



Universidad de Buenos Aires
Facultad de Ciencias Exactas y Naturales
Departamento de Física

Tesis de Licenciatura

**Soluciones cosmológicas de teorías
alternativas de gravedad con redes
neuronales**

Luca J. Gómez Bachar

Directora: Dra. Susana Landau

Codirector: Dr. Pavlos Protopapas

28 de marzo de 2024

TEMA: COSMOLOGIA Y MACHINE LEARNING

ALUMNO: LUCA J. GOMEZ BACHAR

LU N^o: 599/18

LUGAR DE TRABAJO: Facultad de Ciencias Exactas y Naturales

DIRECTORA DEL TRABAJO: Dra. Susana Landau

CO-DIRECTOR DEL TRABAJO: Dr. Pavlos Protopapas

FECHA DE INICIACIÓN: MARZO 2023

FECHA DE FINALIZACIÓN: MARZO 2024

FECHA DE EXAMEN: MARZO 2024

INFORME FINAL APROBADO POR:

Autor:

Jurado

Director:

Jurado

Profesor:

Jurado

Resumen

La cosmología moderna ha logrado una descripción precisa del Universo a través del modelo estándar Λ CDM. El éxito de dicho modelo consiste en poder explicar la mayoría de los datos observacionales actuales. Sin embargo, presenta también algunos problemas tanto del punto de vista teórico como con la contrastación de los datos. Por este motivo, se estudian modelos cosmológicos alternativos. En este contexto, la inferencia de parámetros juega un rol central en la cosmología moderna, debido a que la determinación de estos permite realizar tests de distintos modelos cosmológicos, algunos de los cuales introducen soluciones a las mencionadas fallas del modelo estándar.

En los últimos tiempos, las técnicas de machine learning han sido aplicadas con éxito al campo de la astrofísica y cosmología. Entre los métodos que se han desarrollado, existe uno que permite resolver ecuaciones diferenciales sin requerir de los métodos numéricos tradicionales. En esta tesis se entrenaron redes neuronales artificiales con el objetivo de resolver la ecuación diferencial que describe la evolución de las perturbaciones de materia en el Universo. Se estudiaron dos modelos cosmológicos diferentes: el modelo estándar de la cosmología (Λ CDM) y un modelo fenomenológico de gravedad modificada. En ambos casos se obtuvieron redes neuronales que presentaron errores menores al 1 % para todo el espacio de parámetros estudiado, comparando sus soluciones con las obtenidas a través de métodos numéricos tradicionales. Para poder emplear las soluciones de las redes en un análisis estadístico se utilizó el método bundle en el entrenamiento. Este método consiste en que las redes aprendan a resolver la ecuación, no solo en términos de la variable independiente, sino también en función de los parámetros que se desean inferir. Las soluciones obtenidas se emplearon para realizar la inferencia de parámetros cosmológicos a través del método de integración de Montecarlo mediante cadenas de Markov (MCMC) donde se obtuvieron estimaciones para la densidad de energía en forma de materia Ω_{m0} y del parámetro σ_8 que son compatibles con estimaciones actuales.

Agradecimientos

"Hay que tener equilibrio, porque la derrota te muestra lo peor de vos y la victoria te confunde"

Marcelo D. Gallardo

A veces cierro los ojos, me imagino no haber hecho el enorme esfuerzo que requiere terminar una carrera y me agradezco haber seguido el camino, pero en la confusión de la victoria debo reconocer que este logro no hubiera sido posible sin la ayuda de una enorme cantidad de gente. En primer lugar quiero agradecerle enormemente a mis viejos, Javier y Valeria, que siempre me bancaron a fondo, así como a mis hermanas, Belu, Sofi y Clari que con todo el cariño que me dan me empujan a seguir, aunque una de ellas sea un bot en Stumble. Me es imposible no agradecerle a mi madrina Gaby, quien además de apoyarme incondicionalmente me regaló la compu con la que fue posible hacer gran parte de la carrera, y en particular este trabajo. También le agradezco a mis tíos y tias, primos y primas, con quiénes he compartido gran parte de este camino. Tampoco puedo olvidarme de mis abuelas Alicia y Elena, que nunca se negaron a ayudarme con una comida, o con una frase de aliento.

Pasando al ámbito académico, le agradezco enormemente a Susana por la paciencia, el conocimiento, las charlas y su gran experiencia; pero sobre todas las cosas le agradezco por su gran valor humano que hizo de este año de trabajo algo totalmente placentero, enriquecedor y productivo. Gracias a ella aprendí una gran cantidad de cosas respecto a la cosmología, a las redes, y a la vida académica de la investigación, además de siempre haberme sentido acompañado. En este contexto, no puedo dejar de agradecerle a Pavlos, quien a distancia y con un gran sentido del humor, me abrió las puertas a un mundo hermoso y desconocido para mí: el de la inteligencia artificial. También le agradezco a Augusto y Claudia, quienes han sido de gran ayuda a lo largo de esta tesis. Además, debo agradecer a quienes me ayudaron enormemente a entrar al mundo de la cosmología: Diana, cuyas clases me mostraron lo lindo que es este campo; Nahue, quien con su paciencia, buena onda, sinceridad y experiencia me abrió las puertas a entender cosas que no son nada fáciles, además de guiarme y aconsejarme; y Mati, quien también me ayudó muchísimo y demostró ser una gran persona en el camino. A esta gente se suma el hermoso grupo de cosmo del D.F., quiénes siempre me abrieron las puertas y me hicieron sentir parte de ellos, tanto en el día a día como en los Journals. Espero que la victoria de haber terminado la licenciatura no me confunda, y en ese caso espero que la gran cantidad de amigos y amigas que me dio la carrera estén ahí como siempre. Tengo que agradecer profundamente a Maia que siempre estuvo ahí para hacer todo más divertido y

llevadero (salvo para la defensa, cof,cof); al filósofo contemporáneo (a quien no voy a nombrar porque se agranda y es infumable) que, de forma gentil y glamorosa, me dijo "dejá de llorar y estudia, vivís una vida de privilegios" cuando estaba en la Noriega reflexionando si dar o no el final de teo 3; a Andy "Juanfer" Tovar, que de Juanfer tiene todo salvo los dotes futbolísticos; así como también a Sulei, Manu, Dany, Nico, Lulú, Pedro, Marian y todas las demás personas que hicieron de estos años una experiencia hermosa.

Finalmente, le agradezco profundamente a toda la gente del Club Comunicaciones que siempre me bancó, desde mis amigos y compañeros de toda la vida quienes no merecen ni necesitan ser nombrados, hasta todos los chicos y chicas que formaron parte de todo el trabajo de estos años, con quiénes aprendí tantas cosas... Me hubiera sido imposible llegar hasta acá sin ese cable a tierra, sin los jueves anti-pala y sin el apoyo incondicional que me hacen sentir en el día a día.

Como último comentario, le agradezco a la educación pública, incluyendo a todxs lxs profes que tuve, ya que desde el jardín de infantes hasta la licenciatura en física jamás pagué un peso por estudiar y formarme. Entre toda la gente que forma parte de este recorrido, quiero destacar a Lucía Famá que hizo que un año entero de laboratorio sea súper interesante y llevadero. Sin lugar a dudas llegar a este punto no hubiera sido posible sin este sistema público, el cual debemos reivindicar y defender.

"Viniste sin temores. No te arrepientas, sin importar lo que suceda. Este es el duelo que querías."

Monkey D. Luffy

"Ha sido una historia hermosísima"

Marcelo D. Gallardo

Índice de figuras

2.1.	Evolución del radio de Hubble comóvil en función de a . Las líneas verticales representan, en azul el final de inflación ($a_e = 10^{-25}$), en naranja a_{eq} y en verde a_{eq}^Λ . El valor de $(aH)^{-1}$ representa una escala importante durante la evolución del Universo.	33
2.2.	Evolución del radio de Hubble comóvil en función de a , junto con un modo de Fourier de ejemplo. En esta escala los modos se representan a través de cada λ_{com} como líneas horizontales. Las líneas verticales representan el valor de a en el cual el modo sale del horizonte (izquierda) y el valor en que entra (derecha).	34
2.3.	Representación pictórica de la región de valores de k que se pueden estudiar con la ecuación 2.43. El modo en naranja corresponde a aquel que entra al horizonte cuando se imponen las condiciones iniciales, mientras que el que está en verde corresponde al primer modo no lineal. En gris todos los valores que se pueden estudiar mediante el modelo teórico empleado.	35
2.4.	G_{eff}/G en función del factor de escala para varios valores de k dentro del régimen de estudio. Se observa que para valores pequeños de a $G_{\text{eff}} \rightarrow G$, lo cual se espera en las teorías $f(R)$, mientras que en el Universo tardío hay diferencias apreciables.	41
2.5.	Diferencias porcentuales entre el valor de G_{eff}/G para varios valores de k dentro del rango de estudio, los cuales no se incluyen por ser irrelevantes. Se observan diferencias de hasta el 30 %, lo cual permite concluir que en esta teoría la dependencia en los modos de Fourier de la ecuación diferencial es no trivial.	42
3.1.	Estructura de una neurona artificial. Esta recibe entradas, o inputs, los cuales se afectan a través de un factor multiplicativo llamado peso. Luego de sumar todos los inputs con su correspondiente peso al bias, o sesgo, se aplica una función de activación para obtener el resultado final de esta neurona.	46
3.2.	Estructura de ejemplo de una red neuronal con tres entradas, dos capas ocultas con cuatro neuronas cada una y una capa externa con una única neurona. Se observa que todas las neuronas están conectadas con todas de la siguiente capa, lo cual significa que este esquema corresponde a una red de conexión completa.	47
3.3.	A la izquierda la tangente hiperbólica, a la derecha la función SiLU. En ambos casos las funciones son no lineales lo cual evita el colapso de la red a una única neurona. Figura extraída del libro [1].	49

3.4.	Esquema general de una red neuronal con $r + 1$ entradas, S capas internas con K_j neuronas en cada capa ($j = 1, \dots, S$) y una única neurona en la capa externa.	55
3.5.	Ejemplo de red neurona con la arquitectura customizada. En la figura se ve una capa externa con una única entrada y la primer capa oculta con K_1 neuronas. Además de los pesos se incluye en cada neurona un valor t_j que corresponde al valor central de una gaussiana y además la cantidad σ_j que corresponde a la desviación estándar de esta gaussiana.	57
6.1.	Solución numérica para δ y δ' en función del factor de escala a para el modelo Λ CDM con $\Omega_{m0} = 0,272$	78
6.2.	Solución numérica para δ comparada con una función lineal en función del factor de escala a para el modelo Λ CDM con $\Omega_{m0} = 0,272$	78
6.3.	$f\sigma_8$ vs a obtenido con el método numérico en función del factor de escala, junto con los datos de la tabla 5.1. Los parámetros cosmológicos empleados para el cálculo corresponden a $\Omega_{m0} = 0,272$ y $\sigma_8 = 0,8$	79
6.4.	$f\sigma_8$ vs z obtenido con el método numérico en función del corrimiento al rojo, junto con los datos de la tabla 5.1. Los parámetros cosmológicos empleados para el cálculo corresponden a $\Omega_{m0} = 0,272$ y $\sigma_8 = 0,8$. Esta figura puede compararse directamente con la literatura [2].	80
6.5.	Solución numérica para δ y δ' en función del factor de escala a para el modelo fenomenológico de gravedad modificada. Los parámetros corresponden a $\Omega_{m0} = 0,21$, $\sigma_8 = 0,88$ y $g_a = -1,16$	81
6.6.	$f\sigma_8$ vs z para el modelo fenomenológico de gravedad modificada, junto con los datos de la tabla 5.1. Esta figura puede compararse directamente con la literatura [2]. Los parámetros corresponden a $\Omega_{m0} = 0,21$, $\sigma_8 = 0,88$ y $g_a = -1,16$	81
6.7.	Funciones de pérdida para entrenamientos sin bundle llevados a cabo mediante la arquitectura habitual con función de activación tangente hiperbólica (a) y SiLU (b); y arquitectura customizada con tangente hiperbólica (c) y SiLU (d).	82
6.8.	Errores porcentuales (ecuación 6.1) en función del factor de escala para entrenamientos sin bundle llevados a cabo mediante la arquitectura habitual con función de activación tangente hiperbólica (a) y SiLU (b); y arquitectura customizada con tangente hiperbólica (c) y SiLU (d).	84
6.9.	Funciones de pérdida para entrenamientos con bundle llevados a cabo mediante la arquitectura habitual con función de activación tangente hiperbólica (a) y SiLU (b); y arquitectura customizada con tangente hiperbólica (c) y SiLU (d).	85
6.10.	Errores porcentuales de $\delta(a = 1)$ y δ'_{\max} en función de Ω_{m0} para entrenamientos sin bundle llevados a cabo mediante la arquitectura habitual con función de activación tangente hiperbólica (a) y SiLU (b); y arquitectura customizada con tangente hiperbólica (c) y SiLU (d).	87

6.11. Errores porcentuales para δ (a) y δ' (b) en función del parámetro del bundle (Ω_{m0}) y del factor de escala empleando la función de activación SiLU y la arquitectura customizada.	87
6.12. Función de pérdida para un entrenamiento con bundle realizado para el modelo fenomenológico de gravedad modificada. Para dicho entrenamiento se empleó la arquitectura customizada y la función de activación SiLU.	88
6.13. Errores porcentuales para δ evaluado en $a = 1$ (hoy), y el valor máximo para δ'	89
6.14. Errores porcentuales de δ y δ' en función del factor de escala y Ω_{m0} para dos valores del parámetro g_a : $-0,9$ y $-0,8$	90
6.15. Cálculo del χ^2 mediante el método numérico.	91
6.16. Cálculo del χ^2 mediante el método de las redes neuronales.	91
6.17. Cadenas de Markov obtenidas mediante el método de MCMC para ambos métodos en el modelo Λ CDM. En ambos casos se observa que las cadenas no chocan con los priors, lo cual es importante para garantizar una estimación correcta.	92
6.18. Distribuciones posterior y contornos de confianza para el modelo Λ CDM obtenidas mediante el método de MCMC.	93
6.19. Comparación entre los parámetros obtenidos por medio de la inferencia con el MCMC y los reportados en Planck 2018 [3].	94
6.20. Distribuciones posterior y contornos de confianza para el modelo fenomenológico de gravedad modificada obtenidas mediante el método de MCMC.	96
6.21. G_{eff}/G vs a . Se observa que las correcciones introducidas dentro del intervalo en que se encuentran los datos observacionales de $f\sigma_8$ son tales que $G_{\text{eff}}/G \approx 1$, lo cual explica la descorrelación entre g_a y los otros parámetros del modelo.	97
7.1. Puntos en el espacio de parámetros que cumplen la condición $14,25 < \chi^2 < 14,30$. En rojo se resalta al conjunto de parámetros que se eligió como de mejor ajuste debidoa que representa el mínimo global de χ^2	104
7.2. Cortes bidimensiones de la figura 7.1. En (a) y (c) se puede observar una delimitación entre el mínimo de mejor ajuste y los siguientes mínimos. De esta manera se definió el prior en g_a como el intervalo delimitado entre las dos líneas punteadas.	104
7.3. Cadenas de Markov obtenidas mediante el método de MCMC para ambos métodos en el modelo fenomenológico de gravedad modificada.	105

Índice de tablas

5.1. Datos observacionales de la cantidad $f\sigma_8$. Las columnas corresponden a: un índice, el nombre del experimento en cuestión, el redshift para el cual se mide, el valor reportado del observable junto con su error, la referencia y el valor fiducial de los parámetros adoptados para la medición.	73
6.1. Estimación a priori de los parámetros mediante el método de máxima verosimilitud para el método numérico tradicional y para el método de las redes neuronales.	92
6.2. Intervalos de confianza para el modelo Λ CDM obtenidos con el método numérico tradicional y el método de las redes neuronales.	93
6.3. Estimación a priori mediante el criterio de máxima verosimilitud para el modelo fenomenológico de gravedad modificada para los dos métodos empleados.	96
6.4. Intervalos de confianza obtenidos para el modelo fenomenológico de gravedad modificada utilizando el método numérico tradicional y el método de las redes neuronales.	97
6.5. Tiempos en los cuales se completó el proceso de MCMC, para el modelo estándar Λ CDM y el modelo fenomenológico de gravedad modificada.	98
7.1. Valores de χ^2 en función de los parámetros del modelo. Se observa una degeneración en el parámetro g_a , dado a que dentro de variaciones de aproximadamente el 0,3% de χ^2 se ven variaciones de hasta aproximadamente el 150% en los valores de dicho parámetro.	103

Índice general

1. Introducción	9
2. Modelo teórico	17
2.1. Ecuaciones de fondo cosmológico en Λ CDM	17
2.2. Ecuación de perturbaciones de materia en Λ CDM	20
2.3. Las condiciones iniciales en Λ CDM	25
2.3.1. Etapas de evolución del Universo	25
2.3.2. Dinámica cosmológica de un Universo dominado por materia	26
2.3.3. Evolución de las perturbaciones de materia en un Universo dominado por materia	27
2.4. El observable $f\sigma_8$	29
2.5. Los modos de Fourier dentro de la aproximación sub-horizon	31
2.6. Ecuaciones de fondo en $f(R)$	35
2.7. Ecuación de perturbaciones en $f(R)$	36
2.8. Modelo de Hu-Sawicki	39
2.9. Modelo fenomenológico de perturbaciones de materia con gravedad modificada . .	42
3. Método de las redes neuronales artificiales	45
3.1. Neuronas artificiales	45
3.2. Estructura de la red neuronal	46
3.3. Las funciones de activación	48
3.4. Aprendizaje de la red neuronal	49
3.4.1. La función de pérdida	50
3.4.2. El problema de optimización	52
3.5. Resolución de ecuaciones diferenciales	53
3.6. El método bundle	54
3.7. La arquitectura customizada	56
4. Implementaciones del método de las redes neuronales	59
4.1. Implementación a los modelos teóricos	59
4.1.1. Implementación a la ecuación para perturbaciones de materia en Λ CDM .	60
4.1.2. Implementación a la ecuación para perturbaciones de materia en el modelo fenomenológico	65

4.2. Análisis estadístico con cadenas de Markov MonteCarlo (MCMC) usando las soluciones obtenidas con el método de redes neuronales	66
5. Datos	71
5.1. Detalles observacionales de $f\sigma_8$	71
5.2. Detalles de implementación de los datos	74
5.2.1. Correcciones fiduciales para el cálculo de χ^2	74
5.2.2. Correlación entre los datos	75
6. Resultados	77
6.1. Soluciones numéricas	77
6.1.1. La solución numérica para Λ CDM	77
6.1.2. La solución numérica para el modelo fenomenológico	80
6.2. Entrenamientos sin bundle para Λ CDM	82
6.3. Soluciones bundle para los modelos cosmológicos	84
6.3.1. Método bundle para Λ CDM	84
6.3.2. Método bundle para el modelo fenomenológico de gravedad modificada . .	88
6.4. Inferencia de parámetros cosmológicos	89
6.4.1. Estimación de los parámetros en Λ CDM	89
6.4.2. Estimación de los parámetros en el modelo fenomenológico de gravedad modificada	95
6.5. Tiempos computacionales en el MCMC	98
7. Conclusiones	101
Apéndice A: estimación del prior de g_a	103

Capítulo 1

Introducción

El estudio del Universo posee un rol importante en la sociedad desde los comienzos de esta misma. Preguntas referidas a nuestro porvenir o hacia nuestro rol en el escenario cósmico ocupan a científicos y pensadores desde hace mucho tiempo, dado que comprender el Universo en que se nos es dado ser no solo es fascinante, sino que constituye un tipo de conocimiento que puede cambiar nuestra forma de percibir el mundo.

A principios del siglo XX, cuando los astrónomos comenzaron a observar la distribución de las galaxias en el Universo se dieron cuenta de que, a grandes escalas, parecían estar distribuidas de manera uniforme en el cielo. Esta observación fue un punto de partida importante para la cosmología y sentó las bases para postular una de las hipótesis más cruciales en el estudio del cosmos: el principio cosmológico. Este principio establece que a gran escala espacial el Universo es isótropo y homogéneo, donde la isotropía refiere a que el espacio es igual en todas las direcciones y homogeneidad significa equivalencia entre todos los puntos.

En 1920 con el desarrollo de la teoría de la relatividad general el estudio del Universo se colocó dentro de un marco teórico robusto que permite hacer predicciones sobre la evolución del mismo. Uno de los primeros hitos importantes en la evolución de la cosmología moderna fue la solución cosmológica de la ecuación de campo de Einstein presentada por Alexander Friedmann en 1922 y Georges Lemaître en 1927, la cual se apoya fuertemente en el principio cosmológico. Estos modelos propusieron un universo en expansión, una idea radical en ese momento, que contradecía la concepción estática del mismo. El descubrimiento fortuito de la expansión del universo por Edwin Hubble en la década de 1920 [4] proporcionó un apoyo observacional crucial para esta teoría. Hubble observó que todas las galaxias se alejan de nosotros, y a su vez que, las galaxias más distantes se alejan más rápido, llegando algunas a moverse a velocidades superlumínicas. Esto contradice a la relatividad especial en el contexto de un Universo estático, motivo que puede llevar a postular que el Universo se encuentra en expansión. Esto explica porque todas las galaxias se alejan de nosotros y también las velocidades superlumínicas [5]. Einstein no concordaba con la hipótesis de un universo en expansión, y en búsqueda de la descripción de un Universo estático, introdujo una constante a sus ecuaciones: la constante cosmológica (Λ). Al incluir esta constante se obtiene un universo estático, pero inestable, motivo por el cual el mismo Einstein decidió descartarla.

En la década de 1930, Lemaître quien era físico y sacerdote, propuso una idea revolucionaria

que cambiaría para siempre nuestra comprensión del Universo: la teoría del Big Bang. Esta teoría sugiere que el universo no es estático ni eterno, sino que tuvo un comienzo definido en un evento que Lemaître describió como una "singularidad cósmica". En otras palabras, toda la materia, energía y espacio-tiempo del universo estaban concentrados en un punto infinitesimal de densidad y temperatura infinitas. La teoría del Big Bang proporcionó un marco teórico unificado para comprender varios fenómenos observados en el universo. En particular, sirve como condición inicial para la solución cosmológica mencionada previamente y además ayudó a explicar la expansión cósmica observada por Hubble y otros astrónomos. También, predijo la existencia de la radiación cósmica de fondo de microondas (FCR), un remanente de las altas energías del Big Bang que fue descubierto décadas más tarde y reforzó la validez de esta teoría.

Con el devenir de nuevas tecnologías que permitieron el desarrollo de nuevos y mejores instrumentos, la cantidad y la calidad de los datos observacionales experimentó un aumento significativo. Estas nuevas observaciones permiten contrastar las diversas predicciones de la teoría y determinar el valor de los parámetros cosmológicos. De esta manera, la cosmología se convirtió en una rama científica de precisión. Por ejemplo, la década de 1960 fue testigo del descubrimiento de la mencionada radiación cósmica de fondo de microondas [6], lo que sustentó de manera contundente la teoría del Big Bang. Luego de su descubrimiento mediciones más precisas realizadas a finales del siglo XX acerca de la intensidad de la luz proveniente del FCR en función de su longitud de onda [7] muestran una gran coincidencia con un espectro de cuerpo negro, lo cual da un soporte aún más sólido al modelo propuesto por Lemaitre. [8].

La medición del FCR sustenta además la predicción de Hubble de un Universo en expansión. En el contexto de la teoría de la relatividad general y asumiendo un Universo con materia y radiación, el Universo se expande de forma desacelerada. En contraposición a estas predicciones, a finales de siglo XX se observó, mediante el estudio de supernovas tipo Ia (SNIa), que el Universo se expande de manera acelerada [9]. En el contexto del modelo cosmológico de ese momento la forma más simple de explicar este tipo de expansión era incluir nuevamente la constante cosmológica (Λ) en las ecuaciones, no para describir un Universo estático, sino para explicar la aceleración de la expansión. No obstante esta constante no posee un valor medido precisamente, como por ejemplo la velocidad de la luz en vacío. Si se adjudica la existencia de la constante cosmológica a la energía del vacío y se realiza la estimación de valor de la misma usando la teoría cuántica de campos, se obtiene un resultado que difiere entre 60 y 120 órdenes de magnitud con el valor medido de esta constante. Estas discrepancias entre el valor teórico de Λ y su correspondiente observación experimental dependen de la escala de corte que se emplee al realizar los cálculos mediante una teoría cuántica de campos, resultando en 60 órdenes de magnitud si se emplea como corte la escala de Planck, mientras que los 120 órdenes de discrepancia se obtienen empleando como escala de corte a aquella que corresponde a supersimetría [10–13]. Este resultado pone en tensión la propia definición de Λ y se lo llama el problema de la constante cosmológica [10]. A pesar de esta cuestión, la constante cosmológica sigue siendo a día de hoy la manera estándar de comprender el Universo, dado que el modelo estándar de la cosmología actual se apoya fuertemente en su existencia.

Por otro lado, observaciones pertinentes a las galaxias mostraron que estas parecen tener

más materia que la se puede observar. Un ejemplo de estas observaciones corresponde a las curvas de rotación de las galaxias, donde se estudia la velocidad de rotación según la distancia al centro galáctico [14]. En dichas curvas se observa que estas velocidades no se corresponden con la materia luminosa (también llamada materia bariónica) que se detecta con los telescopios. Esto motivó a incluir en la descripción del Universo un tipo particular de materia que no tiene interacción electromagnética con otras partículas, y, por lo tanto, no absorbe ni emite radiación electromagnética: la materia oscura fría (CDM) [15]. Se la llama materia oscura porque únicamente interactúa con el resto de los componentes del Universo a través de la gravedad, y, por lo tanto, como no absorbe ni emite radiación electromagnética, no se la puede detectar con los telescopios de la misma forma que la materia luminosa. El hecho de considerar materia oscura fría se debe a cuestiones de consistencia que refieren a la masa de los constituyentes y a sus tasas de interacción con el resto de los componentes de la sopa cósmica, significando "fría" que las partículas de la materia oscura son no relativistas al momento de desacoplarse del FCR. Sin embargo, este tipo de materia aún no fue detectado mediante experimentos, lo cual implica que no se dispone de una descripción microscópica precisa de su constitución. De manera adicional a la mencionada evidencia proveniente de las curvas de rotación de las galaxias, existe una gran cantidad de evidencia empírica acerca de la existencia de la materia oscura. Por ejemplo, no es posible explicar las mediciones correspondientes al FCR si no se postula materia oscura como un ingrediente más de la sopa cósmica [3, 16, 17]. Tampoco pueden explicarse las observaciones a través de SNIa sin postular la existencia de este tipo de materia [9, 18, 19]. Otra forma de justificar la existencia de este tipo de materia es a través de las lentes gravitacionales ya que la distribución de la materia oscura puede ser inferida a partir de los efectos gravitacionales observados en las imágenes distorsionadas de galaxias de fondo [20, 21]. Como ejemplo final, el choque del cúmulo Bala, que corresponde a dos galaxias que colisionaron hace aproximadamente 150 millones de años. A partir de comparar observaciones del choque del cúmulo obtenidas con lentes gravitacionales con las respectivas obtenidas con rayos X, se puede observar que la materia bariónica se dispersa, pero los halos de materia oscura se atraviesan sin interactuar [22]. Esta es una prueba fuerte de la existencia de la materia oscura, debido a que se observa específicamente como, a diferencia de la materia bariónica que se dispersa por colisiones, hay una cantidad de materia que no se puede ver, pero se atraviesa sin otra interacción que la gravitatoria.

La inclusión de la constante cosmológica para explicar la expansión acelerada del Universo, y de la materia oscura fría para explicar, entre otras cosas, las curvas de rotación de galaxias da lugar al modelo estándar de la cosmología: el modelo Λ CDM, donde Λ refiere a la constante cosmológica y CDM a la materia oscura fría. Este modelo es estándar en la comunidad científica debido a la gran cantidad de evidencia experimental que lo respalda [3, 23, 24]. Como un ejemplo se encuentra la ya mencionada medición del FCR, donde el hecho de obtener un espectro de cuerpo negro con una temperatura bien definida respalda al modelo Λ CDM. Sin embargo, como el FCR consta de luz que se propaga desde el Universo temprano, las pequeñas anisotropías e inhomogeneidades que esta luz posee tienen mucha información del Universo en ese momento. Utilizando esas pequeñas fluctuaciones como condición inicial, y evolucionando los componentes del cosmos con las ecuaciones del modelo Λ CDM se puede simular la formación de estructuras

galácticas a gran escala. Dichas simulaciones están en un acuerdo muy fuerte con las estructuras que se miden a día de hoy, lo cual también es una validación fuerte del modelo estándar [25, 26]. Otro sustento del modelo que no puede ser ignorado refiere a la nucleosíntesis primordial (BBN), la cual consiste en la formación de núcleos atómicos durante la evolución del Universo. En el contexto de BBN hay predicciones como la abundancia relativa de ${}^4\text{He}$ primordial, y de otros elementos livianos, las cuales también están en fuerte acuerdo con las observaciones, lo cual otorga aún más respaldo al modelo ΛCDM [27]. El acuerdo en el valor de los de parámetros cosmológicos obtenidos con observaciones independientes confirma la consistencia del modelo [3]. Las mencionadas pruebas observacionales del modelo estándar, junto a otras que no han sido mencionadas, le otorgan a ΛCDM una posición privilegiada dentro de los modelos cosmológicos que describen el Universo, aunque dicho modelo presenta problemas (algunos de los cuales son discutidos en esta tesis) que introducen la posibilidad de que haya física aún no entendida en nuestra descripción del cosmos.

Como se mencionó en el párrafo anterior, existen problemas que, a pesar de la gran cantidad de evidencia que respalda a ΛCDM , ponen en tensión al modelo y motivan a buscar alternativas para describir al Universo. El ya mencionado problema de la constante cosmológica es un ejemplo, pero no es el único. La naturaleza desconocida de la materia oscura también constituye un problema en sí mismo, dado que a pesar de necesitarla en el modelo no hay aún un experimento que pueda detectarla directamente. Otro problema del modelo cosmológico estándar es la llamada tensión de Hubble [28]. El valor del actual parámetro de Hubble H_0 que se ha obtenido mediante el FCR asumiendo un modelo cosmológico estándar [29] no concuerda con el inferido de forma independiente del modelo, con datos de supernovas de tipo Ia y Cefeidas [29, 30]. Esta tensión es relevante porque para realizar la medición a través de SNIa no se asume una dependencia fuerte en ningún modelo cosmológico, motivo por el cual la discrepancia puede deberse a fallas en ΛCDM . Este problema es fuertemente discutido en la literatura pero no hay acuerdo acerca de la discrepancia en las estimaciones, aunque la mayoría de las discusiones se inclinan por errores sistematicos en la estimación de las supernovas [31–33]. Otro problema teórico que acecha al modelo estándar es el llamado "problema de la coincidencia", el cual consiste en que en la actualidad la densidad volumétrica de energía correspondiente a la constante cosmológica es igual a la densidad volumétrica de energía referida a la materia oscura, siendo que para que esto sea así se requieren condiciones iniciales muy específicas [34]. Entre los problemas mencionados vale la pena realizar una distinción entre problemas teóricos, tales como el de la constante cosmológica o la coincidencia cósmica; y las tensiones, dentro de las cuales se puede ubicar la tensión de H_0 . Esta distinción es importante debido a que las tensiones son problemas que aparecen al comparar datos con observaciones. En esta tesis se discuten tensiones obtenidas en la estimación de la densidad de energía en forma de materia del Universo (Ω_{m0}).

Posibles alternativas para resolver los problemas mencionados consisten en considerar modelos de energía oscura en los cuales se agrega un nuevo componente al tensor energía momento en vez de agregar la constante cosmológica. Este nuevo componente debería tener una presión negativa para explicar la expansión acelerada del cosmos. Dichos modelos se pueden estudiar considerando a la energía oscura como un fluido o como un campo [18]. Nuevamente, estos mo-

delos tienen como problema que aún no se han realizado mediciones directas referentes a dicha energía oscura, ni en forma de fluido ni tampoco como un campo cuántico. Un enfoque alternativo al de la energía oscura que permite explicar la expansión acelerada del Universo sin la necesidad de introducir contenidos exóticos en el modelo consiste en modificar las ecuaciones de Einstein, a través de modificaciones en la acción de Einstein-Hilbert [35, 36]. Estas modificaciones traen como consecuencia que al realizar las variaciones de la acción respecto de la métrica las ecuaciones dinámicas que se obtienen son distintas a las de Einstein. Las ecuaciones modificadas pueden introducir cambios en la dinámica de las componentes del tensor métrico, logrando así predicciones alternativas al modelo estándar enmarcado en la relatividad general. En este contexto, puede ser posible, por ejemplo, describir la expansión acelerada del Universo sin necesidad de introducir cantidades exóticas tales como la constante cosmológica o la energía oscura. De esta manera se explica dicha expansión acelerada en términos de que las ecuaciones de la relatividad general no modelan correctamente el Universo, y no en términos de cantidades desconocidas cuyas propiedades pueden inferirse a través de las observaciones. Un ejemplo de las mencionadas modificaciones a la acción de Einstein-Hilbert son las llamadas $f(T)$, en las cuales se introduce en la acción una función del escalar de torsión (T) [37–39]. Además se han propuesto otras teorías de gravedad alternativas que proponen soluciones a los problemas mencionados previamente, como por ejemplo la teoría de gravedad unimodular en la que se considera que la densidad total de volumen del espacio-tiempo es una constante [40] o la teoría de gravedad masiva en la que se considera un gravitón con masa no nula [41]. Otra teoría de gravedad modificada es la llamada teoría de Gauss-Bonnet, a veces notada $f(G)$, la cual se enmarca dentro de la geometría diferencial y también incluye modificaciones en la acción de Einstein-Hilbert, logrando modelos cosmológicos alternativos [42].

En este trabajo nos focalizamos en las llamadas teorías $f(R)$, en las cuales se reemplaza al escalar de Ricci (R) en la acción de Einstein-Hilbert por una función arbitraria ($f(R)$) de esta cantidad [43–46]. Estas teorías son capaces de explicar la expansión acelerada del Universo sin considerar más que materia y radiación dentro del mismo. Resulta relevante remarcar que, al fijar la función arbitraria $f(R)$ se obtiene un modelo, habiendo distintos tipos de modelos que se corresponden a distintas formas de la función $f(R)$. En particular, en este trabajo se presenta una discusión sobre el modelo de Hu-Sawicki, el cual consiste en una forma particular para la $f(R)$ y es ampliamente estudiado en la literatura [44]. En particular, no se trabajó con dicho modelo debido a dificultades que exceden los objetivos de esta tesis, pero en su lugar, se estudió un modelo fenomenológico de gravedad modificada [2], el cual sirve como paso previo al estudio de teorías de gravedad modificadas desde primeros principios, tales como las $f(R)$.

Los modelos descriptos hasta ahora permiten escribir ecuaciones diferenciales que describen la dinámica del Universo, la cual se estudia a través de una cantidad física llamada factor de escala ($a(t)$). En términos de esta cantidad, se interpreta que si a crece el Universo se expande y si a disminuye se contrae. No fue trabajo de esta tesis estudiar el comportamiento del fondo cosmológico, ya que esto se estudió en trabajos anteriores [47]. En esta tesis se puso el foco en las perturbaciones de materia, que rompen la hipótesis del principio cosmológico y dan lugar a la formación de estructuras en el Universo. Una motivación para estudiar perturbaciones yace en el

hecho de que no existe una gran diferencia entre el comportamiento del fondo cósmico en Λ CDM y en teorías $f(R)$, mientras que, estas dos teorías predicen una evolución significativamente diferente de las perturbaciones de materia.

En la teoría de perturbaciones de materia se estudia la densidad fraccional de materia ($\delta(a, \vec{x})$), que tiene información sobre la diferencia entre la densidad de materia en algún punto y algún tiempo respecto a la densidad promedio en todo el Universo, para dicho valor de tiempo. En este trabajo se asume que las perturbaciones no afectan a la dinámica del fondo cosmológico y por lo tanto evolucionan dentro de un escenario cósmico descrito por el modelo en cuestión. Lógicamente, la evolución de las perturbaciones depende de parámetros cosmológicos que afectan a la dinámica a gran escala, y, aunque la cantidad $\delta(a, \vec{x})$ no puede medirse directamente, sí se pueden medir observables que contienen información sobre ella y su derivada. En este trabajo se empleó como observable $f\sigma_8(a, \vec{x})$, el cual contiene información sobre el crecimiento de perturbaciones dentro de esferas de radio de 8 Mpc, y tiene una relación directa con $\delta(a, \vec{x})$, su derivada y con el parámetro cosmológico σ_8 . Mediciones recientes de esta cantidad permitieron emplear métodos estadísticos para inferir parámetros y de esta forma testear modelos cosmológicos. Los métodos estadísticos empleados en el trabajo corresponden al de máxima verosimilitud y al de integración de Monte Carlo con cadenas de Markov (MCMC). Sin embargo, para poder realizar la inferencia de parámetros es necesario resolver la ecuación dinámica que describe la evolución de $\delta(a, \vec{x})$ en el tiempo. Los métodos tradicionales para integrar estas ecuaciones consisten en métodos numéricos. Estos métodos requieren fijar todos los parámetros antes de realizar la integración, motivo por el cual al explorar el espacio de parámetros con el MCMC se debe realizar una integración numérica para cada valor de los mismos. Esto representa una dificultad computacional en términos de tiempos de resolución al considerar modelos con varios parámetros libres [48]. Parte de este trabajo consiste en atacar las mencionadas dificultades, a través del empleo de herramientas que surgen desde el campo de la inteligencia artificial.

El Machine Learning, o aprendizaje automático, representa un campo multidisciplinario que se enfoca en el desarrollo de algoritmos y modelos capaces de aprender patrones y realizar tareas específicas sin una programación explícita. En este sentido, el Deep Learning y las Redes Neuronales Artificiales emergen como subdisciplinas poderosas dentro del Machine Learning, permitiendo el aprendizaje de representaciones de datos de manera jerárquica y abstracta. Al emplear redes neuronales (NN) con múltiples capas, el Deep Learning puede capturar y aprender características complejas de los datos, lo que lo hace especialmente efectivo en problemas de visión por computadora, procesamiento de lenguaje natural, reconocimiento de voz y otras áreas. Estas técnicas han revolucionado numerosos campos, ofreciendo soluciones innovadoras y de vanguardia para una variedad de problemas en investigación y aplicaciones industriales. Por ejemplo, el empleo de redes neuronales en el campo de la astrofísica y la cosmología es vasto y variado, cobrando un gran impulso con el desarrollo de los métodos de Machine Learning, siendo algunos ejemplos la predicción de corrimientos al rojo [49], la clasificación de objetos celestes [50], el análisis de datos del telescopio Hubble [51] y simulaciones cosmológicas [52], entre otras.

En este trabajo se emplearon métodos que utilizan redes neuronales (NN) para resolver las ecuaciones diferenciales de los modelos cosmológicos. A diferencia de los métodos numéricos

tradicionales, las redes aprenden a resolver las ecuaciones diferenciales luego de un período de entrenamiento. Durante dicho entrenamiento, se emplean técnicas para implementar el reconocimiento de patrones de las redes neuronales sobre las ecuaciones diferenciales [53]. Una vez entrenadas, las redes conforman una función, a la cual se le da como entrada, o input, un valor de la variable independiente y estas entregan la solución de la ecuación diferencial en dicho punto.

La implementación de las redes trae como una de las principales ventajas el método bundle, el cual consiste en que las redes aprendan a resolver la ecuación no solo en función de la variable independiente, sino que además la resuelvan en función de parámetros que se ven involucrados en las ecuaciones [54]. De esta forma, si se quiere realizar una exploración sobre el espacio de parámetros no es necesario re-entrenar a las redes en cada uno de los valores, sino que al implementar el método bundle se puede evaluar a las redes neuronales en cualquiera de los valores de los parámetros que se hayan incluido durante el entrenamiento. Esto, en algunos casos, permite acelerar el proceso de inferencia de parámetros a través del MCMC [47]. Cabe mencionar también, que dicho entrenamiento no requiere emplear las soluciones numéricas, motivo por el cual representa un método alternativo a los tradicionales. De manera adicional se incluyen en este trabajo resultados correspondientes a un tipo nuevo de arquitectura de redes neuronales cuyo desarrollo sigue en marcha por parte del grupo de trabajo del co-director de esta tesis: la arquitectura customizada.

No obstante, el método de las redes neuronales no tiene, a día de hoy, una manera de estimar el error cometido en las soluciones. Para poder realizar estas estimaciones y garantizar soluciones confiables, sí se emplearon los métodos numéricos tradicionales para calcular errores porcentuales utilizando a estos últimos como referencia.

Respecto a la estructura de esta tesis, en el capítulo 2 se describen en detalle los modelos teóricos estudiados durante el trabajo, junto con discusiones referidas a las condiciones iniciales de las ecuaciones y el rango de validez de las aproximaciones realizadas. El método de las redes neuronales es detallado en el capítulo 3, mientras que en el capítulo 4 se discute la implementación específica de los métodos de las redes sobre los modelos teóricos. En el capítulo 5 se introducen los datos observacionales empleados correspondientes al observable $f\sigma_8$, mientras que los resultados se comentan en el capítulo 6, quedando las conclusiones en el capítulo 7.

Capítulo 2

Modelo teórico

En este capítulo se describen los modelos teóricos utilizados en esta tesis. Cabe destacar que se han estudiado dos modelos distintos: el modelo estándar de la cosmología (Λ CDM) y un modelo fenomenológico basado en teorías de gravedad modificada. En cada uno de los modelos ha sido necesario entender tanto el comportamiento del fondo cosmológico, como la evolución de las perturbaciones de materia, donde la expresión "fondo cosmológico" refiere a la dinámica del Universo en sí mismo. Como se detalla en esta primera sección, esta dinámica puede entenderse a través de la evolución de una cantidad física positiva llamada factor de escala, la cual describe como se modifican las distancias físicas.

2.1. Ecuaciones de fondo cosmológico en Λ CDM

Como se describió en la introducción el modelo cosmológico estándar incluye una constante cosmológica (Λ) [9] para explicar la expansión acelerada del Universo y materia oscura fría (CDM) [15] para explicar una gran cantidad de datos astrofísicos y cosmológicos. Dicho modelo se enmarca en el contexto de la teoría de la relatividad general, motivo por el cual es necesario un tensor métrico para realizar una descripción del espacio-tiempo. Para construir este objeto en el escenario cosmológico, es necesario considerar el principio cosmológico [55], el cual establece que el espacio es isótropo y homogéneo. Esta hipótesis está bien justificada a gran escala (del orden de 100 Mpc [56]), y teniendo esto en consideración se puede escribir el elemento de línea correspondiente a la métrica de Friedman-Robertson-Walker (FRW) en coordenadas esféricas [57]

$$ds^2 = -dt^2 + a(t)^2 \left[\frac{dr^2}{1 - Kr^2} + r^2(d\theta^2 + \text{sen}^2\theta d\phi^2) \right], \quad (2.1)$$

donde $a(t)$ es el factor de escala descrito anteriormente, mientras que K corresponde a la curvatura del espacio-tiempo, siendo $K > 0$ un universo esférico y cerrado, $K < 0$ un universo hiperbólico y abierto, y $K = 0$ un universo plano. En la ecuación 2.1 se puede ver que el factor de escala multiplica al diferencial de longitud, por lo cual se entiende que la evolución del factor de escala afecta a las distancias en el Universo. De esta forma, la dinámica de esta cantidad representa la evolución del espacio, ya sea su contracción o su expansión. Es importante mencionar que al estar trabajando en el contexto de la relatividad general, las

componentes del tensor métrico pueden cambiar de forma según el sistema de referencia. El tensor métrico descrito en la ecuación 2.1 corresponde al llamado sistema de referencia comóvil, el cual corresponde a observadores que están en reposo respecto de la grilla cósmica que conforma al espacio-tiempo. Esta mención es importante, dado que las coordenadas espaciales en el frame comóvil (\vec{r}) se relacionan con las coordenadas físicas (\vec{x}) como

$$\vec{x}(t) = a(t)\vec{r}, \quad (2.2)$$

donde se ve explícitamente que las distancias comóviles son constantes mientras que las físicas cambian a medida que evoluciona el Universo. Es importante enfatizar que el tensor métrico de la ecuación 2.1 es la propuesta más general posible que captura las posibles geometrías cosmológicas, contemplando la hipótesis del principio cosmológico [58], es decir, no existen otros tensores métricos que puedan capturar la isotropía y homogeneidad del espacio-tiempo.

Además, mediante el factor de escala se puede definir una cantidad importante en la cosmología: el corrimiento al rojo, también llamado redshift. Este se define como

$$z = \frac{1}{a} - 1, \quad (2.3)$$

donde se ve que si $a = 1$, entonces $z = 0$. Para describir la evolución del factor de escala son necesarias ecuaciones dinámicas. En el contexto de la relatividad general, estas ecuaciones se obtienen variando la acción de Einstein-Hilbert

$$S = \frac{1}{16\pi G} \int dx^4 \sqrt{-g}(R - 2\Lambda) + S_m, \quad (2.4)$$

donde R representa al escalar de Ricci, Λ es la constante cosmológica, G es la constante de gravitación de Newton y g es el determinante de la métrica. Además S_m es una acción que corresponde a la materia. Al realizar la variación de la ecuación 2.4 respecto de la métrica, se obtienen las ecuaciones de campo de Einstein

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi GT_{\mu\nu}, \quad (2.5)$$

las cuales son diez ecuaciones diferenciales de segundo orden para las componentes del tensor métrico. En Λ CDM la evolución del espacio-tiempo está descrita por la evolución del factor de escala, y por lo tanto estas ecuaciones describen la evolución de $a(t)$. Sin embargo, para poder escribir las ecuaciones es necesario dar una prescripción para el contenido energético del Universo, es decir, es necesario dar una expresión para el tensor de energía-momento ($T_{\mu\nu}$).

El modelo más sencillo consiste en asumir que el contenido energético del Universo se comporta como un fluido ideal. De esta forma, la expresión para el tensor de energía-momento es sencilla y resulta

$$T_{\mu\nu} = \text{diag}(\rho, p, p, p), \quad (2.6)$$

donde ρ representa la densidad volumétrica de energía y p la presión del fluido. Resulta importante notar que este tensor es isótropo en la parte espacial, lo cual es necesario para que

sea consistente con el principio cosmológico.

En el modelo Λ CDM se asume que el Universo consta de dos tipos de energía: materia no relativista y radiación. En ambos casos es posible escribir una ecuación de estado que relaciona la densidad volumétrica de energía con la presión del fluido

$$P = \omega\rho \quad (2.7)$$

donde ω vale $\frac{1}{3}$ para la radiación y 0 para la materia. Además, se definen ρ_Λ y $P_\Lambda = -\rho_\Lambda$, la densidad de energía y la presión asociadas a la constante cosmológica, respectivamente. Estas cantidades sirven para tratar a la constante cosmológica como un contenido energético más del Universo.

De esta manera, se pueden introducir las expresiones de la métrica (2.1) y la expresión del tensor de energía-momento (2.6) en las ecuaciones de campo de Einstein (2.5) para obtener las ecuaciones de Friedman. Asumiendo un universo con parte espacial plana ($K = 0$), se obtiene

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}(\rho_m + \rho_r + \rho_\Lambda) \quad (2.8a)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho_m + \rho_r + \rho_\Lambda + 3(P_m + P_r + P_\Lambda)). \quad (2.8b)$$

La ecuación 2.8a corresponde a $\mu = \nu = 0$, mientras que la ecuación 2.8b corresponde a $\mu = \nu = 1, 2, 3$. El hecho de asumir un Universo plano ($K = 0$) se justifica fuertemente a través de mediciones [3, 59].

Estas ecuaciones son suficientes para describir la evolución del Universo, pero puede ser útil considerar una ecuación adicional, la cual resulta linealmente dependiente de estas, derivada a través de la conservación del tensor de energía-momento (es decir, planteando $T_{;\nu}^{\mu\nu} = 0$) que resulta ser

$$\dot{\rho} + 3H(\rho + P) = 0. \quad (2.9)$$

Aquí se definió el parámetro de Hubble

$$H = \frac{\dot{a}}{a}. \quad (2.10)$$

La ecuación 2.9 es útil junto con la ecuación de estado para hallar como dependen las distintas densidades de energía con el factor de escala. En efecto, usando las ecuaciones de estado resulta

$$\rho_m(a) = \frac{\rho_{m0}}{a^3}, \quad \rho_r(a) = \frac{\rho_{r0}}{a^4}, \quad \rho_\Lambda(a) = \rho_{\Lambda 0}, \quad (2.11)$$

donde ρ_{m0} , ρ_{r0} y $\rho_{\Lambda 0}$ son las densidades volumétricas de energía de materia, radiación y constante cosmológica actuales, respectivamente, es decir $\rho_{j0} = \rho_j(a(t_{\text{hoy}} = 1))$. Cabe mencionar que se define que el valor actual del factor de escala se define como $(a(t_{\text{hoy}}) = 1)$. En este punto

resulta de utilidad definir las densidades de energía adimensionales. Estas se definen a través de calcular el cociente entre las densidades volumétricas y la densidad volumétrica crítica actual

$$\rho_{c0} = \frac{3H_0^2}{8\pi G}, \quad (2.12)$$

donde H_0 es el parámetro de Hubble evaluado hoy, es decir

$$H_0 = H(a = 1). \quad (2.13)$$

Luego, se definen las densidades adimensionales de energía como

$$\Omega_i(a) = \frac{\rho_i(a)}{\rho_c(a)}, \quad (2.14)$$

donde i etiqueta al contenido energético en cuestión. Resulta importante mencionar que, si la suma de todas las densidades de energía que se consideran es igual a la densidad crítica, esto quiere decir que el universo posee una geometría plana. Las observaciones están de acuerdo con que esto ocurre, y por eso es que en el modelo estándar de la cosmología se asume $K = 0$ [3, 59].

De esta definición es directo ver que se puede expresar a las densidades volumétricas de energía en términos de las densidades adimensionales evaluadas hoy. De esta forma resulta que

$$\rho_m(a) = \frac{\Omega_{m0}\rho_{c0}}{a^3}, \quad \rho_r(a) = \frac{\Omega_{r0}\rho_{c0}}{a^4}, \quad \rho_\Lambda(a) = \Omega_{\Lambda0}\rho_{c0}. \quad (2.15)$$

Usando la ecuación 2.15, junto con la definición de la densidad crítica (2.12), se puede reescribir la ecuación 2.8a como

$$H(a) = \frac{\dot{a}}{a} = H_0 \sqrt{\frac{\Omega_{m0}}{a^3} + \frac{\Omega_{r0}}{a^4} + \Omega_{\Lambda0}}. \quad (2.16)$$

Por otro lado, usando las ecuaciones de estado (2.7), también resulta útil reescribir la ecuación 2.8b como

$$\frac{\ddot{a}}{a} = -\frac{H_0^2}{2} \left[\frac{\Omega_{m0}}{a^3} + \frac{2\Omega_{r0}}{a^4} - 2\Omega_{\Lambda0} \right]. \quad (2.17)$$

Las ecuaciones 2.16 y 2.17 describen la evolución del fondo cósmico en el modelo Λ CDM y fueron de vital importancia para el trabajo. No obstante, como se mencionó en la introducción, el objetivo de esta tesis fue estudiar la dinámica de las perturbaciones de materia.

2.2. Ecuación de perturbaciones de materia en Λ CDM

Hasta aquí se ha detallado la dinámica del Universo a gran escala, y se han detallado ecuaciones que describen un universo isótropo y homogéneo, las cuales están en concordancia con el principio cosmológico. Sin embargo, al observar galaxias y/o cúmulos de galaxias, se ve que éstas forman estructuras que no respetan el principio ya mencionado [60]. Esto no representa un problema, ya que, como se ha mencionado previamente, las hipótesis de isotropía y homogeneidad

son válidas a escalas mayores que las galácticas.

Este hecho obliga a romper las hipótesis de isotropía y homogeneidad si se quiere estudiar la formación de estructuras. Con esta motivación, se ha desarrollado una teoría de perturbaciones que permite estudiar la evolución de la materia en el cosmos, y de esta forma poder comprender cómo se forman las ya mencionadas estructuras galácticas [61].

La derivación que sigue puede encontrarse, sin considerar la energía oscura, en [62]. Se considera que la materia en el Universo se comporta como un fluido. Luego, se consideran volúmenes físicos del Universo que son menores que la distancia de Hubble

$$|\vec{x}| \ll H^{-1}, \quad (2.18)$$

y, de esta forma, se asume que las velocidades son pequeñas comparadas con la velocidad de la luz. Es necesario aclarar que la aproximación 2.18 restringe el estudio a perturbaciones con longitudes características que estén dentro del radio de Hubble, lo cual se discute en profundidad más adelante. Gracias a estas consideraciones, se pueden utilizar las ecuaciones de fluidos no relativistas para estudiar la evolución de la materia. De esta forma, una de las ecuaciones a utilizar es la de Euler

$$\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \vec{\nabla}) \vec{v} = -\frac{-\vec{\nabla} p}{\rho_m} - \vec{\nabla} \Phi, \quad (2.19)$$

donde $v(\vec{x}, t)$ es la velocidad en la posición \vec{x} al tiempo t , $p(\vec{x}, t)$ es la presión, $\rho_m(\vec{x}, t)$ es la densidad de masa, $\Phi(\vec{x}, t)$ es el potencial gravitatorio y $\vec{\nabla} = \frac{\partial}{\partial \vec{x}}$. Es importante remarcar que se está asumiendo la hipótesis de que la materia evoluciona como un fluido de manera independiente al resto de las especies que conforman el Universo. El potencial gravitatorio se determina a través de la ecuación de Poisson

$$\nabla^2 \Phi = 4\pi G \rho_m, \quad (2.20)$$

y además también es importante incluir la ecuación de continuidad

$$\frac{\partial \rho_m}{\partial t} + \vec{\nabla} \cdot (\rho_m \vec{v}) = 0. \quad (2.21)$$

Se toma como solución de orden cero a esta ecuación el flujo de Hubble: $\rho_m(\vec{x}, t) = \rho_{mh}(t)$ y $\vec{v}_h = H\vec{x}$, donde el subíndice h representa la solución de orden cero. Es importante notar que estas cantidades son isótropas y homogéneas, por lo cual, para estudiar el problema de interés, se definen las cantidades perturbadas a primer orden como

$$\vec{v} = \vec{v}_h + \delta \vec{v} \quad \rho_m = \rho_{mh} + \delta \rho_m \quad p = p_h + \delta p \quad \Phi = \Phi_h + \delta \Phi, \quad (2.22)$$

donde puede mostrarse que $\Phi_h = (2\pi G/3)\rho_{mh}|\vec{x}|^2$, y se asume que las densidades son pequeñas comparadas con los valores a orden cero. Escribiendo estas definiciones en la ecuación 2.19, y usando que las cantidades a orden cero la satisfacen, se obtiene la ecuación de Euler a primer orden en las perturbaciones

$$\frac{d\delta\vec{v}}{dt} + \frac{\dot{a}}{a}\delta\vec{v} = -\frac{\vec{\nabla}\delta p}{\rho_{mh}} - \vec{\nabla}\delta\Phi, \quad (2.23)$$

donde se utilizó la siguiente relación

$$(\delta\vec{v} \cdot \vec{\nabla})\vec{v}_h = H\delta\vec{v}, \quad (2.24)$$

y $\frac{d}{dt}$ refiere a

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \vec{v}_h \cdot \vec{\nabla}, \quad (2.25)$$

la derivada total respecto del tiempo para un observador que no se mueve respecto de la solución de orden cero del fluido. Así mismo, la ecuación de continuidad (2.21) a primer orden resulta

$$\frac{d\delta\rho_m}{dt} + \rho_{mh}\vec{\nabla} \cdot \delta\vec{v} + 3H\delta\rho_m = 0, \quad (2.26)$$

mientras que la ecuación de Poisson a primer orden se puede expresar

$$\nabla^2\delta\Phi = 4\pi G\delta\rho_m. \quad (2.27)$$

A este punto resulta útil definir la perturbación fraccional de densidad de materia

$$\delta_m(\vec{x}, t) = \frac{\rho_m(\vec{x}, t)}{\rho_{mh}(t)}, \quad (2.28)$$

definición que permite reescribir las ecuaciones de continuidad (2.21) y de Poisson (2.20) de la siguiente manera

$$\dot{\delta}_m = -\vec{\nabla} \cdot \delta\vec{v} \quad (2.29a)$$

$$\nabla^2\Phi = 4\pi G\rho_{mh}\delta_m, \quad (2.29b)$$

donde el punto representa $\frac{d}{dt}$. Por otro lado, si se define además la velocidad del sonido en el fluido cosmológico como

$$\delta P = C_s^2\delta\rho, \quad (2.30)$$

con C_s la velocidad del sonido en el fluido cosmológico, junto con las ecuaciones 2.29 en la ecuación de Euler a primer orden (2.23), se puede obtener una única ecuación diferencial para $\delta_m(t, \vec{x})$

$$\ddot{\delta}_m(t, \vec{x}) + 2\frac{\dot{a}}{a}\dot{\delta}_m(t, \vec{x}) = (4\pi G\rho_{mh} + C_s^2\nabla^2)\delta_m(t, \vec{x}). \quad (2.31)$$

Debe notarse que el laplaciano de la ecuación deriva respecto de las coordenadas físicas, no las comóviles. Resulta que $\vec{x}(t) = a(t)\vec{r}$, donde \vec{r} es una coordenada comóvil. Es importante

tener estas cuestiones en consideración a la hora de obtener la ecuación 2.31.

A continuación consideramos que la velocidad del sonido es nula, dado que la materia tiene presión nula. De esta manera es posible reescribir la ec. 2.31 en el espacio de Fourier como:

$$\ddot{\hat{\delta}}_m(t, \vec{k}) + 2\frac{\dot{a}}{a}\dot{\hat{\delta}}_m(t, \vec{k}) - 4\pi G\rho_{mh}\hat{\delta}_m(t, \vec{k}) = 0, \quad (2.32)$$

donde k corresponde al valor del número de onda medido en el sistema de referencia comóvil, lo cual no se aclara por cuestiones de notación debido a que todos los k que aparecen involucrados son comóviles. Puede resultar confuso el hecho de que, en el espacio de coordenadas la dependencia de δ_m es en coordenadas físicas y aquí el número de onda corresponde a las coordenadas comóviles. El hecho es que, para escribirlo de esta forma se realiza la transformación $\vec{k} = a(t)\vec{k}_{\text{fis}}(t)$, la cual no es evidente en la ecuación 2.32 debido a que este término desaparece al considerar $C_S = 0$. Por otro lado, al estar en el espacio de Fourier se ha definido

$$\hat{\delta}_m(t, \vec{k}) = \iiint dx^3 \delta_m(a, \vec{x}) e^{-i\vec{k}\cdot\vec{r}}. \quad (2.33)$$

Resulta importante notar que la aproximación de fluido empleada (ec. 2.18) es válida, únicamente, para modos cuya longitud de onda es menor que el radio de Hubble, donde la longitud de onda del modo en cuestión corresponde a

$$\lambda_{\text{com}} = \frac{2\pi}{k}, \quad (2.34)$$

donde λ_{com} hace énfasis en que la longitud de onda está medida en el sistema de referencia comóvil. Luego, $\lambda_{\text{fis}}(t) = a(t)\lambda_{\text{com}}$ y la condición análoga a la hipótesis impuesta al principio de la derivación (ec. 2.18) corresponde a

$$\lambda_{\text{fis}} \ll H^{-1}. \quad (2.35)$$

Los modos de Fourier que cumplen con esta condición se llaman "modos sub-horizon", porque son los modos que se encuentran dentro del radio de Hubble. También es importante remarcar que, en esta aproximación, la ecuación diferencial no depende del modo de Fourier (\vec{k}) a considerar, dado que en este caso, la única derivada espacial desaparece al considerar un fluido sin presión. De esta manera resulta importante remarcar que, a pesar de no tener dependencia explícita en los modos de Fourier, las aproximaciones realizadas restringen los modos que pueden describirse con la ecuación 2.32.

Por otro lado, resulta conveniente reemplazar las derivadas temporales en la ecuación 2.32 por derivadas respecto del factor de escala. Para ello se emplea la transformación

$$\frac{d}{dt} = \dot{a} \frac{d}{da}, \quad (2.36)$$

que a su vez implica

$$\frac{d^2}{dt^2} = \ddot{a} \frac{d}{da} + \dot{a}^2 \frac{d^2}{da^2}. \quad (2.37)$$

Con esto en consideración, la ecuación 2.32 resulta

$$\dot{a}^2 \hat{\delta}_m''(a, \vec{k}) + \left(\ddot{a} + 2 \frac{\dot{a}^2}{a} \right) \hat{\delta}_m'(a, \vec{k}) - 4\pi G \rho_{mh} \hat{\delta}_m(a, \vec{k}) = 0, \quad (2.38)$$

donde la prima denota la derivada respecto del factor de escala. Sin embargo, esta ecuación involucra tanto al parámetro de Hubble $H = \frac{\dot{a}}{a}$, como a la derivada segunda del factor de escala respecto del tiempo. Para eliminar estas dependencias, se pueden usar las ecuaciones del fondo cosmológico. Resulta importante mencionar que se está asumiendo que las perturbaciones de materia no afectan a la evolución del fondo, ya que en caso de no ser cierto, se deberían derivar nuevas ecuaciones para el fondo utilizando la métrica de FRW perturbada. En este trabajo se asume que la evolución dinámica del Universo no se ve afectada por las perturbaciones de materia, y, por lo tanto, las ecuaciones de Friedmann son válidas. Además, resulta importante aclarar que se está asumiendo que la evolución de las perturbaciones de materia es independiente del resto de los componentes de la sopa cósmica, lo cual se evidencia en el hecho de tener la ecuación para $\hat{\delta}_m(a, \vec{k})$ desacoplada de las ecuaciones de otros contenidos. Esto vale únicamente para modos sub-horizon. Luego, utilizando las ecuaciones del fondo (2.16,2.17) se puede mostrar que

$$\dot{a}^2 = \frac{H_0^2 \Omega_{r0}}{a^2} \left[1 + \frac{a}{a_{\text{eq}}} + \alpha \left(\frac{a}{a_{\text{eq}}} \right)^4 \right] \quad (2.39a)$$

$$\left(\ddot{a} + 2 \frac{\dot{a}^2}{a} \right) = \frac{H_0^2 \Omega_{r0}}{2a^3} \left[2 + 3 \frac{a}{a_{\text{eq}}} + 6\alpha \left(\frac{a}{a_{\text{eq}}} \right)^4 \right], \quad (2.39b)$$

donde se han definido

$$a_{\text{eq}} = \frac{\Omega_{r0}}{\Omega_{m0}} \quad \alpha = a_{\text{eq}}^4 \frac{\Omega_{\Lambda 0}}{\Omega_{r0}}. \quad (2.40)$$

Si se usa que $\rho_{mh} = \Omega_{m0} \rho_c a^{-3}$ (ec. 2.11) se obtiene la ecuación que describe la dinámica de las perturbaciones de materia

$$\hat{\delta}_m''(a, \vec{k}) + a_{\text{eq}} \left[\frac{2 + 3 \frac{a}{a_{\text{eq}}} + 6\alpha \left(\frac{a}{a_{\text{eq}}} \right)^4}{2a \left(1 + \frac{a}{a_{\text{eq}}} + \alpha \left(\frac{a}{a_{\text{eq}}} \right)^4 \right)} \right] \hat{\delta}_m'(a, \vec{k}) - \left[\frac{3a_{\text{eq}}}{2a \left(1 + \frac{a}{a_{\text{eq}}} + \alpha \left(\frac{a}{a_{\text{eq}}} \right)^4 \right)} \right] \hat{\delta}_m(a, \vec{k}) = 0. \quad (2.41)$$

Esta es una ecuación de segundo orden que gobierna la evolución de la densidad fraccional de perturbaciones de materia, a medida que evoluciona el factor de escala. Otra cosa importante que debe tenerse en cuenta, es que, dado que se asumió que $C_s = 0$, la ecuación no depende del modo de Fourier (\vec{k}). Por este motivo puede realizarse la descomposición

$$\hat{\delta}_m(a, \vec{k}) = \delta_m(a) \tilde{F}(\vec{k}), \quad (2.42)$$

y reemplazando esta en la ecuación 2.41 se obtiene una ecuación para la parte que no

depende de los modos de Fourier

$$\delta_m''(a) + a_{\text{eq}} \left[\frac{2 + 3 \frac{a}{a_{\text{eq}}} + 6\alpha \left(\frac{a}{a_{\text{eq}}} \right)^4}{2a \left(1 + \frac{a}{a_{\text{eq}}} + \alpha \left(\frac{a}{a_{\text{eq}}} \right)^4 \right)} \right] \delta_m'(a) - \left[\frac{3a_{\text{eq}}}{2a \left(1 + \frac{a}{a_{\text{eq}}} + \alpha \left(\frac{a}{a_{\text{eq}}} \right)^4 \right)} \right] \delta_m(a) = 0. \quad (2.43)$$

La descomposición 2.42 se discute en más detalle en la siguiente sección, donde además se discuten las condiciones iniciales, las cuales surgen de resolver las ecuaciones para las perturbaciones de todas las especies que conforman el contenido energético del Universo en el momento en consideración. Se puede ver que al realizar esto para el universo dominado por la materia, δ_m es proporcional al factor de escala. Las condiciones iniciales empleadas en el trabajo son [2]

$$\delta_m(a_i) = a_i \quad (2.44a)$$

$$\delta_m'(a_i) = 1, \quad (2.44b)$$

con $a_i = 10^{-3}$. Es importante remarcar que este valor del factor de escala corresponde a tiempos donde el Universo está dominado por la materia, y, como δ_m es proporcional al factor de escala, las condiciones iniciales se justifican. No obstante, en la siguiente sección se incluye una discusión más detallada de este asunto.

2.3. Las condiciones iniciales en Λ CDM

Al finalizar la sección anterior se describieron las condiciones iniciales empleadas en el trabajo, las cuales corresponden a una situación en la cual la cantidad $\delta_m(a) = a$ (ec. 2.61). Obtener estas condiciones iniciales no es trivial, pero es sustancialmente más sencillo en Λ CDM que en modelos más complicados. En esta sección se discute como obtener estas condiciones y que aproximaciones son necesarias para ello.

2.3.1. Etapas de evolución del Universo

En primer lugar, es necesario detallar la evolución del Universo según distintos regímenes en los cuales dominan los distintos tipos de contenido energético considerados. Juntando las ecuaciones 2.14 y 2.15, se puede observar que

$$\Omega_m(a) = \frac{\Omega_{m0}}{a^3} \quad \Omega_r(a) = \frac{\Omega_{r0}}{a^4} \quad \Omega_\Lambda(a) = \Omega_{\Lambda0}. \quad (2.45)$$

Además, en un Universo plano ($K = 0$), vale también que

$$\Omega_m(a) + \Omega_r(a) + \Omega_\Lambda(a) = 1. \quad (2.46)$$

De esta forma, se puede ver que, en función de los valores de las densidades de energía

hoy, a medida que evoluciona el Universo puede modificarse cual es el contenido energético que dispone de una densidad de energía mayor. Por ejemplo, cuando $a \rightarrow 0$, se puede ver que

$$\Omega_r(a) > \Omega_m(a) > \Omega_\Lambda(a), \quad (2.47)$$

mientras que, dependiendo de los valores de las densidades hoy, puede ocurrir que cuando a crece, alguna de las densidades de energía supere a alguna de las otras. En efecto, si se asume que $\Omega_r(a_{\text{eq}}) = \Omega_m(a_{\text{eq}})$, entonces el valor del factor de escala que cumple con esta condición es

$$a_{\text{eq}} = \frac{\Omega_{r0}}{\Omega_{m0}}, \quad (2.48)$$

donde asumiendo valores cosmológicos usuales ($\Omega_{r0} = 3,58 \times 10^{-5}$ [63] y $\Omega_{m0} = 0,315$ [3], por ejemplo) se obtiene $a_{\text{eq}} = 1,13 \times 10^{-4}$.

Esto, junto con el orden evidenciado en 2.47, significa que para tiempos cósmicos en los cuales $a < a_{\text{eq}}$ vale que $\Omega_r(a) > \Omega_m(a)$, mientras que si $a > a_{\text{eq}}$ ocurre que $\Omega_r(a) < \Omega_m(a)$. Esto se interpreta como que, en el caso $a < a_{\text{eq}}$ el Universo es dominado por la radiación, mientras que si $a > a_{\text{eq}}$ domina la materia.

Lo discutido en los párrafos anteriores permite, en una primera aproximación, estudiar la evolución temporal del cosmos partiendo dicha evolución en tramos dominados por distintos tipos de energía, y así simplificar las ecuaciones para poder obtener soluciones analíticas.

2.3.2. Dinámica cosmológica de un Universo dominado por materia

Al considerar un Universo dominado por materia la ecuación de Friedmann 2.16 resulta

$$\frac{\dot{a}}{a} = H_0 \sqrt{\Omega_{m0}} a^{-\frac{3}{2}}, \quad (2.49)$$

donde se han despreciado las contribuciones de la radiación y la constante cosmológica. Al resolver esta ecuación utilizando como condición inicial la usual en el modelo de Big Bang, la cual dice que $a(t=0) = 0$, se obtiene

$$a^{\text{mat}}(t) = C_m t^{\frac{2}{3}}, \quad (2.50)$$

donde $C_m = \sqrt{H_0 \sqrt{\Omega_{m0}}}$ es constante.

Un comentario relevante, es que este régimen vale si $a > a_{\text{eq}}$, pero además debe tenerse en cuenta que en el Universo tardío comienza a dominar la constante cosmológica. El valor del factor de escala a partir del cual domina Λ se puede calcular con el mismo razonamiento empleado para a_{eq} (ec. 2.48) y corresponde a

$$a_{\text{eq}}^\Lambda = \left(\frac{\Omega_{m0}}{\Omega_{\Lambda0}} \right)^{\frac{1}{3}}. \quad (2.51)$$

De esta forma, el régimen en el cual domina la materia corresponde a

$$a_{\text{eq}} < a < a_{\text{eq}}^\Lambda, \quad (2.52)$$

donde asumiendo valores cosmológicos estándar (por ejemplo $\Omega_{m0} = 0,315$ [3] y $\Omega_{\Lambda 0} = 0,686$ [3]) se obtiene $a_{\text{eq}}^{\Lambda} = 0,77$.

2.3.3. Evolución de las perturbaciones de materia en un Universo dominado por materia

Previamente en este capítulo se mostró que la ecuación diferencial que gobierna la dinámica de δ_m puede ser escrita, tanto en función de derivadas temporales, como de derivadas respecto del factor de escala. Como al estudiar la evolución por regímenes se obtienen expresiones explícitas para $a(t)$, se considera la ecuación para δ_m que contiene derivadas respecto del tiempo (2.32). En rigor y considerando la dependencia en los modos de Fourier, la ecuación se escribe como

$$\ddot{\delta}_m(t, \vec{k}) + 2\frac{\dot{a}}{a}\dot{\delta}_m(t, \vec{k}) - 4\pi G\rho_{mh}\hat{\delta}_m(t, \vec{k}) = 0. \quad (2.53)$$

Es importante remarcar que aquí se ha utilizado, al igual que en la sección anterior, que la materia evoluciona como un fluido de forma independiente a los otros componentes del Universo.

Además, es importante notar que la ecuación 2.53 es invariante de escala. Esto quiere decir que en la ecuación no aparece explícitamente \vec{k} . Este hecho permite concluir que los modos de Fourier estudiados en esta aproximación evolucionan de manera independiente, lo cual implica que la evolución no mezcla modos y, de esta manera se puede condensar la dependencia en \vec{k} en una función independiente del tiempo. Con esto en mente puede realizarse la factorización

$$\hat{\delta}_m(t, \vec{k}) = \delta_m(t)\tilde{F}(\vec{k}). \quad (2.54)$$

Así es que se puede estudiar el comportamiento de $\delta_m(t)$ en función de los distintos regímenes comentados anteriormente. En particular, se estudiará únicamente el régimen dominado por materia, debido a que en este trabajo se impusieron condiciones iniciales en $a_i = 10^{-3}$. Esto se justifica notando que como $\Omega_{r0} \ll \Omega_{m0}$, entonces $a_{\text{eq}} \ll a_i$.

Al imponer la descomposición 2.54 en la ecuación 2.53, y utilizar las expresiones para $a(t)$ correspondientes al período dominado por materia obtenidas anteriormente (ec. 2.50), la ecuación a resolver corresponde a

$$\ddot{\delta}_m(t) + \frac{4}{3t}\dot{\delta}_m(t) - \frac{2}{3t^2}\delta_m(t) = 0, \quad (2.55)$$

donde la dependencia en \vec{k} ha desaparecido al cancelarse $\tilde{F}(\vec{k})$. Esta ecuación tiene como solución general

$$\delta_m(t) = c_1 t^{-1} + c_2 t^{\frac{2}{3}}. \quad (2.56)$$

Como t^{-1} decae en el tiempo, se dice que este término corresponde al modo decreciente. Entonces, la solución que se considera, teniendo en cuenta que en el régimen dominado por materia vale que $a(t) = t^{\frac{2}{3}}$, es

$$\delta_m(a) = c_2 a. \quad (2.57)$$

Es de vital importancia entender que la constante c_2 sí tiene dependencia en los modos de Fourier. Esto ocurre, dado que la dinámica de los modos varía al entrar o salir del radio de Hubble. De esta forma, cada modo puede tener una constante distinta, motivo por el cual a partir de este momento se la nota como $c_2(\vec{k})$. Si se deseara obtener el valor de esta constante, sería necesario estudiar la evolución de los modos desde el período dominado por radiación, utilizando como condición inicial aquella que se obtiene al final del período de inflación. Sin embargo, debido a que en los modelos estudiados en este trabajo la ecuación es independiente del modo k no fue necesario estudiar en detalle esta constante, ya que se puede utilizar la descomposición 2.54 para absorber $c_2(\vec{k})$ dentro de $\tilde{F}(\vec{k})$, y de esta manera definir

$$F(\vec{k}) = \frac{1}{c_2(\vec{k})} \tilde{F}(\vec{k}), \quad (2.58)$$

de forma tal que

$$\hat{\delta}(a, \vec{k}) = \delta_m(a) F(\vec{k}), \quad (2.59)$$

donde con la redefinición, en el período dominado por materia, se puede considerar como una buena aproximación

$$\delta_m(a) = a. \quad (2.60)$$

En este punto es que se realiza otra aproximación: se asume que en el tiempo inicial para el cual se integra la ecuación diferencial el Universo está dominado por materia, de forma tal que la radiación ya es despreciable, y que la constante cosmológica (Λ) aún no es relevante. De este modo, al comenzar a integrar la ecuación y utilizar que esta última es invariante de escala, se puede emplear la definición 2.59, y bajo la aproximación mencionada, ocurre que en el valor inicial del factor de escala ($a_i = 10^{-3}$), se puede imponer que

$$\delta_m(a_i) = a_i \quad (2.61a)$$

$$\delta'_m(a_i) = 1, \quad (2.61b)$$

siendo estas las condiciones iniciales presentadas al final de la sección anterior.

Resulta relevante enfatizar que en la obtención de estas condiciones iniciales ha sido crucial el hecho de que la ecuación diferencial sea invariante de escala (es decir, no dependa explícitamente de \vec{k}). Si esto no fuera así, obtener las condiciones iniciales representaría una tarea más compleja dado que no sería posible realizar la factorización (ec. 2.54), siendo además la ecuación análoga a 2.55 explícitamente dependiente de \vec{k} . En las teorías $f(R)$ mencionadas en la introducción esta dificultad se suma a la propia dificultad de las ecuaciones de fondo cosmológico y perturbaciones, y como resultado de estas complejidades, no hay, a día de hoy, una expresión

analítica para las condiciones iniciales de las perturbaciones de materia en dichas teorías. Esto se discutirá al final de este capítulo.

Por otro lado, el observable empleado en este trabajo para inferir parámetros cosmológicos tiene la propiedad de que, en caso de que sea posible realizar la factorización 2.54, la propia naturaleza de dicho observable elimina la dependencia en los modos de Fourier. Por lo tanto, en los modelos en los cuales la ecuación resulta invariante de escala, no es necesario estudiar la dinámica de $F(\vec{k})$, y más aún, no es necesario obtener explícitamente el valor de $c_2(\vec{k})$, dificultad que se evitó durante este trabajo. Esto se discute en la siguiente sección.

2.4. El observable $f\sigma_8$

No es posible medir directamente la función $\delta_m(a, \vec{x})$. En cambio, en este trabajo, se empleó como observable cosmológico una cantidad que se relaciona directamente, tanto con $\delta_m(a, \vec{x})$ como con su derivada. Dicho observable se define como [2]

$$f\sigma_8(a, \vec{x}) = f(a, \vec{x})\sigma(a, \vec{x}), \quad (2.62)$$

donde en este caso

$$f(a, \vec{x}) = \frac{d\ln(\delta_m(a, \vec{x}))}{d\ln(a)} \quad (2.63)$$

es la tasa de crecimiento, y

$$\sigma(a, \vec{x}) = \sigma_8 \frac{\delta_m(a, \vec{x})}{\delta_m(1, \vec{x})} \quad (2.64)$$

representa las fluctuaciones rms dependientes del corrimiento al rojo del campo de densidad lineal dentro de esferas de radio $R = 8h^{-1}$ Mpc, mientras que σ_8 representa su valor hoy. Utilizando la regla de la cadena, puede escribirse el observable como

$$f\sigma_8(a, \vec{x}) = \frac{\sigma_8}{\delta_m(1, \vec{x})} a\delta'_m(a, \vec{x}), \quad (2.65)$$

donde $\delta_m(a, \vec{x})$ representa la anti-transformada de Fourier de $\hat{\delta}_m(a, \vec{k})$, y lo propio ocurre con la derivada.

Si se incluye la dependencia de $\hat{\delta}_m(a, \vec{k})$ en los modos de Fourier de manera factorizada (como en 2.59), puede observarse que el observable elimina por completo la dependencia en \vec{k} . Para ver esto, debe notarse que, para el cálculo del observable, es necesario primero anti-transformar Fourier la solución de la ecuación diferencial. Es decir, las funciones que entran en el cálculo del observable son

$$\delta_m(a, \vec{x}) = \frac{1}{2\pi} \iiint d^3k \hat{\delta}_m(a, \vec{k}) e^{-i\vec{k}\cdot\vec{r}} \quad \delta'_m(a, \vec{x}) = \frac{1}{2\pi} \iiint d^3k \hat{\delta}'_m(a, \vec{k}) e^{-i\vec{k}\cdot\vec{r}}. \quad (2.66)$$

En este punto resulta de vital importancia la factorización propuesta en 2.59, debido a que,

en este caso la derivada cumple

$$\hat{\delta}'_m(a, \vec{k}) = F(\vec{k})\delta'_m(a), \quad (2.67)$$

ya que el modo de Fourier \vec{k} no depende del factor de escala. Con esto en mente, se pueden reescribir las ecuaciones de 2.66 como

$$\delta_m(a, \vec{x}) = \frac{\delta_m(a)}{2\pi} \iiint d^3k F(\vec{k}) e^{-i\vec{k}\cdot\vec{x}} \quad \delta'_m(a, \vec{x}) = \frac{\delta_m(a)}{2\pi} \iiint d^3k F(\vec{k}) e^{-i\vec{k}\cdot\vec{x}}, \quad (2.68)$$

donde debe notarse que la parte que depende del factor de escala puede salir de la integral en \vec{k} . Definiendo una función de \vec{r} como

$$G(\vec{r}) = \frac{1}{2\pi} \iiint d^3k F(\vec{k}) e^{-i\vec{k}\cdot\vec{r}}, \quad (2.69)$$

es directo ver que el observable definido en 2.70, elimina la información de dicha función, ya que, en general

$$f\sigma_8(a, \vec{r}) = \frac{\sigma_8}{\hat{\delta}_m(1, \vec{r})} a \hat{\delta}'_m(a, \vec{r}) = \frac{\sigma_8}{\delta_m(1)} a \delta'_m(a) \frac{G(\vec{r})}{G(\vec{r})} = \frac{\sigma_8}{\delta_m(1)} a \delta'_m(a) = f\sigma_8(a). \quad (2.70)$$

Es importante remarcar que para que esto sea así debe cumplirse que la ecuación diferencial no dependa explícitamente de los modos de Fourier, lo cual vale en el caso de Λ CDM. Por otro lado, el régimen estudiado por la ecuación 2.43, como se discutió previamente, corresponde a los modos sub-horizon. Esta condición restringe el valor de las longitudes de onda que pueden estudiarse en este contexto, ya que solo aquellos modos que hayan ingresado al horizonte en valores de $a < a_i$ cumplen esta restricción. El análisis vale para todos los modos con longitud de onda menor a λ_{com_i} , y por lo tanto vale para modos tales que

$$k_i = \frac{2\pi}{\lambda_{\text{com}_i}} < k. \quad (2.71)$$

En la siguiente sección se discute el resultado, pero el valor que se obtiene corresponde a

$$k_i \approx 10^{-3} \text{ h Mpc}^{-1}. \quad (2.72)$$

Por otro lado, se sabe que valores de \vec{k} cercanos a $k_{NL} = 0.2 \text{ h Mpc}^{-1}$ hoy tienen un comportamiento no lineal [8]. De esta forma, los modos de Fourier que son sub-horizon durante todo el período de integración y, por lo tanto, pueden enmarcarse en el contexto de estudio de este trabajo son aquellos que cumplen que

$$k_i < k < k_{NL}, \quad (2.73)$$

y para estos modos vale la ecuación de las perturbaciones de la materia 2.43. Además, como esta ecuación no depende explícitamente de \vec{k} , también es posible factorizar la dependencia en

\vec{k} de $\delta_m(t, \vec{k})$ y, así, las condiciones iniciales corresponden a las mencionadas previamente. En la siguiente sección se describe en detalle el cálculo de los modos de Fourier que son descriptos por la ecuación diferencial 2.43.

2.5. Los modos de Fourier dentro de la aproximación sub-horizon

En la sección anterior se mencionó que, a pesar de que la ecuación 2.43 no depende explícitamente de k , las aproximaciones realizadas restringen los modos de Fourier que pueden estudiarse mediante dicha ecuación. Esto se impone al considerar a la materia como un fluido no relativista mediante la aproximación 2.18, donde se asume que los volúmenes físicos considerados son menores que el radio de Hubble

$$|\vec{x}| \ll H^{-1}, \quad (2.74)$$

donde es importante mencionar que H^{-1} corresponde al radio de Hubble físico. Imponer esta restricción permite estudiar la evolución de la materia como un fluido mediante la ecuación de Euler no relativista, pero lógicamente restringe los modos de Fourier que pueden estudiarse. Esto es así debido a que los modos descriptos en este régimen corresponden a aquellos tales que

$$\lambda_{\text{fis}} \ll H^{-1}, \quad (2.75)$$

donde λ_{fis} refiere a la longitud de onda física. Esta longitud de onda se relaciona con la longitud de onda comóvil como

$$\lambda_{\text{fis}} = a\lambda_{\text{com}}. \quad (2.76)$$

Es importante notar que el valor de λ_{com} es constante en el sistema de referencia comóvil. Por otro lado, a través de esta longitud de onda puede definirse el número de onda comóvil

$$k = \frac{2\pi}{\lambda_{\text{com}}}, \quad (2.77)$$

donde no se utiliza el sub-índice $_{\text{com}}$ en k debido a que, como se mencionó previamente, todos los k involucrados en las ecuaciones corresponden al número de onda en el sistema de referencia comóvil. Por completitud, puede definirse también un número de onda físico como

$$k_{\text{fis}} = \frac{k}{a}. \quad (2.78)$$

Para comprender qué modos están bien descriptos por las aproximaciones realizadas es necesario comprender qué modos están dentro del horizonte desde el momento en que se comienza a integrar la ecuación diferencial. Para realizar este cálculo, es útil definir al radio de Hubble comóvil

$$r_H = \frac{1}{aH}. \quad (2.79)$$

Además, al igual que para las condiciones iniciales, se estudia la evolución del Universo a través de los distintos regímenes. De esta forma, el radio comóvil de Hubble cuando el Universo está dominado por radiación es

$$H_{\text{rad}} \propto a^{-2} \rightarrow \log(r_{H_{\text{rad}}}) \propto \log(a), \quad (2.80)$$

en el momento dominado por materia es

$$H_{\text{mat}} \propto a^{-\frac{3}{2}} \rightarrow \log(r_{H_{\text{mat}}}) \propto \frac{1}{2} \log(a), \quad (2.81)$$

y finalmente cuando domina la constante cosmológica

$$H_{\Lambda} \propto a^{-1} \rightarrow \log(r_{H_{\Lambda}}) \propto -\log(a). \quad (2.82)$$

Además, es importante mencionar que en el modelo Λ CDM se introduce un mecanismo previo al dominio de radiación llamado "inflación" [64]. Durante dicho período se asume que

$$H_{\text{inf}} = \text{cte}, \quad (2.83)$$

y de ese modo resulta que, al igual que para el caso de la constante cosmológica

$$H_{\text{inf}} \propto a^{-1} \rightarrow \log(r_{H_{\text{inf}}}) \propto -\log(a). \quad (2.84)$$

En la figura 2.1 se puede ver un gráfico esquemático del comportamiento del radio comóvil de Hubble durante la evolución del Universo, en escala logarítmica.

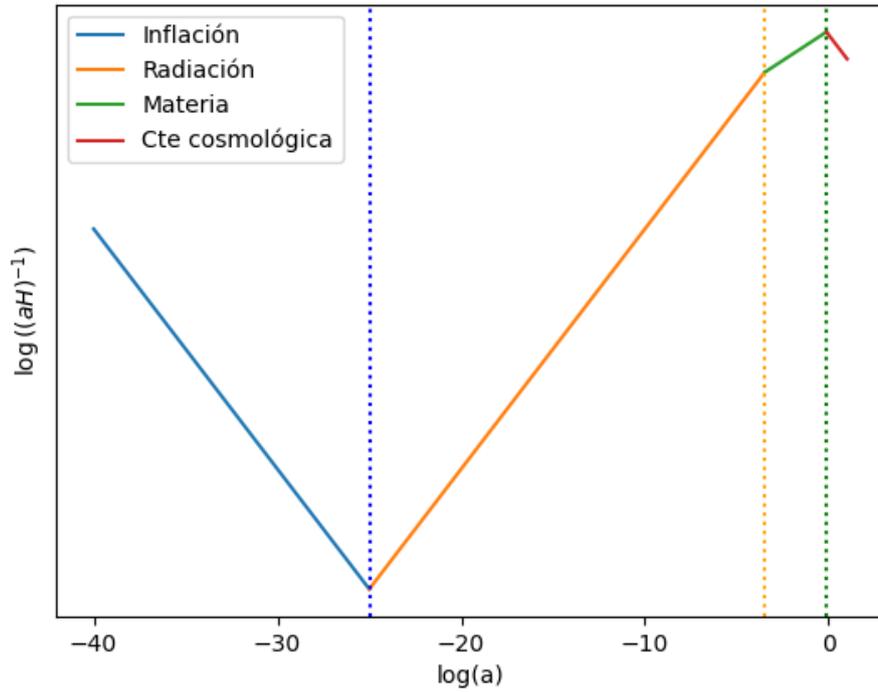


Figura 2.1: Evolución del radio de Hubble comóvil en función de a . Las líneas verticales representan, en azul el final de inflación ($a_e = 10^{-25}$), en naranja a_{eq} y en verde a_{eq}^Λ . El valor de $(aH)^{-1}$ representa una escala importante durante la evolución del Universo.

Este gráfico permite comprender como se modifica el radio comóvil de Hubble a través de la evolución del cosmos. En dicho gráfico, el valor de λ_{com} de los modos de Fourier son rectas con valor constante, ya que, como se mencionó, la longitud de onda comóvil es constante en el sistema de referencia correspondiente. Por ejemplo, en la figura 2.2 se puede ver un modo con algún valor de λ_{com} tal que sale del horizonte durante inflación y vuelve a entrar durante el período dominado por la radiación.

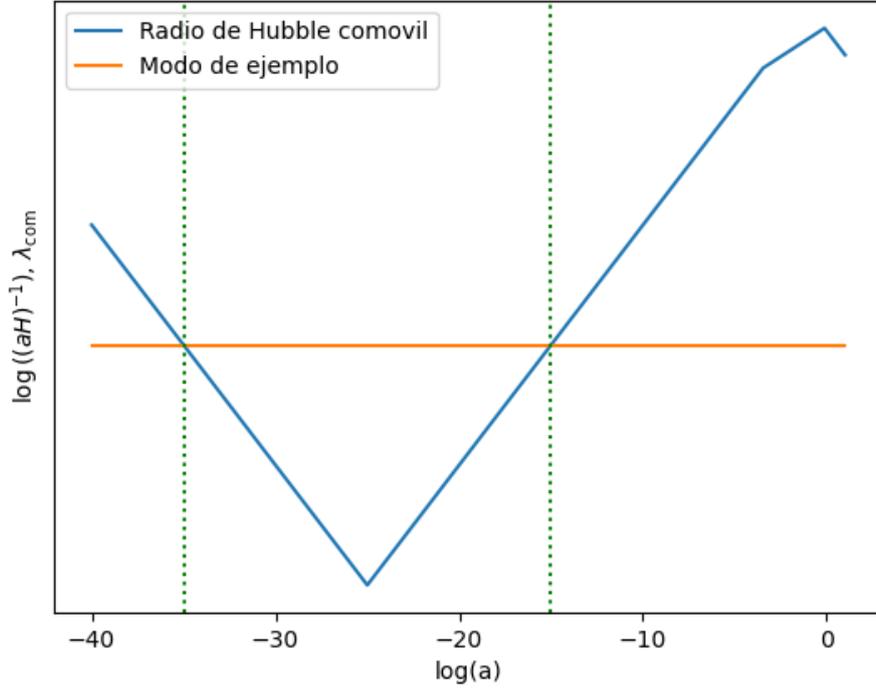


Figura 2.2: Evolución del radio de Hubble comóvil en función de a , junto con un modo de Fourier de ejemplo. En esta escala los modos se representan a través de cada λ_{com} como líneas horizontales. Las líneas verticales representan el valor de a en el cual el modo sale del horizonte (izquierda) y el valor en que entra (derecha).

En este punto es necesario calcular el valor de k_i que corresponde al modo que ingresa al horizonte cuando $a = a_i$, que es el momento en el cual se imponen las condiciones iniciales. Para ello simplemente debe calcularse $\lambda_{\text{com},i}$ tal que

$$\lambda_{\text{com},i} = \frac{1}{a_i H(a_i)}. \quad (2.85)$$

Emplando la ecuación 2.77 es directo notar que

$$k_i = \frac{a_i H(a_i)}{2\pi} \approx 10^{-3} \text{ h Mpc}^{-1}, \quad (2.86)$$

donde para calcular $H(a_i)$ se utilizó la ecuación 2.16 con $\Omega_{m0} = 0,315$ [3], $\Omega_{\Lambda 0} = 0,628$ [3] y $\Omega_{r0} = 5,37 \times 10^{-5}$ [63]. Este valor corresponde al modo de Fourier que ingresa al horizonte justo en el momento en que se comienza la integración de la ecuación. Luego, todos los modos con longitud de onda comóvil menor, i.e., k mayor ya están dentro del horizonte cuando comienza la integración y, por ende, son correctamente descriptos bajo las aproximaciones realizadas.

Por otro lado, al considerar longitudes de onda cada vez más pequeñas la aproximación lineal empleada en la ecuación de Euler deja de valer. Para calcular el valor de k en el cual las no linealidades cobran importancia se plantea el momento en que comienza a valer que

$$\delta_m \approx 1, \quad (2.87)$$

y se obtiene que este valor corresponde a [8]

$$k_{NL} \approx 0,2 \text{ h Mpc}^{-1}. \quad (2.88)$$

De esta forma, los modos de Fourier que se describen mediante las aproximaciones empleadas durante este trabajo corresponden a

$$k_i = 10^{-3} \text{ h Mpc}^{-1} < k < 0,2 \text{ h Mpc}^{-1} = k_{NL}, \quad (2.89)$$

los cuales se pueden ver esquemáticamente en la figura 2.3.

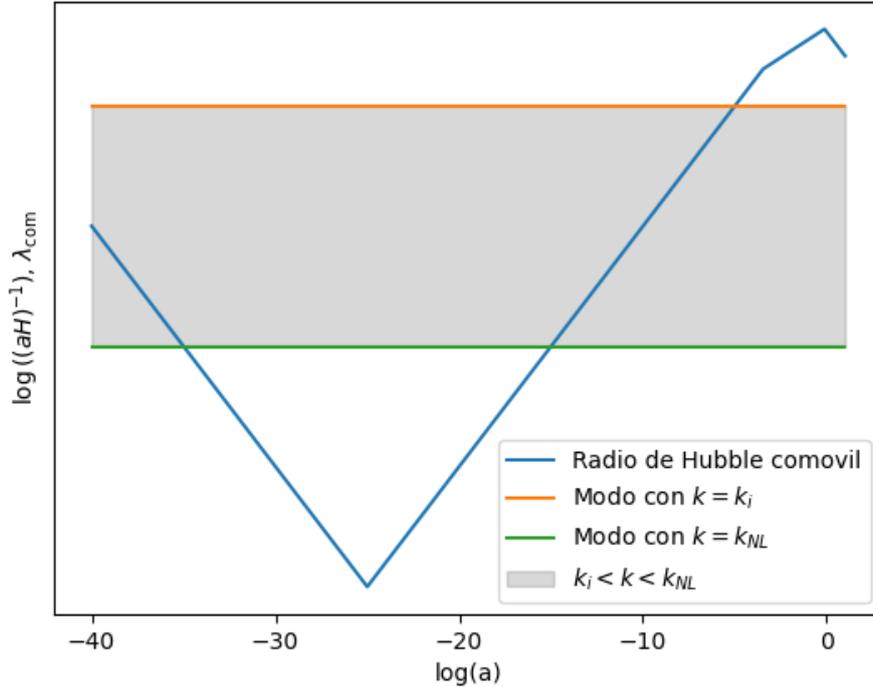


Figura 2.3: Representación pictórica de la región de valores de k que se pueden estudiar con la ecuación 2.43. El modo en naranja corresponde a aquel que entra al horizonte cuando se imponen las condiciones iniciales, mientras que el que está en verde corresponde al primer modo no lineal. En gris todos los valores que se pueden estudiar mediante el modelo teórico empleado.

Toda la discusión realizada hasta aquí corresponde al modelo estándar de la cosmología Λ CDM, pero el objetivo de este trabajo fue estudiar modelos alternativos que pudieran explicar las observaciones cosmológicas sin recurrir a la constante cosmológica. En lo que sigue del capítulo se discuten las teorías $f(R)$, que dan lugar a modelos de gravedad modificada.

2.6. Ecuaciones de fondo en $f(R)$

Una posible explicación de la expansión acelerada del Universo, diferente al modelo de la constante cosmológica, consiste en asumir una teoría alternativa a la relatividad general para describir la interacción gravitacional [36]. En esta tesis nos focalizamos en un conjunto de modelos

cosmológicos alternativos basados en las teorías $f(R)$, donde se reemplaza el escalar de Ricci en la acción de Einstein-Hilbert (2.4) por una función arbitraria de esta cantidad [65]. De esta forma, la acción de la teoría resulta

$$S = \frac{1}{16\pi G} \int dx^4 \sqrt{-g} f(R) + S_m. \quad (2.90)$$

Del mismo modo que en Relatividad General, variando la acción respecto de la métrica se obtienen las ecuaciones de campo de la teoría

$$R_{\mu\nu} F(R) - \frac{1}{2} g_{\mu\nu} f(R) + (g_{\mu\nu} \square - \nabla_\mu \nabla_\nu) F(R) = 8\pi G T_{\mu\nu}, \quad (2.91)$$

donde \square es el d'alembertiano, ∇_μ es la derivada covariante y se emplea la notación

$$F(R) = \frac{df}{dR}, \quad F_{,R}(R) = \frac{dF}{dR} = \frac{d^2 f}{dR^2}, \quad F_{,RR}(R) = \frac{d^3 f}{dR^3}. \quad (2.92)$$

Con las mismas hipótesis cosmológicas discutidas al principio de este capítulo, y procediendo de una manera análoga al caso de Relatividad General, se pueden obtener ecuaciones dinámicas para el factor de escala de la métrica de FRW dentro de la teoría $f(R)$

$$H^2 = \frac{1}{3F} \left[8\pi G \sum_i \rho_i + \frac{RF - f}{2} - 3H \dot{R} F_{,R} \right] \quad (2.93a)$$

$$2\dot{H} + 3H^2 = -\frac{1}{F} \left[-\frac{RF - f}{2} + F_{,RR} \dot{R}^2 + (\ddot{R} + 2H\dot{R}) F_{,R} \right], \quad (2.93b)$$

donde por notación se han obviado las dependencias explícitas en R de f y sus derivadas.

En primera instancia, puede verse que es necesario fijar la función arbitraria $f(R)$ para obtener un modelo particular y poder resolver estas ecuaciones. Además, aparecen involucradas derivadas temporales del escalar de Ricci, lo cual indica que deben introducirse ecuaciones adicionales para poder resolver el sistema [66]. No fue trabajo de esta tesis resolver el fondo cosmológico para una teoría $f(R)$, sino que la intención fue emplear las ecuaciones 2.93 para estudiar la evolución de las perturbaciones de materia, asumiendo ya resulta la dinámica del factor de escala (i.e, la dinámica de $H(a)$). En una tesis anterior [47], se empleó el método de las redes neuronales utilizado en esta tesis y descrito en el capítulo 3 para resolver las ecuaciones correspondientes al fondo cosmológico en teorías $f(R)$. El avance en esta tesis es emplear el mismo método para resolver las ecuaciones de perturbaciones de materia en dichas teorías.

2.7. Ecuación de perturbaciones en $f(R)$

Considerar perturbaciones de materia en una teoría de gravedad modificada del tipo $f(R)$ no es tan sencillo como en el modelo de fluido no relativista de Λ CDM. No obstante, puede escribirse una ecuación similar a 2.32, donde toda la información sobre la teoría queda condensada dentro de una función efectiva $G_{\text{eff}}(\vec{k}, a)$ que cumple el rol de la constante de gravitación de Newton [67].

En la literatura referenciada los autores consideraron un campo escalar adicional acoplado

al fondo cosmológico [67]. Como en este trabajo no se incluyó ningún campo escalar en la dinámica cosmológica, no se tendrá en consideración dicho campo para discutir las ecuaciones de perturbaciones de materia.

Se considera una métrica perturbada con perturbaciones escalares Φ , Ψ en un gauge longitudinal

$$ds^2 = -(1 + 2\Phi)dt^2 + a(t)^2(1 - 2\Psi)\delta_{ij}dx^i dx^j \quad (2.94)$$

Además, se considera un tensor de energía momento perturbado

$$T_0^0 = -(\rho_m + \delta\rho_m) \quad (2.95a)$$

$$T_\alpha^0 = -\rho_m v_{m,\alpha}, \quad (2.95b)$$

donde v_m es un potencial de velocidad. Las ecuaciones perturbadas, obtenidas a partir de las ecuaciones de Einstein y en el espacio de Fourier son [68, 69]

$$3H(\dot{\Psi} + H\Phi) + \frac{k^2}{a^2}\Psi + \frac{1}{2F} \left[-3H\delta\dot{F} + \left(3H^2 + 3\dot{H} - \frac{k^2}{a^2} \right) \delta F + 3\dot{F}(\dot{\Psi} + H\Phi) + 3H\dot{F}\Phi + \delta\rho_m \right] = 0, \quad (2.96)$$

$$\Psi = \Phi + \frac{\delta F}{F}, \quad (2.97a)$$

$$\delta\dot{\rho}_m + 3H\delta\rho_m = \rho_m \left(3\dot{\Psi} - \frac{k^2}{a} v_m \right), \quad (2.97b)$$

$$\dot{v}_m + H v_m = \frac{1}{a} \Phi, \quad (2.97c)$$

donde \vec{k} es el número de onda. En este contexto se define la densidad fraccional de materia invariante de gauge en el espacio de Fourier como

$$\hat{\delta}_m(t, \vec{k}) = \frac{\delta\rho_m}{\rho_m} + 3Hv, \quad (2.98)$$

donde $v = av_m$, y además es interesante notar que el nuevo término $3Hv$ aparece por no estar trabajando en el gauge comovil, sino en un gauge longitudinal. Entonces las ecuaciones 2.97b, 2.97c se reescriben como

$$\dot{\hat{\delta}}_m(t, \vec{k}) = -\frac{k^2}{a^2}v + 3(\Psi + Hv), \quad (2.99a)$$

$$\dot{v} = \Phi, \quad (2.99b)$$

De las cuales se obtiene

$$\ddot{\delta}_m(t, \vec{k}) + 2H\dot{\delta}_m(t, \vec{k}) + \frac{k^2}{a^2}\Phi = 3\ddot{B} + 6H\dot{B}, \quad (2.100)$$

donde se definió $B = \Psi + Hv$. En este punto es útil, al igual que en Λ CDM, considerar la aproximación sub-horizon. En dicha aproximación se asume que todos los modos de Fourier están dentro del radio de Hubble, y de esa forma vale que $k^2 \gg a^2 H^2$.

En la ecuación 2.100 los términos del lado derecho aportan contribuciones de orden $H^2\Psi$ [67], por lo que son despreciables frente a los términos de orden $\frac{k^2}{a^2}\Phi$ en la aproximación sub-horizon. De esta forma resulta

$$\ddot{\delta}_m(t, \vec{k}) + 2H\dot{\delta}_m(t, \vec{k}) + \frac{k^2}{a^2}\Phi = 0 \quad (2.101)$$

Lo que falta es encontrar una relación entre Φ y δ_m . Para ello, se considera la ecuación 2.96 bajo la aproximación sub-horizon, la cual resulta

$$\frac{k^2}{a^2}\Psi = \frac{1}{2F} \left(\frac{k^2}{a^2}\delta F - \delta\rho_m \right). \quad (2.102)$$

Usando la ecuación 2.97a en esta última resulta

$$\frac{k^2}{a^2}\Phi = -\frac{k^2}{2a^2} \frac{\delta F}{F} - \frac{1}{2F}\delta\rho_m. \quad (2.103)$$

Por otro lado, como $F = F(R)$ resulta que

$$\delta F = F_{,R}\delta R, \quad (2.104)$$

mientras que δR puede calcularse como la perturbación a primer orden del escalar de Ricci explícitamente a partir de la métrica y vale [67]

$$\delta R = 2 \left[-3(\ddot{\Psi} + 4H\dot{\Psi} + H\dot{\Phi} + \dot{H}\Phi + 4H^2\Phi) + \left(\frac{k^2}{a^2} - 3\dot{H} \right) \Phi - 2\frac{k^2}{a^2}\Psi \right]. \quad (2.105)$$

Nuevamente, al usar la aproximación sub-horizon en esta expresión, junto con la expresión 2.97a se obtiene que

$$\delta R = -2\frac{k^2}{a^2} \left(\Phi + 2\frac{\delta F}{F} \right). \quad (2.106)$$

Introduciendo en esta última expresión las ecuaciones 2.103 y 2.104

$$\delta R = -\frac{k^2}{a^2} \frac{1}{1 + 4\frac{k^2}{a^2} \frac{F_{,R}}{F}} \Phi. \quad (2.107)$$

Operando a partir de la ecuación 2.103 usando las expresiones obtenidas, se llega a la relación deseada

$$\frac{k^2}{a^2}\Phi = -\frac{\rho_m}{2F} \frac{1 + 4\frac{k^2}{a^2} \frac{F_{,R}}{F}}{1 + 3\frac{k^2}{a^2} \frac{F_{,R}}{F}}, \quad (2.108)$$

donde se utilizó que en la aproximación sub-horizon $\hat{\delta}_m \approx \frac{\delta\rho_m}{\rho_m}$, y a partir de esta se puede

escribir la ecuación 2.101 como

$$\ddot{\delta}_m(t, \vec{k}) + 2\frac{\dot{a}}{a}\dot{\delta}_m - 4\pi\frac{G_{\text{eff}}(t, \vec{k})}{G}\rho_{mh}\hat{\delta}_m(t, \vec{k}) = 0, \quad (2.109)$$

donde se ha definido la expresión

$$\frac{G_{\text{eff}}}{G} = \frac{1}{8\pi F} \frac{1 + 4\frac{k^2}{a^2}m}{1 + 3\frac{k^2}{a^2}m}, \quad (2.110)$$

con

$$m = \frac{F_{,R}}{F}. \quad (2.111)$$

El modelo Λ CDM se recupera al considerar

$$8\pi f(R) = R - 2\Lambda, \quad (2.112)$$

donde vale que $F = f'(R) = \frac{1}{8\pi}$, así resulta $m = 0$ y $G_{\text{eff}} = G$.

Con la misma transformación para las derivadas temporales empleada en el modelo anterior (2.36,2.37), se puede escribir la ecuación 2.109 como

$$\dot{a}^2\hat{\delta}_m''(a, \vec{k}) + \left(\ddot{a} + 2\frac{\dot{a}^2}{a}\right)\hat{\delta}_m'(a, \vec{k}) - 4\pi\rho_{m0}\frac{G_{\text{eff}}(t, \vec{k})}{G}\hat{\delta}_m(a, \vec{k}) = 0. \quad (2.113)$$

Para continuar el procedimiento realizado en Λ CDM, deberían emplearse las ecuaciones de Friedmann modificadas (2.93) para eliminar \ddot{a} y \dot{a} . Sin embargo, resulta importante notar cierto detalle sobre las condiciones iniciales.

Previamente en este capítulo se discutió la obtención de las condiciones iniciales en Λ CDM. En dicha discusión se utilizó que la ecuación diferencial 2.43 es invariante de escala para poder factorizar la dependencia en \vec{k} y en a de δ_m . Sin embargo, esta ecuación depende explícitamente de \vec{k} , y esto no permite realizar la factorización, debido a que, durante la evolución, los distintos modos se acoplan. Esto, sumado a la complejidad de las ecuaciones en $f(R)$, representa una dificultad teórica en la obtención de las condiciones iniciales que escapó a las dificultades de esta tesis, debido a que no hay a día de hoy una expresión analítica para dichas condiciones iniciales en este contexto de gravedad modificada.

No obstante, para realizar esta discusión es necesario introducir el modelo de Hu-Sawicki [44], que se enmarca dentro de las teorías $f(R)$, y estudiar la dependencia en \vec{k} de la ecuación 2.113 con las expresiones de este modelo en particular.

2.8. Modelo de Hu-Sawicki

Como se mencionó previamente, al fijar el valor de la $f(R)$ se describe un modelo particular dentro de las teorías $f(R)$. El modelo de Hu-Sawicki [70] consiste en reemplazar el escalar de Ricci en la acción de Einstein-Hilbert por la función [71]

$$8\pi f(R) = R - \frac{2\Lambda}{1 + \left(\frac{b\Lambda}{R}\right)^n}, \quad (2.114)$$

donde b es un parámetro del modelo que mide la desviación de Λ CDM y n es una constante que se suele elegir $n = 1$. Además, Λ es tal que si se elige $b = 0$ se recupera el modelo Λ CDM. Esto implica una relación entre los parámetros cosmológicos [47]

$$\Lambda = 3H_0^2(1 - \Omega_{m0}). \quad (2.115)$$

Cabe aclarar que la forma de la función descripta en la ecuación 2.114 no corresponde al modelo original, sino que es una parametrización que permite realizar desarrollos perturbativos para $b \ll 1$ y así obtener soluciones analíticas para realizar comparaciones.

Esta función, junto con sus derivadas se usa para calcular $\frac{G_{\text{eff}}}{G}(a, \vec{k})$ mediante la ecuación 2.110. En la literatura [2, 72] se asume que dicha función no tiene dependencia en los modos de Fourier y se procede mediante la misma factorización empleada en Λ CDM, tanto para la ecuación diferencial como para las condiciones iniciales. En esta tesis se decidió revisar dicho argumento.

Para estudiar dicha dependencia, la ecuación 2.110 se puede reescribir como

$$\frac{G_{\text{eff}}}{G} = \frac{1}{F} \left[1 + \frac{\frac{k^2}{a^2} m}{1 + 3\frac{k^2}{a^2} m} \right], \quad (2.116)$$

donde simplemente se ha utilizado que $1 + 3 = 4$ en el numerador. Luego se puede simplificar un poco más la expresión para obtener

$$\frac{G_{\text{eff}}}{G} = \frac{1}{F} \left[1 + \frac{1}{3 + \frac{a^2}{mk^2}} \right]. \quad (2.117)$$

En la figura 2.4 se puede ver un gráfico de $\frac{G_{\text{eff}}}{G}$ en función del factor de escala para varios valores de k , todos dentro del rango $[10^{-3}, 0.2]$ h Mpc $^{-1}$. Para realizar dicha figura se utilizaron como valores de los parámetros $b = 0,1$, $H_0 = 70 \frac{\text{km}}{\text{s}} \text{Mpc}^{-1}$ y $\Omega_{m0} = 0,272$.

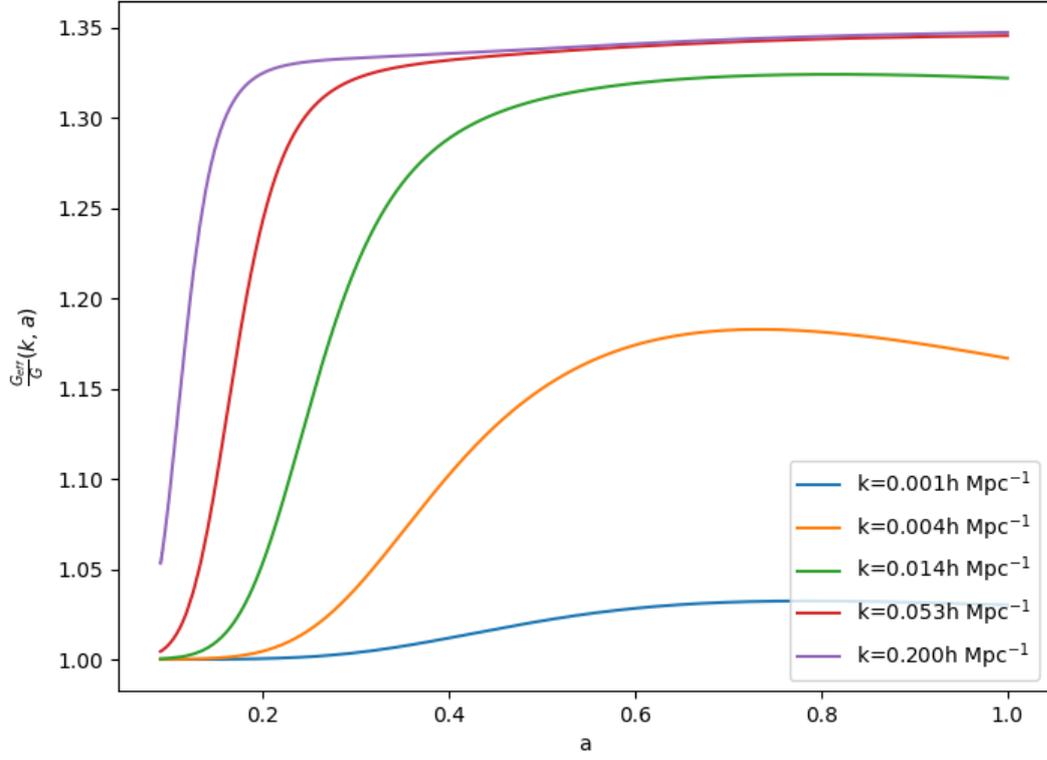


Figura 2.4: G_{eff}/G en función del factor de escala para varios valores de k dentro del régimen de estudio. Se observa que para valores pequeños de a $G_{\text{eff}} \rightarrow G$, lo cual se espera en las teorías $f(R)$, mientras que en el Universo tardío hay diferencias apreciables.

En la figura se observa que efectivamente $\frac{G_{\text{eff}}}{G}$ tiene una dependencia no trivial en k dentro del rango estudiado. Para confirmar esto en la figura 2.5 se muestran las diferencias porcentuales empleando como referencia el valor $k_{\text{ref}} = 10^{-3} \text{ h Mpc}^{-1}$, donde dichas diferencias se calculan como

$$\text{diff } \%_{k_j} = 100 \frac{|G_{\text{eff}}(k_j) - G_{\text{eff}}(k_{\text{ref}})|}{|G_{\text{eff}}(k_{\text{ref}})|} \quad (2.118)$$

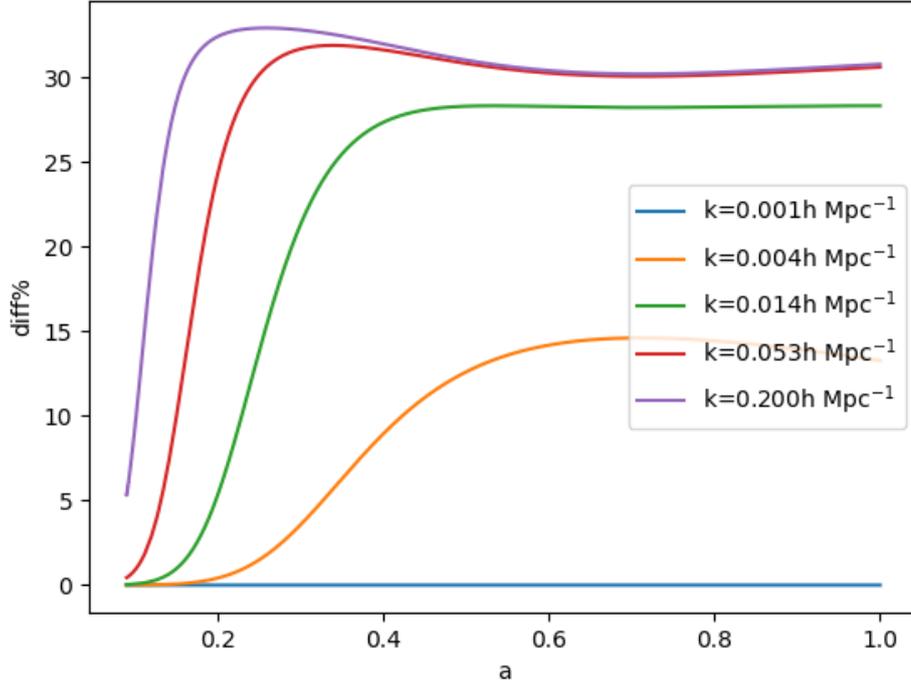


Figura 2.5: Diferencias porcentuales entre el valor de G_{eff}/G para varios valores de k dentro del rango de estudio, los cuales no se incluyen por ser irrelevantes. Se observan diferencias de hasta el 30 %, lo cual permite concluir que en esta teoría la dependencia en los modos de Fourier de la ecuación diferencial es no trivial.

En la figura se puede ver que el valor de $\frac{G_{\text{eff}}}{G}$ varía hasta un 30 % comparando a los modos que se estudian durante el trabajo. Esta dependencia en k no permite realizar la factorización empleada en Λ CDM dado a que la ecuación diferencial mezcla los modos durante la evolución. Por este motivo, para imponer condiciones iniciales en este contexto es necesario evolucionar las ecuaciones de cada especie desde el espectro primordial obtenido desde inflación hasta el momento en que se desea comenzar la integración. Esto implica que dichas condiciones iniciales tendrán una dependencia no trivial en los modos de Fourier, lo cual representa una dificultad que escapa a los límites de esta tesis.

Por este motivo, se decidió realizar un paso intermedio, el cual corresponde a estudiar un modelo fenomenológico que permite capturar la esencia de las teorías de gravedad modificadas, pero sin las dificultades discutidas hasta ahora. Esto se puede ver como un paso previo para calibrar el método de las redes antes de aplicarlo a un modelo de gravedad modificada proveniente de primeros principios.

2.9. Modelo fenomenológico de perturbaciones de materia con gravedad modificada

Debido a las dificultades discutidas previamente, nos proponemos estudiar la evolución de las perturbaciones de la materia de un modelo fenomenológico que en algún límite coincide con

un modelo basado en teorías $f(R)$.

En primer lugar, se asume que la dinámica del fondo cosmológico está descrita por la teoría Λ CDM. Luego, se introduce una expresión para la constante de gravitación de Newton sugerida en [2]

$$\frac{G_{\text{eff}}}{G} = 1 + g_a (1 - a)^n - g_a (1 - a)^{2n}, \quad (2.119)$$

donde g_a y n son parámetros libres del modelo. En este trabajo se asumió $n = 2$ [2]. Esta parametrización cumple con algunas condiciones necesarias para cualquier modelo de gravedad modificada:

- $G_{\text{eff}} > 0$ para que la energía gravitatoria sea definida positiva.
- $G_{\text{eff}}/G = 1,09 \pm 0,20$ para estar en acuerdo con los límites impuestos por la nucleosíntesis primordial (BBN).
- $G_{\text{eff}}(a = 1)/G = 1$, el valor medido para G_{eff} hoy tiene que ser igual a G .

Para estudiar la evolución de las perturbaciones de materia se toma la ecuación 2.113 teniendo en cuenta el valor de G_{eff} de este modelo, pero en lugar de usar las ecuaciones de Friedmann correspondientes al modelo particular de $f(R)$, se utilizan las ecuaciones de Friedmann de Λ CDM (2.8). De esta forma, el procedimiento es análogo a lo que ya se ha discutido previamente, y la ecuación obtenida es la misma, con la diferencia de que ahora las constantes G no se cancelan: la que viene de la ecuación de Euler corresponde a G_{eff} , mientras que la que viene de las ecuaciones de Friedmann corresponde a G .

Así, se obtiene la ecuación que describe la dinámica de δ_m en este modelo fenomenológico

$$\hat{\delta}_m''(a, \vec{k}) + a_{\text{eq}} \left[\frac{2 + 3\frac{a}{a_{\text{eq}}} + 6\alpha \left(\frac{a}{a_{\text{eq}}}\right)^4}{2a \left(1 + \frac{a}{a_{\text{eq}}} + \alpha \left(\frac{a}{a_{\text{eq}}}\right)^4\right)} \right] \hat{\delta}_m'(a, \vec{k}) - \left[\frac{3a_{\text{eq}}}{2a \left(1 + \frac{a}{a_{\text{eq}}} + \alpha \left(\frac{a}{a_{\text{eq}}}\right)^4\right)} \right] \frac{G_{\text{eff}}}{G}(a) \hat{\delta}_m(a, \vec{k}) = 0, \quad (2.120)$$

la cual, al igual que en Λ CDM es invariante de escala. De esta forma, empleando la factorización 2.59 se puede escribir la ecuación diferencial para la parte de $\hat{\delta}_m''(a, \vec{k})$ que solo depende de a

$$\delta_m''(a) + a_{\text{eq}} \left[\frac{2 + 3\frac{a}{a_{\text{eq}}} + 6\alpha \left(\frac{a}{a_{\text{eq}}}\right)^4}{2a \left(1 + \frac{a}{a_{\text{eq}}} + \alpha \left(\frac{a}{a_{\text{eq}}}\right)^4\right)} \right] \delta_m'(a) - \left[\frac{3a_{\text{eq}}}{2a \left(1 + \frac{a}{a_{\text{eq}}} + \alpha \left(\frac{a}{a_{\text{eq}}}\right)^4\right)} \right] \frac{G_{\text{eff}}}{G}(a) \delta_m(a) = 0. \quad (2.121)$$

Además, asumiendo que la dinámica del Universo hasta el momento de comenzar a integrar las ecuaciones es igual a Λ CDM, las condiciones iniciales son

$$\delta_m(a_i) = a_i \tag{2.122a}$$

$$\delta'_m(a_i) = 1, \tag{2.122b}$$

con $a_i = 10^{-3}$, donde vale la misma discusión realizada para el modelo estándar de la cosmología. Resulta importante remarcar que en este modelo, el fondo cosmológico corresponde a Λ CDM, mientras que la modificación de la gravedad se codifica en la parametrización empleada de $G_{\text{eff}}(a)$. Este es el motivo por el cual el modelo es fenomenológico, y además resulta un modelo de juguete intermedio entre Λ CDM y una teoría $f(R)$.

Capítulo 3

Método de las redes neuronales artificiales

El aprendizaje automático ("machine learning") es una rama de la inteligencia artificial que se enfoca en desarrollar algoritmos y modelos capaces de aprender patrones a partir de datos, permitiendo a las máquinas realizar tareas sin ser programadas de manera explícita [73]. Dentro del machine learning, el aprendizaje profundo ("deep learning") emerge como una disciplina especializada que utiliza arquitecturas de redes neuronales profundas para aprender representaciones jerárquicas complejas de los datos, permitiendo la extracción de características sofisticadas [74]. Las redes neuronales, inspiradas en el funcionamiento del cerebro humano, son la piedra angular de estas disciplinas, y consisten en estructuras interconectadas de nodos (neuronas) que procesan y transforman la información [75]. Este paradigma ha revolucionado la capacidad de las máquinas para abordar tareas complejas, como reconocimiento de imágenes, procesamiento del lenguaje natural y toma de decisiones, propulsando así el avance continuo de la inteligencia artificial y sus aplicaciones prácticas en diversas áreas.

En este capítulo se detallarán aspectos acerca del método de las redes neuronales (NN's), el cual fue utilizado para resolver las ecuaciones diferenciales detalladas en el capítulo 2. El tópico de las redes neuronales es extenso de por sí, motivo por el cual en esta tesis se presentan los aspectos que fueron relevantes durante el trabajo. Para trabajar con las NN's se emplearon las librerías de python neurodiffeq [76] y pytorch [77].

3.1. Neuronas artificiales

Como su nombre lo indica, las redes neuronales constan de neuronas individuales que se conectan formando una red. Las neuronas toman valores, realizan cálculos y devuelven un nuevo valor. Un ejemplo de neurona artificial puede verse en la figura 3.1.

Se puede observar que, en este ejemplo, la neurona artificial toma cuatro valores, multiplica a cada valor por un peso, y luego los suma. Luego de sumar los valores iniciales afectándolos con cada peso, la neurona le suma al resultado un sesgo (también llamado bias), para luego, finalmente, aplicar al resultado final una función de activación. Matemáticamente, el resultado

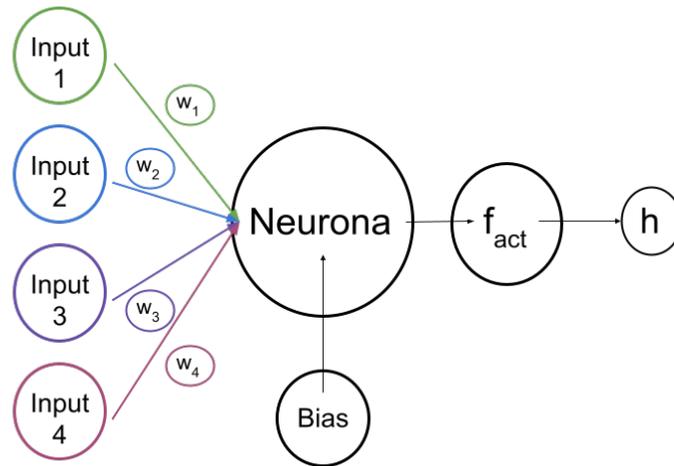


Figura 3.1: Estructura de una neurona artificial. Esta recibe entradas, o inputs, los cuales se afectan a través de un factor multiplicativo llamado peso. Luego de sumar todos los inputs con su correspondiente peso al bias, o sesgo, se aplica una función de activación para obtener el resultado final de esta neurona.

(h) de esta neurona puede expresarse como

$$h = f_{\text{act}}\left(\sum_{i=1}^4 I_i * w_i + b\right), \quad (3.1)$$

donde f_{act} representa la función de activación, I_i representa las entradas, o inputs de la red, w_i los pesos y b el bias de la neurona. Es importante mencionar que cada neurona tiene su propio bias y función de activación, mientras que por cada conexión entre neuronas hay asociado un peso. Estos valores son los que se ajustan durante el entrenamiento de la red, y son los responsables de que la red arroje un resultado u otro.

Una red neuronal consta de varias neuronas conectadas entre sí. Al considerar esto, resulta de gran importancia la función de activación, la cual en general es la misma para todas las neuronas de la red, y cumple la función de evitar que la red colapse a una única neurona. Esto es importante porque una red con una sola neurona tiene una limitada capacidad de aprendizaje [1, 78].

3.2. Estructura de la red neuronal

La red neuronal es un conjunto de neuronas artificiales conectadas entre sí de forma tal que los resultados de cada neurona constituyen los valores que toman las otras. En este trabajo se utilizaron redes en las cuales la información siempre avanza ("feed-forward networks") [79], es decir, las primeras neuronas toman los inputs, realizan sus cálculos, pasan sus valores a las neuronas siguientes y así hasta obtener el resultado final.

Un esquema de la estructura empleada para estas redes se puede ver en la figura 3.2.

En el ejemplo de la figura, se ve una red que toma tres valores iniciales, o inputs, luego tiene dos capas internas con cuatro neuronas en cada una, y finalmente tiene una capa externa con una neurona que entrega el resultado final. De manera adicional, se puede ver que todas las

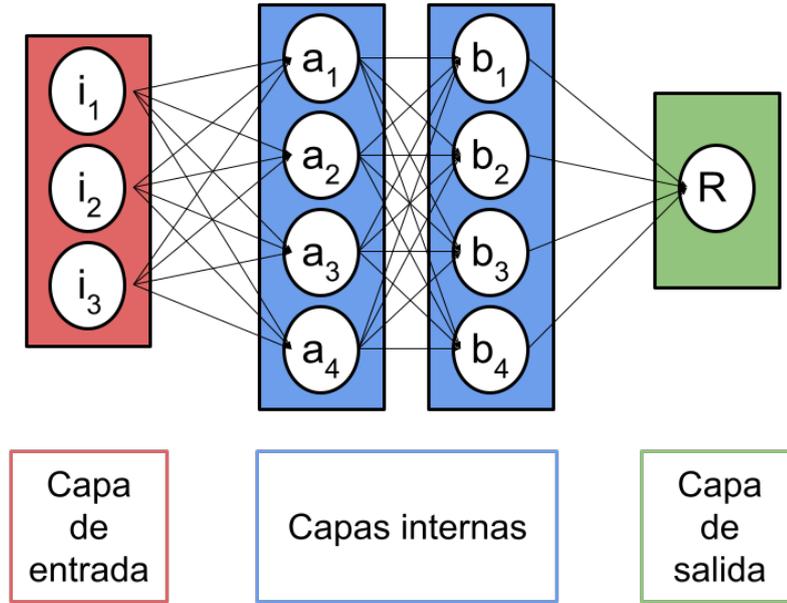


Figura 3.2: Estructura de ejemplo de una red neuronal con tres entradas, dos capas ocultas con cuatro neuronas cada una y una capa externa con una única neurona. Se observa que todas las neuronas están conectadas con todas de la siguiente capa, lo cual significa que este esquema corresponde a una red de conexión completa.

neuronas de una capa se conectan con todas las neuronas de la capa siguiente. Este tipo de red se empleó en el trabajo, y se llama de conexión completa ("fully connected") [74].

Para expresar como trabaja la red se puede escribir la ecuación para cada neurona de las capas internas. Por ejemplo, para la primer capa interna resulta que

$$a_j = f_{\text{act}}\left(\sum_{k=1}^3 i_k w_{a_{j_k}} + b_{a_j}\right), \quad (3.2)$$

donde se está asumiendo que todas las neuronas de la primer capa interna tienen la misma función de activación. Además, se ve que, al recibir tres inputs, cada neurona tiene tres pesos asociados y su bias. Es importante remarcar que cada una de las cuatro neuronas de la capa a tiene cuatro pesos distintos, es decir, en la primer capa interna hay doce pesos $w_{a_{j_k}}$ diferentes.

Por otro lado, para la segunda capa interna se puede escribir que

$$b_j = f_{\text{act}}\left(\sum_{k=1}^4 a_k w_{b_{j_k}} + b_{b_j}\right), \quad (3.3)$$

donde nuevamente se asumió la misma función de activación para todas las neuronas. Además, es explícito notar que los inputs de la segunda capa son los resultados de la primera, y que en este caso, al haber cuatro neuronas que reciben cuatro inputs, hay un total de dieciséis pesos en la segunda capa interna.

Finalmente, el resultado de la red se puede calcular como

$$R = f_{\text{act}}\left(\sum_{k=1}^4 b_k w_{R_k} + b_R\right), \quad (3.4)$$

donde debe notarse que los valores b_k son los calculados en 3.3, y estos a su vez se calculan a partir de 3.2. De esta manera, se evidencia el trabajo explícito que realiza la red para operar sobre los inputs y obtener el resultado. También se muestra explícitamente que el resultado depende de los pesos internos asociados a las conexiones neuronales y los bias de cada neurona, a los que se los llama parámetros internos de la red y son los valores que se modifican durante el entrenamiento para lograr el resultado deseado. En esta red que se utilizó de ejemplo, puede verse que hay un total de 32 pesos internos y 9 bias (uno por cada neurona). En la siguiente sección se discutirá como el entrenamiento modifica estos valores para así lograr una red eficiente.

Como se mencionó en la sección anterior, si se utiliza una función de activación lineal siempre es posible reemplazar todas las neuronas de las capas internas por una única neurona con pesos a determinar. Esto se debe a que todos los cálculos realizados son lineales, y es algo a evitar, ya que una única neurona tiene una capacidad de aprendizaje menor que una red de varias neuronas. Además, la función de activación debe ser inyectiva para no dar un mismo resultado para dos valores diferentes del cálculo de las neuronas. En este trabajo se emplearon como funciones de activación la tangente hiperbólica y la función SiLU, las cuales se comentan en la siguiente sección.

Como comentario final en esta sección referida a la estructura de las redes, se suele llamar arquitectura de la red a la cantidad de capas internas, y a la cantidad de neuronas que contiene cada capa interna. En el ejemplo, la arquitectura de la red corresponde a dos capas internas con cuatro neuronas. Esto se puede denotar como (4,4).

3.3. Las funciones de activación

Al discutir acerca de la estructura de las redes neuronales se mencionó la importancia de la función de activación, la cual, al ser no lineal, evita que una red con una gran cantidad de neuronas colapse a una red con una única neurona [1]. En este trabajo se emplearon dos tipos diferentes de función de activación: la tangente hiperbólica y la función SiLU [80], a la cual a veces se la suele llamar "swish"¹ [81], cuyos gráficos se muestran en la figura 3.3.

¹Existen pequeñas diferencias entre la función SiLU y la swish, pero a efectos prácticos son iguales.

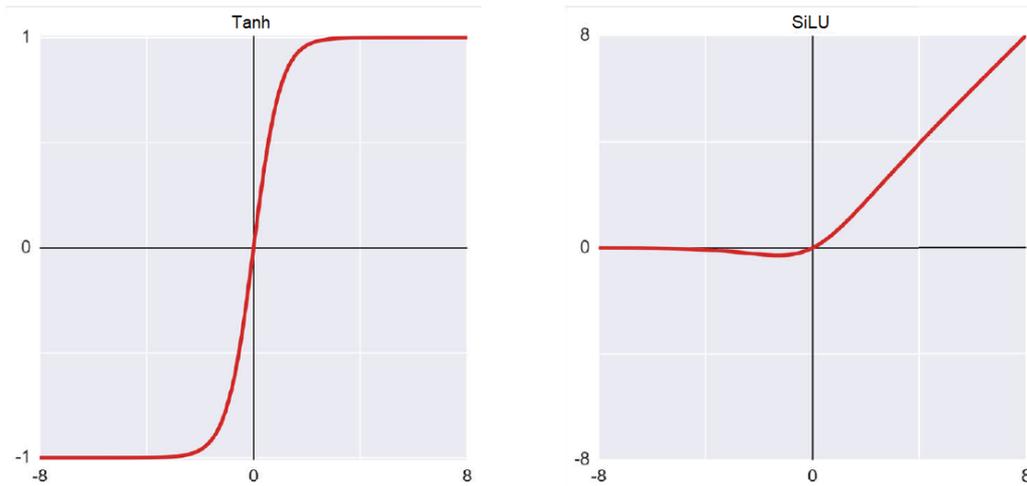


Figura 3.3: A la izquierda la tangente hiperbólica, a la derecha la función SiLU. En ambos casos las funciones son no lineales lo cual evita el colapso de la red a una única neurona. Figura extraída del libro [1].

Como se mencionó previamente, la función de activación se aplica luego de que las neuronas realicen los cálculos sobre los inputs empleando los pesos. Entonces, el valor que llega a la función de activación es una combinación lineal de las entradas de cada neurona. Se puede ver en la figura 3.3 que la tangente hiperbólica se satura rápidamente para valores con módulo mayor a uno. Este fue uno de los motivos por el cual se reescaló la variable independiente en la ecuación diferencial del modelo a un intervalo cuyo módulo sea menor o igual que uno, para de esta forma garantizar que las funciones de activación no se saturan.

En el caso de la función SiLU, esta se define como

$$\text{SiLU} = x\sigma_{\text{sig}}(x), \quad (3.5)$$

donde $\sigma_{\text{sig}}(x)$ corresponde a la función sigmoide. La mencionada diferencia con la función Swish consiste en que esta función se define como

$$\text{swish} = x\sigma_{\text{sig}}(\beta x), \quad (3.6)$$

donde β es un parámetro que se ajusta durante el entrenamiento. Como se mencionó previamente, en este trabajo se emplearon Tanh y SiLU, cuyas comparaciones se pueden ver en el capítulo de Resultados.

3.4. Aprendizaje de la red neuronal

Hasta ahora se ha mencionado como opera una neurona artificial, y luego como se conectan estas para obtener un resultado que depende de múltiples parámetros. En esta sección se detalla la forma en que estos parámetros internos de la red (los pesos y los bias) se ajustan para optimizar el resultado.

Se puede pensar que una red neuronal es una función que toma n inputs y devuelve un

número. Como se mencionó en la sección anterior, esta función depende explícitamente de los parámetros internos y de los inputs. Sea $\vec{\Theta}$ un vector que contiene a todos los parámetros internos (es decir, a todos los pesos y los bias), y \vec{x} un vector que contiene todos los inputs de la red, se puede pensar a la red neuronal como una función $F(\vec{x}, \vec{\Theta}) : \mathbb{R}^n \rightarrow \mathbb{R}$.

El objetivo de entrenar una red neuronal es, para una dada arquitectura, hallar el vector $\vec{\Theta}$ que mejor resuelve un dado problema. Para lograr esto, se puede definir una función de pérdida, también llamada "loss", que cuantifique que tan diferente es el resultado que obtiene la red respecto de un valor de referencia.

3.4.1. La función de pérdida

Antes de continuar, es necesario aclarar que en el campo del aprendizaje automático existen distintos tipos de entrenamiento para las redes neuronales. Por ejemplo, se pueden dividir estos en dos grandes grupos: entrenamientos supervisados y no supervisados. En el aprendizaje supervisado, la red se entrena con pares de entrada-salida, mientras que en el aprendizaje no supervisado, la red debe descubrir patrones en los datos sin recibir información sobre las salidas deseadas.

En el presente trabajo se emplearon las redes para resolver ecuaciones diferenciales mediante un método no supervisado, lo cual implica que no se requieren datos externos para entrenarlas. Esto quiere decir que no se utilizan las soluciones numéricas de las ecuaciones en el entrenamiento de las redes, aunque luego sí se utilizaran esas soluciones para chequear los resultados obtenidos.

Es posible definir la función de pérdida, también llamada loss, de forma directa utilizando la misma ecuación que se quiere resolver. Por ejemplo, sea que se quiere resolver una ecuación para $x(t)$ del tipo

$$x'(t) = x(t). \quad (3.7)$$

Lo primero a notar, es que esta ecuación puede reescribirse como

$$x'(t) - x(t) = 0. \quad (3.8)$$

Luego, supongamos que la red neuronal entrega un valor x_N y puede calcular su derivada x'_N . En este caso, el valor de entrada de la red (el input) corresponderá al tiempo t . De esta manera puede calcularse una cantidad L de la forma

$$L = (x'_N - x_N)^2, \quad (3.9)$$

donde a la cantidad $x'_N - x_N$ se la suele llamar el residuo. Es directo notar que si la red obtiene la solución exacta a la ecuación, el residuo es cero y $L = 0$. Luego, cuanto más difiera la solución de la red de la solución exacta de la ecuación, se obtendrán valores más altos de L .

Esta definición de L constituye un posible ejemplo de función de pérdida. No obstante, es posible notar que esta L evalúa a la solución de la red en un único valor del tiempo, lo cual no es ideal al considerar un intervalo temporal.

Yendo a un caso general, si se desea resolver una ecuación diferencial de primer orden² del tipo

$$x' - f(x, t) = 0, \quad (3.10)$$

puede pensarse a la red como una función que toma un valor de tiempo t y entrega un valor de la función x . Es decir, en este caso la red neuronal constituye una función $F(t, \vec{\Theta}) : \mathbb{R} \rightarrow \mathbb{R}$.

Para construir una función de pérdida que permita que la red aprenda a resolver la ecuación en todo el intervalo, puede proponerse una del tipo

$$L = \sum_{i=1}^{N_{batch}} (x'_{N_i} - f(x_{N_i}, t_i))^2, \quad (3.11)$$

donde el subíndice i etiqueta distintos valores de la variable independiente t . Además, se puede observar que se han considerado N_{batch} valores distintos de la variable independiente para construir esta función de pérdida. De esta forma, al aumentar el valor de N_{batch} , se aumenta la cantidad de puntos en donde se evalúa la red para calcular la función de pérdida. A esta cantidad de puntos N_{batch} se la llama "Batch Size", y es un parámetro importante del entrenamiento. Es importante tener en consideración que estos puntos se toman de manera aleatoria con un muestreo lineal en cada interacción del entrenamiento, y, de esta forma, no suelen ser los mismos puntos en cada instancia del mismo.

De manera adicional, puede introducirse un nuevo factor en la función de pérdida que favorezca el aprendizaje de la red cerca de la condición inicial. Si en el problema planteado se tiene que

$$x(t_i) = x_i, \quad (3.12)$$

entonces puede considerarse

$$L = \sum_{j=1}^{N_{batch}} (x'_{N_j} - f(x_{N_j}, t_j))^2 e^{-\omega|t-t_i|}, \quad (3.13)$$

donde para valores mayores de ω se obtiene una red que ve su aprendizaje favorecido para valores de la variable independiente cercanos al valor en que está impuesta la condición inicial.

Hasta aquí se ha comentado la definición de la función de pérdida empleada en el trabajo. Sin embargo, esta definición no dice nada sobre el mecanismo por el cual la red modifica los parámetros para lograr el menor valor posible de L , ni tampoco se ha garantizado que la solución cumpla con la condición inicial. Con respecto a la minimización de la función de pérdida, el método usual consiste en emplear el descenso de gradiente estocástico, lo cual se explica en la siguiente subsección. Por otro lado, para lograr que la solución cumpla las condiciones iniciales se emplea una reparametrización de la salida de las redes neuronales en la cual se impone explícitamente dicha condición. Esta reparametrización, consiste básicamente en transformar la

²Solamente se especifican ecuaciones de primer orden, ya que al considerar ecuaciones de órdenes más altos siempre pueden reescribirse como un sistema de ecuaciones, todas de orden uno.

salida \tilde{x} de las redes de la forma

$$\tilde{x} \rightarrow x_i + (1 - e^{-t-t_i})x_N, \quad (3.14)$$

donde x_i es la condición inicial impuesta en el valor $t = t_i$ y x_N es la salida de las redes neuronales. Esto se explica en detalle más adelante, y en el siguiente capítulo se discute la implementación precisa al problema en cuestión.

3.4.2. El problema de optimización

En la sección anterior se ha presentado una definición de la función de pérdida que garantiza que, si la solución provista por la red neuronal se acerca a la solución exacta, entonces esta función de pérdida se acerca a cero.

El objetivo ahora es describir un mecanismo por el cual se pueda lograr que la solución provista por las redes se acerque lo más posible a la solución exacta. Para ello, se emplea el método del descenso de gradiente estocástico [82]. Para dar una idea de como funciona este algoritmo, lo primero que debe notarse es que la función de pérdida definida en 3.13 es, explícitamente, una función $L(\vec{\Theta})$, donde debe recordarse que $\vec{\Theta}$ representa a los parámetros internos de la red. De esta forma, en cada iteración del proceso de entrenamiento puede calcularse el gradiente

$$\vec{g} = \vec{\nabla}_{\vec{\Theta}} L(\vec{\Theta}), \quad (3.15)$$

y de esta forma puede realizarse la transformación del vector de parámetros internos de forma tal que

$$\vec{\Theta} - \varepsilon \vec{g} \rightarrow \vec{\Theta}, \quad (3.16)$$

donde ε es un nuevo parámetro importante en el entrenamiento, llamado tasa de aprendizaje, o "learning rate". Mediante este proceso se espera lograr que, en cada interacción, la red encuentre un nuevo vector $\vec{\Theta}$ de manera tal que la función de pérdida sea menor que la anterior.

Es interesante notar que la tasa de aprendizaje representa que tan brusca es la transformación de los parámetros internos de la red en cada interacción. Se puede pensar que, un learning rate más grande logra un aprendizaje más rápido, mientras que un learning rate más pequeño genera un aprendizaje más lento, pero que puede ser más preciso que el anterior [83]. Este descenso de gradiente se logra en la práctica a través de los algoritmos que son llamados optimizadores. Los optimizadores implementan la minimización de la función de pérdida y pueden ser más complejos que únicamente el algoritmo de descenso de gradiente. En el trabajo se empleó el optimizador "ADAM" [84]. Resulta importante mencionar que a los parámetros que corresponden al proceso de entrenamiento, tales como el batch size o el learning rate se los suele llamar hiperparámetros, y son aquellos que se fijan antes de comenzar y no se modifican durante el mismo.

3.5. Resolución de ecuaciones diferenciales

Hasta ahora se han comentado aspectos sobre el funcionamiento de las neuronas, la arquitectura de las redes neuronales, el proceso de entrenamiento y los hiperparámetros que controlan este último. Se ha logrado escribir una función de pérdida que logra cuantificar cuanto difiere la solución de la red neuronal respecto de la solución exacta. No obstante, a pesar de que en toda la discusión se ha enfocado el eje en que las redes que se emplearon en el trabajo están destinadas a resolver ecuaciones diferenciales, aún no se ha discutido en detalle como se han utilizado para lograr dicho objetivo.

En el presente trabajo se han estudiado ecuaciones diferenciales lineales de segundo orden. Sin embargo, para aplicar el método de las redes se ha convertido, cada ecuación de segundo orden, en un sistema de dos ecuaciones de primer orden.

Para comprender esto, se puede considerar la ecuación lineal más general de segundo orden para $x(t)$

$$x'' + h(t)x' + k(t)x = 0, \quad (3.17)$$

con condiciones iniciales $x(t_i) = x_i$ y $x'(t_i) = x'_i$, donde es importante notar que los coeficientes de la ecuación solamente dependen de la variable independiente t . Luego se realiza el cambio de variables

$$y = x', \quad (3.18)$$

logrando así reescribir la ecuación 3.17 como

$$y' + h(t)y + k(t)x = 0. \quad (3.19)$$

Se puede ver que ahora la ecuación es de primer orden. El precio a pagar por reducir el orden de la ecuación es que, ahora, esta ecuación involucra dos funciones desconocidas: x e y . Entonces, es análogo resolver la ecuación 3.17, a resolver el sistema

$$x' - y = 0. \quad (3.20a)$$

$$y' + h(t)y + k(t)x = 0 \quad (3.20b)$$

De esta forma se puede transformar cualquier ecuación de segundo orden en dos ecuaciones de primer orden. Además, las condiciones iniciales se transforman de manera directa

$$x(t_i) = x_i \quad \rightarrow \quad x(t_i) = x_i \quad (3.21a)$$

$$x'(t_i) = x'_i \quad \rightarrow \quad y(t_i) = y_i = x'_i. \quad (3.21b)$$

En este trabajo, se ha procedido de esta forma para resolver, para cada ecuación de se-

gundo orden, un sistema de dos ecuaciones diferenciales de primer orden. Respecto a las redes neuronales, se ha utilizado una red para cada ecuación. Es decir, para resolver una ecuación de segundo orden, se han entrenado de manera simultánea dos redes neuronales diferentes: una que devuelve la función y otra que devuelve su derivada.

Para garantizar que se cumplan las condiciones iniciales se puede emplear una reparametrización de la solución provista por las redes neuronales como la que se mencionó previamente (ec. 3.14). Si se piensa en un sistema de ecuaciones como 3.20 con condiciones iniciales como las de 3.21, las redes neuronales entregan soluciones x_N e y_N . Si se considera el caso de x_N , puede definirse, como en [85]

$$\tilde{x}(t) = x_i + (1 - e^{-(t-t_i)})x_N(t), \quad (3.22)$$

de donde se ve directamente que $\tilde{x}(t_i) = x_i$. De la misma forma se considera

$$\tilde{y}(t) = y_i + (1 - e^{-(t-t_i)})y_N(t), \quad (3.23)$$

que cumple que $\tilde{y}(t_i) = y_i$. Si para cada red se utilizan estas reparametrizaciones para construir la función de pérdida de la misma forma que en 3.13, se garantiza que la solución que se obtiene de las redes cumpla las condiciones iniciales. De esta forma, la función de pérdida resulta

$$L(\tilde{x}, \tilde{y}, t) = \sum_{j=1}^{N_{batch}} \left(Res_1(\tilde{x}_j, \tilde{y}_j, t_j)^2 + Res_2(\tilde{x}_j, \tilde{y}_j, t_j)^2 \right) e^{-\omega|t-t_i|}, \quad (3.24)$$

Es importante aclarar que con este método, durante el entrenamiento se modifican los parámetros internos para ajustar x_N e y_N , pero al momento de realizar cálculos, deben realizarse las transformaciones 3.22 y 3.23 para obtener la solución correcta. Esta solución cumple exactamente las condiciones iniciales, y dependiendo de la calidad del entrenamiento será más o menos similar a la solución exacta fuera de $t = t_i$.

Cabe mencionar que esta imposición de las condiciones iniciales se llevó a cabo mediante la función `BundleIVP` de la librería `neurodifreq`.

3.6. El método bundle

Hasta este punto se ha detallado un método que permite entrenar una red neuronal para que adquiera la capacidad de resolver ecuaciones diferenciales. Sin embargo, se ha omitido, por simplicidad, una cuestión cuya relevancia significa una ventaja significativa del método con respecto a los métodos numéricos tradicionales, el método `bundle` [54].

Notese que en la ecuación 3.17 no se ha asumido ninguna dependencia en parámetros del modelo teórico empleado. Si se considera entonces una nueva ecuación a resolver, la cual depende de estos, por ejemplo

$$x'' + h(t, \vec{\alpha})x' + k(t, \vec{\alpha})x = 0, \quad (3.25)$$

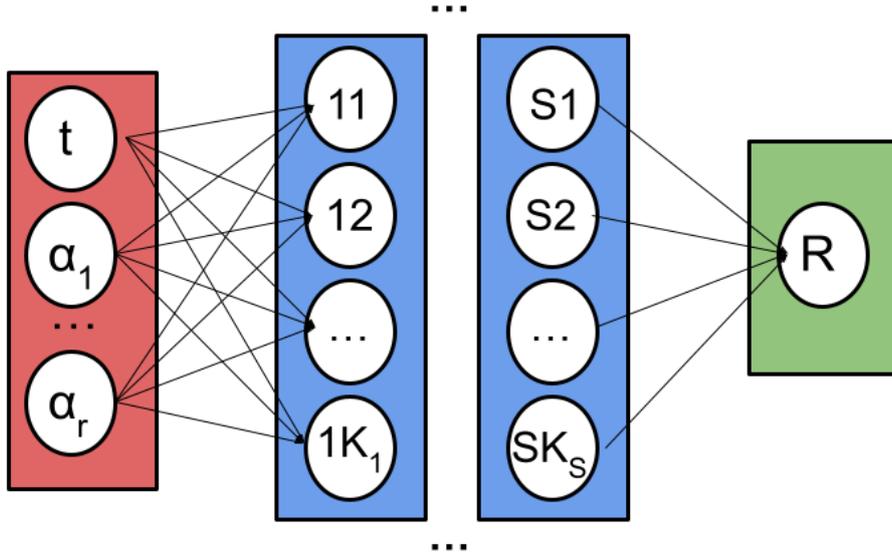


Figura 3.4: Esquema general de una red neuronal con $r + 1$ entradas, S capas internas con K_j neuronas en cada capa ($j = 1, \dots, S$) y una única neurona en la capa externa.

donde $\vec{\alpha}$ representa un vector de parámetros del modelo teórico. Entonces para integrar la ecuación a través de métodos numéricos tradicionales debe especificarse el valor de cada uno de los α_i . Es decir, debe integrarse la ecuación tantas veces como vectores $\vec{\alpha}$ se consideren. Esto representa un gasto computacional alto a la hora de realizar un análisis estadístico que explore el espacio de los parámetros integrando la ecuación con métodos numéricos tradicionales. En el método de las redes neuronales es posible tratar este problema de una forma que podría garantizar eficiencia computacional a través del llamado método bundle. Dicho método consiste en realizar el entrenamiento utilizando a cada valor de los parámetros del modelo como un input de la red. De esta forma, se logra entrenar una red cuya solución puede evaluarse, no solo en distintos valores de la variable independiente, sino que además puede evaluarse en distintos valores de cada parámetro que se incluya en el bundle.

Por ejemplo, si se considera un vector con r parámetros del bundle del tipo $\vec{\alpha} = (\alpha_1, \dots, \alpha_r)$, y una red neuronal con S capas internas y K_j neuronas en la capa interna j -ésima ($j = 1, \dots, S$), un esquema posible con conexión completa de esta red se puede ver en la figura 3.4.

En el caso de la reparametrización que impone las condiciones iniciales, al considerar una solución bundle se considera

$$\tilde{x}(t, \vec{\alpha}) = x_i(\vec{\alpha}) + (1 - e^{-(t-t_i)}) x_N(t, \vec{\alpha}) \quad (3.26a)$$

$$\tilde{y}(t, \vec{\alpha}) = y_i(\vec{\alpha}) + (1 - e^{-(t-t_i)}) y_N(t, \vec{\alpha}), \quad (3.26b)$$

donde ahora las condiciones iniciales x_i e y_i pueden depender de los parámetros del modelo.

Por otro lado, si se considera el sistema de ecuaciones de primer orden análogo a 3.25, y se definen los residuos como la cantidad por la que las redes difieren de la solución exacta, es decir

$$\tilde{x}' - \tilde{y} = Res_1, \quad (3.27a)$$

$$\tilde{y}' + h(t, \vec{\alpha})\tilde{y} + k(t, \vec{\alpha})\tilde{x} = Res_2, \quad (3.27b)$$

entonces la función de pérdida que debe considerarse a la hora de realizar el entrenamiento es, en lugar de 3.13,

$$L(\tilde{x}, \tilde{y}, t, \vec{\alpha}) = \sum_{j=1}^M \left(Res_1(\tilde{x}_j, \tilde{y}_j, t_j, \vec{\alpha}_j)^2 + Res_2(\tilde{x}_j, \tilde{y}_j, t_j, \vec{\alpha}_j)^2 \right) e^{-\omega|t-t_j|}, \quad (3.28)$$

donde el subíndice j en cada dependencia de los residuos representa el valor que ha tomado la red para cada muestra en una misma interacción del entrenamiento. Esto significa que M representa el total de puntos en que se ha evaluado la red para una única iteración. En el caso sin bundle, M representa el batch size, sin embargo, en el caso bundle, deben tomarse muestras del dominio de $\vec{\alpha}$, por lo que M será el producto de los batch size de la variable independiente y de cada parámetro del modelo.

Por ejemplo, si se supone un modelo con dos parámetros y la variable temporal, y se toma un batch size de 16, tanto para los parámetros como para t , entonces $M = 16 * 16 * 16$. Una aclaración pertinente es que no es necesario tomar el mismo batch size para todas las entradas de la red, estos pueden ser diferentes entre sí.

De esta forma, la función de pérdida definida en 3.28 permite llevar a cabo un entrenamiento donde ahora, la red, no solamente se entrena para distintos valores de la variable independiente t , sino que además se entrena para distintos valores de los parámetros del modelo $\vec{\alpha}$. Además, se ha incluido el factor de peso ω , el cual puede ayudar al entrenamiento a privilegiar puntos cercanos a la condición inicial, donde si se elige $\omega = 0$ este factor desaparece y no hay privilegio en ningún punto del dominio temporal.

En resumen, se ha descrito un método que permite obtener una red que puede ser evaluada para cualquier valor de los parámetros del modelo incluidos en el rango de entrenamiento. Esto significa que a la hora de explorar el espacio de parámetros, en lugar de realizar cálculos costosos computacionalmente, únicamente debe evaluarse una función. En trabajos anteriores del grupo de investigación se ha visto que esto permite reducir significantivamente los tiempos computacionales del análisis estadístico con cadenas de Markov [47, 86].

En el trabajo se ha empleado este método a través de la librería neurodiffee para los modelos teóricos descritos previamente.

3.7. La arquitectura customizada

Lo descrito hasta este momento en este capítulo corresponde a lo que, durante el trabajo, llamaremos arquitectura tradicional o habitual. En esta sección se discute una implementación en la arquitectura de las redes neuronales que puede suponer una ventaja respecto a la tradicional: la arquitectura customizada, cuyo desarrollo sigue en curso a través del grupo de trabajo del

co-director de esta tesis.

El objetivo de esta nueva arquitectura consiste en optimizar el proceso de entrenamiento a través de descorrelacionar los pesos de la primer capa oculta, restringiendo cada neurona a un sub-intervalo más pequeño. Para comprender esto, se asume una red neuronal con un único input y K_1 neuronas en la primer capa oculta. Esta estructura se ve en la figura 3.5.

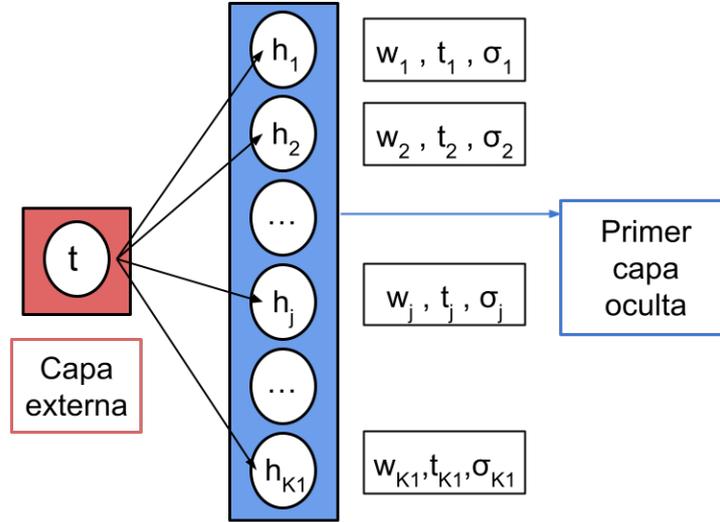


Figura 3.5: Ejemplo de red neuronal con la arquitectura customizada. En la figura se ve una capa externa con una única entrada y la primer capa oculta con K_1 neuronas. Además de los pesos se incluye en cada neurona un valor t_j que corresponde al valor central de una gaussiana y además la cantidad σ_j que corresponde a la desviación estándar de esta gaussiana.

En la figura se observa que cada neurona tiene un peso asociado al input, pero además se introdujeron dos nuevos valores: t_j y σ_j , donde $j = 1, 2, \dots, K_1$. Los valores t_j son nuevos hiperparámetros, es decir, se fijan antes de comenzar el entrenamiento, mientras que los valores σ_j corresponden a parámetros internos de la red que se ajustan durante el mismo.

Lo novedoso en esta arquitectura es que, al calcular el valor que obtiene la neurona j -ésima de la primer capa oculta, se multiplica al peso w_j por una distribución gaussiana de valor medio t_j y desviación estándar σ_j . De este modo, al introducir un input t la neurona j -ésima calcula su valor h_j como

$$h_j = f_{\text{act}}(N(t, t_j, \sigma_j)tw_j + b_j), \quad (3.29)$$

donde $N(t, t_j, \sigma_j)$ representa una función gaussiana de valor medio t_j , desviación estándar σ_j evaluada en el punto t , y, al igual que antes, w_j y b_j son el peso y el bias asociados a la neurona j -ésima, respectivamente. Es importante remarcar que para cada neurona hay un valor establecido de t_j . Para establecer este valor se toman s puntos equiespaciados dentro del intervalo de entrenamiento de la variable independiente, donde $s = K_1$. Luego, se asigna un t_j a cada neurona de la primer capa oculta. Esta modificación se aplica únicamente sobre las neuronas de la primer capa, resultando los cálculos realizados por las neuronas de las demás capas idénticos a lo comentado anteriormente en este capítulo. En este trabajo se emplearon

redes con $K_1 = 256$ neuronas en la primer capa, y se tomó $s = 256$, resultando entonces cada neurona con un propio valor central para su gaussiana. Se asume que trabajar con redes con $1 \ll K_1 = s$ tiene mejores resultados debido a que se realiza una mejor subdivisión del intervalo de entrenamiento de la variable independiente.

El hecho de asociar una gaussiana a cada neurona de la primer capa implica que, en función de la desviación estándar σ_j correspondiente, dicha neurona vea restringido el intervalo en el cual afecta al resultado. Por ejemplo, si la primer neurona dispone de una gaussiana muy centralizada cerca del valor t_1 (es decir, con $\sigma_1 \ll \sigma_{k \neq 1}$), entonces esta neurona tendrá un peso importante al computar valores de la variable independiente cercanos a t_1 , pero no tendrá relevancia al calcular valores donde t se aleje de dicho valor. Esto descorrelaciona los pesos de la primer capa, lo que se espera que optimice el proceso de entrenamiento.

Al considerar el problema bundle, la arquitectura customizada se emplea de la misma manera, pero asignando una gaussiana no solo a la variable independiente, sino también a todas las entradas de la red. De esta forma, si los inputs corresponden a la variable independiente t y a r parámetros del bundle α_j , con $j = 1, \dots, r$, se considera el vector de $r+1$ entradas $\vec{\zeta} = (t, \alpha_1, \dots, \alpha_r)$. Entonces, en cada neurona j -ésima de la primer capa se asocia una gaussiana para cada input ζ_k . De este modo, el cálculo del valor h_j de la neurona j -ésima se realiza

$$h_j = f_{\text{act}}(N(t, t_{jk}, \sigma_{jk})tw_j + b_j), \quad (3.30)$$

donde, como se mencionó, ahora cada gaussiana tiene dos índices $j = 1, \dots, K_1$ referido a la neurona y $k = 1, \dots, r + 1$ a la entrada correspondiente.

Es importante recordar que esta arquitectura customizada está siendo desarrollada y por este motivo puede presentar cambios o alternativas, tales como incluir a los valores t_j dentro de los parámetros del entrenamiento, o fijar el valor de las desviaciones σ_j . En este trabajo se compararon los resultados de esta nueva arquitectura con la tradicional, y los resultados se comentan en el capítulo correspondiente.

Capítulo 4

Implementaciones del método de las redes neuronales

Hasta ahora se han presentado los modelos teóricos considerados durante este trabajo. Además, se ha discutido el método de aprendizaje automático empleado para desarrollar entrenamientos de redes neuronales que aprendan a resolver las ecuaciones descritas en los modelos teóricos. En este capítulo se detalla la implementación específica del método de las redes neuronales a los modelos teóricos considerados. Además, en esta tesis se incluye la implementación de dicho método al análisis estadístico realizado mediante las cadenas de Markov (MCMC). Cabe destacar que todos los entrenamientos correspondientes a las redes neuronales se llevaron a cabo en una GPU, mediante una tarjeta gráfica Nvidia A100. Por otro lado, los procesos de MCMC se llevaron a cabo en una CPU con procesador Intel(R) Core(TM) i7-8550U 1.99 GHz.

4.1. Implementación a los modelos teóricos

En las secciones anteriores se ha presentado el método de las redes neuronales para una ecuación diferencial lineal de segundo orden genérica, como por ejemplo 3.25. En esta sección se presenta la implementación del método a las ecuaciones correspondientes a los modelos teóricos estudiados en este trabajo.

Resulta importante remarcar que, en ambos casos considerados, y en el caso de cualquier ecuación diferencial a resolver, se detiene el entrenamiento de las redes y se asume que éstas ya están entrenadas cuando la función de pérdida es lo suficientemente baja, lo cual en la práctica implica un descenso de más de cuatro órdenes de magnitud a partir del comienzo del entrenamiento. Este comentario resulta pertinente, ya que no hay un criterio unificado sobre el valor de la función de pérdida necesario para finalizar el entrenamiento. En este trabajo se han dado por finalizados los entrenamientos cuando las soluciones obtenidas alcanzaron diferencias porcentuales menores al 1% con respecto a las soluciones numéricas. Sin embargo, en el caso de desconocer la solución numérica, un criterio razonable para detener el entrenamiento y considerar que la solución obtenida es correcta, es que la función pérdida haya disminuido en 4 o 5 órdenes de magnitud respecto de su valor al comenzar el entrenamiento.

4.1.1. Implementación a la ecuación para perturbaciones de materia en Λ CDM

Previamente se ha mostrado que la ecuación para estudiar la evolución de las perturbaciones de materia en el universo es 2.43

$$\delta_m''(a) + a_{\text{eq}} \left[\frac{2 + 3\frac{a}{a_{\text{eq}}} + 6\alpha \left(\frac{a}{a_{\text{eq}}}\right)^4}{2a \left(1 + \frac{a}{a_{\text{eq}}} + \alpha \left(\frac{a}{a_{\text{eq}}}\right)^4\right)} \right] \delta_m'(a) - \left[\frac{3a_{\text{eq}}}{2a \left(1 + \frac{a}{a_{\text{eq}}} + \alpha \left(\frac{a}{a_{\text{eq}}}\right)^4\right)} \right] \delta_m(a) = 0. \quad (4.1)$$

Es directo ver que esta ecuación tiene la forma genérica discutida previamente para el método de las redes (3.25)

$$x'' + h(t, \vec{\alpha})x' + k(t, \vec{\alpha})x = 0, \quad (4.2)$$

con la diferencia de que $x \rightarrow \delta$ y $t \rightarrow a$.

Con respecto a los parámetros, durante el trabajo se empleó un valor fijo para $\Omega_{r0} = 5,38 \times 10^{-5}$ [63], se empleó un valor usual para $h = 0,7$, y además se asumió un universo plano, motivo por el cual el único parámetro cosmológico a determinar es Ω_{m0} , pues una vez determinado este, $\Omega_{\Lambda0} = 1 - \Omega_{r0} - \Omega_{m0}$. En la ecuación 2.43 el parámetro Ω_{m0} aparece explícitamente en la definición de a_{eq} y α .

De esta manera, la ecuación a resolver con las redes es de la forma

$$\delta'' + h(a, \Omega_{m0})\delta' + k(a, \Omega_{m0})\delta = 0, \quad (4.3)$$

donde la prima denota derivadas respecto del factor de escala a y el único parámetro del bundle es Ω_{m0} .

Sin embargo, la ecuación escrita de esta manera presentó varios problemas a la hora de realizar el entrenamiento de las redes, lo que motivó a realizar múltiples cambios de variables.

Primer cambio de variables: $N = \ln(a)$

Por un lado, como se discutió en la sección correspondiente a las condiciones iniciales, estas se establecen en $a_i = 10^{-3}$, mientras que la parte de la solución que debe emplearse se halla cerca de $a_f = 1$. Esto quiere decir que las redes deben entrenarse, por lo menos, en el intervalo $[10^{-3}, 1]$. Como la forma más sencilla de realizar el sampleo de la variable independiente a la hora de realizar el entrenamiento es lineal, este cambio de órdenes de magnitud es difícil de captar en el entrenamiento. Por este motivo, se realizó el cambio de variables

$$N = \ln(a), \quad (4.4)$$

donde \ln es el logaritmo natural, y por lo tanto el intervalo de entrenamiento resulta $[\ln(10^{-3}), 0]$. Es directo mostrar que con este cambio ocurre que

$$\frac{d}{da} = e^{-N} \frac{d}{dN} \quad \rightarrow \quad \frac{d^2}{da^2} = e^{-2N} \left(\frac{d^2}{dN^2} - \frac{d}{dN} \right). \quad (4.5)$$

Si se aplica esta transformación a la ecuación, y a las condiciones iniciales, resulta entonces

$$\frac{d^2 \delta}{dN^2} + \left(\frac{1 + 4\alpha \left(\frac{e^N}{a_{\text{eq}}} \right)^3}{2 \left(1 + \frac{a_{\text{eq}}}{e^N} + \alpha \left(\frac{e^N}{a_{\text{eq}}} \right)^3 \right)} \right) \frac{d\delta}{dN} - \frac{3e^N}{2a_{\text{eq}} \left(1 + \frac{e^N}{a_{\text{eq}}} + \alpha \left(\frac{e^N}{a_{\text{eq}}} \right)^4 \right)} \delta = 0, \quad (4.6)$$

con las condiciones iniciales

$$\delta(N_i) = e^{N_i} \quad (4.7a)$$

$$\frac{d\delta}{dN}(N_i) = e^{N_i}, \quad (4.7b)$$

donde se ha definido $N_i = \ln(a_i) = \ln(10^{-3})$.

Esta ecuación tampoco pudo resolverse por medio del método de las redes, por, al menos, dos problemas distintos. Lo primero que se notó es que el intervalo de la variable independiente corresponde a, aproximadamente, $[-6,93, 0]$. Se ha mencionado previamente que una de las funciones de activación empleadas fue la tangente hiperbólica, la cual puede saturarse para valores que se alejen del intervalo $[-1, 1]$. para evitar esto, se reescaló la variable N con el objetivo de que esto no suceda.

Segundo cambio de variables: reescalo de N

Con el fin de llevar la variable independiente al intervalo $[-1, 0]$, se consideró la nueva variable

$$\hat{N} = \frac{N}{n_i}, \quad (4.8)$$

donde se ha definido $n_i = |N_i|$. De esta forma se obtienen las transformaciones para las derivadas

$$\frac{d}{dN} = \frac{1}{n_i} \frac{d}{d\hat{N}} \quad \rightarrow \quad \frac{d^2}{dN^2} = \frac{1}{a_i^2} \frac{d^2}{d\hat{N}^2}. \quad (4.9)$$

Aplicando esta transformación a la ecuación diferencial 4.6 se obtiene

$$\frac{d^2 \delta}{d\hat{N}^2} + \left(\frac{n_i (1 + 4\alpha \left(\frac{e^{\hat{N}n_i}}{a_{\text{eq}}} \right)^3)}{2 \left(1 + \frac{a_{\text{eq}}}{e^{\hat{N}n_i}} + \alpha \left(\frac{e^{\hat{N}n_i}}{a_{\text{eq}}} \right)^3 \right)} \right) \frac{d\delta}{d\hat{N}} - \frac{3e^{\hat{N}n_i} n_i^2}{2a_{\text{eq}} \left(1 + \frac{e^{\hat{N}n_i}}{a_{\text{eq}}} + \alpha \left(\frac{e^{\hat{N}n_i}}{a_{\text{eq}}} \right)^4 \right)} \delta = 0, \quad (4.10)$$

con las condiciones iniciales

$$\delta(\hat{N}_i) = e^{n_i} \quad (4.11a)$$

$$\frac{d\delta}{d\hat{N}}(\hat{N}_i) = n_i e^{n_i}, \quad (4.11b)$$

donde ahora $\hat{N}_i = -1$ y el intervalo de la variable independiente corresponde a $[-1, 0]$.

Este reescalo tampoco fue suficiente para lograr un entrenamiento adecuado de las redes. Un estudio simplificado de la ecuación mostró que esta tiene un comportamiento similar a una exponencial, y realizando estudios simplificados con las redes, se observó que estas tienen problemas para entrenarse cuando las soluciones cambian varios órdenes de magnitud en el intervalo de integración, que es el caso de una solución exponencial.

Tercer cambio de variables: $X = \ln(\delta)$

Para analizar el motivo por el cual los entrenamientos de las redes no podían completarse de manera satisfactoria, se asumieron coeficientes constantes para la ecuación 2.43. Como se mencionó previamente, esta ecuación con coeficientes constantes mostró un comportamiento exponencial en la solución. Esto motivó a realizar un nuevo cambio de variables, en el cual se pudiera lograr que la solución de la ecuación diferencial obtenga valores cuyos órdenes de magnitud no muestren cambios grandes. Para ello, se consideró como nueva variable dependiente en la ecuación 4.10

$$X = \ln(\delta). \quad (4.12)$$

Para implementar una transformación de la variable dependiente, debe considerarse nuevamente la regla de la cadena. De esta forma se obtiene que

$$\frac{d\delta}{d\hat{N}} = \frac{d\delta}{dX} \frac{dX}{d\hat{N}}, \quad (4.13)$$

donde $\frac{d\delta}{dX} = e^X = \delta$. De esta forma, y aplicando nuevamente la regla de la cadena para el caso de la derivada segunda, se obtiene que

$$\frac{d\delta}{d\hat{N}} = e^X \frac{dX}{d\hat{N}} \rightarrow \frac{d^2\delta}{d\hat{N}^2} = e^X \left(\frac{d^2X}{d\hat{N}^2} + \left(\frac{dX}{d\hat{N}} \right)^2 \right). \quad (4.14)$$

Es directo notar que todos los factores e^X se cancelaran al introducir estas transformaciones en la ecuación 4.10. A pesar de esto, también se puede ver que este cambio de variables introduce un término no lineal en la ecuación. Sin embargo, por la naturaleza del método, esto no representará un problema para las redes.

Aplicando el nuevo cambio de variables a la ecuación 4.10 se obtiene

$$\frac{d^2 X}{d\hat{N}^2} + \left(\frac{dX}{d\hat{N}} \right)^2 + \left(\frac{n_i(1 + 4\alpha(\frac{e^{\hat{N}n_i}}{a_{\text{eq}}})^3)}{2(1 + \frac{a_{\text{eq}}}{e^{\hat{N}n_i}} + \alpha(\frac{e^{\hat{N}n_i}}{a_{\text{eq}}})^3)} \right) \frac{dX}{d\hat{N}} - \frac{3e^{\hat{N}n_i}n_i^2}{2a_{\text{eq}}(1 + \frac{e^{\hat{N}n_i}}{a_{\text{eq}}} + \alpha(\frac{e^{\hat{N}n_i}}{a_{\text{eq}}})^4)} = 0, \quad (4.15)$$

con las condiciones iniciales

$$X(\hat{N}_i) = -n_i = X_i \quad (4.16a)$$

$$\frac{dX}{d\hat{N}}(\hat{N}_i) = n_i = Y_i. \quad (4.16b)$$

Esta forma para la ecuación diferencial trajo como beneficio que la variable dependiente X no cambie en varios órdenes de magnitud. De esta manera, se observó que fue posible completar de manera satisfactoria el entrenamiento de las redes neuronales.

Implementación al método bundle

Como ya se ha discutido, definiendo

$$Y = \frac{dX}{d\hat{N}}, \quad (4.17)$$

puede transformarse la ecuación de segundo orden 4.15 al sistema

$$\frac{dX}{d\hat{N}} - Y = 0 \quad (4.18a)$$

$$\frac{dY}{d\hat{N}} + Y^2 + h(\hat{N}, \Omega_{m0})Y + k(a, \Omega_{m0}) = 0, \quad (4.18b)$$

donde

$$h(\hat{N}, \Omega_{m0}) = \left(\frac{n_i(1 + 4\alpha(\frac{e^{\hat{N}n_i}}{a_{\text{eq}}})^3)}{2(1 + \frac{a_{\text{eq}}}{e^{\hat{N}n_i}} + \alpha(\frac{e^{\hat{N}n_i}}{a_{\text{eq}}})^3)} \right), \quad (4.19)$$

y

$$k(\hat{N}, \Omega_{m0}) = -\frac{3e^{\hat{N}n_i}n_i^2}{2a_{\text{eq}}(1 + \frac{e^{\hat{N}n_i}}{a_{\text{eq}}} + \alpha(\frac{e^{\hat{N}n_i}}{a_{\text{eq}}})^4)}. \quad (4.20)$$

Resulta importante además recordar que

$$a_{\text{eq}}(\Omega_{m0}) = \frac{\Omega_{r0}}{\Omega_{m0}} \quad \alpha(\Omega_{m0}) = a_{\text{eq}}(\Omega_{m0})^4 \frac{\Omega_{\Lambda 0}}{\Omega_{r0}} \quad \Omega_{\Lambda 0}(\Omega_{m0}) = 1 - \Omega_{r0} - \Omega_{m0}, \quad (4.21)$$

dejando en evidencia que el único parámetro relevante en la ecuación es Ω_{m0} . Además, también es importante recordar que \hat{N} pertenece al intervalo $[-1, 0]$, y que $n_i = |\ln(10^{-3})|$.

De esta manera, se definen los residuos de las ecuaciones, necesarios para la definición de la función de pérdida, como

$$\frac{d\tilde{X}}{d\hat{N}} - \tilde{Y} = Res_1 \quad (4.22a)$$

$$\frac{d\tilde{Y}}{d\hat{N}} + \tilde{Y}^2 + h(\hat{N}, \Omega_{m0})\tilde{Y} + k(a, \Omega_{m0}) = Res_2, \quad (4.22b)$$

donde se han utilizado las condiciones iniciales 4.16 para realizar las reparametrizaciones

$$\tilde{X} = X_i + \left(1 - e^{-(\hat{N}_i - \hat{N})}\right) X_N \quad (4.23a)$$

$$\tilde{Y} = Y_i + \left(1 - e^{-(\hat{N}_i - \hat{N})}\right) Y_N. \quad (4.23b)$$

De esta forma, la función de pérdida que debe minimizarse para entrenar una red neuronal que resuelva la ecuación 4.15 es

$$L(\tilde{X}, \tilde{Y}, \hat{N}, \Omega_{m0}) = \sum_{j=1}^M \left(Res_1(\tilde{X}_j, \tilde{Y}_j, \hat{N}_j, \Omega_{m0_j})^2 + Res_2(\tilde{X}_j, \tilde{Y}_j, \hat{N}_j, \Omega_{m0_j})^2 \right) e^{-\omega|\hat{N} - \hat{N}_i|}, \quad (4.24)$$

donde, como se discutió previamente, M es el producto de los batch sizes correspondientes a las variables \hat{N} y Ω_{m0} .

Durante este trabajo, se procedió minimizando el valor medio de la función de pérdida definida en 4.30, es decir, se minimizó la cantidad

$$\langle L(\tilde{X}, \tilde{Y}, \hat{N}, \Omega_{m0}) \rangle = \frac{1}{M} L(\tilde{X}, \tilde{Y}, \hat{N}, \Omega_{m0}). \quad (4.25)$$

La solución del método bundle

Una vez finalizado el entrenamiento, se obtienen dos redes neuronales (una para cada ecuación del sistema de primer orden 4.18), las cuales pueden ser evaluadas para valores dentro del rango de entrenamiento, tanto para \hat{N} , como para Ω_{m0} . Reestableciendo el valor de las reparametrizaciones, y restituyendo los cambios de variables realizados, se obtienen como resultado del entrenamiento las funciones $\delta(a, \Omega_{m0})$ y $\delta'(a, \Omega_{m0})$.

Cabe mencionar que no hay, hasta ahora, un método para estimar el error de la solución obtenida durante el entrenamiento. Por este motivo, se comparó a las soluciones obtenidas de las redes neuronales con las soluciones numéricas obtenidas mediante métodos de integración numéricos tradicionales. En este trabajo se empleó el método Runge-Kutta de orden 5.

4.1.2. Implementación a la ecuación para perturbaciones de materia en el modelo fenomenológico

Ya se ha discutido que, en el modelo fenomenológico estudiado, la modificación de la ecuación diferencial que controla la evolución de la densidad fraccional de perturbaciones de materia consiste en simplemente agregar el factor $\frac{G_{\text{eff}}}{G}(a, g_a)$. Por este motivo, todos los cambios de variables efectuados en Λ CDM se llevan a cabo de manera directa, y de esta forma el sistema a resolver corresponde a

$$\frac{dX}{d\hat{N}} - Y = 0 \quad (4.26a)$$

$$\frac{dY}{d\hat{N}} + Y^2 + h(\hat{N}, \Omega_{m0})Y + k(a, \Omega_{m0})\frac{G_{\text{eff}}}{G}(\hat{N}, g_a) = 0, \quad (4.26b)$$

donde nuevamente

$$h(\hat{N}, \Omega_{m0}) = \left(\frac{n_i(1 + 4\alpha(\frac{e^{\hat{N}n_i}}{a_{\text{eq}}})^3)}{2(1 + \frac{a_{\text{eq}}}{e^{\hat{N}n_i}} + \alpha(\frac{e^{\hat{N}n_i}}{a_{\text{eq}}})^3)} \right), \quad (4.27)$$

y

$$k(\hat{N}, \Omega_{m0}) = -\frac{3e^{\hat{N}n_i}n_i^2}{2a_{\text{eq}}(1 + \frac{e^{\hat{N}n_i}}{a_{\text{eq}}} + \alpha(\frac{e^{\hat{N}n_i}}{a_{\text{eq}}})^4)}. \quad (4.28)$$

Es importante notar que se agregó un nuevo parámetro al bundle: g_a . De esta forma, el procedimiento es el mismo, pero teniendo en cuenta que ahora el vector de parámetros del modelo $\vec{\alpha} = (\Omega_{m0}, g_a)$. Con esta consideración el procedimiento es análogo: se definen los residuos

$$\frac{d\tilde{X}}{d\hat{N}} - \tilde{Y} = Res_1 \quad (4.29a)$$

$$\frac{d\tilde{Y}}{d\hat{N}} + \tilde{Y}^2 + h(\hat{N}, \Omega_{m0})\tilde{Y} + k(a, \Omega_{m0})\frac{G_{\text{eff}}}{G}(\hat{N}, g_a) = Res_2 \quad (4.29b)$$

con la misma reparametrización empleada que en Λ CDM, luego se calcula la función de pérdida definida como

$$L(\tilde{X}, \tilde{Y}, \hat{N}, \Omega_{m0}, g_a) = \sum_{j=1}^M \left(Res_1(\tilde{X}_j, \tilde{Y}_j, \hat{N}_j, \Omega_{m0_j}, g_{a_j})^2 + Res_2(\tilde{X}_j, \tilde{Y}_j, \hat{N}_j, \Omega_{m0_j}, g_{a_j})^2 \right) e^{-\omega|\hat{N} - \hat{N}_j|}, \quad (4.30)$$

donde ahora el valor de M corresponde al producto de los batch sizes de las cantidades \hat{N} , Ω_{m0} y g_a , y finalmente se minimiza la cantidad

$$\langle L(\tilde{X}, \tilde{Y}, \hat{N}, \Omega_{m0}, g_a) \rangle = \frac{1}{M} L(\tilde{X}, \tilde{Y}, \hat{N}, \Omega_{m0}, g_a). \quad (4.31)$$

El cambio puede parecer sencillo de implementar, pero el costo está en agregar un parámetro más al bundle, ya que esto multiplica al número de cálculos que debe realizar la red neuronal en cada iteración del entrenamiento por un factor igual al batch size asociado al nuevo parámetro incluido al modelo.

Al igual que para Λ CDM, finalizado el entrenamiento, reestablecida la reparametrización y reestituidos los cambios de variables, se obtienen como resultado las funciones $\delta(a, \Omega_{m0}, g_a)$ y $\delta'(a, \Omega_{m0}, g_a)$.

Estas funciones de salida se emplean para construir al observable $f\sigma_8$ y realizar de esta manera un análisis estadístico a través de cadenas de Markov.

4.2. Análisis estadístico con cadenas de Markov Monte-Carlo (MCMC) usando las soluciones obtenidas con el método de redes neuronales

Como se ha discutido previamente, con las soluciones obtenidas a las ecuaciones diferenciales estudiadas ($\delta(a, \Omega_{m0}, g_a)$ y $\delta'(a, \Omega_{m0}, g_a)$) se puede construir el observable $f\sigma_8$. El objetivo de este trabajo es, no solo implementar el entrenamiento a estas ecuaciones, sino que además incluye utilizar estas soluciones para realizar un análisis estadístico que permita testear modelos cosmológicos y establecer límites sobre los parámetros libres de los mismos. En esta tesis se llama \mathcal{D} a los datos medidos y \vec{p} a los parámetros libres del modelo, los cuales no necesariamente están todos en el bundle de las redes neuronales, como se discute a lo largo de esta sección.

En el trabajo se empleó un enfoque Bayesiano para realizar la estimación de los parámetros cosmológicos. Desde la perspectiva bayesiana, no hay una distinción fundamental entre las mediciones y parámetros de un modelo estadístico: todos son considerados cantidades “desconocidas”, y se les puede asociar un concepto de probabilidad [87]. El objeto de la inferencia bayesiana es la distribución a posteriori (o posterior), la cual es una distribución de probabilidad conjunta que representa la probabilidad de obtener los parámetros \vec{p} dado el conjunto de datos \mathcal{D} . En efecto, si la distribución de probabilidad conjunta es $P(\mathcal{D}, \vec{p})$, esta distribución consta de dos partes: una distribución a priori $P(\vec{p})$ y una función de verosimilitud (o likelihood), la cual suele notarse como $P(\mathcal{D}|\vec{p}) = \mathcal{L}$. La función de verosimilitud expresa, dado un problema físico descrito por los parámetros \vec{p} , que tan probable es obtener el conjunto de datos \mathcal{D} . Es una función de los parámetros y no tiene las propiedades de una distribución de probabilidad. Habiendo medido \mathcal{D} , se usa el teorema de Bayes para obtener la distribución de \vec{p} condicionada a los datos

$$P(\vec{p}|\mathcal{D}) = \frac{P(\mathcal{D}|\vec{p})P(\vec{p})}{P(\mathcal{D})}, \quad (4.32)$$

donde $P(\mathcal{D}|\vec{p})$ es la mencionada likelihood, $P(\vec{p})$ es la probabilidad que se le asigna a priori al valor de los parámetros llamada distribución a priori (o prior) y $P(\mathcal{D})$ es la probabilidad de los datos, la cual en general, dada la likelihood y los priors, se calcula como un factor de

normalización. Esta distribución de probabilidad es la distribución a posteriori (o posterior), y representa la distribución de probabilidad de que el valor verdadero de los parámetros del modelo sea \vec{p} dados los datos provistos, es decir, constituye la probabilidad de que los parámetros tengan un cierto valor a partir de un conjunto de datos. Cabe mencionar que las distribuciones a priori, o priors, ($P(\vec{p})$) juegan un rol importante en la exploración del espacio de parámetros ya que restringen los valores posibles [88]. Elegir esta distribución constituye una decisión del analista de los datos en el proceso de inferencia. Durante este trabajo se emplearon priors uniformes para todos los parámetros, es decir, se eligió un intervalo, y se asumió que, dentro de ese intervalo, la distribución a priori tiene un valor constante, mientras que fuera de dicho intervalo esta probabilidad vale cero. La elección de las distribuciones a priori estuvo ligada al intervalo de entrenamiento de las redes para los parámetros del bundle, dado que estas deben poder evaluarse para cualquier valor de los parámetros dentro de los priors elegidos. Además, la integral que aparece en el denominador de la ecuación 4.32 se interpreta como la probabilidad de haber medido los datos ($P(\mathcal{D})$), pero no suele calcularse ya que al conocer los priors y la likelihood esta constituye un factor de normalización que garantiza que

$$\int \vec{d}p P(\vec{p}|\mathcal{D}) = 1. \quad (4.33)$$

En este capítulo ya se han mencionado parámetros cosmológicos que se incluyen en el bundle de las redes neuronales, no obstante, puede haber parámetros del modelo que no jueguen ningún rol en las ecuaciones diferenciales, como por ejemplo σ_8 . Este parámetro cosmológico se infiere a través del análisis estadístico, y se ve involucrado en el cálculo del observable $f\sigma_8$, pero no es un parámetro del bundle de las redes neuronales ya que no aparece en la ecuación diferencial del modelo. Hecho este comentario, se utilizaron las salidas de las redes para calcular el observable según la ecuación 2.70

$$f\sigma_8(a) = a \frac{\sigma_8}{\delta_m(1)} \delta'_m(a), \quad (4.34)$$

y de esta manera llevar a cabo la inferencia de parámetros mediante las cadenas de Markov [89].

En general, el proceso de inferencia con el método MCMC consiste en explorar el espacio de parámetros para samplear la distribución posterior, empleando un criterio basado en maximizar la likelihood [90]. Esta función se relaciona con una cantidad importante en la estadística llamada χ^2 mediante

$$\mathcal{L} = e^{-\chi^2}, \quad (4.35)$$

donde es importante notar que si se minimiza la cantidad χ^2 , se maximiza la función de verosimilitud. Luego, la cantidad χ^2 se calcula como

$$\chi^2 = \sum_{jk} V_j^t C_{jk}^{-1} V_k, \quad (4.36)$$

donde C_{ij} es la matriz de covariancia que incluye los errores individuales de los datos en su

diagonal y las correlaciones entre los mismos en los términos fuera de la diagonal, y el vector V_i se calcula como [2]

$$V^j(\mathcal{D}, \vec{p}) = f\sigma_8^j - \text{cociente}(\mathcal{D}^j, \vec{p}^j) f_{s8}^{Teo}(\mathcal{D}^j, \vec{p}^j). \quad (4.37)$$

Se puede notar que este vector cuantifica la diferencia entre las mediciones del observable ($f\sigma_8^j$) y el cálculo teórico del mismo, en función de los datos y de los parámetros considerados ($f_{s8}^{Teo}(\mathcal{D}^j, \vec{p}^j)$). Esta diferencia se calcula teniendo en consideración un cociente calculado como

$$\text{cociente}(\mathcal{D}^j, \vec{p}^j) = \frac{H(\mathcal{D}^j, \vec{p}^j) d_L(\mathcal{D}^j, \vec{p}^j)}{H^{\text{fid}}(\mathcal{D}^j) d_L^{\text{fid}}(\mathcal{D}^j)}, \quad (4.38)$$

donde $H(\mathcal{D}^j)$ es el parámetros de Hubble, $d_L(\mathcal{D}^j)$ es la distancia luminosa, mientras que $^{\text{fid}}$ significa fiducial e implica evaluar estas funciones en los parámetros fiduciales, que dependen del experimento realizado para obtener los datos. Esta cantidad se discute con más detalle en el siguiente capítulo, pero de forma sencilla es una corrección que se introduce dado que para realizar las mediciones se debe asumir una cosmología fiducial (se observa que si los parámetros \vec{p} son iguales a los parámetros asumidos en la medición, el cociente da 1 y no hay corrección). Es directo notar que, dadas las propiedades de la matriz C_{ij} , la cantidad χ^2 es una cantidad positiva.

El proceso denominado Markov Chain Monte Carlo (MCMC) consiste en generar una caminata al azar de las cadenas sobre el espacio de parámetros, donde cada paso consiste en un eslabón de la cadena. De esta forma se genera una secuencia de variables aleatorias $\{X_0, X_1, X_2, \dots\}$ de forma tal que para $j > 0$ el siguiente estado X_{j+1} se muestree de una distribución $\tilde{P}(X_{j+1}|X_j)$ que depende solo del estado actual de la cadena X_j , esta secuencia se llama cadena de Markov y $\tilde{P}(X_{j+1}|X_j)$ es la probabilidad de transición de la cadena. En estas condiciones, la cadena perderá información del estado inicial X_0 mientras evoluciona, y la probabilidad de transición de la cual se muestrean los valores tenderá a una distribución estacionaria $\phi(\cdot)$. El algoritmo de Metropolis-Hastings es el que garantiza que esta distribución estacionaria sea, en efecto, la distribución a posteriori que se desea obtener [87].

La implementación de este algoritmo consiste en que, en cada paso j , el siguiente estado X_{j+1} se elige primero muestreando un punto candidato Y de una distribución propuesta $q(\cdot|X_j)$, que por ejemplo, puede ser una distribución normal de variables múltiples, con valor medio X_j y una matriz de covarianza fija. Para el punto candidato Y se calcula una probabilidad $\alpha(X_j, Y)$ tal que

$$\alpha(X_j, Y) = \min\left(1, \frac{P(Y)q(X_j|Y)}{P(X_j)q(Y|X_j)}\right). \quad (4.39)$$

A la vez se elige un número aleatorio entre 0 y 1 y se compara dicho número con $\alpha(X_j, Y)$. Si el número es menor a $\alpha(X_j, Y)$ se toma $X_{j+1} = Y$, sino se toma $X_{j+1} = X_j$ luego se incrementa j y se continua. Cabe destacar que $q(\cdot|X_j)$ puede tener cualquier forma, y en cualquier caso la distribución estacionaria de la cadena será la distribución posterior $P(\vec{p}|\mathcal{D})$. Los valores de la cadena que se utilizan para construir la posterior se toman luego de m iteraciones de "burn-in",

donde se asume que la cadena ya perdió información del estado inicial. Entonces los puntos $\{X_j; j = m + 1, \dots, n\}$ se extraeran de la distribución estacionaria $\phi(\cdot)$ para representar la distribución a posteriori, donde se excluyeron puntos que aún contienen información sobre los valores iniciales de las cadenas.

El problema fundamental de la inferencia a partir de simulaciones de cadenas de Markov es que siempre habrá regiones del espacio de parámetros que no serán cubiertas por una cadena de longitud finita. A medida que la simulación avanza, la propiedad ergódica de la cadena de Markov hace que eventualmente se cubra toda la región, pero en el corto plazo, las simulaciones no pueden decir nada acerca de regiones que la cadena aún no haya explorado. Esto es un problema cuando la convergencia es lenta. La mejor manera de notar la convergencia lenta es examinando múltiples simulaciones independientes. También es importante que los puntos iniciales de cada una de las cadenas estén suficientemente separados [87]. En este trabajo se empleó un criterio de convergencia para garantizar que las cadenas hayan convergido, de forma tal que la inferencia de los parámetros sea exitosa [91].

De esta forma, mediante el MCMC, se obtienen las distribuciones a posteriori para los parámetros inferidos. Para obtener una estimación de cada parámetro, se buscó un intervalo dentro del cual haya un 68 % de probabilidades de hallar su valor. Para ello, lo primero que se debe realizar es marginalizar la posterior obtenida sobre todos los parámetros que no corresponden al parámetro que se desea inferir. Por ejemplo, si de todos los parámetros del modelo se desea inferir el intervalo de confianza del parámetro p_j , la distribución marginalizada se calcula como la integral de la distribución a posteriori sobre todos los valores de los otros parámetros p_k con $k \neq j$. Es decir, la distribución marginalizada para p_j se calcula como

$$P^{\text{marg}}(p_j|\mathcal{D}) = \int \prod_{k \neq j} dp_k P(\vec{p}|\mathcal{D}). \quad (4.40)$$

Esta distribución marginalizada cumple las propiedades de una distribución de probabilidades para el parámetro p_j . Lo siguiente es imponer un criterio para definir un intervalo de confianza. En este trabajo se definió como intervalo de confianza aquel rango dentro del cual hay un 68 % de probabilidad de que se encuentre el valor del parámetro.

Entonces, se buscan los valores p_{j1} y p_{j2} tales que

$$\int_{p_{j1}}^{p_{j2}} dp_j P^{\text{marg}}(p_j|\mathcal{D}) = 0,68, \quad (4.41)$$

y se define el intervalo de confianza para el parámetro p_j como $[p_{j1}, p_{j2}]$. Esto se realiza para todos los parámetros del modelo, marginalizando siempre sobre todos los otros.

Con este método es posible aplicar el método de las redes neuronales a la inferencia de parámetros cosmológicos. Nuevamente, resulta importante destacar que al explorar el espacio de parámetros, en cada paso de las cadenas, estas deben computar un valor del observable $f\sigma_8$. Al emplear métodos numéricos tradicionales, esto implica integrar las ecuaciones en cada paso del análisis estadístico, mientras que para las redes neuronales, el cálculo del observable requiere únicamente evaluar la función obtenida mediante el entrenamiento. Esto último puede repre-

sentar una ventaja en términos de tiempos computacionales del método de las redes neuronales sobre los métodos numéricos tradicionales.

En este trabajo se empleó el MCMC para obtener las distribuciones a posteriori de los parámetros cosmológicos de cada modelo empleando tanto las soluciones numéricas como las soluciones de las redes neuronales. Como se mencionó previamente, en todos los casos se emplearon priors uniformes, por lo que para los parámetros relevantes de los modelos estudiados se tomaron intervalos dentro de los cuales se asume que, a priori, todos los valores tienen la misma probabilidad. Estos intervalos corresponden a

$$I_{\Omega_{m0}} = [0,1; 0,7], \quad I_{\sigma_8} = [0,6; 1,3], \quad I_{g_a} = [-1,0; -0,7]. \quad (4.42)$$

Capítulo 5

Datos

Hasta este momento se han detallado los modelos teóricos y se ha discutido sobre los métodos de inteligencia artificial empleados durante el trabajo. Además, se comentaron implementaciones del método de las redes neuronales a los modelos teóricos, y al análisis estadístico para inferir parámetros cosmológicos. Para realizar dicha inferencia, se discutió el método de las cadenas de Markov. En dicha discusión, se comentó que es necesario disponer de un juego de datos para poder obtener las distribuciones a posteriori de los parámetros cosmológicos. En este capítulo se describen los datos observacionales que se utilizan para contrastar con las predicciones teóricas de los modelos estudiados en esta tesis. Es decir, se discute la física y el proceso de medición del observable $f\sigma_8$ detallado en la ecuación 2.70. Finalmente, se discute de manera más precisa la ec. 5.10 empleada para calcular la cantidad χ^2 durante el proceso del MCMC.

5.1. Detalles observacionales de $f\sigma_8$

El observable empleado para inferir los parámetros y testear modelos cosmológicos en este trabajo corresponde a $f\sigma_8$. La expresión teórica de dicha cantidad es (ec. 2.70)

$$f\sigma_8(a, \vec{x}) = f(a, \vec{x})\sigma(a, \vec{x}) \quad (5.1)$$

La cantidad $f\sigma_8$ es una medida fundamental en cosmología que captura la evolución de las estructuras cósmicas a lo largo del tiempo. La tasa de crecimiento

$$f(a, \vec{x}) = \frac{d \ln \delta_m(a, \vec{x})}{d \ln a} \quad (5.2)$$

describe cómo las perturbaciones de la densidad de materia en el Universo cambian con el tiempo, indicando la capacidad de la gravedad para atraer la materia y formar estructuras. Por otro lado,

$$\sigma_8(a, \vec{x}) = \sigma_8 \frac{\delta(a, \vec{x})}{\delta(1, \vec{x})} \quad (5.3)$$

mide la dispersión de la densidad de masa en regiones esféricas de 8 megaparsecs, representando la amplitud de las fluctuaciones de densidad a gran escala. Multiplicar f por σ_8 normaliza

la tasa de crecimiento, centrando la atención en la evolución relativa de las perturbaciones.

Se mostró, en el desarrollo 2.70, que, en los casos en que puede factorizarse $\hat{\delta}(a, \vec{k}) = F(\vec{k})\delta(a)$, la expresión del observable no depende del modo de Fourier y corresponde a

$$f\sigma_8(a) = a \frac{\sigma_8}{\delta_m(1)} \delta'_m(a). \quad (5.4)$$

La determinación de $f\sigma_8$ está relacionada con el análisis de las distorsiones en el espacio de redshift (ec. 2.3). Para ella, se recopilan catálogos de galaxias con posiciones y redshifts, y se calcula la función de correlación de dos puntos para entender la distribución espacial y correlaciones. El análisis de redshift espacial permite entender cómo las velocidades peculiares afectan las posiciones aparentes, generando distorsiones en la función de correlación.

Las distorsiones del espacio de corrimiento al rojo (RSD) son una herramienta de observación clave para estudiar la naturaleza de la energía oscura, ya que trazan el campo de velocidad de la materia a través de las velocidades peculiares de las galaxias. Permiten medir la tasa de crecimiento de la estructura a través de un aumento de la agrupación de materia a lo largo de la línea de visión [92]. Por este motivo las RSD permiten discriminar entre los diferentes modelos físicos de energía oscura, ya que los modelos que comparten la misma historia de expansión a menudo predicen diferentes valores de la estructura f (e.g. [93]). En la actualidad se han desarrollado dos métodos de medición [94]: uno basado en métodos estadísticos y otro que emplea las velocidades peculiares, las cuales son directamente proporcionales a la tasa de crecimiento f [95,96], y por lo tanto constituyen un punto clave en la medición del observable. En el método estadístico, dado un catálogo de galaxias a un determinado redshift se utiliza que la función de correlación en el espacio de corrimientos al rojo depende de $\beta = f/b$, donde b es un parámetro de sesgo de cada galaxia [97] y f es la ya mencionada tasa de crecimiento. En escalas grandes se asume que $b = \sigma_{8g}/\sigma_8$, donde σ_8 es el parámetro mencionado previamente en este trabajo y σ_{8g} se extrae directamente del catálogo de galaxias. Entonces, se pueden combinar los observables para obtener

$$f\sigma_8 = \beta\sigma_{8g}. \quad (5.5)$$

El segundo método se basa en medir directamente las velocidades peculiares midiendo la distancia de las galaxias a través de candelas estándar y comparando esa distancia con su redshift. Las velocidades medidas se pueden comparar con las que se predicen a través de un campo de densidad de galaxias $\delta_g(\vec{x})$ derivado de un catálogo de redshifts independiente, ya que se sabe que las velocidades peculiares se relacionan con esta cantidad como [98]

$$v(\vec{x}) = \frac{H_0}{4\pi} \frac{f}{b} \int_0^\infty d^3r' \delta_g(\vec{x}') \frac{\vec{x}' - \vec{x}}{|\vec{x}' - \vec{x}|^3}, \quad (5.6)$$

esta comparación permite obtener β para luego obtener un valor del observable $f\sigma_8$ de manera análoga a la mencionada para el primer método.

Los valores observacionales empleados en el trabajo se pueden ver en la tabla 5.1, la cual corresponde a una recopilación realizada sobre un conjunto de datos más grande, pero que tenían

Tabla 5.1: Datos observacionales de la cantidad $f\sigma_8$. Las columnas corresponden a: un índice, el nombre del experimento en cuestión, el redshift para el cual se mide, el valor reportado del observable junto con su error, la referencia y el valor fiducial de los parámetros adoptados para la medición.

Index	Set de datos	z	$f\sigma_8(z)$	Refs.	Params. fid.
1	6dFGS+SnIa	0.02	0.428 ± 0.0465	[99]	$(\Omega_m, h, \sigma_8) = (0.3, 0.683, 0.8)$
2	SnIa+IRAS	0.02	0.398 ± 0.065	[100], [101]	$(\Omega_m, \Omega_K) = (0.3, 0)$
3	2MASS	0.02	0.314 ± 0.048	[102], [101]	$(\Omega_m, \Omega_K) = (0.266, 0)$
4	SDSS-veloc	0.10	0.370 ± 0.130	[103]	$(\Omega_m, \Omega_K) = (0.3, 0)$
5	SDSS-MGS	0.15	0.490 ± 0.145	[104]	$(\Omega_m, h, \sigma_8) = (0.31, 0.67, 0.83)$
6	2dFGRS	0.17	0.510 ± 0.060	[105]	$(\Omega_m, \Omega_K) = (0.3, 0)$
7	GAMA	0.18	0.360 ± 0.090	[106]	$(\Omega_m, \Omega_K) = (0.27, 0)$
8	GAMA	0.38	0.440 ± 0.060	[106]	
9	SDSS-LRG-200	0.25	0.3512 ± 0.0583	[107]	$(\Omega_m, \Omega_K) = (0.25, 0)$
10	SDSS-LRG-200	0.37	0.4602 ± 0.0378	[107]	
11	BOSS-LOWZ	0.32	0.384 ± 0.095	[108]	$(\Omega_m, \Omega_K) = (0.274, 0)$
12	SDSS-CMASS	0.59	0.488 ± 0.060	[109]	$(\Omega_m, h, \sigma_8) = (0.307115, 0.6777, 0.8288)$
13	WiggleZ	0.44	0.413 ± 0.080	[110]	$(\Omega_m, h) = (0.27, 0.71)$
14	WiggleZ	0.60	0.390 ± 0.063	[110]	
15	WiggleZ	0.73	0.437 ± 0.072	[110]	
16	Vipers PDR-2	0.60	0.550 ± 0.120	[111]	$(\Omega_m, \Omega_b) = (0.3, 0.045)$
17	Vipers PDR-2	0.86	0.400 ± 0.110	[111]	
18	FastSound	1.40	0.482 ± 0.116	[112]	$(\Omega_m, \Omega_K) = (0.270, 0)$

una mayor correlación entre sí. Dentro del conjunto de datos empleados, los únicos tres resultados que están correlacionados corresponden al experimento WiggleZ, cuya matriz de correlación es conocida [2]. En la tabla además pueden verse los valores fiduciales de los parámetros cosmológicos empleados para realizar su medición, ya que para realizar estas mediciones es necesario asumir una cosmología. En general, se asume Λ CDM y algún valor fiducial para los parámetros que conforman dicho modelo. Esta cosmología asumida resulta necesaria dado que, durante el proceso de medición es necesario, por un lado, convertir redshift a distancia luminosa, mientras que también es necesario asumir un valor de H_0 , lo cual se ve explícitamente en la ecuación 5.6. El valor fiducial relevante para este trabajo corresponde a Ω_{m0} , cuya relevancia se discute en la siguiente sección.

5.2. Detalles de implementación de los datos

En la sección anterior se presentó la tabla de datos empleada para realizar la inferencia de parámetros. En esta sección se discuten detalles pertinentes al uso de dicha tabla. Cuando se discutió el proceso de inferencia a través del MCMC, se mencionó que, una cantidad relevante durante la determinación de parámetros corresponde al

$$\chi^2 = \sum_{jk} V_j^t C_{jk}^{-1} V_k, \quad (5.7)$$

donde V_i es un vector que mide la diferencia entre el valor teórico de la cantidad $f\sigma_8$ y su valor experimental, pesado por el cociente; mientras que C_{jk} corresponde a la matriz de covarianza de los datos. En las siguientes subsecciones se comentan detalles técnicos sobre las correcciones fiduciales y sobre la correlación entre los datos, ambos puntos claves para realizar el cálculo de χ^2 y de esta forma llevar a cabo la inferencia de parámetros a través del MCMC.

5.2.1. Correcciones fiduciales para el cálculo de χ^2

La cantidad V_j se define como [2]

$$V^j = f\sigma_8^j - \text{cociente}(z^j, \Omega_{m0}) f_{s8}^{\text{Teo}}(z^j, \Omega_{m0}), \quad (5.8)$$

donde la etiqueta j refiere al número de índice del dato de la tabla 5.1, Ω_{m0} es el valor de dicha cantidad empleada para el cálculo y evaluar $f_{s8}^{\text{Teo}}(z^j, \Omega_{m0})$ es directo usando la expresión 5.4 y que

$$a^j = \frac{1}{1 + z^j}. \quad (5.9)$$

Por otro lado, el cociente es un factor de corrección que se incluye debido a la cosmología asumida durante la determinación del observable. Es decir, esta cantidad corrige el valor del χ^2 teniendo en cuenta que el valor del parámetro Ω_{m0} incluido en el cálculo es distinto al valor fiducial asumido para obtener cada dato de la tabla 5.1. Dicho valor se calcula como

$$\text{cociente}(z^j, \Omega_{m0}) = \frac{H(z^j, \Omega_{m0}, H_0) d_L(z^j, \Omega_{m0}, H_0)}{H^{\text{fid}}(z^j, H_0^{\text{fid}}) d_L^{\text{fid}}(z^j, H_0^{\text{fid}})}, \quad (5.10)$$

donde H y d_L son el parámetro de Hubble y la distancia luminosa, y $^{\text{fid}}$ refiere a evaluar estas funciones en el valor de Ω_{m0}^{fid} que se ve en la tabla 5.1. Además, se hace explícita la dependencia de las funciones en H_0 y H_0^{fid} , pero es importante remarcar que debido a que $H \propto H_0$ y $d_L \propto H_0^{-1}$, el cociente no depende de dicho valor. Asumir una cosmología para realizar las mediciones resulta inevitable debido a que, como se comentó al principio de este capítulo, en la medición de $f\sigma_8$ se ven involucradas mediciones de velocidad y distancia, lo que explica que el cociente dependa de H y d_L . Es importante mencionar que d_L depende explícitamente de H_0 , por lo que se puede asumir que al fijar una cosmología fiducial se calcula primero H_0 , y luego con este valor sumado al modelo asumido se obtiene d_L . Es directo notar que si ocurre que Ω_{m0} es igual al valor fiducial

de Ω_{m0} correspondiente al dato j -ésimo de la tabla, entonces el valor del cociente es 1 y no hay corrección.

Durante la exploración del espacio de parámetros se utilizan distintos valores de Ω_{m0} , y debe quedar claro que el cociente es un vector que asigna un valor a cada dato de la tabla 5.1, pero que cambia estos valores según el valor asumido de Ω_{m0} .

Por otro lado, en el cálculo de χ^2 se ve involucrada además la matriz de covarianza de los datos, la cual no es trivial al considerar correlación entre los datos.

5.2.2. Correlación entre los datos

Al comienzo de este capítulo se detalló la expresión para efectuar el cálculo de la cantidad χ^2 (ec. 5.7) y se comentó que en dicho cálculo se ve involucrada la matriz de covarianza. Al considerar datos que no están correlacionados, la matriz de covarianza es

$$C_{jk} = \text{diag}(\sigma_1^{-1}, \sigma_2^{-1}, \dots, \sigma_N^{-1}), \quad (5.11)$$

donde N es el número de datos disponible, y σ_j es el error asociado al dato j -ésimo. En la tabla 5.1 están detallados los errores en la medición de la cantidad $f\sigma_8$ para cada experimento. Debido al procedimiento de medición, resulta necesario considerar que hay correlación entre los datos correspondientes al experimento "WiggleZ" (datos número 13, 14 y 15) [2]. Que los datos de distintos experimentos estén correlacionados implica que los valores reportados de $f\sigma_8$, y los errores, no son independientes entre sí. En estos casos, la matriz de covarianza no es diagonal y debe ser reportada junto con los datos. En este caso en particular, la matriz de covarianza para expresar la correlación entre estos tres datos es [2]

$$C_{jk}^{\text{WiggleZ}} = 10^{-3} \begin{pmatrix} 6,400 & 2,570 & 0,000 \\ 2,570 & 3,969 & 2,540 \\ 0,000 & 2,540 & 5,184 \end{pmatrix}. \quad (5.12)$$

Como el resto de los datos no están correlacionados entre sí, ni tampoco con los del experimento WiggleZ, la matriz de covarianza total para calcular el χ^2 es

$$C_{jk} = \begin{pmatrix} \sigma_1^{-1} & 0 & \dots & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2^{-1} & \dots & 0 & 0 & 0 & 0 & 0 \\ \dots & \dots & \dots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{ij}^{\text{WiggleZ}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{16}^{-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{17}^{-1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{18}^{-1} & 0 \end{pmatrix}. \quad (5.13)$$

Esta es la expresión que debe utilizarse para calcular la cantidad χ^2 mediante la ecuación 5.7. Un último comentario corresponde al hecho de que en el cálculo de dicha cantidad se necesita invertir esta matriz, ya que en la expresión aparece C_{jk}^{-1} . Todo lo detallado en este capítulo sirve para calcular la función de verosimilitud necesaria para realizar el MCMC como

$$\mathcal{L} = e^{-\chi^2}. \quad (5.14)$$

Como se mencionó en el capítulo anterior, esta cantidad cuantifica, luego de haber realizado una medición, que tan probable era obtener ese resultado dados los parámetros del modelo cosmológico que se está asumiendo. Es una función de los parámetros y no representa una distribución de probabilidad, dado que no está normalizada. A pesar de esto, esta función contiene mucha información, ya que el juego de parámetros que la haga alcanzar su máximo valor corresponde a aquel que constituye una mayor probabilidad de haber medido los datos que se están estudiando. En el siguiente capítulo se presentan los resultados obtenidos, dentro de los cuales jugó un rol importante el proceso de estimación de parámetros cosmológicos a través de maximizar la likelihood (o minimizar el χ^2), dado que esta estimación a priori fue relevante para imponer los priors en el método del MCMC.

Es importante aclarar que el conjunto de datos empleado (tabla 5.1) no corresponde al más actual. En este contexto, el grupo de trabajo de la directora de esta tesis está trabajando en una actualización de los datos, cuyo empleo a futuro puede resultar prometedor.

Capítulo 6

Resultados

El objetivo de esta tesis fue resolver las ecuaciones dinámicas para la evolución de las perturbaciones de materia en el Universo mediante el método de las redes neuronales, y utilizar estas soluciones para realizar estimaciones de parámetros cosmológicos. Esto se realizó para dos modelos diferentes: el modelo estándar de la cosmología Λ CDM y un modelo fenomenológico de gravedad modificada. En este capítulo se muestran los resultados obtenidos, donde resulta importante mencionar que antes de poder estudiar las soluciones de las redes neuronales fue necesario garantizar una solución numérica precisa. Esta solución resulta importante ya que el método de las redes neuronales no dispone aún un mecanismo propio para estimar el error cometido en las soluciones. Debido a esto, la forma de estimar el error empleada en el trabajo consistió en calcular el error porcentual comparando con la solución numérica como

$$err \% (a_j) = \frac{|\delta_{NN}(a_j) - \delta_{Num}(a_j)|}{|\delta_{Num}(a_j)|} * 100, \quad (6.1)$$

donde es necesario observar que dicho error está evaluado en cada punto a_j de la integración numérica. Por este motivo, el primer paso para poder garantizar un entrenamiento adecuado de las redes requiere de una solución numérica que permita realizar una comparación certera, la cual se discute a continuación.

6.1. Soluciones numéricas

Desarrollar integradores numéricos no fue trabajo de esta tesis. En este caso, se empleó como integrador numérico un algoritmo de Runge-Kutta de orden 5, implementado por medio de la librería Scipy a través del método "solve_ivp".

6.1.1. La solución numérica para Λ CDM

Como se ha discutido anteriormente, para realizar la integración numérica de la ecuación debe fijarse el valor de todos los parámetros del modelo. En la figura 6.1 se puede ver un ejemplo de la solución obtenida donde se fijó $\Omega_{m0} = 0,272$.

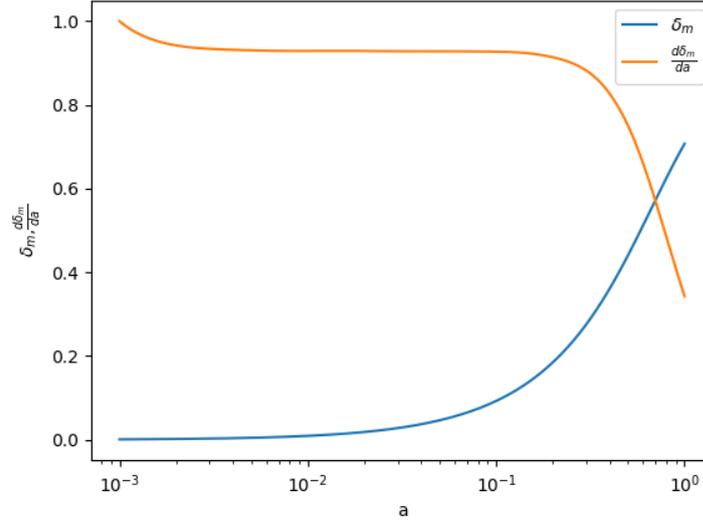


Figura 6.1: Solución numérica para δ y δ' en función del factor de escala a para el modelo Λ CDM con $\Omega_{m0} = 0,272$.

Esta solución numérica puede compararse con la aproximación analítica discutida en el capítulo 2, donde se ha detallado que, durante el período dominado por materia, δ_m tiene un comportamiento lineal respecto del factor de escala. En la figura 6.2 se puede ver dicha comparación en una escala que permite apreciar la concordancia con la solución analítica aproximada, la cual permite ganar confianza en la solución numérica obtenida. Es importante remarcar que la solución se aproxima a la lineal al comienzo del período de integración y luego se separa, debido a que a medida que crece el factor de escala la contribución de la constante cosmológica comienza a cobrar relevancia.

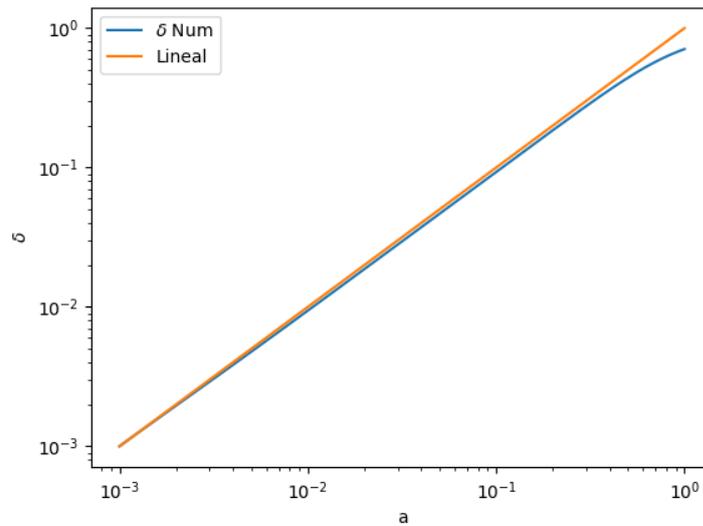


Figura 6.2: Solución numérica para δ comparada con una función lineal en función del factor de escala a para el modelo Λ CDM con $\Omega_{m0} = 0,272$.

Luego se emplearon los valores numéricos de δ y δ' para calcular el observable $f\sigma_8$, utilizando

la ecuación 2.70 y se comparará gráficamente esta estimación numérica de $f\sigma_8$ con los datos observacionales. Un gráfico que contiene el observable calculado numericamente para $\Omega_{m0} = 0,272$ y $\sigma_8 = 0,8$, junto con los datos de la tabla 5.1 se puede ver en la figura 6.3.

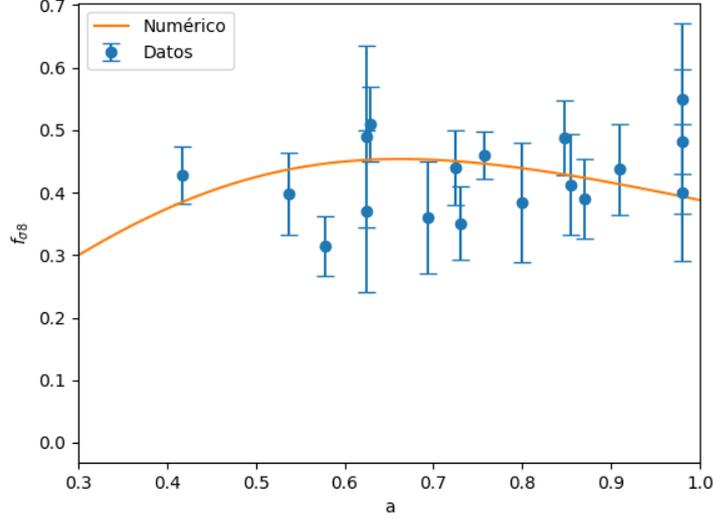


Figura 6.3: $f\sigma_8$ vs a obtenido con el método numérico en función del factor de escala, junto con los datos de la tabla 5.1. Los parámetros cosmológicos empleados para el cálculo corresponden a $\Omega_{m0} = 0,272$ y $\sigma_8 = 0,8$.

En la literatura estos gráficos suelen realizarse en función del corrimiento al rojo, definido como

$$z = \frac{1}{a} - 1, \quad (6.2)$$

motivo por el cual se realizó la transformación correspondiente para obtener la figura 6.4, la cual puede compararse directamente con la literatura [2].

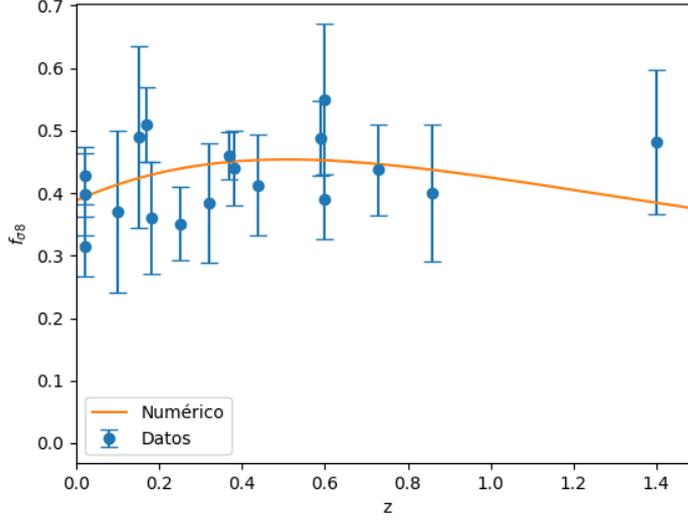


Figura 6.4: $f\sigma_8$ vs z obtenido con el método numérico en función del corrimiento al rojo, junto con los datos de la tabla 5.1. Los parámetros cosmológicos empleados para el cálculo corresponden a $\Omega_{m0} = 0,272$ y $\sigma_8 = 0,8$. Esta figura puede compararse directamente con la literatura [2].

Dada la comparación realizada con la solución analítica aproximada, se concluyó que, para el modelo Λ CDM, las soluciones numéricas son razonables para comparar a las soluciones obtenidas mediante el método de las redes neuronales. Las comparaciones visuales con figuras de la literatura también se utilizaron como chequeo.

6.1.2. La solución numérica para el modelo fenomenológico

Para el modelo fenomenológico se procedió de la misma forma que en la sección anterior. En primer lugar, se obtuvieron las soluciones para δ y δ' , chequeando que en el caso $g_a = 0$ se obtuvieron las mismas soluciones que para Λ CDM. En la figura 6.5 se pueden ver las soluciones obtenidas para los valores de los parámetros reportados en [2] ($\Omega_{m0} = 0,21$, $\sigma_8 = 0,88$ y $g_a = -1,16$).

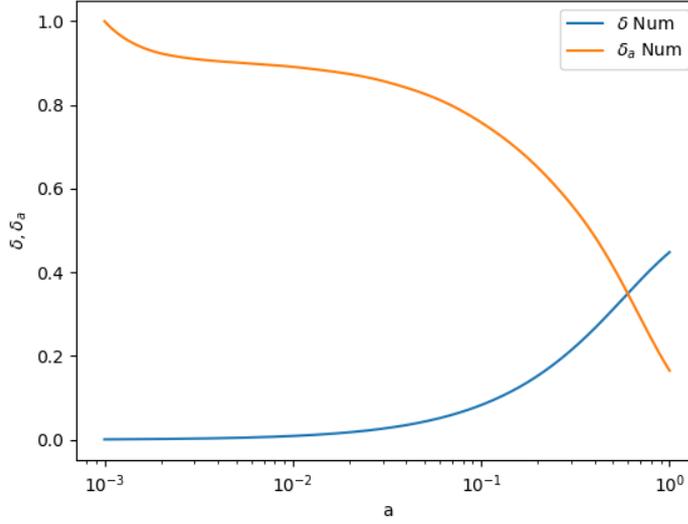


Figura 6.5: Solución numérica para δ y δ' en función del factor de escala a para el modelo fenomenológico de gravedad modificada. Los parámetros corresponden a $\Omega_{m0} = 0,21$, $\sigma_8 = 0,88$ y $g_a = -1,16$.

Con estas soluciones se calculó el observable para los valores $\Omega_{m0} = 0,21$, $\sigma_8 = 0,88$ y $g_a = -1,16$ [2], obteniendo el resultado de la figura 6.6.

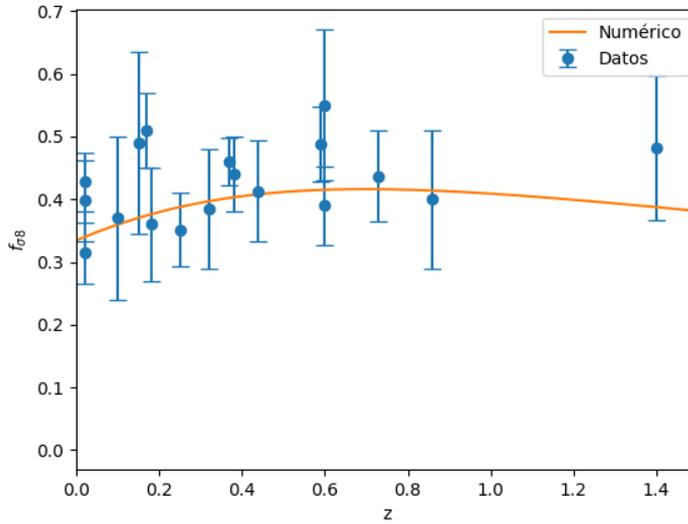


Figura 6.6: $f\sigma_8$ vs z para el modelo fenomenológico de gravedad modificada, junto con los datos de la tabla 5.1. Esta figura puede compararse directamente con la literatura [2]. Los parámetros corresponden a $\Omega_{m0} = 0,21$, $\sigma_8 = 0,88$ y $g_a = -1,16$.

Al igual que para el modelo Λ CDM, se concluyó que las soluciones obtenidas numéricamente están en condiciones de utilizarse como referencia para el método de las redes neuronales.

6.2. Entrenamientos sin bundle para Λ CDM

Una vez obtenida una solución numérica confiable, se realizó el entrenamiento más sencillo de las redes neuronales: el caso sin bundle. Como se discutió en el capítulo 3, este caso corresponde al entrenamiento de redes neuronales que aprenden a resolver la ecuación diferencial 4.15 para un valor fijo del parámetro del modelo. Para probar distintas alternativas, se entrenaron redes empleando dos funciones de activación: la tangente hiperbólica, y la función SiLU. Además, se emplearon dos arquitecturas distintas: la habitual, y la customizada. En la figura 6.7 se pueden ver las funciones de pérdida (loss) correspondientes a estas cuatro redes neuronales, donde se utilizó un batch size de 32, un learning rate de 10^{-3} y una estructura interna de tres capas de (256,128,64) neuronas artificiales cada una con conexión completa. Además, luego de probar varios valores se utilizó un valor de $\omega = 0$, es decir, no se favoreció el entrenamiento de las redes cerca de las condiciones iniciales. Cabe mencionar que estos cuatro entrenamientos tardaron en completarse entre 7 y 11 minutos.

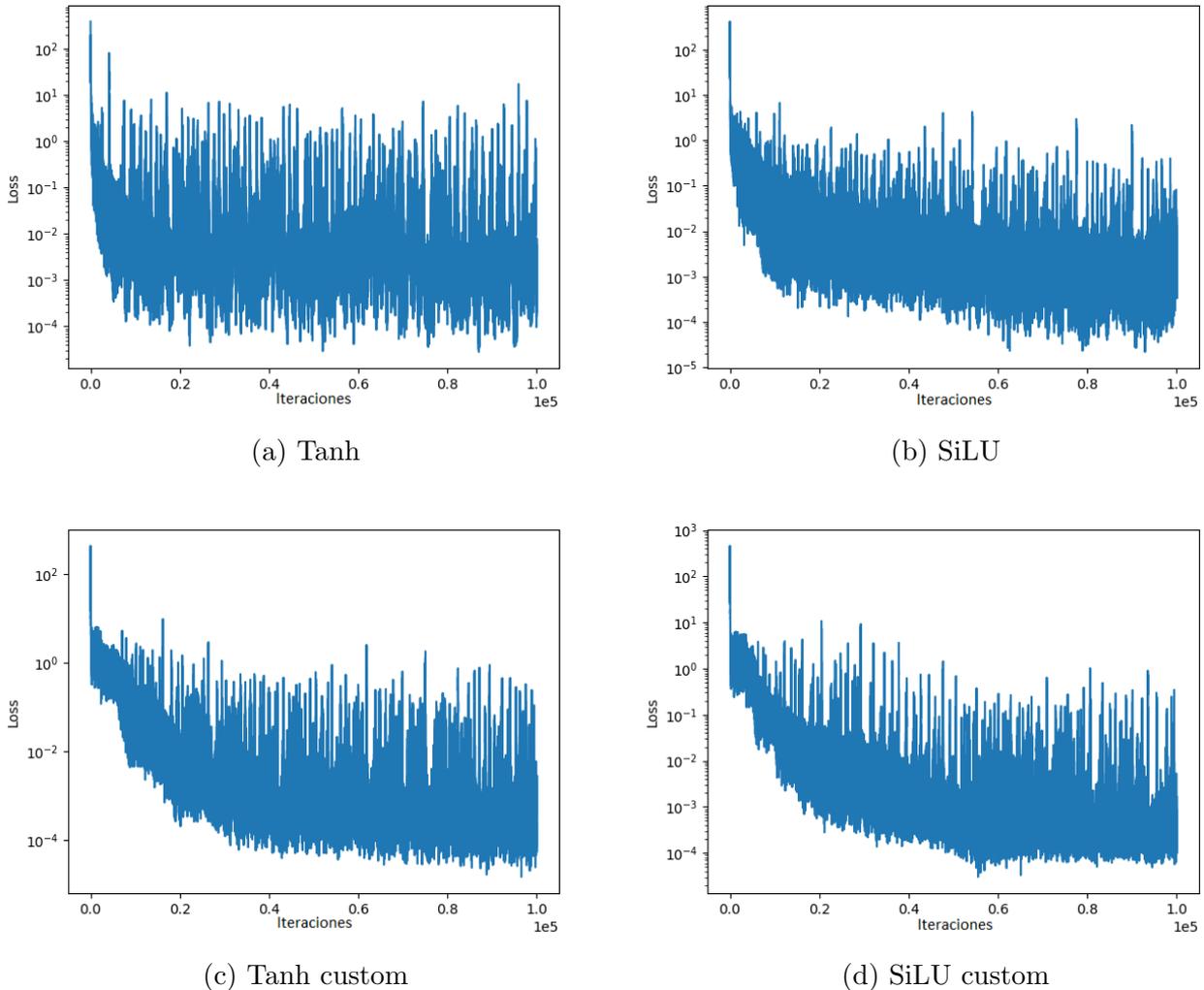


Figura 6.7: Funciones de pérdida para entrenamientos sin bundle llevados a cabo mediante la arquitectura habitual con función de activación tangente hiperbólica (a) y SiLU (b); y arquitectura customizada con tangente hiperbólica (c) y SiLU (d).

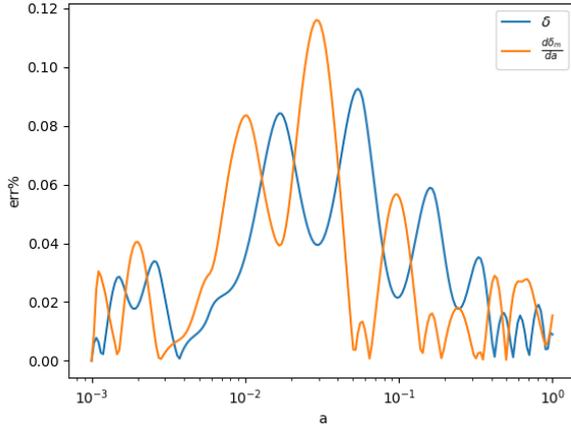
Es importante remarcar que cada iteración corresponde a un valor de los parámetros internos de las redes. Dicho esto, los valores fijados al finalizar el entrenamiento son aquellos que corresponden al valor más bajo de la función de pérdida, por lo tanto, el valor representativo de esta función para caracterizar el entrenamiento es el valor mínimo alcanzado entre todas las iteraciones. Además de esto, el gráfico de la función de pérdida a través de las iteraciones permite entender la dinámica del proceso de aprendizaje. Por ejemplo, en el caso (a) se puede ver que las redes rápidamente alcanzaron un valor que se mantuvo aproximadamente constante. En este caso, no se espera que la red mejore el aprendizaje con más iteraciones. Si se quiere obtener un mejor resultado con estas redes una opción puede ser achicar el learning rate o agrandar el batch size. En los otros casos se observa que las redes necesitan más iteraciones para alcanzar valores bajos, pero también se observa que, por ejemplo en el caso (b), hasta las últimas iteraciones empleadas las redes mantienen una tendencia al aprendizaje, por lo cual el resultado puede mejorar con los mismos hiperparámetros y más interacciones.

Se puede ver que en todos los casos se obtuvo un descenso de entre seis y siete órdenes de magnitud, lo cual es indicio de aprendizaje en las redes. Por otro lado, empleando la arquitectura habitual y la función de activación SiLU se logró el valor más bajo de dicha función. No obstante, el valor más bajo de la función de pérdida no permite caracterizar el error cometido por las redes al resolver las ecuaciones. Esto se corresponde al hecho de que, esta cantidad, se calcula como el promedio de los residuos de las ecuaciones sobre todos los valores tomados en cada iteración (es decir, el promedio sobre el batch). Debido a esto, para cuantificar el error de la solución provista por la red neuronal se calcula el error porcentual comparando con la solución numérica, empleando la ecuación 6.1. Estas diferencias porcentuales se pueden ver en la figura 6.8.

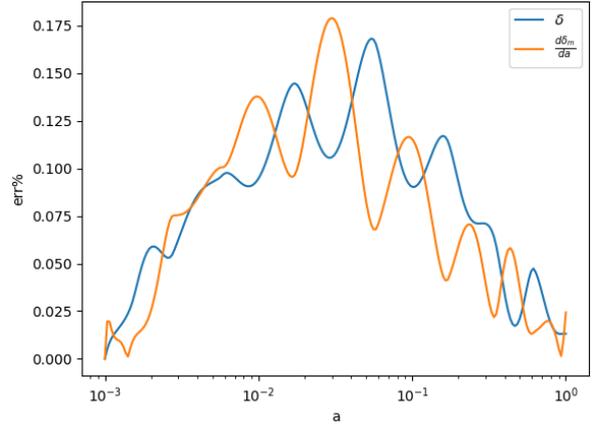
Estas figuras son las que permiten concluir que el entrenamiento de las redes fue exitoso. En todos los casos el error es menor al 0,25 %, cuando se asume como un resultado exitoso aquel que tenga una diferencia porcentual menor al 1 %. Resulta importante notar que en la condición inicial el error es del 0 %, lo cual quiere decir que las redes cumplen exactamente la condición inicial. Esto significa que la reparametrización impuesta funciona correctamente.

Por otro lado, es importante notar que las redes que corresponden al valor más bajo de la función de pérdida no se corresponden a un valor mínimo de diferencias porcentuales, dado que estas son mínimas en el caso (a), mientras que la función de pérdida más baja se logró en el caso (b). Esto refuerza la idea de que un valor menor de esta función no necesariamente se corresponde con una mejor solución, aunque el descenso de dicha función sí represente un indicio de aprendizaje en las redes.

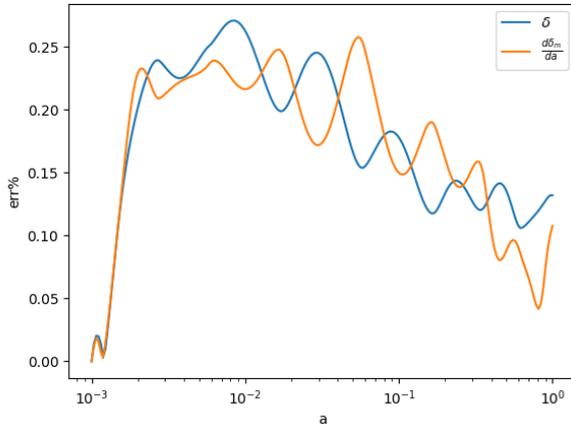
Para concluir esta sección, es necesario observar que el entrenamiento para el método sin bundle es el más sencillo de llevar a cabo, y, por lo tanto, los resultados obtenidos no necesariamente se trasladan a los entrenamientos con el método bundle, aunque sirven de punto de partida.



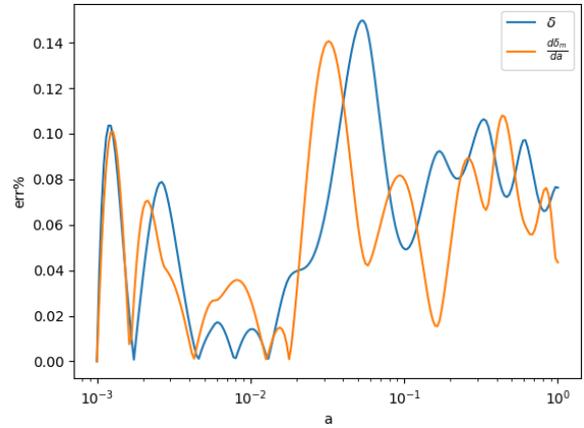
(a) Tanh



(b) SiLU



(c) Tanh custom



(d) SiLU custom

Figura 6.8: Errores porcentuales (ecuación 6.1) en función del factor de escala para entrenamientos sin bundle llevados a cabo mediante la arquitectura habitual con función de activación tangente hiperbólica (a) y SiLU (b); y arquitectura customizada con tangente hiperbólica (c) y SiLU (d).

6.3. Soluciones bundle para los modelos cosmológicos

En esta sección se discuten las soluciones obtenidas mediante el método bundle de las redes neuronales para los modelos teóricos estudiados: el modelo Λ CDM y el modelo fenomenológico de gravedad modificada.

6.3.1. Método bundle para Λ CDM

Se empleó el método bundle para resolver la ecuación diferencial 4.15, que corresponde a la evolución de las perturbaciones en el modelo Λ CDM. En este caso los inputs de la red son la variable independiente con las transformaciones pertinentes (\hat{N}) y la densidad de energía en forma de materia actual (Ω_{m0}).

En todos los casos se utilizó un batch size de 32, un learning rate de 10^{-3} y una estructura interna de tres capas de (256,128,64) neuronas artificiales cada una, con conexión completa entre

ellas.

Funciones de activación y arquitecturas para el caso bundle en Λ CDM

Al igual que en el caso sin bundle, se entrenaron redes neuronales empleando como funciones de activación la tangente hiperbólica y la función SiLU, con una arquitectura normal y con la arquitectura customizada. Los gráficos de las funciones de pérdida durante los entrenamientos se pueden ver en la figura 6.9. Cabe mencionar que estos cuatro entrenamientos tardaron en completarse entre una hora y media y dos horas.

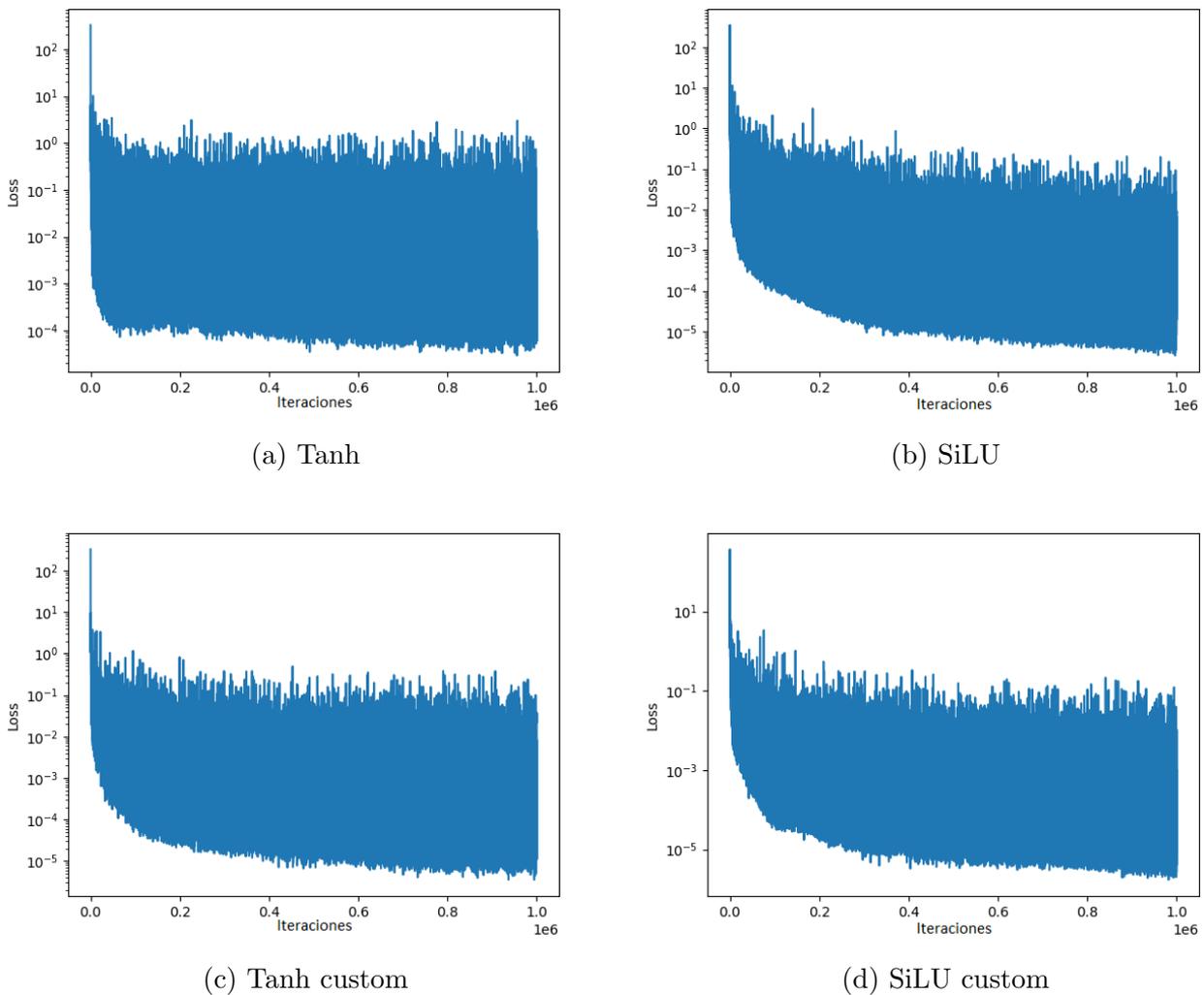


Figura 6.9: Funciones de pérdida para entrenamientos con bundle llevados a cabo mediante la arquitectura habitual con función de activación tangente hiperbólica (a) y SiLU (b); y arquitectura customizada con tangente hiperbólica (c) y SiLU (d).

En la figura se observa que con la arquitectura habitual y la tangente hiperbólica (a) se alcanza rápidamente, al igual que en el caso sin bundle, un valor bajo de la función de pérdida que luego se mantiene aproximadamente constante. Sin embargo, este valor es por lo menos un orden de magnitud más alto que en los otros tres casos. Este es un ejemplo de aprendizaje rápido, pero menos efectivo que el resto. Por otro lado, se observa que la arquitectura customizada

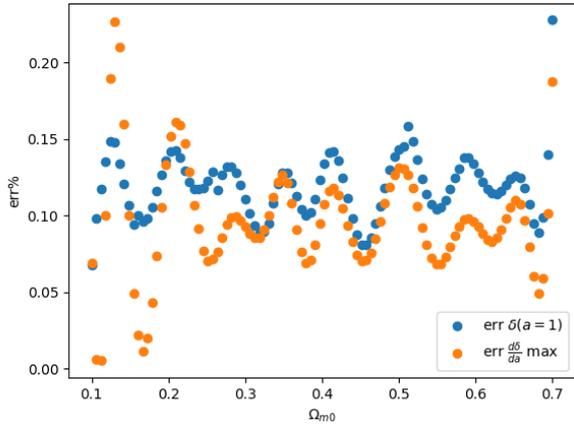
mejoró el aprendizaje en este caso, dado que logró una loss un orden de magnitud menor. De manera adicional, en los casos (b), (c) y (d) se observa que las redes mantienen una tendencia al aprendizaje hasta el final del entrenamiento, lo cual significa que, en caso de no obtener una solución satisfactoria, esta se puede mejorar empleando más iteraciones. Es importante remarcar que, para lograr funciones de pérdida del mismo orden de magnitud que para el caso sin bundle, fue necesario emplear un millón de iteraciones, mientras que en el caso anterior se utilizaron cien mil. Como se discutió en la sección anterior, el valor de la función de pérdida no permite estimar el error cometido en las soluciones. En la siguiente sección se discute el error cometido por las soluciones.

Error porcentual de la solución

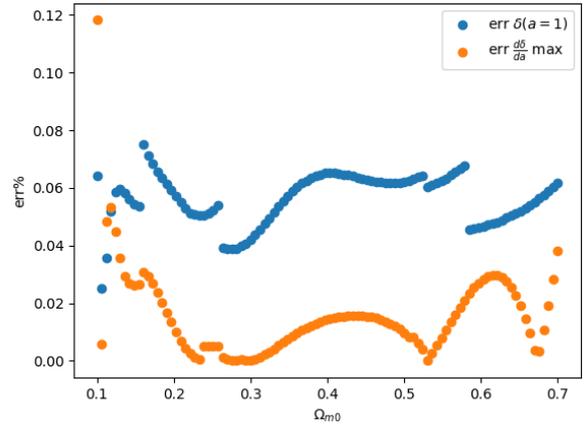
Al considerar el método bundle, se pueden construir gráficos como los de la figura 6.8 para cada valor de Ω_{m0} . Por este motivo, en la figura 6.10 se presenta el error porcentual en función de Ω_{m0} , donde para δ se evalúa en $a = 1$ (es decir, hoy) y para δ' se reporta el valor máximo que alcanza dicha cantidad. El motivo por el cual se presenta el error para δ en $a = 1$ corresponde al hecho de que, en el cálculo del observable (ec 2.70) para δ resulta relevante únicamente su valor actual, mientras que δ' se evalúa en el valor del factor de escala que corresponde a los datos, y, para no subestimar el error, se reporta el valor máximo.

Se puede ver que todas las diferencias porcentuales relevantes en el cálculo del observable están por debajo del 0,25% para todas las redes entrenadas, lo cual significa que todas las redes completaron un entrenamiento eficiente. Por otro lado, las redes que presentan un menor error porcentual son las que se entrenaron con la función de activación SiLU y la arquitectura customizada, motivo por el cual, a partir de ahora, se continuará únicamente con estas redes. No obstante, es importante remarcar que se podría haber empleado cualquiera de estas arquitecturas, ya que no hay diferencias relevantes entre ellas.

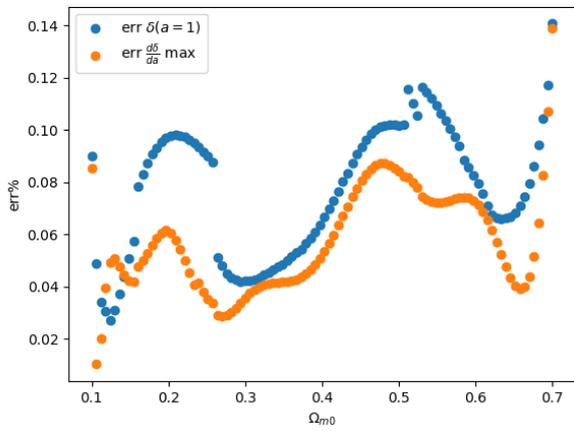
Finalmente, para estudiar los errores porcentuales cometidos en todo el rango de integración y así disponer de mayor información sobre la solución obtenida, en la figura 6.11 se puede ver un mapa en tres dimensiones en el cual se presenta el valor de dicho error en función de todos los valores de a y de Ω_{m0} . Este mapa se muestra únicamente para las redes entrenadas con la arquitectura customizada y con la función de activación SiLU. Se observa que para todos los valores de las dos entradas de las redes el error porcentual es menor al 0.1%, lo cual se interpreta prácticamente como una solución exacta.



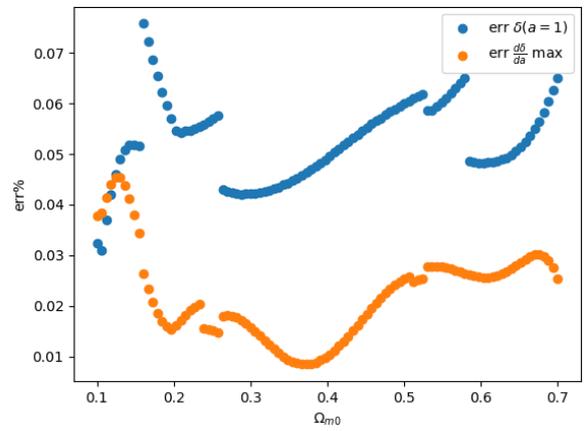
(a) Tanh



(b) SiLU

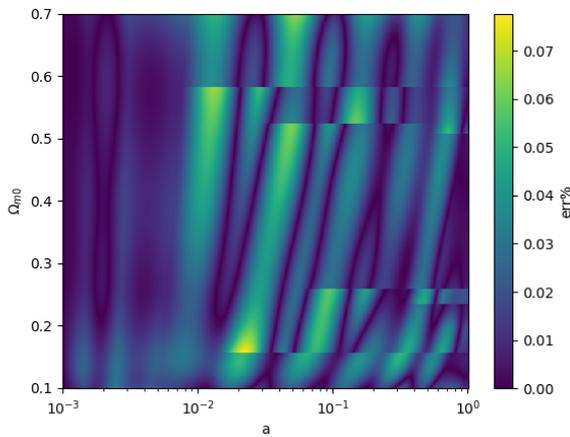


(c) Tanh custom

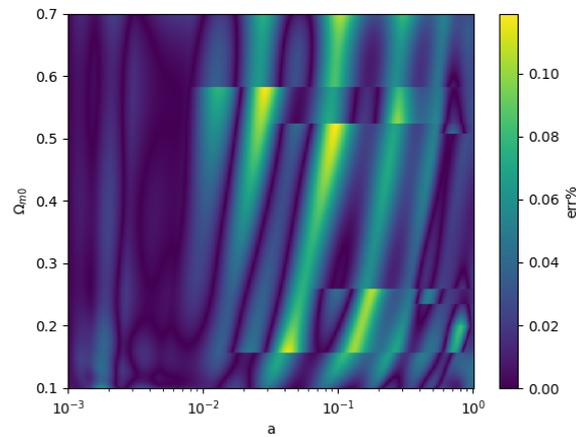


(d) SiLU custom

Figura 6.10: Errores porcentuales de $\delta(a = 1)$ y δ'_{\max} en función de Ω_{m0} para entrenamientos sin bundle llevados a cabo mediante la arquitectura habitual con función de activación tangente hiperbólica (a) y SiLU (b); y arquitectura customizada con tangente hiperbólica (c) y SiLU (d).



(a) δ



(b) δ'

Figura 6.11: Errores porcentuales para δ (a) y δ' (b) en función del parámetro del bundle (Ω_{m0}) y del factor de escala empleando la función de activación SiLU y la arquitectura customizada.

6.3.2. Método bundle para el modelo fenomenológico de gravedad modificada

Al considerar el método bundle para entrenar redes que resuelvan la ecuación 2.121 se procedió de la misma forma que para el modelo Λ CDM, con la diferencia de que hay un nuevo parámetro cosmológico en el bundle: g_a . Se empleó como función de activación la función SiLU y la arquitectura customizada, con un batch size de 32, un learning rate de 10^{-3} y una estructura interna de tres capas de (256,128,64) neuronas artificiales cada una, con conexión completa entre ellas. El gráfico de la función de pérdida se puede ver en la figura 6.12. Cabe mencionar que este entrenamiento tardó en completarse siete horas.

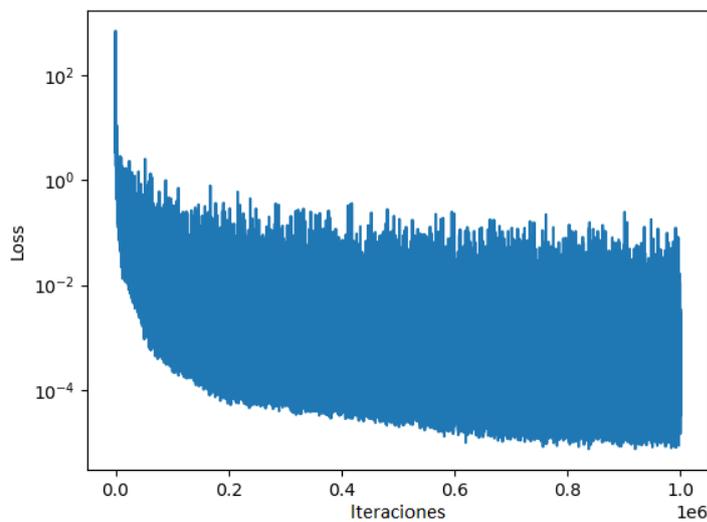


Figura 6.12: Función de pérdida para un entrenamiento con bundle realizado para el modelo fenomenológico de gravedad modificada. Para dicho entrenamiento se empleó la arquitectura customizada y la función de activación SiLU.

En la figura se ve que en este entrenamiento la función de pérdida se vio afectada en un descenso de entre seis y siete órdenes de magnitud, lo cual es indicio de un aprendizaje exitoso.

Error porcentual de la solución

Para estudiar el error cometido por las redes neuronales en este modelo fenomenológico, se comparó el valor de las redes con el numérico donde, para δ se reporta el valor hoy ($a = 1$), mientras que para δ' se reporta el valor máximo. Esto es así por el mismo motivo que antes: en el observable 2.70 entran en juego el valor de $\delta(a = 1)$, mientras que δ' se evalúa en varios puntos, y de esta forma reportar el error máximo no permite subestimar dicha incerteza. En la figura 6.13b se pueden ver mapas que muestran estos errores en función de los parámetros del bundle.

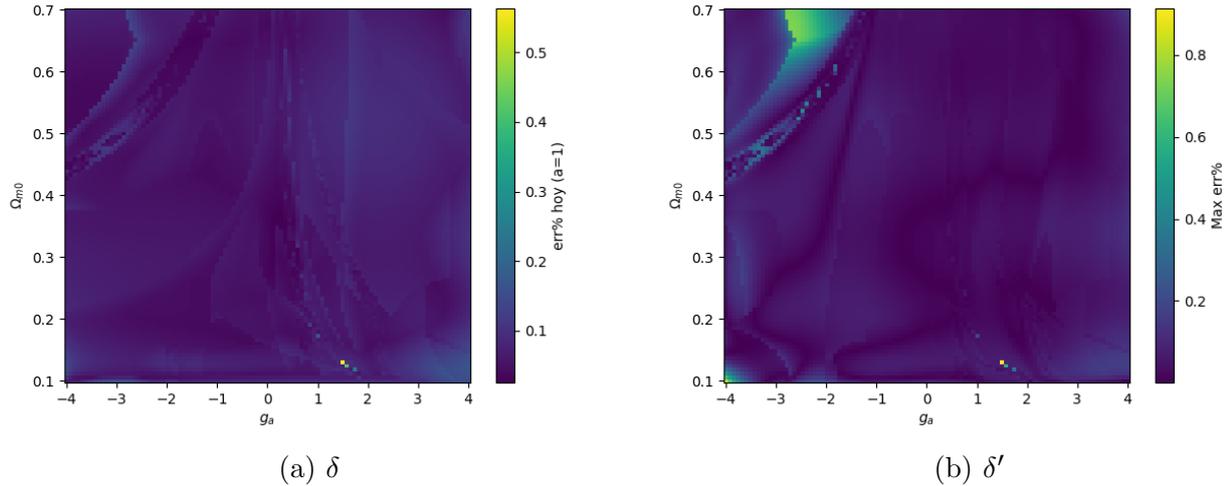


Figura 6.13: Errores porcentuales para δ evaluado en $a = 1$ (hoy), y el valor máximo para δ' .

En este caso también se observa que las diferencias porcentuales de las redes son menores al 0,8%, cumpliendo de este modo el criterio del 1%. Este resultado permite concluir que el entrenamiento de las redes fue exitoso y que se pueden emplear para la estimación de los parámetros cosmológicos. De manera complementaria, en la figura 6.14 se muestran mapas para el error porcentual para dos valores fijos de g_a : $-0,9$ y $-0,8$.

En la figura se observan errores porcentuales menores al 1% en todos los casos, lo cual, nuevamente significa una gran concordancia entre ambos métodos.

6.4. Inferencia de parámetros cosmológicos

En esta sección se describe una estimación a priori de los parámetros cosmológicos mediante el criterio de la máxima verosimilitud y la inferencia de los mismos a través del método de MCMC para los dos modelos teóricos estudiados en este trabajo.

6.4.1. Estimación de los parámetros en Λ CDM

Hasta ahora se ha discutido la capacidad de las redes neuronales para resolver las ecuaciones diferenciales correspondientes a la evolución de la densidad fraccional de perturbaciones de materia. Sin embargo, el objetivo del trabajo consiste en emplear estas soluciones para determinar parámetros cosmológicos a partir de datos observacionales, y de esta forma testear distintos modelos teóricos.

Un primer paso, rápido y sencillo, que sirve como resultado de chequeo para los análisis posteriores consiste en maximizar la función de verosimilitud, o likelihood. Este método es análogo a minimizar la cantidad χ^2 definida a través de la ecuación 5.7. La idea es que, los valores de los parámetros que minimizan esta cantidad, suelen llamarse "parámetros de mejor ajuste", y corresponden al juego de parámetros que maximizan la probabilidad de haber medido los datos en cuestión. Además, obtener un valor que sea consistente con el modelo Λ CDM constituye un chequeo del trabajo realizado.

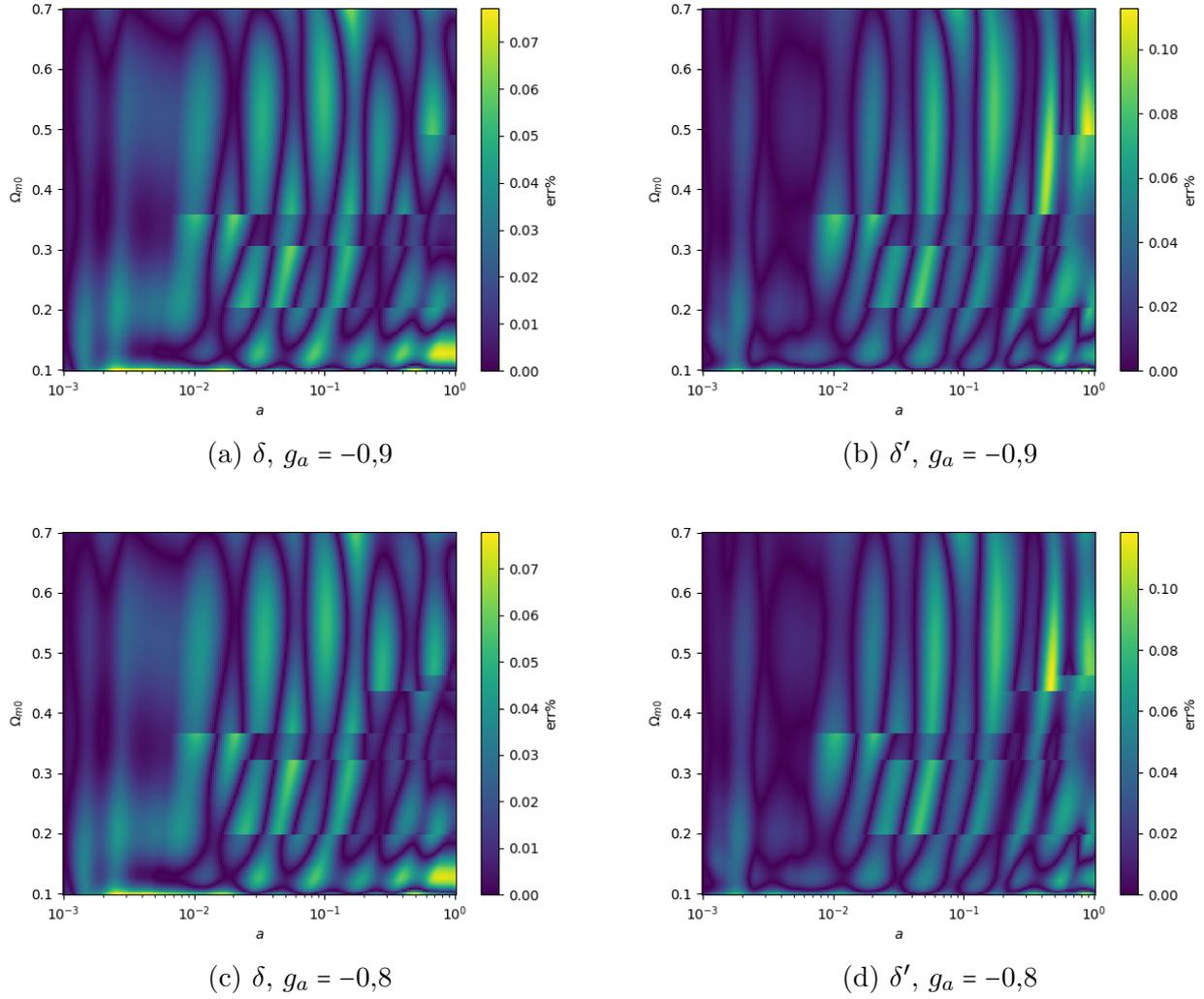


Figura 6.14: Errores porcentuales de δ y δ' en función del factor de escala y Ω_{m0} para dos valores del parámetro g_a : $-0,9$ y $-0,8$.

Estimación mediante el criterio de la máxima verosimilitud

Empleando la expresión 5.7, junto con los datos de la tabla 5.1, la definición del cociente (5.10) y la matriz de covarianza (5.13) para los datos correlacionados, es posible calcular el valor de la cantidad $\chi^2(\Omega_{m0}, \sigma_8)$. Resulta de interés mencionar que al realizar esto con el método numérico, para cada elección de los parámetros cosmológicos se debe realizar la integración con el algoritmo de Runge-Kutta

En la figura 6.15 se ve dicho valor, donde se impuso como valor máximo $\chi^2 = 20$ para una mayor resolución cerca de los parámetros de mejor ajuste.

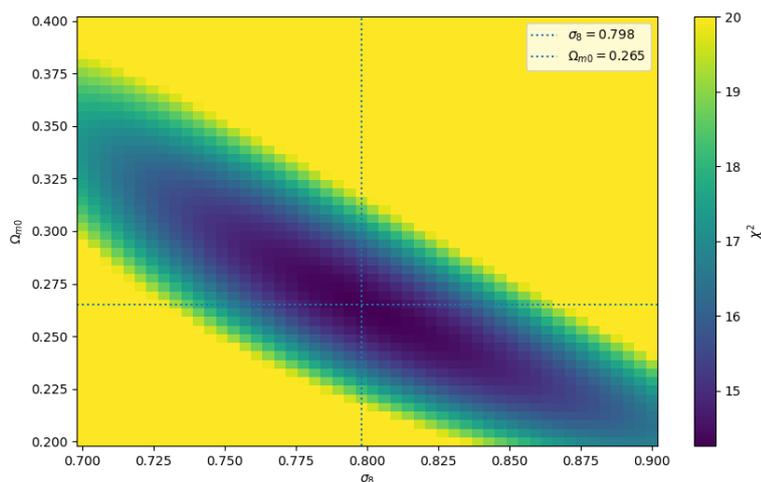


Figura 6.15: Cálculo del χ^2 mediante el método numérico.

Se obtuvo que el valor de los parámetros que mejor se ajustan a los datos corresponde a $\Omega_{m0} = 0,261$ y $\sigma_8 = 0,802$, los cuales son valores que tienen sentido en el contexto de la literatura [16]. No obstante, es necesario remarcar que estos valores tienen sentido únicamente como chequeo de seguridad, y no constituyen una predicción del trabajo, dado que no se les ha asociado un valor de incerteza.

Para estudiar la cantidad χ^2 con el método de las redes se realizó lo mismo que con el método numérico, donde ahora, para valor de los parámetros, en lugar de realizar una integración numérica, se evaluaron las redes neuronales en los valores correspondientes. El resultado se puede ver en la figura 6.16.

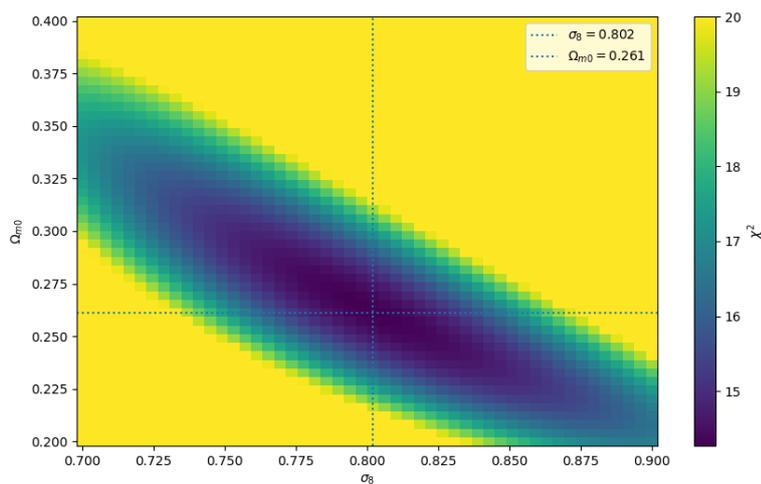


Figura 6.16: Cálculo del χ^2 mediante el método de las redes neuronales.

Como era de esperar dado el pequeño error porcentual obtenido, los resultados obtenidos por ambos métodos son prácticamente iguales. Esto nuevamente confirma que el entrenamiento de las redes fue exitoso.

De esta forma, a través de la estimación a priori mediante el método de la máxima verosimilitud se obtuvo

-	Ω_{m0}	σ_8
Num	0.261	0.802
NN	0.261	0.802

Tabla 6.1: Estimación a priori de los parámetros mediante el método de máxima verosimilitud para el método numérico tradicional y para el método de las redes neuronales.

Inferencia de parámetros mediante el MCMC

Luego de realizar una estimación a priori mediante el criterio de la maximización de la likelihood, se obtuvieron las distribuciones a posteriori de los parámetros cosmológicos de cada modelo a través del método de MCMC. En la figura 6.17 se pueden ver las cadenas obtenidas mediante ambos métodos. Cabe mencionar que para implementar el MCMC se empleó la librería emcee [113], que introduce los "samplers" mencionados en [114].

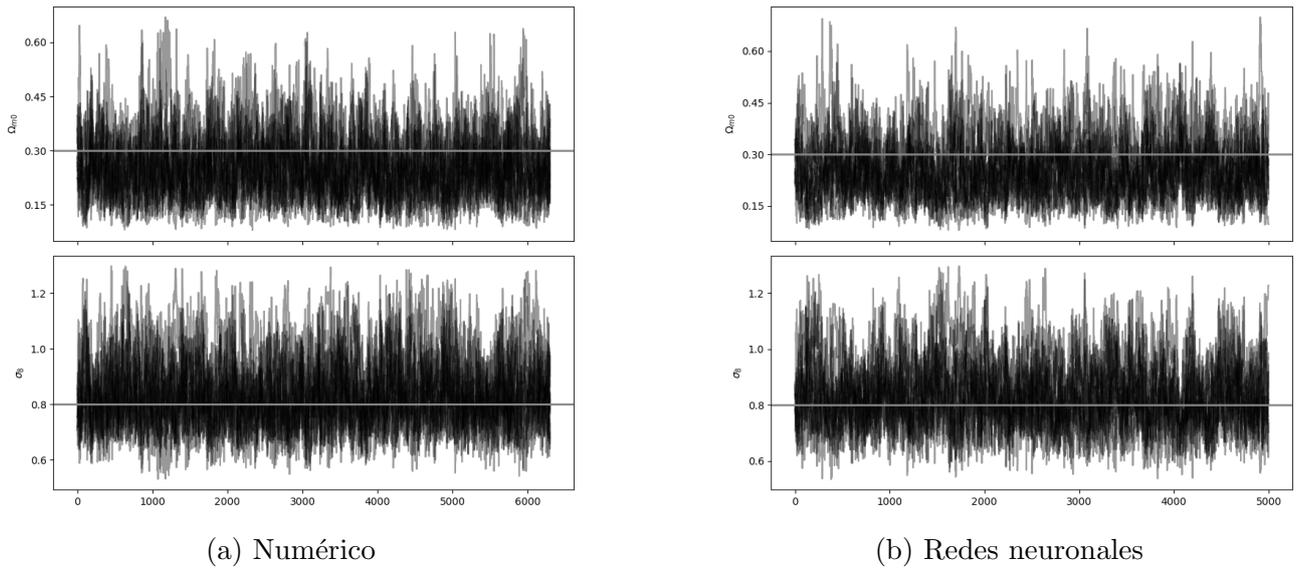


Figura 6.17: Cadenas de Markov obtenidas mediante el método de MCMC para ambos métodos en el modelo Λ CDM. En ambos casos se observa que las cadenas no chocan con los priors, lo cual es importante para garantizar una estimación correcta.

En la figura 6.18 se pueden ver, superpuestas, las distribuciones posterior tanto para el método numérico como para el método de las redes neuronales. Para realizar estos gráficos se empleó la librería de python getdist [115]. En oscuro se muestran los contornos correspondientes al 68% de confianza, mientras que la parte clara corresponde al 95% de confianza.

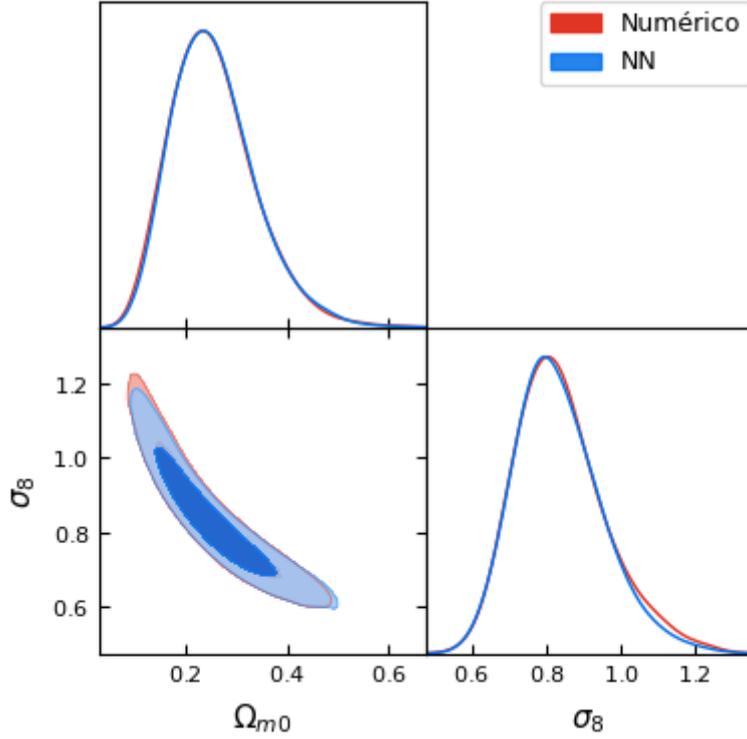


Figura 6.18: Distribuciones posterior y contornos de confianza para el modelo Λ CDM obtenidas mediante el método de MCMC.

En la figura 6.18 se puede ver que los parámetros Ω_{m0} y σ_8 están fuertemente correlacionados. Esto se observa en el gráfico correspondiente a los contornos, dado que una variación en uno de los parámetros necesariamente implica una variación en el otro. Esta correlación se entiende notando que la solución de la ecuación del modelo (δ_m) depende de Ω_{m0} , luego como para construir el observable se multiplica esta solución por el parámetro σ_8 resulta intuitivo que el observable tenga dependencia en un factor multiplicativo que involucre ambos parámetros. Este hecho introduce una correlación entre Ω_{m0} y σ_8 a la hora de calcular $f\sigma_8$, la cual es evidente en la figura 6.18. Empleando el criterio mencionado en el capítulo de implementaciones, se obtienen los intervalos de confianza para los parámetros cosmológicos, los cuales se muestran en la tabla 6.2.

-	Ω_{m0}	σ_8
Num	$0,253^{+0,029}_{-0,044}$	$0,839^{+0,040}_{-0,064}$
NN	$0,255^{+0,029}_{-0,044}$	$0,835^{+0,041}_{-0,061}$

Tabla 6.2: Intervalos de confianza para el modelo Λ CDM obtenidos con el método numérico tradicional y el método de las redes neuronales.

Los resultados obtenidos mediante ambos métodos son prácticamente indistinguibles entre sí, lo cual permite concluir que el objetivo de emplear las redes neuronales para realizar la inferencia de parámetros fue exitoso. Además, la estimación para los valores de Ω_{m0} y σ_8 son compatibles con resultados similares [2, 72]. A pesar de esto, resulta importante notar que, al

emplear la misma tabla de datos, la estimación debería ser idéntica a la de la literatura [2]. Las diferencias obtenidas se atribuyen al valor de la cantidad Ω_{r0} . En este trabajo se empleó $\Omega_{r0} = 5,38 \times 10^{-5}$ [63], mientras que dada la fecha de publicación del artículo correspondiente el valor empleado pudo ser otro. La importancia del valor de Ω_{r0} radica fundamentalmente en la expresión de $a_{\text{eq}} = \Omega_{r0}/\Omega_{m0}$, que constituye una cantidad relevante en las ecuaciones, tanto por su propia aparición, como además también dentro de la cantidad $\alpha = a_{\text{eq}}^4 \Omega_{\Lambda 0}/\Omega_{m0}$.

De mayor interés resulta notar la diferencia entre la inferencia realizada y el valor obtenido mediante el FCR asumiendo Λ CDM, reportado en la colaboración de Planck 2018 [3]: $\Omega_{m0} = 0,315 \pm 0,007$ y $\sigma_8 = 0,811 \pm 0,006$. Para discutir los resultados obtenidos en los análisis estadísticos es útil calcular el χ^2 que corresponde a cada uno de ellos. Con los datos que se usaron en este trabajo se obtiene, para los parámetros de Planck $\chi^2(0,315; 0,811) = 21,663$, para los valores obtenidos mediante el MCMC resulta $\chi^2(0,255; 0,835) = 14,993$, mientras que para los parámetros de mejor ajuste $\chi^2(0,261; 0,802) = 14,272$. En la figura 6.19 se puede ver la comparación entre las predicciones del observable $f\sigma_8$ correspondientes a los valores de Planck 2018 y a los parámetros inferidos mediante el MCMC.

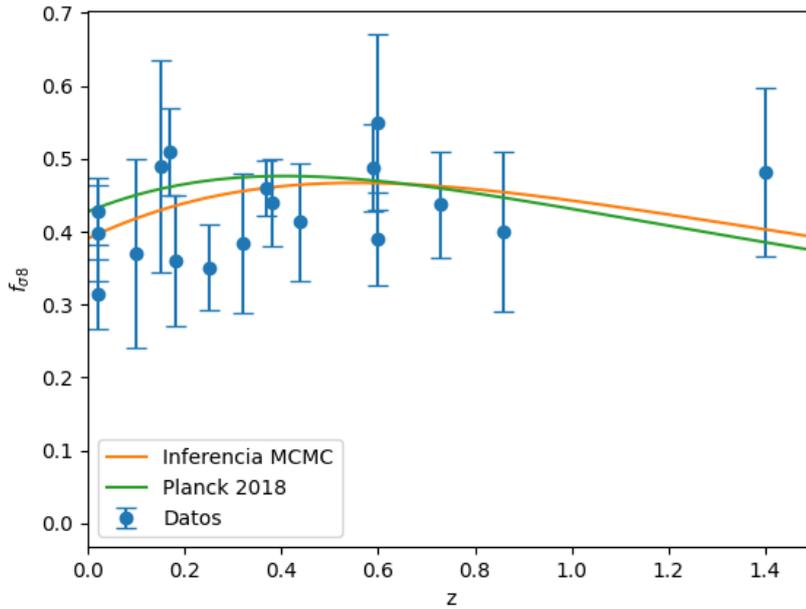


Figura 6.19: Comparación entre los parámetros obtenidos por medio de la inferencia con el MCMC y los reportados en Planck 2018 [3].

En la figura se ve que la curva que corresponde a los parámetros de Planck 2018 parece alejarse de los datos más que la que corresponde a la inferencia realizada, lo cual se corresponde con los valores obtenidos de la cantidad χ^2 y permite ganar confianza en el resultado obtenido. Se concluye entonces que mediante la inferencia de parámetros llevada a cabo mediante el MCMC con las redes neuronales se obtuvo un valor de Ω_{m0} que, al igual que en la literatura [2, 72] no presenta un acuerdo en el intervalo del 68% de confianza con el valor reportado por la colaboración de Planck 2018 [3]. Para cuantificar esta diferencia, se puede emplear una regla rápida [28] que consiste en calcular el número de sigmas entre dos resultados $x_1 = x_m^1 \pm \sigma_1$ y

$x_2 = x_m^2 \pm \sigma_2$ como

$$N_\sigma = \frac{|x_m^1 - x_m^2|}{\sqrt{\sigma_1^2 + \sigma_2^2}}. \quad (6.3)$$

Empleando la ecuación 6.3 se ve que el resultado obtenido mediante la inferencia a través del MCMC y la colaboración de Planck 2018 corresponde a $N_\sigma = 1,34$ para el parámetro Ω_{m0} , mientras que en el caso de σ_8 se obtiene $N_\sigma = 0,38$. Este breve análisis permite concluir que las diferencias encontradas no constituyen una inconsistencia para el modelo, donde se puede interpretar como inconsistencia a una diferencia de más de 3σ . No obstante es necesario aclarar que el estudio de tensiones dentro de los modelos cosmológicos es vasto, y entrar en esas discusiones está más allá de los objetivos de esta tesis. Herramientas más sofisticadas para estudiar estos aspectos se pueden encontrar en la literatura [28, 116, 117], y un trabajo a futuro prometedor consiste en emplear dichas herramientas para estudiar este tipo de discrepancias.

6.4.2. Estimación de los parámetros en el modelo fenomenológico de gravedad modificada

Al igual que para Λ CDM, se calculó el valor de la cantidad χ^2 definida en 5.7 para distintos valores de los parámetros con el fin de hallar aquellos de máxima verosimilitud, para luego realizar la inferencia de los parámetros mediante el MCMC. En este caso existe un problema de degeneración en χ^2 dentro del espacio de parámetros, lo cual complica llevar a cabo la inferencia. En este contexto, la estimación mediante máxima verosimilitud representa un punto crucial para poder obtener las distribuciones a posteriori y de esta forma conseguir los intervalos de confianza para los parámetros del modelo.

Estimación mediante el criterio de máxima verosimilitud

A diferencia del modelo Λ CDM no es posible representar los valores de χ^2 en un gráfico, dado que en este modelo hay tres parámetros. Además se observó una degeneración en los valores del mínimo χ^2 en el espacio de parámetros, la cual también se reporta en la literatura [2]. Por este motivo, la estimación a priori mediante el criterio de máxima verosimilitud presenta valores de la likelihood muy similares, para conjuntos de parámetros muy diferentes entre sí. Resulta importante remarcar que, dentro de la degeneración se encuentra el valor de $g_a = 0$, el cual corresponde al modelo Λ CDM y presenta valores de Ω_{m0} y σ_8 que son compatibles con la inferencia realizada para dicho modelo. A pesar de la degeneración, se eligieron como parámetros de mejor ajuste, tanto para el método numérico como para el método de las redes, los valores $\Omega_{m0} = 0,252$, $\sigma_8 = 0,882$ y $g_a = -0,882$, los cuales corresponden al mínimo global de la cantidad χ^2 . El valor de mejor ajuste del parámetro g_a corresponde a una elección importante, dado que impone condiciones fuertes a su distribución a priori, lo cual se discute en el apéndice A. Al igual que en el modelo estándar de la cosmología, se obtuvieron resultados idénticos en ambos modelos, al menos al considerar las primeras tres cifras después de la coma. Esto era de esperar al considerar que los errores porcentuales de las soluciones fueron pequeños, de esta manera en

la tabla 6.3 se ven los parámetros elegidos como mejor ajuste para ambos métodos.

-	Ω_{m0}	σ_8	g_a
Num	0.252	0.882	-0.882
NN	0.252	0.882	-0.882

Tabla 6.3: Estimación a priori mediante el criterio de máxima verosimilitud para el modelo fenomenológico de gravedad modificada para los dos métodos empleados.

Inferencia de parámetros mediante el MCMC

Como se mencionó previamente, imponer las distribuciones a priori para los parámetros correspondientes a este modelo no resulta tan directo como en Λ CDM. Esto es así debido a que la degeneración que presenta el parámetro g_a genera problemas en las cadenas de Markov al explorar el espacio de parámetros. En el apéndice A se discute la elección para la distribución a priori de g_a . Luego, empleando los priors uniformes de la ecuación 4.42

$$I_{\Omega_{m0}} = [0,1; 0,7], \quad I_{\sigma_8} = [0,6; 1,3], \quad I_{g_a} = [-1,0; -0,7], \quad (6.4)$$

se llevó a cabo el proceso de MCMC, mediante el cual se obtuvieron las distribuciones a posteriori para ambos métodos, las cuales se pueden ver en la figura 6.20.

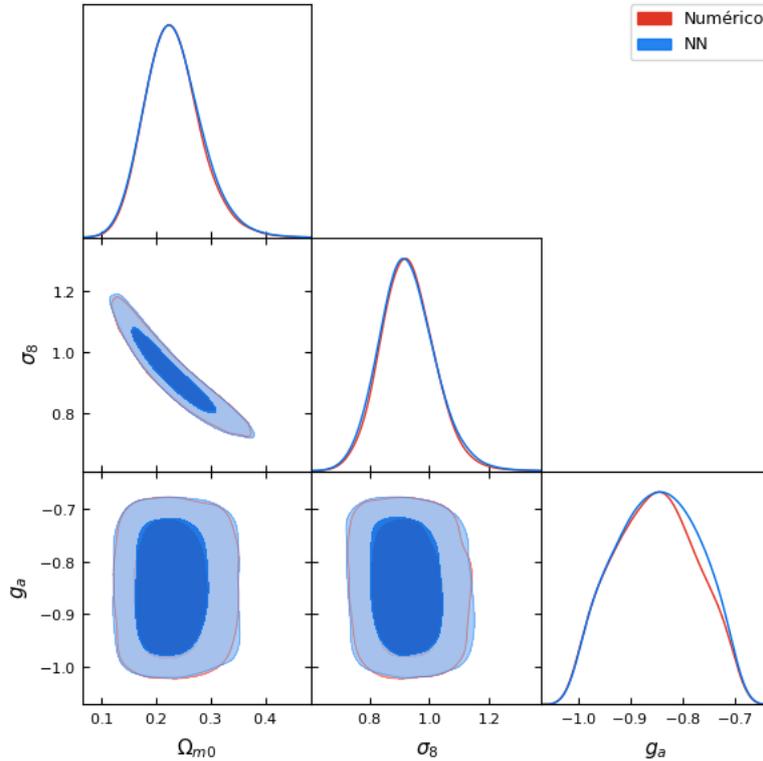


Figura 6.20: Distribuciones posterior y contornos de confianza para el modelo fenomenológico de gravedad modificada obtenidas mediante el método de MCMC.

En la figura 6.20 se ve que, al igual que en Λ CDM, hay correlación entre Ω_{m0} y σ_8 , la cual se entiende de la misma forma que en el modelo estándar. No obstante, el nuevo parámetro g_a

no muestra correlación con ninguno de los otros dos parámetros. Esto se entiende en los gráficos de contornos dado que, modificar el valor de g_a no impone modificaciones ni en Ω_{m0} , ni en σ_8 . Se atribuyó esta falta de correlación al hecho de que, este parámetro, entra en la expresión 2.119

$$\frac{G_{\text{eff}}}{G} = 1 + g_a (1 - a)^n - g_a (1 - a)^{2n}, \quad (6.5)$$

siendo el menor valor de a para los datos estudiados $a_{\text{min}} = 0,44$ y donde se asume $n = 2$. En este contexto, se puede pensar que la parametrización sugerida introduce correcciones pequeñas, dado que $G_{\text{eff}}/G \approx 1$, lo cual se puede ver en la figura 6.21 para los valores de g_a que delimitaron las distribuciones a priori. Cualquier otro valor de g_a en el rango estudiado mediante el MCMC define una curva que se encuentra entre las dos que se muestran en la figura.

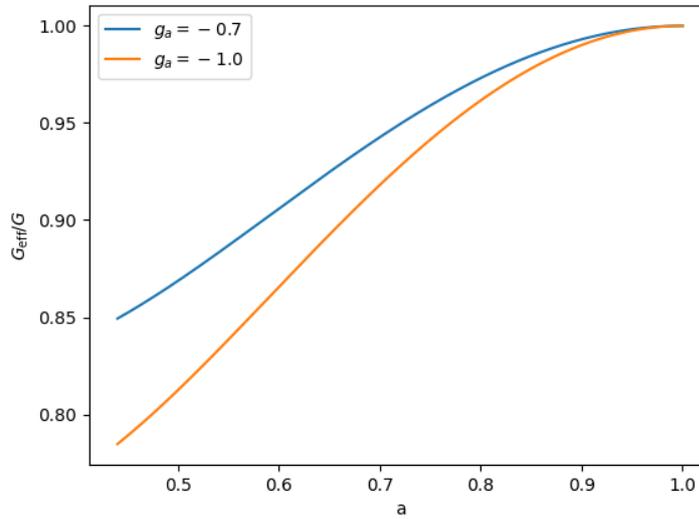


Figura 6.21: G_{eff}/G vs a . Se observa que las correcciones introducidas dentro del intervalo en que se encuentran los datos observacionales de $f\sigma_8$ son tales que $G_{\text{eff}}/G \approx 1$, lo cual explica la descorrelación entre g_a y los otros parámetros del modelo.

Mediante las distribuciones a posteriori (figura 6.20) se realizó la inferencia de los parámetros, la cual se presenta en la tabla 6.4.

-	Ω_{m0}	σ_8	g_a
Num	$0,231^{+0,019}_{-0,025}$	$0,929^{+0,036}_{-0,043}$	$-0,849^{+0,039}_{-0,039}$
NN	$0,231^{+0,020}_{-0,025}$	$0,929^{+0,036}_{-0,044}$	$-0,849^{+0,040}_{-0,040}$

Tabla 6.4: Intervalos de confianza obtenidos para el modelo fenomenológico de gravedad modificada utilizando el método numérico tradicional y el método de las redes neuronales.

De esta forma, al igual que para el modelo Λ CDM, se concluye que fue posible emplear el método de las redes neuronales en la estimación de parámetros cosmológicos en este modelo fenomenológico. De esta forma, el siguiente paso consiste en emplear este método en un modelo que se enmarque dentro de las teorías $f(R)$, para lo cual se deben resolver las dificultades planteadas en el capítulo 2 para dicho modelo.

Un comentario pertinente corresponde a que, al igual que en el modelo Λ CDM, no se obtuvo una estimación idéntica a la de la literatura [2]. De la misma forma que en dicho modelo, se atribuyen las diferencias en la inferencia de parámetros al valor empleado de Ω_{r0} , el cual se modificó desde la publicación del artículo mencionado. En este trabajo se empleó el valor reportado en [63].

Finalmente, es importante discutir la comparación de los resultados obtenidos mediante el modelo fenomenológico con resultados de la literatura. En este caso no es posible comparar los resultados, ni de Ω_{m0} , ni de σ_8 con Planck 2018 debido a que los resultados de dicha colaboración se obtienen asumiendo Λ CDM. Como no hay mediciones de estos parámetros que se realicen en el contexto del modelo estudiado con un juego de datos distinto al utilizado en este trabajo no es posible realizar comparaciones como en el caso de Λ CDM. Sí se pueden comparar estos valores con los de [2] ($\Omega_{m0} \sim 0,2$, $\sigma_8 \sim 1,0$ y $g_a = -1,156 \pm 0,341$), que son consistentes. Esto no es sorprendente dado que se obtienen con los mismos datos, pero con un valor diferente de Ω_{r0} .

6.5. Tiempos computacionales en el MCMC

Hasta este punto se mostraron los resultados obtenidos mediante el método de las redes neuronales y el método numérico tradicional. En esta sección se discuten los tiempos computacionales que se obtuvieron al realizar el proceso de MCMC. En la tabla 6.5 se muestran valores aproximados de los tiempos de corrida, en ambos modelos y mediante ambos métodos. Es importante remarcar que dado que las soluciones de ambos métodos tienen diferencias menores al 1%, se espera que las cadenas convergan con un número similar de iteraciones sin importar el método en cuestión. Esto debe ser así debido a que se recorre el mismo espacio de parámetros con una likelihood prácticamente idéntica. Se observó que esto fue así, y de esta forma se atribuyen las diferencias en los tiempos computacionales únicamente a cuestiones de cómputo en cada método.

-	Λ CDM	Mod. fenom.
Num	$\sim 10'$	$\sim 20'$
NN	$\sim 20'$	$\sim 30'$

Tabla 6.5: Tiempos en los cuales se completó el proceso de MCMC, para el modelo estándar Λ CDM y el modelo fenomenológico de gravedad modificada.

Se observa que el método numérico tradicional tardó la mitad que el método de las redes para el modelo estándar. Además, se observa que en el modelo fenomenológico el numérico tardó 2/3 respecto al tiempo que llevó el método con las redes. El hecho de que el método tradicional se lleve a cabo más rápido también se observó en un trabajo anterior del grupo [86] y se asocia a dos motivos: la optimización del integrador de solve_ivp respecto a los códigos empleados en esta tesis, y al hecho de ser un sistema simple de dos ecuaciones. A pesar de esto, se observó que la diferencia de tiempos se acortó al introducir un parámetro más, lo cual también se observó en el trabajo mencionado anteriormente, donde en sistemas más complicados de cinco ecuaciones las redes superaron en tiempos a los métodos numéricos tradicionales. Se concluye que el método

de las redes debe ser optimizado, pero también se supone que en modelos más complejos, como por ejemplo el de Hu-Sawicki, en el cual se introduce una mayor cantidad de parámetros, el método de las redes puede introducir ventajas importantes. Esto último es algo que también se observó en trabajos anteriores del grupo, y motiva a continuar estudiando este método con el fin de optimizarlo y emplearlo en modelos donde la cantidad de parámetros libres permita obtener ventajas respecto a los métodos numéricos tradicionales.

Capítulo 7

Conclusiones

En este trabajo se logró aplicar exitosamente el método de las redes neuronales (NN) para resolver la ecuación diferencial que describe la evolución de las perturbaciones de materia. En un trabajo anterior [47] se empleó dicho método a las ecuaciones de fondo cosmológico. En este trabajo se emplearon las redes en el contexto del modelo estándar de la cosmología (Λ CDM) y en un modelo fenomenológico de gravedad modificada. En cada caso se realizó una comparación del método de las redes con un método numérico tradicional (Runge-Kutta de orden 5), obteniendo diferencias porcentuales que en todos los casos fueron menores al 1%, donde además estas diferencias podrían reducirse aún más en caso de ser necesario. De manera adicional, se incluyó durante el entrenamiento de las redes neuronales una arquitectura customizada cuyo desarrollo aún sigue en marcha por el grupo de trabajo del co-director de esta tesis.

Con las soluciones obtenidas se realizaron análisis estadísticos mediante el método MCMC que permitieron inferir los parámetros cosmológicos de cada modelo. Con dicho método se estimaron los valores de las cantidades Ω_{m0} y σ_8 para el modelo Λ CDM, mientras que en el modelo fenomenológico se estimó además un parámetro característico llamado g_a . En ambos modelos los resultados para el método numérico y el método de las redes fueron concordantes, lo cual era de esperar dada la precisión de las soluciones. En Λ CDM se obtuvo $\Omega_{m0} = 0,255^{+0,029}_{-0,044}$ y $\sigma_8 = 0,835^{+0,041}_{-0,061}$ mediante el método de las redes, mientras que en el modelo fenomenológico los resultados obtenidos empleando las redes neuronales fueron $\Omega_{m0} = 0,231^{+0,020}_{-0,025}$, $\sigma_8 = 0,929^{+0,036}_{-0,044}$ y $g_a = -0,849^{+0,040}_{-0,040}$.

Los valores inferidos para los parámetros se compararon, en el caso de Λ CDM con la colaboración Planck 2018, en la cual se asume el modelo estándar de la cosmología. Empleando una regla sencilla del análisis estadístico se concluyó que todos los resultados obtenidos son consistentes, dado que ninguno alcanza el valor de 3σ respecto del valor de referencia empleado en cada caso. No se pudieron comparar los valores inferidos en el modelo fenomenológico con unos obtenidos a través de otro juego de datos, dado que no se disponen resultados para los parámetros en el contexto del modelo en cuestión. Sin embargo los resultados obtenidos son consistentes con una estimación realizada mediante los mismos datos, pero con un valor diferente de Ω_{r0} .

Además se observó que a través de los métodos numéricos tradicionales, se completó el MCMC en menos tiempo que a través del método de las redes neuronales. En el caso de Λ CDM el numérico tardó la mitad que las redes, mientras que para el modelo fenomenológico (donde se

agrega un nuevo parámetro) esta diferencia se redujo a que el numérico tardó dos terceras partes respecto al tiempo que le llevó a las redes neuronales. Se atribuye esta diferencia a dos aspectos: a la optimización del método `solve_ivp` respecto a los códigos empleados en esta tesis, y al hecho de estar integrando un sistema sencillo de dos ecuaciones, dado que, en trabajos anteriores [86] se observó que para sistemas de cinco ecuaciones el método de las redes neuronales superó en tiempos computacionales al método numérico tradicional. A pesar de esto, el hecho de que la diferencia se haya reducido al agregar un parámetro al modelo muestra una tendencia hacia el hecho de que, en modelos con más parámetros libres, el método de las redes neuronales puede presentar ventajas respecto a los métodos numéricos.

Por otro lado, en esta tesis se presenta una discusión acerca de la dificultad para estimar las condiciones iniciales de la ecuación diferencial que describe la evolución de la densidad de materia en el contexto de teorías $f(R)$. Esta dificultad se sorteó considerando el modelo fenomenológico, pero es relevante a la hora de estudiar modelos de gravedad modificada provenientes de primeros principios. Se observó que las hipótesis usuales acerca de las condiciones iniciales no se aplican y resulta necesario realizar una integración previa de todas las ecuaciones acopladas para poder obtener las condiciones iniciales correctas.

De esta forma, se proponen tres líneas de trabajo a futuro las cuales son prometedoras y tienen como punto de partida los temas abordados en esta tesis:

- Por un lado resulta prometedor estudiar la evolución de perturbaciones de materia en el contexto de teorías $f(R)$ empleando el cálculo correcto para las condiciones iniciales. Este cálculo puede llevarse a cabo mediante códigos de público acceso que resuelven las ecuaciones de la cosmología acopladas.
- Otra línea de trabajo prometedora consiste en la estimación de errores en el método de las redes neuronales. Como se discutió en este trabajo, dicho método no tiene a día de hoy un método automatizado para realizar dicha estimación, y por eso resulta crucial la comparación con métodos numéricos tradicionales. En este momento está en desarrollo un método de cuantificación de incertezas a través del grupo de trabajo del co-director de esta tesis. Aplicar dicho método a las soluciones obtenidas y estudiar la estimación de los errores también representa un campo de trabajo prometedor dentro del machine learning.
- Finalmente, puede ser de utilidad trabajar en la optimización del método de las redes neuronales con el fin de mejorar los tiempos computacionales en modelos con pocos parámetros libres.

Apéndice A: estimación del prior de g_a

En la tabla 7.1 se muestran valores de χ^2 todos dentro del intervalo [14,25;14,30] junto con los parámetros empleados para el cálculo, siendo este llevado a cabo mediante el método de las redes neuronales. Este cálculo se llevó a cabo para 35 valores de cada parámetro dentro de los intervalos $I_{\Omega_{m0}} = [0,1;0,5]$, $I_{\sigma_8} = [0,6;1]$ y $I_{g_a} = [-3;3]$ Es importante notar que el rango de variación elegido para el valor de χ^2 corresponde a, aproximadamente, 0,05 sobre 14,27, siendo así un porcentaje pequeño del 0,3%, mientras que la diferencia entre los valores de los parámetros que figuran en la tabla es notoria, sobretodo para el caso de g_a . Con las elecciones para la exploración del espacio de parámetros mencionadas previamente, se encontraron 16 juegos de parámetros que cumplieron estar dentro del rango de variación elegido para χ^2 , de los cuales solo algunos se reportan en la tabla 7.1, donde aparecen en orden ascendente del valor de χ^2 .

χ^2	Ω_{m0}	σ_8	g_a
14.252	0.252	0.882	-0.882
14.265	0.252	0.870	-0.705
14.267	0.264	0.823	-0.352
14.275	0.264	0.800	0.000
14.280	0.264	0.788	0.176
14.287	0.264	0.776	0.352
14.288	0.241	0.976	-1.588

Tabla 7.1: Valores de χ^2 en función de los parámetros del modelo. Se observa una degeneración en el parámetro g_a , dado a que dentro de variaciones de aproximadamente el 0,3% de χ^2 se ven variaciones de hasta aproximadamente el 150% en los valores de dicho parámetro.

Para poder realizar una exploración precisa y obtener una distribución a posteriori que permita describir un intervalo de confianza robusto para este parámetro, fue necesario tomar ciertas precauciones a la hora de definir su distribución a priori. Para fijar el prior sobre g_a se procedió de la siguiente manera: en primer lugar se centró el intervalo en el valor de g_a que se eligió como de mejor ajuste, es decir, en $g_a = -0,882$. Luego, para extenderlo, se estudió la distribución de mínimos de χ^2 en el espacio de parámetros, donde se asumieron como mínimos todos los valores calculados de dicha cantidad que se encontraron entre 14.25 y 14.30. Este estudio permitió restringir el prior en g_a para lograr aislar un único mínimo dentro de dicho intervalo, siendo este único mínimo el que corresponde a los parámetros de mejor ajuste. En la figura 7.1 se puede ver la distribución de estos puntos en el espacio de parámetros, donde el punto que corresponde a los parámetros de mejor ajuste se marca en un color diferente.

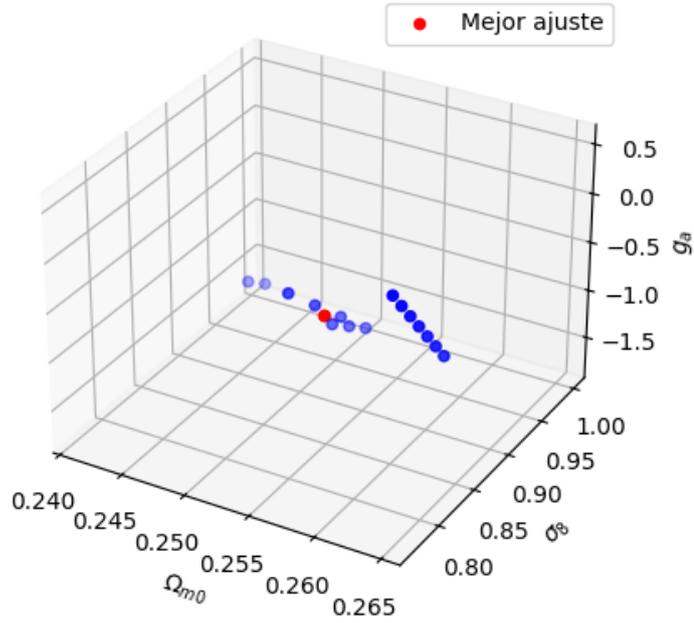


Figura 7.1: Puntos en el espacio de parámetros que cumplen la condición $14,25 < \chi^2 < 14,30$. En rojo se resalta al conjunto de parámetros que se eligió como de mejor ajuste debido a que representa el mínimo global de χ^2 .

En la figura 7.2 se muestran los tres gráficos bidimensionales relevantes para entender la distribución de puntos en el espacio de parámetros.

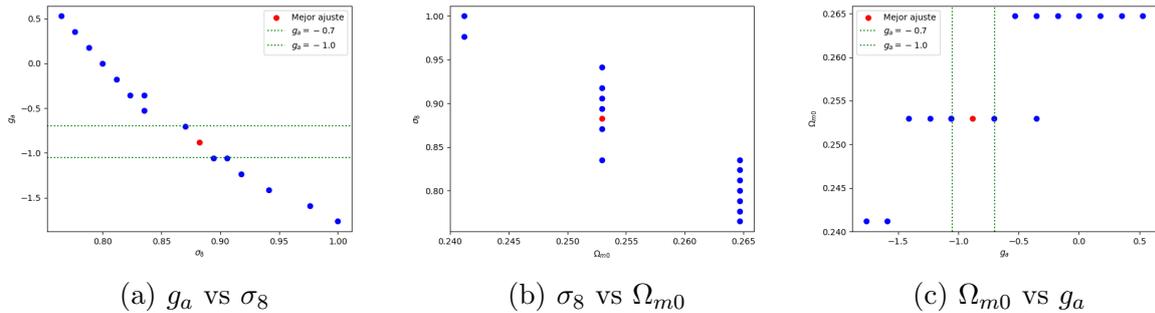


Figura 7.2: Cortes bidimensionales de la figura 7.1. En (a) y (c) se puede observar una delimitación entre el mínimo de mejor ajuste y los siguientes mínimos. De esta manera se definió el prior en g_a como el intervalo delimitado entre las dos líneas punteadas.

En las figuras 7.2a y 7.2c se puede ver marcada una delimitación entre el mínimo que corresponde a los parámetros de mejor ajuste, el cual corresponde al mínimo global, y los mínimos locales que le siguen. Se tomó un prior para g_a de manera tal de dejar dentro de dicho intervalo únicamente al mínimo de mejor ajuste. De esta forma, el prior empleado corresponde a

$$I_{g_a} = [-1, 0, -0, 7], \quad (7.1)$$

mientras que los priors correspondientes a Ω_{m0} y a σ_8 corresponden a los mencionados al final del capítulo de implementaciones (ec. 4.42), es decir

$$I_{\Omega_{m0}} = [0,1; 0,7] \quad I_{\sigma_8} = [0,6; 1,3]. \quad (7.2)$$

Con estas consideraciones, las cadenas de Markov correspondientes a este MCMC se pueden ver en la figura 7.3.

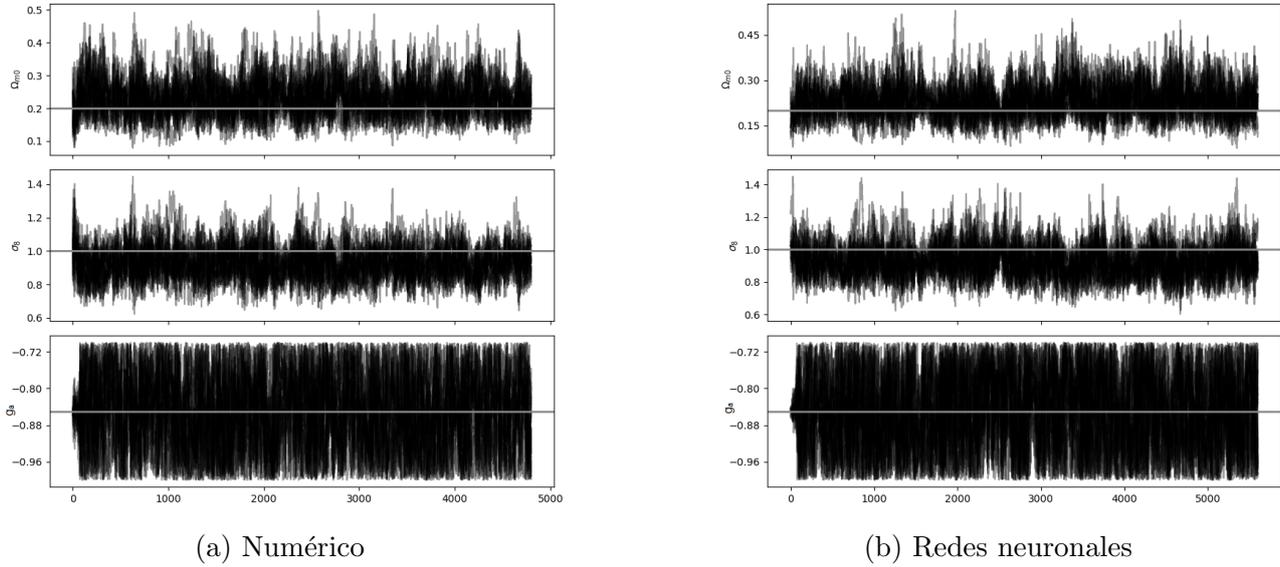


Figura 7.3: Cadenas de Markov obtenidas mediante el método de MCMC para ambos métodos en el modelo fenomenológico de gravedad modificada.

Se observa que las cadenas chocan contra el prior de g_a , lo cual es de esperar debido a la restricción que se impuso alrededor de un mínimo y la degeneración en este parámetro que presenta el modelo.

Bibliografía

- [1] Andrew Glassner. Deep learning, vol. 1: From basics to practice.
- [2] Savvas Nesseris, George Pantazis, and Leandros Perivolaropoulos. Tension and constraints on modified gravity parametrizations of $G_{\text{eff}}(z)$ from growth rate and planck data. *Phys. Rev. D*, 96:023542, Jul 2017.
- [3] Planck Collaboration. Planck 2018 results. vi. cosmological parameters. *Astronomy & Astrophysics*, 641:A6, 2020.
- [4] Edwin Hubble. A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Sciences*, 15:168–173, 1929.
- [5] Tamara M. Davis and Charles H. Lineweaver. Expanding confusion: Common misconceptions of cosmological horizons and the superluminal expansion of the universe. *Publications of the Astronomical Society of Australia*, 21(1):97–109, 2004.
- [6] A.A. Penzias and R.W. Wilson. A measurement of excess antenna temperature at 4080 mc/s. *The Astrophysical Journal*, 142:419, 1965.
- [7] D. J. Fixsen, E. S. Cheng, J. M. Gales, J. C. Mather, and R. A. Shafer. The cosmic microwave background spectrum from the full coBE FIRAS data set. *The Astrophysical Journal*, 473:576–587, 1996.
- [8] Fabian Schmidt Scott Dodelson. *Modern Cosmology*. Academic Press, San Diego, CA, 2021.
- [9] Adam G. Riess, Alexei V. Filippenko, Peter Challis, and et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal*, 116(3):1009–1038, 1998.
- [10] Steven Weinberg. The cosmological constant problem. *Reviews of Modern Physics*, 61(1):1–23, 1989.
- [11] Raphael Bousso. The cosmological constant. *General Relativity and Gravitation*, 40:607–637, 2008.
- [12] Sean M Carroll. The cosmological constant. *Living reviews in relativity*, 4(1):1–56, 2001.

- [13] Varun Sahni and Alexei Starobinsky. The case for a positive cosmological λ -term. *International Journal of Modern Physics D*, 9(04):373–443, 2000.
- [14] WJG De Blok, Fabian Walter, Elias Brinks, C Trachternach, SH Oh, and Robert C Kennicutt. High-resolution rotation curves and galaxy mass models from things. *The Astronomical Journal*, 136(6):2648, 2008.
- [15] P. J. E. Peebles. Cosmological consequences of a dark matter halo with a very cold dense component. *The Astrophysical Journal*, 263:L1–L5, 1982.
- [16] Gary Hinshaw, D Larson, Eiichiro Komatsu, David N Spergel, CLaa Bennett, Joanna Dunkley, MR Nolta, M Halpern, RS Hill, N Odegard, et al. Nine-year wilkinson microwave anisotropy probe (wmap) observations: cosmological parameter results. *The Astrophysical Journal Supplement Series*, 208(2):19, 2013.
- [17] E. Komatsu and et al. Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. *arXiv preprint arXiv:1001.4538*, 2011.
- [18] Joshua A. Frieman, Michael S. Turner, and Dragan Huterer. Dark energy and the accelerating universe. *Annual Review of Astronomy and Astrophysics*, 46:385–432, 2008.
- [19] S. Perlmutter and et al. Measurements of Omega and Lambda from 42 High-Redshift Supernovae. *arXiv preprint astro-ph/9812133*, 1999.
- [20] J.-P. Kneib and et al. The Mass Distribution of MS 2137-23 Derived from Joint Analysis of Strong Lensing, X-Ray and Sunyaev-Zel’dovich Effect Observations. *arXiv preprint astro-ph/0211483*, 2003.
- [21] R. Massey and et al. Dark matter maps reveal cosmic scaffolding. *arXiv preprint arXiv:1003.5594*, 2010.
- [22] D. Clowe and et al. A Direct Empirical Proof of the Existence of Dark Matter. *arXiv preprint astro-ph/0608407*, 2006.
- [23] D. J. Eisenstein, I. Zehavi, D. W. Hogg, and et al. Detection of the baryon acoustic peak in the large-scale correlation function of sdss luminous red galaxies. *The Astrophysical Journal*, 633:560–574, 2005.
- [24] D. N. Spergel, R. Bean, O. Doré, and et al. Three-year wilkinson microwave anisotropy probe (wmap) observations: Implications for cosmology. *The Astrophysical Journal Supplement Series*, 170:377–408, 2007.
- [25] Volker Springel, Simon DM White, Adrian Jenkins, Carlos S Frenk, Naoki Yoshida, Liang Gao, Julio Navarro, Robert Thacker, Darren Croton, John Helly, et al. Simulating the joint evolution of quasars, galaxies and their large-scale distribution. *Nature*, 435:629–636, 2005.

- [26] Mark Vogelsberger, Simon DM White, J S Bolton, C S Frenk, Doug Potter, Tom Abel, and Romain Teyssier. Properties of galaxies reproduced by a hydrodynamic simulation. *Nature*, 509:177–182, 2014.
- [27] Gary Steigman. Primordial nucleosynthesis: Successes and challenges. *International Journal of Modern Physics E*, 20(11):2249–2309, 2011.
- [28] Matías Leizerovich, Susana J. Landau, and Claudia G. Scóccola. Tensions in cosmology: a discussion of statistical tools to determine inconsistencies, 2023.
- [29] A. G. Riess, D. O. Jones, W. Yuan, and et al. Cosmic distances calibrated to 1 *The Astrophysical Journal*, 908(1):13, 2021.
- [30] Adam G. Riess, Wenlong Yuan, Lucas M. Macri, Dan Scolnic, Dillon Brout, Stefano Casertano, David O. Jones, Yukei Murakami, Gagandeep S. Anand, Louise Breuval, Thomas G. Brink, Alexei V. Filippenko, Samantha Hoffmann, Saurabh W. Jha, W. D’arcy Kenworthy, John Mackenty, Benjamin E. Stahl, and WeiKang Zheng. A comprehensive measurement of the local value of the hubble constant with 1 km s⁻¹ mpc⁻¹ uncertainty from the hubble space telescope and the sh0es team. *The Astrophysical Journal Letters*, 934(1):L7, July 2022.
- [31] Wendy L. Freedman. Measurements of the hubble constant: Tensions in perspective*. *The Astrophysical Journal*, 919(1):16, September 2021.
- [32] Adam G. Riess and Louise Breuval. The local value of h₀, 2023.
- [33] Nils Schöneberg, Guillermo Franco Abellán, Andrea Pérez Sánchez, Samuel J. Witte, Vivian Poulin, and Julien Lesgourgues. The olympics: A fair ranking of proposed models. *Physics Reports*, 984:1–55, October 2022.
- [34] Hermano ES Velten, RF Vom Marttens, and Winifried Zimdahl. Aspects of the cosmological “coincidence problem”. *The European Physical Journal C*, 74:1–8, 2014.
- [35] P Avelino, T Barreiro, CS Carvalho, A da Silva, FSN Lobo, P Martin-Moruno, JP Mimoso, NJ Nunes, D Rubiera-Garcia, D Saez-Gomez, et al. Unveiling the dynamics of the universe. *symmetry* 8 (8), 70 (2016). *arXiv preprint arXiv:1607.02979*.
- [36] Timothy Clifton, Pedro G. Ferreira, Antonio Padilla, and Constantinos Skordis. Modified gravity and cosmology. *Physics Reports*, 513(1):1–189, 2012. Modified Gravity and Cosmology.
- [37] Christian G Böhmer and Erik Jensko. Modified gravity: A unified approach. *Physical Review D*, 104(2):024010, 2021.
- [38] Gabriel R. Bengochea and Rafael Ferraro. Dark torsion as the cosmic speed-up. *Physical Review D*, 79(12), June 2009.

- [39] Rafael Ferraro and Franco Fiorini. Modified teleparallel gravity: Inflation without an inflaton. *Physical Review D*, 75(8), April 2007.
- [40] Raúl Carballo-Rubio, Luis J Garay, and Gerardo García-Moreno. Unimodular gravity vs general relativity: a status report. *Classical and Quantum Gravity*, 39(24):243001, 2022.
- [41] Kurt Hinterbichler. Theoretical aspects of massive gravity. *Reviews of Modern Physics*, 84(2):671, 2012.
- [42] Baojiu Li, John D Barrow, and David F Mota. Cosmology of modified gauss-bonnet gravity. *Physical Review D*, 76(4):044027, 2007.
- [43] Alexei A Starobinsky. Disappearing cosmological constant in $f(R)$ gravity. *JETP letters*, 86:157–163, 2007.
- [44] Wayne Hu and Ignacy Sawicki. Parametrized post-friedmann framework for modified gravity. *Physical Review D*, 76(10):104043, 2007.
- [45] Guido Cognola, E Elizalde, S Nojiri, SD Odintsov, L Sebastiani, and Sergio Zerbini. Class of viable modified $f(R)$ gravities describing inflation and the onset of accelerated expansion. *Physical Review D*, 77(4):046009, 2008.
- [46] E Elizalde, S Nojiri, SD Odintsov, Lorenzo Sebastiani, and Sergio Zerbini. Nonsingular exponential gravity: A simple theory for early-and late-time accelerated expansion. *Physical Review D*, 83(8):086006, 2011.
- [47] Augusto Tomás Chantada. Soluciones de la evolución tardía del universo con redes neuronales, 2022.
- [48] C. W. Smith and E. Abeysinghe. Proceedings of the practice and experience in advanced research computing 2017 on sustainability, success and impact. In *PEARC17*, New York, NY, USA, 2017. Association for Computing Machinery.
- [49] Oya Deniz Beygu, Istemi E. Dereli, and Zafer Koç. Photometric redshifts with artificial neural networks. *Monthly Notices of the Royal Astronomical Society*, 471(3):3494–3506, 2017.
- [50] Armin Rest, John-Paul N. White, Douglas L. Tucker, and et al. Deep-see: Convolutional neural networks for classifying irregularly and chaotically exploding stars. *The Astrophysical Journal*, 834(1):83, 2016.
- [51] Benedetta Vulcani, Giovanni Cresci, and et al. The hubble space telescope frontier fields: Predictions for mass distributions and observable properties of high-redshift galaxies. *Monthly Notices of the Royal Astronomical Society*, 460(3), 2016.
- [52] D. Gruen, J. B. Abdalla, and et al. Deep learning and curvelets for redshift estimation from multi-band photometry: A study in the dark energy survey. *Monthly Notices of the Royal Astronomical Society*, 465(2), 2017.

- [53] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.
- [54] Cedric Flamant, Pavlos Protopapas, and David Sondak. Solving differential equations using neural network solution bundles. *arXiv preprint arXiv:2006.14372*, 2020.
- [55] Edwin Hubble. A homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae. *Monthly Notices of the Royal Astronomical Society*, 84:665–673, 1924.
- [56] Ruth Durrer. The homogeneous and isotropic universe.
- [57] R. C. Tolman. Cosmological models with variable scale-factor and cosmological constant. *Physical Review*, 38:1758–1775, 1931.
- [58] E. Lifshitz and I. Khalatnikov. Investigations in relativistic cosmology. *Advances in Physics*, 12:185–249, 1963.
- [59] Mihran Vardanyan, Roberto Trotta, and Joseph Silk. How flat can you get? a model comparison perspective on the curvature of the universe. *Monthly Notices of the Royal Astronomical Society*, 397(1):431–444, 2009.
- [60] Christos G. Tsagas, Anthony Challinor, and Roy Maartens. Relativistic cosmology and large-scale structure. *Physics Reports*, 465(2):61–147, 2008.
- [61] Chung-Pei Ma and Edmund Bertschinger. Cosmological perturbation theory in the synchronous and conformal newtonian gauges. *arXiv preprint astro-ph/9506072*, 1995.
- [62] Marc Kamionkowski. Week 4: The growth of density perturbations, 2008. *Cosmology*, Ay127, Spring 2008.
- [63] R. L. Workman and Others. Review of Particle Physics. *PTEP*, 2022:083C01, 2022.
- [64] Alan H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Physical Review D*, 23:347–356, 1981.
- [65] H. A. Buchdahl. Non-Linear Lagrangians and Cosmological Theory. *Monthly Notices of the Royal Astronomical Society*, 150(1):1–8, 09 1970.
- [66] Álvaro de la Cruz-Dombriz and Antonio Dobado. $f(r)$ gravity without a cosmological constant. *Phys. Rev. D*, 74:087501, Oct 2006.
- [67] Shinji Tsujikawa. Matter density perturbations and effective gravitational constant in modified gravity models of dark energy. *Phys. Rev. D*, 76:023514, Jul 2007.
- [68] Jai-chan Hwang and Hyerim Noh. Cosmological perturbations in a generalized gravity including tachyonic condensation. *Phys. Rev. D*, 66:084009, Oct 2002.

- [69] Jai-chan Hwang and Hyerim Noh. Classical evolution and quantum generation in generalized gravity theories including string corrections and tachyons: Unified analyses. *Phys. Rev. D*, 71:063536, Mar 2005.
- [70] Wayne Hu and Ignacy Sawicki. Models of Λ CDM cosmic acceleration that evade solar system tests. *Physical Review D*, 76(6), September 2007.
- [71] Spyros Basilakos, Savvas Nesseris, and Leandros Perivolaropoulos. Observational constraints on viable Λ CDM parametrizations with geometrical and dynamical probes. *Physical Review D*, 87(12), June 2013.
- [72] P. Bessa, M. Campista, and A. Bernui. Observational constraints on starobinsky $f(r)$ cosmology from cosmic expansion and structure growth data. *The European Physical Journal C*, 82(6):506, 6 2022.
- [73] Pedro Domingos. A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78–87, 2012.
- [74] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 1097–1105, 2012.
- [75] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986.
- [76] Feiyu Chen, David Sondak, Pavlos Protopapas, Marios Mattheakis, Shuheng Liu, Devansh Agarwal, and Marco Di Giovanni. Neurodiffeq: A python package for solving differential equations with neural networks. *Journal of Open Source Software*, 5(46):1931, 2020.
- [77] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.
- [78] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 807–814, 2010.
- [79] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

- [80] Prajit Ramachandran, Barret Zoph, and Quoc V Le. Searching for activation functions. *arXiv preprint arXiv:1710.05941*, 2017.
- [81] Prajit Ramachandran, Barret Zoph, and Quoc V Le. Swish: a self-gated activation function. *arXiv preprint arXiv:1710.05941*, 2017.
- [82] Sashank J. Reddi, Satyen Kale, and Sanjiv Kumar. On the convergence of adam and beyond. In *International Conference on Learning Representations (ICLR)*, 2018.
- [83] Léon Bottou, Frank E. Curtis, Jorge Nocedal, and Richard H. Byrd. Optimization methods for large-scale machine learning. *SIAM Review*, 60(2):223–311, 2018.
- [84] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015.
- [85] Marios Mattheakis, David Sondak, Akshunna S. Dogra, and Pavlos Protopapas. Hamiltonian neural networks for solving equations of motion. *Phys. Rev. E*, 105:065305, Jun 2022.
- [86] Augusto T. Chantada, Susana J. Landau, Pavlos Protopapas, Claudia G. Scóccola, and Cecilia Garraffo. Cosmology-informed neural networks to solve the background dynamics of the universe. *Phys. Rev. D*, 107:063523, Mar 2023.
- [87] Claudia Graciela Scóccola. Variación de la constante de estructura fina y de la masa del electrón en el universo primitivo, 2009.
- [88] Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. The importance of the likelihood function in bayesian analysis. *Bayesian Analysis*, 1(3):369–376, 2006.
- [89] Steve Brooks, Andrew Gelman, Galin L. Jones, and Xiao-Li Meng. Markov chain monte carlo: Stochastic simulation for bayesian inference. *Statistical Science*, 18(3):335–358, 2003.
- [90] Persi Diaconis and Joseph B. Kadane. Markov chain monte carlo in practice: A roundtable discussion. *The American Statistician*, 49(2):131–144, 1995.
- [91] Jonathan Goodman and Jonathan Weare. Ensemble samplers with affine invariance. *Communications in applied mathematics and computational science*, 5(1):65–80, 2010.
- [92] Nick Kaiser. Clustering in real space and in redshift space. *Monthly Notices of the Royal Astronomical Society*, 227(1):1–21, 1987.
- [93] Eric V Linder and Adrian Jenkins. Cosmic structure growth and dark energy. *Monthly Notices of the Royal Astronomical Society*, 346(2):573–583, 2003.
- [94] Michael J Hudson and Stephen J Turnbull. The growth rate of cosmic structure from peculiar velocities at low and high redshifts. *The Astrophysical Journal Letters*, 751(2):L30, 2012.

- [95] L. Guzzo, M. Pierleoni, B. Meneux, E. Branchini, O.èvre@, C. Marinoni, B. Garilli, J. Blai-zot, G. De Lucia, A. Pollo, H. J. McCracken, D. Bottini, V. Le Brun, D. Maccagni, J. P. Picat, R. Scaramella, M. Scodreggio, L. Tresse, G. Vettolani, A. Zanichelli, C. Adami, S. Arnouts, S. Bardelli, M. Bolzonella, A. Bongiorno, A. Cappi, S. Charlot, P. Ciliegi, T. Contini, O. Cucciati, S. de la Torre, K. Dolag, S. Foucaud, P. Franzetti, I. Gavignaud, O. Ilbert, A. Iovino, F. Lamareille, B. Marano, A. Mazure, P. Memeo, R. Merighi, L. Moscardini, S. Paltani, R.ò@, E. Perez-Montero, L. Pozzetti, M. Radovich, D. Vergani, G. Zamorani, and E. Zucca. A test of the nature of cosmic acceleration using galaxy redshift distortions. , 451(7178):541–544, January 2008.
- [96] Arthur Kosowsky and Suman Bhattacharya. A future test of gravitation using galaxy cluster velocities. , 80(6):062003, September 2009.
- [97] Nick Kaiser. Clustering in real space and in redshift space. , 227:1–21, July 1987.
- [98] Phillip James Edwin Peebles. *Principles of physical cosmology*, volume 27. Princeton university press, 1993.
- [99] Dragan Huterer, Daniel Shafer, Daniel Scolnic, and Fabian Schmidt. Testing λ cdm at the lowest redshifts with sn ia and galaxy velocities. *arXiv preprint arXiv:1606.06268*, 2016.
- [100] Stephen J Turnbull, Michael J Hudson, Hume A Feldman, Malcolm Hicken, Robert P Kirshner, and Richard Watkins. Cosmic flows in the nearby universe from type ia super-novae. *The Astrophysical Journal*, 569(1):L19, 2002.
- [101] Michael J Hudson and Stephen J Turnbull. The growth rate of cosmic structure from peculiar velocities at low and high redshifts. *The Astrophysical Journal*, 496(2):L13, 1998.
- [102] Marc Davis, Adi Nusser, Karen Masters, Christopher Springob, John P Huchra, and Gerard Lemson. Local gravity versus local velocity: Solutions for β and nonlinear bias. *The Astrophysical Journal*, 703(2):L53, 2009.
- [103] Martin Feix, Adi Nusser, and Enzo Branchini. Local gravity versus local velocity: Solutions for β and nonlinear bias. *Physical Review Letters*, 115:011301, 2015.
- [104] Cullan Howlett, Ashley Ross, Lado Samushia, Will Percival, and Marc Manera. The clustering of the sdss main galaxy sample–ii. mock galaxy catalogues and a measurement of the growth of structure from redshift space distortions at $z = 0,15$. *Monthly Notices of the Royal Astronomical Society*, 449(1):848–869, 2015.
- [105] Yong-Seon Song and Will J Percival. Reconstructing the history of structure formation using redshift distortions. *Journal of Cosmology and Astroparticle Physics*, 2009(10):004, 2009.
- [106] Chris Blake and et al. Galaxy and mass assembly (gama): improved cosmic growth mea-surements using multiple tracers of large-scale structure. *Monthly Notices of the Royal Astronomical Society*, 462(4):4240–4261, 2016.

- [107] Lado Samushia, Will J Percival, and Alwise Raccanelli. Interpreting large-scale redshift-space distortion measurements. *Monthly Notices of the Royal Astronomical Society*, 439(4):3504–3520, 2014.
- [108] Ariel G Sanchez and et al. The clustering of galaxies in the sdss-iii baryon oscillation spectroscopic survey: cosmological implications of the full shape of the clustering wedges in the data release 10 and 11 galaxy samples. *Monthly Notices of the Royal Astronomical Society*, 464(2):1640–1655, 2017.
- [109] Chia-Hsun Chuang and et al. The clustering of galaxies in the sdss-iii baryon oscillation spectroscopic survey: single-probe measurements from cmass anisotropic galaxy clustering. *Monthly Notices of the Royal Astronomical Society*, 466(1):762–784, 2017.
- [110] Chris Blake and et al. The wigglez dark energy survey: Joint measurements of the expansion and growth history at $z < 1$. *Monthly Notices of the Royal Astronomical Society*, 415(3):2892–2909, 2011.
- [111] A. Pezzotta and et al. The vimos public extragalactic redshift survey (vipers): The growth of structures at $0,5 < z < 1,2$ from redshift-space distortions in the clustering of the pdr-2 final sample. 2016. arXiv preprint arXiv:1607.05380.
- [112] Teppei Okumura and et al. The subaru fmos galaxy redshift survey (fastsound). iv. new constraint on gravity theory from redshift space distortions at $z \sim 1,4$. *Publications of the Astronomical Society of Japan*, 68(3):38, 2016.
- [113] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman. emcee: The MCMC hammer. *Publications of the Astronomical Society of the Pacific*, 125:306, 2013.
- [114] J. Goodman and J. Weare. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, 5:65, 2010.
- [115] Antony Lewis. Getdist: a python package for analysing monte carlo samples, 2019.
- [116] Marco Raveri and Wayne Hu. Concordance and discordance in cosmology. *Physical Review D*, 99(4):043506, 2019.
- [117] Marco Raveri, Georgios Zacharegkas, and Wayne Hu. Quantifying concordance of correlated cosmological data sets. *Physical Review D*, 101(10):103527, 2020.

Tesis disponible bajo Licencia Creative Commons

Atribución – No Comercial – Compartir Igual (by-nc-sa) 2.5 Argentina

Buenos Aires, 2024