un planeta dado, teniendo así una descripción mucho más detallada y realista.

Evidentemente, este modelo está dejando de lado cosas como las excentricidades de las órbitas o incluso las pequeñas inclinaciones de los planos orbitales de los planetas. Sin embargo, aun si se toman en cuenta, la precesión teórica de la órbita de Mercurio difiere de la observada, siendo éstas 5.32 arcseg/año y 5.74 arcseg/año, respectivamente. El hecho de que sea la precesión del planeta más cercano al Sol la que no encaja con lo observado es un fuerte indicio de que tal vez la teoría de Newton tiene sus limitaciones en ciertas situaciones, como se discutirá en el Cap. 3.

Después de publicar su **Teoría de la Relatividad Especial** (RE) [5], Einstein introdujo una descripción de la dinámica de los cuerpos que se mueven a velocidades relativas cercanas a la de la luz. Esta nueva descripción rompió con ideas que se tenían desde la época de Newton, siendo el mejor ejemplo que en la primera no existe un tiempo absoluto, dejando de lado el concepto de instantaneidad absoluta. Una década después, la teoría se generalizó, por lo que recibe el nombre de **Relatividad General** (RG) [4]. En esta nueva teoría se da una descripción totalmente diferente a la gravedad: ya no es una fuerza, sino la consecuencia de la deformación de un nuevo concepto llamado *espacio-tiempo*.

La RG dio un sinnúmero de predicciones como la desviación de la luz al pasar cerca de cuerpos masivos [4], la precesión natural de órbitas [4], o incluso nuevos objetos astrofísicos que desafían completamente nuestro entendimiento de la teoría: los agujeros negros [6]. Históricamente, la comprobación de la desviación de la luz en 1919 por Eddington fue la que catapultó a Einstein a la fama mundial por su nueva descripción del universo [7]. Además, aplicando su nuevo formalismo fue capaz de explicar con suma precisión el avance del perihelio de la órbita de Mercurio; obtuvo un valor de 5.75 arcseg/año, siendo un valor casi idéntico al observado [8]. De esta manera, la RG fue capaz de explicar la dinámica del Sistema Solar de manera excelente; por otro lado, los recientes avances en la observación astronómica han permitido poner a prueba sus límites en regímenes de campos gravitacionales más fuertes, como lo es el centro galáctico [9] [10].

Finalmente, una de las predicciones más sorprendentes de la RG es la existencia de agujeros negros, i.e. objetos tan masivos que no dejan que ni siquiera la luz escape de ellos. No obstante, es importante mencionar que esta idea es algo que ha existido desde la física newtoniana. En el S.

XVIII, el naturalista John Michell y el matemático Pierre Simon Laplace se dieron cuenta de que al aplicar las leyes de la dinámica de Newton junto con su teoría de la gravedad se podría tener un objeto tan masivo que su velocidad de escape sería mayor que la de la luz [11]; es claro que en ese entonces no se tenía evidencia de que la luz fuera afectada por la gravedad como sucede con otros objetos, pero de ser así, toda luz emitida en la superficie estaría atrapada gravitacionalmente. John Michell pensó este escenario para estrellas con una densidad de masa muy grande, por lo que acuñó el término **estrella oscura** [11]. Hoy en día sabemos que ésta no sería la explicación del origen de dichos objetos, aunque la existencia de los agujeros negros tiene cada vez más evidencia a favor [12] y podría existir uno con millones de veces la masa del Sol más cerca de lo que pensamos.

1.2. Descubrimiento de Sgr A*

Se había observado que existían fuentes de ondas de radio en los lóbulos de galaxias distantes, pero también se encontró una fuente importante de estas ondas en la constelación de Sagitario, hacia el centro de la Vía Láctea. Debido a que esta fuente era la más intensa, se denominó como Sagitario A (Sgr A). Al principio no se contaba con la resolución suficiente para poder analizar estas regiones de una mejor manera, pero tiempo después, y con el desarrollo de la radio interferometría, se observó que en realidad Sgr A tenía varios componentes [12]. Entre estas componentes están Sgr A-Este y Sgr A-Oeste. La primera de éstas corresponde a un remanente de una supernova, mientras que la segunda es una fuente de *bremstrahlung*² de gas ionizado. En 1974, se encontró una fuente de ondas de radio con un diámetro menor a 1 arcseg y fue nombrada Sgr A* [13]. La Fig. 1.1 muestra la región central de Sgr A, donde se puede observar un punto más intenso que corresponde a Sgr A*.

²El *bremstrahlung* es radiación electromagnética que emiten partículas cargadas debido a la desaceleración producida por un campo eléctrico de otra partícula.

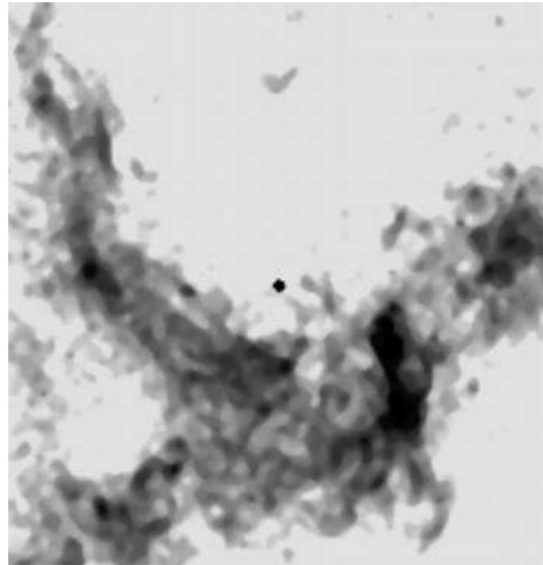


Figura 1.1: Vista de los 0.25 pc centrales del centro galáctico. El punto más oscuro en la imagen corresponde a la fuente Sgr A*. Imagen tomada de [12].

1.2.1. La naturaleza de Sgr A*

Por medio de observaciones de las velocidades de los cúmulos de gas, y suponiendo que estos están en órbitas circulares, se pudo concluir que deberían orbitar una masa central de varios millones de masas solares [14], además de estrellas a una distancia menor a 1 pc cuyas masas deberían sumar varios millones de masas solares. Estos resultados mostraron que, en efecto, hay una gran concentración de masa en nuestro centro galáctico. Sin embargo, una distribución de $\approx 10^7$ estrellas en el centro todavía no podía ser descartada como la respuesta a esta alta concentración de masa [12].

El estudio de la radiación emitida por gases tiene sus limitaciones, pues estos pueden interactuar de otras formas además de la gravitacional; por otro lado, la principal interacción entre las estrellas es gravitacional. En un principio, la presencia de un cúmulo de estrellas en el centro galáctico pasó desapercibida porque la luz visible emitida por éstas es absorbida por todo el gas en el centro galáctico y todo el que se encuentra entre la Tierra y Sgr A*. No obstante, estas estrellas pueden ser observadas en el espectro infrarrojo, pues las ondas en este rango no son totalmente absorbidas. De esta manera, la observación de estas estrellas dio lugar a otro método para estudiar la masa en el centro de la Vía Láctea. La Fig. 1.2 muestra las posiciones observadas de diferentes estrellas desde 1995 hasta 2003 [15].

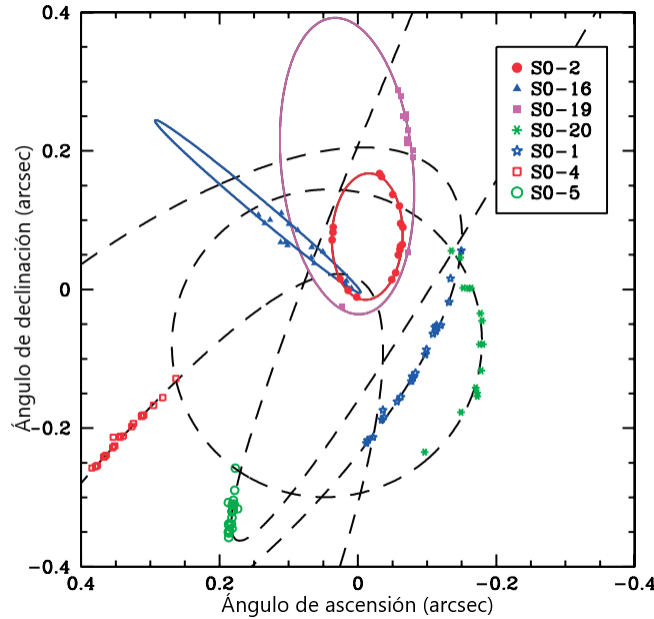


Figura 1.2: Posiciones astrométricas de siete estrellas en el centro galáctico tomadas entre 1995 y 2003. Figura tomada de [15].

De las observaciones de las posiciones estelares y los ajustes a sus órbitas se puede notar que las estrellas permanecen en órbitas elípticas casi perfectas, por lo que la mayoría de la masa, la cual no es directamente observable, debería estar contenida dentro de un radio de alrededor de 100 AU (≈ 0.0005 pc) [12]. Resulta que además los vectores de aceleración de las estrellas apuntan a un centro gravitacional común muy cerca de la posición de la fuente Sgr A*. Es importante notar que la posibilidad de que esta fuente sea en realidad un objeto lejano que se proyecte en el cielo de tal manera que se observe muy cerca del foco de las órbitas estelares es demasiado remota. Además, por medio del estudio de cuásares distantes es posible determinar que Sgr A* tiene una dinámica que coincide con la del centro galáctico; esto muestra que efectivamente esta fuente se encuentra en la Vía Láctea y que además está muy cerca del centro de ésta.

Una de las maneras más importantes de estudiar a Sgr A* y sus propiedades es por medio del estudio de las órbitas estelares a su alrededor. Entre todas las estrellas que se encuentran en el cúmulo del centro galáctico, una de las más estudiadas es la estrella llamada S2 (en la literatura también es conocida como S0-2). Para esta estrella en particular se ha estudiado su corrimiento al rojo gravitacional en 2019 [10] y la precesión de su órbita en 2020 [9]; ambos estudios han sido totalmente consistentes con la RG. En particular, su precesión se ha observado que es de $\delta\phi \approx 12$ arcmin por periodo orbital, lo cual es totalmente consistente con el espacio-tiempo de Schwarzschild

de la RG³. A través de los años se han observado numerosas posiciones de S2 a lo largo de su órbita alrededor de Sgr A*, lo cual permite ajustar modelos a estos datos con el fin de estudiar la naturaleza de Sgr A* como candidato a agujero negro súpermasivo (*SMBH*, por sus siglas en inglés). Además, es importante mencionar que estos estudios de las estrellas orbitando Sgr A* valieron para la mitad del Premio Nobel de 2020.

La Fig. 1.3 muestra 45 posiciones de S2 observadas entre 1995 y 2018, cubriendo más de un ciclo orbital [10]. Lo que hace especial a esta estrella es la cantidad de datos que se tienen, la cercanía de su pericentro con respecto a Sgr A* (≈ 120 AU), y su velocidad orbital máxima (≈ 7700 km/s) [9]; es por esto mismo que S2 es una excelente candidata para estudiar la validez de la RG en el régimen gravitacional presente en el centro galáctico.

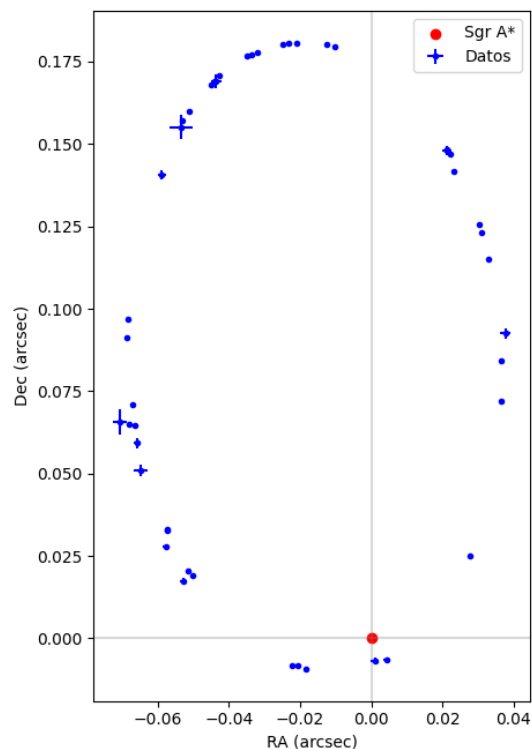


Figura 1.3: Posiciones de la estrella S2 observadas entre 1995 y 2018. Cada posición tiene su incertidumbre asociada y en el centro del sistema coordinado se marca a Sgr A*. Figura realizada con los datos proporcionados en [10].

³En la literatura es común encontrar que se utiliza este espacio-tiempo principalmente por sus simetrías y la resolución actual de los datos.

De esta manera, las órbitas estelares alrededor de Sgr A* nos brindan una herramienta para estudiar el comportamiento de la RG en un régimen no estudiado antes y explorar sus límites, tal como pasó con la gravitación de Newton el siglo anterior. Además, el ajuste de órbitas a los datos, que es el tema principal de este trabajo, nos abre las puertas a una nueva forma de poner a prueba diversas teorías o propuestas sobre la naturaleza de Sgr A*, haciendo que las órbitas estelares sean una nueva herramienta para estudiar la naturaleza del espacio-tiempo.

El objetivo principal de esta tesis es estudiar la naturaleza de Sgr A* por medio del análisis de la órbita de la estrella S2 en el marco de la Relatividad General utilizando métodos estocásticos; la meta es obtener las distribuciones de los valores de los parámetros que describen a este sistema. Para esto, los Capítulos 2 y 3 actúan como marco teórico, donde en el primero introducimos toda la herramienta matemática necesaria de geometría diferencial para así dedicar el segundo a la formulación de la Relatividad General como teoría física haciendo uso de los conceptos matemáticos previamente desarrollados. Además, en el Cap. 3 hablaremos en particular de la solución de Schwarzschild y de las ecuaciones que rigen el movimiento de una partícula en un espacio-tiempo con simetría esférica.

En el Capítulo 4 explicamos los métodos de Monte Carlo como herramienta para resolver problemas usando variables aleatorias y las Cadenas de Markov como procesos estocásticos markovianos. Luego, mostramos cómo la unión de estos da origen a los llamados métodos MCMC, los cuales son utilizados para obtener las distribuciones de valores de los parámetros de un modelo para ajustar a un conjunto de datos; en esta parte hacemos explícita la conexión entre estos métodos y el problema particular de la órbita de la estrella S2 alrededor de Sgr A*.

Una vez explicadas la parte teórica de la RG y la solución de Schwarzschild, y contando con los conocimientos de los métodos MCMC, la primera parte del Capítulo 5 sirve para describir la construcción y el funcionamiento del algoritmo de cómputo desarrollado para esta tesis. Después, la segunda parte del mismo está destinada a presentar los resultados obtenidos en este trabajo. Finalmente, en el Capítulo 6 escribimos las conclusiones obtenidas a partir de los resultados y describimos las ventajas y limitaciones de los métodos utilizados para obtenerlos.

Capítulo 2

Variedades y curvatura

En este capítulo se introducen las herramientas matemáticas necesarias para entender la teoría de la Relatividad General de manera semiformal; en particular, para una lectura más especializada en los temas de Geometría Diferencial y Variedades se recomiendan los libros de Wald [16] y Chandrasekhar [17]. A lo largo de este capítulo se dan temas como variedades, tensores, curvatura y geodésicas, haciendo énfasis en los cálculos y el sentido geométrico de algunas definiciones, como el transporte paralelo, por ejemplo; un libro que en particular sigue este enfoque es *Gravitation* [18] de donde se tomaron algunos ejemplos y descripciones.

2.1. Variedades

Definición 2.1.1 *Una variedad, denotada como \mathcal{M} , real, n -dimensional, y de clase \mathcal{C}^m , es un conjunto de puntos y una colección de subconjuntos $\{O_\alpha\}$ con $\alpha \in I$ un índice, tales que:*

1. *Cada punto $p \in \mathcal{M}$ pertenece al menos a un O_α ; dicho de otra forma, la colección $\{O_\alpha\}$ cubre totalmente a \mathcal{M} .*
2. *Para cada α existe un mapeo $\Psi_\alpha : O_\alpha \rightarrow U_\alpha$, con U_α un subconjunto abierto en \mathbb{R}^n que es uno-a-uno y suprayectivo.*
3. *Supongamos que $\exists \alpha, \beta$ tales que $O_\alpha \cap O_\beta \neq \emptyset$. Entonces se puede considerar el mapeo $\Psi_\beta \circ \Psi_\alpha^{-1}$ que toma puntos en $\Psi_\alpha[O_\alpha \cap O_\beta] \subset U_\alpha$ y los manda a $\Psi_\beta[O_\alpha \cap O_\beta] \subset U_\beta$. En este sentido, el mapeo $\Psi_\beta \circ \Psi_\alpha^{-1}$ es una función de \mathbb{R}^n a \mathbb{R}^n que pedimos sea de clase \mathcal{C}^m .*

Como se puede observar de la Definición 2.1.1, la variedad no es más que un conjunto de puntos que cumplen con ciertas características. Esto implica para que cumpla otras más particulares, se le tienen que definir las estructuras deseadas. Todas estas estructuras se le irán dando a lo largo de este capítulo.

Los mapeos Ψ_α , cuya acción se puede observar en la Fig. 2.1, son llamados **cartas** o **sistemas coordenados**, y el conjunto de todas las cartas es llamado **atlas**. De esta manera, se suele decir que el atlas cubre a toda la variedad.

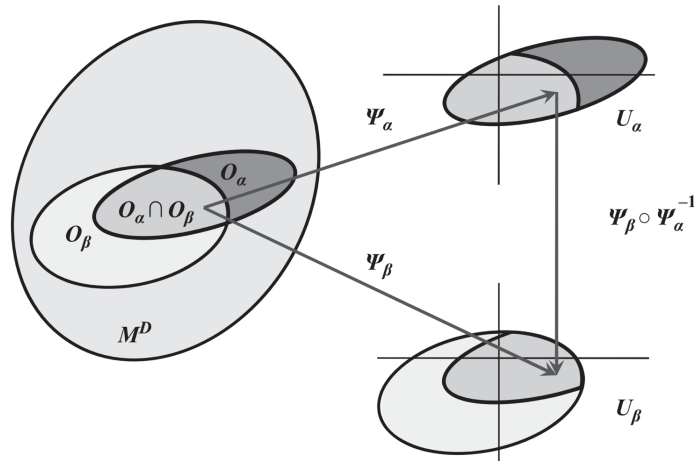


Figura 2.1: Acción de los mapeos dados en la Definición 2.1.1. Los conjuntos U_α, U_β son abiertos en \mathbb{R}^n , mientras que los conjuntos O_α, O_β son conjuntos de puntos de la variedad. Imagen tomada de [19].

Si x^μ , con $\mu = 1, 2, \dots, n$ son n números que describen puntos en U_α , y $x^{\mu'}$ son n números que describen puntos en U_β , entonces la transición de un sistema coordenado al otro está dado por funciones de la forma $x^{\mu'} = x^{\mu'}(x^\mu)$, recordando que se tiene un mapeo de clase C^m que los conecta. Cabe mencionar que las variedades relevantes en la RG son infinitamente diferenciables o de clase C^∞ .

Definición 2.1.2 Sean dos variedades \mathcal{M} y \mathcal{M}' de clase C^∞ . Un **difeomorfismo** es un mapeo $f : \mathcal{M} \rightarrow \mathcal{M}'$ que es uno-a-uno, suprayectivo y con inversa de clase C^∞ .

Si bien en la definición se entiende que se necesitan varios conjuntos $\{O_\alpha\}$ para cubrir a toda la variedad, en realidad puede suceder que una variedad pueda ser cubierta por una sola carta. Por ejemplo, considérese el conjunto de puntos en \mathbb{R}^n con carta \mathbb{R}^n . Se puede tomar como mapeo a la

identidad $\mathcal{I} : \mathbb{R}^n \rightarrow \mathbb{R}^n$. En este sentido, se satisfacen cada una de las condiciones de la Definición 2.1.1. Sin embargo, y como es de esperarse, son pocas las variedades que pueden ser cubiertas de esta manera.

Por ejemplo, sea la 2-esfera unitaria, \mathbb{S}^2 , la cual está definida por el conjunto de puntos

$$\mathbb{S}^2 := \{(x^1, x^2, x^3) \in \mathbb{R}^3 \mid (x^1)^2 + (x^2)^2 + (x^3)^2 = 1\} .$$

En este caso, se definen los abiertos [16]

$$O_i^\pm := \{(x^1, x^2, x^3) \in \mathbb{S}^2 \mid \pm x^i > 0\} ,$$

que corresponden a seis hemisferios en la esfera. Es claro que $\{O_i^\pm\}$ cubren a \mathbb{S}^2 . Ahora, los mapeos $\Psi : O_i^\pm \rightarrow U_i \in \mathbb{R}^2$ pueden ser construidos como las proyecciones a $D = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$, que corresponde al disco unitario bidimensional. Finalmente, la composición $\Psi_i^\pm \circ (\Psi_j^\pm)^{-1}$ es de clase \mathcal{C}^∞ , por lo que \mathbb{S}^2 es una variedad real 2-dimensional. Otros ejemplos de variedades 2-dimensionales son el toro y el cilindro, cada una con sus respectivas cartas definidas.

Sean dos variedades \mathcal{M} y \mathcal{M}' de dimensionalidades n y n' , respectivamente. Se puede construir el espacio de puntos $\mathcal{M} \times \mathcal{M}'$ de los pares $(p, p') \in \mathcal{M} \times \mathcal{M}'$, que corresponde al producto cartesiano entre las variedades, con $p \in \mathcal{M}$ y $p' \in \mathcal{M}'$ tal que se tiene una variedad de dimensión $n + n'$. Para esto, se definen los abiertos $O_{\alpha\beta} := O_\alpha \times O_\beta$ y $U_{\alpha\beta} := U_\alpha \times U_\beta$ con el mapeo $\Psi_{\alpha\beta}(p, p') = [\Psi_\alpha(p), \Psi_\beta(p')]$. De estas definiciones se tiene que $\Psi_{\alpha\beta} : O_{\alpha\beta} \subset \mathcal{M} \times \mathcal{M}' \rightarrow U_{\alpha\beta} \subset \mathbb{R}^{n+n'}$; este proceso permite construir variedades de dimensiones mayores a partir de otras variedades, o también para entender una variedad dada como el producto cartesiano de variedades distintas.

2.1.1. Vectores y campos en la variedad

Definición 2.1.3 Sea \mathcal{F} la colección de funciones \mathcal{C}^∞ de \mathcal{M} a \mathbb{R} . Se define el **vector tangente** en $p \in \mathcal{M}$ como un mapeo $\mathbf{v} : \mathcal{F} \rightarrow \mathbb{R}$ que es lineal y sigue la regla de Leibniz. Entonces, si $a, b \in \mathbb{R}$ y $f, g \in \mathcal{F}$, se satisface:

1. $\mathbf{v}(af + bg) = a\mathbf{v}(f) + b\mathbf{v}(g)$,
2. $\mathbf{v}(fg) = f(p)\mathbf{v}(g) + g(p)\mathbf{v}(f)$.

Si \mathbf{v}_p es un vector tangente al punto $p \in \mathcal{M}$, se sobrentiende que cualquier función $f \in \mathcal{F}$ se evalúa en p . De esta manera, un vector tangente a un punto de la variedad toma funciones y arroja

escalares. Al conjunto de vectores tangentes a un punto p se le denota $T_p\mathcal{M}$. Ahora, si se definen la suma de vectores y el producto por un escalar como:

- $(\mathbf{v}_1 + \mathbf{v}_2)f = \mathbf{v}_1(f) + \mathbf{v}_2(f)$,
- $\mathbf{v}(af) = a\mathbf{v}(f)$,

con $\mathbf{v}_1, \mathbf{v}_2 \in T_p\mathcal{M}$ y $a \in \mathbb{R}$, se tiene que $T_p\mathcal{M}$ es un espacio vectorial, el cual es comúnmente llamado **espacio tangente** en p . Debido a que el dominio de los vectores son las funciones \mathcal{C}^∞ , dada una $f \in \mathcal{F}$, dos vectores diferentes en $T_p\mathcal{M}$ pueden dar valores reales diferentes; esto es, en general puede suceder que $\mathbf{v}_1(f) \neq \mathbf{v}_2(f)$. Además, si \mathcal{M} tiene dimensión n , se satisface que $\dim(T_p\mathcal{M}) = n$ [16].

Sabiendo que se tiene una estructura de espacio vectorial, el siguiente paso es dar una base para sus elementos. Una base natural para estos es la llamada **base coordinada**, dada por las funciones $X_\mu : \mathcal{F} \rightarrow \mathbb{R}$ definidas como

$$X_\mu(f) := \left. \frac{\partial}{\partial x^\mu} (f \circ \Psi^{-1}) \right|_{\Psi(p)} , \quad (2.1)$$

con $f \in \mathcal{F}$. Es importante notar que se está evaluando el punto $\Psi(p) \in \mathbb{R}^n$ y que X_μ lo está mandando a \mathbb{R} . En general se suelen denotar estos elementos como $X_\mu \equiv \partial/\partial x^\mu \equiv \partial_\mu$. Debido a que es una base, los vectores tangentes en p pueden ser escritos como una combinación lineal

$$\mathbf{v} = \sum_{\mu} v^\mu X_\mu = v^\mu \partial_\mu , \quad (2.2)$$

donde en la segunda igualdad se omite la suma al utilizar la convención de suma de Einstein. En la expresión anterior se hace evidente que el vector no es un objeto por sí mismo debido a que la parcial debe actuar sobre una función. Otra cosa a resaltar es la dependencia de la expresión (2.1) de la carta correspondiente; es decir, si se tiene otra carta Ψ' , entonces $\{X_\mu\} \neq \{X'_\mu\}$ con las funciones X'_μ definidas con la carta primada; no obstante, se pueden relacionar ambas bases siguiendo

$$X_\mu = \sum_{\nu} \left. \frac{\partial x'^{\nu}}{\partial x^\mu} \right|_{\Psi(p)} X'_\nu , \quad (2.3)$$

con x'^{ν} la componente ν del mapeo $\Psi' \circ \Psi^{-1}$. Con base en lo anterior, se tiene que dadas dos cartas diferentes, Ψ y Ψ' , la relación entre las componentes de un vector \mathbf{v} es

$$v'^{\nu} = \sum_{\mu} v^\mu \frac{\partial x'^{\nu}}{\partial x^\mu} , \quad (2.4)$$

donde los coeficientes corresponden a $v^\mu = \mathbf{v}(x^\mu \circ \Psi)$. La Ec. (2.4) recibe el nombre de **regla de transformación de vectores**. Esto se sigue de que el vector \mathbf{v} es un objeto independiente de la base que se elija, por lo que $\mathbf{v} = \sum v^\mu X_\mu$, pero también $\mathbf{v} = \sum v'^\nu X'_\nu$, y utilizando la expresión para el cambio de los elementos de la base se obtiene la transformación de las componentes de \mathbf{v} de una carta a otra. Debido a que usamos cartas, las componentes del vector solamente van a estar definidas donde las coordenadas estén definidas. Más aún, la transformación de coordenadas requiere que ambos sistemas coordenados estén definidos en dicha región.

Sean dos puntos $p, q \in \mathcal{M}$. Debido a que los vectores tangentes se definen de manera puntual sobre la variedad, dos vectores en puntos diferentes viven en espacios tangentes diferentes. Además, si bien en \mathbb{R}^n se puede definir un vector que conecte dos puntos diferentes, esta estructura se pierde en la variedad, por lo que en general no existe un vector que conecte dos puntos sobre ésta. Sin embargo, el espacio tangente mantiene la noción de desplazamientos infinitesimales.

Sean p y q dos puntos en la variedad infinitesimalmente cercanos; es decir, dada una carta Ψ , sean (x^α) las coordenadas de p y $(x^\alpha + dx^\alpha)$ las coordenadas de q . Definimos el **vector de desplazamiento infinitesimal**, $d\mathbf{x} \in T_p\mathcal{M}$, como aquel cuya acción sobre una función escalar f es¹

$$d\mathbf{x}(f) = df|_{p \rightarrow q} = f(q) - f(p) . \quad (2.5)$$

La elección de notación para el vector de desplazamiento infinitesimal se sigue de la notación que se suele utilizar para cambios infinitesimales en un vector en el espacio plano y porque en la base coordenada las componentes de este vector van a ser los incrementos infinitesimales dx^α , i.e., $d\mathbf{x} = dx^\alpha \partial_\alpha$. Es importante notar que en \mathbb{R}^n se tiene la estructura de espacio vectorial, por lo que la noción de un vector que conecte puntos infinitesimalmente cercanos es natural. No obstante, en una variedad no se cuenta con dicha estructura, como ya se comentó con anterioridad, y es por lo mismo que en este caso un vector que “conecte” dos puntos cercanos debe darse por medio de una función evaluada en dichos puntos.

Una **curva suave** γ en una variedad \mathcal{M} es un mapeo suave, es decir, \mathcal{C}^∞ , de un intervalo $I \subset \mathbb{R}$ a la variedad [16]; entonces $\gamma : I \subset \mathbb{R} \rightarrow \mathcal{M}$. Dado un punto $p \in \mathcal{M}$ por el cual pasa γ se puede asociar un vector tangente $\boldsymbol{\xi} \in T_p\mathcal{M}$ con dicha curva. Una vez que se escoge una carta Ψ , la curva γ en \mathcal{M} se mapea a una curva $x = x^\mu(\lambda)$ en \mathbb{R}^n . Definimos el vector tangente a γ en p como aquel

¹Al ser p y q puntos infinitesimalmente cercanos, existe una única curva γ que los une. Otra manera de definir el vector de desplazamiento infinitesimal es $d\mathbf{x} := \mathbf{v}d\lambda$, con \mathbf{v} el vector tangente en p y $d\lambda$ el incremento infinitesimal del parámetro de γ . Para un desarrollo más profundo siguiendo este enfoque se recomienda [20].

con las componentes

$$\xi^\mu := \left. \frac{dx^\mu}{d\lambda} \right|_{\lambda_0}, \quad (2.6)$$

con λ_0 el valor en I que corresponde a p . Entonces la componente μ del vector tangente a γ no es más que la derivada de la componente μ de la función que mapea el intervalo a \mathbb{R}^n evaluada en λ_0 . Sea una función escalar $\varphi : \mathcal{M} \rightarrow \mathbb{R}$. Un vector tangente define el operador diferencial $\xi = \xi^\mu \partial_\mu$ que actúa sobre φ como $\xi^\mu \partial_\mu \varphi$.

Un **campo tangente**, \mathbf{v} , en \mathcal{M} es la asignación de un vector tangente $\mathbf{v}_p \in T_p\mathcal{M}$ para cada $p \in \mathcal{M}$. Decimos que el campo tangente es suave si para cada función suave $f \in \mathcal{F}$, $\mathbf{v}(f)$ también lo es. Esto implica que \mathbf{v} es suave si y sólo si las componentes v^μ lo son [16].

2.1.2. Formas lineales

Un tipo de operación que se puede realizar sobre los vectores definidos en la sección anterior es asignarlos a un número real. Lo interesante de esta operación es que se puede definir de tal manera que sea de forma lineal.

Definición 2.1.4 Sea $p \in \mathcal{M}$ y $T_p\mathcal{M}$ su espacio tangente. Definimos una **forma lineal** como el mapeo lineal

$$\begin{aligned} \omega : T_p\mathcal{M} &\rightarrow \mathbb{R} \\ \mathbf{v} &\mapsto \omega(\mathbf{v}), \end{aligned}$$

con $\mathbf{v} \in T_p\mathcal{M}$. Sean $\mathbf{v}, \mathbf{u} \in T_p\mathcal{M}$ y $\lambda \in \mathbb{R}$. Como ω es un mapeo lineal se satisface

$$\omega(\lambda\mathbf{v} + \mathbf{u}) = \lambda\omega(\mathbf{v}) + \omega(\mathbf{u}).$$

Es importante resaltar que la notación $\omega(\mathbf{v})$ hace explícito el hecho de que es la forma lineal la que está actuando sobre el vector \mathbf{v} ; sin embargo, otra notación común en la literatura es $\langle \cdot, \cdot \rangle$, por lo que una forma lineal actuando sobre un vector también suele escribirse como $\langle \omega, \mathbf{v} \rangle$ [17] [18]. A lo largo de este trabajo se usarán ambas notaciones según cuál sea más ilustrativa en cada caso.

Como una forma lineal, ω , actúa sobre $T_p\mathcal{M}$, se puede observar que entonces ésta también se define de manera local en la variedad. Además, dadas dos formas lineales, ω y π , y $\mathbf{v} \in T_p\mathcal{M}$, en general se tiene que $\langle \omega, \mathbf{v} \rangle \neq \langle \pi, \mathbf{v} \rangle$. Si se definen la suma de formas lineales y el producto por un

escalar, $\lambda \in \mathbb{R}$, como

$$\begin{aligned}\langle \boldsymbol{\omega} + \boldsymbol{\pi}, \mathbf{v} \rangle &= \langle \boldsymbol{\omega}, \mathbf{v} \rangle + \langle \boldsymbol{\pi}, \mathbf{v} \rangle, \\ \langle \lambda \boldsymbol{\omega}, \mathbf{v} \rangle &= \lambda \langle \boldsymbol{\omega}, \mathbf{v} \rangle,\end{aligned}$$

se tiene que el conjunto de formas lineales actuando sobre $T_p\mathcal{M}$ forma un espacio vectorial; este espacio es llamado **espacio dual** de $T_p\mathcal{M}$ y se denota por $T_p^*\mathcal{M}$.

Sea una carta Ψ que actúa sobre el conjunto de puntos $\{x^\mu\}$ y la base coordenada $(\boldsymbol{\partial}_\mu)$ de $T_p\mathcal{M}$ asociada a ésta. Entonces existe una única base, $(d\mathbf{x}^\alpha)$, que satisface

$$\langle d\mathbf{x}^\alpha, \boldsymbol{\partial}_\mu \rangle = \delta_\mu^\alpha, \quad (2.7)$$

donde δ_μ^α es la delta de Kronecker. La base $(d\mathbf{x}^\alpha)$ es llamada **base dual** de $(\boldsymbol{\partial}_\mu)$. Debido a que es una base, cualquier forma lineal $\boldsymbol{\omega}$ puede expandirse en términos de ésta, por lo que se tiene que $\boldsymbol{\omega} = \omega_\alpha d\mathbf{x}^\alpha$. Utilizando esto, se puede ver que la acción de la forma lineal sobre un vector \mathbf{v} es

$$\begin{aligned}\langle \boldsymbol{\omega}, \mathbf{v} \rangle &= \langle \omega_\alpha d\mathbf{x}^\alpha, v^\mu \boldsymbol{\partial}_\mu \rangle \\ &= \omega_\alpha v^\mu \langle d\mathbf{x}^\alpha, \boldsymbol{\partial}_\mu \rangle \\ &= \omega_\alpha v^\mu \delta_\mu^\alpha \\ &= \omega_\alpha v^\alpha.\end{aligned} \quad (2.8)$$

De la expresión anterior se puede ver que efectivamente $\langle \boldsymbol{\omega}, \mathbf{v} \rangle \in \mathbb{R}$. Sea ahora una base (\mathbf{e}_j) de $T_p\mathcal{M}$. Entonces $\langle \boldsymbol{\omega}, \mathbf{v} \rangle = \langle \boldsymbol{\omega}, v^i \mathbf{e}_i \rangle = v^i \langle \boldsymbol{\omega}, \mathbf{e}_i \rangle$. Definamos $\langle \boldsymbol{\omega}, \mathbf{e}_i \rangle := \omega_i$ como los números a los cuales son mapeados los elementos de la base del espacio tangente bajo la acción de $\boldsymbol{\omega}$. De esta manera, se tiene que $\langle \boldsymbol{\omega}, \mathbf{v} \rangle = \omega_i v^i$, en concordancia con la Ec. (2.8). Como base para $T_p^*\mathcal{M}$ podemos escoger las formas lineales $(\tilde{\mathbf{e}}^i)$ que mapean un vector a su i -ésima componente, es decir, $\langle \tilde{\mathbf{e}}^i, \mathbf{v} \rangle = v^i$, implicando que $\langle \tilde{\mathbf{e}}^i, \mathbf{e}_j \rangle = \delta_j^i$. Con base en esto,

$$\begin{aligned}\langle \boldsymbol{\omega}, \mathbf{v} \rangle &= \omega_i v^i \\ &= \omega_i \langle \tilde{\mathbf{e}}^i, v^j \mathbf{e}_j \rangle \\ &= \omega_i \langle \tilde{\mathbf{e}}^i, \mathbf{v} \rangle \\ &= \langle \omega_i \tilde{\mathbf{e}}^i, \mathbf{v} \rangle,\end{aligned} \quad (2.9)$$

de donde vemos que dada una base $(\tilde{\mathbf{e}}^i)$ para el espacio dual, cualquier forma lineal se puede expandir como $\boldsymbol{\omega} = \omega_i \tilde{\mathbf{e}}^i$, donde los elementos de la base son aquellos que arrojan la componente respectiva

de los vectores. Las igualdades en (2.8) son un caso particular de lo anterior donde la base del espacio tangente es precisamente la base coordenada. Además, las componentes de una forma lineal se transforman como

$$\omega'_\nu = \sum_\mu \omega_\mu \frac{\partial x^\mu}{\partial x'^\nu}. \quad (2.10)$$

Como ω es una combinación lineal de las \tilde{e}^i , que además satisfacen $\langle \tilde{e}^i, e_j \rangle = \delta_j^i$, se tiene que $\dim(T_p^* \mathcal{M}) = \dim(T_p \mathcal{M})$. La relación $e_i \leftrightarrow \tilde{e}^i$ da origen a un isomorfismo entre $T_p \mathcal{M}$ y $T_p^* \mathcal{M}$, por lo que ambos espacios tienen la misma estructura, pero no hay que olvidar que se tiene una dependencia en la elección de base. Al igual que un campo vectorial es la asignación de un vector a cada $p \in \mathcal{M}$, también se le puede asociar una forma lineal a cada punto en la variedad. Además, a las formas lineales se les suele denotar también por 1-formas y estos nombres se utilizan de manera indistinta.

En cuanto al espacio dual, podemos también hacer un mapeo de sus elementos a los reales igual que como se hizo con los vectores por medio de las formas lineales. Así, se tendría un espacio doble dual denotado $T_p^{**} \mathcal{M}$ cuyos elementos mapean formas lineales a los reales. Sea entonces $\Omega \in T_p^{**} \mathcal{M}$ y $\omega \in T_p^* \mathcal{M}$, por lo que $\Omega(\omega) \in \mathbb{R}$. Sin embargo, para cada Ω se puede asociar un $v \in T_p \mathcal{M}$ tal que los valores reales $\Omega(\omega), \omega(v)$ sean iguales. De esta manera se tiene un mapeo² de elementos del espacio doble dual al espacio tangente que es uno-a-uno y además suprayectivo porque $\dim(T_p^{**} \mathcal{M}) = \dim(T_p^* \mathcal{M})$. Por estas razones vemos que el espacio doble dual no es nada nuevo, sino que de hecho corresponde a $T_p \mathcal{M}$.

2.1.3. Tensores

Definición 2.1.5 Sea un punto $p \in \mathcal{M}$. Definimos el **tensor** de tipo (k, l) , con $k, l \in \mathbb{N}$, como el mapeo multilinear

$$\begin{aligned} \mathbf{T} : \underbrace{T_p^* \mathcal{M} \times \dots \times T_p^* \mathcal{M}}_{k \text{ veces}} \times \underbrace{T_p \mathcal{M} \times \dots \times T_p \mathcal{M}}_{l \text{ veces}} &\rightarrow \mathbb{R} \\ (\omega_1, \dots, \omega_k, v_1, \dots, v_l) &\mapsto T(\omega_1, \dots, \omega_k, v_1, \dots, v_l) \end{aligned}$$

con $\omega_i \in T_p^* \mathcal{M}$ y $v_j \in T_p \mathcal{M}$.

²La definición y propiedades de este mapeo van más allá de los objetivos de este trabajo. Sin embargo, se recomiendan [16, 17] para un desarrollo más profundo al respecto.

También se dice que el tensor de tipo (k, l) es k -veces contravariante y l -veces covariante, y que su orden es $k + l \in \mathbb{N}$. Además, que sea multilineal significa que

$$T(\omega_1, \dots, \omega_k, \alpha v + \beta u, v_2, \dots, v_l) = \alpha T(\omega_1, \dots, \omega_k, v, v_2, \dots, v_l) + \beta T(\omega_1, \dots, \omega_k, u, v_2, \dots, v_l)$$

para cada una de sus entradas y con $\alpha, \beta \in \mathbb{R}$. Denotamos a la colección de tensores de orden (k, l) como $\mathcal{T}(k, l)$. Definiendo el producto por un escalar y la suma de tensores de orden (k, l) de manera usual, esta colección es un espacio vectorial cuya dimensión es n^{k+l} .

Con base en la Definición 2.1.5, podemos ver que $\omega : T_p\mathcal{M} \rightarrow \mathbb{R}$ es un tensor de tipo $(0, 1)$ y $v : T_p^*\mathcal{M} \rightarrow \mathbb{R}$ es un tensor de tipo $(1, 0)$; esto explica la primera notación dada para las formas lineales actuando sobre los vectores. Sea un tensor de tipo $(1, 1)$, $\mathbf{T} : T_p^*\mathcal{M} \times T_p\mathcal{M} \rightarrow \mathbb{R}$. Si fijamos $v \in T_p\mathcal{M}$, tenemos que $\mathbf{T}(\cdot, v)$ es un mapeo de $T_p^*\mathcal{M}$ a los reales, por lo que es un vector. Así, tenemos que un tensor de tipo $(1, 1)$ puede verse como un mapeo lineal de $T_p\mathcal{M}$ en sí mismo, y de manera similar para $\omega \in T_p^*\mathcal{M}$ fija.

Sean un tensor, \mathbf{T} , de tipo (k, l) , y otro tensor \mathbf{S} , de tipo (r, m) . Se puede construir un tensor de tipo $(k + r, l + m)$ mediante el producto exterior de estos, denotado por $\mathbf{T} \otimes \mathbf{S}$, que actúa sobre $k + r$ covectores y $l + m$ vectores como

$$\mathbf{T} \otimes \mathbf{S}(\omega^1, \dots, \omega^{k+r}, v_1, \dots, v_{l+m}) = \left\{ \mathbf{T}(\omega^1, \dots, \omega^k, v_1, \dots, v_l) \right\} \left\{ \mathbf{S}(\omega^{k+1}, \dots, \omega^{k+r}, v_{l+1}, \dots, v_{l+m}) \right\} .$$

En la expresión anterior tenemos que lo que está dentro de los corchetes es un número real, por lo que se tiene la multiplicación usual entre números reales. De lo anterior se puede observar que se pueden construir tensores de orden arbitrario mediante los productos tensoriales necesarios de vectores y covectores.

Dada una base (e_i) de $T_p\mathcal{M}$ y una base (\tilde{e}^j) de $T_p^*\mathcal{M}$, cualquier tensor de orden $(k + l)$ se puede expandir como

$$\mathbf{T} = T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_l} e_{\alpha_1} \otimes \dots \otimes e_{\alpha_k} \otimes \tilde{e}^{\beta_1} \otimes \dots \otimes \tilde{e}^{\beta_l} \quad (2.11)$$

donde $(e_{\alpha_1} \otimes \dots \otimes e_{\alpha_k} \otimes \tilde{e}^{\beta_1} \otimes \dots \otimes \tilde{e}^{\beta_l})$ actúa sobre $(\omega_1, \dots, \omega_k, v_1, \dots, v_l)$ como

$$\begin{aligned} (e_{\alpha_1} \otimes \dots \otimes e_{\alpha_k} \otimes \tilde{e}^{\beta_1} \otimes \dots \otimes \tilde{e}^{\beta_l})(\omega_1, \dots, \omega_k, v_1, \dots, v_l) &= \langle \omega^1, e_{\alpha_1} \rangle \dots \langle \omega^k, e_{\alpha_k} \rangle \langle \tilde{e}^{\beta_1}, v_1 \rangle \dots \langle \tilde{e}^{\beta_l}, v_l \rangle \\ &= \omega_{\alpha_1}^1 \omega_{\alpha_2}^2 \dots \omega_{\alpha_k}^k v_1^{\beta_1} v_2^{\beta_2} \dots v_l^{\beta_l} . \end{aligned}$$

Los n^{k+l} coeficientes escalares $T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_l}$ son llamados los componentes del tensor \mathbf{T} con respecto a la base $(e_i \otimes \tilde{e}^j)$; estos son únicos y caracterizan de manera completa al tensor. Por ejemplo, sea

un tensor de tipo $(0, 2)$, $\mathbf{g} : T_p\mathcal{M} \times T_p\mathcal{M} \rightarrow \mathbb{R}$, en una variedad de dimensión $n = 4$. En este caso, \mathbf{g} tiene 16 componentes $g_{\alpha\beta}$.

En general, las componentes de un tensor de tipo (k, l) , \mathbf{T} , se transforman de un sistema coordenado a otro mediante la **regla de transformación de tensores**, la cual establece que

$$T'^{\mu'_1 \dots \mu'_k}_{\nu'_1 \dots \nu'_l} = \sum_{\mu_1 \dots \mu_l=1}^n T^{\mu_1 \dots \mu_k}_{\nu_1 \dots \nu_l} \frac{\partial x'^{\mu_1}}{\partial x_{\mu_1}} \cdots \frac{\partial x'^{\nu_l}}{\partial x_{\nu_l}}. \quad (2.12)$$

De la expresión anterior se puede observar que si las componentes de un tensor se anulan en un sistema coordenado, entonces en cualquier otro también lo harán. Además, esta expresión es consistente con las Ecs. (2.4) y (2.10), pues los vectores y covectores son casos particulares de tensores.

Definición 2.1.6 *Un campo tensorial de tipo (k, l) es la asignación de un tensor $\mathbf{T} \in \mathcal{T}(k, l)$ para cada $p \in \mathcal{M}$.*

Por convención, un campo escalar en \mathcal{M} es un campo tensorial de tipo $(0, 0)$.

2.1.4. Tensor Métrico

Definición 2.1.7 *Un tensor métrico, \mathbf{g} , en \mathcal{M} es un campo tensorial que satisface:*

1. \mathbf{g} es un campo tensorial de tipo $(0, 2)$. Entonces, dado $p \in \mathcal{M}$, \mathbf{g} es un mapeo $\mathbf{g} : T_p\mathcal{M} \times T_p\mathcal{M} \rightarrow \mathbb{R}$.
2. \mathbf{g} es simétrico, i.e. $\mathbf{g}(\mathbf{v}, \mathbf{u}) = \mathbf{g}(\mathbf{u}, \mathbf{v})$ para cualesquiera $\mathbf{v}, \mathbf{u} \in T_p\mathcal{M}$.
3. \mathbf{g} es no-degenerado: si $\mathbf{g}(\mathbf{u}, \mathbf{v}) = 0$ para cualquier $\mathbf{u} \in T_p\mathcal{M}$, entonces \mathbf{v} es el vector cero.

Recordemos que, dada la estructura que tenemos hasta ahora en \mathcal{M} , se pierde la noción de desplazamiento desde un punto p a un punto arbitrario q . Sin embargo, el espacio tangente preserva la noción de desplazamientos infinitesimales alrededor del punto p por medio del vector $d\mathbf{x}$. De manera intuitiva, la métrica nos da el cuadrado del desplazamiento infinitesimal asociado al vector $d\mathbf{x}$, lo cual es claro más adelante al ver cómo actúa \mathbf{g} sobre $d\mathbf{x}$. Así, la métrica se puede pensar como un producto interior de elementos del espacio tangente a cada punto de la variedad, aunque no necesariamente es positivo definido, y en los casos que se analizan en este trabajo de hecho no lo es.

Decimos que dos vectores, \mathbf{u}, \mathbf{v} , son ortogonales si y sólo si $\mathbf{g}(\mathbf{v}, \mathbf{u}) = 0$. Además, si no hay ambigüedad en la métrica que se está utilizando, el producto escalar entre dos elementos se suele denotar como

$$\mathbf{v} \cdot \mathbf{u} := \mathbf{g}(\mathbf{v}, \mathbf{u}) . \quad (2.13)$$

Dada una base $(\tilde{\mathbf{e}}^i)$ de $T_p^*\mathcal{M}$, las componentes de \mathbf{g} representan los coeficientes en la expansión

$$\mathbf{g} = g_{\alpha\beta} \tilde{\mathbf{e}}^\alpha \otimes \tilde{\mathbf{e}}^\beta . \quad (2.14)$$

Con base en la expresión (2.14) para \mathbf{g} , se puede ver que su acción sobre dos vectores \mathbf{v}, \mathbf{u} está dada por

$$\begin{aligned} \mathbf{g}(\mathbf{v}, \mathbf{u}) &= g_{\alpha\beta} \langle \tilde{\mathbf{e}}^\alpha, \mathbf{v} \rangle \langle \tilde{\mathbf{e}}^\beta, \mathbf{u} \rangle \\ &= g_{\alpha\beta} v^\alpha u^\beta . \end{aligned} \quad (2.15)$$

En las expresiones anteriores se observa que para obtener la expansión de \mathbf{g} en términos de los elementos de una base, o para saber su acción sobre vectores, se necesitan conocer las componentes $g_{\alpha\beta}$. Estas componentes están definidas como $g_{\alpha\beta} = \mathbf{g}(\mathbf{e}_\alpha, \mathbf{e}_\beta)$. Si se trabaja con la base coordinada (∂_μ) asociada al sistema coordinado (x^μ) , entonces la acción de la métrica sobre $(d\mathbf{x}, d\mathbf{x})$ da el cuadrado del desplazamiento infinitesimal, esto es,

$$\begin{aligned} ds^2 &:= \mathbf{g}(d\mathbf{x}, d\mathbf{x}) \\ &= g_{\alpha\beta} \langle \tilde{\mathbf{e}}^\alpha, dx^\mu \partial_\mu \rangle \langle \tilde{\mathbf{e}}^\beta, dx^\nu \partial_\nu \rangle \\ &= g_{\alpha\beta} dx^\mu dx^\nu \langle \tilde{\mathbf{e}}^\alpha, \partial_\mu \rangle \langle \tilde{\mathbf{e}}^\beta, \partial_\nu \rangle \\ &= g_{\alpha\beta} dx^\mu dx^\nu \delta_\mu^\alpha \delta_\nu^\beta \\ &= g_{\alpha\beta} dx^\alpha dx^\beta . \end{aligned} \quad (2.16)$$

Debido a que \mathbf{g} no es positivo definido, ds^2 puede tomar valores negativos a pesar de la notación utilizada. La expresión (2.16) es llamada **elemento de línea**. Para la base dual $(d\mathbf{x}^\alpha)$ se tiene que $\mathbf{g} = g_{\alpha\beta} d\mathbf{x}^\alpha \otimes d\mathbf{x}^\beta$.

Definimos la **signatura** de la métrica como la diferencia del número de componentes positivas y el número de componentes negativas de $(g_{\alpha\beta})$ una vez que la matriz está en su forma diagonal; suele referirse a ella por el valor correspondiente, o bien, especificando el signo de las entradas en su forma diagonal. Es importante resaltar que sin importar la base de $T_p\mathcal{M}$ en la cual se lleve la

matriz a su forma diagonal, la signatura siempre va a ser la misma [20]; esto es, no hay ambigüedad en la signatura de la métrica aunque no se especifique la base que se está utilizando.

Sea el caso de una variedad en 4 dimensiones. Cuando la signatura es $++++$ se tiene una métrica Riemanniana; por otro lado, cuando es $-+++$ se dice que la métrica es **Lorentziana** [16]. En el último caso, al par $(\mathcal{M}, \mathbf{g})$ se le denomina **variedad Lorentziana**. Uno puede tener que $\mathbf{g}(\mathbf{v}, \mathbf{v}) < 0$, a lo cual decimos que \mathbf{v} es temporaloide o tipo tiempo; si $\mathbf{g}(\mathbf{v}, \mathbf{v}) = 0$, entonces \mathbf{v} es un vector nulo; y si $\mathbf{g}(\mathbf{v}, \mathbf{v}) > 0$, entonces \mathbf{v} es espacialoide o tipo espacio. Estos nombres surgen de la signatura de la métrica y de qué parte (espacial o temporal) está "dominando"³.

Como \mathbf{g} es un tensor de orden $(0, 2)$, actúa sobre $T_p\mathcal{M} \times T_p\mathcal{M}$. No obstante, si se deja fijo $\mathbf{v} \in T_p\mathcal{M}$, se tiene un mapeo lineal $\mathbf{g}(\cdot, \mathbf{v}) : T_p\mathcal{M} \rightarrow \mathbb{R}$; esto es, se tiene un mapeo que actúa como covector. Además, al ser \mathbf{g} no-degenerado, este mapeo es inyectivo y suprayectivo; en particular, el inverso existe. Así, el tensor métrico permite crear una correspondencia uno-a-uno entre vectores y covectores.

El hecho de que el tensor métrico sea no degenerado es equivalente a pedir que la matriz $(g_{\alpha\beta})$ sea invertible y que la inversa sea única. Los elementos de la inversa son tales que se satisface $g_{\alpha\beta}g^{\beta\mu} = \delta_{\mu}^{\alpha}$. Esto garantiza que las componentes $g^{\alpha\beta}$ son las componentes del tensor de tipo $(2, 0)$ \mathbf{g}^{-1} dado por $\mathbf{g}^{-1} = g^{\alpha\beta}\mathbf{e}_{\alpha} \otimes \mathbf{e}_{\beta}$; en la base coordenada se tiene entonces que $\mathbf{g}^{-1} = g^{\alpha\beta}\partial_{\alpha} \otimes \partial_{\beta}$ [17].

Sabemos que la métrica permite establecer una relación entre vectores y covectores por medio del mapeo $\mathbf{g}(\cdot, \mathbf{v}) : T_p\mathcal{M} \rightarrow \mathbb{R}$. De esta forma, para un $\mathbf{u} \in T_p\mathcal{M}$ existe una única 1-forma, que denotaremos como \underline{u} , tal que para todo $\mathbf{v} \in T_p\mathcal{M}$ se satisface $\langle \underline{u}, \mathbf{v} \rangle = \mathbf{g}(\mathbf{u}, \mathbf{v})$. Las componentes de esta 1-forma están dadas por

$$u_{\alpha} = g_{\alpha\beta}u^{\beta} . \quad (2.17)$$

Por otro lado, el mapeo también permite el otro sentido. Dada una $\omega \in T_p^*\mathcal{M}$, existe un único vector, el cual denotaremos por $\vec{\omega}$, tal que para todo $\mathbf{v} \in T_p\mathcal{M}$ se satisface $\langle \omega, \mathbf{v} \rangle = \mathbf{g}(\vec{\omega}, \mathbf{v})$, y sus componentes están dadas como

$$\omega^{\alpha} = g^{\alpha\beta}\omega_{\beta} . \quad (2.18)$$

Las Ecs. (2.17) y (2.18) se conocen comúnmente como “bajar” y “subir” índices, respectivamente; entonces decimos que la métrica permite subir y bajar índices de tensores, aunque lo que está pasando es que la métrica y su inversa permiten ir del espacio de vectores al de covectores, y viceversa. Sin embargo, el uso de la misma letra para las componentes hace explícito el isomorfismo inducido por el tensor métrico entre $T_p\mathcal{M}$ y $T_p^*\mathcal{M}$.

³Esto se aclarará cuando analicemos el caso de la métrica en la Relatividad General.

2.2. Curvatura

2.2.1. Derivada covariante

Recordemos que definimos el vector de desplazamiento infinitesimal como $d\mathbf{x} := f(q) - f(p)$ para dos puntos p y q infinitesimalmente cercanos. Sin embargo, no se puede definir un vector que dé el cambio de un campo vectorial entre dos puntos p y q , $d\mathbf{v} := \mathbf{v}(q) - \mathbf{v}(p)$, porque ni siquiera estaría bien definido de esta manera, pues $\mathbf{v}(q)$ y $\mathbf{v}(p)$ son elementos de espacios vectoriales distintos. No se tiene este problema con $d\mathbf{x}$ porque f no depende de vectores base definidos en cada punto de la variedad como lo hacen los vectores.

Aun así, se puede intentar construir un operador de derivada que mantenga la noción de la variación de un campo vectorial. Sea Ψ un sistema coordenado con (∂_α) la base de $T_p\mathcal{M}$. En la región cubierta por esta carta se puede definir un operador de derivada ∂_μ [16], llamado **derivada ordinaria**, que actúa sobre un tensor $T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_l}$ como $\partial_c T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_l}$, produciendo un tensor cuyas componentes son las derivadas parciales de las funciones escalares $T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_l}$, i.e. $\partial T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_l} / \partial x^\sigma$.

No obstante, dada otra carta Ψ' se tendría otra base (∂'_α) , por lo que se tendría otro operador ∂'_α . Esto hace que las componentes de la derivada ordinaria en esta base no coincidan con las componentes en la base de la carta Ψ . De esta manera, el operador de derivada ordinaria es dependiente de las coordenadas y no está asociado de manera natural a la estructura propia de la variedad.

Para arreglar esto, se tiene que agregar una estructura más a la variedad \mathcal{M} que permita definir la variación de campos vectoriales conectando los diferentes espacios tangentes. Sea $\mathfrak{X}(\mathcal{M})$ el conjunto de todos los campos vectoriales suaves en \mathcal{M} . Este conjunto constituye un espacio vectorial infinito-dimensional en \mathcal{M} sobre \mathbb{R} [20].

Definición 2.2.1 Definimos la **conexión afín** en \mathcal{M} como el mapeo

$$\begin{aligned} \nabla : \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) &\rightarrow \mathfrak{X}(\mathcal{M}) \\ (\mathbf{u}, \mathbf{v}) &\mapsto \nabla_{\mathbf{u}} \mathbf{v} \end{aligned}$$

donde se tiene que $\nabla_{\mathbf{u}} \mathbf{v}$ es un campo vectorial suave, el cual se denomina la **derivada covariante de \mathbf{v}** a lo largo de \mathbf{u} . Además, la conexión afín satisface:

1. El mapeo ∇ es bilineal;
2. Sea f un campo escalar en \mathcal{M} . Entonces se tiene

$$\nabla_{f\mathbf{u}} \mathbf{v} = f \nabla_{\mathbf{u}} \mathbf{v} ;$$

3. ∇ cumple con la regla de Leibniz:

$$\nabla_{\mathbf{u}}(f\mathbf{v}) = \langle \nabla f, \mathbf{u} \rangle \mathbf{v} + f \nabla_{\mathbf{u}} \mathbf{v} .$$

Notemos que la segunda propiedad no se desprende de la primera porque f no es un real, sino una función escalar en \mathcal{M} . Ahora, como $\nabla_{\mathbf{u}} \mathbf{v}$ es un campo vectorial, dado un $p \in \mathcal{M}$ también está definido en los puntos alrededor de éste. Sin embargo, la segunda propiedad garantiza que $\nabla_{\mathbf{u}} \mathbf{v}$ dependa solamente de $\mathbf{u}(p)$ y no del comportamiento de \mathbf{u} alrededor de p [20].

Dada la notación $\nabla_{\mathbf{u}} \mathbf{v}$ se entiende que se está derivando \mathbf{v} a lo largo de la dirección de \mathbf{u} . De esta manera, podemos definir la variación de un campo vectorial \mathbf{v} entre dos puntos cercanos como

$$d\mathbf{v} := \nabla_{d\mathbf{x}} \mathbf{v} , \quad (2.19)$$

donde $d\mathbf{x}$ es el vector de desplazamiento infinitesimal que conecta ambos puntos. Entendiendo la derivada covariante como el cambio del campo en una cierta dirección, la expresión (2.19) permite hacer la conexión de la derivada usual como cambio infinitesimal con el cambio del campo de manera local en la variedad. Si \mathbf{v} no cambia mientras se desplaza a lo largo del vector $d\mathbf{x}$, entonces $d\mathbf{v} = 0$ y se dice que \mathbf{v} es transportado paralelamente con respecto a la conexión ∇ .

El sentido geométrico del transporte paralelo es conectar los espacios tangentes de puntos infinitesimalmente cercanos con el objetivo de tener una forma clara de conectar y comparar vectores de espacios tangentes arbitrariamente alejados. Sea $d\mathbf{x}$ el vector de desplazamiento infinitesimal entre los puntos p y q , y sea (\mathbf{e}_j) la base de $T_p \mathcal{M}$ tal que el vector \mathbf{v} se ve como $\mathbf{v} = v^i \mathbf{e}_i$. Ahora, estos vectores base se transportan paralelamente como $\nabla_{d\mathbf{x}} \mathbf{e}_j = 0$, resultando en los vectores $(\tilde{\mathbf{e}}_j)$ que ahora son la base de $T_q \mathcal{M}$. De esta manera, el vector \mathbf{v} en q toma la forma $\mathbf{v} = \tilde{v}^i \tilde{\mathbf{e}}_i$. No obstante, como p y q están infinitesimalmente cerca, se tiene que $v^i = \tilde{v}^i$, es decir, los valores numéricos de las componentes del vector son los mismos en ambos espacios tangentes después de transportar paralelamente los vectores base de $T_p \mathcal{M}$ [18].

Sean A y B los puntos inicial y final de una curva \mathcal{L} . Si $\mathbf{v}_A := \mathbf{v}(A)$ y $\mathbf{v}_B := \mathbf{v}(B)$ son los vectores inicial y final, entonces puede suceder que al transportar paralelamente el vector a lo largo de la curva al final se tenga $\mathbf{v}(A) \neq \mathbf{v}(B)$. Esto se puede deber precisamente a la curvatura de la variedad, ya que si fuera plana se tendría paralelismo absoluto sobre ella. La condición $\nabla_{\mathbf{u}} \mathbf{v} = 0$ no implica que al final se tenga el mismo vector \mathbf{v} , pues esto se da particularmente en los puntos infinitesimalmente cercanos, como ya se comentó. Un ejemplo de esto es transportar \mathbf{v} a lo largo de un paralelo sobre la esfera. Al final de dar una vuelta ($\phi = 0 \rightarrow \phi = 2\pi$), se tiene que \mathbf{v} ha sido rotado [21].

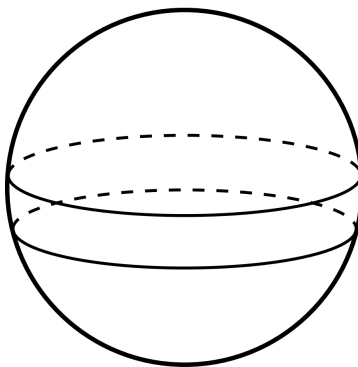


Figura 2.2: Paralelos sobre la esfera. Imagen tomada de [21].

Definimos un **campo coordinado** o **campo de cartas** sobre la variedad \mathcal{M} de dimensión n como la n -tupla de campos vectoriales en \mathcal{M} , (e_α) , tales que $(e_\alpha(p))$ es una base para $T_p\mathcal{M}$. Al ser cada e_α un campo vectorial, se le puede aplicar la derivada covariante con respecto a otro campo vectorial; en particular, se puede obtener su derivada covariante con respecto a otro campo e_β . Así, se tiene que

$$\nabla_{e_\beta} e_\alpha := \Gamma_{\alpha\beta}^\mu e_\mu, \quad (2.20)$$

con $\Gamma_{\alpha\beta}^\mu$ los coeficientes de la expansión de $\nabla_{e_\beta} e_\alpha$ en la base $(e_\alpha(p))$, llamados **coeficientes de la conexión**. Una conexión afín está definida totalmente por sus coeficientes de la conexión. Como los vectores base de los espacios tangentes están definidos de manera puntual, tiene sentido pensar que no son los mismos en diferentes puntos y que varían a través de la variedad. Un ejemplo particular de lo anterior son los vectores base de \mathbb{R}^2 en coordenadas polares [22]. Así, los coeficientes de la conexión cuantifican los cambios que sufren los vectores base al ser transportados de un punto a otro con respecto a una determinada dirección en la base de los campos vectoriales mismos.

En una variedad pseudo-Riemanniana (\mathcal{M}, g) existe una única conexión afín, ∇ , tal que

1. ∇ no tiene torsión, i.e. si f es un campo escalar en \mathcal{M} , entonces

$$\nabla_a \nabla_b f = \nabla_b \nabla_a f; \quad (2.21)$$

2. La derivada covariante del tensor métrico es cero, i.e.

$$\nabla g = 0. \quad (2.22)$$

La conexión afín que satisface lo anterior se denomina **conexión de Levi-Civita**. En una base coordenada (∂_α) se tiene que $\nabla_{\partial_\alpha} f = \partial f / \partial x^\alpha$ para una función escalar. El que la conexión no tenga torsión quiere decir que las derivadas covariantes conmutan al actuar en funciones. La prueba de la unicidad del operador ∇ que satisface la segunda propiedad se encuentra en [16].

Dada una carta (x^α) en \mathcal{M} , los coeficientes de la conexión de Levi-Civita asociados con dichas coordenadas reciben el nombre de **símbolos de Christoffel** y pueden ser calculados en términos de derivadas parciales de la métrica como

$$\Gamma^\mu_{\alpha\beta} = \frac{1}{2} g^{\mu\nu} \left(\frac{g_{\nu\beta}}{\partial x^\alpha} + \frac{g_{\nu\alpha}}{\partial x^\beta} - \frac{g_{\alpha\beta}}{\partial x^\nu} \right). \quad (2.23)$$

Como se puede ver en la expresión (2.23) y de la simetría de la métrica, los símbolos de Christoffel son simétricos con respecto a los índices inferiores, es decir, $\Gamma^\mu_{\alpha\beta} = \Gamma^\mu_{\beta\alpha}$.

Dado un campo tensorial \mathbf{T} de tipo (k, l) se puede definir su **divergencia** con respecto a la conexión ∇ , denotada como $\nabla \cdot \mathbf{T}$, como un campo tensorial con componentes [20]

$$(\nabla \cdot \mathbf{T})^{\alpha_1 \dots \alpha_{k-1}}_{\beta_1 \dots \beta_l} = \nabla_\mu T^{\alpha_1 \dots \alpha_{k-1} \mu}_{\beta_1 \dots \beta_l}. \quad (2.24)$$

En particular, la divergencia de un campo vectorial en coordenadas (x^α) tiene componentes

$$\nabla \cdot \mathbf{v} = \nabla_\mu v^\mu = \frac{\partial v^\mu}{\partial x^\mu} + \Gamma^\mu_{\sigma\mu} v^\sigma, \quad (2.25)$$

donde se puede apreciar que se debe tomar en cuenta la variación de los vectores base, pues ésta puede ser no trivial; de hecho, en general no lo es. En el caso del espacio plano y coordenadas cartesianas se tiene simplemente que $\nabla \cdot \mathbf{v} = \partial_\mu v^\mu$, pues como se tiene un conjunto de vectores constantes, estos no cambian al moverse de un punto de \mathbb{R}^n a otro. En cuanto a la derivada covariante de \mathbf{v} con respecto a la dirección ν , ésta se puede expresar como

$$\nabla_{\mathbf{e}_\nu} \mathbf{v} = \left(\frac{\partial v^\mu}{\partial x^\nu} + \Gamma^\mu_{\alpha\nu} v^\alpha \right) \mathbf{e}_\mu. \quad (2.26)$$

2.2.2. Tensor de Riemann

Sea un circuito cerrado $ABCD$ como se muestra en la Figura 2.3 donde las líneas del circuito están dadas por valores constantes de las coordenadas x^μ y x^ν , con δa^μ y δa^ν infinitesimalmente pequeños.

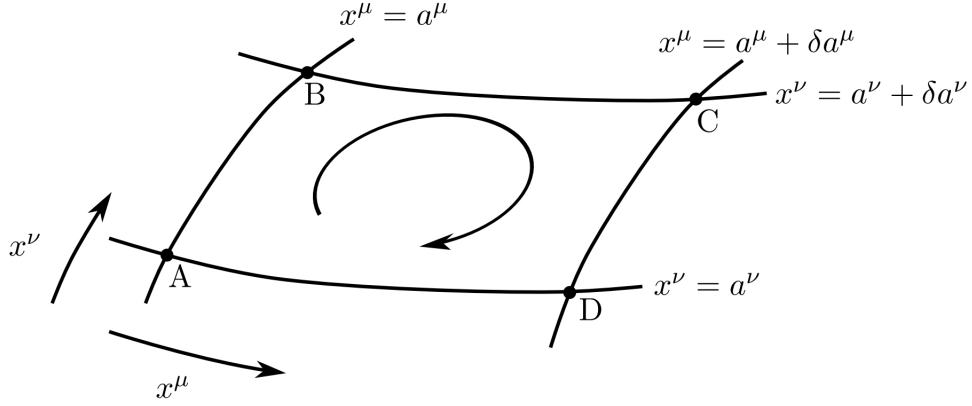


Figura 2.3: Circuito cerrado sobre la variedad definido por valores constantes de las coordenadas. Imagen tomada de [21].

A través de las líneas coordenadas se transporta paralelamente el vector \mathbf{v} . Entonces, para la dirección x^ν con x^μ fija se tiene

$$\nabla_{e_\nu} \mathbf{v} = 0 \Rightarrow \frac{\partial v^\gamma}{\partial x^\nu} = -v^\alpha \Gamma_{\alpha\nu}^\gamma, \quad (2.27)$$

donde se utilizó la Ec. (2.26). Definamos \mathbf{v}_i como el vector \mathbf{v} en la posición inicial. El cambio del vector después del transporte paralelo en el segmento AB está dado por el valor inicial del vector más el cambio del vector sobre la coordenada x^ν , i.e.

$$v^\gamma(B) = v_i^\gamma + \int_A^B \frac{\partial v^\gamma}{\partial x^\nu} dx^\nu = v_i^\gamma - \int_A^B v^\alpha \Gamma_{\alpha\nu}^\gamma dx^\nu. \quad (2.28)$$

De manera similar se pueden calcular los cambios del vector al ir de un punto del circuito al siguiente:

$$v^\gamma(C) = v^\gamma(B) - \int_B^C v^\alpha \Gamma_{\alpha\mu}^\gamma dx^\mu, \quad (2.29)$$

$$v^\gamma(D) = v^\gamma(C) - \int_C^D v^\alpha \Gamma_{\alpha\nu}^\gamma dx^\nu, \quad (2.30)$$

$$v_f^\gamma = v^\gamma(D) - \int_D^A v^\alpha \Gamma_{\alpha\mu}^\gamma dx^\mu. \quad (2.31)$$

El cambio total de v después de recorrer el circuito completo es $\delta v^\gamma = v_f^\gamma - v_i^\gamma$, que utilizando los valores de los cambios en cada segmento es

$$\delta v^\gamma = - \int_A^B v^\alpha \Gamma^\gamma_{\alpha\nu} dx^\nu - \int_B^C v^\alpha \Gamma^\gamma_{\alpha\mu} dx^\mu - \int_C^D v^\alpha \Gamma^\gamma_{\alpha\nu} dx^\nu - \int_D^A v^\alpha \Gamma^\gamma_{\alpha\mu} dx^\mu . \quad (2.32)$$

Los valores dentro de la integral se deben calcular tomando la coordenada correspondiente constante y considerando que δa^ν y δa^μ son muy pequeños. Así, por ejemplo,

$$\left(v^\alpha \Gamma^\gamma_{\alpha\nu} \right) \Big|_{x^\mu = a^\mu + \delta a^\mu} \approx \left(v^\alpha \Gamma^\gamma_{\alpha\nu} \right) \Big|_{x^\mu = a^\mu} + \delta a^\mu \frac{\partial}{\partial x^\mu} \left(v^\alpha \Gamma^\gamma_{\alpha\nu} \right) \Big|_{x^\mu = a^\mu} , \quad (2.33)$$

por lo que el cambio del vector sobre las líneas con x^ν fija, que corresponde a los segmentos AB y CD , es

$$\begin{aligned} & - \int_A^B v^\alpha \Gamma^\gamma_{\alpha\nu} dx^\nu - \int_C^D v^\alpha \Gamma^\gamma_{\alpha\nu} dx^\nu \approx \\ & - \int_A^B dx^\nu \left(v^\alpha \Gamma^\gamma_{\alpha\nu} \right) \Big|_{x^\mu = a^\mu} - \int_C^D dx^\nu \left(v^\alpha \Gamma^\gamma_{\alpha\nu} \right) \Big|_{x^\mu = a^\mu} - \delta a^\mu \int_C^D dx^\nu \frac{\partial}{\partial x^\mu} \left(v^\alpha \Gamma^\gamma_{\alpha\nu} \right) \Big|_{x^\mu = a^\mu} , \end{aligned} \quad (2.34)$$

y usando

$$\int_A^B dx^\nu = \delta a^\nu , \quad (2.35)$$

$$\int_C^D dx^\nu = -\delta a^\nu , \quad (2.36)$$

donde el signo menos surge de que el circuito CD se recorre en el sentido opuesto, se tiene que

$$- \int_A^B v^\alpha \Gamma^\gamma_{\alpha\nu} dx^\nu - \int_C^D v^\alpha \Gamma^\gamma_{\alpha\nu} dx^\nu \approx \delta a^\nu (-v^\alpha \Gamma^\gamma_{\alpha\nu} + v^\alpha \Gamma^\gamma_{\alpha\nu}) + \delta a^\mu \delta a^\nu \frac{\partial}{\partial x^\mu} \left(v^\alpha \Gamma^\gamma_{\alpha\nu} \right) , \quad (2.37)$$

donde se utilizó la suposición de que $v^\alpha \Gamma^\gamma_{\alpha\nu}$ y sus derivadas casi no cambian en la dirección x^ν . Realizando el mismo análisis para los demás segmentos se tiene

$$\begin{aligned} \delta v^\gamma & \approx \delta a^\mu \delta a^\nu \frac{\partial}{\partial x^\mu} \left(v^\alpha \Gamma^\gamma_{\alpha\nu} \right) - \delta a^\mu \delta a^\nu \frac{\partial}{\partial x^\nu} \left(v^\alpha \Gamma^\gamma_{\alpha\mu} \right) \\ & = \delta a^\mu \delta a^\nu \left[\Gamma^\gamma_{\alpha\nu,\mu} - \Gamma^\gamma_{\sigma\nu} \Gamma^\sigma_{\alpha\mu} - \Gamma^\gamma_{\alpha\mu,\nu} + \Gamma^\gamma_{\sigma\mu} \Gamma^\sigma_{\alpha\nu} \right] v^\alpha , \end{aligned} \quad (2.38)$$

donde se utiliza la notación con *coma* para la derivada ordinaria, i.e. $\Gamma^\gamma_{\alpha\nu,\mu} \equiv \frac{\partial}{\partial x^\mu} (\Gamma^\gamma_{\alpha\nu})$. Definiendo el **tensor de Riemann** o **tensor de curvatura** como aquel con componentes

$$R^\gamma_{\alpha\mu\nu} := \Gamma^\gamma_{\alpha\nu,\mu} - \Gamma^\gamma_{\sigma\nu}\Gamma^\sigma_{\alpha\mu} - \Gamma^\gamma_{\alpha\mu,\nu} + \Gamma^\gamma_{\sigma\mu}\Gamma^\sigma_{\alpha\nu} , \quad (2.39)$$

se puede expresar el cambio de v como

$$\delta v^\gamma = R^\gamma_{\alpha\mu\nu} v^\alpha \delta a^\mu \delta a^\nu . \quad (2.40)$$

Debido a que las componentes del tensor de Riemann tienen derivadas de los símbolos de Christoffel, éste incluye segundas derivadas de la métrica. Otra definición del tensor de curvatura, que de hecho es superior a la anterior porque sí es manifiestamente covariante, es a través del conmutador de derivadas covariantes, $[\nabla_\mu, \nabla_\nu] := \nabla_\mu \nabla_\nu - \nabla_\nu \nabla_\mu$, como

$$[\nabla_\mu, \nabla_\nu] \omega^\gamma = R^\gamma_{\alpha\mu\nu} \omega^\alpha \quad \forall \omega \in \mathfrak{X}(\mathcal{M}) , \quad (2.41)$$

que es conocida como la **identidad de Ricci**. Esta definición permite ver al tensor de Riemann como la medida de la falta de conmutatividad de las derivadas covariantes, que de alguna manera es lo que ocurre en el desarrollo del transporte paralelo en el circuito cerrado de la Fig. 2.3.

El tensor de Riemann permite hacer una clara distinción entre una variedad plana y una curva, pues en la primera se tiene una definición global de paralelismo, i.e. un vector puede ser transportado paralelamente alrededor de cualquier curva y terminar igual. Así, en todos los puntos de \mathcal{M}

$$R^\gamma_{\alpha\mu\nu} = 0 \iff \text{variedad plana} .$$

2.2.3. Planitud local

Sea η el tensor métrico en una variedad Lorentziana plana. En su representación matricial se tiene que $(\eta_{\alpha\beta}) = \text{diag}(-1, +1, +1, +1)$. Ahora, una variedad localmente debe verse plana pues es homeomorfa a \mathbb{R}^n a través de las cartas. De hecho, dada una matriz simétrica, ésta siempre se puede llevar a una representación diagonal con entradas ± 1 o 0, dependiendo de sus eigenvalores. Y como $(g_{\alpha\beta})$ tiene signatura $-+++$ y es simétrica, entonces siempre se puede encontrar una transformación tal que, en ese punto,

$$(g_{\alpha\beta}) = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \end{pmatrix} = (\eta_{\alpha\beta}) . \quad (2.42)$$

Esto implica que dado un punto $p \in \mathcal{M}$ siempre se puede hacer que $(g_{\alpha\beta})$ en ese punto sea precisamente $(\eta_{\alpha\beta})$, y que en los alrededores sea casi $(\eta_{\alpha\beta})$ con correcciones de segundo orden en las coordenadas. De hecho, la curvatura va a determinar el peso de dichas correcciones, por lo que en los alrededores de p se va a tener

$$g_{\alpha\beta} \approx \eta_{\alpha\beta} - \frac{1}{3} R_{\alpha\beta\mu\nu} x^\mu x^\nu , \quad (2.43)$$

donde $x = p$ es el origen del sistema coordenado y x^μ, x^ν son pequeños. Este aspecto de planitud local se puede formular como [22]:

$$g_{\alpha\beta}(p) = \eta_{\alpha\beta} ; \quad (2.44)$$

$$\frac{\partial}{\partial x^\lambda} g_{\alpha\beta}(p) = 0 \quad \forall \alpha, \beta, \lambda ; \quad (2.45)$$

$$\frac{\partial}{\partial x^\sigma} \left(\frac{\partial}{\partial x^\lambda} g_{\alpha\beta}(p) \right) \neq 0 \quad \text{para algunos } \alpha, \beta, \lambda, \sigma , \quad (2.46)$$

en el punto p que se considera. El que se satisfagan las primeras dos hace más fáciles los cálculos de manera local en este sistema coordenado; en particular, por ejemplo, hace que los $\Gamma_{\mu\nu}^\gamma$ se anulen en p . Si bien los resultados que se obtengan de esta forma serán en este sistema coordenado particular, también se pueden generalizar a cualquier otro sistema coordenado. Esto se debe a que los tensores se transforman de tal manera que las ecuaciones tensoriales que son válidas en un sistema coordenado lo seguirán siendo en cualquier otro siempre y cuando estos se relacionen por medio de un difeomorfismo⁴.

2.2.4. Simetrías y propiedades del tensor de Riemann

Sea un punto $p \in \mathcal{M}$ y un sistema coordenado tal que $g^{\alpha\beta}_{,\sigma} = 0$. Pero sabemos que no podemos hacer las segundas derivadas todas nulas. Por lo tanto, en esta localidad se tiene que

$$\Gamma_{\mu\nu,\sigma}^\gamma = \frac{1}{2} g^{\alpha\beta} (g_{\beta\mu,\nu\sigma} + g_{\beta\nu,\mu\sigma} - g_{\mu\nu,\beta\sigma}) , \quad (2.47)$$

⁴Es importante mencionar que los símbolos de Christoffel no se transforman como tensores.

y como $\Gamma^\alpha_{\mu\nu} = 0$, el tensor de Riemann se vuelve

$$\begin{aligned}
R^\beta_{\alpha\mu\nu} &= \Gamma^\beta_{\alpha\nu,\mu} - \Gamma^\beta_{\alpha\mu,\nu} \\
&= \frac{1}{2}g^{\beta\gamma}(g_{\gamma\alpha,\nu\mu} + g_{\nu\gamma,\alpha\mu} - g_{\alpha\nu,\gamma\mu} - g_{\gamma\alpha,\nu\mu} - g_{\mu\gamma,\alpha\nu} + g_{\alpha\mu,\gamma\nu}) \\
&= \frac{1}{2}g^{\beta\gamma}(g_{\nu\gamma,\alpha\mu} - g_{\alpha\nu,\gamma\mu} + g_{\alpha\mu,\gamma\nu} - g_{\mu\gamma,\alpha\nu}) .
\end{aligned} \tag{2.48}$$

De esta manera obtenemos una expresión para el tensor de Riemann de manera local. Bajando el primer índice con la métrica se llega a

$$\begin{aligned}
R_{\alpha\beta\mu\nu} &= g_{\alpha\lambda}R^\lambda_{\beta\mu\nu} \\
&= \frac{1}{2}g_{\alpha\lambda}g^{\lambda\gamma}(g_{\nu\gamma,\beta\mu} - g_{\beta\nu,\gamma\mu} + g_{\beta\mu,\gamma\nu} - g_{\mu\gamma,\beta\nu}) \\
&= \frac{1}{2}\delta^\gamma_\alpha(g_{\nu\gamma,\beta\mu} - g_{\beta\nu,\gamma\mu} + g_{\beta\mu,\gamma\nu} - g_{\mu\gamma,\beta\nu}) \\
&= \frac{1}{2}(g_{\nu\alpha,\beta\mu} - g_{\beta\nu,\alpha\mu} + g_{\beta\mu,\alpha\nu} - g_{\mu\alpha,\beta\nu}) .
\end{aligned} \tag{2.49}$$

Debido a la simetría del tensor métrico se puede ver que se satisfacen las siguientes simetrías:

$$R_{\alpha\beta\mu\nu} = -R_{\beta\alpha\mu\nu} , \tag{2.50}$$

$$R_{\alpha\beta\mu\nu} = -R_{\alpha\beta\nu\mu} , \tag{2.51}$$

$$R_{\alpha\beta\mu\nu} = R_{\mu\nu\alpha\beta} . \tag{2.52}$$

Como estas expresiones son tensoriales, son válidas en cualquier sistema coordenado y en cualquier punto de la variedad. Esta es la ventaja que se comentó en la subsección anterior de facilitar los cálculos para después generalizarlos. Además, se satisface la llamada *propiedad cíclica*, la cual es

$$R_{\alpha\beta\mu\nu} + R_{\alpha\nu\beta\mu} + R_{\alpha\mu\nu\beta} = 0 . \tag{2.53}$$

En principio, para $n = 4$ el tensor $R^\alpha_{\beta\mu\nu}$ (y $R_{\alpha\beta\mu\nu}$) tiene $4^4 = 256$ componentes, pero debido a sus simetrías y a la propiedad cíclica se tienen en realidad sólo 20 componentes independientes. Para d dimensiones son $\frac{1}{12}d^2(d^2 - 1)$ componentes independientes.

Localmente se tiene

$$R_{\alpha\beta\mu\nu,\lambda} = \frac{1}{2}(g_{\nu\alpha,\beta\mu\lambda} - g_{\beta\nu,\alpha\mu\lambda} + g_{\beta\mu,\alpha\nu\lambda} - g_{\mu\alpha,\beta\nu\lambda}) , \tag{2.54}$$

y permutando algunos índices se verifica que

$$R_{\alpha\beta\mu\nu;\lambda} + R_{\alpha\beta\lambda\mu;\nu} + R_{\alpha\beta\nu\lambda;\mu} = 0 . \quad (2.55)$$

Ahora, como en este punto y en esta elección de coordenadas se satisface $\Gamma^\mu_{\alpha\beta} = 0$, entonces la expresión anterior es equivalente a

$$R_{\alpha\beta\mu\nu;\lambda} + R_{\alpha\beta\lambda\mu;\nu} + R_{\alpha\beta\nu\lambda;\mu} = 0 . \quad (2.56)$$

En la Ec. (2.56) se utilizó la notación de *punto y coma* para denotar a la derivada covariante. Esta expresión, llamada **Identidad de Bianchi**, es tensorial, por lo que es igual de válida sobre toda la variedad y en cualquier sistema coordenado.

El tensor de Riemann se puede contraer para tener un tensor de orden menor. En particular, se pueden contraer el primer y el tercer índice, haciendo

$$R_{\alpha\beta} := R^\mu_{\alpha\mu\beta} = g^{\mu\nu} R_{\nu\alpha\mu\beta} , \quad (2.57)$$

que recibe el nombre de **tensor de Ricci**, que es un tensor de orden $(0, 2)$ simétrico, i.e. $R_{\alpha\beta} = R_{\beta\alpha}$. Otras posibles contracciones se hacen cero o dan como resultado $\pm R_{\alpha\beta}$ debido a la antisimetría de $R_{\alpha\beta\mu\nu}$. En este sentido, se puede decir que el tensor de Ricci es la única contracción del tensor de Riemann [22].

Otra definición importante es el **escalar de Ricci** o **escalar de curvatura**, el cual está dado por la traza del tensor de Ricci con respecto a la métrica, es decir,

$$R := g^{\alpha\beta} R_{\alpha\beta} . \quad (2.58)$$

Como la métrica conmuta con la derivada covariante, se puede contraer la Ec. (2.56) como

$$g^{\alpha\beta} (R_{\alpha\beta\mu\nu;\lambda} + R_{\alpha\beta\lambda\mu;\nu} + R_{\alpha\beta\nu\lambda;\mu}) = R_{\beta\nu;\lambda} - R_{\beta\lambda;\nu} + R^\mu_{\beta\nu\lambda;\mu} = 0 . \quad (2.59)$$

Contrayendo una vez más se tiene

$$\begin{aligned}
g^{\beta\nu}(R_{\beta\nu;\lambda} - R_{\beta\lambda;\nu} + R^{\mu}_{\beta\nu\lambda;\mu}) &= R_{;\lambda} - R^{\nu}_{\lambda} + R^{\mu\nu}_{\nu\lambda;\mu} \\
&= R_{;\lambda} - R^{\nu}_{\lambda;\nu} - R^{\nu\mu}_{\nu\lambda;\mu} \\
&= R_{;\lambda} - R^{\nu}_{\lambda;\nu} - R^{\mu}_{\lambda;\mu} \\
&= R_{;\lambda} - 2R^{\nu}_{\lambda;\nu} \\
&= R_{;\lambda} - 2g_{\lambda\mu}R^{\nu\mu}_{;\nu} = 0 .
\end{aligned} \tag{2.60}$$

Esto implica que

$$\begin{aligned}
\frac{1}{2}R_{;\lambda} - g_{\lambda\mu}R^{\mu\nu}_{;\nu} &= 0 \\
\Rightarrow \frac{1}{2}g^{\lambda\mu}R_{;\lambda} - R^{\mu\nu}_{;\nu} &= 0 \\
\Rightarrow \frac{1}{2}g^{\mu\nu}R_{;\nu} - R^{\mu\nu}_{;\nu} &= 0 \\
\Rightarrow \left(\frac{1}{2}g^{\mu\nu}R - R^{\mu\nu}\right)_{;\nu} &= 0 .
\end{aligned} \tag{2.61}$$

Definiendo el **tensor de Einstein** como aquel con componentes

$$G^{\mu\nu} := R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R , \tag{2.62}$$

que es un tensor simétrico, se llega a que

$$(-G^{\mu\nu})_{;\nu} = 0 \quad \Rightarrow \quad G^{\mu\nu}_{;\nu} = 0 . \tag{2.63}$$

De esta manera se tiene que el tensor de Einstein construido a partir del tensor de Ricci y el escalar de curvatura tiene en automático divergencia cero. Además, por ser simétrico tiene solamente 10 componentes independientes en 4 dimensiones.

2.2.5. Geodésicas

Sea un vector \mathbf{v} y γ una curva con vector tangente \mathbf{u} . La condición de transporte paralelo de \mathbf{v} a lo largo de γ es $\nabla_{\mathbf{u}}\mathbf{v} = 0$, la cual es una condición local, i.e. dado un punto p , \mathbf{v} se transporta al siguiente punto de tal manera que no cambie y se mantenga paralelo a sí mismo. Si \mathbf{u} es el vector tangente a la curva, entonces $\mathbf{u} = \frac{d\mathbf{x}}{d\lambda}$ con λ el parámetro de la curva. Como pedimos que el vector

no cambie,

$$\frac{d\mathbf{v}}{d\lambda} = \frac{dx^\alpha}{d\lambda} \frac{\partial \mathbf{v}}{\partial x^\alpha} = u^\alpha \mathbf{v}_{,\alpha} = 0 . \quad (2.64)$$

Al trabajar localmente, los símbolos de Christoffel son cero, por lo que la condición de transporte paralelo se puede reescribir en coordenadas como

$$u^\alpha v^\beta_{,\alpha} = u^\alpha v^\beta_{;\alpha} = \frac{dx^\alpha}{d\lambda} \frac{\partial v^\beta}{\partial x^\alpha} + \Gamma^\beta_{\nu\alpha} \frac{dx^\alpha}{d\lambda} v^\nu = 0 . \quad (2.65)$$

Definición 2.2.2 Una curva γ en una variedad pseudo-Riemanniana $(\mathcal{M}, \mathbf{g})$ es llamada una **geodésica** si y sólo si admite una parametrización \mathcal{P} cuyo vector tangente asociado, \mathbf{u} , se transporta paralelamente a sí mismo a lo largo de γ :

$$\nabla_{\mathbf{u}} \mathbf{u} = 0 . \quad (2.66)$$

La parametrización \mathcal{P} es llamada **parametrización afín** y el parámetro λ es el **parámetro afín**.

Cualesquiera dos parametrizaciones afines de una geodésica γ están relacionadas por $\lambda' = a\lambda + b$, con $a, b \in \mathbb{R}$, que es llamada una transformación afín. Las componentes de $\nabla_{\mathbf{u}} \mathbf{u}$ son

$$\begin{aligned} u^\beta u^\alpha_{;\beta} &= u^\beta (u^\alpha_{,\beta} + \Gamma^\alpha_{\gamma\beta} u^\gamma) \\ &= \frac{dx^\beta}{d\lambda} \left(\frac{\partial}{\partial x^\beta} \frac{dx^\alpha}{d\lambda} + \Gamma^\alpha_{\gamma\beta} \frac{dx^\gamma}{d\lambda} \right) \\ &= \frac{dx^\beta}{d\lambda} \frac{\partial}{\partial x^\beta} \frac{dx^\alpha}{d\lambda} + \Gamma^\alpha_{\gamma\beta} \frac{dx^\gamma}{d\lambda} \frac{dx^\beta}{d\lambda} \\ &= \frac{d^2 x^\alpha}{d\lambda^2} + \Gamma^\alpha_{\gamma\beta} \frac{dx^\gamma}{d\lambda} \frac{dx^\beta}{d\lambda} . \end{aligned} \quad (2.67)$$

De esta manera, la ecuación de la geodésica se transforma en ecuaciones diferenciales para (x^μ) :

$$\frac{d^2 x^\alpha}{d\lambda^2} + \Gamma^\alpha_{\gamma\beta} \frac{dx^\gamma}{d\lambda} \frac{dx^\beta}{d\lambda} = 0 . \quad (2.68)$$

Sea \mathbf{u} un campo vectorial tangente a una geodésica γ , la cual tiene una parametrización afín con parámetro λ . Tenemos que la derivada de $\mathbf{g}(\mathbf{u}, \mathbf{u})$ con respecto a λ es

$$\begin{aligned}
 \frac{d}{d\lambda}(\mathbf{g}(\mathbf{u}, \mathbf{u})) &= \mathbf{u}(\mathbf{g}(\mathbf{u}, \mathbf{u})) \\
 &= \nabla_{\mathbf{u}}(\mathbf{g}(\mathbf{u}, \mathbf{u})) \\
 &= u^\mu (\mathbf{g}(\mathbf{u}, \mathbf{u}))_{;\mu} \\
 &= u^\mu (g_{\rho\sigma} u^\rho u^\sigma)_{;\mu} \\
 &= u^\mu (g_{\rho\sigma})_{;\mu} u^\rho u^\sigma + u^\mu g_{\rho\sigma} (u^\rho)_{;\mu} u^\sigma + u^\mu g_{\rho\sigma} u^\rho (u^\sigma)_{;\mu}.
 \end{aligned} \tag{2.69}$$

Como $\nabla \mathbf{g} = 0$ con respecto a cualquier vector y \mathbf{u} satisface la ecuación de la geodésica, entonces

$$\frac{d}{d\lambda}(\mathbf{g}(\mathbf{u}, \mathbf{u})) = 0 \Rightarrow \mathbf{g}(\mathbf{u}, \mathbf{u}) = \text{cte}. \tag{2.70}$$

La consistencia del valor de $\mathbf{g}(\mathbf{u}, \mathbf{u})$ a lo largo de la geodésica garantiza que el vector tangente siempre sea del mismo tipo: tiempo, espacio, o nulo. Así, cualquier geodésica en $(\mathcal{M}, \mathbf{g})$ pertenece a uno de los siguiente tipos:

1. Geodésica tipo tiempo (vector tangente tipo tiempo);
2. Geodésica tipo espacio (vector tangente tipo espacio);
3. Geodésica nula (vector tangente nulo).

De esta forma se ve que las geodésicas son curvas especiales en la variedad no sólo porque transportan paralelamente a su vector tangente, sino porque además éste es siempre del mismo tipo. Por otro lado, el vector tangente a una curva cualquiera puede cambiar de tipo a lo largo de ésta. Más adelante se mencionará con mayor detalle que las geodésicas son curvas especiales porque también extremizan la distancia entre dos puntos de la variedad.

Una geodésica γ se llama **inextendible** o **maximal** si no existe ninguna otra geodésica γ' tal que $\gamma \subset \gamma'$ y $\gamma \neq \gamma'$. Además, una geodésica es **completa** si el intervalo de valores de su parámetro afín λ son los números reales, i.e. $I = \mathbb{R}$. De esta manera, toda geodésica completa es inextendible [20]. Una variedad $(\mathcal{M}, \mathbf{g})$ es **geodésicamente completa** si toda geodésica inextendible es también completa. El caso en el que esto no se satisface implica que las geodésicas “terminan”, es decir, no hay geodésicas que las contengan pero su parámetro llega hasta un valor finito.

Capítulo 3

Relatividad General

A lo largo del Capítulo 2 se desarrollaron conceptos matemáticos de Geometría Diferencial sin hacer en ningún momento mención o conexión con algún concepto físico. La Sección 3.1 tiene como objetivo la construcción de la teoría de la Relatividad General (RG) desde la postulación de sus principios utilizando todos los conceptos introducidos en el Cap. 2, con el fin de que se tenga un puente entre lo físico y lo matemático. Una vez que se postulan los principios de la RG, se da una sección completa para entender el papel que juega la curvatura del *espacio-tiempo* en el concepto de gravedad, en particular en el caso de un campo gravitacional débil. Luego, entendiendo este caso en concreto, se pasa a la justificación de las Ecuaciones de Campo de Einstein, las cuales muestran la conexión entre la curvatura (estructura matemática) con la gravedad (manifestaciones físicas de la teoría). Finalmente, se estudia el caso de un sistema estático y con simetría esférica y se habla de sus consecuencias en la causalidad entre eventos y cómo éstas dan origen a lo que se conoce como *agujeros negros*, así como las ecuaciones que rigen el movimiento de una partícula que orbita dicho objeto; evidentemente, este proceso se hace para darnos las herramientas necesarias para poder describir matemáticamente la órbita de la estrella S2 alrededor de Sgr A*.

3.1. Principios y construcción

Para tener una teoría física con base en algún formalismo matemático se debe establecer una relación entre los conceptos matemáticos y los fenómenos físicos que se busca describir. Aunado a esto, se debe tener una imagen clara del entorno físico para así poder conectar lo observado con el lenguaje y conceptos matemáticos. Con este propósito se considera lo siguiente [23]:

El espacio-tiempo es una variedad diferenciable 4-dimensional real con una métrica Lorentziana.

Cada punto del espacio-tiempo se conoce como **evento** y representa un punto en el espacio en algún instante del tiempo; además, le estamos asociando una estructura matemática particular. Esta consideración permite utilizar toda la maquinaria matemática descrita en el Capítulo 2 para dar una descripción del universo y la relación entre sus eventos. Como se irá viendo a lo largo de este capítulo, todas las definiciones y resultados del capítulo anterior tienen una correspondencia física en este marco. Por ejemplo, la estructura diferenciable de la variedad, i.e., el espacio-tiempo, hace posible la descripción de sistemas físicos por medio de ecuaciones diferenciales; pero también se tendrán consecuencias físicas cruciales debido a la curvatura, como hacer que la estructura causal sea no trivial.

Para continuar con la construcción de la teoría hace falta todavía introducir principios sobre los cuales se formulará. Uno crucial en la teoría de la gravitación es el **Principio de Equivalencia Débil** [4], el cual establece que la masa inercial, m_i , es equivalente a la masa gravitacional, m_g . Esto permite que la ecuación de movimiento de la gravitación Newtoniana se reduzca a

$$\ddot{\mathbf{x}} = -\nabla\phi(\mathbf{x}) , \quad (3.1)$$

con $\phi(\mathbf{x})$ el potencial gravitacional Newtoniano. Esto indica que, dado un campo gravitacional, las partículas están sujetas a la misma aceleración, sin importar sus masas. Esta observación es de suma importancia porque permite ver que el movimiento en caída libre es en realidad algo universal. Debido a esta característica del movimiento en un campo gravitacional, uno siempre puede ir a un sistema coordinado en caída libre en el cual las partículas¹ cercanas no están aceleradas; como no están aceleradas, deben moverse en línea recta, al menos de manera local. Este hecho permite establecer una clara correspondencia entre el movimiento de partículas en un campo gravitacional y el concepto de geodésica. Considerando todo lo anterior, el Principio de Equivalencia Débil se puede escribir como [22]:

Las partículas en caída libre se mueven a través del espacio-tiempo siguiendo geodésicas.

Es importante resaltar que por *caída libre* se entiende que las partículas no están siendo afectadas por otras fuerzas, sino que solamente sienten la influencia de la gravedad. Un experimento pensado (*Gedankenexperiment*, por su nombre en alemán) que arroja un poco de luz sobre la naturaleza de la gravedad es el siguiente: sea una caja sellada con dos partículas dentro en presencia del campo

¹Siempre nos referiremos a partículas de prueba.

gravitacional terrestre. Si la caja es pequeña y las partículas están en caída libre, en esta pequeña región éstas se mueven en líneas paralelas. Por otro lado, si la caja es lo suficientemente grande y las partículas están bastante alejadas, entonces sus trayectorias dejan de ser paralelas y comienzan a acercarse, como se observa en la Fig. 3.1.

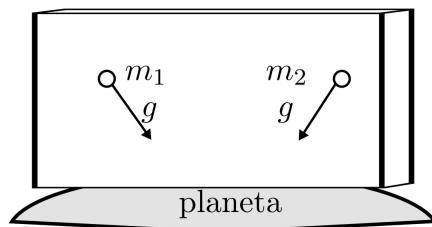


Figura 3.1: Experimento pensado sobre el papel de la gravedad y las trayectorias de partículas. Las trayectorias de dos partículas alejadas en presencia de un campo gravitacional dejan de ser paralelas. Imagen tomada de [21].

Este pequeño experimento pensado permite ver que en realidad la gravedad no introduce nada nuevo de manera local, y con base en esta conclusión se puede formular el llamado **Principio de Equivalencia Fuerte** [22]:

Cualquier experimento físico local que no incluya la gravedad va a arrojar los mismos resultados si se realiza en un marco de referencia en caída libre o en el espacio-tiempo plano de la Relatividad Especial.

Este principio implica que cualquier ley física local expresada de manera tensorial en Relatividad Especial tiene la misma forma en un marco de referencia inercial en el espacio-tiempo curvo, puesto que por la planitud local de la variedad siempre se puede hacer que $g_{\mu\nu} = \eta_{\mu\nu}$ en cualquier punto (Ver Sec. 2.2.3). Esta discusión permite desechar por completo la idea de la gravedad como una fuerza a distancia tal como lo pensaba Newton, y más bien verla como la consecuencia de la curvatura del espacio-tiempo. Entonces se puede concluir que

$$\text{Gravedad} \equiv \text{Curvatura} .$$

3.1.1. Trayectorias a través del espacio-tiempo

En la subsección anterior vimos que en la Relatividad General la gravedad es la consecuencia de la curvatura del espacio-tiempo y que las partículas libres siguen geodésicas. Estas conclusiones nos

permiten estudiar de manera más precisa cómo es que se mueven partículas a través del espacio-tiempo, por lo que podremos describir matemáticamente el movimiento de la estrella S2 alrededor de Sgr A*.

Definimos la **línea de mundo** de una partícula puntual como la curva γ que sigue esa partícula a través del espacio-tiempo. De esta forma, podemos pensar a γ como el conjunto de eventos, o bien, el conjunto de posiciones conforme el tiempo evoluciona. Como estamos describiendo el movimiento de una partícula, γ debe ser forzosamente una curva causal, es decir, tipo tiempo o nula. La dinámica de la partícula está totalmente descrita por su 4-momento, \mathbf{p} , que es un campo vectorial tangente a γ que en cada punto apunta hacia el futuro.

Para el caso de partículas masivas se tiene que γ es una curva tipo tiempo, i.e., $\mathbf{g}(\mathbf{p}, \mathbf{p}) < 0$, y además el tiempo propio², τ , de la partícula puede ser el parámetro afín. Por otro lado, para partículas sin masa, γ es una curva nula y entonces $\mathbf{g}(\mathbf{p}, \mathbf{p}) = 0$; en ambos casos se define la **masa** de la partícula como $m = \sqrt{-\mathbf{p} \cdot \mathbf{p}}$ [20]. La condición de que γ sea tipo tiempo o nula implica que la curva siempre debe estar dentro del cono de luz³ o a lo largo de éste, respectivamente, como se muestra en la Fig. 3.2.

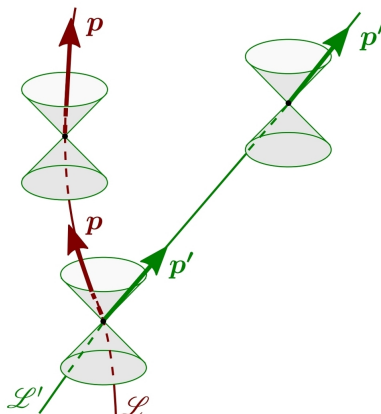


Figura 3.2: La curva \mathcal{L} representa la línea de mundo de una partícula masiva, por lo que siempre debe permanecer dentro del cono de luz en cada punto. Por otro lado, la curva \mathcal{L}' es la línea de mundo de una partícula sin masa, haciendo que ésta siempre esté a lo largo del cono de luz. Imagen tomada de [20].

²El tiempo propio es el tiempo de un reloj que se mueve con el marco de referencia de dicha partícula.

³Dado un evento, su cono de luz es la superficie que contiene todos los demás que están causalmente conectados con éste [21].

Si se define la acción de una partícula libre y puntual como

$$S = \int_A^B m d\tau = \int_{\lambda_A}^{\lambda_B} m \sqrt{-\mathbf{g} \left(\frac{d\mathbf{x}}{d\lambda}, \frac{d\mathbf{x}}{d\lambda} \right)} d\lambda, \quad (3.2)$$

utilizando el Principio de Acción Estacionaria se puede observar que pedir $\delta S = 0$ implica las Ecs. (2.68). Dicho de otra forma, las curvas γ que extremizan la acción (3.2) son precisamente las geodésicas. En la definición de la acción se toma $d\mathbf{x}$ como el vector de desplazamiento infinitesimal a lo largo de γ que apunta hacia el futuro, y tomando la 4-velocidad como $\mathbf{u} := d\mathbf{x}/d\tau$ se tiene que es tangente a la curva por construcción, haciendo que sea colineal con \mathbf{p} .

3.2. Ecuaciones de campo

Pensando que efectivamente hay una conexión entre la curvatura del espacio-tiempo y la gravedad, un campo gravitacional débil debe implicar una curvatura pequeña, es decir, una ligera desviación de un espacio-tiempo plano. Para estudiar este caso tomemos las siguientes suposiciones:

1. Estamos en un marco donde las partículas son no-relativistas, i.e.,

$$|u^i| \ll c, \text{ con } c \text{ la velocidad de la luz y } u^i \text{ la } i\text{-ésima componente espacial del vector;} \quad (3.3)$$

2. El espacio-tiempo es casi plano, por lo que la métrica se puede aproximar como

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad |h_{\mu\nu}| \ll 1; \quad (3.4)$$

3. El campo gravitacional es estacionario, por lo que existe alguna coordenada x^0 tal que

$$g_{\mu\nu,0} = 0. \quad (3.5)$$

Para analizar el comportamiento de partículas libres en esta situación debemos estudiar la ecuación de la geodésica implementando las suposiciones anteriores. Comenzando por la parte de los

símbolos de Christoffel se tiene

$$\begin{aligned}
\Gamma^{\mu}_{\alpha\beta} \frac{dx^{\alpha}}{d\tau} \frac{dx^{\beta}}{d\tau} &= \Gamma^{\mu}_{00} \left(\frac{dx^0}{d\tau} \right)^2 + 2\Gamma^{\mu}_{0i} \frac{dx^0}{d\tau} \frac{dx^i}{d\tau} + \Gamma^{\mu}_{ij} \frac{dx^i}{d\tau} \frac{dx^j}{d\tau} \\
&= \left(\Gamma^{\mu}_{00} + 2\Gamma^{\mu}_{0i} \frac{\partial x^i}{\partial x^0} + \Gamma^{\mu}_{ij} \frac{\partial x^i}{\partial x^0} \frac{\partial x^j}{\partial x^0} \right) \left(\frac{dx^0}{d\tau} \right)^2 \\
&= \left(\Gamma^{\mu}_{00} + 2\Gamma^{\mu}_{0i} \frac{u^i}{c} + \Gamma^{\mu}_{ij} \frac{u^i}{c} \frac{u^j}{c} \right) \left(\frac{dx^0}{d\tau} \right)^2 \\
&\approx \Gamma^{\mu}_{00} \left(\frac{dx^0}{d\tau} \right)^2,
\end{aligned} \tag{3.6}$$

$$\Rightarrow \frac{d^2 x^{\mu}}{d\tau^2} + \Gamma^{\mu}_{00} \left(\frac{dx^0}{d\tau} \right)^2 = 0. \tag{3.7}$$

En el desarrollo se utilizó la simetría de los símbolos de Christoffel y la suposición de que la partícula es no-relativista, por lo que $u^{\alpha}/c \ll 1$. Ahora, los términos Γ^{μ}_{00} se pueden escribir bajo las suposiciones como

$$\Gamma^{\mu}_{00} = \frac{1}{2} g^{\lambda\mu} (g_{\lambda 0,0} + g_{0\lambda,0} - g_{00,\lambda}) = -\frac{1}{2} g^{\lambda\mu} g_{00,\lambda}. \tag{3.8}$$

Por otro lado, para seguir satisfaciendo $g^{\mu\lambda} g_{\lambda\nu} = \delta^{\mu}_{\nu}$ se debe tomar $\eta^{\mu\lambda} h_{\lambda\nu} = \eta_{\lambda\nu} h^{\mu\lambda}$ con $g^{\mu\lambda} = \eta^{\mu\lambda} - h^{\mu\lambda}$. Así,

$$\begin{aligned}
\Gamma^{\mu}_{00} &= -\frac{1}{2} g^{\lambda\mu} g_{00,\lambda} \\
&\approx -\frac{1}{2} (\eta^{\mu\lambda} - h^{\mu\lambda}) (\eta_{00} + h_{00})_{,\lambda} \\
&= -\frac{1}{2} (\eta^{\mu\lambda} \eta_{00,\lambda} + \eta^{\mu\lambda} h_{00,\lambda} - h^{\mu\lambda} \eta_{00,\lambda} - h^{\mu\lambda} h_{00,\lambda}) \\
&= -\frac{1}{2} (\eta^{\mu\lambda} h_{00,\lambda} - h^{\mu\lambda} h_{00,\lambda}),
\end{aligned} \tag{3.9}$$

donde se utilizó el hecho de que $\eta_{00,\lambda} = 0$. Manteniéndonos a primer orden en las desviaciones tenemos $\Gamma^{\mu}_{00} \approx -\frac{1}{2} \eta^{\mu\lambda} h_{00,\lambda}$. Entonces la ecuación de la geodésica se vuelve

$$\frac{d^2 x^{\mu}}{d\tau^2} - \frac{1}{2} \eta^{\mu\lambda} h_{00,\lambda} \left(\frac{dx^0}{d\tau} \right)^2 = 0. \tag{3.10}$$

Tomando $\mu = 0$ tenemos que

$$\begin{aligned} \frac{d^2 x^0}{d\tau^2} - \frac{1}{2} \left(\frac{dx^0}{d\tau} \right)^2 \left(\eta^{00} h_{00,0} + \eta^{01} h_{00,1} + \eta^{02} h_{00,2} + \eta^{03} h_{00,3} \right) &= 0, \\ \Rightarrow \frac{d^2 x^0}{d\tau^2} = 0 &\Rightarrow \frac{dt}{d\tau} = \text{cte}. \end{aligned} \quad (3.11)$$

Por otro lado, si $\mu = i$,

$$\begin{aligned} \frac{d^2 x^i}{d\tau^2} &\approx \frac{1}{2} h_{00,i} \left(\frac{dx^0}{d\tau} \right)^2 = \frac{1}{2} h_{00,i} c^2 \left(\frac{dt}{d\tau} \right)^2, \\ \Rightarrow \frac{d^2 x^i}{dt^2} &\approx \frac{1}{2} c^2 h_{00,i}. \end{aligned} \quad (3.12)$$

Como estamos en el caso del campo débil, se debe recuperar la gravitación newtoniana, y por lo tanto $h_{00} = 2\phi/c^2$ para tener $d^2 x^i/dt^2 = -\phi_{,i}$ con $\phi = -GM/r$ el potencial gravitacional newtoniano de una distribución de masa M esférica y a una distancia r de la misma. De hecho, en general se tiene $h_{ii} = 2\phi/c^2$ [21]. Notando esto, el elemento de línea en este caso toma la forma

$$ds^2 = - \left(1 - \frac{2GM}{c^2 r} \right) (dx^0)^2 + \left(1 + \frac{2GM}{c^2 r} \right) (dx^i)^2. \quad (3.13)$$

El hecho de que $h_{\mu\nu} = 2\mathbb{I}_{\mu\nu}\phi/c^2$ muestra la conexión entre gravedad y curvatura que se ha ido desarrollando a lo largo de este trabajo. Además, permite ver de manera explícita que un campo débil genera una pequeña desviación de un espacio-tiempo plano, pues si $|\phi| \ll 1$, entonces $|h_{\mu\nu}| \ll 1$.

El siguiente paso es postular una ley que muestre cómo las fuentes del campo gravitacional generan la métrica. El análogo en el caso Newtoniano es la densidad de masa, ρ , la cual genera el potencial ϕ por medio de la ecuación $\nabla^2 \phi = 4\pi G\rho$. En este caso, la fuente de gravedad (la curvatura) debe estar relacionada con la densidad de masa; sin embargo, en la Relatividad General se está permitiendo cualquier sistema coordenado, por lo que la densidad de masa por sí misma no es la respuesta, pues implicaría un sistema coordenado preferido. La manera de resolver esto es mediante el tensor de energía-momento, \mathbf{T} , como la fuente del campo gravitacional [22]. El tensor de energía-momento es un tensor con dos índices y además es simétrico.

Ahora, como se habla de la curvatura del espacio-tiempo como la fuente de los campos gravitacionales, entonces esta ley debe conectar el tensor de energía-momento con algún o algunos términos relacionados con la curvatura, es decir, con segundas derivadas de la métrica. Para esto, introducimos un operador diferencial tensorial que actúa sobre la métrica, $\mathbf{O}(\mathbf{g})$. Así, la generalización de la

teoría gravitacional se vuelve

$$\mathbf{O}(g) = k\mathbf{T} , \quad (3.14)$$

con k una constante, y como \mathbf{O} está igualado a \mathbf{T} , debe producir un tensor de orden $(2, 0)$. Ahora, como hablamos de curvatura, las componentes de \mathbf{O} deben ser combinaciones de $g_{\mu\nu,\lambda\sigma}$, $g_{\mu\nu,\lambda}$ y $g_{\mu\nu}$. Con esto en mente, cualquier tensor de la forma $O^{\mu\nu} = R^{\mu\nu} + \alpha g^{\mu\nu} R + \Lambda g^{\mu\nu}$, con α, Λ constantes satisface lo anterior [22]. No obstante, el Principio de Equivalencia exige una conservación local de energía y momento, lo cual se traduce en

$$\nabla\mathbf{T} = 0 \Rightarrow T^{\mu\nu}_{;\mu} = 0 . \quad (3.15)$$

Debido a que se está tomando la Ecuación (3.14) como base y que \mathbf{T} debe satisfacer (3.15), entonces \mathbf{O} también debe satisfacerlo. Recordando que el tensor de Einstein, $\mathbf{G} = \mathbf{R} - \frac{1}{2}R\mathbf{g}$, satisface $\nabla\mathbf{G} = 0$ para una métrica arbitraria, se concluye que $\alpha = -\frac{1}{2}$, y como $\nabla\mathbf{g} = 0$, Λ puede tomar cualquier valor⁴. Así, se llega a las **Ecuaciones de campo de Einstein**, que en componentes son

$$G^{\mu\nu} + \Lambda g^{\mu\nu} = \frac{8\pi G}{c^4} T^{\mu\nu} . \quad (3.16)$$

El valor $k = 8\pi G/c^4$ resulta de pedir que las ecuaciones (3.16) den los valores correctos en los límites Newtonianos [22]. Es de suma importancia resaltar que las ecuaciones de campo también pueden obtenerse del principio variacional extremizando la acción de Einstein-Hilbert [20] (de igual manera se tiene que postular el Lagrangiano).

3.3. Solución de Schwarzschild

Muchos objetos astronómicos en el Universo pueden ser modelados como objetos esféricos; un ejemplo de lo anterior son las estrellas como el Sol, o incluso los mismos planetas. Es por esto que es relevante el estudio de sistemas dotados con simetría esférica en el contexto de la gravitación, y en particular en la Relatividad General, pues recordemos que buscamos describir el movimiento de una estrella alrededor de un objeto que, por simplicidad, podríamos suponerlo como esféricamente simétrico. Para poder hacer esto, primero se tiene que trabajar con coordenadas adecuadas para dicha simetría. En coordenadas (t, r, θ, ϕ) el espacio-tiempo plano o de Minkowski puede ser descrito

⁴En este contexto, Λ es llamada **constante cosmológica**.

por medio del elemento de línea

$$ds^2 = -c^2 dt^2 + dr^2 + r^2(d\theta^2 + \sin^2 \theta d\phi^2) . \quad (3.17)$$

Haciendo $r = \text{cte}$ y $t = \text{cte}$ el elemento de línea (3.17) toma la forma $ds^2 = r^2(d\theta^2 + \sin^2 \theta d\phi^2) := r^2 d\Omega^2$, que corresponde al elemento de línea de una 2-esfera. Es importante resaltar que cualquier superficie 2-dimensional que tenga un elemento de línea como el anterior donde r^2 es independiente de θ y ϕ tiene la geometría intrínseca de una 2-esfera.

Cualquier superficie con $r = \text{cte}$ y $t = \text{cte}$ es entonces una 2-esfera de área $4\pi r^2$, donde r es llamada **coordenada de curvatura** o **coordenada de área**. Estos nombres muestran que esta coordenada no está necesariamente relacionada con la distancia radial a algún origen, sino que es una etiqueta para las esferas que folian al espacio-tiempo. De hecho, el que un espacio-tiempo tenga simetría esférica no implica que necesariamente exista un punto en el centro.

Sean las esferas con r y $r + dr$. En principio, cada una de éstas tiene su propio sistema (θ, ϕ) , pero lo ideal es que estos estén relacionados entre sí de una manera que sea sencillo trabajar. Para esto, pedimos que la línea $\theta = \text{cte}$, $\phi = \text{cte}$ sea ortogonal a ambas esferas; este proceso es equivalente a pedir que los polos de las esferas con r y $r + dr$ estén alineados, por lo que el sistema de coordenadas (θ, ϕ) sea el mismo para ambas. Como esta línea tiene un vector tangente \mathbf{e}_r y los vectores \mathbf{e}_θ y \mathbf{e}_ϕ están sobre la esfera, la ortogonalidad se traduce en pedir que $\mathbf{g}(\mathbf{e}_\theta, \mathbf{e}_r) \equiv g_{\theta r} = 0$ y $\mathbf{g}(\mathbf{e}_\phi, \mathbf{e}_r) \equiv g_{\phi r} = 0$. Con estas consideraciones, el elemento de línea para un espacio-tiempo con simetría esférica toma la forma

$$ds^2 = g_{00}c^2 dt^2 + 2g_{0r}drdt + 2g_{0\theta}d\theta dt + 2g_{0\phi}d\phi dt + g_{rr}dr^2 + r^2 d\Omega^2 , \quad (3.18)$$

donde las $g_{\mu\nu}$ son funciones de las coordenadas. El siguiente paso para simplificar la situación es pedir que el espacio-tiempo sea también estático. Un espacio-tiempo con esta característica es aquel que admite una coordenada temporal que tenga las propiedades: (1) que sea estacionario, y (2) el espacio-tiempo es invariante ante la transformación de inversión temporal $t \rightarrow -t$. Debido a que estamos admitiendo estaticidad⁵, la segunda propiedad implica que $g_{0r} = g_{0\theta} = g_{0\phi} = 0$ para que el elemento de línea permanezca invariante. Además, la simetría esférica del espacio-tiempo y la primera propiedad permiten hacer $g_{00} := -A(r)$ y $g_{rr} := B(r)$ [22], de tal manera que el elemento (3.18) se simplifica a

$$ds^2 = -A(r)c^2 dt^2 + B(r)dr^2 + r^2 d\Omega^2 . \quad (3.19)$$

⁵Recordemos que no es lo mismo que un espacio-tiempo sea estático a que sea estacionario. En el segundo caso no se puede hacer una inversión temporal, y un ejemplo es el espacio-tiempo de Kerr [24].

Tomando en consideración la forma de ds^2 , la matriz asociada a la métrica en las coordenadas (t, r, θ, ϕ) se puede escribir como $(g_{\mu\nu}) = \text{diag}(-A(r), B(r), r^2, r^2 \sin^2 \theta)$. Si pedimos que sea en el vacío, $T_{\mu\nu} = 0$ y las ecuaciones de campo (3.16) sin constante cosmológica se vuelven

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 0. \quad (3.20)$$

Sacando la traza con respecto a $g^{\mu\nu}$ y considerando que $g_{\mu\nu}g^{\mu\nu} = 4$ con n la dimensión del espacio-tiempo, las ecuaciones en vacío son

$$0 = g^{\mu\nu}R_{\mu\nu} - \frac{1}{2}g^{\mu\nu}g_{\mu\nu}R = R - \frac{4}{2}R = -R \quad (3.21)$$

$$\Rightarrow R = 0 \quad \Rightarrow R_{\mu\nu} = 0. \quad (3.22)$$

Conociendo la forma de $(g_{\mu\nu})$, y sabiendo que se satisface (3.22), se tienen ecuaciones diferenciales en términos de las funciones $A(r)$ y $B(r)$. Al resolver dichas ecuaciones y aplicar las condiciones de frontera adecuadas se llega a las expresiones de $A(r)$ y $B(r)$. Para ver esto, primero notemos que el tensor de Ricci puede escribirse en componentes como

$$R_{\mu\nu} = \left(\Gamma^{\sigma}_{\mu\nu,\sigma} + \Gamma^{\sigma}_{\sigma\lambda}\Gamma^{\lambda}_{\mu\nu} - \Gamma^{\sigma}_{\sigma\nu,\mu} - \Gamma^{\sigma}_{\mu\lambda}\Gamma^{\lambda}_{\sigma\nu} \right), \quad (3.23)$$

por lo que, por ejemplo,

$$\begin{aligned} R_{00} &= \Gamma^{\sigma}_{00,\sigma} + \Gamma^{\sigma}_{\sigma\lambda}\Gamma^{\lambda}_{00} - \Gamma^{\sigma}_{\sigma 0,0} + \Gamma^{\sigma}_{0\lambda}\Gamma^{\lambda}_{\sigma 0} \\ &= \Gamma^r_{00,r} + \Gamma^{\sigma}_{\sigma r}\Gamma^r_{00} - 2\Gamma^0_{0\lambda}\Gamma^{\lambda}_{00} \\ &= \left(\frac{A'}{2B} \right)' + \left(\frac{A}{2A} + \frac{B'}{2B} + \frac{2}{r} \right) \left(\frac{A'}{2B} \right) - 2 \left(\frac{A'}{2A} \right) \left(\frac{A'}{2B} \right). \end{aligned} \quad (3.24)$$

Procediendo del mismo modo se llega a

$$R_{00} = \frac{A''}{2B} + \frac{A'}{rB} - \frac{A'}{4B} \left(\frac{A'}{A} + \frac{B'}{B} \right), \quad (3.25)$$

$$R_{rr} = -\frac{A''}{2A} + \frac{B'}{rB} + \frac{A'}{4A} \left(\frac{A'}{A} + \frac{B'}{B} \right), \quad (3.26)$$

$$R_{\theta\theta} = 1 - \frac{1}{B} - \frac{r}{2B} \left(\frac{A'}{A} - \frac{B'}{B} \right), \quad (3.27)$$

$$R_{\phi\phi} = \sin^2 \theta R_{\theta\theta}, \quad (3.28)$$

mientras que todos las demás componentes de R se anulan por los símbolos de Christoffel que involucran y la forma de $(g_{\mu\nu})$ [2]. Dadas las expresiones anteriores se tiene que

$$\begin{aligned}\frac{R_{00}}{A} &= \frac{A''}{2AB} + \frac{A'}{rAB} - \frac{A'}{4AB} \left(\frac{A'}{A} + \frac{B'}{B} \right), \\ \frac{R_{rr}}{B} &= -\frac{A''}{2AB} + \frac{B'}{rB^2} + \frac{A'}{4AB} \left(\frac{A'}{A} + \frac{B'}{B} \right) \\ \Rightarrow \frac{R_{00}}{A} + \frac{R_{rr}}{B} &= \frac{A'}{rAB} + \frac{B'}{rB^2} = \frac{1}{rB} \left(\frac{A'}{A} + \frac{B'}{B} \right).\end{aligned}$$

En particular, sabemos que $R_{\mu\nu} = 0$, por lo que $R_{00} = R_{rr} = 0$. Entonces

$$\frac{1}{rB} \left(\frac{A'}{A} + \frac{B'}{B} \right) = 0 \quad \Rightarrow \quad \frac{A'}{A} = -\frac{B'}{B}. \quad (3.29)$$

De esta manera, para $R_{\theta\theta}$ tenemos

$$R_{\theta\theta} = 1 - \frac{1}{B} - \frac{r}{2B} \left(-\frac{B'}{B} - \frac{B'}{B} \right) = 1 - \frac{1}{B} + \frac{rB'}{B^2} = 0 \quad (3.30)$$

$$\Rightarrow 1 = \frac{1}{B} - \frac{rB'}{B^2} = \frac{1}{B} + r \left(\frac{1}{B} \right)' := y(r) + ry'(r). \quad (3.31)$$

Resolviendo la ecuación diferencial (3.31) se obtiene que $y(r) = 1 + \frac{b}{r} = \frac{1}{B(r)}$ donde b es la constante de integración. Además, la condición $\frac{A'}{A} + \frac{B'}{B} = 0$ se puede reescribir como $A'B + B'A = 0$, la cual se resuelve si $AB = k$ con k una constante. Esto implica que $A(r) = k \left(1 + \frac{b}{r} \right)$. Finalmente, tenemos que el elemento de línea (3.19) debe tomar la forma de (3.13) cuando el campo gravitacional sea débil, por lo que se tiene que $b = -\frac{2GM}{c^2}$ y $k = 1$. Así, el elemento de línea de este sistema, llamado **solución de Schwarzschild**, tiene la forma

$$ds^2 = - \left(1 - \frac{2GM}{c^2 r} \right) c^2 dt^2 + \left(1 - \frac{2GM}{c^2 r} \right)^{-1} dr^2 + r^2 d\Omega^2. \quad (3.32)$$

La importancia del intervalo (3.32) radica en que si bien el intervalo (3.13) describe la situación de un campo gravitacional débil, la solución de Schwarzschild es válida para cualquier régimen gravitatorio, i.e., cualquier intensidad del campo gravitacional; es precisamente el caso de un campo intenso el que nos interesa estudiar en este trabajo.

3.3.1. Cantidades conservadas

Cuando se analiza un sistema, siempre es bueno buscar simetrías ya que éstas facilitan el análisis de su dinámica. En particular, un espacio-tiempo puede estar dotado de ciertas simetrías, las cuales, como se verá a continuación, dan origen a cantidades conservadas en el movimiento de partículas. Para ver esto, supongamos que una partícula se mueve a través de una geodésica, por lo que las componentes de su 4-momento satisfacen $p^\alpha p^\mu{}_{;\alpha} = 0$, pues \mathbf{p} es el vector tangente asociado (ver Sec. 2.2.5). Si bajamos el índice μ y expandimos la expresión utilizando la forma de la derivada covariante tenemos

$$0 = p^\alpha \left(p_{\mu,\alpha} - \Gamma^\beta_{\mu\alpha} p_\beta \right) = p^\alpha p_{\mu,\alpha} - \Gamma^\beta_{\mu\alpha} p_\beta p^\alpha \quad (3.33)$$

$$\Rightarrow p^\alpha p_{\mu,\alpha} = \Gamma^\beta_{\mu\alpha} p_\beta p^\alpha . \quad (3.34)$$

Analicemos primero el lado derecho de la Ec. (3.34). Si sustituimos la forma del símbolo de Christoffel se tiene

$$\begin{aligned} \Gamma^\beta_{\mu\alpha} p^\alpha p_\beta &= \frac{1}{2} g^{\beta\lambda} (g_{\lambda\mu,\alpha} + g_{\alpha\lambda,\mu} - g_{\mu\alpha,\lambda}) p^\alpha p_\beta \\ &= \frac{1}{2} (g_{\lambda\mu,\alpha} + g_{\alpha\lambda,\mu} - g_{\alpha\mu,\lambda}) p^\alpha p^\lambda . \end{aligned} \quad (3.35)$$

Notemos que tenemos una antisimetría bajo la permutación de los índices α y λ en el primer y tercer término dentro del paréntesis. Esto implica que al expandir la suma esos términos se van a cancelar, resultando en

$$\Gamma^\beta_{\mu\alpha} p^\alpha p_\beta = \frac{1}{2} g_{\alpha\lambda,\mu} p^\alpha p^\lambda \quad \Rightarrow \quad p^\alpha p_{\mu,\alpha} = \frac{1}{2} g_{\alpha\lambda,\mu} p^\alpha p^\lambda . \quad (3.36)$$

La importancia del resultado anterior radica en que la forma de las componentes de la métrica y su dependencia con respecto a ciertas coordenadas juega un papel en las derivadas de las componentes del 4-momento. Para poder apreciar esto de una manera más clara notemos que $p^\alpha = mu^\alpha = m(dx^\alpha/d\tau)$ para partículas masivas de masa m , por lo que

$$m \frac{dx^\alpha}{d\tau} \frac{\partial}{\partial x^\alpha} p_\mu = m \frac{d}{d\tau} p_\mu = \frac{1}{2} g_{\alpha\lambda,\mu} p^\alpha p^\lambda . \quad (3.37)$$

Si el espacio-tiempo puede ser descrito por un sistema coordenado en el cual la métrica no dependa de la coordenada x^μ , entonces $g_{\alpha\lambda,\mu} = 0$ y $p_\mu = \text{cte}$. Es importante resaltar que se tiene

una dependencia del sistema coordenado elegido para poder ver las simetrías, pues no existe en general una conservación global para cualquier observador con un sistema de referencia arbitrario.

Para ver las simetrías y cantidades conservadas en acción, regresemos al caso de la solución de Schwarzschild. En este caso, el elemento de línea (3.32) describe un espacio-tiempo en simetría esférica en unas coordenadas (t, r, θ, ϕ) , donde se puede apreciar que los elementos de la métrica no dependen de las coordenadas t y ϕ . Así, en el movimiento a través de una geodésica se conservan las cantidades $p_0 \equiv p_t$ y p_ϕ dadas por

$$p_0 = g_{00}p^0 = - \left(1 - \frac{2GM}{c^2 r} \right) mc \frac{dt}{d\tau} := -m \frac{e}{c} = \text{cte} , \quad (3.38)$$

$$p_\phi = g_{\phi\phi}p^\phi = r^2 \sin^2 \theta m \frac{d\phi}{d\tau} = r^2 m \frac{d\phi}{d\tau} := ml = \text{cte} , \quad (3.39)$$

que corresponden a la energía y el momento angular, respectivamente; las cantidades e y l son por unidad de masa. Además, se utilizó el hecho de que la simetría esférica permite escoger el plano ecuatorial $\theta = \pi/2$ sin pérdida de generalidad, como se verá a continuación.

3.3.2. Trayectorias de partículas

Recordando el objetivo de nuestro trabajo, el siguiente paso es encontrar la forma de describir de manera matemática el movimiento de la estrella S2 alrededor de Sgr A*. Para esto, vamos a decir de manera explícita que consideraremos a S2 como una partícula de prueba que sigue una geodésica tipo tiempo dictada por el campo gravitacional de Sgr A*, el cual tomaremos que está dado por la solución de Schwarzschild. Cabe resaltar que Sgr A* puede tener rotación, pero en este análisis se está ignorando.

El espacio-tiempo de Schwarzschild es poco general por todas las simetrías que tiene; sin embargo, es esto mismo lo que simplifica el estudio de las trayectorias en éste. En el caso general se tienen que integrar cada una de ecuaciones de las geodésicas, Ec. (2.68), de cada coordenada, pero en este caso particular se cuenta con el número de cantidades conservadas suficiente para volver el problema completamente integrable; esto quiere decir que se tienen cuatro cantidades conservadas y son cuatro las funciones coordenadas que se requieren.

Lo primero que veremos es que las órbitas en este espacio-tiempo permanecen en un plano. Debido a la simetría esférica siempre se puede hacer una rotación para que la partícula comience en el plano $\theta = \pi/2$, por lo que se tienen las condiciones iniciales $(\tau_0, r_0, \theta_0 = \pi/2, \phi_0 = 0)$. Además, uno puede hacer otra rotación que preserve la posición del punto inicial de la partícula, por lo que hacemos $u^\theta(\tau_0) = 0$, de tal forma que la partícula comience con una velocidad inicial tangencial al

plano ecuatorial. La ecuación de la geodésica para la coordenada θ es

$$\ddot{\theta} + \Gamma^{\theta}_{\mu\nu} \dot{x}^{\mu} \dot{x}^{\nu} = 0 . \quad (3.40)$$

Notemos que si μ o ν son igual a θ , la condición inicial que impusimos hace que el segundo término se anule por la velocidad inicial. Por otro lado, si μ y ν son distintos de θ , entonces los símbolos de Christoffel se anulan. Esto demuestra que siempre se tiene $\ddot{\theta} = 0$, y como la velocidad inicial es tangencial al plano, la partícula siempre permanece en el plano ecuatorial. Este hecho también se sigue de la conservación del momento angular [16].

El hecho de que las trayectorias permanezcan en un plano permite reducir la dimensionalidad del problema a un espacio-tiempo de dimensión 3 con elemento de línea

$$ds^2 = - \left(1 - \frac{r_S}{r}\right) c^2 dt^2 + \left(1 - \frac{r_S}{r}\right)^{-1} dr^2 + r^2 d\phi^2 , \quad (3.41)$$

donde hemos definido el **radio de Schwarzschild** como $r_S := 2GM/c^2$. La última constante para hacer el problema completamente integrable es la normalización de la 4-velocidad

$$u_{\mu} u^{\mu} = g_{\mu\nu} u^{\mu} u^{\nu} = -c^2 . \quad (3.42)$$

Con esta constante y la forma de la métrica se tiene

$$- \left(1 - \frac{2GM}{c^2 r}\right) c^2 \dot{t}^2 + \left(1 - \frac{2GM}{c^2 r}\right)^{-1} \dot{r}^2 + r^2 \dot{\phi}^2 = -c^2 , \quad (3.43)$$

donde las derivadas denotadas por punto son con respecto al tiempo propio τ . Utilizando las cantidades conservadas (3.38) y (3.39) se llega a

$$\begin{aligned} -c^2 &= - \left(1 - \frac{2GM}{c^2 r}\right) c^2 \left[\frac{e^2}{c^4} \left(1 - \frac{2GM}{c^2 r}\right)^{-2} \right] + \left(1 - \frac{2GM}{c^2 r}\right)^{-1} \dot{r}^2 + r^2 \left(\frac{l}{r^2}\right)^2 \\ &= - \left(1 - \frac{2GM}{c^2 r}\right)^{-1} \frac{e^2}{c^2} + \left(1 - \frac{2GM}{c^2 r}\right)^{-1} \dot{r}^2 + \frac{l^2}{r^2} \\ &\Rightarrow \frac{e^2}{c^2} = c^2 - \frac{2GM}{r} + \dot{r}^2 + \frac{l^2}{r^2} \left(1 - \frac{2GM}{c^2 r}\right) \\ &\Rightarrow \frac{e^2/c^2 - c^2}{2} = \frac{1}{2} \dot{r}^2 + \frac{1}{2} \left[-\frac{2GM}{r} + \frac{l^2}{r^2} - \frac{2GMl^2}{c^2 r^3} \right] . \end{aligned}$$

Como e es constante, se puede definir una nueva constante $\mathcal{E} := \frac{e^2/c^2 - c^2}{2}$ y un potencial efectivo $V_{eff}(r)$ como

$$V_{eff}(r) := -\frac{GM}{r} + \frac{l^2}{2r^2} - \frac{GMl^2}{c^2 r^3}. \quad (3.44)$$

Con esto se llega a una ecuación que se asemeja a la de movimiento radial

$$\mathcal{E} = \frac{1}{2}\dot{r}^2 + V_{eff}(r). \quad (3.45)$$

Los primeros dos términos de (3.44) representan la atracción gravitacional y la repulsión centrífuga, respectivamente, y ambos aparecen en el caso Newtoniano de la gravedad; no obstante, el tercer término es nuevo y corresponde a la corrección relativista. La principal consecuencia de éste es que bajo ciertas condiciones la partícula caerá a la fuente del potencial gravitacional incluso si tiene momento angular. Esta situación no sucede en el caso newtoniano ya que en éste siempre se puede contrarrestar la atracción gravitacional si se tiene el momento angular suficiente [2].

De la forma del potencial efectivo (3.44) se puede observar que los parámetros que modifican la barrera de potencial son únicamente M y l , y en particular el cociente cl/GM . En la Fig. 3.3 se muestran gráficas de V_{eff} para diferentes valores de l/M (Se toman unidades naturales, i.e., $c = G = 1$). Además, se muestra el caso Newtoniano, donde se puede apreciar que en éste se tiene una barrera de potencial infinitamente alta, mientras que en el caso relativista la barrera tiende a $-\infty$ con $r \rightarrow 0$; ésta es la consecuencia del término de la corrección relativista.

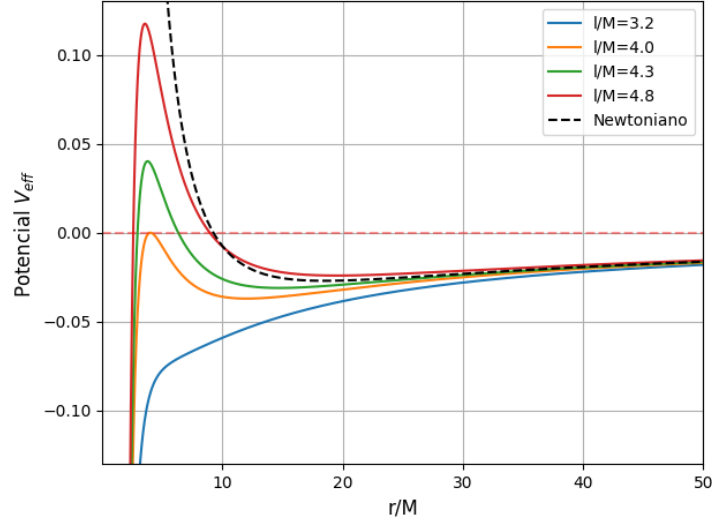


Figura 3.3: Potenciales efectivos para distintos valores del cociente l/M ; además, se presenta el potencial Newtoniano con $l/M = 4.3$. El caso relativista presenta una barrera de altura finita, mientras que el caso clásico tiene una barrera infinita. Cabe resaltar que las gráficas están con unidades naturales ($G = c = 1$).

Notemos que el término \dot{r}^2 siempre es mayor o igual a cero, por lo que de la ecuación de movimiento radial, Ec. (3.45), se tiene que $\mathcal{E} - V_{eff}$ también debe ser siempre mayor o igual a cero. Esto implica que sólo están permitidas trayectorias en las cuales la energía satisface $\mathcal{E} \geq V_{eff}$. Esto da origen a cuatro tipos de trayectorias en este espacio-tiempo:

1. Si $\mathcal{E} > V_{eff}(r) \forall r$, entonces una partícula que viene de $r = \infty$ va a colisionar con la fuente gravitacional en $r = 0$;
2. Si $\mathcal{E} \geq V_{eff}(r)$ para un cierto rango $[r_1, r_2]$, la partícula describe una órbita acotada entre r_1 y r_2 , llamados puntos de retorno y se tiene $V_{eff}(r_1) = V_{eff}(r_2) = \mathcal{E}$;
3. Si $\mathcal{E} \geq V_{eff}(r)$ en $[r_{min}, \infty)$, la partícula llega a un punto de máximo acercamiento, r_{min} con $V_{eff}(r_{min}) = \mathcal{E}$, y después se aleja arbitrariamente;
4. Si $\mathcal{E} = V_{eff}(r)$ para algún r , y es mínimo o máximo local, se tiene una órbita circular, la cual puede ser estable o inestable, respectivamente.

Es importante estudiar y graficar los casos 2 y 4 porque recordemos que queremos analizar la órbita de la estrella S2, la cual sabemos por los datos que no pertenece a los casos 1 ni 3. A

continuación, la Fig. 3.4 muestra los potenciales y órbitas descritas por una partícula en dichos casos.

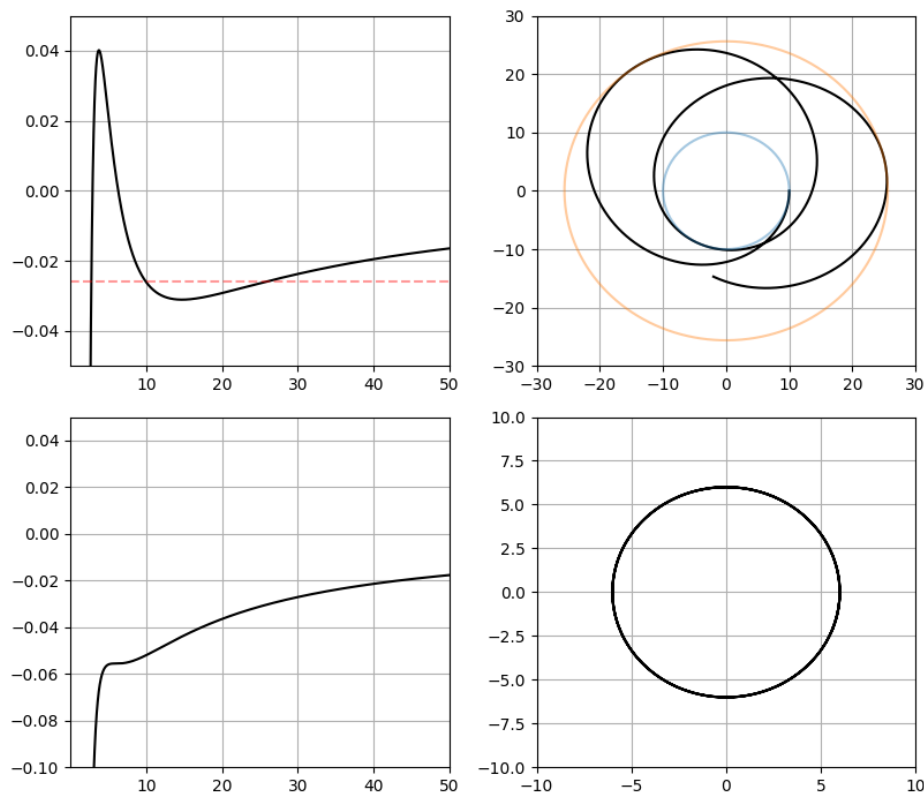


Figura 3.4: Representaciones de los potenciales y órbitas para los casos 2 y 4. La fila superior muestra primero el potencial efectivo para el caso $l/M = 4.3$ y $\mathcal{E} = -0.026$ y a la derecha la órbita descrita por una partícula en dichas condiciones; se puede observar que la partícula se mantiene en una órbita (en negro) acotada por los puntos de retorno $r_{min} = 10M$ (en azul) y $r_{max} = 25.6M$ (en naranja). Por otro lado, la segunda fila representa el caso del potencial de una órbita circular estable localizada en $r = 6M$ cuando $l/M = \sqrt{12}$.

En un principio, las cantidades conservadas (3.38), (3.39) y la ecuación (3.45) permiten describir el movimiento de una partícula a través del espacio-tiempo al resolver las ecuaciones diferenciales. Sin embargo, como nuestro objetivo es darle ecuaciones diferenciales a la computadora para obtener la trayectoria de la estrella S2, es más conveniente trabajar con una forma equivalente para el tiempo coordenado y la coordenada radial. Primero, definamos $N := 1 - 2GM/c^2r$, por lo que la

normalización de la 4-velocidad es

$$-c^2 = -Nc^2\dot{t}^2 + N^{-1}\dot{r}^2 + r^2\dot{\phi}^2 \quad (3.46)$$

$$= -Nc^2\dot{t}^2 + N^{-1}\dot{r}^2 + \frac{l^2}{r^2} \quad (3.47)$$

$$\Rightarrow 1 = N\dot{t}^2 - N^{-1}\frac{\dot{r}^2}{c^2} - \frac{l^2}{c^2r^2} \quad (3.48)$$

$$\Rightarrow N\dot{t}^2 = 1 + N^{-1}\frac{\dot{r}^2}{c^2} + \frac{l^2}{c^2r^2} \quad (3.49)$$

$$\Rightarrow \dot{t}^2 = N^{-1} + N^{-2}\frac{\dot{r}^2}{c^2} + N^{-1}\frac{l^2}{c^2r^2} \quad (3.50)$$

$$\Rightarrow \dot{t} = \left\{ \frac{1}{\left(1 - \frac{2GM}{c^2r}\right)} + \frac{\dot{r}^2}{c^2\left(1 - \frac{2GM}{c^2r}\right)^2} + \frac{l^2}{c^2r^2\left(1 - \frac{2GM}{c^2r}\right)} \right\}^{1/2}. \quad (3.51)$$

La razón por la cual escribimos cada uno de los términos de manera explícita es porque ésta es una de las ecuaciones que se implementarán en el código (ver Sec. 5.1.1). De esta manera, la Ec. (3.51) permite establecer la relación entre el tiempo propio y el tiempo coordinado en función de los parámetros del sistema (M, l), y las posiciones y velocidades de la partícula. Es importante mencionar que este desarrollo se puede generalizar a cualquier espacio-tiempo estacionario, pues solamente se requieren las componentes $g_{\mu\nu}$ al utilizar la normalización de la 4-velocidad. Por otro lado, se puede obtener una ecuación diferencial de segundo orden en r al derivar la Ec. (3.45) y usar la conservación de la energía ($d\mathcal{E}/d\tau = 0$) como

$$\begin{aligned} 0 &= \frac{1}{2} \frac{d}{d\tau} \dot{r}^2 + \frac{d}{d\tau} V_{eff}(r) \\ &= \frac{1}{2} \frac{d}{d\tau} \dot{r}^2 + \frac{d}{d\tau} \left(-\frac{GM}{r} + \frac{l^2}{2r^2} - \frac{GMl^2}{c^2r^3} \right) \\ &= \frac{1}{2} (2\dot{r}\ddot{r}) + \frac{GM}{r^2} \dot{r} - \frac{l^2}{r^3} \dot{r} + \frac{3GMl^2}{c^2r^4} \dot{r} \\ &= \dot{r}\ddot{r} + \dot{r} \left(\frac{GM}{r^2} - \frac{l^2}{r^3} + \frac{3GMl^2}{c^2r^4} \right) \\ &= \dot{r} \left(\ddot{r} + \frac{GM}{r^2} - \frac{l^2}{r^3} + \frac{3GMl^2}{c^2r^4} \right). \end{aligned} \quad (3.52)$$

Notemos que al obtener \ddot{r} estamos quitando la energía \mathcal{E} de los parámetros necesarios para describir la órbita⁶. El caso $\dot{r} = 0$ se satisface en los puntos de retorno de la órbita, mientras que en todos los demás se tiene que la segunda derivada debe satisfacer

$$\ddot{r} = -\frac{GM}{r^2} + \frac{l^2}{r^3} - \frac{3GMl^2}{c^2 r^4}. \quad (3.53)$$

Así, obtenemos la ecuación diferencial de segundo orden en r que rige el movimiento de las órbitas. Obtener esta ecuación resulta más conveniente por las condiciones iniciales que se conocen en el problema de la estrella S2 (para más detalles al respecto y la explicación de por qué resulta éste un mejor método véase la Sec. 5.1.1).

3.4. Agujeros negros

3.4.1. Singularidades

Al observar la forma del elemento de línea de Schwarzschild, Ec. (3.32), podemos notar que existen dos puntos singulares: $r = 2GM/c^2$ y $r = 0$; en ambos algún elemento de la métrica se anula o diverge, ya sea g_{rr} o g_{tt} . Estos dos puntos reciben el nombre de **singularidades**, y existen dos tipos: de coordenadas y espacio-temporales o geométricas. Las primeras se deben simplemente a una mala elección de coordenadas, por lo que el espacio-tiempo es completamente regular ahí; por otro lado, las segundas no son removibles en ningún sistema coordenado e indican que realmente la geometría del espacio-tiempo es singular. Como veremos a continuación, la métrica de Schwarzschild dada como en la Ec. (3.32) cuenta con una singularidad de coordenadas y una espacio-temporal.

Un excelente ejemplo para entender las singularidades de coordenadas es precisamente el espacio Euclídeo en coordenadas esféricas. Si pensamos en una 2-esfera y utilizamos coordenadas esféricas para describirla, notaremos que el polo norte es un punto singular. Dada una esfera de radio r , los puntos (θ, ϕ) permiten describir los diferentes puntos sobre ella; no obstante, la familia de puntos $(\theta = 0, 0 \leq \phi \leq 2\pi)$ describen todos el mismo punto: el polo norte. Esto se nota, por ejemplo, en los mapas de la Tierra porque tanto el polo norte como el polo sur se convierten cada uno en una línea, cosa que no es cierta de la geometría de la esfera. Así, tenemos un sistema coordenado que no logra describir perfectamente todos los puntos de la esfera, teniendo singularidades de coordenadas como el polo norte, pero el polo norte en sí mismo es un punto regular de ésta.

⁶Este hecho será de suma importancia cuando hablemos de la integración de la órbita de la estrella S2 y los parámetros que necesitamos para describirla.

Comencemos estudiando el punto $r = 2GM/c^2$. Para darnos una idea de la naturaleza de la singularidad podemos estudiar cantidades físicas de observadores que se acerquen a dicho punto. Algo que podemos investigar es el tiempo propio que le tomaría a un observador en caída radial llegar desde un punto $r = R > 2GM/c^2$ hasta $r = 2GM/c^2$. Primero, notemos que la distancia coordenada es una distancia finita, la cual está dada por

$$\Delta r = \int_R^{2GM/c^2} g_{rr} dr = \int_R^{2GM/c^2} \frac{dr}{\left(1 - \frac{2GM}{c^2 r}\right)}. \quad (3.54)$$

La integral dada por (3.54) es finita para cualquier valor de R [18], por lo que esperaríamos que al observador le tome un tiempo propio finito llegar a $2GM/c^2$. Para poder calcular el tiempo $\Delta\tau$ recordemos que la Ec.(3.45) relaciona $d\tau$ y dr . Suponiendo que la partícula cae de forma radial, i.e., no hay momento angular, la expresión para el $d\tau$ es

$$d\tau = -\frac{dr}{\left[2\left(\mathcal{E} + \frac{GM}{r}\right)\right]^{1/2}} \Rightarrow \Delta\tau = -\int \frac{dr}{\left[2\left(\mathcal{E} + \frac{GM}{r}\right)\right]^{1/2}}, \quad (3.55)$$

donde se escoge el signo $(-)$ para indicar que el observador cae. Notemos que al integrar entre R y $2GM/c^2$ para obtener $\Delta\tau$ el valor se mantiene finito [22]; dicho en otras palabras, a un observador siempre le toma un tiempo propio finito llegar a $2GM/c^2$. Incluso uno podría integrar hasta cualquier $0 < r < 2GM/c^2$ y la integral seguiría dando un valor finito. Esto nos indica que el espacio-tiempo parece permanecer regular.

Lo que falta es analizar el comportamiento del tiempo coordenado en la misma situación, es decir, cómo percibiría un observador al infinito la caída radial de una partícula. Lo que necesitamos ahora es relacionar dt con dr . Una manera de hacerlo es ver que $U^0 = g^{00}U_0 = -g^{00}\mathcal{E}/c$ nos conecta $d\tau$ y dt , y como ya vimos la relación entre $d\tau$ y dr , entonces podemos obtener una ecuación que nos relacione dr y dt . Como $U^0 = dt/d\tau$, esta discusión nos permite escribir

$$\frac{dt}{d\tau} = -\frac{\mathcal{E}/c}{\left(1 - \frac{2GM}{c^2 r}\right)} \Rightarrow dt = -\frac{d\tau \mathcal{E}/c}{\left(1 - \frac{2GM}{c^2 r}\right)}, \quad (3.56)$$

y como ya vimos la expresión de $d\tau$ en términos de dr finalmente obtenemos

$$dt = \left(-\frac{\mathcal{E}/c}{1 - \frac{2GM}{c^2 r}}\right) \frac{dr}{\left[2\left(\mathcal{E} + \frac{GM}{r}\right)\right]^{1/2}}. \quad (3.57)$$

Analizando el comportamiento de la integral de la Ec. (3.57) uno puede observar que el valor diverge conforme $r \rightarrow 2GM/c^2$ [22]. Si bien vimos que el tiempo propio es finito, el resultado anterior nos indica que el tiempo coordenado diverge conforme la partícula se acerca a $r = 2GM/c^2$. De esta manera, existe una discrepancia entre lo que perciben un observador en caída libre y uno en el infinito. Aunque esto parece alarmante, recordemos que las coordenadas (t, r, θ, ϕ) fueron escogidas para describir bien el espacio-tiempo lejos, por lo que este fallo al intentar describir $r = 2GM/c^2$ nos muestra una limitante de las coordenadas mismas, pero ¿de dónde proviene el infinito en el tiempo coordenado? Para responder esta pregunta analicemos los conos de luz de las trayectorias radiales.

Al considerar trayectorias nulas y radiales estamos pidiendo que se satisfaga $ds^2 = 0$ y $d\theta = d\phi = 0$. Así, el elemento de línea se vuelve

$$0 = - \left(1 - \frac{2GM}{c^2 r} \right) dt^2 + \frac{dr^2}{1 - \frac{2GM}{c^2 r}} \quad (3.58)$$

$$\Rightarrow \frac{dt}{dr} = \pm \left(\frac{1}{1 - \frac{2GM}{c^2 r}} \right) . \quad (3.59)$$

En la Ec. (3.59) obtenemos un (\pm) . El signo $(+)$ indica que conforme avanza el tiempo la coordenada r aumenta, es decir, se tienen trayectorias salientes; por otro lado, si se tiene $(-)$, cuando t avanza la coordenada r disminuye, indicando que se tienen trayectorias entrantes. El comportamiento de los conos de luz se puede apreciar en el diagrama $t - r$ de la Fig. 3.5.

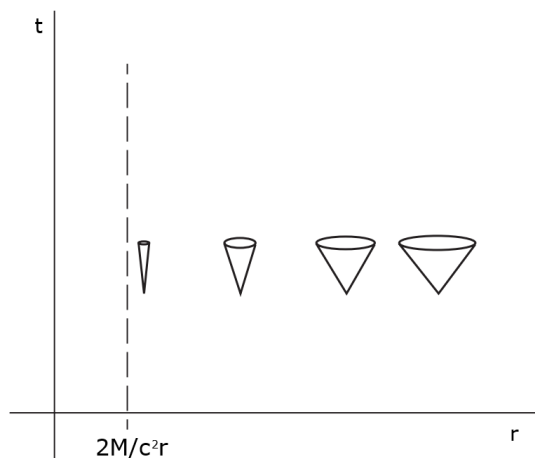


Figura 3.5: Diagrama $t - r$ que muestra los conos de luz de partículas en caída radial hacia $r = 2GM/c^2$. Se puede observar que al acercarse a dicho punto los conos de luz se comienzan a cerrar, implicando que cada vez les toma más tiempo coordenado caer. Imagen tomada de [22].

En la Fig. 3.5 podemos apreciar que muy lejos de $r = 2GM/c^2$ los conos de luz se abren en 45° , justo como en el espacio-tiempo plano de Minkowski; esto es totalmente normal porque tenemos un espacio-tiempo asintóticamente plano. Sin embargo, conforme nos acercamos a $r = 2GM/c^2$ los conos de luz comienzan a cerrarse. Esto se debe a que la pendiente de los conos está dada por el lado derecho de la Ec. (3.59), la cual diverge si $r \rightarrow 2GM/c^2$. Que los conos de luz se cierren se puede interpretar como que a la partícula en caída libre le toma cada vez más tiempo coordinado hacer que su coordenada radial disminuya; este es el origen de la divergencia discutida en la Ec. (3.57). Finalmente, en el límite cuando $r = 2GM/c^2$ los conos de luz se cierran completamente y las trayectorias nulas quedan atrapadas, es decir, sin disminuir ni aumentar su coordenada radial.

Toda esta discusión nos indica que realmente se tiene una singularidad de coordenadas en dicho punto; no obstante, falta determinar qué pasa en la otra singularidad de la métrica, i.e., en $r = 0$. Para ver esto, podemos estudiar si es una singularidad originada por la elección de coordenadas o si no puede ser removida en ningún marco de referencia. Esto se logra viendo el comportamiento de alguna cantidad que sea invariante de coordenadas. La invariante que analizaremos esta dada por la contracción del tensor de Riemann consigo mismo, llamada **invariante** o **escalar de Kretschmann**. Esta cantidad está dada como [25]

$$K := R_{\alpha\beta\mu\nu}R^{\alpha\beta\mu\nu} = \frac{48G^2M^2}{c^4r^6}. \quad (3.60)$$

Al ser un escalar, su valor no va a cambiar cuando cambiemos de sistema coordinado; esto permite ver propiedades independientemente del marco de referencia que se elija. Como podemos ver de la Ec. (3.60), el escalar de Kretschmann diverge con $r \rightarrow 0$; esto nos indica que la singularidad $r = 0$ es en realidad una singularidad espacio-temporal. Además, de la misma expresión podemos corroborar que $r = 2GM/c^2$ no es una singularidad espacio-temporal, pues el escalar no diverge ahí.

3.4.2. Horizonte de eventos

Anteriormente, vimos que la métrica de Schwarzschild tiene dos singularidades: $r = 2GM/c^2$ y $r = 0$, siendo la primera de coordenadas y la segunda espacio-temporal. Como sabemos que se tiene una singularidad coordinada, el siguiente paso sería encontrar un sistema coordinado en el cual se pueda remover. Para esto existen dos cambios de coordenadas comunes: las coordenadas de Eddington-Finkelstein y las coordenadas de Kruskal-Szekeres⁷.

⁷Estos cambios de coordenadas no se discutirán a detalle en este trabajo. Sin embargo, para un estudio más profundo de las mismas se recomiendan las siguientes referencias: [2, 18, 19, 22, 25].

En el elemento de línea (3.32) podemos observar que la componente g_{rr} de la métrica se indetermina en la singularidad de coordenadas $r = 2GM/c^2$. Vamos ahora a definir una nueva coordenada, v , que nos va a permitir deshacernos de la singularidad. Definamos

$$v := t + f(r) \Rightarrow dv = dt + f' dr . \quad (3.61)$$

De esta manera,

$$\begin{aligned} \left(1 - \frac{2GM}{c^2 r}\right) dt^2 &= \left(1 - \frac{2GM}{c^2 r}\right) (dv - f' dr)^2 \\ &= \left(1 - \frac{2GM}{c^2 r}\right) (dv^2 - 2f' dvdr + (f')^2 dr^2) . \end{aligned} \quad (3.62)$$

Sustituyendo en la métrica y definiendo $N := 2GM/c^2 r$ tenemos

$$\begin{aligned} ds^2 &= -N(dv^2 - 2f' dvdr + (f')^2 dr^2) + \frac{dr^2}{N} + r^2 d\Omega^2 \\ &= -Ndv^2 + 2Nf' dvdr + dr^2(-N(f')^2 + \frac{1}{N}) + r^2 d\Omega^2 , \end{aligned} \quad (3.63)$$

donde vemos que se elimina la singularidad si se satisface

$$N(f')^2 = \frac{1}{N} . \quad (3.64)$$

Debido a la potencia cuadrática en f' se tienen dos opciones de signo. Escogiendo el signo positivo,

$$f' = \frac{1}{1 - 2GM/c^2 r} = 1 + \frac{2GM/c^2 r}{r - 2GM/c^2 r} \quad (3.65)$$

$$\Rightarrow v = t + r + \frac{2GM}{c^2 r} \ln \left(\frac{r - 2GM/c^2 r}{2GM/c^2 r} \right) . \quad (3.66)$$

Como al eliminar la singularidad al remover dr^2 surge un término $dvdr$, observamos que al hacer este cambio de coordenadas terminamos con una métrica que ya no es diagonal. No obstante, la ventaja de lo anterior es que ahora todos los coeficientes de ésta están bien definidos. Además,

$$\det g = -r^4 \sin^2 \theta , \quad (3.67)$$

que es diferente de cero para toda $r > 0$, excepto en los polos ($\theta = 0, \pi$), que son las singularidades usuales de las coordenadas esféricas. Debido a esto, sabemos que la signatura de la métrica no cambia al cruzar $r = 2GM/c^2$, por lo que tenemos que g es una métrica Lorentziana bien definida en el conjunto $\{v \in \mathbb{R}, r \in (0, \infty)\} \times \mathbb{S}^2$ [25]. Cabe remarcar que, si bien al principio dijimos que íbamos a realizar un simple cambio de coordenadas, lo que hicimos fue una extensión analítica del espacio-tiempo original. A pesar de que en este trabajo no se discutirán las coordenadas de Kruskal-Szekeres, éstas también representan una extensión analítica del espacio-tiempo de Schwarzschild.

Escribiendo el elemento de línea de manera explícita tenemos

$$ds^2 = - \left(1 - \frac{2GM}{c^2 r} \right) dv^2 + 2dvdr + r^2 d\Omega^2, \quad (3.68)$$

con lo cual podemos analizar las trayectorias nulas radiales, es decir, el caso cuando $ds^2 = 0$ y $d\Omega^2 = 0$. Aquí se tienen dos soluciones [18]:

$$\frac{dv}{dr} = 0, \quad (3.69)$$

$$\frac{dv}{dr} = \frac{2}{1 - 2GM/c^2 r}. \quad (3.70)$$

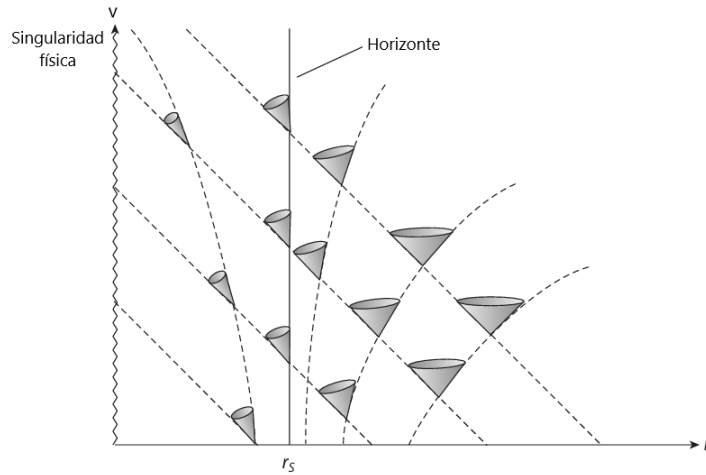


Figura 3.6: Diagrama de las trayectorias nulas radiales en el plano (v, r) . Se puede observar que lejos de r_s los conos de luz se abren a 45° , pero conforme se acercan comienzan a inclinarse. Imagen tomada de [2].

La Fig. 3.6 muestra la gráfica de los casos anteriores mostrando los conos de luz. Podemos observar que lejos de $2GM/c^2$ los conos de luz se abren a 45° , justo como pasa en el espacio-tiempo plano; sin embargo, conforme la partícula comienza a acercarse a esta superficie, el cono de luz se “inclina”. En $r = 2GM/c^2$ el cono se ha inclinado de tal forma que una trayectoria nula se dirige a la singularidad espacio-temporal y la otra está justo sobre el radio de Schwarzschild, por lo que el futuro de cualquier partícula masiva queda dentro de esta superficie. El hecho de que los conos de luz se inclinen de esta forma hace que cualquier partícula con masa o sin masa que cruce el radio de Schwarzschild esté destinada a caer a la singularidad; dicho de otra forma, cualquier partícula que cruce el radio de Schwarzschild desde r mayores ya no podrá salir. Así, la región $\{r \leq 2GM/c^2\}$ para la métrica (3.68) puede ser definida como una región de un **agujero negro** en el siguiente sentido [25]:

Es una región la cual un observador o señal puede cruzar de afuera hacia dentro, pero de la cual no puede salir nunca.

Este hecho implica que el espacio-tiempo tiene una región que está causalmente desconectada del resto, pues toda curva tipo tiempo o nula que esté dentro de ésta no puede escapar a las demás regiones del espacio-tiempo (fuera del agujero negro). Como ya se había comentado que un observador en caída radial alcanza la singularidad en un tiempo propio finito, y recordando lo visto sobre geodésicas en la Subsec. 2.2.5, podemos decir que su geodésica es maximal pero no completa. Así, el espacio-tiempo se vuelve una variedad geodésicamente incompleta.

Capítulo 4

Métodos MCMC

En el Cap. 3 se dieron las bases de la Relatividad General y, más en concreto, se habló de las ecuaciones que describen el movimiento de partículas alrededor de un sistema estático y con simetría esférica. Dichas ecuaciones son suficientes para simular las órbitas de las partículas y poder estudiar sus propiedades, como simular la órbita de la estrella S2; sin embargo, el caso es más complejo cuando se busca ajustar órbitas estelares a datos observacionales. En este capítulo se desarrollan los principios de la probabilidad y se habla de la estadística Bayesiana como método de inferencia para ajustar modelos a conjuntos de datos. Más adelante, en la Sec. 4.2, se da una breve introducción a los Métodos de Monte Carlo como métodos que usan variables aleatorias para resolver problemas. Luego, se introduce el concepto de Cadenas de Markov como procesos estocásticos incluyendo un ejemplo que permite entender de manera clara sus principios. Finalmente, se muestra cómo la unión de los Métodos de Monte Carlo y las Cadenas de Markov dan origen a los métodos MCMC; en particular, se habla del uso del algoritmo de Random Walk Metropolis-Hastings (RWMH) para el caso de órbitas estelares. Los conceptos aquí mencionados se irán definiendo a lo largo de todo este capítulo. La importancia de desarrollar estas herramientas se verá en el Cap. 5 cuando se hable en concreto de su aplicación para el ajuste de los datos observacionales de la órbita de S2.

4.1. Estadística Bayesiana

Supongamos que contamos con un conjunto de datos D de un sistema y queremos proponer un modelo M que ajuste a dichos datos. En principio, el modelo podría depender de un conjunto de parámetros, por lo que $M = M(\theta)$, con θ un vector de parámetros que vive en el espacio de

parámetros \mathcal{X} . En este sentido, cada punto en \mathcal{X} representa un conjunto particular de valores para los distintos parámetros, y cada $\theta \in \mathcal{X}$ va a generar cierta información con respecto al sistema.

Un ejemplo de lo anterior, y que de hecho sirve de antesala para el problema de la órbita de S2, sería la ecuación de un círculo en \mathbb{R}^2 ,

$$(x - h)^2 + (y - k)^2 = R^2 . \quad (4.1)$$

En este caso, nuestro modelo cuenta con tres parámetros: dos puntos espaciales para localizar el centro del círculo (h, k) , y su radio R ; por lo que la dimensión de \mathcal{X} es 3, y $\theta = (h, k, R) \in \mathcal{X}$. El conjunto de datos, D , puede representar puntos en \mathbb{R}^2 que pertenecen a un círculo cuya ecuación se desconoce.

Si nosotros proponemos un cierto θ , podemos comparar el círculo generado con los datos a ajustar mediante una prueba de tipo χ^2 para determinar qué tan cercanos son los puntos del círculo generado a los datos. Para el caso de dos conjuntos de datos esto se puede escribir como

$$\text{Likelihood} = -\frac{1}{2} \sum_i \left[\left(\frac{x - \bar{x}_i}{\sigma_{x_i}} \right)^2 + \left(\frac{y - \bar{y}_i}{\sigma_{y_i}} \right)^2 \right] , \quad (4.2)$$

donde $\sigma_{x_i}, \sigma_{y_i}$ representan la incertidumbre de los datos, y los puntos x, y son los puntos generados por nuestro modelo dado el vector θ correspondiente; en este caso $\bar{x}_i, \bar{y}_i \in D$. Es importante notar que la suma va sobre los datos, ya que en principio por cada dato corresponde un punto, entonces la suma está bien definida¹. La Ec. (4.2) representa una medida para saber qué tan bueno es el conjunto de valores para reproducir los datos observados, y notemos que la diferencia entre punto y dato está pesada por el error del segundo, haciendo que un mejor modelo sea aquel que ajusta mejor a los datos que tienen menor incertidumbre.

Sin embargo, la Ec. (4.2) representa una medida individual para ese modelo particular, pero muchas veces lo que se quiere es conocer la distribución que tienen los valores de los parámetros para determinar en qué rango se encuentran los valores que mejor reproducen los datos. El problema de determinar la probabilidad de que una hipótesis sea cierta dado un conjunto de datos es el origen de la **estadística Bayesiana**.

Para entender la estadística Bayesiana se tienen que entender primero los fundamentos de la probabilidad; estos pueden ser descritos por medio de teoría de conjuntos y determinados axiomas. Definamos el conjunto de muestras S como los posibles valores que puede tomar una variable. Sea

¹No siempre es éste el caso, y, de hecho, no lo es para la caso de la órbita de S2. Para arreglarlo tendremos que introducir en el siguiente capítulo un mecanismo para tener el mismo número de puntos que de datos observacionales.

$A \subset S$ un subconjunto de S . Entonces a cada $A \subset S$ se le asigna un valor real $P(A)$, llamada **probabilidad** de A , que satisface:

1. $P(A) \geq 0$;
2. $P(S) = 1$;
3. Si A, B son dos subconjuntos disjuntos de S , es decir, $A \cap B = \emptyset$, entonces se satisface que $P(A \cup B) = P(A) + P(B)$.

Pensando en términos de probabilidad, el axioma 1 implica que no hay probabilidades negativas; el segundo axioma nos dice que la variable toma un valor y S contiene todos los posibles valores que toma A ; y finalmente, el último axioma implica que la probabilidad de que uno de dos subconjuntos sucedan es simplemente la suma de las probabilidades de que cada uno suceda; en la terminología de la estadística a los subconjuntos se les suele llamar **eventos**. Decimos que A, B son dos eventos **independientes** si $P(A \cap B) = P(A)P(B)$; en otras palabras, la probabilidad de que sucedan dos eventos independientes es el producto de sus probabilidades.

Sean A y B dos eventos. Podemos definir la probabilidad de que suceda A dado B , llamada **probabilidad condicional** y denotada $P(A|B)$, como

$$P(A|B) := \frac{P(A \cap B)}{P(B)}. \quad (4.3)$$

De la Ec. (4.3) puede verse que la probabilidad condicional de dos eventos independientes es simplemente la probabilidad del evento; es decir, si A, B son dos eventos independientes, entonces la probabilidad de que suceda A dado B es simplemente la probabilidad de que suceda A , pues al ser independientes, el hecho de que ocurra B no afecta el que ocurra o no A .

Utilizando la Ec. (4.3) se puede escribir la probabilidad condicional de que ocurra B dado A y notando que $A \cap B = B \cap A$, entonces

$$P(B|A)P(A) = P(A|B)P(B) \quad (4.4)$$

$$\Rightarrow P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \quad (4.5)$$

El resultado (4.5) es conocido como **Teorema de Bayes** y muestra que la probabilidad condicional de que ocurra A dado B está relacionada con la probabilidad de que ocurra B dado A .

En muchos casos, y particularmente en literatura de Física, A representa un posible resultado de un experimento, y entonces se entiende $P(A)$ como la probabilidad de que el resultado A salga

en la medición del experimento. Desde el punto de vista **frecuentista**, $P(A)$ se interpreta como el número de veces que se obtiene el resultado A si el experimento se realiza en número infinito de veces en condiciones iguales [26]. Así, se tiene que $P(A)$ se puede expresar como el límite

$$P(A) = \lim_{n \rightarrow \infty} \frac{\text{veces que se obtiene } A}{n}, \quad (4.6)$$

con n el número de veces que se realiza el experimento.

No obstante, se puede definir una **probabilidad subjetiva** en la cual A es una hipótesis que puede ser cierta o falsa y $P(A)$ representa el grado de creencia de que A sea verdadera. Debido a que ahora se tratará con hipótesis, conviene hacer un cambio de notación $A \rightarrow H$. Así, dada una hipótesis H , se debe satisfacer la **regla de la suma** [27]

$$P(H) + P(\bar{H}) = 1, \quad (4.7)$$

donde \bar{H} significa que la proposición o hipótesis H es falsa. De esta manera, la regla de la suma indica que una hipótesis es cierta o no. Bajo esta perspectiva, el Teorema de Bayes, que ahora toma la forma

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}, \quad (4.8)$$

se puede interpretar como la probabilidad de que una hipótesis H sea verdadera dados unos datos D . Las diferentes partes de la Ec. (4.8) se pueden interpretar como:

1. $P(D|H)$: Probabilidad de que se observen los datos suponiendo que la hipótesis H es verdadera, y es conocida como **Likelihood**;
2. $P(H)$: Probabilidad de que la hipótesis sea verdadera sin evidencia previa, y es llamado **Probabilidad a priori** (por simpleza se referirá a ella como **prior**);
3. $P(D)$: Probabilidad de medir evidencia. Resulta ser una constante de normalización, y se le suele llamar **Evidencia**.

Considerando que $P(D)$ es una constante, el Teorema de Bayes se puede escribir como

$$P(H|D) \propto P(D|H)P(H), \quad (4.9)$$

por lo que la probabilidad de que una hipótesis sea cierta dados unos datos es proporcional al producto entre la probabilidad de que los datos sean observados suponiendo H cierta y el conocimiento previo que teníamos sobre la certeza de la hipótesis.

El Teorema de Bayes es fundamental cuando se busca hacer inferencia a partir de datos. Resulta que este teorema plantea las bases para la teoría necesaria en la estadística Bayesiana. Recordemos que lo que queremos es obtener los valores de parámetros que mejor ajustan la órbita de la estrella S2 a los datos astronómicos, haciendo que este problema sea por naturaleza un problema de inferencia Bayesiana. Es importante aclarar que si bien el Teorema de Bayes nos da los fundamentos matemáticos para poder hablar de inferencia de este tipo, más adelante veremos que en este caso lo que nos va a dar las distribuciones posteriores de los parámetros es la implementación de los métodos MCMC.

4.2. Métodos de Monte Carlo

En la sección anterior vimos cómo el Teorema de Bayes permite conectar probabilidades condicionales de una hipótesis y datos. Ahora lo que toca hacer es introducir una herramienta poderosa que se basa en variables aleatorias para obtener la respuesta de diversos problemas.

Los Métodos de Monte Carlo son técnicas computacionales utilizadas para resolver problemas por medio de variables aleatorias. En muchos casos, estos métodos permiten obtener aproximaciones a respuestas de problemas invirtiendo el enfoque: en lugar de estimar cantidades aleatorias por medios deterministas, usan variables aleatorias para estimar cantidades deterministas [28]. Por sí mismos, los Métodos de Monte Carlo parecen no ayudar en el problema de obtener valores de los parámetros de la órbita que ajusten bien a datos; no obstante, son uno de los dos ingredientes necesarios para la construcción de un método que sí va a ser capaz de encontrar valores que hagan que los modelos ajusten a los datos. Es por esto que es importante entender la naturaleza de los Métodos de Monte Carlo con ejemplos sencillos que muestren su alcance.

El ejemplo clásico de lo anterior es utilizar cantidades aleatorias para aproximar el valor de π . Consideremos una región cuadrada en \mathbb{R}^2 en la cual está inscrito un círculo de radio 1. En este caso, la razón entre las áreas del círculo y del cuadrado es

$$\frac{\text{Área círculo}}{\text{Área cuadrado}} = \frac{\pi r^2}{(2r)^2} = \frac{\pi}{4}. \quad (4.10)$$

De esta manera, tenemos que el valor de π está conectado con un aspecto geométrico entre el cuadrado y el círculo. Para aplicar los Métodos de Monte Carlo supongamos una “lluvia ideal” sobre el cuadrado, es decir, la probabilidad de que caiga una gota en una región \mathcal{A} es proporcional al área \mathcal{A} e independiente de la posición; los posibles valores que tendría la posición de la gota son precisamente las posiciones (x, y) dentro del cuadrado. Si se dejan caer muchas gotas, la relación

entre el número total de éstas que cayeron y las que cayeron dentro del círculo se aproxima a la Ec. (4.10). La Fig. 4.1 muestra esquemáticamente este proceso.

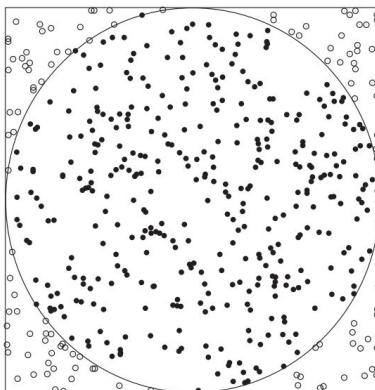


Figura 4.1: Simulación de la caída de gotas sobre un cuadrado con un círculo inscrito para aproximar el valor de π . La relación entre el número total de gotas que cayeron y las que cayeron dentro del círculo se relaciona con el valor de π . Imagen tomada de [28].

El ejemplo de la “lluvia ideal” se puede ver computacionalmente por medio de número pseudo-aleatorios generados por un programa. En este caso, cada coordenada de una gota es una muestra aleatoria extraída de una distribución uniforme $\mathcal{U}_{[0,1]}$, y cada gota representa una iteración del programa. A continuación se muestra el pseudocódigo en el cual se “dejan caer” N gotas, i.e. se realizan N iteraciones:

Algorithm 1 Aproximación del valor de π

```

k=0                                     ▷ Variable para contar las gota dentro del círculo
for  $i = 0$  to  $N$  do
  Sampleo  $x$  de  $\mathcal{U}_{[0,1]}$ 
  Sampleo  $y$  de  $\mathcal{U}_{[0,1]}$ 
  if  $(x, y)$  está en el círculo then
    k+=1
  end if
end for
 $\pi \approx 4k/N$ 

```

Es claro que aproximar el valor de π es un ejercicio sencillo pero educativo del uso de variables aleatorias para intentar obtener el valor de una cantidad. Es precisamente esto lo que motiva el uso de métodos de Monte Carlo, siendo posiblemente el más extendido y utilizado el de aproximar el valor de una integral.

En muchos problemas físicos y matemáticos la solución se plantea como una integral de alguna función. Sin embargo, en la mayoría de los casos la integral es demasiado complicada como para obtener un resultado de manera analítica; en estas situaciones se busca entonces resolver el problema de manera numérica. Una manera de obtener el valor de una integral es por medio de la llamada **integración de Monte Carlo**, la cual se basa en números aleatorios e iteraciones para aproximar el resultado. Para esto, supongamos que queremos conocer el valor de la integral

$$I = \int_a^b f(x) dx . \quad (4.11)$$

Lo que podemos hacer es expresar la integral en términos de la esperanza de una variable x que tiene una distribución $\rho(x)$ como

$$E(f(x)) = \int_a^b f(x)\rho(x) dx \quad (4.12)$$

El caso más sencillo es el de una distribución uniforme del intervalo (a, b) , es decir, $\rho(x) \equiv U(a, b) = 1/(b - a)$. Tomando esta consideración,

$$E(f(x)) = \int_a^b f(x) \frac{1}{b-a} dx . \quad (4.13)$$

De esta manera, el problema de estimar el valor de la integral se reduce a estimar $E(f(x))$. Un resultado que permite esto es el **Teorema del Valor Medio** [29]:

Teorema 4.2.1 *Si f es una función continua sobre $[a, b]$, entonces existe un $c \in [a, b]$ tal que*

$$f(c) = \frac{1}{b-a} \int_a^b f(x) dx . \quad (4.14)$$

Este resultado sugiere que tomemos $E(f(x)) = f(c)$. Con esto en mente, se puede aproximar $E(x)$ por medio de un promedio de valores de f sobre el intervalo a integrar como

$$E(f(x)) \approx \frac{1}{n} \sum_{i=1}^n f(x_i) . \quad (4.15)$$

Así, de acuerdo a la Ec. (4.13) se tiene que el valor de la integral se puede aproximar numéricamente como

$$\int_a^b f(x) dx \approx \frac{b-a}{n} \sum_{i=1}^n f(x_i), \quad (4.16)$$

donde los valores x_i se toman de manera aleatoria sobre el intervalo con la distribución uniforme $U(a, b)$. La igualdad en la Ec. (4.16) se da cuando $n \rightarrow \infty$, lo cual permite que el valor estimado se acerque arbitrariamente al valor real siempre y cuando se tome un número suficiente de muestras. Este método es excelente para usarlo computacionalmente, pues muchas veces se puede escoger un número lo suficientemente grande de muestras para hacer que el valor obtenido esté dentro de un rango de tolerancia determinado.

Definiendo $g(x) := f(x)/\rho(x)$ el error de la aproximación se puede estimar como

$$\begin{aligned} \text{error} &= \left| \frac{1}{n} \sum_{i=1}^n g(x_i) - I \right| \\ &= \left| \frac{\sigma_g}{\sqrt{n}} \left(\frac{\sum_{i=1}^n g(x_i) - nI}{\sigma_g \sqrt{n}} \right) \right| \\ &\approx \left| \frac{\sigma_g}{\sqrt{n}} \mathcal{N}(0, 1) \right|, \end{aligned} \quad (4.17)$$

donde $\sigma_g^2 = \int (g(x) - I)^2 \rho(x) dx$ y $\mathcal{N}(0, 1)$ representa una distribución normal con promedio 0 y varianza 1. La última aproximación se sigue del **Teorema del Límite Central** [30]. Entonces se puede concluir que el error decrece como $1/\sqrt{n}$. Otro comentario importante del desarrollo anterior es que el método de Monte Carlo está fuertemente influenciado por la distribución con densidad $\rho(x)$.

Una de las mayores ventajas del método de integración de Monte Carlo es la facilidad con la que se puede implementar y la cantidad de problemas que se pueden resolver con él. El algoritmo para resolver la integral de una función $f(x)$ es entonces:

1. Obtener n muestras aleatorias dentro del rango (a, b) .
2. Evaluar la función a integrar en cada una de éstas.
3. Promediar, es decir, calcular el lado derecho de Ec. (4.16).

De esta manera, se puede aproximar el valor de la integral; sin embargo, el precio a pagar por esta simplicidad es la convergencia del método, pues ésta va como $1/\sqrt{n}$. Esto muchas veces implica que para obtener la precisión deseada se tiene que generar un número demasiado grande de muestras.

Ahora, otra característica importante de este método es que al tomar una distribución uniforme para obtener las muestras, la estimación de la integral no está sesgada, haciendo que termine convergiendo al valor real de ésta [31].

Tanto el cálculo del valor de π como la integración de Monte Carlo son ejemplos por excelencia de Métodos de Monte Carlo y, además, muestran lo útiles que pueden ser las variables aleatorias para resolver distintos problemas. No obstante, en muchos casos se requiere de algún mecanismo para que los valores de una variable aleatoria sigan una cierta distribución según la situación, y este mecanismo resulta ser el otro de los ingredientes que se mencionaron al inicio de la sección.

4.3. Cadenas de Markov

Lo característico de la integración de Monte Carlo es que las muestras que se toman para promediar son independientes entre sí, es decir, el elemento x_i no está relacionado con el elemento x_{i-1} . Sin embargo, se pueden construir procesos estocásticos en los cuales el estado i -ésimo depende de la historia del sistema, i.e. de valores anteriores de las variables. Un caso particular de esto es cuando el estado x_{t+1} es extraído de una distribución $P(x_{t+1}|x_t)$ que depende solamente del estado actual, x_t . Estos procesos son llamados **Cadenas de Markov** [32]. De esta manera, una cadena de Markov nos permite extraer valores de la variable aleatoria de una distribución que solamente depende del paso anterior y no de la historia completa de la cadena. La distribución $P(\cdot|\cdot)$ se llama **kernel de transición** de la cadena, y cuando es independiente de t se dice que la cadena es homogénea en el tiempo. En el caso del método de integración Monte Carlo se tiene una variable aleatoria X cuyos valores son los valores dentro del intervalo (a, b) , pero dicho método no constituye una cadena de Markov porque la probabilidad de moverse a otro punto es independiente del estado anterior, pues se trabaja con una distribución uniforme.

Sea un sistema con una variable aleatoria X que puede tomar los valores x_1, x_2, \dots, x_k . Al conjunto de posibles valores, $S = \{x_i\}$, se le llama **conjunto de estados** [26]; entonces en cada iteración la variable X toma un valor de S . En el caso de una Cadena de Markov se tiene que en cada iteración la variable toma un valor siguiendo una distribución que depende solamente del valor actual de X .

Sea una Cadena de Markov homogénea para una variable aleatoria X , es decir, las probabilidades $P(\cdot|\cdot)$ son independientes del tiempo. De esta manera, podemos denotar la probabilidad de ir del estado i al j como

$$p_{ij} := P(X_{t+1} = j | X_t = i) , \quad (4.18)$$

donde por ser homogénea no importa en cuál paso de la cadena se cambie de estado y las probabilidades p_{ij} están definidas sin ambigüedad. No hay que confundir la homogeneidad de una cadena de Markov con la uniformidad de la distribución $U(a, b)$ en el caso de la integración de Monte Carlo, ya que la primera significa que la probabilidad de ir del estado i al estado j , p_{ij} , es la misma sin importar en qué paso de la cadena se llegue al estado i pero sigue habiendo una dependencia del estado anterior. Por otro lado, en la integración de Monte Carlo la variable toma un valor en la iteración siguiente sin importar el valor actual de ésta.

Sea un sistema con una variable aleatoria X que puede tomar tres diferentes valores, i.e. el espacio de muestras es $S = \{x_1, x_2, x_3\}$. Supongamos que en la iteración i la variable está en el estado x_2 , por lo que para el siguiente paso puede pasar a los estados x_1, x_3 o incluso quedarse en el estado x_2 , y la probabilidad con la que lo va a hacer está determinada por las probabilidades p_{21}, p_{23}, p_{22} , respectivamente.

Para un caso más general, consideremos ahora el espacio de estados de una variable con $k + 1$ estados posibles, por lo que $S = \{x_0, x_1, x_2, x_3, \dots, x_k\}$. Se puede definir la probabilidad de ir desde cualquier estado a cualquier otro de la misma manera que en Ec. (4.18), y se obtiene la **matriz de transición** de la cadena, cuya entrada (i, j) representa la probabilidad de ir del estado i al j en el siguiente paso, es decir,

$$\mathbf{P} = \begin{bmatrix} p_{00} & p_{01} & p_{02} & \dots & p_{0k} \\ p_{10} & p_{11} & p_{12} & \dots & p_{1k} \\ p_{20} & p_{21} & p_{22} & \dots & p_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ p_{k0} & p_{k1} & p_{k2} & \dots & p_{kk} \end{bmatrix} . \quad (4.19)$$

Es importante resaltar que como cada entrada representa una probabilidad, se debe tener que $p_{ij} \geq 0$. Además, la fila i representa todas las probabilidades de ir del estado i a cualquier otro, por lo que se satisface que la suma sobre cualquier fila de la matriz \mathbf{P} debe ser igual a 1. Finalmente, la probabilidad de ir del estado i al j en m pasos es simplemente la entrada (ij) de la matriz

$$\mathbf{P}^m = \underbrace{\mathbf{P} \cdot \dots \cdot \mathbf{P}}_{m \text{ veces}} .$$

Sea una cadena con $k + 1$ estados posibles, es decir, $S = \{x_0, x_1, x_2, x_3, \dots, x_k\}$. En el caso en que no se conozca con certeza el estado inicial de la cadena, sino que éste obedezca una distribución de probabilidad, se puede definir la **distribución inicial** como el vector fila π_0 donde cada entrada representa la probabilidad de que la cadena empiece en dicho estado, es decir,

$$\pi_0 = (q(0), q(1), q(2), \dots, q(k)) , \quad (4.20)$$

tal que $q(i) = P(X_0 = i)$ y $\sum_i q(i) = 1$. Es fácil ver que la distribución para el instante t es simplemente $\pi_t = \pi_0 P^t$, pues por inducción se tiene que $\pi_1 = \pi_0 P$, y entonces $\pi_2 = \pi_1 P = \pi_0 P^2$.

Es importante resaltar que la distribución π_t y la matriz de transición juegan un papel fundamentalmente diferente en una cadena de Markov. Por un lado, la primera representa las probabilidades para cada uno de los posibles estados de que la variable X tome dichos valores en la iteración t ; por el otro, la matriz P representa la evolución de la cadena y conecta las probabilidades de la variable entre iteraciones. Dicho de otra forma, para una cadena homogénea las entradas de P no cambian nunca pero los valores de π pueden cambiar en cada paso.

La pregunta natural es: ¿después de suficientes iteraciones es posible que los valores de π ya no cambien? En otras palabras, después de un tiempo suficiente de evolución del sistema, ¿las probabilidades de los diferentes estados de X tienden a algún valor? La respuesta es sí. Dada una distribución inicial π_0 y una matriz de transición P , la cadena va a tener una **distribución límite**, π , dada como

$$\pi = \lim_{t \rightarrow \infty} \pi_0 P^t = \lim_{t \rightarrow \infty} \pi_0 p^t(i, j) . \quad (4.21)$$

La existencia de esta distribución límite está garantizada por el siguiente teorema [26]:

Teorema 4.3.1 *Sea una cadena de Markov $\{X_n\}$ con matriz de transición P . Si esta matriz satisface:*

1. *Irreductibilidad: para cualquier estado de la cadena hay una probabilidad de moverse a otro estado del sistema;*
2. *Aperiodicidad: la cadena no debe quedarse atrapada en ciclos de valores del espacio de parámetros,*

entonces existe una única distribución estacionaria π para cualesquiera i, j , es decir,

$$\lim_{t \rightarrow \infty} p^t(i, j) = \pi(i, j) . \quad (4.22)$$

Para entender mejor el significado de una distribución límite supongamos que tenemos un sistema con una variable aleatoria con tres posibles estados, i.e. $S = \{x_1, x_2, x_3\}$ y que la matriz de transición está dada por

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0.1 & 0.9 \\ 0.6 & 0.4 & 0 \end{bmatrix} . \quad (4.23)$$

Vemos que en efecto se satisface que la suma de las entradas de cada fila es la unidad. Además, existen pasos prohibidos, pues las entradas con 0 implican que es imposible que la variable pase de cierto estado a otro. Este sistema puede visualizarse por medio del diagrama de transición en la Fig. 4.2 en donde cada nodo representa los posibles estados y las flechas las probabilidades de pasar de un estado al otro; las probabilidades nulas implican que no hay flecha que conecte los estados.

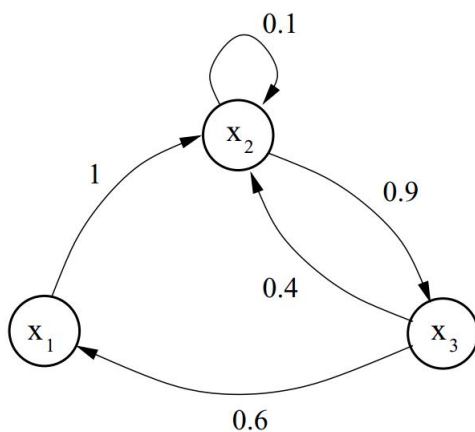


Figura 4.2: Gráfica de transición de una variable aleatoria con tres posibles estados. Los nodos representan los estados posibles de la variable y las conexiones entre nodos las probabilidades de moverse entre dichos estados. Imagen tomada de [31].

Supongamos que los estados tienen una distribución inicial $\pi_0 = (0.5, 0.2, 0.3)$. Así, para el tiempo $t = 1$ la distribución para los tres estados está dada por $\pi_1 = \pi_0 P$, con la matriz P dada como la Ec. (4.23). Realizando la multiplicación se tiene que $\pi_1 = (0.18, 0.64, 0.18)$. Después de suficientes pasos se observa que la distribución de los estados tiende a $\pi = \lim_{t \rightarrow \infty} \pi_0 P^t = (0.2, 0.4, 0.4)$. De esta manera, sin importar en cuál estado se empiece, se puede asegurar que para t suficientemente grande los estados tienen una probabilidad bien determinada y la variable aleatoria X está regida por una distribución bien definida.

Finalmente, lo que hace muy poderosas a las cadenas de Markov es que si satisfacen la dos condiciones del Teorema 4.3.1, entonces la distribución límite de la variable X es independiente de su estado inicial [27]. Esto se traduce en que para cualquier distribución inicial de los estados, π_0 , se tiene que

$$\lim_{t \rightarrow \infty} |\pi_0 P^t - \pi| = 0, \quad (4.24)$$

con π la distribución límite de la cadena y P su matriz de transición. De esta manera se puede ver explícitamente que lo que determina la distribución final de la variable es solamente la matriz P .

Hasta el momento sólo hemos hablado del caso de variables aleatorias discretas, pero para nuestro propósito vamos a necesitar variables aleatorias continuas, i.e. variables cuyo espacio de muestras es, por ejemplo, un intervalo $[a, b]$; en este sentido, para toda iteración la variable X_i puede tomar cualquier valor real dentro del intervalo. Debido a lo anterior, el kernel de transición no puede ser una matriz, sino que ahora se tiene una densidad de probabilidad $K(x, y) := \text{Prob}(x_{t+1} = y | x_t = x)$ [27]. Cuando hablamos de espacios de muestras discretos notamos que las filas de P tenían que sumar la unidad; ahora, como se tiene una densidad de probabilidad, se va a tener que satisfacer que $\int dy K(x, y) = 1$ sobre todos los posibles valores. No obstante, los resultados de esta sección se generalizan fácilmente al caso de una variable continua notando que en lugar de P se tiene K .

4.4. Métodos MCMC aplicados al ajuste de órbitas estelares

Una de las aplicaciones más importantes de las cadenas de Markov es en los llamados Métodos de Monte Carlo basados en Cadenas de Markov (*MCMC*, por sus siglas en inglés). Estos métodos se utilizan para tomar muestras de una distribución que se desconoce y se componen de dos partes: un método de muestreo aleatorio, y una cadena de Markov con una distribución límite particular. De todos los métodos MCMC, el más conocido es el Algoritmo de Metropolis-Hastings (MH) [33], y muchos de los algoritmos modernos pueden verse como un caso particular de éste. El Algoritmo de MH se puede dividir en dos grandes pasos:

1. Proponer un nuevo valor de la variable aleatoria (lo llamaremos **candidato**) x_{t+1} de la distribución $K(x_t, x_{t+1})$, y
2. Ver si el candidato es bueno o no, es decir, si el nuevo modelo reproduce mejor los datos o no.

El paso 1 muestra claramente que se está trabajando con un proceso Markoviano pues la forma de moverse en el espacio de muestras es debido a una distribución que depende solamente del valor actual de la variable. Con esto se puede ver que el nombre de Monte Carlo en los métodos se debe a que el candidato x_{t+1} se obtiene de manera aleatoria de una distribución.

Existen diferentes formas de proponer la densidad $K(x_t, x_{t+1})$, lo cual da origen a una familia de sub-algoritmos de MH. El sub-algoritmo que se discute en este trabajo es el Random Walk Metropolis-Hastings (RWMH), en el cual la distribución $K(x_t, x_{t+1})$ es una distribución normal centrada en x_t y con varianza σ^2 , i.e. $K(x_t, x_{t+1}) \equiv \mathcal{N}(x_t, \sigma^2)$.

Regresando al problema del ajuste de la órbita de S2, en nuestro caso tenemos que el modelo que describe al sistema son las ecuaciones de la geodésica en el espacio-tiempo de Schwarzschild, y

los parámetros θ corresponden al conjunto (M, l) , los cuales además no son completamente independientes, como se discutirá en el capítulo siguiente². Aquí ya podemos comenzar a aterrizar todos los conceptos que hemos visto a lo largo de este capítulo. Por ejemplo, comenzamos hablando del problema del ajuste de la ecuación de un círculo generado por un modelo $M(\theta)$ a datos en \mathbb{R}^2 . Para el caso de órbitas estelares el proceso es a grandes rasgos muy parecido: escogemos valores de (M, l) , generamos la órbita correspondiente y realizamos una prueba de tipo χ^2 como en la Ec. (4.2).

Este proceso es claro cuando queremos analizar un único θ ; sin embargo, el objetivo es realizar muchas pruebas con distintos θ para intentar encontrar los valores que mejor reproducen los datos. Es aquí donde entran en juego los procesos estocásticos. Si tratamos nuestros parámetros como variables aleatorias continuas X_i , entonces podemos hacer uso de los métodos de Monte Carlo para intentar encontrar los mejores modelos para la órbita de S2. Un primer intento podría ser definir un intervalo dentro del cual sepamos que deben estar M y l , i.e. $M \in [M_{inf}, M_{sup}]$ y $l \in [l_{inf}, l_{sup}]$; luego, en cada iteración extraer un par (M, l) donde cada uno esté dentro de su rango correspondiente, y calcular la Likelihood (Ec. (4.2)). La distribución de la cual se podrían extraer podría ser, por simplicidad, una distribución uniforme. No obstante, el mayor inconveniente de esta construcción es que los métodos de Monte Carlo dan muestras independientes. Esto implica que realmente se tiene un método muy poco eficiente porque es igual de probable en la siguiente iteración que se obtenga un modelo malo incluso si se acaba de obtener un θ que reproduce bien los datos, haciendo que nos alejemos de las zonas en \mathcal{X} de mejor ajuste a los datos.

Teniendo en cuenta lo anterior, la mejor solución a nuestro alcance es introducir cadenas de Markov en el proceso de selección de los θ . Antes de continuar, es importante introducir el concepto de **caminante**. En este contexto, un caminante sería un proceso que evalúa un θ por iteración o paso; el origen del nombre viene de que la cadena explora el espacio \mathcal{X} en busca de las regiones con los valores de θ que mejor ajustan a los datos. El siguiente paso es proponer el kernel de transición, $K(\cdot, \cdot)$, que van a utilizar nuestros caminantes para explorar el espacio. Aquí podemos utilizar el planteamiento del algoritmo RWMH en el cual el kernel es $\mathcal{N}(x_t, \sigma^2)$. Es claro que esto es un kernel de una cadena porque la distribución está centrada en el valor actual de las variables, y es estacionario porque σ^2 no depende de t , pues es una Gaussiana que simplemente se desplaza sin deformarse.

Para entender mejor el funcionamiento de este kernel veamos lo siguiente. Supongamos que en la iteración t tenemos un conjunto $\theta_t = (M_t, l_t)$. Para la iteración $t + 1$ el caminante va a escoger de manera aleatoria los valores M_{t+1} y l_{t+1} de las distribuciones $\mathcal{N}(M_t, \sigma_M^2)$ y $\mathcal{N}(l_t, \sigma_l^2)$, respectiva-

²En el siguiente capítulo veremos que l y M no son completamente independientes; además, el modelo requiere de más parámetros, pues estamos ocupando las ecuaciones de la geodésica en el marco de referencia de Sgr A* y los datos son las posiciones estelares vistas desde la Tierra.

mente, donde σ_M^2 y σ_l^2 son las varianzas que proponemos para cada parámetro. Si proponemos una varianza muy pequeña, entonces el movimiento del caminante está muy restringido; por otro lado, si proponemos una varianza muy grande, el caminante tiene un movimiento menos preciso ya que tendrá entonces un rango más amplio de posibles valores en cada paso.

Ahora, ¿cómo sabemos que $\theta_{t+1} = (M_{t+1}, l_{t+1})$ genera un modelo que ajusta mejor a los datos? Dada la forma de la Ec. (4.2), si tenemos dos conjuntos de valores, θ_t y θ_{t+1} , el mejor conjunto va a ser aquel que resulte en un valor de Likelihood más cercano a cero. En caso de que θ_{t+1} genere un modelo que tenga una peor Likelihood que θ_t , para la iteración $t + 1$ se redefinen las variables aleatorias como $M_{t+1} = M_t$ y $l_{t+1} = l_t$; esto lo que indica es que el caminante no se mueve y debe escoger un conjunto θ que sea mejor que el actual. Por otro lado, si θ_{t+1} sí dio un mejor valor de Likelihood, significa que el caminante se acerca a la región de \mathcal{X} deseada y continúa por ese camino. Cabe resaltar que se pueden implementar muchos caminantes, cada uno siguiendo su respectivo proceso de exploración y selección; de hecho, en este trabajo se implementarán más de un caminante. El proceso individual de cada caminante se ilustra mejor en la Fig. 4.3.

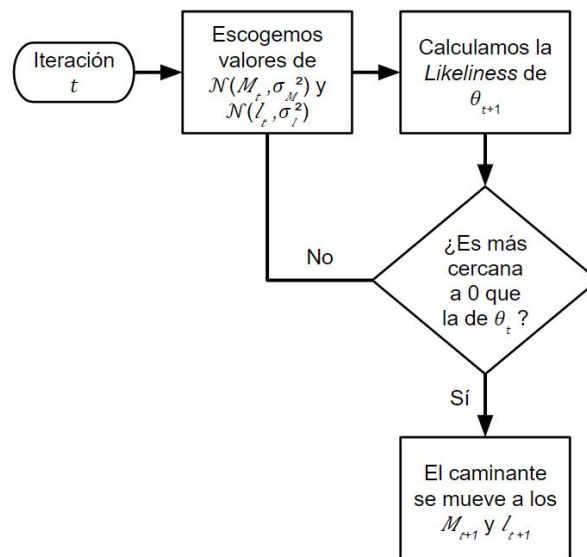


Figura 4.3: Diagrama de flujo del proceso de selección de un nuevo conjunto de parámetros θ del modelo.

Pareciera que con esto tenemos ya todos los ingredientes para construir el código y echarlo a andar; sin embargo, no es así. Todavía tenemos que resolver el problema de que todas las discusiones hasta este punto han sido suponiendo que la órbita en el marco de referencia de Sgr A* es igual a como se observa en la Tierra, lo cual no es cierto, pero esto se resolverá en el siguiente capítulo.

Capítulo 5

Implementación y resultados

En este capítulo veremos cómo se implementan todos los conocimientos y herramientas desarrollados en los capítulos anteriores para el ajuste de la órbita de la estrella S2. Además, se resuelve el problema de generar la órbita en el marco de referencia de Sgr A* y luego proyectarla al plano del cielo para compararla con los datos observacionales. Luego, se explica la limitación del uso de las ecuaciones diferenciales del Cap. 3 para poder comparar las simulaciones con las observaciones y se muestra la solución encontrada. Después, se especifican todos los datos requeridos por el código computacional para poner en práctica los métodos MCMC en nuestro caso particular. Finalmente, se muestran los resultados obtenidos y se discute su significado.

5.1. Implementación del código

5.1.1. Integración y proyección de la órbita

En un principio, las órbitas en un plano 2D pueden ser generadas simplemente utilizando las Ecs. (3.45) y (3.39) e integrándolas numéricamente con las condiciones iniciales necesarias. En este caso, se necesitaría una condición inicial para la posición radial, r_0 , y una para el ángulo azimutal inicial, ϕ_0 . Teniendo estas condiciones se puede utilizar cualquier paquetería en PYTHON que integre numéricamente, y haciendo la transformación a coordenadas cartesianas como $x = r \cos \phi$, $y = r \sin \phi$ se tendría una gráfica bidimensional de la órbita (una gráfica 2D es suficiente ya que las órbitas en Schwarzschild permanecen en un plano). No obstante, en lugar de (3.45), en este trabajo se utilizará la ecuación de segundo orden (3.53), la cual involucra \ddot{r} , y a continuación veremos las razones de esto.

Si bien la integración de la ecuación (3.53) requiere de otra condición inicial, que en este caso sería \dot{r}_0 , la naturaleza del problema permite hacer simplificaciones al respecto. Si se conoce el punto de máximo acercamiento (el periapsis), entonces lo elegimos como punto inicial para la integración y lo denotamos r_0 . Como r_0 corresponde a un punto de retorno, entonces, por definición, la condición inicial para \dot{r}_0 es trivialmente $\dot{r}_0 = 0$. Otra ventaja de elegir la ecuación (3.53) es que ésta no involucra la “energía” \mathcal{E} y por lo tanto es innecesario ajustarla como un parámetro más de la órbita. Notemos que la elección de resolver la ecuación de segundo orden es válida en virtud de que se tienen datos observacionales de la estrella que lo permiten; en general, utilizar la ecuación diferencial de segundo orden podría no ser conveniente si no se conoce o estima el valor del periapsis del objeto que orbita.

Es importante resaltar que al seguir estos pasos se está obteniendo la órbita en el marco de referencia del centro de masa, que si el objeto central es muy masivo con respecto al que orbita, se puede tomar como el marco con el objeto masivo en el origen. Esto es relevante ya que como se quiere comparar la órbita con los datos observados, entonces es necesario proyectar la órbita obtenida al plano del cielo de la Tierra. Para realizar esto, se tiene que definir la orientación del plano sobre el cual está la órbita con respecto al plano celeste, lo cual requiere de dos parámetros: la **inclinación** (i), y la **longitud del nodo ascendente** (Ω).

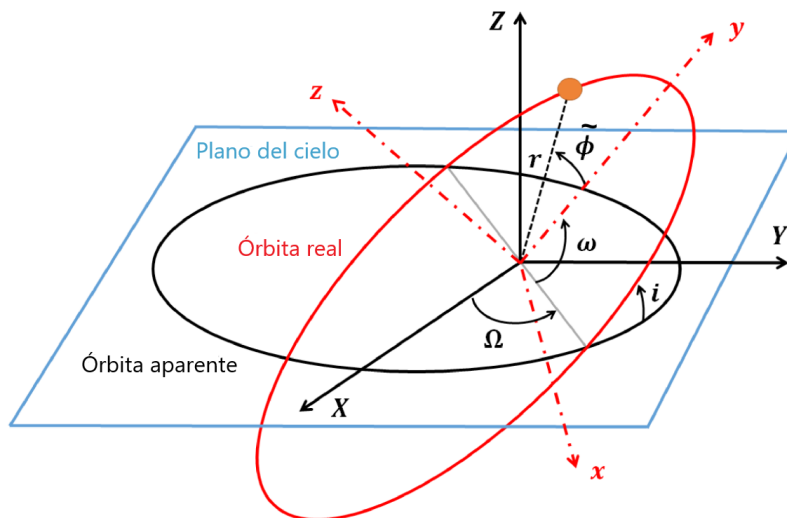


Figura 5.1: Elementos para la orientación y proyección de la órbita real a la órbita aparente sobre el plano del cielo. Se muestran los diferentes ángulos que permiten orientar tanto el plano de la órbita como la órbita misma sobre éste.

La línea de intersección entre el plano del cielo y el plano de la órbita de la estrella se llama **nodo ascendente**; en la Fig. 5.1 se observa como una línea gris. El siguiente paso es medir el ángulo que forma esta línea con respecto al marco de referencia del plano celeste, X, Y, Z ; este ángulo se llama **longitud del nodo ascendente** y está representado por la letra Ω . El último paso para orientar el plano de la órbita es medir el ángulo que forma éste con respecto al plano celeste; este ángulo se llama **inclinación de la órbita** y está representado por la letra i . Finalmente, el **argumento del periapsis**, ω , define el ángulo que forma el periapsis de la órbita medido desde el nodo ascendente (línea gris) [3]. La Fig. 5.1 muestra esquemáticamente la definición de estos ángulos.

Si x, y representan las coordenadas de la órbita real, entonces las componentes X, Y en el plano del cielo se obtienen por medio de las llamadas **constantes de Thiele-Innes** dadas como [10]

$$X = Bx + Gy \quad , \quad Y = Ax + Fy \quad , \quad (5.1)$$

donde las constantes A, B, F, G están dadas por

$$A = a(\cos \Omega \cos \omega - \sin \omega \cos i \sin \Omega), \quad (5.2)$$

$$B = a(\sin \Omega \cos \omega + \cos \Omega \sin \omega \cos i), \quad (5.3)$$

$$F = a(-\cos \Omega \sin \omega - \sin \Omega \cos \omega \cos i), \quad (5.4)$$

$$G = a(-\sin \Omega \sin \omega + \cos \Omega \cos \omega \cos i), \quad (5.5)$$

$$a = \frac{206265 \text{ arcs}}{r_d}, \quad (5.6)$$

con r_d la distancia entre nuestro marco de referencia y Sgr A*, es decir, al centro galáctico; el papel de la constante a es escalar la órbita real a segundos de arco (un radián es igual a 206265 arcs). Esto implica que hasta ahora se requiere de tres ángulos (ω, Ω, i) y la distancia a Sgr A*, r_d , para poder proyectar la órbita, más los parámetros para generarla (M, l, r_0, ϕ_0) .

El procedimiento antes mencionado es suficiente si lo que se busca es proyectar una órbita cualquiera al plano del cielo, pero resulta que hacer esto no nos da la información completa del sistema. El problema de integrar solamente las ecuaciones radial y angular radica en que diferentes conjuntos de parámetros pueden generar órbitas demasiado similares, por lo que al proyectarlas sería difícil distinguirlas. Para ver esto, pensemos en dos sistemas gravitacionales de una partícula con masa m orbitando un objeto mucho más masivo y se estima el valor del periapsis r_0 ; sin embargo, el sistema 1 es tal que se tienen valores M_1 y l_1 , mientras que en el sistema 2 son M_2 y l_2 . Uno siempre puede escoger pares de valores M_1, l_1 y M_2, l_2 de tal forma que las órbitas sean parecidas y hasta indistinguibles. Por ejemplo, dado un r_0 , se puede tener una órbita con M grande, lo cual

implicaría un l también grande para que la partícula masiva resista la atracción gravitacional; por otro lado, se podría generar una órbita muy parecida con M y l menores. Sin embargo, la clave está en el periodo orbital de ambos sistemas, que es el elemento que marcaría la diferencia.

Si se tienen datos astronómicos de la órbita que se busca estudiar, entonces las posiciones del objeto están marcadas por las fechas en las cuales fueron observadas. Así, solamente un par de valores de M, l van a reproducir las observaciones, ya que otro momento angular y otra masa van a implicar un periodo mayor o menor, y si bien la órbita descrita sería similar, los puntos no van a coincidir con las fechas. Entonces la ambigüedad se quita si además se introduce como restricción que las posiciones de la órbita simulada coincidan en fechas con los datos.

A pesar de que parece resuelto el problema, no hay que olvidar que las ecuaciones de movimiento están dadas como ecuaciones diferenciales en términos del tiempo propio, es decir, del tiempo en el marco de referencia del objeto que orbita, mientras que los datos tienen fechas del tiempo del marco de referencia desde el cual se observa la órbita. En el caso concreto que se estudia en este trabajo, los datos tienen la fechas según el tiempo en el sistema de la Tierra y la órbita se está integrando según el tiempo propio de la estrella S2. Evidentemente, el marco de referencia de la Tierra no es comóvil con S2, entonces los tiempos registrados se asocian a la coordenada temporal del espacio-tiempo y no al tiempo propio de la estrella. La Ec. (3.51) representa la forma de conectar τ con el tiempo en nuestro marco. Para poder introducir esta “restricción del tiempo” es necesario integrar también la Ec. (3.51) y mapear los puntos simulados con sus respectivos t para comparar. El resolver esta ecuación diferencial introduce otra condición inicial necesaria: t_0 . De esta manera, nuestro modelo requiere de nueve parámetros, los cuales son: $r_d, (\omega, \Omega, i), (M, l, r_0, t_0, \phi_0)$.

5.1.2. Interpolación de la órbita

El último problema a resolver es que debido a que se están integrando numéricamente las ecuaciones de movimiento, realmente no se tiene una expresión analítica que permita conocer los valores X, Y para cualquier valor de t . Esto es, después de integrar y proyectar la órbita, se terminan teniendo arreglos (*arrays* de PYTHON) de valores donde a cada t le corresponde un par de valores X, Y . Sin embargo, debido a la naturaleza del método utilizado, es prácticamente imposible que el conjunto de valores integrados para t corresponda exactamente con los valores de las fechas de las observaciones. La Fig. 5.2 muestra las posiciones observadas, x_{obs}, y_{obs} , según la fecha en la cual se tomaron las mediciones. Esta Figura es relevante porque permite discernir las posiciones de la estrella en función del tiempo, cosa que no se puede hacer con las posiciones como en la Fig. 1.3, y entonces va a permitir asegurarnos de que efectivamente obtenemos el par (M, l) correcto en el sentido discutido en la subsección anterior.

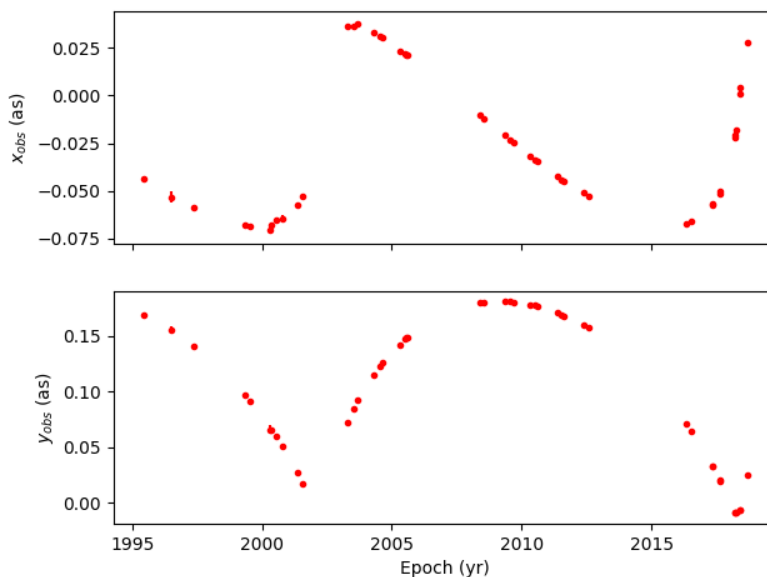


Figura 5.2: Posiciones de la estrella S2 observadas en el cielo desde 1995 hasta 2018. Las posiciones en el cielo están dadas en segundos de arco (as) y además se muestra la incertidumbre de cada medición.

Para arreglar esta situación se recurrió a realizar un proceso de interpolación de las órbitas proyectadas. Debido a que se integran conjuntamente las ecuaciones para \ddot{r} , $\dot{\phi}$ y \dot{t} , y después se proyecta la órbita, para cada valor de t corresponde una posición en el plano del cielo. Este hecho se puede explotar para determinar cuáles serían las posiciones correspondientes de la órbita proyectada en los tiempos de las observaciones y comparar con respecto a x_{obs} , y_{obs} . Para realizar la interpolación se utilizó la función `INTERP` de la paquetería `NUMPY`. Dicha función necesita tres entradas: los tiempos en los cuales se quieren interpolar los datos, los tiempos obtenidos mediante la integración, y las posiciones integradas; y realiza una interpolación lineal de punto a punto conocidos e interpola por una línea recta entre éstos.

La interpolación lineal no es ideal cuando el conjunto de valores de la función a interpolar son pocos; sin embargo, si el conjunto de puntos simulados es suficientemente grande, la distancia entre los puntos adyacentes es pequeña y la aproximación global mejora aún usando una interpolación lineal. Para observar esto, la Fig. 5.3 muestra la interpolación de la función $\sin(x)$ en dos casos: en el primero se cuenta con 10 valores simulados, mientras que en el segundo se cuenta con el doble. Es claro que en el segundo la interpolación aproxima más satisfactoriamente la forma de la curva en todo el intervalo considerado debido al tamaño del conjunto de punto simulados.

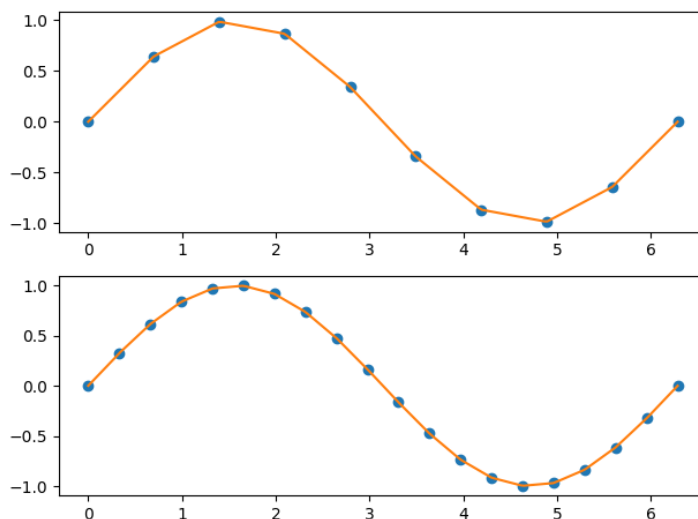


Figura 5.3: Interpolación de la función $\sin(x)$ en dos casos: con 10 valores simulados, y con 20 valores simulados. En el primero es notorio que la interpolación lineal no es ideal; por otro lado, en el segundo caso se puede observar que la interpolación produce una gráfica que se asemeja más a la función analítica en todo el dominio considerado.

En el caso de la órbita, lo que se busca es tomar la órbita obtenida numéricamente e interpolarla en las fechas de observación; esto permitiría poder comparar de manera directa nuestro modelo teórico con las posiciones observadas. Para este efecto, lo ideal es que la órbita proyectada cuente con el mayor número posible de puntos para que la interpolación sea satisfactoria y los puntos obtenidos sean lo más cercano a los valores que se tendrían si se tuviera una expresión analítica para estos. Debido al poder computacional actual, se puede integrar sin mayor problema una órbita que cuente con varios miles de puntos minimizando al máximo la apariencia quebrada por la linealidad de la interpolación, justo como pasa en la Fig. 5.3.

Finalmente, una vez que se han interpolado las posiciones se tiene un arreglo de la misma dimensionalidad que el arreglo de las fechas. Así, para cada fecha se tienen las posiciones observadas en el cielo y las posiciones correspondientes a la órbita simulada con cierto conjunto de valores de los parámetros $\theta = (M, l, \dots)$. Esto permite comparar directamente qué tan bien se acerca la órbita simulada a la observada y aceptar o descartar el conjunto de parámetros θ . La Fig. 5.4 muestra un diagrama de flujo que representa esquemáticamente el proceso mencionado.

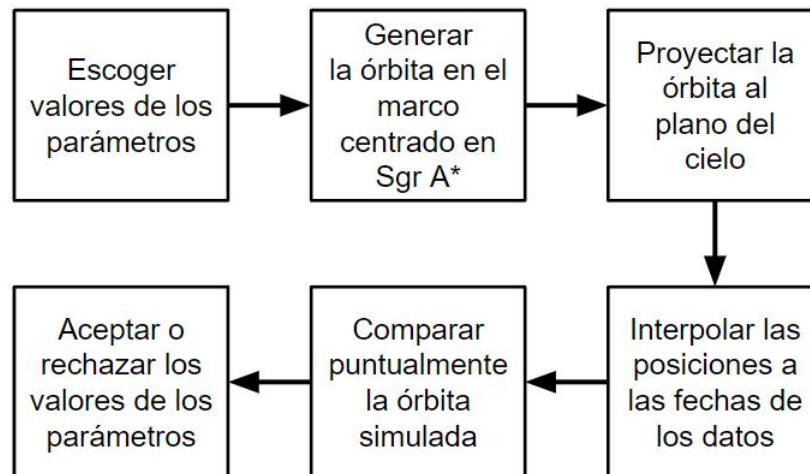


Figura 5.4: Diagrama de flujo del proceso para comparar las posiciones integradas con las observadas utilizando una interpolación de las primeras.

5.1.3. Implementación de las cadenas

El diagrama 5.4 muestra esquemáticamente el proceso para determinar si un conjunto de parámetros del modelo es bueno describiendo los datos observados, pero lo que se busca es determinar la distribución de los valores que mejor lo hacen; en otras palabras, se quiere conocer la distribución posterior de éstos. Para determinar esto se puede utilizar un método MCMC con un cierto algoritmo tal que que las distribuciones límites de las cadenas sean las distribuciones posteriores de los modelos. Es importante recordar que en el Cap. 4 se introdujo el Teorema de Bayes, el cual nos da los fundamentos matemáticos para la inferencia Bayesiana; sin embargo, la parte práctica que en este trabajo nos genera la distribución posterior es la implementación de las cadenas.

En este trabajo se implementó un algoritmo RWMH para buscar los valores de los parámetros que mejor se ajustan a los datos. Para correr el código se necesita especificar los siguientes elementos:

1. Priors (Valores propuestos de los parámetros y sus límites inferior y superior).
2. Kernel de la transición.
3. Número de caminantes.
4. Número de iteraciones para cada caminante.
5. Criterio para aceptar nuevos valores.

La Tabla 5.1 muestra los valores propuestos para los parámetros y los límites que se establecieron. Además, como se está utilizando el algoritmo RWMH, el kernel para cada parámetro es una gaussiana de la forma $\mathcal{N}(x, \sigma^2)$, donde x representa el valor del parámetro; la Tabla 5.1 también muestra la varianza en cada caso. Cada órbita generada está compuesta por 5,000 puntos con el objetivo de obtener una interpolación satisfactoria en el sentido discutido en la subsección anterior.

Parámetro	Prior	Límite inferior	Límite superior	σ	Unidad
r_d	8100	7800	8300	20	pc
l	1.408	1.35	1.45	0.01	$\times 10^{-13}$ pc ² /año
M	4.2	3.9	4.5	0.01	$\times 10^6 M_\odot$
r_0	5.5	5.3	5.8	0.01	$\times 10^{-4}$ pc
t_0	1986	1985	1987	0.1	año
ϕ_0	12.557	12.5	12.6	0.001	rad
i	134.567	130	140	0.001	°
ω	66.263	60	72	0.001	°
Ω	227.171	270	234	0.001	°

Tabla 5.1: Datos necesarios de los parámetros que se ajustan con el código; se incluyen su prior, sus límites, y la desviación estándar de su kernel asociado.

Es importante mencionar la elección particular de valores de la Tabla 5.1. Para la distancia al centro galáctico, r_d , el prior es un valor intermedio entre las estimaciones más recientes [9] y algunas no tan recientes [34], mientras que los límites son valores un tanto arbitrarios que permiten considerar un rango aceptable de movimiento de los caminantes pero para los cuales se sigue teniendo convergencia. Si se eligiera un rango más amplio, la convergencia podría requerir de más iteraciones; por otro lado, un rango más reducido podría favorecer uno de los límites o dejar fuera el valor al cual convergería si se permitieran más iteraciones. Para el parámetro de la masa, M , se siguieron los mismos criterios que para r_d . Ahora, para el pericentro, r_0 , se tiene como límite superior la distancia mínima a la cual ha sido observada S2 ($120 \text{ UA} \approx 5.8 \times 10^{-4} \text{ pc}$) [9], mientras que el límite inferior representa un valor en el cual la estrella ya no describe una órbita acotada y cae rápidamente a Sgr A* (considerando valores aproximados de M y dándole un margen de 10 UA para acercarse). Los últimos valores de priors que fueron tomados o basados directamente de la literatura fueron los argumentos orbitales (i, ω, Ω) [9]; de igual manera, los límites fueron escogidos para dar un rango suficiente a explorar a los caminantes pero que se siga teniendo convergencia con el número de iteraciones utilizadas.

El método para obtener priors y límites fue diferente para los parámetros l , t_0 , ϕ_0 . El prior de la magnitud del momento angular, que recordemos es por unidad de masa, se basa en un cálculo dado como el producto $l = r_d v_d$, con r_d el valor estimado del periapsis y v_d la rapidez orbital en dicho punto [9]; sus límites se basan en simulaciones individuales en donde el límite inferior representa aproximadamente el momento angular con el cual la estrella ya no resiste la atracción gravitacional de Sgr A*, y el límite superior el valor en el cual ésta logra escapar y describe una órbita abierta. Los valores t_0 y ϕ_0 fueron escogidos por medio de simulaciones individuales probando valores que más o menos reprodujeran la órbita en el plano del cielo y las posiciones a través de los años. Los límites para t_0 se basan en que la estimación del periodo orbital de S2 es de aproximadamente 16 años con la primera observación del periapsis en 2002 [9]; los límites para ϕ_0 son valores medianamente arbitrarios en los cuales se sigue teniendo convergencia con el hardware utilizado.

Finalmente, el valor de todas las σ es de cierta manera arbitrario. Todos fueron escogidos teniendo en mente las cifras significativas de los parámetros, pero sobre todo el número de iteraciones permitido por el hardware. Por ejemplo, se sabe que $r_d \approx 8$ kpc y el rango permitido para los caminantes es de 500 pc. Si σ fuera mucho menor, el paso de los caminantes estaría más restringido y serían necesarias más iteraciones para que el valor converja. Ahora, una σ mucho mayor podría comenzar a ser del orden del rango, haciendo que muchos valores sean permitidos por paso, afectando la convergencia. Es claro que $\sigma = 20$ pc no es un valor especial, pues se tiene cierta flexibilidad al escogerlo siempre y cuando se mantenga la convergencia con el número de iteraciones que el hardware permita.

El código se corre con 3,000 caminantes y con 7,000 iteraciones. Además, al comenzar el código se le permite a cada caminante explorar el espacio de parámetros por 500 iteraciones; este proceso es comúnmente llamado **burning-in** y permite desechar las muestras iniciales que posiblemente estén lejos de las distribuciones de equilibrio de los parámetros. Finalmente, el criterio que siguen las cadenas es el de la Ec. (4.2), el cual se explicó a detalle en la Sec. 4.4.

Cada caminante va guardando en una lista los conjuntos de valores de los parámetros que va aceptando y la Likelihood asociada a cada uno. Al finalizar se juntan todas las listas guardadas por cada caminante y se obtiene una lista global de todos los valores de θ *aceptados*. El mejor modelo que ajusta a los datos es aquel con el valor absoluto menor de Likelihood¹.

¹Este valor se puede saber haciendo que PYTHON tome el valor de Likelihood máximo; recordemos que a PYTHON se le debe pedir el valor mayor de la lista porque la Likelihood, Ec. (4.2), está dada con un signo (-), entonces va a interpretar como “valor mayor” aquel con un valor absoluto menor de la lista.

5.2. Resultados

5.2.1. Distribuciones posteriores de los parámetros

Recordemos que nuestro espacio de parámetros es de 9 dimensiones, pues estamos ajustando 9 parámetros. De esta forma, y debido a que los priors son todas distribuciones normales, la distribución posterior (dp) del modelo es una distribución gaussiana 9-dimensional. La Fig. 5.5 muestra las proyecciones de esta gaussiana en cada uno de los ejes de los parámetros, o dicho de otra manera, las distribuciones posteriores de cada parámetro. Las líneas punteadas en cada dp representan las regiones con percentiles de 16 y 86, por lo que la región entre éstas determina una distancia de 2σ .

Para entender mejor lo que nos dice la Fig. 5.5 analicemos el caso de la primera columna. Hasta arriba, lo que sería la primera fila, se puede apreciar un histograma de todos los valores aceptados para el parámetro r_d (hasta abajo se indica el parámetro). Además, en el límite con muchos caminantes este histograma se refinará hasta el punto de ser una distribución; así, el histograma representa la dp para r_d . Ahora, en todas las columnas encontramos histogramas, uno para cada parámetro con sus respectivos percentiles.

Pasando a la segunda fila de la primera columna tenemos una proyección 2D del histograma para r_d y del histograma del parámetro que se encuentra a la derecha, que en este caso sería el momento angular, l . De esta manera, la Fig. 5.5 muestra las dp de cada uno de los parámetros y las proyecciones de los histogramas de cualesquiera dos parámetros. Por ejemplo, la proyección 2D de la última fila de la primera columna es la proyección de los histogramas de r_d y Ω . Entonces la Fig. 5.5 muestra tanto las dp como las proyecciones entre éstas para todos los parámetros del modelo.

Estas proyecciones en un plano 2D permiten analizar si existen correlaciones entre diferentes parámetros; una correlación entre parámetros se revelaría en la gráfica como la proyección respectiva alargada a lo largo de una línea recta (en todos los casos obtenemos figuras cercanas a un círculo). Es justo aquí donde podemos observar de manera directa las limitaciones del hardware. La proyección para los parámetros M y r_d tiene forma parecida a un círculo, por lo que se concluye que no existe correlación entre estos; no obstante, se sabe que dicha correlación existe [35]. De hecho, no se obtuvieron correlaciones entre ningún par de parámetros; esto se sigue de que un valor de un cierto parámetro no implica un valor particular de otro, sino que puede tomar cualquier valor, como se puede observar en las proyecciones 2D de las dp. Se espera que al aumentar el número de iteraciones para poder restringir los rangos de los parámetros estas correlaciones comiencen a aparecer.

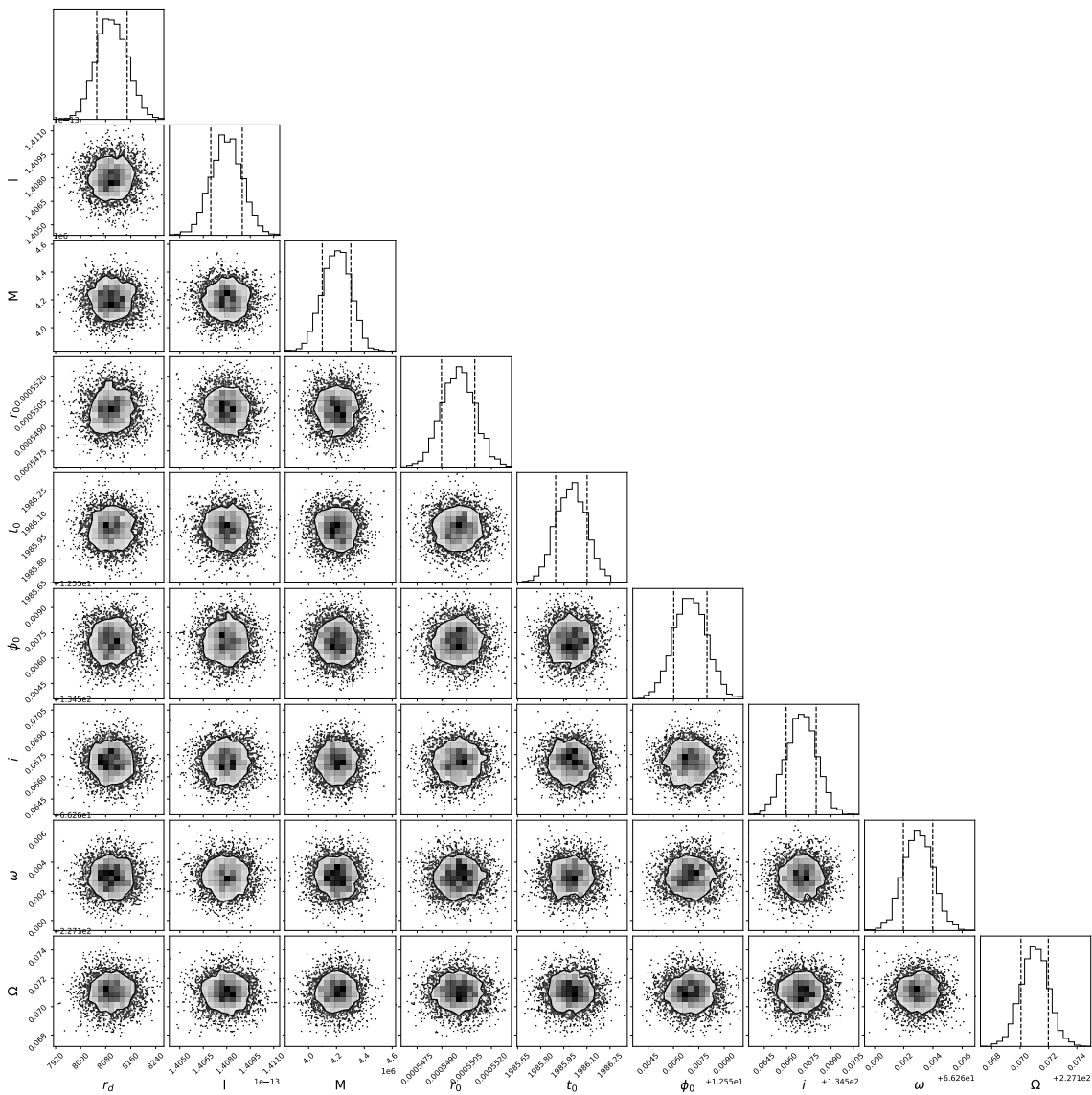


Figura 5.5: Proyecciones de las distribuciones posteriores de los valores de los parámetros. Las líneas punteadas muestran los percentiles de 16 y 84, por lo que la región entre éstas representa 2σ .

5.2.2. Mejor modelo del sistema

A continuación se presenta el mejor modelo utilizando los parámetros y valores establecidos en la Subsección 5.1.3. Primero, la Tabla 5.2 muestra los valores de los parámetros del mejor modelo encontrado (recordemos que la σ de cada uno está dado en la Tabla 5.1). No está de más recor-

dar que el objetivo de este trabajo es obtener las distribuciones posteriores de los parámetros, no necesariamente el conjunto de parámetros que mejor reproduce los datos.

Parámetro	Valor	σ	Unidad
r_d	8129	20	pc
l	1.407	0.01	$\times 10^{-13}$ pc ² /año
M	4.19	0.01	$\times 10^6 M_\odot$
r_0	5.501	0.01	$\times 10^{-4}$ pc
t_0	1986.2	0.1	año
ϕ_0	12.556	0.001	rad
i	134.566	0.001	°
ω	66.264	0.001	°
Ω	227.172	0.001	°

Tabla 5.2: Tabla con los valores de todos los parámetros ajustados. Se muestran sus valores, el valor de la desviación estándar y sus unidades.

Ahora, lo que podemos hacer es comparar nuestros resultados con los resultados obtenidos en otros trabajos, en particular, en (Abuter, R. 2020) [9]. Debido a que nuestro método para obtener la órbita es diferente, sólo podemos comparar los parámetros comunes. La Tabla 5.3 muestra esta comparación.

Parámetro	Tesis	Abuter, R. 2020	Dif%
M	4.19	4.261	1.6 %
r_d	8129	8246.7	1.4 %
i	134.566	134.567	7.4×10^{-4} %
ω	66.264	66.263	1.5×10^{-3} %
Ω	227.172	227.171	4.4×10^{-4} %

Tabla 5.3: Comparación de los valores de los parámetros para los mejores modelos entre este trabajo y el reportado en la literatura [9].

Al analizar la Tabla 5.3 es importante recordar que muchos de los priors fueron tomados de este otro trabajo, por lo que la cercanía entre nuestro mejor modelo y ese se debe en gran parte a eso. Además, al comparar los valores de la Tabla 5.1 y los valores encontrados para el mejor modelo notamos que no existe una diferencia considerable; esto se debe en particular a la calidad de los

priors escogidos. Si de antemano escogemos priors que por sí mismos reproducen bien los datos, entonces los caminantes van a permanecer cerca de esa zona.

Utilizando los valores de la Tabla 5.2 procedimos a integrar numéricamente la ecuaciones de movimiento para graficar la órbita de dicho modelo en el plano del cielo sobreponiendo los datos observacionales. La Fig. 5.6 muestra lo anterior. Es importante resaltar que el modelo no ajusta bien a los puntos cercanos a la posición de Sgr A*; no obstante, esto no es de extrañar y no debería ser preocupante puesto que esas posiciones corresponden a los momentos en los cuales S2 pasó cerca de Sgr A*, la cual es por sí misma una fuente de emisión de radiación. Esto quiere decir que estos datos en particular podrían ser mediciones de otros objetos o de las emisiones de Sgr A*.

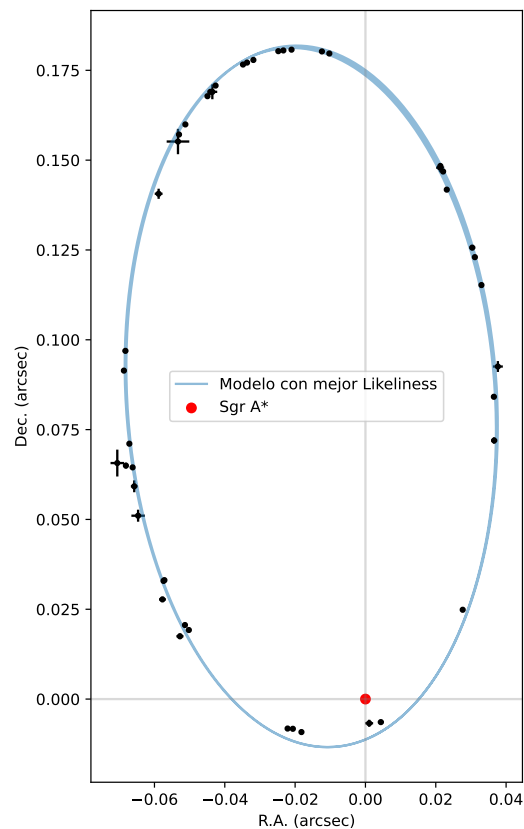


Figura 5.6: Órbita del mejor modelo encontrado comparado con las posiciones de S2 en el cielo en el sistema coordenado de ascensión recta (R.A.) y declinación (Dec.). Se muestra a Sgr A* en el origen del sistema coordenado; los puntos negros representan las posiciones observadas y su respectiva incertidumbre.

Además de la órbita del mejor modelo, también se pueden graficar las posiciones simuladas en el eje x (R.A.) y en el eje y (Dec.) en función de las fechas de la observaciones, lo cual se muestra en las Figs. 5.7 y 5.8. Son precisamente estas gráficas las que nos aseguran que el par de valores obtenidos (M, l) es el que describe al sistema porque la órbita simulada de la estrella pasa por los puntos de las observaciones en las fechas correspondientes a los datos.

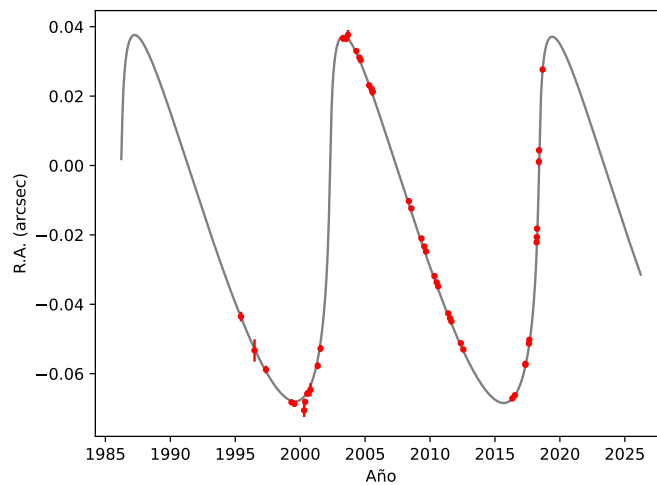


Figura 5.7: Valores de R.A. (*right ascension*, que corresponde al eje x en el plano del cielo centrado en Sgr A*) de la estrella en función de los años de observación. La línea continua representa la simulación de las posiciones con el mejor modelo.

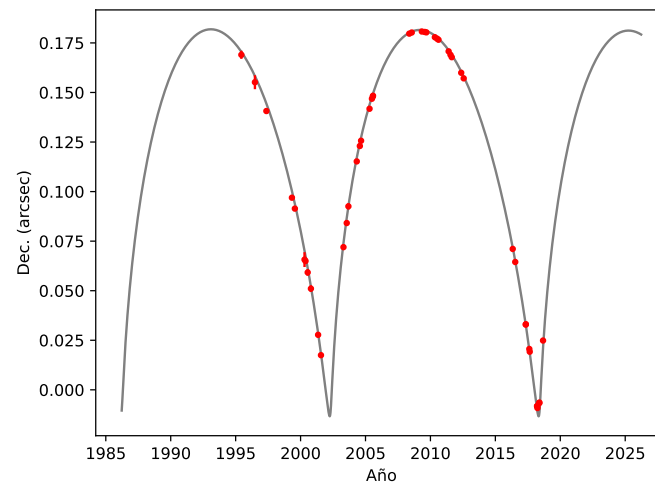


Figura 5.8: Valores de Dec. (*declination*, que corresponde al eje y en el plano del cielo centrado en Sgr A*) de la estrella en función de los años de observación. La línea continua representa la simulación de las posiciones con el mejor modelo.

Capítulo 6

Conclusiones

La implementación de métodos MCMC al análisis de órbitas estelares no es algo nuevo; de hecho, es una de las maneras más comunes de analizarlas. No obstante, en este trabajo se aprovecharon las simetrías del sistema para realizar un código que hiciera uso explícito de ellas. Una de las ventajas principales de esto radica en el uso de la ecuación de segundo grado para r . Esta derivación permite quitar el parámetro \mathcal{E} de todas las ecuaciones de movimiento, lo cual hace que ya no sea un parámetro a ajustar en nuestro modelo, a diferencia de otros trabajos [36].

Además, la integración de la órbita en el marco de referencia de Sgr A* evita la introducción de parámetros en el modelo como el semi eje mayor o la excentricidad, los cuales se han utilizado en otros trabajos [9] [10]. El obtener la órbita en el marco de referencia de Sgr A* y proyectarla por medio de las constantes de Thiele-Innes resulta más natural porque no se está imponiendo que la órbita sea elíptica. De esta manera, el proceso presentado en este trabajo no impone explícita o implícitamente ninguna característica de la órbita, dando así una completa libertad en las órbitas que se pueden obtener y una mayor certeza de los valores obtenidos. Si bien estos desarrollos imponen otras condiciones iniciales necesarias, i.e., ϕ_0 y t_0 , éstas pueden ser deducidas del mismo conjunto de datos, como se comentó en la Subsec. 5.1.3.

Otra ventaja interesante de este método es su versatilidad y el provecho que se le puede sacar. Notemos que el código sólo requiere de ecuaciones diferenciales de movimiento y la ecuación que conecte el tiempo propio con el tiempo coordinado; esto quiere decir que siempre y cuando se puedan obtener éstas, el método va a poder seguir siendo utilizado. Por ejemplo, un siguiente paso obvio sería trabajar con un espacio-tiempo más general como sería Kerr [24], o incluso se podría ir más allá e intentar ajustar la órbita a espacio-tiempos más exóticos predichos por teorías alternativas de gravedad. Visto de esta forma, este trabajo presenta los fundamentos necesarios para poder explorar

alternativas a la naturaleza de Sgr A*. Este trabajo se debe pensar como guía explícita de todo lo necesario para implementar métodos MCMC al análisis de órbitas y, además, su aplicación directa en el espacio-tiempo de Schwarzschild aprovechando todas las simetrías del mismo.

Es importante resaltar que este método no es un *fitter*¹, sino un *sampler*. Esto quiere decir que no busca el mejor conjunto de parámetros para describir la órbita, sino que obtiene la distribución que los describe. Esto es crucial porque indica que para describir los datos uno puede tener un porcentaje de certeza de que los parámetros están dentro de cierto rango (test de σ); este hecho se puede utilizar cuando se trabaja con espacio-tiempos más exóticos para imponer constricciones en los valores de sus parámetros.

Incluso con las ventajas que presenta, no se puede olvidar que la calidad de los resultados está fuertemente atada al hardware en el cual se corre el código. Un claro ejemplo de lo anterior es que el equipo utilizado para los resultados de este trabajo era una laptop, la cual fue llevada al límite de rendimiento durante las corridas y aún así no se obtuvieron las correlaciones esperadas entre ciertos parámetros. Dicho de otra manera, si se quisieran resultados más precisos y con una menor desviación estándar, entonces se debería recurrir a un equipo más potente o directamente a un clúster. Empero, es interesante observar la calidad de los resultados obtenidos en este trabajo incluso con la limitación del equipo, la cual se refleja en la Fig. 5.6 y en las diferencias porcentuales de la Tabla 5.3. Además de poder explorar más espacio-tiempos, el siguiente paso desde el punto de vista computacional sería correr el código en un clúster y comparar los resultados obtenidos de esa forma con los que se obtuvieron en este trabajo.

Otra simplificación que se utilizó en el desarrollo de este trabajo fue suponer que Sgr A* permanece inmóvil en el plano del cielo y además se despreció cualquier efecto que pueda causar la velocidad relativa entre nuestro marco de referencia y el centro galáctico. Se ha observado que Sgr A* presenta un leve movimiento, también llamado *drift*, el cual tendría que ser considerado. Cabe aclarar que estos movimientos son muy pequeños, del orden de 0.005 mas/año [9]. De esta manera, el no tomarlos en cuenta en nuestro desarrollo no perjudica los resultados obtenidos, y el tomarlos en cuenta resultaría en un refinamiento pequeño.

Ahora, la naturaleza estocástica del método hace que los valores del mejor modelo no sean reproducibles en un sentido determinista; incluso otra corrida usando el mismo código con los mismos valores de priors y límites para los parámetros puede arrojar valores diferentes para el nuevo mejor modelo que se encuentre. El origen de esto es precisamente la naturaleza estocástica del método. Lo que sí es reproducible es la forma de la distribución posterior de cada parámetro; con diferentes corridas se obtienen distribuciones gaussianas con la misma varianza pero con un valor

¹Un ejemplo clásico de un *fitter* sería el método de mínimos cuadrados para ajustar puntos a una línea recta.

medio que puede estar ligeramente desplazado (de nueva cuenta esto viene de la parte estocástica). Lo que no sucede es que tras una nueva corrida, los valores medios de las distribuciones normales o los valores para el mejor modelo estén considerablemente alejados a los que se encontraron en una corrida anterior.

Finalmente, se podría llegar a argumentar que el uso de la RG en el estudio de las órbitas de estrellas tipo S, como la estrella analizada en este trabajo, es innecesario y que bastaría con un análisis Newtoniano del sistema. Sin embargo, en los últimos años, y tras mediciones y métodos de observación cada vez más precisos, se ha podido detectar una combinación de corrimiento al rojo relativista y gravitacional, con el cual se puede excluir un modelo Newtoniano con una significación estadística de 5σ [10]. Además, se ha podido detectar una precesión de la órbita de S2 que es totalmente consistente con el modelo de Schwarzschild [9]. De esta manera, podemos ver que la evidencia apunta a que un análisis Newtoniano sería inadecuado porque dicho modelo está ya desechado en este régimen y lo que se debe estudiar son los límites y la validez de la RG.

Índice de figuras

1.1. Vista de los 0.25 pc centrales del centro galáctico. El punto más oscuro en la imagen corresponde a la fuente Sgr A*. Imagen tomada de [12].	11
1.2. Posiciones astrométricas de siete estrellas en el centro galáctico tomadas entre 1995 y 2003. Figura tomada de [15].	12
1.3. Posiciones de la estrella S2 observadas entre 1995 y 2018. Cada posición tiene su incertidumbre asociada y en el centro del sistema coordinado se marca a Sgr A*. Figura realizada con los datos proporcionados en [10].	13
2.1. Acción de los mapeos dados en la Definición 2.1.1. Los conjuntos U_α, U_β son abiertos en \mathbb{R}^n , mientras que los conjuntos O_α, O_β son conjuntos de puntos de la variedad. Imagen tomada de [19].	16
2.2. Paralelos sobre la esfera. Imagen tomada de [21].	29
2.3. Circuito cerrado sobre la variedad definido por valores constantes de las coordenadas. Imagen tomada de [21].	31
3.1. Experimento pensado sobre el papel de la gravedad y las trayectorias de partículas. Las trayectorias de dos partículas alejadas en presencia de un campo gravitacional dejan de ser paralelas. Imagen tomada de [21].	43
3.2. La curva \mathcal{L} representa la línea de mundo de una partícula masiva, por lo que siempre debe permanecer dentro del cono de luz en cada punto. Por otro lado, la curva \mathcal{L}' es la línea de mundo de una partícula sin masa, haciendo que ésta siempre esté a lo largo del cono de luz. Imagen tomada de [20].	44

3.3. Potenciales efectivos para distintos valores del cociente l/M ; además, se presenta el potencial Newtoniano con $l/M = 4.3$. El caso relativista presenta una barrera de altura finita, mientras que el caso clásico tiene una barrera infinita. Cabe resaltar que las gráficas están con unidades naturales ($G = c = 1$).	56
3.4. Representaciones de los potenciales y órbitas para los casos 2 y 4. La fila superior muestra primero el potencial efectivo para el caso $l/M = 4.3$ y $\mathcal{E} = -0.026$ y a la derecha la órbita descrita por una partícula en dichas condiciones; se puede observar que la partícula se mantiene en una órbita (en negro) acotada por los puntos de retorno $r_{min} = 10M$ (en azul) y $r_{max} = 25.6M$ (en naranja). Por otro lado, la segunda fila representa el caso del potencial de una órbita circular estable localizada en $r = 6M$ cuando $l/M = \sqrt{12}$	57
3.5. Diagrama $t - r$ que muestra los conos de luz de partículas en caída radial hacia $r = 2GM/c^2$. Se puede observar que al acercarse a dicho punto los conos de luz se comienzan a cerrar, implicando que cada vez les toma más tiempo coordenado caer. Imagen tomada de [22].	61
3.6. Diagrama de las trayectorias nulas radiales en el plano (v, r) . Se puede observar que lejos de r_S los conos de luz se abren a 45° , pero conforme se acercan comienzan a inclinarse. Imagen tomada de [2].	64
4.1. Simulación de la caída de gotas sobre un cuadrado con un círculo inscrito para aproximar el valor de π . La relación entre el número total de gotas que cayeron y las que cayeron dentro del círculo se relaciona con el valor de π . Imagen tomada de [28].	72
4.2. Gráfica de transición de una variable aleatoria con tres posibles estados. Los nodos representan los estados posibles de la variable y las conexiones entre nodos las probabilidades de moverse entre dichos estados. Imagen tomada de [31].	78
4.3. Diagrama de flujo del proceso de selección de un nuevo conjunto de parámetros θ del modelo.	81
5.1. Elementos para la orientación y proyección de la órbita real a la órbita aparente sobre el plano del cielo. Se muestran los diferentes ángulos que permiten orientar tanto el plano de la órbita como la órbita misma sobre éste.	84
5.2. Posiciones de la estrella S2 observadas en el cielo desde 1995 hasta 2018. Las posiciones en el cielo están dadas en segundos de arco (as) y además se muestra la incertidumbre de cada medición.	87

5.3. Interpolación de la función $\sin(x)$ en dos casos: con 10 valores simulados, y con 20 valores simulados. En el primero es notorio que la interpolación lineal no es ideal; por otro lado, en el segundo caso se puede observar que la interpolación produce una gráfica que se asemeja más a la función analítica en todo el dominio considerado. 88

5.4. Diagrama de flujo del proceso para comparar las posiciones integradas con las observadas utilizando una interpolación de las primeras. 89

5.5. Proyecciones de las distribuciones posteriores de los valores de los parámetros. Las líneas punteadas muestran los percentiles de 16 y 84, por lo que la región entre éstas representa 2σ 93

5.6. Órbita del mejor modelo encontrado comparado con las posiciones de S2 en el cielo en el sistema coordenado de ascensión recta (R.A.) y declinación (Dec.). Se muestra a Sgr A* en el origen del sistema coordenado; los puntos negros representan las posiciones observadas y su respectiva incertidumbre. 95

5.7. Valores de R.A. (*right ascension*, que corresponde al eje x en el plano del cielo centrado en Sgr A*) de la estrella en función de los años de observación. La línea continua representa la simulación de las posiciones con el mejor modelo. 96

5.8. Valores de Dec. (*declination*, que corresponde al eje y en el plano del cielo centrado en Sgr A*) de la estrella en función de los años de observación. La línea continua representa la simulación de las posiciones con el mejor modelo. 97

Bibliografía

- [1] Isaac Newton. *Philosophiae Naturalis Principia Mathematica*. 1687.
- [2] A. Zee. *Einstein Gravity in a nutshell*. Princeton University Press, 2013.
- [3] R. Fitzpatrick. *An Introduction to Celestial Mechanics*. Cambridge University Press, 2012.
- [4] A. Einstein. Die Grundlage der allgemeinen Relativitätstheorie. *Annalen der Physik*, 7:769–822, 1916.
- [5] A. Einstein. Zur Elektrodynamik bewegter Körper. *Annalen der Physik*, 10:891–921, 1905.
- [6] Karl Schwarzschild. On the gravitational field of a mass point according to Einstein’s theory. *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)*, 1916:189–196, 1916.
- [7] F. W. Dyson, A. S. Eddington, and C. Davidson. A Determination of the Deflection of Light by the Sun’s Gravitational Field, from Observations Made at the Total Eclipse of May 29, 1919. *Phil. Trans. Roy. Soc. Lond. A*, 220:291–333, 1920.
- [8] A. Einstein. The foundation of the general theory of relativity. *Annalen Phys.*, 49(7):769–822, 1916.
- [9] R. Abuter et al. [GRAVITY]. Detection of the Schwarzschild precession in the orbit of the star S2 near the Galactic centre massive black hole. *Astron. Astrophys*, 636:L5, 2020.
- [10] Tuan Do et al. Relativistic redshift of the star S0-2 orbiting the Galactic center supermassive black hole. *Science*, 365(6454):664–668, 2019.
- [11] J. Michell. On the Means of Discovering the Distance, Magnitude, c. of the Fixed Stars, in Consequence of the Diminution of the Velocity of Their Light, in Case Such a Diminution Should be Found to Take Place in any of Them, and Such Other Data Should be Procured from Observations, as Would be Farther Necessary for That Purpose. By the Rev. John Michell, B.

- D. F. R. S. In a Letter to Henry Cavendish, Esq. F. R. S. and A. S. *Philosophical Transactions of the Royal Society of London*, 74:35–57, 1784.
- [12] M. J. Reid. Is there a supermassive black hole at the center of the milky way? *International Journal of Modern Physics D*, 18(06):889–910, jun 2009.
- [13] B. Balick and R. L. Brown. Intense sub-arcsecond structure in the galactic center. , 194:265–270, December 1974.
- [14] L. F. Rodriguez and E. J. Chaisson. The temperature and dynamics of the ionized gas in the nucleus of our galaxy. *Astrophysical Journal*, 228(1):734–739, march 1979.
- [15] A. M. Ghez, S. Salim, S. D. Hornstein, A. Tanner, J. R. Lu, M. Morris, E. E. Becklin, and G. Duchene. Stellar orbits around the galactic center black hole. *The Astrophysical Journal*, 620(2):744–757, feb 2005.
- [16] R. M. Wald. *General Relativity*. The University of Chicago Press, 1984.
- [17] S. Chandrasekhar. *The Mathematical Theory of Black Holes*. International series of monographs on physics. Oxford University Press, Oxford, 1983.
- [18] C. Misner, K. Thorne, and J. Wheeler. *Gravitation*. W. H. Freeman and Company, 1973.
- [19] V. Frolov and A. Zelnikov. *Introduction to Black Hole Physics*. Oxford University Press, 2011.
- [20] É.ourgoulhon. *Geometry and physics of black holes: Lecture notes*. 2023.
- [21] S. Ramos Sánchez. *Relatividad para futuros físicos*. Prensa de Ciencias, UNAM, 2018.
- [22] B. Schutz. *A First course un General Relativity*. Cambridge University Press, 2009.
- [23] Vila G. Romero, G. *Introduction to Black Hole Astrophysics*. Springer, 2014.
- [24] Roy P. Kerr. Gravitational field of a spinning mass as an example of algebraically special metrics. *Phys. Rev. Lett.*, 11:237–238, Sep 1963.
- [25] P. Chrusciel. *Geometry of Black Holes*. International series of monographs on physics. Oxford University Press, Oxford, 2020.
- [26] Valenzuela M. Markov chains and applications. *SELECCIONES MATEMÁTICAS*, 9(1):53–78, 2022.

-
- [27] Sanjib Sharma. Markov Chain Monte Carlo Methods for Bayesian Data Analysis in Astronomy. *Annual Review of Astronomy and Astrophysics*, 55:213–259, 2017.
- [28] A. M. Johansen. *International Encyclopedia of Education: Monte Carlo Methods*. Elsevier, 2010.
- [29] J. Stewart. *Cálculo de una variable: Trascendentes tempranas*. Cengage Learning, 2018.
- [30] D. V. Schroeder. *An Introduction to Thermal Physics*. Oxford University Press, 2021.
- [31] C. Andrieu, N. de Freitas, A. Doucet, and et al. An Introduction to MCMC for Machine Learning. *Machine Learning*, 50:5–43, 2003.
- [32] S. Richardson W. R. Gilks and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Springer Science+Business Media, 1996.
- [33] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.
- [34] F. Eisenhauer, R. Schdel, R. Genzel, T. Ott, M. Tecza, R. Abuter, A. Eckart, and T. Alexander. A geometric determination of the distance to the galactic center. *The Astrophysical Journal*, 597(2):L121–L124, oct 2003.
- [35] A. M. Ghez, S. Salim, N. N. Weinberg, J. R. Lu, T. Do, J. K. Dunn, K. Matthews, M. R. Morris, S. Yelda, E. E. Becklin, T. Kremenek, M. Milosavljevic, and J. Naiman. Measuring distance and properties of the milky way’s central supermassive black hole with stellar orbits. *The Astrophysical Journal*, 689(2):1044–1062, dec 2008.
- [36] Parth Bambhaniya, Ashok B. Joshi, Dipanjan Dey, Pankaj S. Joshi, Arindam Mazumdar, Tomohiro Harada, and Ken ichi Nakao. Relativistic orbits of s2 star in the presence of scalar field. 2022.