



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

ANÁLISIS Y DESCRIPCIÓN DE LA DISTRIBUCIÓN
ESPECTRAL DE ENERGÍA DEL BLÁZAR W COMAE

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

LICENCIADO EN CIENCIAS DE LA COMPUTACIÓN

P R E S E N T A :

JUAN CARLOS MONTERO SANTIAGO

TUTOR

DR. NISSIM ILLICH FRAIJA CABRERA



CIUDAD UNIVERSITARIA, Cd. Mx., 2019



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.

A mi madre

Agradecimientos

En primera instancia a todos los docentes que me acompañaron durante este proceso; a mi tutor el Dr. Nissim Fraija por apoyarme académicamente, por sus consejos y sabiduría; a mi amigo el Mtro. Antonio Galván, agradezco que me brindara su gran conocimiento técnico para realizar este trabajo; agradezco a ambos su ayuda durante toda mi estancia en el Instituto de Astronomía.

A mis padres por confiar, por los valores y principios que me han inculcado. Dedico este trabajo a mi madre María Josefa Santiago, pues es en mayor parte gracias a ella que logré llegar hasta aquí. También agradezco a todos los miembros de mi familia que siempre estuvieron ahí para apoyarme: a mi padre Francisco Montero, a mi hermana Nielsen Montero, a mis tíos Anabel Gutiérrez y Edson Santiago, a mis primas Nazaret y Ximena Santiago, y a mi abuela Felisa Perea.

Finalmente, a mis amigos de la vida y de la facultad, por el apoyo diario. A mis amigos Aarón Grajales, Alexis Rodríguez, Arturo Calva, Bernardo Barragán, Brandon Padilla, Carlos Aguilar, César Camacho, Daniel Gómez, Felipe Cova, Francisco Gámez, Gabriel Calderón, Héctor Santaella, Jaime Argueta, Jonathan González, Juan Garduño, Luis Bernabé, Luis Flores, Max Salazar, Rafael García, Rodrigo Alvar, Rudy Matías, Santiago Ley y Valeria Gómez por su apoyo constante.

Índice general

Agradecimientos	II
1. Introducción	1
2. Núcleos de galaxias activos	3
2.1. Estructura	3
2.2. Clasificación	4
2.3. Blázares	5
2.3.1. Modelo de la emisión de los blázares	6
2.3.2. Componente de baja frecuencia	7
2.3.3. Componente de alta energía	8
3. Descripción de los procesos radiativos	10
3.1. Componente de sincrotrón	10
3.2. Componente de <i>synchrotron self-Compton</i>	13
3.3. Ajustes Prácticos	13
4. Implementación	15
4.1. Descenso por el gradiente	15
4.1.1. Descripción	16
4.1.2. Limitaciones	16
4.1.3. Función de error	17
4.1.4. Vectorización	19
4.2. Optimización de descenso por el gradiente	20

4.2.1. Descenso por el gradiente con momento	20
4.2.2. <i>Adaptive Gradient Descent</i> o AdaGrad	21
4.2.3. <i>Root Mean Square Propagation</i> o RMSProp	23
4.2.4. <i>Adaptive Moment Estimation</i> o ADAM	23
4.3. Criterios de parada	24
4.4. Implementación del Flujo	25
4.4.1. Cython	25
4.5. Implementación del Ajuste.	27
5. Resultados	31
5.1. W Comae	31
5.1.1. Detección en rayos gamma	31
5.1.2. Clasificación e Importancia	32
6. Conclusiones	39
Bibliografía	41

Índice de figuras

2.1. Esquema del paradigma actual para los AGNs (no a escala). Rodeando el agujero negro del centro hay un brillante disco de acreción. Las líneas de emisión se producen en las nubes orbitando sobre el disco y tal vez también el disco en sí. Un toro de polvo oscurece las líneas de emisión para un punto de vista transversal. Los <i>jets</i> emanan de la región cercana al agujero negro, inicialmente a velocidades relativistas ([UP95]).	4
2.2. Ejemplo simple de un SED.	7
2.3. Comparación de los SEDs entre tres blázares diferentes. FSRQ = 3C 279 ([Har+01]), LBL = BL Lacertae ([Böt+03]) y HBL = Mrk 501 ([Pet+00])	8
4.1. Visualización del método de descenso por el gradiente. En este ejemplo, se quiere aproximar una recta que pasa por el origen, por lo tanto solo se tiene un parámetro: la pendiente θ_1 . En la derecha está la aproximación obtenida en cada iteración junto con el valor de θ_1 , marcado con puntos negros está el objetivo. A la izquierda se muestra la función de error para todos los parámetros posibles, y marcados con puntos están los valores del error en cada iteración. Se puede observar claramente cómo el método se hace más lento al acercarse al mínimo.	17
4.2. Comparación de tiempos de ejecución utilizado para obtener el n-ésimo número de la serie de Fibonacci entre <i>Cython</i> y <i>Python</i>	26

5.1. Posición aproximada de W Comae en coordenadas galácticas	32
5.2. Mapa del cielo visto por VERITAS al detectar W Comae a muy altas energías ([Acc+08]).	33
5.3. Ronda 1 de ajuste. Iteraciones totales: 59. El ajuste convergió a un mínimo local con un error de ≈ 0.03 . Parámetros finales: $\delta_D = 37.13$, $B = 42.7$, $\log(K_e) = 42.72$, $p = 2.318$, $\log(\gamma_{break}) = 4.59$, $\log(\gamma_2) = 6.614$	34
5.4. Ronda 2 de ajuste. Iteraciones totales: 31. El ajuste convergió a un mínimo local con un error de ≈ 0.056 . Parámetros finales: $\delta_D = 36.57$, $B = 18.12$, $\log(K_e) = 42.569$, $p = 2.262$, $\log(\gamma_{break}) = 4.92$, $\log(\gamma_2) = 6.24$	35
5.5. Ronda 3 de ajuste. Iteraciones totales: 91. El ajuste convergió a un mínimo local con un error de ≈ 0.015 . Parámetros finales: $\delta_D = 20.225$, $B = 71.11$, $\log(K_e) = 44.25$, $p = 2.15$, $\log(\gamma_{break}) = 4.322$, $\log(\gamma_2) = 5.49$.	36
5.6. Ronda 4 de ajuste. Iteraciones totales: 86. El ajuste convergió a un mínimo local con un error de ≈ 0.0086 . Parámetros finales: $\delta_D = 34.752$, $B = 85.85$, $\log(K_e) = 44.1$, $p = 2.25$, $\log(\gamma_{break}) = 3.964$, $\log(\gamma_2) = 5.541$.	37
5.7. Ronda de ajuste de SSC. Iteraciones totales: 112. El ajuste convergió a un mínimo local con un error de ≈ 0.041 . Parámetros finales: $\log(t_{v,min}) = 3.78$	38
5.8. SED final para el blazar W Comae	38

Índice de tablas

3.1. Coeficientes para la aproximación de $R(x)$	12
3.2. Comparación entre la expresión exacta de $R(x)$ y la aproximación utilizada	12
5.1. Rangos definidos para cada parámetro. El parámetro inicial se escoge aleatoriamente dentro de estos rangos.	34

Acrónimos

AGN	<i>Active Galactic Nuclei.</i>
BL Lac	BL Lacertae.
EC	<i>External Compton.</i>
FSRQ	<i>Flat Spectrum Radio Quasars.</i>
HESS	<i>High Energy Stereoscopic System.</i>
MAGIC	<i>Major Atmospheric Gamma-ray Imaging Cherenkov Telescope.</i>
SED	<i>Spectral Energy Distribution.</i>
SSC	<i>Synchrotron Self-Compton.</i>
VERITAS	<i>Very Energetic Radiation Imaging Telescope Array System.</i>
VHE	<i>Very High Energies.</i>
VLBI	<i>Very Long Baseline Interferometry.</i>

Capítulo 1

Introducción

Desde la segunda mitad del siglo XX, los núcleos de galaxias activos han sido observados por los astrónomos del mundo. Su enorme luminosidad y características inusuales en sus espectros indican que su fuente de poder no podían ser estrellas ordinarias. En las décadas de los 60s y 70s, observaciones de la recién nacida Astronomía de Rayos-X, demostraron que varias galaxias eran grandes fuentes de radiación de rayos-X y que esta se origina en las regiones internas del disco de acreción de los agujeros negros. Hoy en día, los AGN son un gran tema de investigación astrofísica, tanto teórica como observacional. Una forma de estudiar un objeto de este tipo es con la distribución espectral de energía, con la cual podemos observar a simple vista la densidad de la energía emitida por un objeto, separándose en *bandas* o rangos de energía, con esta gráfica podemos apreciar qué proporción de la energía emitida por nuestro objeto pertenece a cierto tipo; por ejemplo, qué proporción de la emisión corresponde a rayos-X o rayos gamma. Centraremos nuestros esfuerzos en un tipo específico de núcleos de galaxias activos conocidos como *blázares*, expondremos sus características e importancia.

Actualmente, para analizar las características de estos objetos, los investigadores recurren a ajustes arbitrarios con métodos como el ajuste polinomial para ajustar los datos observados por satélites y observatorios. Este tipo de ajustes son suficientes a la hora de describir el comportamiento matemático del espectro, pero nos dejan sin información de las características físicas del objeto en sí y de su entorno. Existen ya

varias propuestas sobre la naturaleza física de las emisiones de los blázares, todas se tratan de ecuaciones bastante complejas y que casi siempre incluyen integrales no elementales. Desde aquí se puede ver ya una oportunidad para las ciencias de la computación de contribuir en este campo, ya que estas ecuaciones casi siempre son resueltas usando métodos de integración numérica.

Sin embargo, existe otra forma en la que las ciencias de la computación pueden agregar valor al estudio de los núcleos de galaxias activos; encontrando un método de ajuste que utilice los modelos teóricos físicos para ajustar la distribución espectral de energía de un blazar. De esta forma al finalizar el ajuste no solo tendremos información de su comportamiento matemático, sino también algunas características físicas del objeto.

Se han publicado los resultados de otros intentos de este tipo; pero en las publicaciones correspondientes, el foco de la investigación está centrado en los modelos físicos y no en el algoritmo de ajuste, incluso acabando sin mencionar el método utilizado. Así que el objetivo de este trabajo es primero: definir y explorar un método de ajuste y sus posibles optimizaciones, y segundo: exponer los parámetros físicos que podemos obtener de estos ajustes. También pondremos atención a que los resultados sean adecuados y que tengan sentido físico, esto con el objetivo extra de aportar a la participación de las ciencias de la computación en el estudio de la astrofísica.

Para hacer más puntuales nuestros objetivos, elegimos el blazar conocido como W Comae. Este objeto es un blazar que fue detectado desde 1971 y cuyas emisiones fueron detectadas de nuevo en 1997, 1998 y 2008. Los datos de estas observaciones son fáciles de encontrar y su distribución espectral de energía tiene una forma muy característica en el resto de blázares. Por esto, lo consideramos un sujeto de prueba ideal para el desarrollo de un algoritmo de ajuste que pueda ser aplicado a cualquier otro blazar.

Capítulo 2

Núcleos de galaxias activos

Los núcleos de galaxias activos o AGNs producen luminosidades muy altas en volúmenes muy concentrados a través de procesos físicos muy diferentes a la fusión nuclear que da poder a las estrellas, su entendimiento es esencial para estudiar el origen y la evolución del universo.

2.1. Estructura

La estructura de los AGNs es conocida de la siguiente manera ([UP95]): en su centro existe un agujero negro súper masivo cuya energía potencial gravitatoria es la fuente de la luminosidad del AGN, el gas que es empujado hacia el agujero negro se asienta en un disco de acreción y se calienta, éste brilla fuertemente en longitudes de onda ultravioletas y tal vez de Rayos X.

Fuertes líneas de emisión óptica y ultravioleta se producen en nubes de gas moviéndose rápidamente en el potencial del agujero negro. La radiación óptica y ultravioleta es oscurecida a lo largo de algunas líneas de visión por un toro o un disco deformado de gas y polvo. Más allá del toro, nubes de gas más lentas producen líneas de emisión con anchos más angostos. Casi en el borde del agujero negro, a través de procesos aún no bien conocidos, una parte del gas sale disparado fuera del disco de acreción formando así chorros o *jets* en direcciones opuestas a velocidades cercanas a la de la

luz. El plasma en el jet, al menos en escalas pequeñas, fluye hacia afuera a velocidades muy altas, emitiendo radiación relativista en dirección hacia adelante.

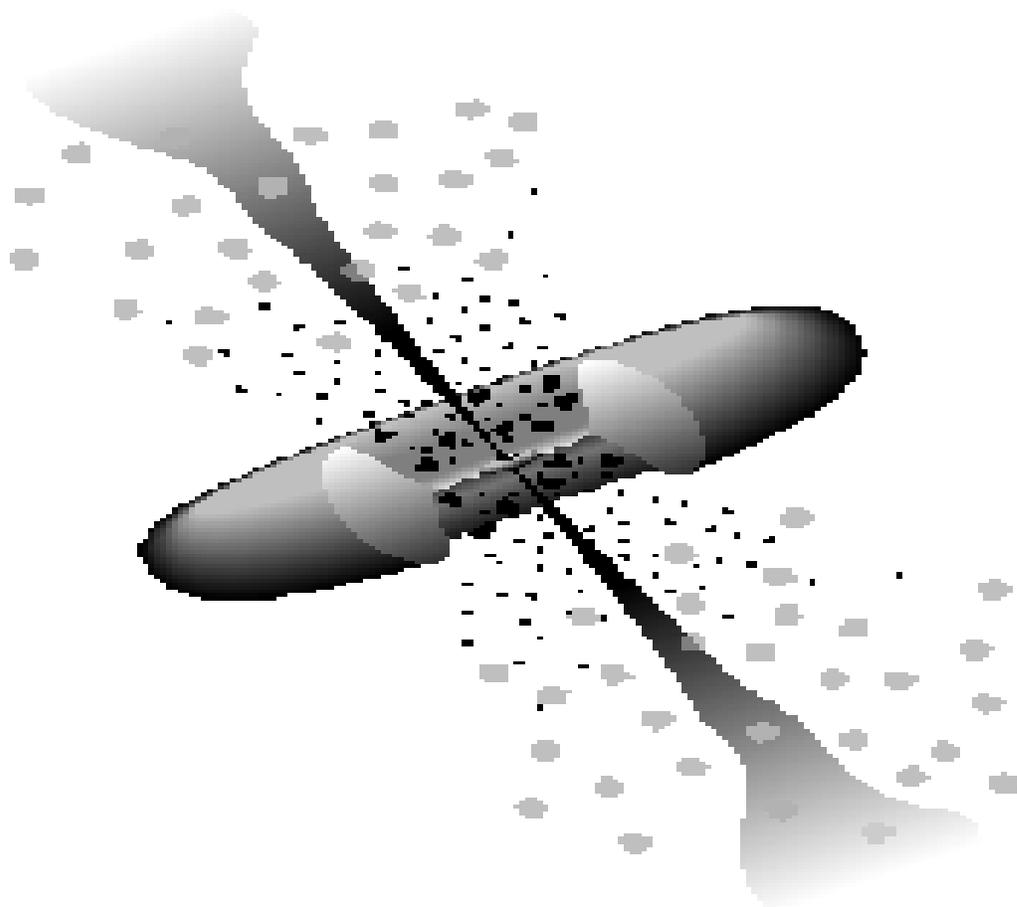


Figura 2.1: Esquema del paradigma actual para los AGNs (no a escala). Rodeando el agujero negro del centro hay un brillante disco de acreción. Las líneas de emisión se producen en las nubes orbitando sobre el disco y tal vez también el disco en sí. Un toro de polvo oscurece las líneas de emisión para un punto de vista transversal. Los *jets* emanan de la región cercana al agujero negro, inicialmente a velocidades relativistas ([UP95]).

2.2. Clasificación

La clasificación general de los AGNs es muy amplia, sin embargo aquí nos interesan especialmente un pequeño grupo de AGNs que tienen características espectrales muy inusuales. Se especula que estos están relacionados por un ángulo pequeño a la línea de visión del observador (casi 0 grados). Estos incluyen los objetos BL Lacertae

o simplemente BL Lac que son AGNs muy ruidosos en radio pero carecen de características de emisión o absorción fuertes. Estos objetos están también caracterizados por una variabilidad muy rápida, polarización inusualmente alta y variable, temperaturas de alto brillo y velocidades superluminales de núcleos con radio compacto. Otros objetos que comparten estas últimas características son llamados cuántares de radio de espectro plano o FSRQ. Colectivamente, los objetos BL Lac y FSRQ son conocidos como **blázares** (que es una combinación de ambos nombres).

2.3. Blázares

En imágenes a longitudes de onda visibles, casi todos los blázares parecen compactos y puntuales, pero gran cantidad de observaciones de alta resolución revelan que estos se encuentran exclusivamente en el centro de galaxias elípticas. Las galaxias huésped de estos objetos son grandes, y a su vez las más grandes tienden a contener los más luminosos. Estas galaxias huésped muestran perfiles de color normales; los colores integrados son consistentes con emisiones cuyo corrimiento al rojo corresponden a una galaxia elíptica evolucionando pasivamente con una población estelar vieja e implica una época de formación estelar inicial de al menos hace ~ 6 billones de años ([Urr+99]).

Las galaxias más grandes son resultado de colisiones y mezclas con muchas galaxias más pequeñas, proceso que ocurría con mucha mayor frecuencia más atrás en el tiempo. Estas colisiones proveían a la creciente galaxia de suficiente gas, de modo que éste alcanzaba más fácilmente el agujero negro de su centro, donde se acumuló en enormes y brillantes discos de acreción como los observados en los FSRQ. Parte del gas da lugar a un jet, mientras que el resto cae en el agujero negro aumentando gradualmente su *spin*.

Al expandirse el universo la densidad de galaxias decreció, y a su vez la frecuencia de estas colisiones. El disco de acreción se ve disminuido con el tiempo, pero lo que queda está orbitando un agujero negro más masivo y con una rotación mucho más

rápida. Estas propiedades permiten a los objetos BL Lac mantener un poderoso jet aun con una cantidad relativamente pobre de material orbitando su agujero negro.

Se podría decir que el agujero negro de un blazar actúa como una batería, ya que la energía resultante de la acreción cuando la galaxia era un FSRQ, se mantenía en forma de la rotación incrementada de un agujero negro más masivo. Cuando el disco de acreción ha casi desaparecido, el blazar hace uso de la energía guardada que le permite continuar operando su jet de partículas y produciendo emisiones de altas energías como un objeto BL Lac ([Fra14]).

2.3.1. Modelo de la emisión de los blázares

La manera en que los blázares producen sus emisiones es un objeto de estudio bastante amplio por sí solo, obtener un modelo físico de esto es esencial para lograr el objetivo propuesto, por lo que en esta sección se expone uno de los modelos más aceptados de estos procesos de emisión.

En base a este modelo teórico se pueden obtener ecuaciones matemáticas que serán el ingrediente principal de este estudio.

Distribución espectral de energía

La distribución espectral de energía (en inglés *Spectral Energy Distribution* o SED), es una gráfica de la energía emitida por un objeto en función de diferentes longitudes de onda de la luz (es decir, el eje X de la gráfica representa diferentes niveles o bandas de energía, estas pueden presentarse en unidades de energía o de frecuencia; mientras que el eje Y representa la densidad de la energía expresada en unidades de $\frac{erg}{cm^2s}$). Al graficar la energía emitida por un objeto astrofísico podemos comparar a simple vista la emisión de éste en las diferentes longitudes de onda; por ejemplo, podemos diferenciar claramente si la mayor parte de la energía viene de emisión ultravioleta o infrarroja. En otras palabras, el SED es la representación gráfica del comportamiento de la energía emitida por un objeto.

El espectro continuo de banda ancha de los blázares está dominado por emisiones

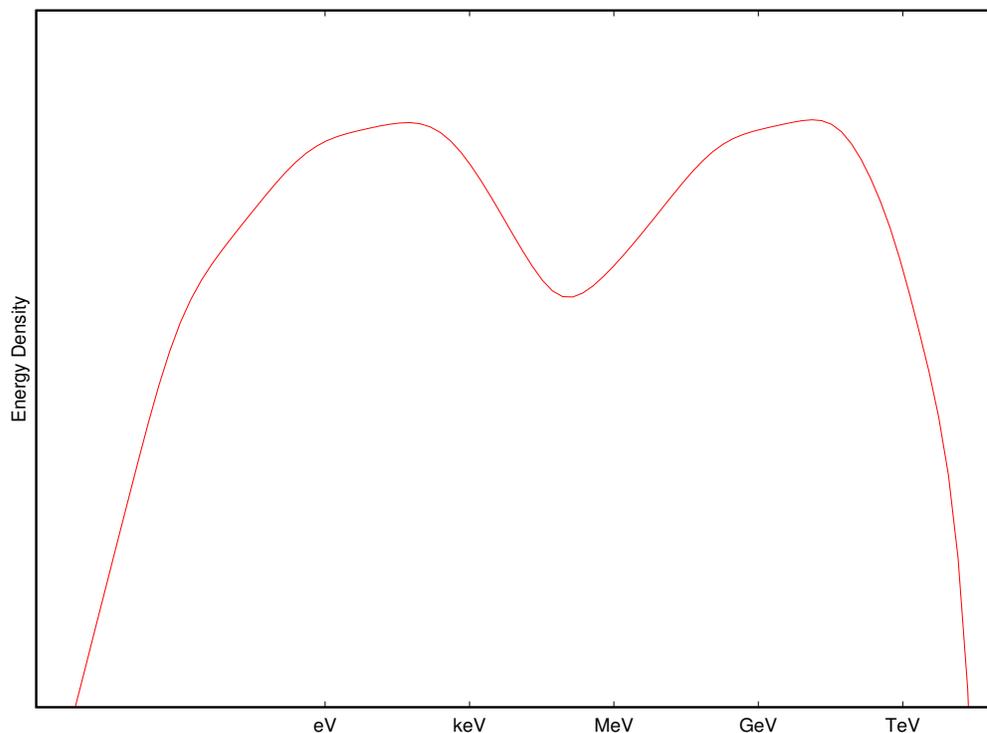


Figura 2.2: Ejemplo simple de un SED.

no térmicas y consiste en dos componentes distintos: un componente de baja energía (que abarca de radio a ultravioleta o rayos X) y un componente de alta energía (rayos X a rayos gamma). A través de la frecuencia más alta o los picos de los flujos $\nu F\nu$, se puede definir una secuencia de subclases de blázares, desde FSRQ a objetos BL Lac con pico en baja frecuencia o *low-frequency peaked BL Lac objects* (LBLs) a objetos BL Lac con pico en alta frecuencia o *high-frequency peaked BL Lac objects* (HBLs). Esta secuencia (FSRQ \rightarrow LBL \rightarrow HBL) se caracteriza, entre otras cosas, por el incremento del pico de la frecuencia de $\nu F\nu$ y un decreciente dominio del flujo en rayos gamma sobre la emisión de baja frecuencia ([Böt06]), (ver Figura 2.3).

2.3.2. Componente de baja frecuencia

Generalmente se considera que el componente de baja frecuencia de la distribución espectral de energía (SED) de los blázares es radiación sincrotrón de electrones ultrarelativistas no térmicos.

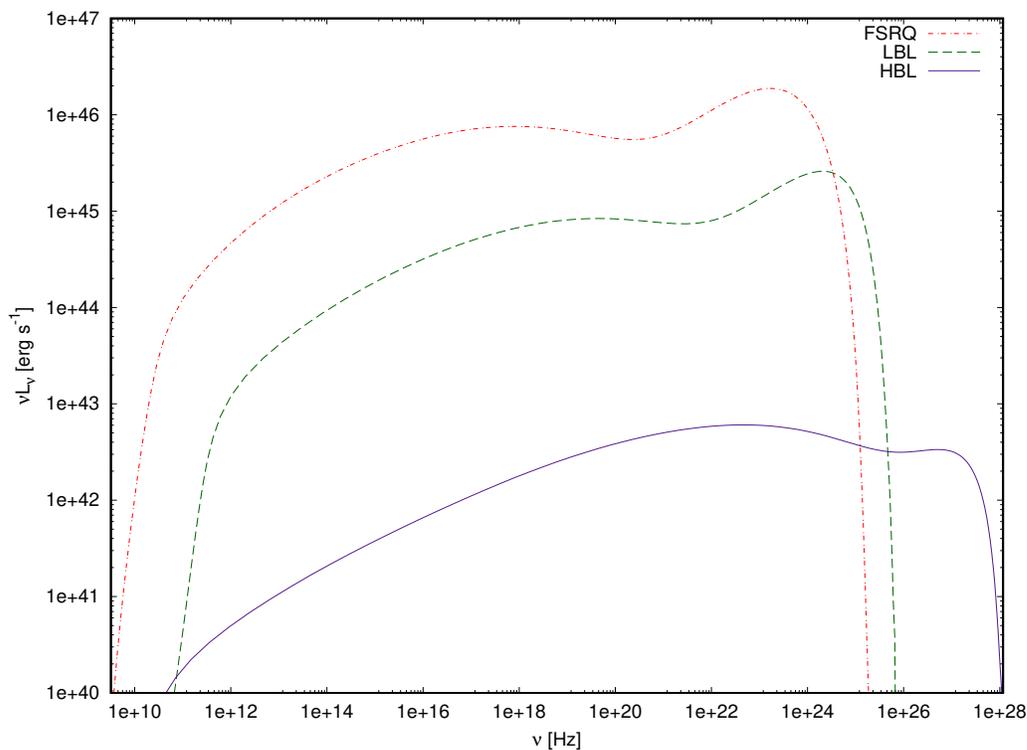


Figura 2.3: Comparación de los SEDs entre tres blázares diferentes. FSRQ = 3C 279 ([Har+01]), LBL = BL Lacertae ([Böt+03]) y HBL = Mrk 501 ([Pet+00])

Radiación sincrotrón

La radiación sincrotrón se refiere a la radiación electromagnética que emiten las partículas cargadas sujetas a una aceleración radial (o bien, aquella aceleración que es perpendicular a su velocidad) ([Eld+47]), en particular cuando dichas partículas son ultra-relativistas. Este fenómeno se produce cuando partículas de alta energía son aceleradas, específicamente electrones que son forzados a moverse en un camino curvo por un campo magnético. La emisión de sincrotrón fue detectada por primera vez en la astrofísica en un jet emitido por el objeto Messier 87 en 1956 ([Bur56]), aunque su papel en este campo había sido predicho desde 1950 ([AH50]).

2.3.3. Componente de alta energía

En cuanto a este componente, existen dos aproximaciones. Si los protones no son acelerados a energías ultra-relativistas, la radiación de alta energía estará dominada

por emisión Compton proveniente de electrones y/o pares ultra-relativistas. En el caso opuesto la emisión de alta energía estará dominada por cascadas iniciadas por pares $p\gamma$, producción de piones así como radiación sincrotrón de protones, π^\pm y μ^\pm .

En la descripción que usaremos aquí nos centraremos en el primer caso, conocido como el modelo leptónico, por lo tanto lo describiremos más a detalle a continuación.

Dispersión Compton inversa

En este modelo, la emisión de alta energía se produce a través de la dispersión Compton de los fotones blandos provenientes de la misma emisión de sincrotrón. Ambos, los fotones de sincrotrón producidos dentro del jet (SSC) y fotones externos (EC) son posibles objetivos para la dispersión Compton ([Böt12]).

La dispersión Compton, o más específico al caso de estudio, Compton Inverso, involucra la dispersión de un fotón de baja energía a una energía más alta por electrones ultrarelativistas, de forma que el electrón pierde energía (de aquí que se le llame inverso, ya que ocurre lo contrario que en la dispersión Compton estándar).

Capítulo 3

Descripción de los procesos radiativos

En este capítulo expondremos las ecuaciones necesarias para poder simular de manera matemática el SED de un blazar, empezando por su componente de baja energía y siguiendo con la componente de alta energía.

3.1. Componente de sincrotrón

Para simular la distribución espectral de energía de un blazar calcularemos (ver [FDB08]) el flujo energético de sincrotrón para una energía ϵ dada, con la ecuación:

$$f_{\epsilon}^{syn} = \frac{\sqrt{3}\delta_D^4 \epsilon' B}{4\pi h d_L^2} \int_1^{\infty} d\gamma' N'_e(\gamma') R(x) \quad (3.1)$$

donde δ_D es el factor de Doppler, B es la intensidad promedio del campo magnético con dirección aleatoria, d_L es la distancia de luminosidad, γ es el factor de Lorentz del fotón emitido y h es la constante de Planck.

Los valores en el marco de referencias del radio de emisión están primados, mientras que los que están en referencia al observador no lo están. En cuanto al resto de la ecuación, $N'_e(\gamma')$ es la distribución de electrones, dada por la siguiente ecuación:

$$N_e(\gamma') = K_e \left[\left(\frac{\gamma'}{\gamma'_{break}} \right)^{-p} H(\gamma'_{break} - \gamma') + \left(\frac{\gamma'}{\gamma'_{break}} \right)^{-(p+1)} H(\gamma' - \gamma'_{break}) \right] H(\gamma'; \gamma'_1, \gamma'_2) \quad (3.2)$$

donde $p, K_e, \gamma_1, \gamma_2$ y γ_{break} son constantes y $H(x)$ denota una función escalón de Heaviside.

$R(x)$ está dado por la ecuación:

$$R(x) = \frac{x}{2} \int_0^\pi d\theta \sin\theta \int_{x/\sin\theta}^\infty dt K_{5/3}(t) \quad (3.3)$$

donde $K_{5/3}$ es la función de Bessel modificada del segundo tipo de orden $5/3$ y

$$x = \frac{4\pi\epsilon' m_e^2 c^3}{3eBh\gamma'^2} \quad (3.4)$$

donde m_e es la masa del electrón, e es la carga fundamental y c es la velocidad de la luz.

La ecuación 3.3 puede ser integrada a

$$R(x) = \frac{1}{2} \pi x [W_{0, \frac{4}{3}}(x) W_{0, \frac{1}{3}}(x) - W_{\frac{1}{2}, \frac{5}{6}}(x) W_{-\frac{1}{2}, \frac{5}{6}}(x)] \quad (3.5)$$

donde $W_{\lambda, \mu}(x)$ denota la función de Whittaker ([CS86]).

La función $R(x)$ puede ser aproximada como:

$$\log(R) = A_0 + A_1 y + A_2 y^2 + A_3 y^3 + A_4 y^4 + A_5 y^5 \quad (3.6)$$

donde $y = \log(x)$. Esta aproximación tiene una precisión del 1% dentro del rango $x < 10^{-2}$ y $x > 10$. Fuera de este rango podemos usar las expresiones asintóticas de las funciones de Whittaker para valores muy pequeños y muy grandes

$$R(x) = \begin{cases} 1.80842x^{1/3} & x \ll 1 \\ \frac{\pi}{2} e^{-2} \left(1 - \frac{99}{162x}\right) & x \gg 1 \end{cases} \quad (3.7)$$

Coefficientes	$20^{-3} \leq x < 10^{-2}$	$10^{-2} \leq x < 10^{-1}$	$10^{-1} \leq x < 1$	$1 \leq x < 4$
A0	0.126382	0.247666	0.53561	0.947533
A1	66.5445	15.5655	1.89341	-0.688081
A2	-11132.9	-344.364	-6.87853	0.204715
A3	1.33967×10^6	4808.73	10.3096	-0.030625
A4	-8.92112×10^7	-35788.2	-7.72996	0.00227635
A5	2.45562×10^9	2,107178	2.31062	-6.68025×10^{-5}

Tabla 3.1: Coeficientes para la aproximación de $R(x)$

x	Aproximación	$R(x)$	x	Aproximación	$R(x)$
1.000×10^{-4}	8.393×10^{-2}	8.375×10^{-2}	1.268×10^{-1}	6.842×10^{-1}	6.844×10^{-1}
1.610×10^{-4}	9.838×10^{-2}	9.809×10^{-2}	2.043×10^{-1}	7.106×10^{-1}	7.113×10^{-1}
2.592×10^{-4}	1.153×10^{-1}	1.148×10^{-1}	3.290×10^{-1}	6.994×10^{-1}	6.986×10^{-1}
4.175×10^{-4}	1.351×10^{-1}	1.344×10^{-1}	5.298×10^{-1}	6.285×10^{-1}	6.292×10^{-1}
6.723×10^{-4}	1.584×10^{-1}	1.572×10^{-1}	8.531×10^{-1}	4.954×10^{-1}	4.957×10^{-1}
1.082×10^{-3}	1.856×10^{-1}	1.837×10^{-1}	1.373×10^0	3.169×10^{-1}	3.167×10^{-1}
1.743×10^{-3}	2.176×10^{-1}	2.144×10^{-1}	2.212×10^0	1.466×10^{-1}	1.454×10^{-1}
2.807×10^{-3}	2.499×10^{-1}	2.500×10^{-1}	3.562×10^0	3.804×10^{-2}	3.956×10^{-2}
4.520×10^{-3}	2.908×10^{-1}	2.908×10^{-1}	5.736×10^0	4.529×10^{-3}	4.667×10^{-3}
7.278×10^{-3}	3.373×10^{-1}	3.372×10^{-1}	9.236×10^0	1.428×10^{-4}	1.447×10^{-4}
1.172×10^{-2}	3.898×10^{-1}	3.895×10^{-1}	1.487×10^1	5.228×10^{-7}	5.257×10^{-7}
1.887×10^{-2}	4.468×10^{-1}	4.473×10^{-1}	2.395×10^1	6.073×10^{-1}	6.086×10^{-1}
3.039×10^{-2}	5.098×10^{-1}	5.094×10^{-1}	3.856×10^1	2.754×10^{-7}	2.757×10^{-7}
4.893×10^{-2}	5.730×10^{-1}	5.733×10^{-1}	6.210×10^1	1.664×10^{-7}	1.665×10^{-7}
7.880×10^{-2}	6.346×10^{-1}	6.344×10^{-1}	1.000×10^2	5.807×10^{-4}	5.808×10^{-4}

Tabla 3.2: Comparación entre la expresión exacta de $R(x)$ y la aproximación utilizada

fuera de $x < 10^{-2}$ y $x > 10$ esta expresión tiene un error de menos del 5%. Utilizaremos ambas expresiones con el fin de reducir el tiempo de cómputo. Los coeficientes de la aproximación se muestran en la tabla 3.1.

Comparando esta aproximación con la ecuación 3.5 encontramos que tiene un error absoluto promedio de 0.005, además de reducir el tiempo de cómputo considerablemente; los resultados de ambas ecuaciones pueden observarse en la tabla 3.2 para distintos valores de x .

3.2. Componente de *synchrotron self-Compton*

En cuanto a la segunda componente, utilizaremos la siguiente expresión del flujo para una energía dada ϵ_s ([FDB08]):

$$f_{\epsilon_s}^{SSC} = \frac{9}{16} \frac{(1+z)^2 \sigma_T \epsilon_s'^2}{\pi \delta_D^2 c^2 t_{v,min}^2} \int_0^\infty d\epsilon' \frac{f_\epsilon^{syn}}{\epsilon'^3} \int_{\gamma'_{min}}^{\gamma'_{max}} d\gamma' \frac{N'_e(\gamma')}{\gamma'^2} F_C(q, \Gamma) \quad (3.8)$$

donde z se refiere al corrimiento al rojo del objeto, σ_T es la sección eficaz de Thomson, δ_D el factor de Doppler y $t_{v,min}$ el tiempo de variabilidad.

En cuanto a $F_C(q, \Gamma)$, denota el *kernel* de dispersión en el régimen de Compton, y es ([FDB08]):

$$F_C(q, \Gamma_e) = \left[2q \ln q + (1+2q)(1-q) + \frac{1}{2} \frac{(\Gamma_e q)^2}{(1+\Gamma_e q)} (1-q) \right] H\left(q; \frac{1}{4\gamma'^2}, 1\right) \quad (3.9)$$

donde

$$q := \frac{\epsilon'_s / \gamma'}{\Gamma_e (1 - \epsilon'_s / \gamma')} \quad (3.10)$$

$$\Gamma_e = 4\epsilon' \gamma' \quad (3.11)$$

Los límites en q (i.e. $\frac{1}{4\gamma'^2} \leq q \leq 1$) implican los límites de integración sobre γ' :

$$\gamma'_{min} = \frac{1}{2} \epsilon'_s \left(1 + \sqrt{1 + \frac{1}{\epsilon' \epsilon'_s}} \right) \quad (3.12)$$

$$\gamma'_{max} = \frac{\epsilon' \epsilon'_s}{\epsilon' - \epsilon'_s} H(\epsilon' - \epsilon'_s) + \gamma'_2 H(\epsilon'_s - \epsilon') \quad (3.13)$$

3.3. Ajustes Prácticos

Ahora expondremos algunos de los ajustes que haremos a las expresiones del flujo, para poder reducir el tiempo de cómputo lo más que podamos. La mayoría de estos ajustes fueron obtenidos de forma analítica, observando las características de las

expresiones anteriores. Primero, en la expresión del flujo de sincrotrón (3.1), podemos notar que la función escalón Heaviside dentro de la expresión de la distribución de electrones (3.2) será igual a 0 para cualquier valor fuera del rango $\gamma'_1 \leq \gamma' \leq \gamma'_2$, por lo que podemos ignorar los valores de γ' fuera de este, dejándonos reducir el rango de integración.

$$f_\epsilon^{syn} = \frac{\sqrt{3}\delta_D^4 \epsilon' B}{4\pi h d_L^2} \int_{\gamma'_1}^{\gamma'_2} d\gamma' N'_e(\gamma') R(x) \quad (3.14)$$

De manera similar, podemos ajustar los límites de integración de la primera integral en la expresión (3.8), ya que la energía máxima para el valor ϵ puede ser expresada en términos del factor de Lorentz con la siguiente ecuación:

$$\gamma'_s = \sqrt{\frac{\epsilon(1+z)}{\delta_D \epsilon_B}} = \sqrt{\frac{\epsilon'}{\epsilon_B}} \quad (3.15)$$

donde $\epsilon_B = B/B_{cr}$, siendo $B_{cr} = 4.414 \times 10^{13} \text{ G}$ el campo magnético crítico. Así podemos despejar la máxima energía reemplazando γ'_s por γ'_2 , obteniendo un nuevo límite superior para la integral de la siguiente forma:

$$\epsilon'_{max} = \gamma_2'^2 \epsilon_B \quad (3.16)$$

De igual forma, podemos hacer más simples los límites de integración de la segunda integral en la ecuación, notando que $\gamma'_{max} \leq \gamma'_2$ por definición. Por lo tanto, la expresión del flujo para la componente de *synchrotron self-Compton* es:

$$f_{\epsilon_s}^{SSC} = \frac{9}{16} \frac{(1+z)^2 \sigma_T \epsilon_s'^2}{\pi \delta_D^2 c^2 t_{v,min}^2} \int_0^{\epsilon'_{max}} d\epsilon' \frac{f_\epsilon^{syn}}{\epsilon'^3} \int_{\gamma'_{min}}^{\gamma'_2} d\gamma' \frac{N'_e(\gamma')}{\gamma'^2} F_C(q, \Gamma) \quad (3.17)$$

Considerando todos estos ajustes, podemos encontrar que los parámetros a ajustar son el factor de Doppler (δ_D), el campo magnético (B), la constante K_e , el índice p , el factor de Lorentz de quiebre (γ_{break}), los factores de Lorentz máximo (γ_2) y mínimo (γ_1), y el tiempo de variabilidad ($t_{v,min}$). Asumimos un factor de Lorentz mínimo de $\gamma_1 = 10^2$, dejándonos con 7 parámetros a ajustar.

Capítulo 4

Implementación

En el capítulo anterior podemos ver que las expresiones para el flujo dependen de muchas literales que no necesariamente conocemos, tales como los valores de γ y el índice espectral p que juntos describen la distribución de electrones. Encontrar estos y otras literales que son desconocidas *a priori* es nuestro objetivo principal, ya que son las características físicas de la fuente de las emisiones que estamos ajustando y su entorno.

Por lo tanto, si nuestro objetivo es encontrar dichos valores debemos encontrar un algoritmo que nos permita trabajar “en reversa” desde los valores del flujo observados, simular para los mismos valores de energía un flujo utilizando nuestras ecuaciones, y comparar ambos hasta encontrar los valores que hacen que su diferencia sea mínima.

4.1. Descenso por el gradiente

Con lo anterior en mente, yo propongo el algoritmo de optimización de descenso por el gradiente, pues podemos definir una función de error entre los flujos (observado y simulado) y minimizarla. Este método es muy utilizado en el campo del aprendizaje automático o *Machine learning* ([And+16]).

4.1.1. Descripción

El método de descenso por el gradiente es un método de optimización iterativo de primer orden (llamado así por que se limita al uso de derivadas de primer orden) cuyo objetivo es encontrar el mínimo de una función. Llamemos a $F(\vec{X})$ a una función multivariable, y \vec{x} un punto en cuya vecindad F está definida y es diferenciable, F decrece de manera más rápida si se va desde \vec{x} hacia la dirección del negativo del gradiente de F en \vec{x} , o bien en la dirección de $-\nabla F(\vec{x})$.

Ya que nos interesa la dirección de este vector y no su magnitud debemos introducir un valor que convierta la magnitud del vector del gradiente en una que podamos utilizar, este valor dependerá siempre del problema a tratar, pero por lo general es un número pequeño y positivo llamado tasa de aprendizaje o *learning rate* que aquí denotaremos con α .

Ahora podemos definir de manera recursiva un vector de entrada para F de la siguiente forma:

$$\vec{x}_{t+1} = \vec{x}_t - \alpha \nabla F(\vec{x}_t) \quad (4.1)$$

Con un α correcto tenemos que $F(\vec{x}_{t+1}) \leq F(\vec{x}_t)$. Así, iniciando con un vector inicial \vec{x}_0 podemos definir de forma iterativa la secuencia monótona decreciente:

$$F(\vec{x}_0) \geq F(\vec{x}_1) \geq F(\vec{x}_2) \geq F(\vec{x}_3) \geq \dots \quad (4.2)$$

hasta lograr la precisión deseada. Es importante recalcar que el método de descenso por el gradiente puede asegurar encontrar un mínimo local, pero no necesariamente un mínimo global.

4.1.2. Limitaciones

El método de descenso por el gradiente tiene la limitación de volverse bastante lento cuando se acerca al mínimo; es decir, su razón de convergencia es inferior a otros

métodos. También estamos en la desventaja de que nuestra función a minimizar debe ser diferenciable en un espacio bien definido.

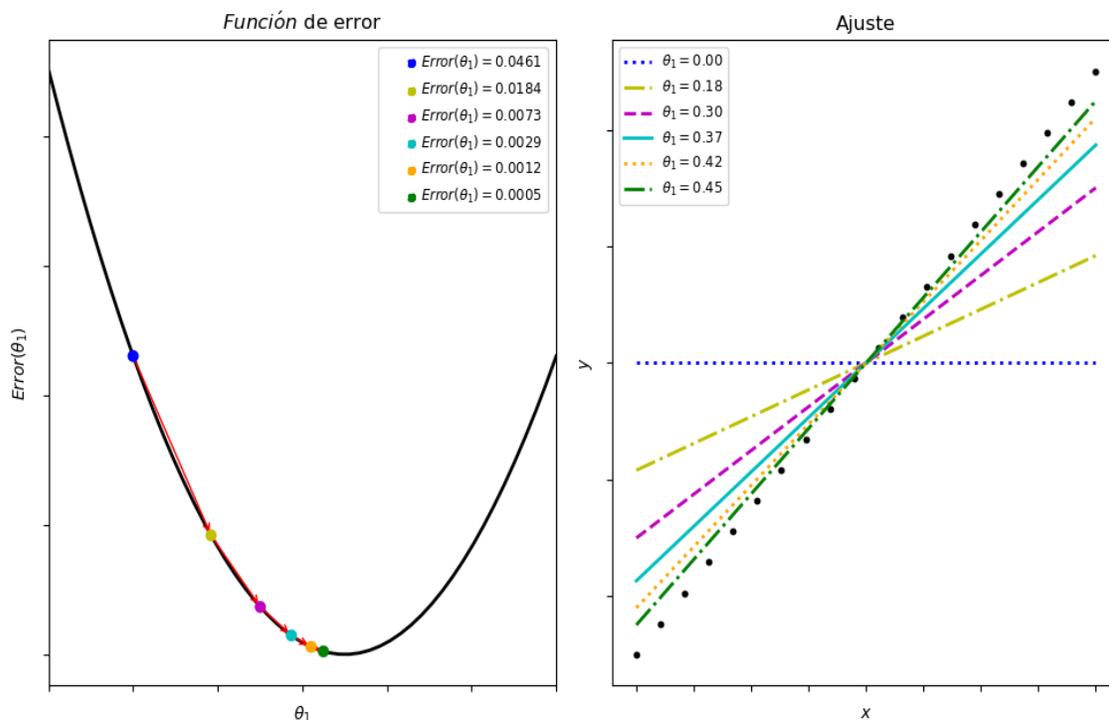


Figura 4.1: Visualización del método de descenso por el gradiente. En este ejemplo, se quiere aproximar una recta que pasa por el origen, por lo tanto solo se tiene un parámetro: la pendiente θ_1 . En la derecha está la aproximación obtenida en cada iteración junto con el valor de θ_1 , marcado con puntos negros está el objetivo. A la izquierda se muestra la función de error para todos los parámetros posibles, y marcados con puntos están los valores del error en cada iteración. Se puede observar claramente cómo el método se hace más lento al acercarse al mínimo.

4.1.3. Función de error

Es necesario definir una función de error que nos permita comparar los flujos que obtengamos, pues es esta función la que será minimizada con descenso por el gradiente. Para esto propongo usar la función de error cuadrático medio o *mean-squared error*. Esta función es la más común a la hora de utilizar descenso por gradiente, pues es siempre positiva y fácil de diferenciar.

Antes de empezar con la función, introduciremos algo de notación; $\vec{Y} = \{y_1, y_2, \dots, y_m\}$

es el vector al que queremos que nuestros resultados se aproximen, $\vec{Y}^p = \{y_1^p, y_2^p, \dots, y_m^p\}$ es el vector predicho por nuestras simulaciones, y por último $\vec{\Theta} = \{\theta_1, \theta_2, \dots, \theta_n\}$ es el vector que contiene los parámetros de nuestra simulación. Éste último es el vector más importante de todo el método, pues son los valores que vamos a cambiar para poder minimizar la función de error.

La función está definida como sigue:

$$E(\vec{\Theta}) = \frac{1}{2m} \sum_{i=1}^m (y_i^p - y_i)^2 \quad (4.3)$$

En nuestro caso particular, el vector \vec{Y} es el flujo observado y el vector \vec{Y}^p se referirá a nuestro flujo simulado, por lo que podemos darle a éste último un nombre más afín: $\vec{F}(\vec{\epsilon}, \vec{\Theta})$ que denota la expresión del flujo correspondiente para un vector de energía $\vec{\epsilon} = \{\epsilon_1, \epsilon_2, \dots, \epsilon_m\}$ y los parámetros $\vec{\Theta}$; de una manera más explícita $\vec{F}(\vec{\epsilon}, \vec{\Theta}) = \{F(\epsilon_1, \vec{\Theta}), F(\epsilon_2, \vec{\Theta}), \dots, F(\epsilon_m, \vec{\Theta})\}$ donde F es la función de flujo. Con estos cambios nuestra función de error es la siguiente:

$$E(\vec{\Theta}) = \frac{1}{2m} \sum_{i=1}^m (F(\epsilon_i, \vec{\Theta}) - y_i)^2 \quad (4.4)$$

Ya que el método de descenso por el gradiente requiere que la función a minimizar (en este caso, nuestra función de error) sea diferenciable y continua al menos en una región de su dominio, es hora de discutir la diferenciableidad de esta función. Es fácil ver que la función de error es diferenciable sí y solo sí la función del flujo $F(\epsilon_i, \vec{\Theta})$ lo es; pero es aquí cuando encontramos el primer problema, observando ambas ecuaciones correspondientes a los flujos de sincrotrón y SSC se puede ver que no son expresiones fáciles de analizar y que obtener de manera analítica sus derivadas parciales sería algo bastante difícil, si no imposible. Por lo tanto, continuaré asumiendo que estas funciones, y por lo tanto la función de error, son diferenciables. Al llevar este método a la práctica con W Comae no ocurrieron errores de diferenciableidad en los puntos utilizados.

Para calcular el gradiente de $E(\vec{\Theta})$ hay que calcular la derivada parcial de esta

sobre los parámetros individuales $\frac{\partial}{\partial \theta_j} E(\vec{\Theta})$. Haciendo uso de la regla de la cadena, la expresión queda como sigue:

$$\frac{\partial}{\partial \theta_j} E(\vec{\Theta}) = \frac{1}{m} \sum_{i=1}^m \left(\frac{\partial}{\partial \theta_j} F(\epsilon_i, \vec{\Theta}) (F(\epsilon_i, \vec{\Theta}) - y_i) \right) \quad (4.5)$$

Ahora sólo hay que encontrar $\frac{\partial}{\partial \theta_j} F(\epsilon_i, \vec{\Theta})$. Una vez más, encontrar las derivadas parciales de las expresiones del flujo es demasiado complejo como para hacerlo analíticamente, por lo que recurriremos a derivarlas de manera numérica más tarde.

Asumiendo esto, definimos de manera recursiva cada parámetro de la siguiente forma:

$$\theta_{j,t+1} = \theta_{j,t} - \alpha \frac{\partial}{\partial \theta_j} E(\vec{\Theta}) \quad (4.6)$$

4.1.4. Vectorización

Podemos simplificar varias de las expresiones pasadas si las definimos como operaciones entre vectores, por ejemplo, la función de error queda definida como:

$$E(\vec{\Theta}) = \frac{1}{2m} \sum_{i=1}^m (\vec{F}(\vec{\epsilon}, \vec{\Theta}) - \vec{Y})_i^2 \quad (4.7)$$

Para definir el gradiente de la función de error, hay que introducir una matriz, que llamaré $\nabla \vec{F}(\vec{\epsilon}, \vec{\Theta})$, cuyas entrada (k, r) corresponden a la derivada parcial de la función del flujo con respecto al k-ésimo parámetro evaluada en el r-ésimo valor de energía. Así tenemos una matriz de dimensión $n \times m$ de la siguiente forma:

$$\nabla \vec{F}(\vec{\epsilon}, \vec{\Theta}) = \begin{bmatrix} \frac{\partial}{\partial \theta_1} F(\epsilon_1, \vec{\Theta}) & \frac{\partial}{\partial \theta_1} F(\epsilon_2, \vec{\Theta}) & \dots & \frac{\partial}{\partial \theta_1} F(\epsilon_m, \vec{\Theta}) \\ \frac{\partial}{\partial \theta_2} F(\epsilon_1, \vec{\Theta}) & \frac{\partial}{\partial \theta_2} F(\epsilon_2, \vec{\Theta}) & \dots & \frac{\partial}{\partial \theta_2} F(\epsilon_m, \vec{\Theta}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial \theta_n} F(\epsilon_1, \vec{\Theta}) & \frac{\partial}{\partial \theta_n} F(\epsilon_2, \vec{\Theta}) & \dots & \frac{\partial}{\partial \theta_n} F(\epsilon_m, \vec{\Theta}) \end{bmatrix}$$

En otras palabras cada columna es el gradiente de F para un valor distinto de energía.

Así, el gradiente del error puede ser definido usando operaciones de matrices como sigue:

$$\nabla E(\vec{\Theta}) = \nabla \vec{F}(\vec{\epsilon}, \vec{\Theta}) \cdot (\vec{F}(\vec{\epsilon}, \vec{\Theta}) - \vec{Y}) \quad (4.8)$$

Y por último la definición recursiva del vector con los parámetros queda como sigue:

$$\vec{\Theta}_{t+1} = \vec{\Theta}_t - \alpha \nabla E(\vec{\Theta}_t) \quad (4.9)$$

La razón por la cual vectorizamos el método de descenso por el gradiente es para facilitar la escritura del código más adelante, ya que es más sencillo trabajar con un par de vectores que con conjuntos de datos que deben ser operados manualmente.

4.2. Optimización de descenso por el gradiente

Con el fin de optimizar lo más posible el ajuste del flujo, podríamos intentar optimizar el algoritmo de descenso por el gradiente. Para esto existen varias estrategias que tienen como objetivo acelerar el tiempo que le toma a descenso por el gradiente converger a un mínimo local. A continuación revisaremos algunas de estas optimizaciones y cómo pueden ser aplicadas a nuestro caso particular.

4.2.1. Descenso por el gradiente con momento

Se puede ver que el método de descenso por el gradiente navega de forma bastante lenta en las regiones de la función cuya pendiente es bastante pequeña, pues el gradiente en estas regiones es de igual forma muy pequeño. Para superar esto, podemos agregar otro factor a la ecuación 4.9 al que llamaremos *momento*.

Esta idea viene de la intuición de tomar una pelota y colocarla en algún punto de una superficie; la pelota rodará hacia abajo acumulando momento y moviéndose cada vez más rápido hasta alcanzar una velocidad terminal. Para conseguir este efecto, es necesario acumular de alguna forma todos los cambios en el gradiente anteriores hasta un punto determinado en el tiempo.

El valor antes mencionado puede ser definido como la suma de todos los cambios anteriores aplicados sobre el vector de parámetros; así para la primera iteración tendremos un paso equivalente a descenso por el gradiente normal, y en las siguientes iremos sumando los cambios en un vector que llamaremos \vec{V}_t :

$$\vec{V}_t = \mu\vec{V}_{t-1} - \alpha\nabla E(\vec{\Theta}_t) \quad (4.10)$$

El valor μ , a veces llamado coeficiente de fricción, es un número entre 0 y 1, por lo general se elige 0.9 (este valor depende siempre del caso y su valor se refleja en la velocidad a la que el algoritmo converge a un mínimo). Así los parámetros cambiarían por cada paso de la siguiente forma:

$$\vec{\Theta}_{t+1} = \vec{\Theta}_t - \vec{V}_t \quad (4.11)$$

Esta optimización del método de descenso por el gradiente aumenta de manera considerable la velocidad en la que se encuentra un mínimo local, aunque incluye el efecto secundario de que el error en un tiempo t no siempre será menor al error en el tiempo $t - 1$, pues conforme se acumula momento el algoritmo puede avanzar hacia una dirección en la que ya había estado avanzando y pasar de largo un mínimo local. La mayoría de las veces el algoritmo vuelve sobre sus pasos por la influencia que el gradiente sigue teniendo sobre el ajuste, pero a veces también este comportamiento nos ayuda a escapar de mínimos locales o *puntos silla*. De igual forma la convergencia en un mínimo local sigue estando garantizada.

4.2.2. *Adaptive Gradient Descent* o AdaGrad

Durante los ajustes del flujo que realizamos, se volvió evidente que no todos los parámetros tienen el mismo impacto sobre la expresión del flujo, algunos como el índice espectral (p) impactan enormemente la forma del SED. Esto se traduce a que el gradiente de la función de error con respecto al resto de los parámetros es casi siempre 0 o un valor muy cercano a 0. Una forma de superar esto es olvidar la idea de una tasa de aprendizaje fija y en cambio utilizar una diferente para cada parámetro;

y no solo eso, también podemos hacer que esta tasa sea variable para cada tiempo t . De esta forma, podríamos hacer que cuando el gradiente de uno de los parámetros que casi no impactan en el error deje de ser cero, el ajuste sea ayudado con una tasa de aprendizaje más grande. La idea entonces es decrecer la tasa de aprendizaje en proporción a qué tan frecuentemente es actualizado un parámetro, así los parámetros que se actualizan más seguido tendrán una tasa de aprendizaje más pequeña, mientras que los que son rara vez actualizados tendrán una tasa de aprendizaje mayor.

Para introducir esta idea, cambiamos de nuevo la forma en la que cambian los parámetros para cada tiempo t ([Gyl18]):

$$\vec{\Theta}_{t+1} = \vec{\Theta}_t - \frac{\alpha}{\sqrt{\epsilon + \Sigma^2 E(\vec{\Theta}_t)}} \odot \nabla E(\vec{\Theta}_t) \quad (4.12)$$

Donde $\Sigma^2 E(\vec{\Theta}_t)$ es sencillamente la suma del cuadrado de todos los gradientes del error desde $\tau = 0$ hasta $\tau = t$ o bien:

$$\Sigma^2 E(\vec{\Theta}_t) = \Sigma^2 E(\vec{\Theta}_{t-1}) + (\nabla E(\vec{\Theta}_t))^2 \quad (4.13)$$

El número ϵ es un valor pequeño entre 0 y 1 cuyo propósito es evitar una división entre 0, y el operador \odot denota el producto entre vectores entrada por entrada.

Ahora α será la tasa de aprendizaje inicial, y con cada paso del ajuste $\Sigma^2 E(\vec{\Theta}_t)$ se hará más y más grande en las entradas correspondientes a parámetros cuyo gradiente se acumula más (es decir, cambian seguido); para estos parámetros el denominador será mayor y por lo tanto la tasa de aprendizaje se reducirá, mientras que en los parámetros donde el gradiente es muy pequeño tendrán el efecto contrario.

Una desventaja de este método es que, dado suficiente tiempo, la tasa de aprendizaje para algún valor que cambia muy seguido puede aproximarse a 0 y esto puede estancar por completo el ajuste de estos parámetros. La siguiente optimización soluciona este problema.

4.2.3. *Root Mean Square Propagation* o RMSProp

Para solucionar el problema de estancamiento de AdaGrad, tenemos que introducir un nuevo factor que reduzca la velocidad en la que la tasa de aprendizaje para algún parámetro alcance el 0; en vez de sumar directamente los cuadrados del gradiente, introduciremos la siguiente expresión:

$$\vec{V}_t = \beta \vec{V}_{t-1} + (1 - \beta) \nabla E(\vec{\Theta}_t)^2 \quad (4.14)$$

y con ella modificaremos la ecuación 4.12:

$$\vec{\Theta}_{t+1} = \vec{\Theta}_t - \frac{\alpha}{\sqrt{\epsilon + \vec{V}_t}} \odot \nabla E(\vec{\Theta}_t) \quad (4.15)$$

De esta forma, las tasas de aprendizaje no caerán a 0 de la misma forma como sucedía en AdaGrad. El valor β es un número entre 0 y 1, y su valor recomendado es casi siempre de 0.9 ([HSS12]).

4.2.4. *Adaptive Moment Estimation* o ADAM

Para maximizar la eficiencia del algoritmo, introduciremos nuevamente el concepto de momento, esta vez aplicado a RMSProp. Al igual que en descenso por el gradiente con momento, el momento será representado por la suma de los cambios aplicados a cada parámetro ([KB14]).

$$\vec{M}_t = \beta_1 \vec{M}_{t-1} + (1 - \beta_1) \nabla E(\vec{\Theta}_t) \quad (4.16)$$

En esta optimización, el ajuste de los parámetros quedará definido como:

$$\vec{\Theta}_{t+1} = \vec{\Theta}_t - \frac{\alpha}{\sqrt{\epsilon + \vec{V}_t}} \odot \vec{M}_t \quad (4.17)$$

En este algoritmo \vec{M}_t representa la media móvil exponencial del gradiente (que es un estimado del promedio del gradiente) y \vec{V}_t representa la media móvil exponencial

del gradiente al cuadrado (o bien un estimado de la varianza del gradiente). Los parámetros β_1 y β_2 (este último tomando el lugar de β en la ecuación 4.14) son valores entre 0 y 1 que definen las tasas de decaimiento exponencial de estas medias. Sin embargo, estas medias son inicializadas como vectores de ceros, lo que resulta en estimaciones con un sesgo hacia el cero. Este sesgo de inicialización puede ser superado con los siguientes ajustes:

$$\hat{M}_t = \frac{\vec{M}_t}{1 - \beta_1^t} \quad (4.18)$$

$$\hat{V}_t = \frac{\vec{V}_t}{1 - \beta_2^t} \quad (4.19)$$

Después de estos ajustes, la ecuación 4.17 queda definida así:

$$\vec{\Theta}_{t+1} = \vec{\Theta}_t - \frac{\alpha}{\sqrt{\epsilon + \hat{V}_t}} \odot \hat{M}_t \quad (4.20)$$

Esta será la forma de descenso por el gradiente que utilizaremos para minimizar el error en la función del flujo en las secciones siguientes.

4.3. Criterios de parada

Es importante decidir en qué momento detener las iteraciones y quedarse con los parámetros obtenidos. Para descenso por el gradiente existen varios criterios, uno bastante obvio es simplemente definir un número de iteraciones máximas, y esperar que sean suficientes para encontrar un mínimo; otra se trata de detenerse cuando la diferencia entre el error de la iteración actual y la pasada sea muy pequeño, definiendo un valor de precisión deseado.

Ya que no podemos asegurar que alguna de las funciones de flujo sean estrictamente convexas, es extremadamente probable que este método encuentre la mayoría de las veces un mínimo local, por lo que en el mejor de los casos se debe de esperar un mínimo local cuyos parámetros correspondan a un resultado lo suficientemente cercano al flujo observado.

Por esto último, optaré por utilizar ambas condiciones antes descritas. En los capítulos siguientes, se expondrá un análisis de los resultados del método de descenso por el gradiente, incluyendo cuántas iteraciones le toma al algoritmo encontrar un mínimo local y cuál es la precisión más adecuada para detener el proceso.

Para la implementación de nuestra simulación utilizaremos *Python*, debido que es fácil trabajar con vectores en este lenguaje y la gran variedad de bibliotecas auxiliares; pero la razón más importante, es la gran popularidad de *Python* en el desarrollo de programas para la investigación en el área de la astrofísica. Haremos uso de la biblioteca *astropy* para auxiliarnos con el uso de unidades físicas, *scipy* para derivar e integrar numéricamente nuestras funciones y *numpy* para las operaciones vectoriales necesarias.

4.4. Implementación del Flujo

No es difícil notar que, aun con los ajustes que hicimos a las expresiones del flujo (3.14 y 3.17), estas son muy complejas, y van a exigir bastante poder de procesamiento para poder ser evaluadas eficazmente. Una de las desventajas de utilizar *Python* es que agrega complejidad extra a los cálculos gracias a su tipado dinámico; si utilizáramos otro lenguaje, como *C*, podríamos hacer uso de tiempos de cómputo más reducidos. Sin embargo, preferimos *Python* por las ventajas antes mencionadas. Por suerte hay una manera de combinar ambas ventajas, y es mediante el uso de *Cython*.

4.4.1. Cython

Cython es un compilador estático para *Python* y el lenguaje extendido *Cython*. Permite crear extensiones de *C* para *Python* de manera muy sencilla a la vez que intenta ser un superconjunto de este último, dejándole ser de alto nivel, orientado a objetos, funcional y dinámico. Una de sus principales características es que da soporte para un tipado estático.

Un programa escrito en *Cython* es traducido a *C* de forma optimizada, y después es compilado en extensiones de *Python*. *Cython* puede compilar casi cualquier *script*

escrito en *Python* y el código generado en *C* supera de forma significativa la velocidad del original haciendo uso del tipado estático opcional, a veces superándolo por varios órdenes de magnitud.

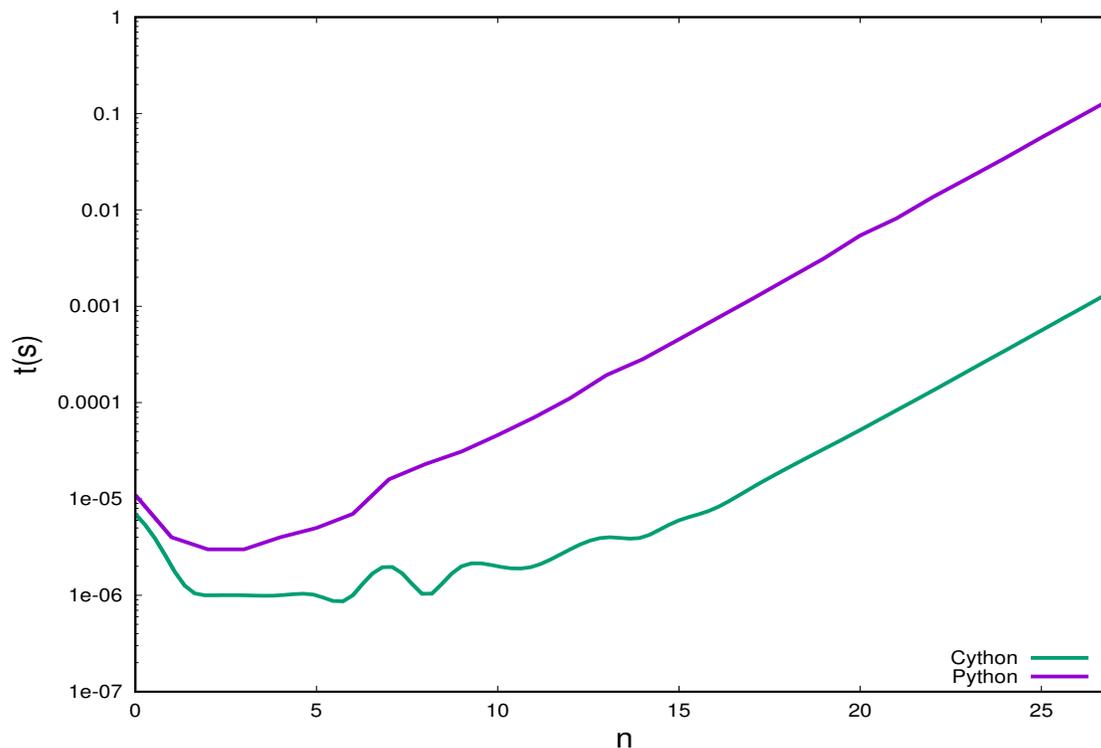


Figura 4.2: Comparación de tiempos de ejecución utilizado para obtener el n -ésimo número de la serie de Fibonacci entre *Cython* y *Python*.

La implementación del flujo utilizada estará escrita con *Cython* para poder aprovechar esta característica, ya que las expresiones del flujo deberán ser evaluadas varias veces para poder realizar el ajuste. El código específico para la simulación del flujo no será expuesto, pues son simplemente las ecuaciones 3.14 y 3.17, junto con todas las funciones auxiliares necesarias para calcularlas. Solo cabe mencionar, que para su integración numérica, hicimos uso de la función *quad* de la biblioteca *scipy*. La implementación del método de ajuste estará escrita en *Python*, pero también podría hacer uso de *Cython*. Esta vez, para facilitar su exposición, permanecerá en *Python* y será descrita en la siguiente sección.

4.5. Implementación del Ajuste.

El ajuste será realizado en dos partes, primero ajustaremos la componente de bajas energías con la expresión del flujo de sincrotrón, intentando hallar los parámetros δ_D , B , K_e , p , γ'_{break} y γ'_2 . Después, usando los parámetros ajustados de la primera parte, ajustaremos la componente de altas energías con la expresión de flujo de *synchrotron self-Compton* hallando el último parámetro $t_{v,min}$.

Como ya habíamos mencionado, utilizaremos el algoritmo Adam; empezando por el método en sí, la implementación es la siguiente:

```

1 def gradientDescent(X,Y,Theta_n,alpha,F,derivatives,max_iter,precision,
    momentum = 0.0,beta = 0.5):
2     error = 1e5
3     iters = 0
4     gradientSum = np.zeros(Theta_n.size)
5     epsilon = np.full(Theta_n.shape,1e-8)
6     lastChange = np.zeros(Theta_n.size)
7
8     for i in range(max_iter):
9         iters+=1
10        currentTheta = Theta_n
11        g,e = gradE(X,Y,currentTheta,F,derivatives)
12
13        #Adam
14        gradientSum = beta*gradientSum + (1-beta)*np.square(g)
15        lastChange = momentum*lastChange + (1-momentum)*g
16
17        #biased 1
18        mt = lastChange/(1-pow(momentum,i+1))
19        #biased 2
20        vt = gradientSum/(1-pow(beta,i+1))
21
22        lerningRate = alpha / np.sqrt((epsilon + vt))
23        Theta_n = currentTheta - lerningRate*mt
24        step = error - e
25

```

```

26     if abs(step) <= precision:
27         print("Precision achieved")
28         break
29
30     error = e
31
32     return [Theta_n, error, iters]
```

Las entradas para esta función son, en el orden que aparecen: el vector de entradas para la función a ajustar, el vector con los datos observados, los parámetros iniciales (Θ_t), la tasa de aprendizaje inicial α , la función a ajustar, un vector de funciones cuyas entradas son las derivadas parciales de F con respecto a cada parámetro, las iteraciones máximas permitidas y la precisión deseada, por último los parámetros *momentum* (β_1) y β_2 . La variable *gradientSum* es usada para ir acumulando la suma de los cuadrados de los gradientes como media móvil exponencial (i.e. \vec{V}_t), mientras que *lastChange* acumula de forma similar el gradiente (i.e. \vec{M}_t). Por último, los ajustes a estas variables se encuentran en la línea 18 y 20. La ecuación del ajuste (4.20) puede ser observada en la línea 23.

La función auxiliar gradE calcula el vector $\nabla E(\vec{\Theta}_t)$ y también el error actual.

```

1 def gradE(X,Y,Theta_n,F,derivatives):
2     m = X.size
3     arrayF = np.zeros((X.shape[0],))
4     for i in range(0,X.shape[0]):
5         arrayF[i] = F(X[i],Theta_n)
6
7     error = Error(X,Y,Theta_n,arrayF)
8     return 1/m*np.dot(dF(X,Theta_n,derivatives),(arrayF-Y)),error
```

En el ciclo que empieza en la línea 4, se llena de forma iterativa un vector con todos los valores simulados para la entrada \vec{X} , o bien $\vec{F}(\vec{\epsilon}, \vec{\Theta})$.

```

1 def dF(X,Theta_n,derivatives):
2     j = Theta_n.shape[0]
3     m = X.shape[0]
4     d = np.zeros((j,m))
```

```

5 for i in range(j):
6     for k in range(m):
7         d[i,k] = derivatives[i](X[k], Theta_n)
8 return d

```

La función auxiliar dF sólo llena iterativamente la matriz $\nabla \vec{F}(\vec{\epsilon}, \vec{\Theta})$ (ver 4.8). Y por último el Error es calculado como sigue:

```

1 def Error(X,Y,Theta_n, arrayF):
2     m = X.shape[0]
3     return np.sum(np.square(arrayF-Y))/(2.*m)

```

Ahora sólo resta definir cuáles son los parámetros iniciales que utilizaremos, obtener algunos datos observados de W Comae, definir las derivadas parciales de las funciones del flujo para cada uno de los parámetros, elegir un α y un número de iteraciones máximas para que el ajuste pueda comenzar.

Otra cosa que cabe aclarar, es que durante los ajustes notamos que los valores que eran resultados de las expresiones del flujo eran de magnitudes que las hacían difíciles de trabajar, siendo a veces de un valor tan pequeño que la precisión de los números de punto flotante de *Python* no los representaba correctamente; por esa razón optamos por cambiar la función a ajustar. En vez de ajustar la expresión del flujo como fue descrita antes (3.14 y 3.17), ajustaremos sus logaritmos (i.e. $\log(f_{\epsilon}^{syn})$ y $\log(f_{\epsilon_s}^{SSC})$). Esto también debe ser considerado a la hora de definir la derivada parcial de la función. Por las mismas razones, algunos parámetros se ajustarán también a sus logaritmos, específicamente $K_e, \gamma'_{break}, \gamma'_2$ y $t_{v,min}$. En el código utilizamos un arreglo llamado *params*, que para la primera parte del ajuste corresponde a $[\delta_D, B, \log(K_e), p, \log(\gamma'_{break}), \log(\gamma'_2)]$ y para la segunda parte simplemente $[t_{v,min}]$.

Para implementar las derivadas parciales utilizaremos la función *derivative* del módulo *scipy*. A continuación un ejemplo de la derivada parcial $\frac{\partial}{\partial p} f_{\epsilon}^{syn}$:

```

1 import numpy as np
2 import astropy.units as u
3 import astropy.constants as c
4 from scipy.misc import derivative
5

```

```

6 def dFp(x, params):
7     a = (x*u.Hz)*c.h
8     b = c.m_e*pow(c.c,2)
9     epsilon = a/(b.to("Hz J s"))
10    return derivative(lambda p: np.log10(Synchrotron(epsilon ,dL, params[0] ,
        params[1],10**params[2] , p ,10**params[4],10**params[5])), params
        [3] , dx=0.5)

```

La implementación del resto de las derivadas parciales es igual a la que se muestra arriba. Estas funciones son colocadas en un arreglo para poder ser entrada de la función *gradientDescent* cuya implementación ya fue descrita.

En cuanto a los parámetros iniciales, optamos por definir un rango para cada parámetro y tomar un valor aleatorio dentro de él como valor inicial. Esta simple solución tiene como ventaja adicional el hacer más probable que evitemos los mínimos locales, ya que si una ronda de ajuste encuentra rápidamente un mínimo local podemos empezar una nueva que, por ser aleatoria, tendrá la posibilidad de encontrar un mínimo local diferente y, con suerte, uno que minimice el error lo más posible.

En cuanto a los criterios de parada, elegimos como precisión deseada 1×10^{-6} y como número de iteraciones máximas 1×10^3 . En nuestros experimentos, el ajuste siempre alcanzaba la precisión deseada antes de realizar el número de iteraciones máximas. El código completo de la implementación puede ser encontrado en este repositorio: github.com/LukaSchnauzer/blazar-sed-adjustment

Capítulo 5

Resultados

5.1. W Comae

El blazar W Comae (también conocido como 1219+285 u ON231) fue descubierto como una fuente de radio en 1971 ([Bro71]). Observaciones interferométricas de muy larga base (VLBI) revelan un jet complejo en W Comae que se extiende hacia el este a $\sim 100^\circ$. También, se encontró que el jet muestra componentes superluminales con una fuerte polarización. La curva de luz histórica de W Comae muestra variaciones a todas las escalas, de días y semanas a meses y años. En enero de 1997 ([GM98]) se observó el valor de brillo más alto para W Comae desde 1940, alcanzando un máximo de 14.2 *mag*. Después, ([Mas+99]) en abril 23 de 1998, el objeto alcanzó un máximo histórico de ~ 12.2 *mag*.

5.1.1. Detección en rayos gamma

La emisión en rayos gamma del blazar W Comae ha sido detectada por el *Energetic Gamma Ray Experiment Telescope*, a bordo del *Compton Gamma-Ray Observatory* en bandas de los 100 MeV hasta 10 GeV ([Har+99]).

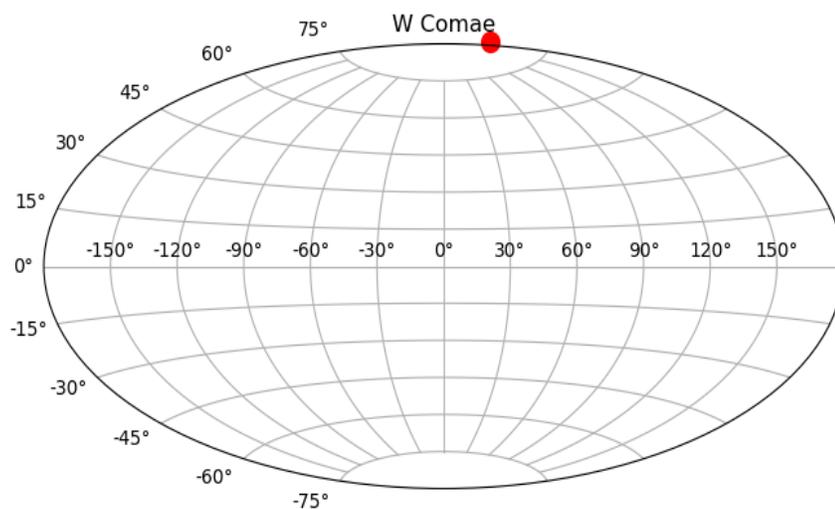


Figura 5.1: Posición aproximada de W Comae en coordenadas galácticas

5.1.2. Clasificación e Importancia

W Comae se clasifica como un BL Lac con pico de frecuencia intermedia o *intermediate frequency-peaked BL Lac object* (IBL) ([Sor+14]).

Este blazar fue considerado un objetivo muy interesante para las observaciones de muy altas energías (VHE) debido a la posibilidad de ser detectado por telescopios Cherenkov tales como el *High Energy Stereoscopic System* (HESS), el telescopio *Major Atmospheric Gamma-ray Imaging Cherenkov Telescope* (MAGIC) y *Very Energetic Radiation Imaging Telescope Array System* (VERITAS). Esta posibilidad fue confirmada cuando en marzo 15 de 2008, pues W Comae fue detectado como un emisor de rayos gamma en VHE por VERITAS ([Acc+08]). Así W Comae se convirtió en el primer IBL en ser detectado en VHE.

Los valores observados que utilizamos para el ajuste de W Comae fueron obtenidos de las observaciones del 15 de marzo de 2008 por VERITAS. Los valores de energía serán extraídos de un archivo y son usados para llenar un arreglo X . De igual forma los valores de flujo son usados para llenar un arreglo Y . Ambos arreglos son entradas de la función *gradientDescent*. Los valores de α para cada parte del ajuste fueron encontrados de forma experimental; para el ajuste de la primera componente dejamos

$\alpha = 1 \times 10^{-1}$ y para la segunda componente $\alpha = 5 \times 10^{-1}$. En cuanto al resto de parámetros, utilizamos sus valores recomendados de $\beta_1 = 0.9$, $\beta_2 = 0.99$ y $\epsilon = 1 \times 10^{-8}$. En la tabla 5.1 se pueden encontrar los rangos utilizados para inicializar los parámetros del flujo. Otros valores obtenidos por medio de observaciones previas son $d_L = 473.3$ *Mpc* y $z \approx 0.0996$.

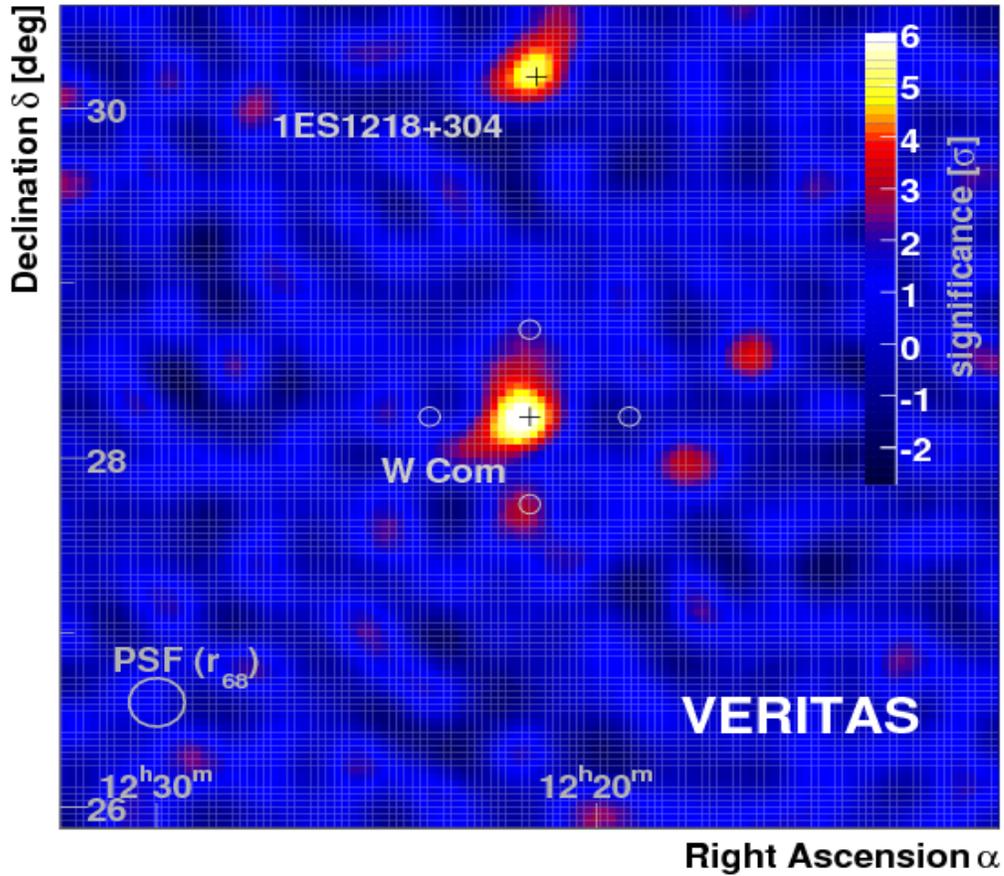


Figura 5.2: Mapa del cielo visto por VERITAS al detectar W Comae a muy altas energías ([Acc+08]).

Presentaremos algunas rondas de ajuste para el flujo de W Comae, estos ajustes fueron realizados de la forma descrita en capítulos anteriores. Para visualizar el avance en el ajuste, graficamos el flujo utilizando los parámetros iniciales junto con el flujo utilizando los parámetros encontrados para el mínimo local al que convergió el ajuste.

Parámetro	Cota Inferior	Cota Superior
δ_D	10	120
B	$10G$	$100G$
$\log K_e$	40	45
p	2.0	3.1
$\log \gamma'_{break}$	4.5	5.6
$\log \gamma'_2$	$\log \gamma'_{break}$	7.0
$\log t_{v,min}$	3.0	6.0

Tabla 5.1: Rangos definidos para cada parámetro. El parámetro inicial se escoge aleatoriamente dentro de estos rangos.

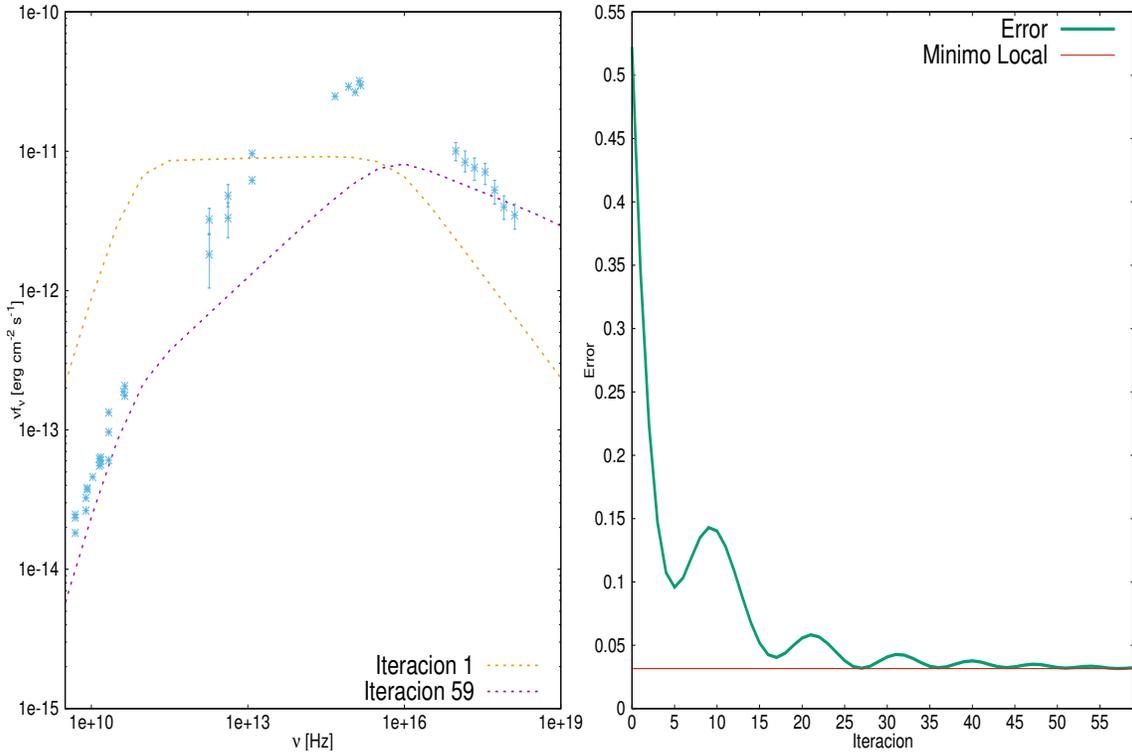


Figura 5.3: Ronda 1 de ajuste. Iteraciones totales: 59. El ajuste convergió a un mínimo local con un error de ≈ 0.03 . Parámetros finales: $\delta_D = 37.13$, $B = 42.7$, $\log(K_e) = 42.72$, $p = 2.318$, $\log(\gamma_{break}) = 4.59$, $\log(\gamma_2) = 6.614$

Junto a esto último, también mostramos cómo evolucionó el error con respecto de cada iteración. En todos los casos, el error presenta un comportamiento casi asintótico hacia un valor de error que corresponde al mínimo local. Como se puede notar, algunas veces este mínimo local no satisface la forma deseada del flujo.

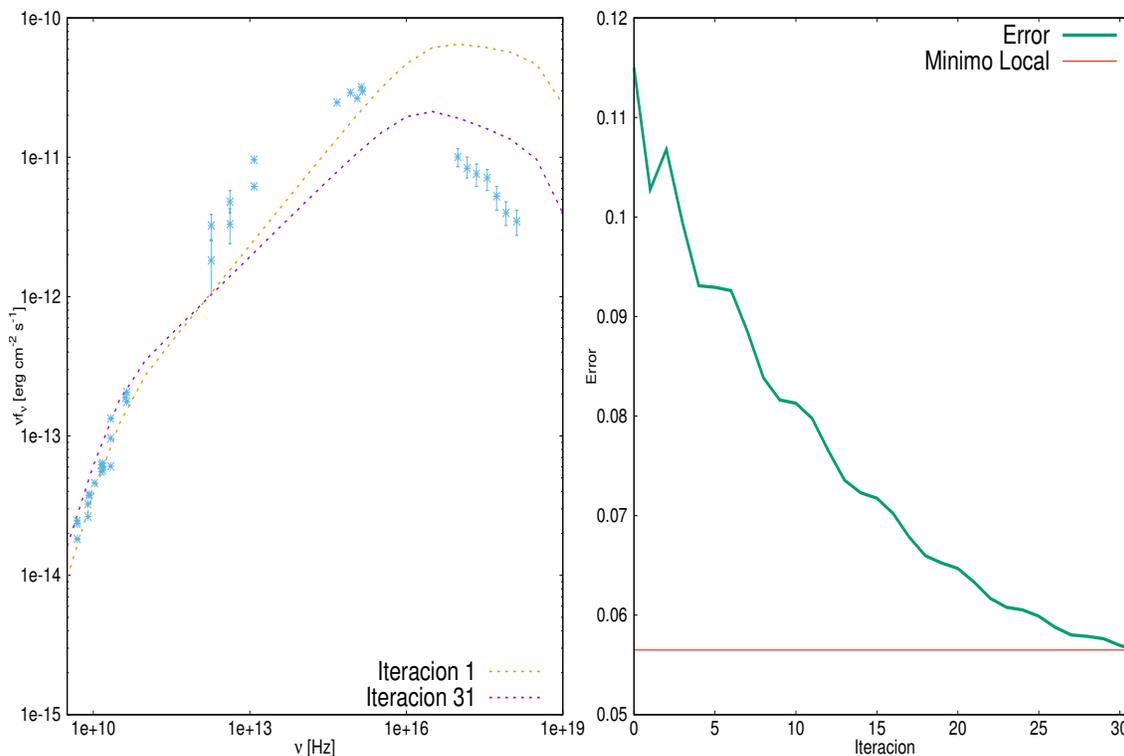


Figura 5.4: Ronda 2 de ajuste. Iteraciones totales: 31. El ajuste convergió a un mínimo local con un error de ≈ 0.056 . Parámetros finales: $\delta_D = 36.57$, $B = 18.12$, $\log(K_e) = 42.569$, $p = 2.262$, $\log(\gamma_{break}) = 4.92$, $\log(\gamma_2) = 6.24$.

En ciertas rondas de ajuste, el error parece *rebotar* alrededor del mínimo local. Esto se debe al factor de momento que introdujimos para optimizar el descenso por el gradiente.

De igual forma, otro comportamiento de descenso por el gradiente se hace aparente. El error cae bastante rápido al principio del ajuste y después desacelera al acercarse más y más al mínimo. Este comportamiento se ve acentuado por las optimizaciones que elegimos realizar; sin estas, deberíamos observar una convergencia más lenta.

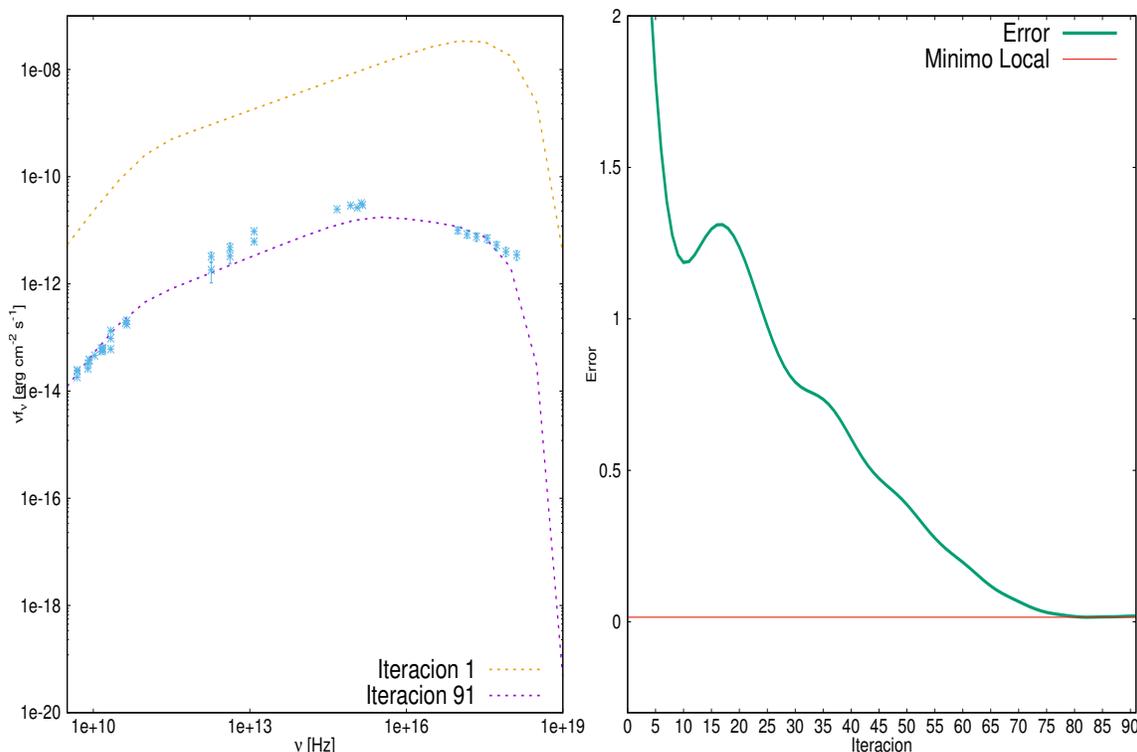


Figura 5.5: Ronda 3 de ajuste. Iteraciones totales: 91. El ajuste convergió a un mínimo local con un error de ≈ 0.015 . Parámetros finales: $\delta_D = 20.225$, $B = 71.11$, $\log(K_e) = 44.25$, $p = 2.15$, $\log(\gamma_{break}) = 4.322$, $\log(\gamma_2) = 5.49$.

El objetivo era realizar varias rondas de ajuste hasta encontrar el mínimo. De todas las rondas que realizamos el valor más pequeño de error que encontramos fue de ≈ 0.0086 . El resultado de este ajuste es también un flujo cuya forma se asemeja bastante al resultado esperado para W Comae.

Una vez que encontramos buenos parámetros para el flujo de sincrotrón, utilizamos los mismos para ajustar el flujo de *synchrotron self-Compton*. El ajuste de esta segunda componente es más exigente en cuanto a recursos, ya que la complejidad de la expresión del flujo es bastante alta. Para esta segunda componente el error más pequeño encontrado es de ≈ 0.041 .

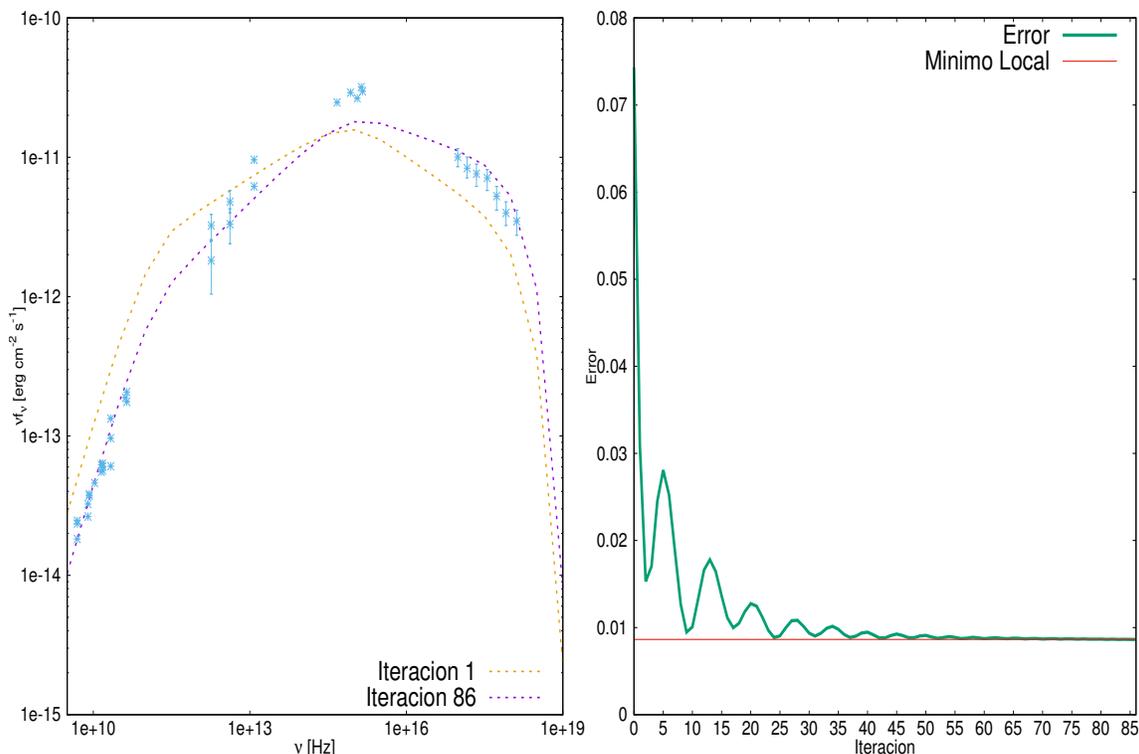


Figura 5.6: Ronda 4 de ajuste. Iteraciones totales: 86. El ajuste convergió a un mínimo local con un error de ≈ 0.0086 . Parámetros finales: $\delta_D = 34.752$, $B = 85.85$, $\log(K_e) = 44.1$, $p = 2.25$, $\log(\gamma_{break}) = 3.964$, $\log(\gamma_2) = 5.541$.

El tiempo que le tomó a nuestro algoritmo ajustar el flujo varió ligeramente en cada ronda; pero en nuestros equipos, el ajuste de la primera componente tomó un tiempo promedio de 6 minutos. Mientras que el de la segunda componente fue de 3 horas y 27 minutos. Es clara la diferencia de complejidad entre ambas partes, aunque como se puede observar en la figura 5.7, el algoritmo encontró rápidamente el mínimo local, pero la influencia del momento en el ajuste le mantuvo rebotando ligeramente alrededor del mínimo; un número de precisión más alto hubiera detenido el algoritmo mucho antes sin afectar la eficacia del resultado final.

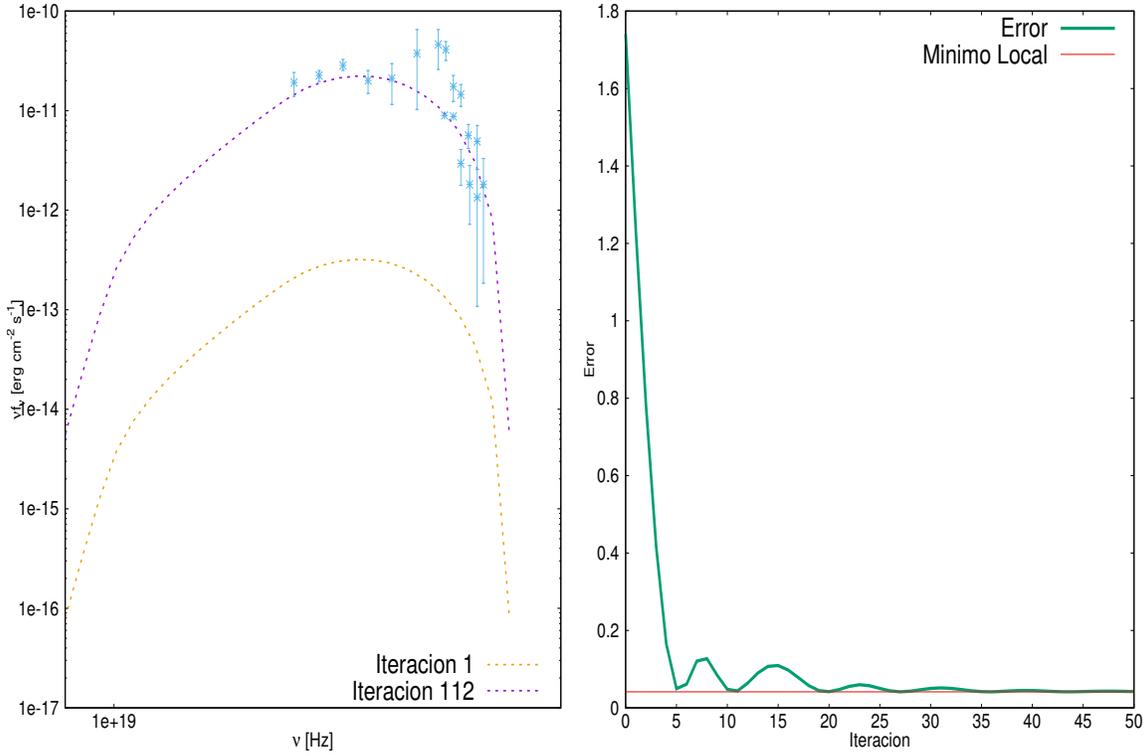


Figura 5.7: Ronda de ajuste de SSC. Iteraciones totales: 112. El ajuste convergió a un mínimo local con un error de ≈ 0.041 . Parámetros finales: $\log(t_{v,min}) = 3.78$.

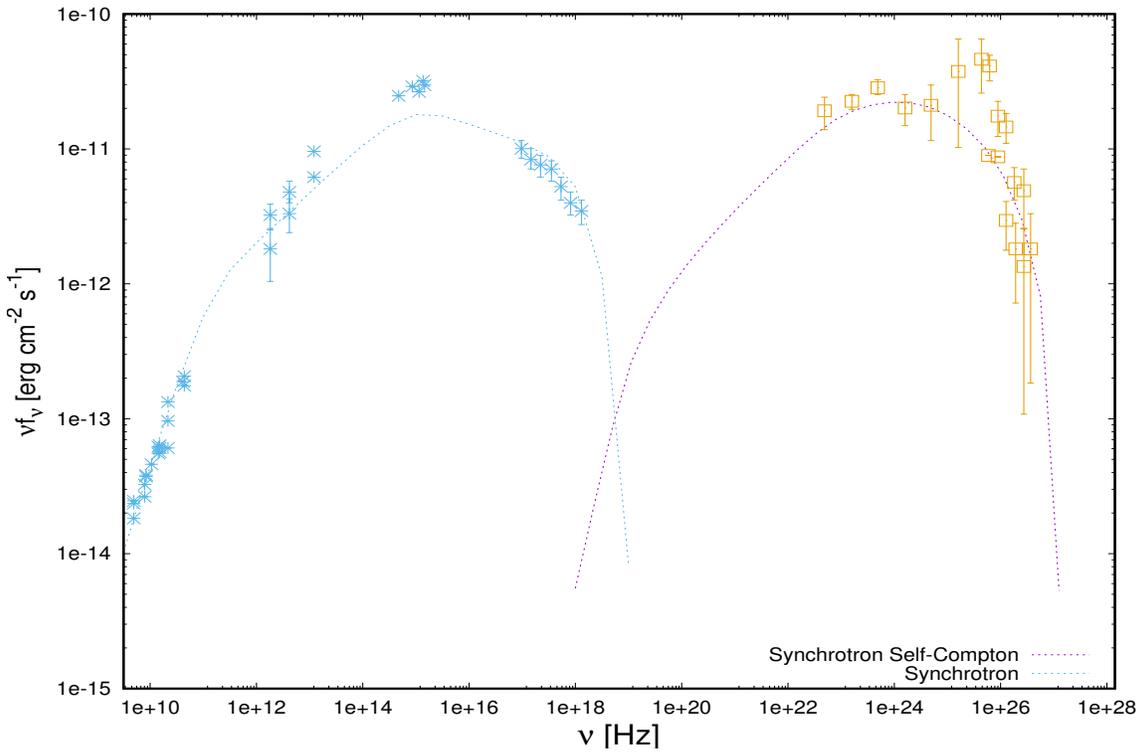


Figura 5.8: SED final para el blazar W Comae

Capítulo 6

Conclusiones

El método de descenso por el gradiente es un algoritmo cuyo principal uso es en el área de *Machine Learning*, pero su utilidad es aplicable en otros campos donde también existen problemas de minimización. Como ya mostramos, obtuvimos resultados satisfactorios al ajustar el flujo de W Comae.

Basado en todas nuestras rondas de ajuste, concluimos que la expresión del flujo contiene muchos mínimos locales. La mayoría de estos no corresponden a un valor de error satisfactorio para ajustar el SED. A pesar de esto, acotando de forma correcta los parámetros iniciales, es posible asegurar encontrar un mínimo local suficientemente acertado, como ya demostramos con nuestros resultados.

Aunque logramos minimizar el error lo más posible, hay que notar que en la segunda componente, la curva no se ajusta muy bien para una parte del flujo, específicamente alrededor de los $10^{23} Hz$. Esta subida de energía podría ser explicada por una componente hadrónica ([KA08]) que no consideramos al realizar la simulación.

Esta misma versión del algoritmo puede ser usada para ajustar el flujo de otros blázares, además de que el tiempo necesario para realizar un ajuste es aceptable. La complejidad del ajuste es del mismo orden que la complejidad de la función de flujo en sí, pues calcular la derivada de primer orden de una función tiene la misma complejidad que ésta. Y es ahí donde otra ventaja de este algoritmo puede ser notada, pues no está limitado a la expresión de flujo que utilizamos aquí; este algoritmo puede recibir cualquier función y realizar un ajuste utilizando la entrada correcta. Por lo

anterior, este mismo método puede ser utilizado para ayudar a describir otros objetos físicos.

Ahora quisiera agregar que, como computólogo, las habilidades de análisis de algoritmos me fueron muy útiles al discutir la complejidad del algoritmo de ajuste, así como la de las funciones de flujo que utilizamos para realizar la simulación. De igual forma, no podría haber realizado las simulaciones necesarias sin la formación matemática que la carrera de Ciencias de la Computación me facilitó. Por último, la educación en física que recibí gracias a algunas materias optativas fue fundamental en la realización de este trabajo.

Bibliografía

- [Eld+47] F. R. Elder y col. «Radiation from Electrons in a Synchrotron. (Inglés) [Radiación de electrones en un sincrotrón]». En: *Physical Review* 71 (1947), pág. 829. DOI: <https://doi.org/10.1103/PhysRev.71.829.5>.
- [AH50] H. Alfvén y N. Herlofson. «Cosmic Radiation and Radio Stars. (Inglés) [Radiación cósmica y estrellas de radio]». En: *Physical Review* 78 (1950), pág. 616. DOI: <https://doi.org/10.1103/PhysRev.78.616>.
- [Bur56] G. R. Burbidge. «On Synchrotron Radiation from Messier 87. (Inglés) [Sobre la radiación de sicrotrón de Messier 87]». En: *Astrophysical Journal* 124 (1956), pág. 416. DOI: <https://doi.org/10.1086/146237>.
- [Bro71] I. W. A. Browne. «Two Bright New Quasi-stellar Radio Sources. (Inglés) [Dos nuevas y brillantes fuentes de radio cuasi-estelares]». En: *Nature* 231 (1971), págs. 515-516. DOI: <https://doi.org/10.1038/231515a0>.
- [CS86] A. Crusius y R. Schlickeiser. «Synchrotron Radiation in Random Magnetic Fields. (Inglés) [Radiación de sincrotrón en campos magnéticos aleatorios]». En: *Astronomy and Astrophysics (ISSN 0004-6361)* 164.2 (1986), págs. L16-L18.
- [UP95] C. M. Urry y Padovani P. «Unified Schemes for Radio-Loud Active Galactic Nuclei. (Inglés) [Esquemas unificados para núcleos galácticos activos de radio fuerte]». En: *Astronomical Society of the Pacific* 107.715 (1995). DOI: <https://doi.org/10.1086/133630>.

- [GM98] Tosti G. y Fiorucci M. «BV R_cI_c monitoring of ON231 during the great outburst in 1994-1997. (Inglés) [Monitoreo de BV R_cI_c de ON231 durante el gran estallido en 1994-1997]». En: *Astronomy and Astrophysics Supplement Series* 130.1 (1998), págs. 109-115. DOI: <https://doi.org/10.1051/aas:1998215>.
- [Har+99] R. C. Hartman y col. «The Third Egret Catalog of High-Energy Gamma-Ray Sources. (Inglés) [El catálogo de Third Egret de fuentes de rayos gamma de alta energía]». En: *The Astrophysical Journal Supplement Series* 123.1 (1999), págs. 79-202. DOI: <https://doi.org/10.1086/313231>.
- [Mas+99] E. Massaro y col. «The extraordinary optical outburst of ON231 (W Com) in spring 1998. (Inglés) [El extraordinario estallido óptico de ON231 (W Com) en la primavera de 1998]». En: *Astronomy and Astrophysics* 342 (1999), págs. L49-L52.
- [Urr+99] C. M. Urry y col. «The Hubble Space Telescope Survey of BL Lacertae Objects. II. Host Galaxies. (Inglés) [La Inspección del Telescopio Espacial Hubble de Objetos BL Lacertae. II. galaxias anfitrionas]». En: *The Astrophysical Journal* 532.2 (1999). DOI: <https://doi.org/10.1086/308616>.
- [Pet+00] D. Petry y col. «Multiwavelength Observations of Markarian 501 during the 1997 High State. (Inglés) [Observaciones de longitud de onda múltiple de Markarian 501 durante el estado alto de 1997]». En: *The Astrophysical Journal* 309.2 (2000). DOI: <https://doi.org/10.1086/308955>.
- [Har+01] R. C. Hartman y col. «Multiepoch Multiwavelength Spectra and Models for Blazar 3C 279. (Inglés) [Espectros y modelos de varias épocas y longitudes de onda múltiples para Blazar 3C 279]». En: *The Astrophysical Journal* 553.2 (2001). DOI: <https://doi.org/10.1086/320970>.
- [Böt+03] M. Böttcher y col. «Coordinated Multiwavelength Observations of BL Lacertae in 2000. (Inglés) [Observaciones coordinadas de longitudes de

- onda múltiples de BL Lacertae en 2000]». En: *The Astrophysical Journal* 596.2 (2003). DOI: <https://doi.org/10.1086/378156>.
- [Böt06] M Böttcher. «Modeling the emission processes in blazars. (Inglés) [Modelado de los procesos de emisión en blazares]». En: *Astrophys Space Sci* 309 (2006), págs. 95-104. DOI: <https://doi.org/10.1007/s10509-007-9404-0>.
- [Acc+08] V. A. Acciari y col. «VERITAS Discovery of >200 GeV Gamma-Ray Emission from the Intermediate-Frequency-Peaked BL Lacertae Object W Comae. (Inglés) [Descubrimiento de VERITAS de una emisión de rayos gamma de >200 GeV del objeto BL Lacertae W Comae de pico de frecuencia intermedia]». En: *The Astrophysical Journal* 684.2 (2008), pág. L73. DOI: <https://doi.org/10.1086/592244>.
- [FDB08] Justin D. Finke, Charles D. Dermer y Markus Böttcher. «Synchrotron Self-Compton Analysis of TeV X-Ray-Selected BL Lacertae Objects. (Inglés) [Análisis de sincrotrón y self-Compton de objetos BL Lacertae seleccionados por rayos X en TeV]». En: *The Astrophysical Journal* 686.1 (2008). DOI: <https://doi.org/10.1086/590900>.
- [KA08] S.R. Kelner y F.A. Ahanorian. «Energy spectra of gamma-rays, electrons and neutrinos produced at interactions of relativistic protons with low energy radiation. (Inglés) [Espectros de energía de rayos gamma, electrones y neutrinos producidos en interacciones de protones relativistas con radiación de baja energía]». En: *Monthly Notices of the Royal Astronomical Society* 302.2 (2008), págs. 373-380. DOI: <https://doi.org/10.1046/j.1365-8711.1999.02132.x>.
- [Böt12] M. Böttcher. «Modeling the Spectral Energy Distributions and Variability of Blazars. (Inglés) [Modelado de distribuciones de energía espectral y variabilidad de los blazares]». En: (2012). DOI: <https://arxiv.org/abs/1205.0539>.

- [HSS12] Geoffrey Hinton, Nitish Srivastava y Kevin Swersky. *Neural Networks for Machine Learning. Lecture 6d: A separate, adaptive learning rate for each connection. (Inglés) [Redes Neuronales para Machine Learning. Lección 6d: Una tasa de aprendizaje adaptativa separada para cada conexión]*. 2012. URL: <http://www.cs.toronto.edu/~hinton/coursera/lecture6/lec6.pdf>.
- [Fra14] Reddy Francis. *Black Hole ‘Batteries’ Keep Blazars Going and Going. (Inglés) [‘Baterias’ de agujero negro mantienen a los Blazars funcionando y funcionando]*. 2014. URL: <https://svs.gsfc.nasa.gov/11563>. (accessed: 05.03.2022).
- [KB14] Diederik Kingma y Jimmy Ba. «Adam: A Method for Stochastic Optimization. (Inglés) [Adam: Un método para optimización estocástica]». En: *International Conference on Learning Representations (2014)*.
- [Sor+14] Marco Sorcia y col. «Rotation of the Optical Polarization Angle Associated with the 2008 γ -ray Flare of Blazar W Comae. (Inglés) [Rotación del ángulo de polarización óptica asociado con la llamarada de rayos γ de 2008 de Blazar W Comae]». En: *The Astrophysical Journal* 794.1 (2014). DOI: <https://doi.org/10.48550/arXiv.1408.1339>.
- [And+16] Marcin Andrychowicz y col. «Learning to learn by gradient descent by gradient descent». En: *Advances in Neural Information Processing Systems*. Ed. por D. Lee y col. Vol. 29. Curran Associates, Inc., 2016. URL: <https://proceedings.neurips.cc/paper/2016/file/fb87582825f9d28a8d42c5e5e5e8Paper.pdf>.
- [Gyl18] Roan Gylberth. *An Introduction to AdaGrad. (Inglés) [Una introducción en AdaGrad]*. 2018. URL: <https://medium.com/konvergen/an-introduction-to-adagrad-f130ae871827>. (accessed: 04.03.2022).