

INSTITUTO DE ASTRONOMÍA UNAM



TESIS DE MAESTRÍA

Optimización de sistemas Astrofísicos y Cosmológicos con Redes Neuronales Físicamente Informadas

Autor:

Juan de Dios Rojas Olvera

Supervisor:

José Alberto Vázquez
González

*Tesis presentada en cumplimiento de los requisitos
de la Maestría en Astrofísica*

Instituto de Astronomía CU

16 de agosto de 2024



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE POSGRADO EN ASTROFÍSICA

En la Universidad Nacional Autónoma de México, en el **INSTITUTO DE ASTRONOMÍA**, a las **11:00** horas del día 11 de diciembre de 2024, el alumno de nacionalidad **mexicana**

JUAN DE DIOS ROJAS OLVERA

se presentó con el fin de sustentar el examen para obtener el grado de

MAESTRO EN CIENCIAS (ASTROFÍSICA)

CONSTANCIA DE ACTA DE GRADO

No. de cuenta
313307579

en la modalidad de graduación por **TESIS**, con el trabajo titulado: **Optimización de sistemas Astrofísicos y Cosmológicos con Redes Neuronales Físicamente Informadas**, del cual fue tutor principal el **DR. JOSÉ ALBERTO VÁZQUEZ GONZÁLEZ**.

El alumno cursó sus estudios en el período **2023-1** a **2024-2**, obtuvo un promedio de **9.25** y cumplió con los requisitos académicos señalados en el plan de estudios **4196** aprobado por el Consejo Universitario.

El jurado resolvió otorgarle la calificación de

APROBADO

le hizo saber el resultado de su examen y le tomó la Protesta Universitaria.

El jurado estuvo integrado por:

- Presidente:** DR. JOSÉ ANTONIO DE DIEGO ONSURBE
- Secretaria:** DRA. ROSA LETICIA BECERRA GODÍNEZ
- Vocal 1:** DR. CHRISTOPHE ROGER JACQUES MORISSET
- Vocal 2:** DR. ISIDRO GÓMEZ VARGAS
- Vocal 3:** DR. SÉBASTIEN MICKAËL MARC FROMENTEAU



PRESIDENTE DEL JURADO



SECRETARIA DEL JURADO



VOCAL 1 DEL JURADO



VOCAL 2 DEL JURADO



VOCAL 3 DEL JURADO

El suscrito coordinador del programa constató que las firmas corresponden al jurado designado



"POR MI RAZA HABLARÁ EL ESPÍRITU"
DR. ARY RODRIGUEZ GONZALEZ
COORDINADOR DEL PROGRAMA

QR de verificación



No. Folio del Programa
4196/2024/18

Cadena de verificación digital

39fc28c307a29f7e3191a2a339b31380b71599b2b3a47511b3df5878176b5424b91f6e9f29f9a671220fc8dd10824cbb9c2cf213e03aa4df0da87913bf7c6f98

Declaración de Autoría

Yo, Juan de Dios Rojas Olvera declaro que esta tesis titulada, «Optimización de sistemas Astrofísicos y Cosmológicos con Redes Neuronales Físicamente Informadas» y el trabajo que se presenta en ella, son de mi autoría. Además declaro:

- Este trabajo se realizó total o principalmente durante mi candidatura para un título de investigación en esta Universidad.
- En caso de que alguna parte de esta tesis haya sido presentada previamente para un título u otra calificación en esta Universidad o cualquier otra institución, esto se ha declarado claramente.
- Siempre que he consultado el trabajo publicado de otros, se atribuye claramente.
- Cuando he citado el trabajo de otros, siempre se indica la fuente. Con la excepción de dichas citas, esta tesis es enteramente mi propio trabajo.
- He reconocido todas las fuentes principales de ayuda.
- Cuando la tesis se base en un trabajo realizado por mí junto con otros, he dejado claro exactamente qué hicieron los demás y cuál ha sido mi contribución.

Firma:

Fecha:

Agradecimientos

“Investigación realizada gracias al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) de la UNAM IN117723. Agradezco a la DGAPA-UNAM la beca recibida”.

INSTITUTO DE ASTRONOMÍA UNAM

Resumen

Optimización de sistemas Astrofísicos y Cosmológicos con Redes Neuronales Físicamente Informadas

Juan de Dios Rojas Olvera

Las Redes Neuronales Artificiales han revolucionado diversos campos, incluyendo la astrofísica y la cosmología. Sin embargo, la mayoría de paradigmas de Redes Neuronales tradicionales no permiten incorporar expresiones matemáticas en su entrenamiento, lo que limita su aplicación en modelos físicos. En este trabajo se propone el uso de Redes Neuronales Físicamente Informadas (PINNs por sus siglas en inglés) como una solución a esta limitante, abriendo así un nuevo abanico de posibilidades para el estudio de fenómenos astrofísicos.

En esta tesis se solucionó el sistema de ecuaciones diferenciales que describe la dinámica del modelo cosmológico $\Lambda - CDM$ utilizando una PINN, mostrando que estas redes pueden incorporar las condiciones iniciales como variables del modelo. A su vez se compararon dos métodos diferentes de entrenar PINN, encontrando una clara ventaja de uno sobre el otro.

También se utilizó una PINN para resolver la integral que define la distancia luminosa en un marco cosmológico de energía oscura paramétrica, obteniendo una solución precisa y computacionalmente eficiente. Esto permite hacer inferencia de parámetros con mayor rapidez en comparación con métodos numéricos convencionales, pues se redujo el tiempo de cómputo necesario en un orden de magnitud $\mathcal{O}(10)$.

Finalmente se construyó una PINN para resolver la ecuación de Klein-Gordon que gobierna la dinámica de Quintaesencia. Esta red permitió modelar la densidad de energía ρ_ϕ , la presión p_ϕ y la ecuación de estado del campo escalar asociado. A diferencia de enfoques tradicionales. La PINN se entrenó sin necesidad de condiciones iniciales al incluir datos observacionales de cronómetros cósmicos en su lugar, lo que elimina problemas asociados a su especificación y que son bien conocidos en la literatura, entregando una reconstrucción de $H(z)$ al mismo tiempo.

Índice general

Índice de figuras	VI
Índice de tablas	VIII
Lista de abreviaturas	IX
1. Introducción	1
2. Breviario cosmológico: las ecuaciones del Universo	4
2.1. Las ecuaciones de Friedmann	4
2.2. La ecuación de continuidad	9
2.3. Ω_i y la ecuación paramétrica de Friedmann	10
2.4. Distancias cosmológicas	11
2.5. Cronómetros cósmicos, supernovas e inferencia de parámetros	13
3. Modelos cosmológicos	17
3.1. Λ -CDM	17
3.1.1. Componentes del universo	18
3.2. Energía oscura paramétrica	20
3.2.1. Dinámica del modelo	21
3.3. Quintaesencia	22
3.3.1. Dinámica del campo escalar	23
4. Redes totalmente conectadas	25
5. Redes Neuronales Físicamente Informadas	28
5.1. Modelando funciones implícitas	28
5.2. Ecuaciones diferenciales	29
5.3. Condiciones iniciales y de frontera	30
5.3.1. Condiciones iniciales y de frontera como componentes del costo.	30
5.3.2. Parametrización	31
5.4. Datos observacionales	33
5.5. Integrales	34
5.6. Sistemas de ecuaciones diferenciales	35

5.7. Parciales	37
5.8. PINN dependiente de las condiciones iniciales o parámetros libres . . .	38
6. Metodología y resultados	41
6.1. Ecuación diferencial con un parámetro libre: reparametrización vs función de costo	41
6.2. Distancias cosmológicas en el modelo CPL: PINN para acelerar inferencia de parámetros	45
6.3. Ecuación de Klein-Gordon en Quintaesencia usando datos de cronómetros cósmicos	52
7. Discusión y conclusiones	58
8. Perspectiva a futuro	60
Referencias	61

Índice de figuras

4.1. Estructura general de una red totalmente conectada o MLP. Donde a^0 representa la capa de entrada de las variables del modelo, las capas ocultas a^l y la salida de la red $ANN(t; \theta)$ donde θ representa el conjunto de parámetros intrínsecos de la red: los pesos y sesgos. La función de activación se aplica a cada una de las capas.	26
5.1. Proceso de entrenamiento de una PINN.	32
5.2. Parametrización de la red para la solución de un sistema de ecuaciones diferenciales con condiciones iniciales y parámetros libres como variables del modelo	40
6.1. Mapas de densidad del error porcentual por cada modelo, bajo el enfoque de la función de costo. Cada una con dos capas ocultas, pero diferente número de nodos por capa. Cada punto del mapa es una combinación del corrimiento al rojo z y la condición inicial $\Omega_{m,0}$	44
6.2. Mapas de densidad del error porcentual por cada modelo, bajo el enfoque de parametrización de la PINN. Cada una con dos capas ocultas, pero diferente número de nodos por capa. Cada punto del mapa es una combinación del corrimiento al rojo z y la condición inicial $\Omega_{m,0}$	45
6.3. Algunas soluciones a la integral que define la distancia comóvil dadas por el modelo de la PINN para diferentes combinaciones de los parámetros libres ω_0, ω_a	47
6.4. Mapa de densidad de error para distintos valores del corrimiento al rojo z y los parámetros ω_0, ω_a	48
6.5. Cadenas resultantes de aplicar el algoritmo metrópolis hastings utilizando los datos de SH0ES, para inferir los parámetros ω_0, ω_a . Ambas cadenas convergieron a puntos mínimos similares, pero en diferentes tiempos.	50
6.6. comparación del ajuste utilizando ambos métodos a los datos del módulo de distancia frente al corrimiento al rojo.	51
6.7. Muestra de cómo una reconstrucción provista por un MLP, donde la red tiende a aprender también el comportamiento ruidoso de los datos.	52

6.8. Solución al sistema de ecuaciones 3.25 dado por la PINN en comparación con la solución numérica. También se muestra el error relativo del modelo dado por la red.	55
6.9. Evolución de la ecuación de estado asociada al campo escalar de quintaesencia, obtenida gracias al modelo provisto por la PINN.	56
6.10. Reconstrucción de la ecuación de Friedmann con base en los datos observacionales de cronómetros cósmicos y la directriz matemática de la ecuación de Klein-Gordon en el marco de Quintaesencia.	56

Índice de tablas

2.1. Cronómetros Cósmicos utilizados para complementar el entrenamiento de una PINN en un universo modelado por una cosmología Q -CDM.	15
6.1. Comparación del tiempo de cómputo	50

Lista de abreviaturas

MLP	Multi Layer Perceptron
PINN	Physics-Informed Neural Network
EDO	Ecuación Diferencial Ordinaria
CC	Cronómetros C3smicos
Λ-CDM	Λ-Cold Dark Matter
SNe Ia	Supernova tipo Ia

Capítulo 1

Introducción

A principios del siglo XXI, dos áreas aparentemente dispares experimentaron un auge sin precedentes: la cosmología observacional y el aprendizaje profundo. En cosmología, la próxima generación de misiones espaciales y observatorios promete arrojar luz sobre enigmas del universo como la distribución de la materia a gran escala, la naturaleza de la energía oscura y la tensión de Hubble. Por su parte, el aprendizaje profundo ha revolucionado diversos campos, desde la visión por computadora hasta el procesamiento del lenguaje natural, impulsando avances notables en el reconocimiento de patrones y la toma de decisiones automatizada.

La astrofísica y la cosmología se enfrentan a una avalancha de datos sin precedentes gracias a de observatorios modernos. Ejemplos notables incluyen el observatorio Vera C. Rubin (2024), que recopilará 20 terabytes de datos por noche, superando al Sloan Digital Sky Survey (2000-2010). El Square Kilometre Array (SKA) (2028) espera generar dos petabytes de datos diarios, mientras que el Very Large Array (ngVLA) (Nueva México) producirá cientos de petabytes anuales. El satélite Euclid (ESA) (2023) y el observatorio Simons (en construcción), entre otros.

En este contexto, el aprendizaje profundo surge como una herramienta indispensable para analizar y comprender esta avalancha de información, pues es menester implementar técnicas estadísticas y computacionales eficientes para procesar, interpretar y extraer la máxima información astrofísica y cosmológica de dichos datos.

El aprendizaje profundo ha mostrado ser un aliado poderoso para la ciencia. En particular, muchos estudios han probado que las redes neuronales son una herramienta poderosa y versátil cuando es utilizada en el tratamiento de datos astrofísicos y cosmológicos: ya sea en problemas de clasificación (Gharat y Dandawate, 2022) (Kang et al., 2023), en hidrodinámica (Dai et al., 2023) simulaciones de N -cuerpos (Chacón et al., 2023) (Chen et al., 2020a), en inferencia estadística (Jimenez-Mesa et al., 2023), en reconstrucción de funciones (Dios Rojas Olvera, Gómez-Vargas y Vázquez, 2022), entre muchos otros. Más aún, dado el crecimiento actual de la Inteligencia Artificial, es de esperar que nuevas aplicaciones sean descubiertas en el futuro que

sumen a todos los ámbitos científicos.

Para sintetizar, esta amalgama de disciplinas tiene un potencial enorme aún en espera de ser explotado, y uno de los nichos de estudio más interesantes que ha surgido en los últimos años es la resolución de ecuaciones diferenciales llevada a cabo utilizando redes neuronales artificiales, o como se les conoce más propiamente: **Redes Neuronales Físicamente Informadas** (o PINN por sus siglas en inglés).

Este tipo de redes ha sido ampliamente utilizado en una enorme cantidad de campos diferentes dentro de la física, incluyendo astrofísica (Moschou et al., 2023), (Martin y Schaub, 2022) y cosmología (Chantada et al., 2023), ofreciendo una alternativa a los métodos numéricos tradicionales y que es capaz de obtener modelos a partir de datos observacionales y además incluyendo información matemática que está detrás del comportamiento del fenómeno en cuestión.

Este tipo de redes es particularmente fascinante en las ciencias físicas debido a que la mayoría de los fenómenos son modelados al analizar cómo cambian con respecto a una variable independiente (generalmente el tiempo) por medio de una ecuación diferencial o mediante la resolución de una integral, y las PINN tienen la capacidad de resolver ambas.

Dicho lo anterior, es importante señalar que sólo para un pequeño grupo de ecuaciones diferenciales (Vuik et al., 2023, p. 1) e integrales (Forsythe, 1970) se cuenta con una solución analítica, es decir, una solución en términos de las funciones elementales conocidas. La basta mayoría de ecuaciones o integrales debe ser resuelta por medio de un método numérico con uso de recursos computacionales.

Si la solución necesita ser evaluada múltiples veces, el esfuerzo computacional requerido por el método numérico al ser aplicado sucesivamente, puede resultar en un tiempo demasiado grande para ser viable, por ejemplo para simulaciones hidrodinámicas de galaxias (Holdship et al., 2021) o al hacer inferencia de parámetros cosmológicos (Spurio Mancini et al., 2022).

Las redes neuronales profundas más simples son las conocidas como perceptrones multicapa (MLP, por sus siglas en inglés), un arreglo de neuronas organizadas en capas. Estas han demostrado poder reducir el tiempo computacional requerido para resolver sistemas de ecuaciones diferenciales incluso al ser evaluadas sobre grandes conjuntos de datos (Dios Rojas Olvera, Gómez-Vargas y Vázquez, 2022). Sin embargo, este método de resolver ecuaciones utilizando redes, pertenece al paradigma del aprendizaje supervisado, lo que implica que para que la red aprenda las soluciones, éstas le deben ser provistas desde el inicio, al obtenerlas con algún método numérico tradicional.

Las PINN tienen muchas ventajas y desventajas respecto a su versión MLP, que serán discutidas a profundidad en los siguientes capítulos. Pero en resumen, los

MLP son considerados *cajas negras* (Buhrmester, Münch y Arens, 2019), debido a que no proveen información física acerca de los datos que están modelando, lo cual es una de sus principales limitaciones para ser utilizadas para describir fenómenos físicos. En contraste, las PINN tienen un cimiento matemático que determina y guía su entrenamiento, ayudando a evitar el sobreajuste al mismo tiempo. Aunque hay otras alternativas para extraer información estadística de las redes neuronales, como las conocidas Redes Neuronales Bayesianas (Jospin et al., 2022), (Fan et al., 2008), este trabajo se centrará únicamente en las PINN.

En astrofísica y cosmología, la delimitación de parámetros mediante observaciones es crucial, lo que requiere técnicas estadísticas en modelos matemáticos que típicamente implican la solución de ecuaciones diferenciales y la reconstrucción de funciones a partir de datos observacionales. Las Redes Neuronales de Físicas Informadas se presentan como una alternativa prometedora a los métodos numéricos tradicionales para resolver estos sistemas, ya que, al ser un modelo matemático continuo, pueden ser manipuladas matemáticamente, derivadas, integradas y evaluadas múltiples veces. Esto las hace especialmente útiles en la inferencia de parámetros cosmológicos, sustituyendo métodos numéricos en procesos como la inferencia variacional o las cadenas de Markov Monte Carlo (MCMC).

El contenido de este trabajo se organiza de la siguiente manera: En el capítulo 2 se expondrán los principios cosmológicos y astrofísicos sobre los que se sustenta la teoría de esta tesis así como las ecuaciones más importantes; en el capítulo 3 se presentarán los tres modelos cosmológicos en los que se desarrollarán los tres problemas de investigación de la tesis; en el capítulo 4 se presentarán de forma breve los fundamentos de las redes neuronales artificiales; luego en el capítulo 5 se hará un amplio desarrollo de los fundamentos de las PINN y el tipo de problemas que pueden resolver; en el capítulo 6 se presentará la metodología aplicada a cada uno de los problemas de investigación que conforman la tesis y sus respectivos resultados; y finalmente en los capítulos 7 y 8 se presentarán las conclusiones y perspectiva a futuro respectivamente.

Todo el código utilizado en el desarrollo de esta tesis puede ser consultado en <https://github.com/JuanDDiosRojas/Master-Thesis.git> y utilizado libremente, se invita a la comunidad a probarlo y/o modificarlo.

Capítulo 2

Breviario cosmológico: las ecuaciones del Universo

El cosmos es todo lo que alguna vez fue, es o será ...

La cosmología es el estudio del universo en su forma más general posible. En otras palabras, la cosmología estudia el comportamiento y evolución de los fenómenos que ocurren a grandes escalas de distancia y de tiempo. Ahí reside la principal diferencia entre astrofísica y cosmología, pues la primera se enfoca en los objetos a pequeña escala (menos de 150 megaparsecs $\sim 3.086 \times 10^{26}$ cm) tales como estrellas, galaxias, o cúmulos de galaxias. Sin embargo, estas disciplinas trabajan de la mano, impulsándose una a otra: lo que se conoce sobre cosmología es gracias a la observación y estudio de los objetos astronómicos, y la cosmología provee un marco teórico que permite interpretar tales observaciones.

¿Por qué el universo es como es? ¿Cómo será dentro de millones de años? Y lo principal, ¿cuáles fueron sus características en el pasado?... incluso solo unos segundos después de su comienzo, ¿si es que tal comienzo ocurrió en primer lugar? Éstas son, en resumen, las preguntas que la cosmología trata de responder. Las observaciones de la estructura a gran escala del universo, la radiación cósmica de fondo, la expansión del universo o las lentes gravitacionales permiten a los cosmólogos encontrar el rango de parámetros para los modelos que gobiernan el comportamiento y la evolución del cosmos. De esa manera, la cosmología puede explicar por qué el universo se comporta como lo hace, inferir cómo era en el pasado y, en última instancia, hacer predicciones sobre su futuro. Sin embargo, antes de intentar responder a estas preguntas, es necesario establecer los principios físicos y matemáticos de esta ciencia.

2.1. Las ecuaciones de Friedmann

La geometría diferencial proporciona las herramientas matemáticas necesarias para estudiar las propiedades intrínsecas del universo. Esta rama de las matemáticas

estudia las propiedades de superficies y objetos curvos, como el espacio-tiempo. Los principios físicos que gobiernan el espacio-tiempo son proporcionados por la teoría de la relatividad especial y general de Einstein. Estas teorías describen cómo la gravedad y la curvatura del espacio-tiempo afectan el movimiento de la materia y la energía. De esta manera, la geometría diferencial y la relatividad de Einstein se convierten en parte fundamental de los cimientos de la cosmología.

Los paradigmas científicos se revolucionaron con las teorías de la relatividad especial y general, publicadas en 1905 y 1915, respectivamente. Estas teorías reemplazaron la concepción newtoniana del espacio y el tiempo como dos entidades absolutas, existentes independientemente de cualquier otra cosa. Los Principia de Newton establecen (Newton, 1687):

'El tiempo absoluto, verdadero y matemático, por sí mismo y por su propia naturaleza fluye equitativamente sin tener en cuenta nada externo...'

'El espacio absoluto, en su propia naturaleza, sin tener en cuenta nada externo, permanece siempre similar e inmóvil...'

En contraste, Einstein Einstein, 1916 mostró que el espacio y el tiempo son manifestaciones de un ente más general y abstracto: el espacio-tiempo. En relatividad general, Einstein propuso que la gravedad no es una fuerza, sino el resultado de los objetos moviéndose a través de trayectorias geodésicas (la trayectoria más corta entre dos puntos sobre una superficie) a lo largo del espacio-tiempo. La curvatura del espacio tiempo se genera por la masa, energía y el estrés contenido en la materia. Este comportamiento se puede resumir en el enunciado:

El espacio-tiempo curvado le dice a la materia cómo moverse; la materia le dice al espacio-tiempo cómo curvarse.

Este comportamiento se sintetiza de forma matemática en las famosas ecuaciones de campo de Einstein (Bonometto, Gorini y Moschella, 2002, pp 29):

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = -8\pi G T_{\mu\nu}, \quad (2.1)$$

donde se ha asumido $c = 1$, $G_{\mu\nu}$ es conocido como el tensor de Einstein, que da una descripción geométrica del espacio-tiempo; $T_{\mu\nu}$ es el tensor de energía-momento; $g_{\mu\nu}$, es el tensor asociado a la métrica del espacio-tiempo, y será detallado más abajo; Λ es una constante añadida por Einstein, denominada comúnmente como *constante cosmológica*; finalmente G es la constante de gravitación. A su vez, el tensor de Einstein se escribe en términos del tensor de Ricci $R_{\mu\nu}$ y el escalar de Ricci R :

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R. \quad (2.2)$$

Para desarrollar las ecuaciones de Einstein es crucial definir un concepto que permita medir ángulos y distancias entre dos puntos sobre una superficie, el espacio-tiempo en este caso. Este concepto es el *intervalo infinitesimal de línea*, denotado por ds^2 , éste se define en términos de los desplazamientos infinitesimales de tiempo y espacio como se muestra en la ecuación 2.3.

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu, \quad (2.3)$$

donde $g_{\mu,\nu}$ es una función del espacio-tiempo conocida como *la métrica* de la superficie y x^μ es el 4-vector de las coordenadas espacio-temporales, con una primera coordenada temporal denotada por 0 y tres componentes espaciales i, j .

Con el fin de construir un modelo cosmológico, ciertas suposiciones sobre la naturaleza del universo deben ser asumidas. La aproximación más simple es asumir que los principios de la relatividad determinan la evolución del cosmos a grandes escalas, y aceptar como válido el *principio cosmológico* (Milne, 1935). Este principio establece que el universo, a grandes escalas (más de 150 Mpc) es homogéneo e isotrópico.

Usando palabras más simples, la homogeneidad implica que las propiedades estadísticas del universo son iguales en cualquier punto, la isotropía indica que las propiedades observadas del universo serán estadísticamente iguales en cualquier dirección. Con estas dos hipótesis la ecuación que define al intervalo se vuelve más simple, y la forma más general para el intervalo infinitesimal de distancia para un espacio-tiempo homogéneo e isotrópico está dada por la ecuación 2.4 (Liddle, 2015),

$$ds^2 = g_{00}dt^2 + 2g_{0i}dtdx^i + g_{ij}dx^i dx^j. \quad (2.4)$$

Nótese que los términos fuera de la diagonal g_{0i} implicarían la existencia de una dirección espacial particular asociada al 3-vector de velocidad v_i . Por lo tanto, con el fin de satisfacer las hipótesis asumidas desde el inicio sobre la isotropía, estos términos de la métrica deben ser establecidos como nulos. Por otro lado, el elemento temporal de la métrica g_{00} puede ser establecido como $g_{00} = -1$ al fijar su valor como el tiempo propio medido en el marco de referencia de los observadores. Esto permite escribir el intervalo en términos de la suma de dos componentes: la temporal dt^2 y la espacial $-dl^2$, como se indica en la ecuación 2.5.

$$ds^2 = dt^2 - g_{ij}dx^i dx^j = dt^2 - dl^2, \quad (2.5)$$

donde dl^2 es la métrica 3-espacial homogénea e isotrópica. Por lo anterior ésta cuenta también con simetría esférica, y por tanto puede ser escrita en términos de las coordenadas esféricas (r, ϕ, θ) como lo indica la ecuación 2.6.

$$dl^2 = a^2 [\lambda^2(r)dr^2 + r^2(d\theta^2 + \sin^2(\theta)d\phi^2)], \quad (2.6)$$

a es conocido como *el factor de escala adimensional*, derivado de considerar la posibilidad de un espacio-tiempo no estático o en expansión (o contracción). Este factor relaciona el cambio entre la distancia propia D entre dos puntos a un tiempo dado t con respecto a la distancia D_0 medida a un tiempo de referencia t_0 , por medio de la ecuación 2.7.

$$D(t) = a(t)D(t_0) = a(t)D_0, \quad (2.7)$$

cabe destacar que, cuando $t = t_0, a = 1$. Por otro lado, λ es una función de r , y cuyos valores posibles pueden ser calculados a través del escalar tres-dimensional de curvatura \mathcal{R} , definido por la ecuación 2.8.

$$\mathcal{R} = \frac{3}{2a^2r^3} \frac{d}{dr} \left[r^2 \left(1 - \frac{1}{\lambda^2} \right) \right]. \quad (2.8)$$

Asumiendo un espacio tiempo homogéneo, la variación de la curvatura con respecto a la posición r no puede ser posible. En consecuencia \mathcal{R} se constriñe a ser constante. Entonces la ecuación 2.8 puede ser igualada a una constante y luego integrada, lo que lleva a la ecuación 2.9.

$$r^2 \left(1 - \frac{1}{\lambda^2} \right) = c_1 r^4 + c_2, \quad (2.9)$$

donde c_1, c_2 son constantes de integración. Además, con el fin de evitar indeterminaciones sobre el término λ cuando $r \rightarrow 0$, c_2 se establece como $c_2 = 0$, por lo que la ecuación 2.9 puede ser escrita como:

$$\lambda = \frac{1}{1 - c_1 r^2}.$$

La constante c_1 se puede remplazar por el escalar k (llamado comúnmente como *la curvatura*). Es posible reescalar r de tal forma que la curvatura tome sólo los valores $k = -1, 0, 1$. Ahora, en la expresión para λ , la métrica resultante de asumir homogeneidad, isotropía y un espacio-tiempo no estático, se puede escribir como en la ecuación 2.10

$$ds^2 = dt^2 - a^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2 \theta d\phi^2) \right]. \quad (2.10)$$

La ecuación 2.10 se conoce como la métrica *The Friedmann-Lemaître-Robertson-Walker (FLRW)*, en honor a las personas que encontraron las soluciones a las ecuaciones de campo de Einstein para un universo con estas características: Friedman, 1922 y Lemaître, 1927. Más tarde, en 1929, basado en el trabajo de Robertson y Walker, las soluciones homogéneas e isotrópicas de las ecuaciones de Einstein salieron a la luz (Jones, 1997). A partir de la ecuación 2.10, las componentes del tensor métrico están dadas por las siguientes expresiones:

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{-a(t)^2}{1-kr^2} & 0 & 0 \\ 0 & 0 & -a(t)^2 r^2 & 0 \\ 0 & 0 & 0 & -a(t)^2 r^2 \sin^2 \theta \end{pmatrix}. \quad (2.11)$$

Haciendo las sustituciones y operaciones correspondientes, las componentes no nulas del tensor de Ricci pueden ser obtenidas:

- $R_{tt} = -3\frac{\ddot{a}}{a}$,
- $R_{rr} = \frac{a\ddot{a}}{1-kr^2} + \frac{2\dot{a}^2}{1-kr^2} + \frac{2k}{1-kr^2}$,
- $R_{\theta\theta} = r^2 a\ddot{a} + 2r^2 \dot{a}^2 + r^2 k$,
- $R_{\phi\phi} = r^2 a\ddot{a} + 2r^2 \dot{a}^2 + 2kr^2$.

Y el valor para el escalar de Ricci:

$$R = -6 \left(\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a} \right)^2 + \frac{k}{a^2} \right). \quad (2.12)$$

Ahora es posible determinar el lado izquierdo de las ecuaciones de Einstein, la parte geométrica. El tensor de energía momento depende del contenido de energía del universo, como una consecuencia de asumir un espacio-tiempo isotrópico, éste será descrito por un tensor de energía momento que sólo dependerá de ρ y de p , la densidad de energía y presión de las componentes del universo. Entonces, este tensor será también una matriz diagonal, cuyas entradas no nulas son:

$$T_{tt} = \rho g_{tt},$$

$$T_{ii} = -p g_{ii}.$$

Sustituyendo estos términos en las ecuaciones de Einstein, se obtiene la ecuación 2.13.

$$R_{tt} - \frac{1}{2} R g_{tt} - \Lambda g_{tt} = 8\pi G \rho g_{tt}, \quad (2.13)$$

lo que es equivalente a:

$$3 \left(\frac{\dot{a}}{a} \right)^2 + \frac{3k}{a^2} - \Lambda = 8\pi G\rho, \quad (2.14)$$

y cuyo desarrollo culmina con:

$$\left(\frac{\dot{a}}{a} \right)^2 = \frac{\kappa_0}{3}\rho - \frac{k}{a^2(t)} + \frac{\Lambda}{3}, \quad (2.15)$$

donde $\kappa_0 = 8\pi G$. Para las tres ecuaciones correspondientes a las componentes espaciales, en cada caso, se obtiene la misma expresión:

$$\frac{-g_{ii}}{a^2(t)} (a\ddot{a} + 2\dot{a}^2 + 2k) - \frac{1}{2}Rg_{ii} - \Lambda g_{ii} = 8\pi G(-p)g_{ii}. \quad (2.16)$$

Donde, nuevamene, el desarrollo de operaciones culmina con:

$$\left(\frac{\ddot{a}}{a} \right) = -\frac{\kappa_0}{6}(\rho + 3p) + \frac{\Lambda}{3}. \quad (2.17)$$

Algunas de estas expresiones son simplificadas considerando a la constante cosmológica Λ como una densidad de un fluido (eso será detallado más adelante), e incluyéndola dentro de la expresión de las demás densidades ρ . Por otro lado, el parámetro de Hubble se define en términos del factor de escala como $H \equiv \frac{\dot{a}}{a}$, este término indica el cambio en la distancia propia entre dos objetos (galaxias o cúmulos de galaxias) con respecto del tiempo, entonces las ecuaciones 2.15, 2.17 se pueden escribir como:

$$H^2 = \frac{\kappa_0}{3}\rho - \frac{k}{a^2(t)}, \quad (2.18)$$

$$\dot{H} + H^2 = \left(\frac{\ddot{a}}{a} \right) = -\frac{\kappa_0}{6}(\rho + 3p). \quad (2.19)$$

Las ecuaciones 2.18, 2.19 son mejor conocidas como *las ecuaciones de Friedmann*.

2.2. La ecuación de continuidad

Las ecuaciones de Friedmann 2.18, 2.19 son independientes y es posible obtener una tercera ecuación al derivar la ecuación 2.18 respecto al tiempo y despejando \dot{H} se obtiene:

$$\dot{H} = \frac{1}{2H} \left(\frac{\kappa_0}{3}\dot{\rho} + \frac{2\dot{a}}{a^3} \right).$$

Esta expresión puede insertarse en la ecuación 2.19, entonces:

$$\frac{1}{2H} \left(\frac{\kappa_0}{3} \dot{\rho} + \frac{2\dot{a}}{a^3} \right) + H^2 = -\frac{\kappa_0}{6}(\rho + 3p).$$

Nuevamente, el término H^2 puede sustituirse utilizando la ecuación 2.18, luego se multiplican ambos lados de la ecuación por $2H$. Desarrollando los términos que queden, se llega fácilmente a la expresión:

$$\dot{\rho} + 3H(\rho + p) = \nabla_{\mu} T^{\mu\nu} = 0, \quad \mu = 0. \quad (2.20)$$

Este resultado se conoce como la *Ecuación de continuidad* o *Ecuación de fluido*, que describe el comportamiento y la evolución del contenido del Universo en función de la densidad y la presión. También suele llamarse la ecuación de conservación de la energía, pues se puede deducir con base en el cambio de energía interna del universo a través de la primera ley de la termodinámica. La ecuación de continuidad es crucial si se quiere inferir cómo se comportará el cosmos en el futuro lejano, o por el contrario, cómo se ha ido transformando desde que ocurrió el Big-Bang. Dicho análisis se puede hacer para cada uno de los componentes ρ_i .

2.3. Ω_i y la ecuación paramétrica de Friedmann

Como se mencionó en la sección 2.1, el valor del parámetro de curvatura k indica el tipo de geometría que toma el espacio-tiempo: Si $k = 0$, se tiene que la curvatura es la de un espacio-tiempo plano; para $k < 0$, se dice que el universo es abierto con geometría hiperbólica; finalmente si $k > 0$, el universo es cerrado y tiene geometría esférica. La curvatura, la densidad de energía del contenido del universo y con el parámetro de expansión están relacionados a través de 2.18. Así, cuando el espacio-tiempo es geoméricamente plano, esta ecuación toma la forma de la ecuación 2.21.

$$H^2 = \frac{\kappa_0}{3}\rho, \quad (2.21)$$

por lo que para cualquier valor de H es posible encontrar una densidad especial para la cuál la curvatura sea cero, este valor es conocido como la *densidad crítica*:

$$\rho_c = \frac{3H^2}{\kappa_0}. \quad (2.22)$$

$$\Omega_i = \frac{\rho}{\rho_c} = \frac{\kappa_0}{3H^2}\rho, \quad (2.23)$$

A partir de la definición de los parámetros adimensionales de densidad (Ecuación 2.23) se puede escribir la ecuación de Friedmann 2.18 como en 2.24,

$$H^2 = \Omega H^2 - \frac{k}{a^2}. \quad (2.24)$$

Lo que permite expresar a k en términos de los parámetros de densidad y de H , al evaluarlos en la actualidad se obtiene:

$$k = H_0^2 (\Omega_0 - 1) a_0^2. \quad (2.25)$$

Así que, con ayuda de la expresión 2.25 se puede escribir la ecuación 2.15 sin depender del parámetro de curvatura de forma explícita:

$$H^2 = \frac{\kappa_0}{3} \rho - \frac{a_0^2}{a^2} H_0^2 (\Omega - 1), \quad (2.26)$$

recordando que el parámetro de expansión adimensional a vale 1 para la época actual y haciendo algunos despejes se obtiene una nueva expresión para la ecuación de Friedmann conocida como la *Forma paramétrica*:

$$\frac{H^2}{H_0^2} = \frac{\rho}{\rho_{c,0}} + \frac{1 - \Omega_0}{a^2}, \quad (2.27)$$

2.4. Distancias cosmológicas

La importancia de la ecuación de Friedmann radica en que es parte de un sistema acoplado de ecuaciones diferenciales a partir del cuál se puede inferir la evolución del universo y su contenido a través del tiempo, pero también describe la expansión del espacio-tiempo y por ende relaciona la velocidad a la que se alejan las galaxias debido una de otra, en función del parámetro a . Por lo que si se tienen estimaciones de H a partir de observaciones, es posible entonces delimitar los parámetros que definen a H de forma teórica. El problema es que el parámetro a no es una observable astrofísica, sin embargo es posible definir a la ecuación de Friedmann en términos de una variable nueva que sí es una observable: el corrimiento al rojo cosmológico z .

La expansión del universo influye sobre la radiación por medio del (redshift) cosmológico, un fenómeno que hace que la luz tienda hacia longitudes de onda más grandes para la luz observada entre dos puntos en función de la distancia entre ellos, el tiempo en que fue emitida y en el que se observa. Este efecto se puede describir con la relación:

$$\frac{\lambda_0}{\lambda} = \frac{a_0}{a}, \quad (2.28)$$

con λ_0 la longitud apreciada y λ la que fue emitida originalmente. Además, el corrimiento al rojo z se define como:

$$z = \frac{\lambda_0 - \lambda}{\lambda}.$$

Juntando ambas expresiones, se obtiene que el corrimiento al rojo en términos del parámetro de expansión es:

$$1 + z = \frac{a_0}{a}. \quad (2.29)$$

Entonces, la ecuación 2.27 se escribe en términos de esta observable como:

$$H(z) = H_0 \sqrt{\frac{\rho(z)}{\rho_{c,0}} + (1 - \Omega_0)(1 + z)^2}. \quad (2.30)$$

Usualmente los términos de la contribución de energía del universo se agrupan en el término $E(z)$, definido por

$$E(z) \equiv \sqrt{\frac{\rho(z)}{\rho_{c,0}} + (1 - \Omega_0)(1 + z)^2},$$

conocido como el parámetro de Hubble adimensional, y a partir de este término se define la *distancia comovil* d_c o distancia entre objetos que no cambia con la expansión del universo, dada por la ecuación 2.31.

$$d_c = d_H \int_0^z \frac{dz'}{E(z')}, \quad (2.31)$$

donde $d_H = \frac{1}{H_0}$ es conocido como el *radio de Hubble*, que mide el radio del universo observable al día de hoy. La *distancia comovil transversa* es la distancia entre dos eventos en el mismo corrimiento al rojo o distancia, pero separados en el cielo por algún ángulo δ_θ , para un universo con geometría plana, esta se define como la ecuación 2.32.

$$d_M = d_c. \quad (2.32)$$

Es posible obtener la separación angular entre los objetos astronómicos de una imagen obtenida de un telescopio y luego obtener la distancia real a la que se encuentran (distancia propia). Para hacer eso se utiliza la cantidad que relaciona el diámetro angular transversal de un objeto con su tamaño angular, es decir su distancia angular d_A con su distancia comovil transversal d_M , dada por la relación:

$$d_A = \frac{d_M}{(1 + z)} = \frac{d_H}{(1 + z)} \int_0^z \frac{dz'}{E(z')}. \quad (2.33)$$

Finalmente está la distancia luminosa, que permite relacionar cantidades astrofísicas muy importantes, como el flujo bolométrico S y la luminosidad bolométrica L por medio de la ecuación 2.34:

$$d_L \equiv \sqrt{\frac{L}{4\pi S'}} \quad (2.34)$$

y además se relaciona con el módulo de distancia μ , es la diferencia de magnitud entre el flujo bolométrico observado de un objeto y lo que sería si estuviera en 10 pc, por medio de la ecuación 2.35.

$$\mu(z) = 25 + 5 \log_{10} d_L(z). \quad (2.35)$$

La distancia luminosa d_L se define en términos de las otras distancias cosmológicas como:

$$d_L = (1+z)d_M = (1+z)^2 d_A = (1+z) \int_0^z \frac{dz'}{E(z')}. \quad (2.36)$$

Todas las distancias cosmológicas aquí mencionadas dependen de la integral del factor de Hubble adimensional. Como se mostrará en el siguiente capítulo, esta integral tiene una forma analítica y cerrada bajo ciertas condiciones sobre el contenido del universo. Pero en los casos generales, la integral sólo puede ser resuelta de forma numérica.

2.5. Cronómetros cósmicos, supernovas e inferencia de parámetros

El parámetro de Hubble H es posiblemente la cantidad física más importante para los cosmólogos, pues permite obtener información muy importante acerca del universo. A partir del parámetro de Hubble se puede inferir la edad del universo, el radio del universo observable y la velocidad de expansión del universo, por supuesto.

Una de las formas más populares para estudiar a la expansión del universo parte de un hecho matemático que se deriva de la propia definición del parámetro de Hubble dada por la ecuación 2.37.

$$H \equiv \frac{\dot{a}}{a}. \quad (2.37)$$

H puede escribirse entonces en términos del corrimiento al rojo a partir de la relación 2.29 y aplicando regla de la cadena, lo que lleva a la ecuación 2.38,

$$H(z) = -\frac{1}{1+z} \frac{dz}{dt}. \quad (2.38)$$

La expresión dt puede entenderse como el envejecimiento diferencial del universo en función del corrimiento al rojo dz . Con el fin de determinar H es posible proponer una aproximación del tipo 2.39,

$$\frac{\Delta z}{\Delta t} \sim \frac{dz}{dt}, \quad (2.39)$$

donde Δz es determinado para una población homogénea de objetos astronómicos que ayude a trazar Δt , una población que permita hacer eso se le conoce como *cronómetros cósmicos* (Borghi, Moresco y Cimatti, 2022).

Los primeros en proponer el uso de CC fueron (Jimenez y Loeb, 2002), quienes propusieron el recurrir a galaxias masivas con evolución pasiva, con corrimientos al rojo altos $z \gtrsim 2.0$ y escalas de tiempo cortas ($< \text{Gyr}$), lo que mantiene la condición 2.39. Sin embargo aunque el corrimiento al rojo z se puede determinar con gran precisión utilizando espectroscopía, determinar la edad de las galaxias es un reto importante, debido a dos factores principales: la complejidad que implica reconstruir su historia de formación estelar y a distintas degeneraciones en los parámetros de dicha población como la composición química, la edad estelar o el tiempo de escala en la formación de la población (Conroy, 2013).

A pesar de los problemas que tiene determinar la edad del universo, la evidencia empírica indica que las galaxias que evolucionan pasivamente y en una escala de tiempo mucho mayor que la diferencia de su edad son los mejores objetos a utilizar como cronómetros cósmicos (Melia y Maier, 2013). En particular las galaxias que se encuentran dentro de cúmulos, en regiones densas y cuya población estelar se formó en $z \gtrsim 2.0$ y han evolucionado pasivamente y sin nuevos brotes de formación estelar.

La importancia de los cronómetros cósmicos reside en que determinar $H(z)$ a partir de ellos no requiere de una cosmología subyacente, puesto que las hipótesis para obtener la ecuación 2.38 son sólo las de homogeneidad e isotropía en un universo descrito por la métrica *FLRW*.

Sin embargo, en este trabajo se va a presentar cómo el entrenamiento de una PINN puede ser complementado al proveerle datos observacionales (cronómetros cósmicos en este caso), por lo que sí se asumirá un universo descrito por una cosmología Q -CDM (descrito en la sección 3.3). Se decidió utilizar cronómetros cósmicos debido a que proveen una relación para $H(z)$ directa y eso es lo que se requiere para el problema de investigación descrito en la sección 6.3. Los CC utilizados en este trabajo y sus respectivas fuentes están ordenados en la tabla 2.1.

Por otro lado, el estudio de la expansión del universo se ha estudiado a través de diferentes eventos astronómicos: como las lentes gravitacionales, las oscilaciones acústicas de bariones o el fondo cósmico de microondas. Sin embargo, una de las maneras más importantes para estudiarla es a través de las supernovas tipo Ia, pues

TABLA 2.1: Cronómetros Cómicos utilizados para complementar el entrenamiento de una PINN en un universo modelado por una cosmología Q -CDM.

z	$H \pm \sigma_H$ (CCB)	$H \pm \sigma_H$ (CCM)	$H \pm \sigma_H$ (CCH)	Ref.
0.07	69 ± 19.6	Zhang et al., 2014
0.09	69 ± 12	Stern et al., 2010
0.12	68.6 ± 26.2	Zhang et al., 2014
0.17	83 ± 8	Stern et al., 2010
0.1797	75 ± 4	Moresco et al., 2012
0.1993	$75 \pm$	Moresco et al., 2012
0.2	72.9 ± 29.6	Zhang et al., 2014
0.27	77 ± 14	Stern et al., 2010
0.28	88.8 ± 36.6	Zhang et al., 2014
0.3519	83 ± 14	88 ± 16	...	Moresco et al., 2012
0.3802	83 ± 13.5	89.2 ± 14.1	...	Moresco et al., 2016
0.4	95 ± 17	Stern et al., 2010
0.4004	77 ± 10.2	82.8 ± 10.6	...	Moresco et al., 2016
0.4247	87.1 ± 11.2	93.7 ± 11.7	...	Moresco et al., 2016
0.4497	92.8 ± 12.9	99.7 ± 13.4	...	Moresco et al., 2016
0.47	89 ± 49.6	Ratsimbazafy:2017vga
0.4783	80.9 ± 9	86.6 ± 8.7	...	Moresco et al., 2016
0.48	97 ± 62	Stern et al., 2010
0.5929	104 ± 13	110 ± 15	...	Moresco et al., 2012
0.6797	92 ± 8	98 ± 10	...	Moresco et al., 2012
0.7812	105 ± 12	88 ± 11	...	Moresco et al., 2012
0.8754	125 ± 17	124 ± 17	...	Moresco et al., 2012
0.88	90 ± 40	Stern et al., 2010
0.9	117 ± 23	Stern et al., 2010
1.037	154 ± 20	113 ± 15	...	Moresco et al., 2012
1.3	168 ± 17	Stern et al., 2010
1.363	160 ± 33.6	160 ± 33.6	...	Moresco, 2015
1.43	177 ± 18	Stern et al., 2010
1.53	140 ± 14	Stern et al., 2010
1.75	202 ± 40	Stern et al., 2010
1.965	186.5 ± 50.4	186.5 ± 50.4	...	Moresco, 2015

fue el estudio de estos fenómenos los que mostraron la expansión acelerada del universo (Riess et al., 1998), (Perlmutter et al., 1999).

Se considera una supernova tipo Ia (SNe Ia) o también llamada supernova nuclear, a la explosión de una enana blanca perteneciente a un sistema binario. Aún no son del todo claros algunos aspectos de este tipo de supernovas, por ejemplo : el tipo espectral de la compañera de la enana blanca (si se trata de una gigante roja u otra enana blanca), el método que se lleva a cabo la transferencia de materia de una a otra, o la densidad de la enana blanca cuando ocurre la explosión (Ruiter, 2019).

A pesar de algunas incógnitas, el uso de SNe Ia es de suma importancia en Astrofísica y Cosmología debido a que son eventos tan brillantes que pueden ser observados a grandes distancias cosmológicas, ocurren con la suficiente frecuencia

como para tener un número grande de datos o que cada una de estas pueden ser estandarizadas con una precisión de ~ 0.1 mag en su brillo o $\sim 5\%$ en distancia (Brout et al., 2022).

Las SNe Ia permiten determinar las distancias cosmológicas gracias a sus luminosidades, el observable que se obtiene a partir de ellas se conoce como el módulo de distancia μ cuya obtención se hace utilizando la relación dada por (Tripp, 1998) y que está determinada por la ecuación 2.40.

$$\mu = m_b - M + \alpha x_1 - \beta c + \Delta M + \Delta B, \quad (2.40)$$

donde c es el color, x_1 el parámetro de la forma de la curva de luz, m_b el log de todo el flujo de normalización, α la relación entre la luminosidad y el estiramiento astronómico, β el coeficiente que relaciona la luminosidad y el color, M la magnitud absoluta en la banda B , ΔB la corrección de distancia basada en simulaciones y ΔM la corrección de distancia basada en la masa de la galaxia anfitriona.

La predicción teórica para $\mu(z)$ es provista por la ecuación del módulo de distancia 2.35 por lo que se puede hacer una estimación de parámetros utilizando las observaciones de SNe Ia y la teoría. En este trabajo se utilizarán los datos provenientes de la combinación de diferentes muestras de supernovas, detalladas en (Scolnic et al., 2018), y de las cuales fueron reducidas a 42 distancias agrupados para distintos intervalos de z .

Dado un vector $\vec{\mu}_{obs}$ que contiene la muestra de estas supernovas, se define $\Delta\vec{\mu}$ en la ecuación 2.41,

$$\Delta\vec{\mu} = \vec{\mu}(z) - \vec{\mu}_{obs}, \quad (2.41)$$

donde $\vec{\mu}(z)$ es la predicción teórica del módulo de distancia dada una cosmología.

El proceso de inferencia consideró como función a optimizar la función de máxima verosimilitud asociada a los datos, definida por la ecuación 2.42.

$$L(\Theta) \propto \exp\left(-\frac{1}{2} \Delta\vec{\mu}^T C^{-1} \Delta\vec{\mu}\right) \quad (2.42)$$

donde C^{-1} es la inversa de la matriz de covarianza asociada a los datos y Θ son los parámetros a inferir. Para simplificar las operaciones, usualmente a la expresión 2.42 se le aplica el logaritmo natural y se multiplica por -1 .

$$-\ln\{L(\Theta)\} = \frac{1}{2} \Delta\mu^T C^{-1} \Delta\mu. \quad (2.43)$$

Así, para inferir los parámetros Θ del modelo cosmológico de interés, la función 6.14 debe ser maximizada.

Capítulo 3

Modelos cosmológicos

3.1. Λ -CDM

Aunque existían soluciones a las ecuaciones de Einstein que permitían un universo en expansión, esta idea no obtuvo demasiada atención en su momento. Fue hasta 1929 cuando Edwin Hubble observó nebulosas extragalácticas y mostró que de hecho el universo se encuentra en expansión, este hecho marcó el nacimiento de la cosmología moderna como una ciencia rigurosa y observacional (Hubble, 1929). Así pues, las observaciones determinarán qué modelos son más o menos plausibles con base en su capacidad de reproducir los datos, pero también con la posibilidad de hacer predicciones.

El modelo actual más famoso y aceptado es el conocido como *Lambda-Cold Dark Matter* Λ -CDM, o modelo cosmológico estándar. El grado de aceptación de este modelo se debe principalmente a dos razones: La primera es por ser el más sencillo en términos de los parámetros necesarios para describirlo y por las ecuaciones que tiene asociadas, puesto que asume un universo homogéneo e isotrópico, la métrica de la que parte es la *FLRW* y sus ecuaciones las descritas en el capítulo 2. La segunda razón y la más importante es que es el modelo que mejor describe a las observaciones cosmológicas con las que se cuenta hoy por hoy:

- Fondo Cósmico de Microondas (CMB): Este modelo predice con precisión la anisotropía a gran escala del CMB, con base en las mediciones hechas por el satélite Planck (Planck Collaboration et al., 2016).
- Abundancia de elementos ligeros: El modelo Lambda-CDM predice la abundancia de elementos ligeros como el D, ^4He , ^7Li (la abundancia primordial de ^3He no es medible) (Signore y Puy, 1999). Teniendo en cuenta la posible peresencia de errores sistemáticos en los datos de Li/H.
- Estructura a gran escala del universo: Se ha mostrado que los parámetros del modelo Λ - CDM estimados a partir de la estructura a gran escala del universo (escalas de 1 a más de 10000 Mpc), concuerdan bien con el espectro de potencia

angular del fondo cósmico de microondas y con los espectros de potencia de las perturbaciones de densidad, estimados a partir de distribuciones espaciales de galaxias, cúmulos de galaxias ricas y de estadísticas de líneas de absorción de Ly $_{\alpha}$ en espectros de cuásares distantes (Apunevych et al., 2008).

- Expansión acelerada del universo: El modelo Λ -CDM explica la expansión acelerada del universo, observada a través de las supernovas tipo Ia (Riess et al., 1998).
- Lente gravitatoria: El modelo Lambda-CDM predice la distorsión de la luz por la gravedad de objetos masivos, lo que se observa en la distribución de las imágenes de lentes gravitacionales (Bartelmann, 2010).

3.1.1. Componentes del universo

El nombre de Λ -Cold dark matter se debe a que este modelo considera que el universo está constituido por los siguientes componentes:

- La materia bariónica: incluye toda la materia compuesta por bariones, es decir aquella formada por protones y neutrones, aunque también se considera a los electrones cuando estos forman parte de los átomos.
- La radiación ρ_r : incluye a las partículas relativistas: los fotones y los neutrinos, por ejemplo.
- La materia oscura fría: es un tipo de materia teorizada por los cosmólogos como no relativista, no-bariónica y sin colisiones. Lo más importante de esta componente es que no interactúa con el resto de los componentes salvo de forma gravitacional. El concepto surgió de algunos problemas reales, como el problema de la rotación de galaxias espirales (Liddle, 2015, pp 63-73). En este trabajo se agruparán la materia oscura y bariónica en un único término ρ_m .
- Constante cosmológica ρ_{Λ} : Esta componente tomada de las ecuaciones de Einstein, se considera como la representación conceptual de la llamada energía oscura (o una parte de ella), la cual es la responsable de la expansión del universo y como será detallado más adelante, es la componente principal del cosmos.

Así, el contenido del universo es la suma de las contribuciones de estas densidades de energía $\rho = \sum_i \rho_i$. Para completar el sistema de ecuaciones formado por 2.20 y 2.27 es necesario agregar una ecuación extra que describe a las componentes de este modelo, cuya naturaleza se asume como la de un fluido perfecto y cada componente tiene asociada una ecuación de estado barotrópica:

$$p_i = (\gamma_i - 1)\rho_i, \quad (3.1)$$

donde la ecuación para radiación es $\gamma_r = 4/3$; la de materia oscura y bariónica, $\gamma_m = 1$; y para la constante cosmológica, $\gamma_\Lambda = 0$. Entonces al sustituir estas expresiones en la ecuación de fluido de cada componente:

$$\dot{\rho}_i + 3H(\rho_i + p_i) = \dot{\rho}_i + 3H\gamma\rho_i = 0, \quad (3.2)$$

donde se obtiene una solución en términos del factor de escala y los parámetros de densidad medidos al día de hoy $\rho_{i,0}$:

$$\rho_i = \rho_{i,0} a^{-3\gamma_i}. \quad (3.3)$$

Sustituyendo cada respectivo valor de γ_i , le evolución de cada componente es:

$$\rho_m = \rho_{m,0} a^{-3}, \quad (3.4)$$

$$\rho_r = \rho_{r,0} a^{-4}, \quad (3.5)$$

$$\rho_\Lambda = \rho_{\Lambda,0}. \quad (3.6)$$

Donde el subíndice 0 indica que es la densidad de energía medida al día de hoy. Con la definiciones de las ecuaciones 3.4, 3.5, 3.6 la ecuación paramétrica de Friedmann en función del corrimiento al rojo z y los parámetros adimensionales de densidad Ω_i :

$$\frac{H^2}{H_0^2} = \Omega_{r,0}(1+z)^4 + \Omega_{m,0}(1+z)^3 + (1 - \Omega_0)(1+z)^2 + \Omega_{\Lambda,0}. \quad (3.7)$$

A pesar de la existencia de las componentes antes mencionadas, en el modelo cosmológico estándar se asume que el tensor de energía momento está descrito únicamente por materia no relativista y constante cosmológica, por lo que la ecuación de Friedmann pasa a ser de la forma 3.8.

$$H = H_0 \sqrt{\Omega_{m,0}(1+z)^3 + \Omega_{\Lambda,0}}. \quad (3.8)$$

Si se evalúa $z = 0$ se puede expresar la ecuación de Friedmann sólo en términos del factor de densidad de materia. Así que la ecuación de Friedmann en este modelo está dada por 3.9.

$$H = H_0 \sqrt{\Omega_{m,0}(1+z)^3 + 1 - \Omega_{m,0}}. \quad (3.9)$$

Dado que para el polvo ρ_m la contribución de su presión es nula $p_m = 0$. La ecuación de continuidad queda definida por el problema de valor inicial 3.10.

$$\frac{dx}{dz} - \frac{3x}{z+1} = 0, \quad x_0 = \Omega_{m,0}. \quad (3.10)$$

3.2. Energía oscura paramétrica

Como se mencionó en la sección 3.1.1, el marco del modelo cosmológico estándar permite atribuirle la expansión acelerada del universo a la contribución de la energía oscura. Lo anterior es un problema, pues aún se ignora mucho sobre la naturaleza de la componente principal del cosmos, con casi 70% (Lahav y Liddle, 2019) (Li et al., 2019) de la contribución energética y que determina la dinámica del universo.

Al tratarse a la energía oscura como un fluido representado por la constante cosmológica Λ , una primera aproximación fuera del modelo estándar es asumir que si dicho fluido es dinámico, su ecuación de estado no puede ser constante y por ende evolucionará con el tiempo cósmico, o en función de z (Liu et al., 2008). Este comportamiento se puede expresar con la relación 3.11.

$$\frac{p(z)}{\rho(z)} = \omega(z). \quad (3.11)$$

Dicho lo anterior, es posible proponer distintas parametrizaciones de la ecuación de estado $\omega(z)$, que dependerán de una serie de parámetros que luego podrán ser inferidos a partir de las observaciones. Se han propuesto muchos modelos para $\omega(z)$ siempre intentando que estos permitan extraer la mayor información relevante de los datos y teniendo en cuenta en todo momento que estos modelos no provienen de una teoría fundamental. Posiblemente la parametrización más conocida sea la propuesta por Chevallier y Polarski, 2001 y por Linder, 2003 de manera independiente en 2005, cuya ecuación de estado se define en la expresión 3.12.

$$\omega = \omega_0 + \omega_a \frac{z}{1+z}. \quad (3.12)$$

Donde ω_0 y ω_a son los parámetros libres a ser delimitados. La popularidad de esta parametrización radica en varios puntos: El primero es su sencillez, pues sólo depende de dos parámetros libres, de los cuales además se cuenta con una interpretación física natural; ω_0 es el valor de la ecuación de estado a $z = 0$, es decir medida al día de hoy; y ω_a determina el comportamiento para grandes corrimientos al rojo. Cabe resaltar que 3.12 está bien definida cuando el corrimiento al rojo tiende a infinito. También se ha probado (Linder, 2007) una relación entre la ecuación de estado global ω y ω_a , dada por la ecuación 3.13:

$$\omega_a = -2\omega'|_{z=1}, \quad (3.13)$$

con $\omega' = \frac{d}{d \ln(a)} \omega$.

La parametrización CPL resulta muy útil en el escenario donde la energía oscura es importante a tiempos cosmológicos tardíos e insignificante a tiempos tempranos. Finalmente, se ha mostrado que CPL no es la parametrización de la que más información se puede obtener acerca de $\omega(z)$, como mostraron (Liu et al., 2008) usando datos de SNIa. Sin embargo, es por las propiedades positivas que se mencionan arriba que esta es una buena primera extensión al modelo estándar y que se decidió utilizarla en este trabajo.

3.2.1. Dinámica del modelo

En el modelo estándar la parametrización de la energía oscura se asume constante, resultando en la evolución de ω_Λ descrita por la ecuación 3.6. En esta ocasión la evolución de esta componente se obtiene al sustituir la ecuación 3.11 en 3.2 se obtiene la ecuación 3.14:

$$\frac{d\rho_\Lambda}{dz} - \frac{3\rho_\Lambda}{1+z} \left(1 + \omega_0 + \frac{\omega_a z}{1+z} \right) = 0. \quad (3.14)$$

Integrando la ecuación anterior se obtiene la evolución para la energía oscura en el marco de modelo CPL, dada por la expresión 3.15.

$$\rho_\Lambda = \rho_{\Lambda,0} (1+z)^{3(1+\omega_0+\omega_a)} \exp\left(-\frac{3\omega_a z}{1+z}\right). \quad (3.15)$$

Adicionalmente como este modelo debe ser trabajado con datos observacionales que se encuentran a bajo corrimiento al rojo, donde la influencia de la densidad de radiación no juega un papel importante, la contribución de esta suele ignorarse. Además al considerar un espacio tiempo con curvatura cero o plano, el término $1 - \Omega_0$ en la ecuación de 2.27 se anula, por lo que la ecuación de Friedmann bajo este par de supuestos, la ecuación paramétrica de Friedmann en el marco CPL queda definida por la ecuación 3.16:

$$H(z) = H_0 \sqrt{\Omega_{m,0} (1+z)^3 + \Omega_{\Lambda,0} (1+z)^{3(1+\omega_0+\omega_a)} \exp\left(-\frac{3\omega_a z}{1+z}\right)}. \quad (3.16)$$

Dada la ecuación de Friedmann en este marco cosmológico, es posible sustituir H en la integral de la ecuación 2.31 para encontrar la distancia comovil en este modelo cosmológico, la ecuación 3.17.

$$d_C = \frac{1}{H_0} \int_0^z \frac{dz'}{\sqrt{\Omega_{m,0}(1+z')^3 + (1-\Omega_{m,0})(1+z')^{3(1+\omega_0+\omega_a)} \exp\left(-\frac{3\omega_a z'}{1+z'}\right)}}. \quad (3.17)$$

Y ahora como se vio en la sección de distancias cosmológicas, dada la distancia comovil se puede calcular la distancia luminosa dada por la relación 2.36.

3.3. Quintaesencia

En el ámbito de la cosmología moderna, el modelo de quintaesencia emerge como una de las propuestas más intrigantes para tratar de explicar algunos hechos observacionales que pusieron discrepancias entre las observaciones de la expansión acelerada del universo, encontrada a finales del siglo XX por «2006 John C. Mather and George F. Smoot» s.f.

En el modelo estándar se considera que la contribución de energía faltante entre la densidad de materia y la densidad crítica es aportada por una constante cosmológica Λ , y pesar del éxito y simpleza del modelo Λ -CDM, no existe una razón definitiva para asumir que dicha energía faltante esté constituida por una energía constante del vacío. Quintaesencia explora la posibilidad de que la energía faltante sea contribuida por una componente espacialmente inhomogénea y que evoluciona con el tiempo, con una presión negativa y una ecuación de estado que difiere de las otras cuatro componentes: bariones, neutrinos, materia oscura y radiación (Caldwell, 2000). Por eso el nombre de una *quinta esencia*.

El nuevo escenario Quintaesencia con materia oscura (Q -CDM) se construye sobre un espacio-tiempo de geometría plana cuyo intervalo infinitesimal de línea está dado por la métrica FLRW y las componentes del universo son la Quintaesencia más todas las partículas del modelo estándar y las componentes del modelo Λ -CDM, salvo la constante cosmológica. Este universo evoluciona desde una época inflacionaria, luego por la etapa de dominación de radiación y materia respectivamente, hasta el tiempo presente de la dominación de Quintaesencia.

la ecuación de estado ω , y junto con el parámetro de densidad de materia $\Omega_m = 1 - \Omega_Q$ son las dos cantidades más importantes para caracterizar esta componente. Quintaesencia se propone como un campo escalar cósmico, pues es una propuesta natural para explorar escenarios más allá del modelo estándar de partículas, pues este tipo de campos son esenciales en la física moderna. El escalar de Higgs, el inflatón o dilatón son algunos de los ejemplos importantes en el esquema actual (Peskin y Schroeder, 1995).

3.3.1. Dinámica del campo escalar

Las ecuaciones de evolución del campo escalar cósmico se pueden determinar al considerar un campo escalar auto-interactuante compuesto por dos partes: una parte homogénea ϕ de fondo y una perturbativa e inhomogénea $\delta\phi$, cuyo lagrangiano asociado está dado por la ecuación 3.18

$$L = \frac{1}{2} \partial_\mu \phi \partial^\mu \phi - V(\phi), \quad (3.18)$$

donde V es el potencial producido por el campo (Caldwell, 2000). Así pues, a partir de una métrica general y una minimización de Euler-Lagrange se llega a la ecuación 3.19.

$$\square\phi - \partial_\phi V = 0. \quad (3.19)$$

Del lagrangiano se puede derivar el tensor de energía-momento para ϕ , teniendo en cuenta que ϕ cumple homogeneidad e isotropía por lo que las parciales espaciales desaparecen, ya que el campo tiene que ser homogéneo en el espacio. Esto lleva a obtener las expresiones para la densidad 3.20 y presión 3.21 para el campo escalar.

$$\rho = \frac{1}{2} \dot{\phi}^2 + V(\phi), \quad (3.20)$$

$$p = \frac{1}{2} \dot{\phi}^2 - V(\phi). \quad (3.21)$$

Al introducir estas componentes en la ecuación de continuidad se obtiene la ecuación 3.22, que también es conocida como la ecuación de *Klein-Gordon*.

$$\ddot{\phi} + 3H\dot{\phi} + \frac{\partial V}{\partial \phi} = 0. \quad (3.22)$$

La expansión del universo bajo este modelo está determinada por las ecuación de Friedmann 3.23.

$$H^2 = \kappa_0^2 \left(\rho_m + \frac{1}{2} \dot{\phi}^2 + V \right). \quad (3.23)$$

La ecuación de Klein-Gordon ha sido resulta utilizando métodos numéricos basados en la teoría de sistemas dinámicos, para distintos potenciales V propuestos (Copeland, Liddle y Wands, 1998). El método consiste en plantear un sistema de dos ecuaciones por dos variables y acoplado. Al separar la ecuación de segundo grado se toma el cambio de variable descrito por 3.24.

$$x = \frac{\sqrt{\kappa_0} \dot{\phi}}{\sqrt{6}H}, \quad y = \frac{\sqrt{\kappa_0 V(\phi)}}{\sqrt{3}H}, \quad (3.24)$$

así que el sistema quedará dado en términos de las ecuaciones 3.25 :

$$\begin{cases} \frac{dx}{dN} = -3x + \frac{\sqrt{6}}{2}\lambda y^2 + \frac{3}{2}x(1 + x^2 - y^2) \\ \frac{dy}{dN} = -\frac{\sqrt{6}}{2}xy\lambda + \frac{3}{2}y(1 + x^2 - y^2) \end{cases} \quad (3.25)$$

donde $N = \ln(a)$. Y con una ecuación de constricción dada por 3.26.

$$\frac{\kappa_0 \rho_m}{3H^2} + x^2 + y^2 = 1. \quad (3.26)$$

Finalmente, el sistema debe ser resuelto definiendo una expresión para el potencial V y lo más importante: un conjunto de condiciones iniciales $\{x_0, y_0\}$. Establecer el conjunto de condiciones iniciales no es una tarea sencilla. En primer lugar el sistema es sensible a las condiciones que sean elegidas (Kujat, Scherrer y Sen, 2006), por lo que en pos de obtener soluciones adecuadas, estas condiciones deben ser cuidadosamente elegidas.

Determinar el valor que deben tomar $\{x_0, y_0\}$ como condiciones iniciales también necesita de asumir distintas hipótesis sobre el modelo y la naturaleza del cosmos. Algunas de ellas son; asumir una forma conveniente para el potencial V , dejando parámetros libres que después pueden ser inferidos con ayuda de datos; definir las condiciones en algún escenario cosmológico que permita evitar tensiones como la de Hubble, por ejemplo en la época de dominación de materia.

Otra manera puede ser la de establecer las condiciones iniciales como parámetros libres, para luego resolver el sistema utilizando métodos estadísticos para encontrar los mejores valores para dichas variables. También se ha utilizado el conocido como método de *shooting o disparo*, donde el sistema se resuelve para alguna combinación de valores iniciales para luego probar qué tan precisas son con el cumplimiento esperado, si no es óptimo, se elige una nueva combinación de valores y el proceso se repite hasta encontrar las soluciones más parecidas a lo deseado según las hipótesis. Todos los escenarios anteriores fueron desarrollados con mayor detalle por (Vázquez et al., 2021).

En este trabajo se abordará la solución a este sistema sin necesidad de establecer las condiciones $\{x_0, y_0\}$, sino fijando las soluciones a partir de incorporar datos observacionales de supernovas en combinación con una red neuronal físicamente informada.

Capítulo 4

Redes totalmente conectadas

Una red totalmente conectada (Figura 4.1) se trata de un arreglo de l capas de neuronas donde una entrada a^0 es procesada hasta generar una salida a^L . La red cuenta con capas entre la salida y la entrada, denominadas capas ocultas, cuya salida es a^l , y se obtiene a partir de la salida de la capa previa a^{l-1} . Las componentes de cada capa l de la red están descritas por la ecuación 4.1.

$$a_j^l = \sigma \left(\sum_i a_i^{l-1} w_{ij}^l + b_j^l \right), \quad (4.1)$$

donde l representa el número de la capa, 0 para la capa de entrada y L para la salida (que en este trabajo también se denota por *ANN*); w_{ij}^l se conoce como el *weight* o peso entre el nodo i en la capa $l - 1$ y el nodo j en la capa l ; y b_j^l se llama *bias* o sesgo. Para simplificar esto, se suele utilizar la notación matricial, por lo que la estructura de las capas de la red neuronal está determinada por la ecuación matricial 4.2:

$$a^l = \sigma(a^{l-1}W^l + b^l), \quad (4.2)$$

donde W^l y b^l son la matriz de pesos y el vector de sesgos de la capa l , y σ es la función de activación.

Como se mencionó antes, la característica principal de una red neuronal es su capacidad de emular funciones, y esto se consigue a partir de un proceso de entrenamiento que consiste en proveerle de dos conjuntos de datos X, Y a la red para que aprenda la función F que subyace entre ambos conjuntos $X \xrightarrow{F} Y$. En machine learning X suele llamarse *features* o características y Y , las etiquetas. Los modelos que requieren de un conjunto de etiquetas para su entrenamiento, es decir que requieren que se les indique los valores que deben emular, se les conoce como modelos supervisados.

Para medir qué tan bien la red emula el comportamiento subyacente entre las características y las etiquetas, se define una función de error o costo, que depende de los parámetros de la red: los pesos W^l y sesgos b^l . La función de costo a utilizar depende de la tarea a realizar; para tareas de regresión, se suele utilizar la función del

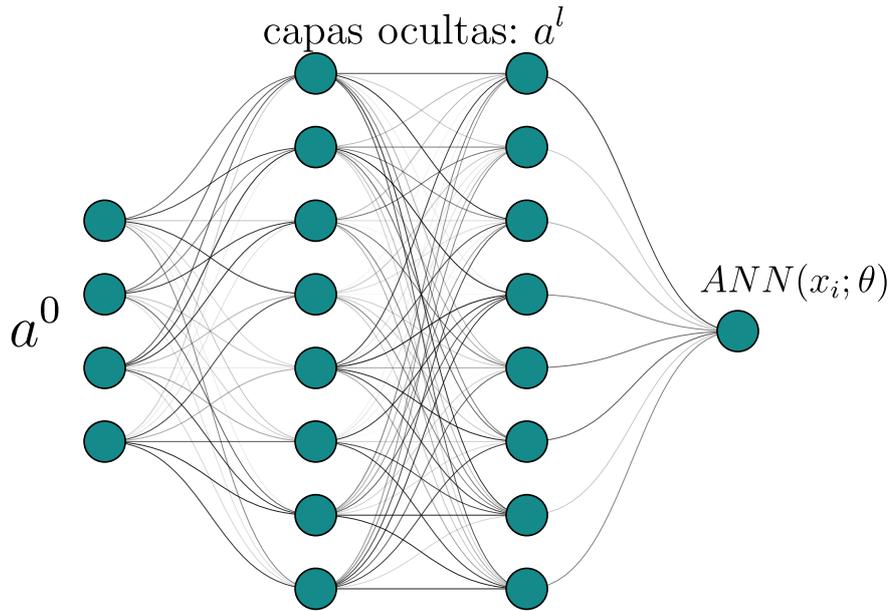


FIGURA 4.1: Estructura general de una red totalmente conectada o MLP. Donde a^0 representa la capa de entrada de las variables del modelo, las capas ocultas a^l y la salida de la red $ANN(t; \theta)$ donde θ representa el conjunto de parámetros intrínsecos de la red: los pesos y sesgos. La función de activación se aplica a cada una de las capas.

error cuadrático medio *MSE* 4.3; para tareas de clasificación, *entropía cruzada* 4.4 es la más socorrida. La función de costo es parte crucial del proceso de entrenamiento, pues es a través de ella que se puede discernir entre un buen modelo y uno que no lo es.

$$\mathcal{L}_{MSE}(a^L, Y) = \frac{1}{N} \sum_{i=1}^N (a_i^L - Y_i)^2, \quad (4.3)$$

$$\mathcal{L}_H(a^L, Y) = -\frac{1}{N} \sum_{i=1}^N \left[Y_i \log(a_i^L) + (1 - Y_i) \log(1 - a_i^L) \right]. \quad (4.4)$$

El proceso de aprendizaje de la red consiste en la modificación de sus pesos y sesgos de tal manera que resulte en la minimización de la función de costo, es decir que se trata de un proceso de optimización. Existen diversos algoritmos para optimizar una función, pero en este trabajo se utilizara una variante del descenso del gradiente, conocida como Adam. Al igual que el descenso del gradiente, Adam es un algoritmo iterativo que requiere el gradiente de la función a minimizar, y por ende las derivadas parciales respecto a cada una de las variables, como se describe en el pseudocódigo 1.

Por lo que la parte crítica del proceso de aprendizaje de la red neuronal es calcular las derivadas parciales de la función de costo respecto a los parámetros de la red neuronal (pesos y sesgos).

Algorithm 1 Algoritmo Adam

```
Inicializar  $\theta, \beta_1, \beta_2, \alpha, \epsilon$ 
for cada iteración  $t$  do
  Calcular el gradiente  $g_t = \nabla_{\theta} F(\theta_{t-1})$ 
  Actualizar:
   $m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$ 
   $v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$ 
  Actualizar  $\theta_t$ 
end for
```

Una vez se logre alcanzar un punto suficientemente bajo de la función de costo, la red fully-connected ofrecerá un modelo que emule de forma adecuada a la relación funcional entre los conjuntos que le fueron provistos. Como se mostró en (Dios Rojas Olvera, Gómez-Vargas y Vázquez, 2022), este tipo de redes neuronales es muy útil para hacer trabajos de regresión de funciones, clasificación de objetos y particularmente para emular las soluciones de ecuaciones diferenciales. En el caso particular de sistemas de ecuaciones diferenciales, una primer aproximación consiste en obtener las soluciones del sistema por medio de un método numérico tradicional, y hacer que la red aprenda directamente de esos datos. Algunos de los problemas de ese método son:

- Se necesita de un método numérico que resuelva el sistema en primer instancia, lo que resulta más complicado si la ecuación diferencial es problemática para los métodos tradicionales.
- El rango del dominio es tan extenso como los datos disponibles para el entrenamiento. Si se quiere ampliar el rango para evaluar el modelo, se deben conseguir más observaciones o generar más datos.
- Para un MLP no existe una forma, a partir sólo de los datos, de inducir información matemática del fenómeno que se quiere modelar. Esto hace que este tipo de redes sea muy sensible al ruido en los datos y se vuelvan propensas al sobre-ajuste.

Capítulo 5

Redes Neuronales Físicamente Informadas

El aprendizaje profundo es un paradigma de la inteligencia artificial encargado del estudio y desarrollo de las técnicas de entrenamiento de las redes neuronales artificiales, un objeto computacional inspirado en el funcionamiento del cerebro, con capas de neuronas interconectadas que procesan información y aprenden patrones. La propiedad fundamental de las redes neuronales es que son un aproximador universal de funciones (Lu y Lu, 2020) y esta habilidad se ve potenciada al utilizar distintas arquitecturas de red que cumplan con objetivos específicos. Las redes neuronales más sencillas son las conocidas como perceptrones multicapa o fully-connected.

La lista anterior sintetiza en pocas palabras la motivación detrás de un paradigma de redes neuronales que se basa en la misma arquitectura de un perceptrón multicapa pero al que se le ha provisto información matemática relacionada con el fenómeno que se quiere modelar. Este paradigma de red se conoce como **Redes Neuronales Físicamente Informadas o PINN (por sus siglas en inglés)**.

5.1. Modelando funciones implícitas

La idea fundamental detrás de las PINN radica en la capacidad de una red neuronal de aproximar funciones, pero a diferencia del aprendizaje supervisado mencionado en el capítulo 4, se propone estructurar una nueva forma de aprendizaje a partir de una expresión matemática y no necesariamente de datos, de tal manera que la red sea capaz de emular una función u inmersa en una función implícita \mathcal{F} cuya forma general está dada por la ecuación 5.1.

$$\mathcal{F}(u, \{t_i\}, \Theta) = 0, \quad (5.1)$$

donde $\{t_i\}, \Theta$ son el conjunto de variables independientes y parámetros libres de la relación respectivamente y $u \equiv u(\{t_i\}, \Theta)$. Así, la intención es que la red neuronal emule la función u a través de pedirle que satisfaga la relación \mathcal{F} , como en 5.2.

$$\mathcal{F}(\text{ANN}(\{t_i\}, \Theta), \{t_i\}, \Theta) = 0. \quad (5.2)$$

Hacer que la red neuronal satisfaga 5.2 se puede lograr modificando la función de costo de la red neuronal definiendo la función de costo 5.3:

$$\mathcal{L}_{\mathcal{F}} = \mathcal{F}(\text{ANN}(\{t_i\}, \Theta), \{t_i\}, \Theta)^2. \quad (5.3)$$

Es de mencionar que hasta aquí se sigue trabajando con la arquitectura de un MLP común y corriente, simplemente se ha definido una nueva función de costo que permite incorporar una expresión matemática de forma directa. A partir de la expresión analítica $\mathcal{L}_{\mathcal{F}}$, se generan los puntos que servirán para el entrenamiento de la red, generándolos siguiendo una cierta distribución (usualmente uniforme) dentro de un cierto rango para las variables $\{t_i\}$, para después continuar con la optimización de la función de costo de la misma manera que para un MLP. Es entonces este proceso de entrenamiento lo que define a una PINN y la diferencia de un MLP convencional.

Al prescindir de las etiquetas para entrenar a esta red neuronal, se tiene un caso de *aprendizaje no supervisado*. Además, en este caso el aprendizaje se obtendrá a partir de una expresión analítica por lo que los puntos que se le proveerán estarán libres de ningún ruido o sesgo observacional. Como no hay restricciones sobre la relación \mathcal{F} más allá de satisfacer las condiciones del teorema fundamental de aproximación (Lu y Lu, 2020), la cantidad de funciones que se pueden emular de esta manera es considerable. Sin embargo hay un tipo en particular que es fundamental para las matemáticas y la física, el caso cuando a \mathcal{F} la definen operadores diferenciales.

5.2. Ecuaciones diferenciales

El caso particular más importante de función implícita se da cuando la relación \mathcal{F} describe una ecuación diferencial o sistema de ecuaciones diferenciales. Por simplicidad, primero se considerará el caso de una ecuación diferencial ordinaria (EDO), la forma más general de una EDO está dada por la ecuación 5.4.

$$\mathcal{F}\left(t, u, \frac{du}{dt}, \frac{d^2u}{dt^2}, \dots, \frac{d^nu}{dt^n}, \Theta\right) = 0. \quad (5.4)$$

Por lo que al sustituir la ecuación 5.4 en la 5.3, se obtiene la función de costo 5.5:

$$\mathcal{L}_{\mathcal{F}} = \mathcal{F}\left(t, \text{ANN}, \frac{d\text{ANN}}{dt}, \frac{d^2\text{ANN}}{dt^2}, \dots, \frac{d^n\text{ANN}}{dt^n}, \Theta\right)^2. \quad (5.5)$$

Así, una vez que se ha minimizado la ecuación 5.5, el modelo provisto por la PINN podrá emular la solución original: $\text{ANN} \sim u$.

5.3. Condiciones iniciales y de frontera

Para que la solución de una ecuación diferencial \mathcal{F} quede completamente determinada, es necesario establecer condiciones extra a la determinada sólo por \mathcal{F} , estas son los valores de la función solución a algunos valores fijos de la(s) variables independientes. Dichas condiciones se denominan condiciones iniciales, cuya expresión general para ecuaciones de una sola variable está dada por 5.6.

$$\frac{d^i u}{dt^i}(t_0) - u_0^{(i)} = 0 \quad i = 0, 1, 2, \dots, n. \quad (5.6)$$

A los sistemas de ecuaciones diferenciales que deben cumplir con condiciones iniciales como las definidas en 5.6 se les denomina *problemas de valor inicial*.

En términos generales la solución de un sistema de ecuaciones diferenciales está definida sobre un dominio $\Omega \subset \mathbb{R}^n$ es común que ciertas condiciones deban ser impuestas sobre la frontera de Ω , $\partial\Omega$. A diferencia de las condiciones iniciales, estas restricciones no se dan sólo en puntos extremos de un intervalo, sino sobre todo un intervalo que parametriza a $\partial\Omega$. Condiciones de este tipo se denominan *condiciones de frontera* y de forma general dependen de las derivadas parciales de la solución, como se representan en la ecuación 5.7.

$$g \left(t_k, u, \frac{\partial^j u}{\partial t_k^j} \right) \Big|_{\partial\Omega} = 0, \quad \forall t_k \in \partial\Omega. \quad (5.7)$$

Dado que siempre se tendrán condiciones iniciales y muy frecuentemente se establecerán restricciones sobre la frontera del dominio, la PINN deberá ser capaz de cumplir tanto con condiciones de frontera como con condiciones iniciales. Existen dos enfoques a través de los cuales se puede forzar a la PINN a cumplir condiciones iniciales o de frontera, y serán detallados a continuación en las secciones 5.3.1 y 5.3.2.

5.3.1. Condiciones iniciales y de frontera como componentes del costo.

Es la optimización de la función de costo lo que determina si el modelo ha aprendido o no. Por lo que es natural afirmar que si la función de costo 5.3 es la que constriñe el comportamiento global de la red, se puede añadir una componente que represente el comportamiento sobre determinados puntos o regiones del dominio. Por lo que se definen las nuevas componentes del costo \mathcal{L}_{u_0} y $\mathcal{L}_{\partial\Omega}$, definidas en las ecuaciones 5.8 y 5.9 que representan las contribuciones de todas las condiciones iniciales y de frontera respectivamente.

$$\mathcal{L}_{u_0} = \sum_i \left(\frac{d^i ANN}{dt^i}(t_0) - u_0^{(i)} \right)^2. \quad (5.8)$$

$$\mathcal{L}_{\partial\Omega} = \sum_k \mathcal{G} \left(t_k, ANN, \frac{\partial^j ANN}{\partial t_k^j} \right)^2 \Big|_{\partial\Omega}. \quad (5.9)$$

Así pues cuando se intente resolver una ecuación diferencial con determinadas condiciones iniciales y de frontera se deberán considerar las contribuciones del costo dadas por la ecuación diferencial $\mathcal{L}_{\mathcal{F}}$, las de las condiciones iniciales \mathcal{L}_{u_0} y las de las constricciones de frontera $\mathcal{L}_{\partial\Omega}$. Por lo tanto, función de costo a optimizar es 5.10, la suma de 5.3, 5.8, 5.9.

$$\mathcal{L} = \mathcal{L}_{\mathcal{F}} + \mathcal{L}_{u_0} + \mathcal{L}_{\partial\Omega}. \quad (5.10)$$

Es importante notar que en términos generales las componentes de la función de costo 5.10 no dependen de las variables independientes, por ejemplo las componentes 5.8 y 5.9 son constantes respecto a dichas variables. Sin embargo, estas funciones de costo sí dependen de los parámetros de la red, por lo que es posible obtener las derivadas parciales respecto a esos parámetros y propagar el error hacia atrás como se indica en la sección 5.

5.3.2. Parametrización

Existe un esquema alternativo al mencionado en la sección 5.3.1 donde la carga de satisfacer las condiciones iniciales o de frontera recaer directamente sobre la función de costo, se trata del descrito por (Lagaris, Likas y Fotiadis, 1998). Este enfoque se basa en asumir que se tiene una ecuación diferencial de la forma 5.4 que precisa de cumplir determinadas condiciones iniciales y de frontera, cuya solución puede expresarse como una función $\Psi(t, \theta)$, donde θ son parámetros ajustables intrínsecos a la solución. Entonces se obtiene la ecuación diferencial 5.11.

$$\mathcal{F} \left(t, \Psi, \frac{d\Psi}{dt}, \frac{d^2\Psi}{dt^2}, \dots, \frac{d^n\Psi}{dt^n}, \Theta \right) = 0. \quad (5.11)$$

No se deben confundir los parámetros Θ, θ : los primeros son los parámetros libres asociados al fenómeno que describe la ecuación diferencial y los segundos son los parámetros ajustables intrínsecos a la función solución. La forma en la que Ψ cumple con las condiciones iniciales y de frontera es definirla como la suma de dos funciones, como se indica en la ecuación 5.12

$$\Psi(t, \theta) = \phi(t) + \psi(t, ANN(t, \theta)), \quad (5.12)$$

donde $\phi(t)$ satisface las condiciones iniciales o de frontera y no contiene parámetros ajustables; $ANN(t, \theta)$ es un perceptrón multicapa, por lo que θ son sus pesos y sesgos;

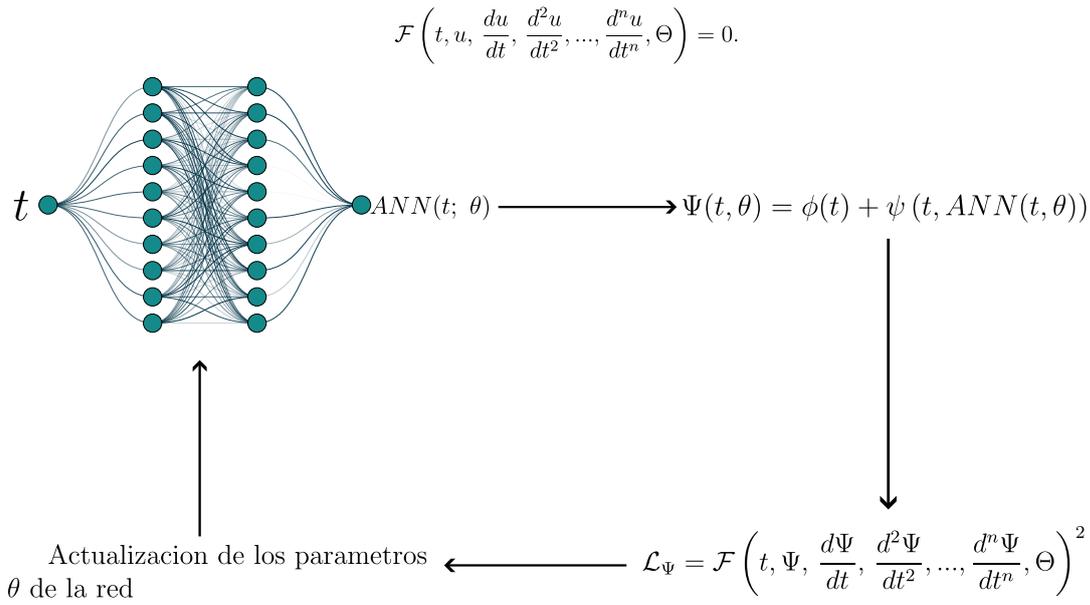


FIGURA 5.1: Proceso de entrenamiento de una PINN.

y ψ se construye de tal manera que no tenga contribución en las condiciones iniciales o de frontera y su forma dependerá de cómo sean las dichas condiciones.

Es entonces la parametrización $\Psi(t, \theta)$ de la red neuronal la que dará la solución de la ecuación diferencial y además dicha solución cumplirá las condiciones iniciales y de frontera del problema. Por lo que la red deberá ser siempre parametrizada por ψ , incluso durante el entrenamiento, como se puede ver en la figura 5.1. La función de costo para el esquema de parametrización está dada entonces por la ecuación 5.13.

$$\mathcal{L}_\Psi = \mathcal{F}\left(t, \Psi, \frac{d\Psi}{dt}, \frac{d^2\Psi}{dt^2}, \dots, \frac{d^n\Psi}{dt^n}, \Theta\right)^2. \quad (5.13)$$

Dado que en este trabajo se abordarán las ecuaciones cosmológicas contenidas en los primeros 4 capítulos, es suficiente definir dos parametrizaciones del tipo Ψ : las correspondientes a los problemas de valor inicial de *Dirichlet* y los problemas de *Neumann*.

En el caso que tomamos una ecuación diferencial de primer orden, el problema de valor inicial se transforma en las ecuaciones 5.14:

$$\begin{cases} \mathcal{F}\left(t, u, \frac{du}{dt}, \Theta\right) = 0, \\ u(t_0) = u_0. \end{cases} \quad (5.14)$$

Para este caso, la forma de construir Ψ es la misma que la propuesta por (Mattheakis et al., 2022). La condición inicial independiente de los parámetros de la red es la constante $\phi(t) = u_0$, y la parte que rige el comportamiento global de la solución,

pero que no contribuye a las condiciones de frontera es $\psi = (1 - e^{-(t-t_0)})ANN(t, \theta)$. Entonces la parametrización de la red para el problema de valor inicial está dado por la ecuación 5.15

$$\Psi(t, \theta) = u_0 + \left(1 - e^{-(t-t_0)}\right) ANN(t, \theta). \quad (5.15)$$

Nótese que es posible elegir distintas parametrizaciones ψ que cumplan con $\Psi(t_0, \theta) = u_0$, pero esta elección cumple dos propiedades interesantes, que debido a la velocidad de convergencia de la exponencial son difícilmente igualables:

$$\left\{ \begin{array}{l} \text{cuando } t \rightarrow \infty, \psi \rightarrow ANN \\ \text{cuando } t \rightarrow t_0, \Psi \rightarrow u_0 \text{ exponencialmente.} \end{array} \right. \quad (5.16)$$

Por otro lado, las condiciones de Neumann para una ecuación diferencial de segundo orden están dadas por 5.17.

$$\left\{ \begin{array}{l} \mathcal{F}\left(t, u, \frac{du}{dt}, \frac{d^2u}{dt^2}, \Theta\right) = 0, \\ u(t_0) = u_0, \\ \left.\frac{du}{dt}\right|_{t_0} = v_0. \end{array} \right. \quad (5.17)$$

Para este caso, la parametrización es derivada de la 5.15, pues así se heredan las propiedades antes mencionadas. Para el caso de un problema de Neumann la parametrización está definida por la ecuación 5.18.

$$\Psi(t, \theta) = u_0 + (t - t_0)v_0 + \left(1 - e^{-(t-t_0)}\right)^2 ANN(t, \theta). \quad (5.18)$$

5.4. Datos observacionales

En el aprendizaje supervisado los datos que se utilizan para el entrenamiento de los modelos generalmente provienen de observaciones, eso implica que los datos tendrán siempre una incertidumbre asociada, sesgos o ruido. Ya sea por la naturaleza intrínseca del fenómeno en sí, por la calidad de los instrumentos o errores humanos cometidos cuando los datos fueron recolectados.

Una mala calidad de datos puede inducir sesgos en el modelo o enseñar comportamientos inexactos sobre el fenómeno real, haciendo que las predicciones hechas con el modelo obtenido sean poco útiles. Una de las principales muestras de este problema es el conocido 'sobre ajuste'.

El sobre ajuste es posiblemente el principal problema cuando se está entrenando una red neuronal, se trata del caso cuando el modelo se ha adaptado demasiado a los datos de entrenamiento, al punto de que es incapaz de generalizar lo aprendido haciendo predicciones sobre elementos fuera del conjunto de entrenamiento.

Las PINN son un tipo de redes neuronales que es robusta ante el ruido o sesgos de los datos observacionales, pues cuenta con una componente muy importante: una directriz matemática. Esta directriz permite guiar el comportamiento global del modelo en algunos puntos clave o incluso el comportamiento global, pero velando por hacer una descripción de los datos lo más acertada posible.

Incluir datos observacionales al entrenamiento de las PINN se consigue a través del método de entrenamiento descrito en la sección 5.3.1. En el caso de una red neuronal totalmente conectada el entrenamiento se rige por la optimización de la función de costo 4.3 y una PINN, por la de la función 5.10. Al considerar la contribución de los datos observacionales a través del error cuadrático medio \mathcal{L}_{MSE} (no necesariamente deber ser el error MSE), esta se puede incluir dentro de la función de costo total 5.10, como se indica en la ecuación 5.19.

$$\mathcal{L} = \alpha_1 \mathcal{L}_{MSE} + \alpha_2 \mathcal{L}_{\mathcal{F}} + \alpha_3 \mathcal{L}_{u_0} + \alpha_4 \mathcal{L}_{\partial\Omega}. \quad (5.19)$$

Los coeficientes α sirven para modular la influencia de cada componente y su impacto ulterior sobre el modelo resultante, entre menor sea su valor, menor será la influencia de esa contribución sobre el comportamiento final de la red. Generalmente el rango de α está dado entre 0 y 1, donde 0 representa la nula contribución sobre el modelo final. El valor de cada uno dependerá, por ejemplo de qué se espera tomar más en cuenta en el modelo final: si se tienen datos muy ruidosos, posiblemente se quiera que $\alpha_1 \ll \alpha_2$.

5.5. Integrales

Las ecuaciones diferenciales son fundamentales en matemáticas y física, pero sus contrapartes operacionales, las integrales son igualmente importantes. Más aún, siempre que se debe abordar una ecuación diferencial de forma analítica, tarde o temprano se llega al punto donde es necesario resolver una integral. Por otro lado, fuera del contexto de las ecuaciones diferenciales, la gran mayoría de las integrales no cuenta con una primitiva analítica, por lo que debe ser abordada de forma numérica. Ya se ha hablado extensamente de los problemas que conlleva cuando un problema se debe tratar de forma computacional, así que en esta sección se limita a mostrar que las PINN con capaces de resolver integrales también.

La integral de una función es, en principio, una ecuación diferencial. Lo anterior es consecuencia del teorema fundamental del cálculo, pues dada cualquier integral de la forma 5.20:

$$\begin{aligned} u(t) &= \int_{t_0}^t h(t', \Theta) dt' + u_0, \\ \frac{du}{dt} &= h(t, \Theta), \end{aligned} \quad (5.20)$$

se puede llevar al problema de valor inicial 5.21, siempre y cuando h sea continua.

$$\begin{cases} \mathcal{F}\left(t, u, \frac{du}{dt}, \Theta\right) = \frac{du}{dt} - h(t, \Theta) = 0, \\ u(t_0) = u_0. \end{cases} \quad (5.21)$$

Y se tiene entonces un problema de valor inicial que puede ser resuelto con cualquiera de los enfoques descritos a profundidad en las secciones anteriores.

5.6. Sistemas de ecuaciones diferenciales

Es muy común que el fenómeno que se desea modelar no esté determinado por una sola ecuación diferencial, si no que sea menester resolver todo un sistema de ecuaciones diferenciales. En general, un sistema de ecuaciones diferenciales ordinarias, con $t \in \mathbb{R}$ y una función vectorial $\vec{u} : \mathbb{R} \rightarrow \mathbb{R}^m$ es de la forma 5.22,

$$\vec{\mathcal{F}}\left(t, \vec{u}, \frac{d\vec{u}}{dt}, \frac{d^2\vec{u}}{dt^2}, \dots, \frac{d^n\vec{u}}{dt^n}, \Theta\right) = \vec{0}, \quad (5.22)$$

donde $\vec{\mathcal{F}}$ es una función vectorial cuyas funciones componentes son de la forma 5.4, el caso para una sólo ecuación diferencial ordinaria. El sistema 5.22 puede expresarse en forma matricial, como se observa en 5.23.

$$\begin{pmatrix} \mathcal{F}_1\left(t, \vec{u}, \frac{d\vec{u}}{dt}, \frac{d^2\vec{u}}{dt^2}, \dots, \frac{d^n\vec{u}}{dt^n}, \Theta\right) \\ \mathcal{F}_2\left(t, \vec{u}, \frac{d\vec{u}}{dt}, \frac{d^2\vec{u}}{dt^2}, \dots, \frac{d^n\vec{u}}{dt^n}, \Theta\right) \\ \vdots \\ \mathcal{F}_m\left(t, \vec{u}, \frac{d\vec{u}}{dt}, \frac{d^2\vec{u}}{dt^2}, \dots, \frac{d^n\vec{u}}{dt^n}, \Theta\right) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (5.23)$$

Con lo descrito hasta ahora, se podría resolver el sistema 5.23 asignando una red neuronal individual a cada una de las componentes de \vec{u} para luego integrar todas en la expresión del sistema. Sin embargo, como todo el sistema está parametrizado por una variable: t , se puede pedir directamente a una red neuronal que emule el comportamiento de la función vectorial solución $ANN \sim \vec{u}$.

$$ANN : t \in \mathbb{R} \rightarrow \mathbb{R}^m.$$

Lo anterior es simplemente una generalización del caso de una ecuación individual. Así que, la función de costo $\mathcal{L}_{\vec{f}}$ para el entrenamiento de la red estaría determinada por las contribuciones de cada una de las ecuaciones diferenciales que conforman el sistema, aplicadas a las componentes de $ANN(t)$, como indica la ecuación 5.24.

$$\mathcal{L}_{\vec{f}} = \sum_k^m \mathcal{L}_k = \sum_k^m \mathcal{F}_k \left(t, ANN, \frac{dANN}{dt}, \frac{d^2ANN}{dt^2}, \dots, \frac{d^n ANN}{dt^n}, \Theta \right)^2. \quad (5.24)$$

El sistema de ecuaciones también puede tener un conjunto de restricciones iniciales o de frontera como las que ya se han comentado aquí. Al igual que con las ecuaciones individuales, en este trabajo sólo se consideraran las condiciones de Dirichlet o Neumann. Los métodos para hacer a la red cumplir las mencionadas condiciones son los mismos que para el caso individual, incluir la restricción en la función de costo (sección 5.3.1) o hacer una parametrización de la red neuronal que fuerce a cumplir las restricciones (sección 5.3.2).

En primer lugar, la generalización de ambos enfoques se hará para cuando se tiene un problema de valor inicial con condiciones de Dirichlet, esto es, una serie de valores dados cuando $\vec{u}(t_0)$ como en la ecuación 5.25.

$$\vec{u}(t_0) = \vec{u}_0 = \begin{pmatrix} u_{1,0} \\ u_{2,0} \\ \vdots \\ u_{m,0} \end{pmatrix}. \quad (5.25)$$

Aquí el término \mathcal{L}_{u_0} se define como la suma de las contribuciones de cada una de las restricciones por componente de $u(t_0)$, como se muestra en 5.26.

$$\mathcal{L}_{u_0} = \sum_k^m (ANN(t_0)_k - u_{k,0})^2, \quad (5.26)$$

donde el subíndice k representa la k -ésima componente del vector. Y la función de costo total puede ser completada según las condiciones adicionales del problema, como en la ecuación 5.10 o en 5.19, si se van a añadir datos observacionales.

En el otro enfoque, se debe realizar una parametrización de la red neuronal, tal que cumpla las condiciones iniciales. De forma natural se puede hacer la generalización de la parametrización 5.15 a su forma vectorial, dada por la ecuación 5.27.

$$\vec{\Psi}(t, \theta) = \underbrace{\vec{u}_0}_{\in \mathbb{R}^m} + \underbrace{(1 - e^{-(t-t_0)})}_{\in \mathbb{R}} \underbrace{ANN(t, \theta)}_{\in \mathbb{R}^m}. \quad (5.27)$$

La parametrización 5.27 además de cumplir con las condiciones iniciales también tiene como componentes las parametrizaciones de cada solución original, de la forma 5.15, por lo que hereda las propiedades mencionadas en 5.16.

$$\vec{u}(t_0) = \begin{pmatrix} u_{1,0} \\ u_{2,0} \\ \vdots \\ u_{m,0} \end{pmatrix}, \quad \left. \frac{d\vec{u}}{dt} \right|_{t_0} = \vec{v}_0 = \begin{pmatrix} v_{1,0} \\ v_{2,0} \\ \vdots \\ v_{m,0} \end{pmatrix}. \quad (5.28)$$

Por otro lado, los métodos para abordar las condiciones de Neuman (ecuación 5.28) se generalizan de manera similar a la mencionada para las condiciones de Dirichlet.

$$\mathcal{L}_{u_0} = \sum_k^m \left[(ANN(t_0)_k - u_{k,0})^2 + \left(\frac{dANN}{dt}(t_0)_k - v_{k,0} \right)^2 \right], \quad (5.29)$$

y su respectiva parametrización está definida por la generalización de 5.18, definida en la ecuación 5.30.

$$\vec{\Psi}(t, \theta) = \vec{u}_0 + (t - t_0)\vec{v}_0 + \left(1 - e^{-(t-t_0)}\right)^2 ANN(t, \theta). \quad (5.30)$$

Así, se ha establecido la metodología para resolver sistemas de ecuaciones diferenciales utilizando PINN, con condiciones de frontera de Dirichlet o de Neumann.

5.7. Parciales

Una ecuación diferencial parcial es una ecuación que depende de múltiples variables independientes y las derivadas respecto a dichas variables. La forma más general de una EDO, dada una función escalar $u : \mathbb{R}^n \rightarrow \mathbb{R}$, con variables $t_1, t_2, \dots, t_n \in \mathbb{R}$ está dada por la ecuación 5.31.

$$\mathcal{F}(t_1, \dots, t_n; u, \frac{\partial u}{\partial t_1}, \dots, \frac{\partial u}{\partial t_n}; \frac{\partial^2 u}{\partial t_1 \partial t_1}, \dots, \frac{\partial^2 u}{\partial t_1 \partial t_n}; \dots) = 0. \quad (5.31)$$

Al igual que para el caso de las ecuaciones ordinarias, una ecuación diferencial parcial debe contar con algunas condiciones para ser determinada en su totalidad. Esto incluye algunas soluciones particulares respecto a alguna de las variables o condiciones iniciales similares a las de 5.8 o de entorno como las definidas en 5.9. Dicho eso, la manera de resolver una ecuación diferencial parcial es exactamente la

misma que cuando se aborda una ecuación diferencial ordinaria, pero en este caso tomando las derivadas parciales en vez de las totales.

5.8. PINN dependiente de las condiciones iniciales o parámetros libres

Dado que un sistema de ecuaciones depende de las condiciones iniciales y de entorno para quedar completamente determinado, cuando se resuelve un sistema de forma numérica, estas deben ser establecidas desde el inicio. Por lo tanto, para distintas combinaciones de condiciones iniciales y de frontera, es necesario implementar el código numérico por completo. Esto puede resultar computacionalmente muy costoso entre más predicciones se requieran.

Por otro lado, en los fenómenos físicos suele haber parámetros libres, representados aquí por el conjunto Θ , y cuyo valor determina los modelos que se obtendrán, ya sea al resolver sistemas de ecuaciones o al ser inferidos a partir de observaciones.

Una de las enormes ventajas que tiene el abordar las soluciones a ecuaciones diferenciales desde las redes neuronales es que, se puede crear una arquitectura de red y un entrenamiento tal que tanto las condiciones iniciales como los parámetros libres puedan formar parte de las variables de la red, y por ende, ser evaluadas en cualesquiera combinaciones de estos (Figura 5.2).

Para mostrar cómo se implementa esto dentro de todo lo que se ha descrito hasta ahora, se considerará el caso donde se tiene un sistema de ecuaciones, que deben cumplir condiciones de frontera de Dirichlet o de Neumann. Se asume también que el sistema tiene un conjunto de parámetros libres asociados Θ .

Lo primero es entender que en el caso cuando se tiene un sistema de ecuaciones diferenciales de la forma 5.22.

$$\vec{\mathcal{F}} \left(t, \vec{u}, \frac{d\vec{u}}{dt}, \frac{d^2\vec{u}}{dt^2}, \dots, \frac{d^n\vec{u}}{dt^n}, \Theta \right) = \vec{0}, \quad (5.32)$$

\vec{u} depende de la variable independiente t , por lo que, si $ANN \sim u$, entonces $ANN \equiv ANN(t)$. Es de notar que desde el inicio del capítulo se estableció que $\vec{\mathcal{F}}$ depende de los parámetros Θ , y estos entran en la ecuación de costo 5.24, pero estos son parámetros libres, por lo que determinan diferentes sistemas de ecuaciones para diferentes combinaciones de valores.

Es posible incluir la dependencia de Θ en la PINN, convirtiendo los parámetros libres en variables independientes, esto es:

$$ANN(t) \longrightarrow ANN(t, \Theta), \quad (5.33)$$

incluir los parámetros como variables implica que estos ya no son constantes dentro de la función \mathcal{F} respecto al cambio de los parámetros de la red θ , pues la medida del cambio en el error depende ahora de su cambio. Lo anterior se refleja cuando se calcula el gradiente del error respecto a Θ , aplicando regla de la cadena se obtiene 5.34.

$$\nabla_{\theta} \mathcal{L}_{\bar{\mathcal{F}}} = \frac{\partial \mathcal{L}_{\bar{\mathcal{F}}}}{\partial ANN} \frac{\partial ANN}{\partial \theta} \frac{\partial \theta}{\partial \Theta}, \quad (5.34)$$

$\frac{\partial \theta}{\partial \Theta}$ indica que ahora durante el proceso de entrenamiento, los parámetros de la red se ajustarán de acuerdo al comportamiento de los parámetros libres. Es decir, que al incluir distintas combinaciones de estos parámetros durante el entrenamiento, el modelo final podrá ser evaluado en esas combinaciones también. Es notar que al agregar más variables, las derivadas ordinarias deben ser ahora consideradas derivadas parciales, por lo que la ecuación 5.32 debe derivar en la función de costo 5.35.

$$\mathcal{L}_{\bar{\mathcal{F}}} = \sum_k^m \mathcal{L}_k = \sum_k^m \mathcal{F}_k \left(t, ANN, \frac{\partial ANN}{\partial t}, \frac{\partial^2 ANN}{\partial t^2}, \dots, \frac{\partial^n ANN}{\partial t^n}, \Theta \right)^2. \quad (5.35)$$

Como síntesis de lo anterior, se puede decir que incluyendo a los parámetros libres como variables de la red, se puede entrenar un modelo que pueda ser evaluado para distintas combinaciones de ellos sin necesidad de resolver el sistema de ecuaciones diferenciales nuevamente, que sería lo necesario en un método numérico tradicional.

Las condiciones iniciales y de frontera abordadas en esta tesis (Dirichlet y Neumann) pueden ser también integradas como variables del modelo provisto por la red neuronal (ver Figura 5.2). Dado que hay dos enfoques para cumplir las condiciones iniciales, ambos permiten ver a las condiciones iniciales como variables del modelo. En este trabajo sólo se considerará el método de parametrización.

La idea tras las condiciones iniciales variables es, en esencia, la misma tras los parámetros libres variables explicados arriba. Estas pasan de ser valores constantes, a variables independientes definidas sobre un rango, que serán evaluadas en la ecuación diferencial. De nuevo, partiendo desde un sistema de ecuaciones diferenciales definido como el sistema 5.32, y asumiendo que se trata de un problema de valor inicial de Dirichlet, pues el de Neumann es completamente análogo, se incluirán las condiciones iniciales como variables de la red neuronal.

$$ANN(t) \longrightarrow ANN(t, \vec{u}_0, \Theta), \quad (5.36)$$

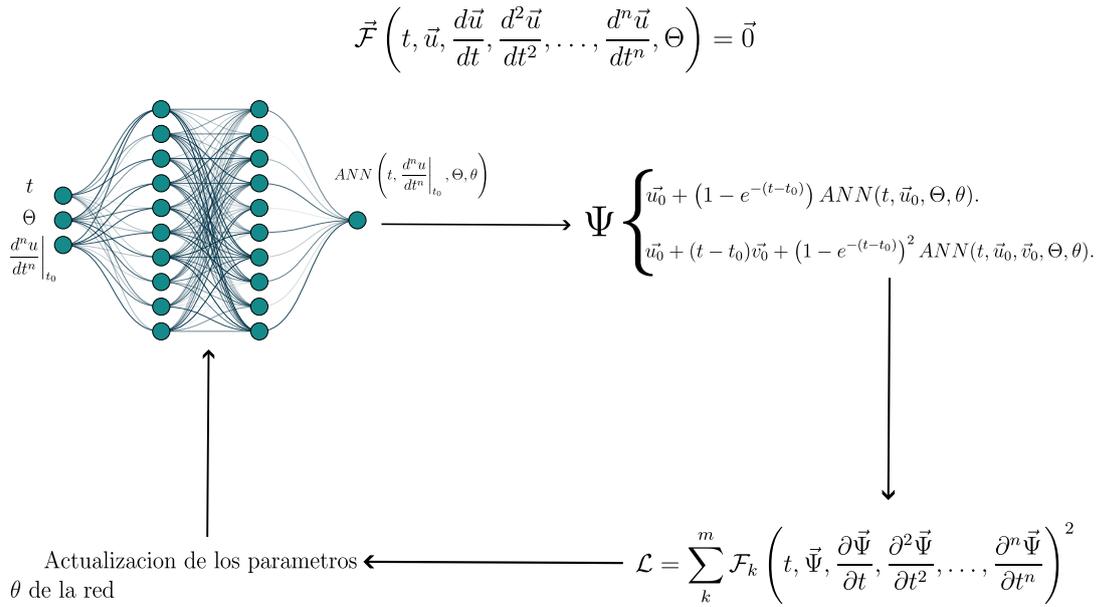


FIGURA 5.2: Parametrización de la red para la solución de un sistema de ecuaciones diferenciales con condiciones iniciales y parámetros libres como variables del modelo

donde la ecuación 5.36 representa que también ya se ha asumido que existen parámetros libres asociados al problema, y que serán incluidos como variables. Luego, con el fin de cumplir las condiciones iniciales del problema se tomará la parametrización 5.27, con las nuevas variables añadidas a la red neuronal.

$$\vec{\Psi}(t, \vec{u}_0, \Theta, \theta) = \vec{u}_0 + \left(1 - e^{-(t-t_0)}\right) ANN(t, \vec{u}_0, \Theta, \theta). \quad (5.37)$$

Luego, será la parametrización 5.37 la que entrará a la función de costo para proveer de las soluciones del sistema, por lo que la función de costo estará definida por 5.38.

$$\mathcal{L} = \sum_k^m \mathcal{F}_k \left(t, \vec{\Psi}(t, \vec{u}_0, \Theta, \theta), \frac{\partial \vec{\Psi}(t, \vec{u}_0, \Theta, \theta)}{\partial t}, \frac{\partial^2 \vec{\Psi}(t, \vec{u}_0, \Theta, \theta)}{\partial t^2}, \dots, \frac{\partial^n \vec{\Psi}(t, \vec{u}_0, \Theta, \theta)}{\partial t^n} \right)^2. \quad (5.38)$$

Donde nuevamente el subíndice k representa la k -ésima ecuación diferencial del sistema.

Capítulo 6

Metodología y resultados

En este capítulo se desarrollan los problemas de investigación relacionados con las aplicaciones cosmológicas de las PINN, explorando cada uno de los casos de ecuaciones diferenciales ampliamente tratados en los capítulos anteriores.

Todo el código utilizado en este trabajo está desarrollado en Python. Aunque existen librerías dedicadas específicamente al entrenamiento de PINN, como `neurodiffeq` desarrollada por Chen et al., 2020b, se decidió abordar cada problema implementando el código desde cero, utilizando únicamente la librería especializada en redes neuronales `Pytorch`.

La decisión de escribir el código desde cero es fruto de priorizar el entendimiento en profundidad y la naturaleza de las PINN, al construir cada parte del proceso de su entrenamiento. Y en la libertad que un código implementado desde cero otorga, como poder extensiones de él, para hacer híbridos de las PINN con otros paradigmas de las redes neuronales, tales como las Bayesianas o las recurrentes, entre otras combinaciones que se listan en (Lawal et al., 2022).

6.1. Ecuación diferencial con un parámetro libre: reparametrización vs función de costo

Para comenzar a describir el background de ecuaciones diferenciales cosmológicas, se tomará el caso cuando el universo está descrito a redshift cercanos, por lo que la contribución de la radiación puede ser despreciada y las componentes que determinan el comportamiento y evolución del cosmos son $\rho_m(z)$ y $\rho_\Lambda(z)$.

La evolución de este escenario estará determinada entonces el problema de valor inicial 6.1.

$$\begin{cases} \mathcal{F}\left(z, u, \frac{du}{dz}\right) = \frac{du}{dz} - \frac{3u}{1+z} = 0, \\ u(z_0 = 0) = \Omega_{m,0}. \end{cases} \quad (6.1)$$

En este caso, como se trata de un problema de valor inicial, se creará un modelo de PINN $ANN(t, \Omega_0)$, capaz de ser evaluado tanto en z , como en $\Omega_{m,0}$. Como es posible

utilizar tanto el enfoque de las condiciones iniciales como el de la reparametrización de ANN . Aquí se utilizarán ambos para mostrar que son efectivos y para comparar el desempeño de cada uno frente al otro.

En primer lugar, para incluir la condición inicial variable dentro de la función de costo, primero se considera la contribución de la ecuación diferencial \mathcal{F} , definida al sustituir la red en 6.1, que deriva en la ecuación 6.2.

$$\mathcal{L}_{\mathcal{F}} = \left(\frac{\partial ANN(z, \Omega_{m,0})}{\partial z} - \frac{3ANN(z, \Omega_{m,0})}{1+z} \right)^2, \quad (6.2)$$

Ahora, para incluir las condiciones iniciales dentro de la función de costo, como se describió en la sección 5.3.1, es necesario considerar la contribución de la condición inicial, sustituyendo la salida de la red neuronal en la ecuación 5.8. Por lo que la contribución de las condiciones iniciales en la función de costo está dada por la ecuación 6.3.

$$\mathcal{L}_{\Omega_{m,0}} = (ANN(0, \Omega_{m,0}) - \Omega_{m,0})^2. \quad (6.3)$$

Es de mencionar que en la práctica se utiliza un conjunto de entrenamiento que contiene n diferentes elementos $x = [z, \Omega_{m,0}]$, con diferentes combinaciones de valores dentro de un determinado rango, por lo que el costo total para el entrenamiento de la PINN está definido por la ecuación 6.4.

$$\mathcal{L} = \frac{1}{n} \sum_x (\mathcal{L}_{\mathcal{F}}) + \mathcal{L}_{\Omega_{m,0}}. \quad (6.4)$$

Por otro lado, para el método de la parametrización, es necesario reparametrizar la red neuronal utilizando la ecuación 5.15. Lo que resulta en la expresión 6.5

$$\Psi(z, \Omega_{m,0}) = \Omega_{m,0} + \left(1 - e^{-(z-z_0)}\right) ANN(z, \Omega_{m,0}), \quad z_0 = 0. \quad (6.5)$$

Como ya se detalló en secciones anteriores, la que será evaluada en la función de costo 5.27 es la ecuación 6.5. Para este caso, como sólo se cuenta con una sola ecuación diferencial, $k = 1$. Al igual que con la ecuación 6.4, se deben considerar todos los n elementos x que entran al entrenamiento de la red, por lo que la función de costo a minimizar para el método de parametrización es:

$$\mathcal{L} = \frac{1}{n} \sum_x \left(\frac{\partial \Psi(z, \Omega_{m,0})}{\partial z} - \frac{3\Psi(z, \Omega_{m,0})}{1+z} \right)^2. \quad (6.6)$$

Es la ecuación 6.6 la que se minimizó para obtener el modelo usando el método de parametrización.

Para comparar ambos enfoques se utilizó el mismo dataset de entrenamiento para cada caso. Los rangos para las variables que se tomaron, fueron $z \in [0, 3]$ y $\Omega_{m,0} \in [0.2, 0.5]$, estos rangos fueron elegidos porque es el rango del corrimiento al rojo donde es válido ignorar la contribución de la radiación, y el rango donde se encuentra el valor de la densidad de materia al día de hoy.

El dataset para el entrenamiento de la red será todas las posibles combinaciones entre los elementos de ambos intervalos: $x = [0.0, 3.0] \times [0.2, 0.5]$. En este caso fueron 100 elementos para z y 40 para $\Omega_{m,0}$, por lo que el conjunto de entrenamiento x contó con 4000 elementos. Adicionalmente de tomar el producto cruz de los intervalos, se hizo una permutación aleatoria de los elementos de x , con el fin de que el modelo reduzca la probabilidad de aprender el orden de los elementos y sólo memorizarlos.

En cuanto al loop de aprendizaje, que es donde se lleva a cabo el proceso de minimización de la función de costo y la actualización de los parámetros de la red, se eligió usar el optimizador Adam con una tasa de aprendizaje variable. Se designó que para que la tasa de aprendizaje lr no fuera una variable al entrenar los modelos, esta tomara tres valores para todos los entrenamientos $lr = 10^{-2}, 10^{-3}, 5 \times 10^{-4}$. Dichas tasas de aprendizaje se implementaron durante $10^4, 4 \times 10^4, 3 \times 10^4$ épocas respectivamente.

Para validar el desempeño de ambos modelos en función de la arquitectura utilizada, se utilizó una arquitectura de dos capas ocultas con una cantidad de nodos variable. La cantidad de nodos fue de 20, 30 y 50 y se hizo una prueba de desempeño por cada una de las arquitecturas para ambos enfoques que llevan a la red a cumplir con las condiciones iniciales: el de parametrización de la red y el de incluir las condiciones iniciales directamente dentro de la función de costo.

La prueba de desempeño consistió en crear una malla con todas las posibles combinaciones entre los intervalos de z y $\Omega_{m,0}$ para los que la red fue entrenada, para posteriormente calcular el error porcentual entre el modelo provisto por la red y la solución real provista por la solución analítica. El desempeño del enfoque de las condiciones iniciales dentro de la función de costo se puede observar en la figura 6.1 y para el enfoque de parametrización en la figura 6.2.

Los resultados mostraron que el enfoque de parametrización (Figura 6.2) es superior al de la función de costo (Figura 6.1) para la cantidad de datos y arquitecturas establecidas, pues para las cuatro arquitecturas diferentes que fueron elegidas, las de parametrización mostraron converger todas con un error porcentual total del 1 % para la de 20 nodos, hasta un error porcentual del 0.2 % para la de 60 nodos. Por otro lado para el enfoque donde las constricciones están en la función de costo, sólo convergieron de forma exitosa las arquitecturas con 20 y 30 nodos, las de 50 y 60 no lograron converger de manera exitosa.

El hecho de que el enfoque de la función de costo "perdiera" en esta comparación

Error porcentual: Λ – CDM costo

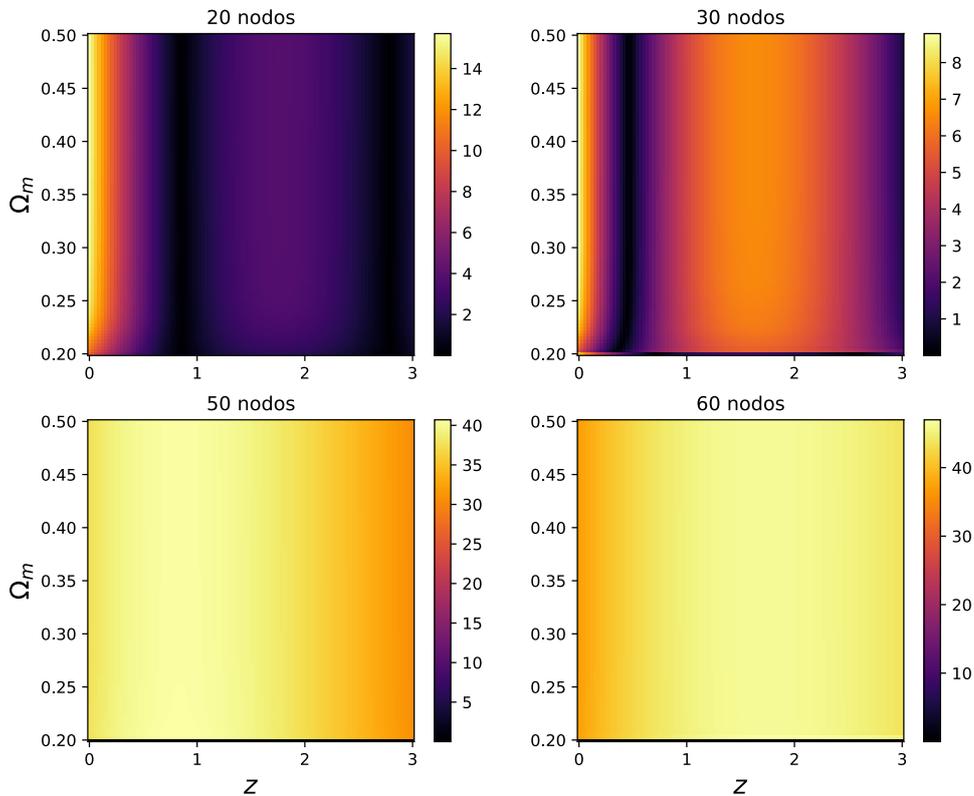


FIGURA 6.1: Mapas de densidad del error porcentual por cada modelo, bajo el enfoque de la función de costo. Cada una con dos capas ocultas, pero diferente número de nodos por capa. Cada punto del mapa es una combinación del corrimiento al rojo z y la condición inicial $\Omega_{m,0}$.

se debe a que entrenar una PINN bajo este enfoque requiere de una cantidad de datos importante, aproximadamente del orden de 10^2 elementos por cada variable independiente, una cantidad mucho menor que la utilizada aquí.

De lo anterior se concluye que una cantidad de nodos alta puede mejorar el desempeño cuando se parametriza la red, aunque se debe considerar un balance entre desempeño y complejidad, pues más nodos implican también una complejidad computacional mayor. Así que la elección dependerá de la precisión requerida en el problema.

El enfoque de la función de costo es inferior bajo las condiciones de entrenamiento establecidas aquí, pero esto se debe a que son necesarios gran cantidad de puntos para entrenar una red que depende completamente de la función de costo. Por lo que este enfoque trabaja mejor cuando se tiene gran cantidad de datos para su entrenamiento, o donde no es posible parametrizar la red neuronal, como en el caso cuando se incluyen datos observacionales prescindiendo de condiciones iniciales.

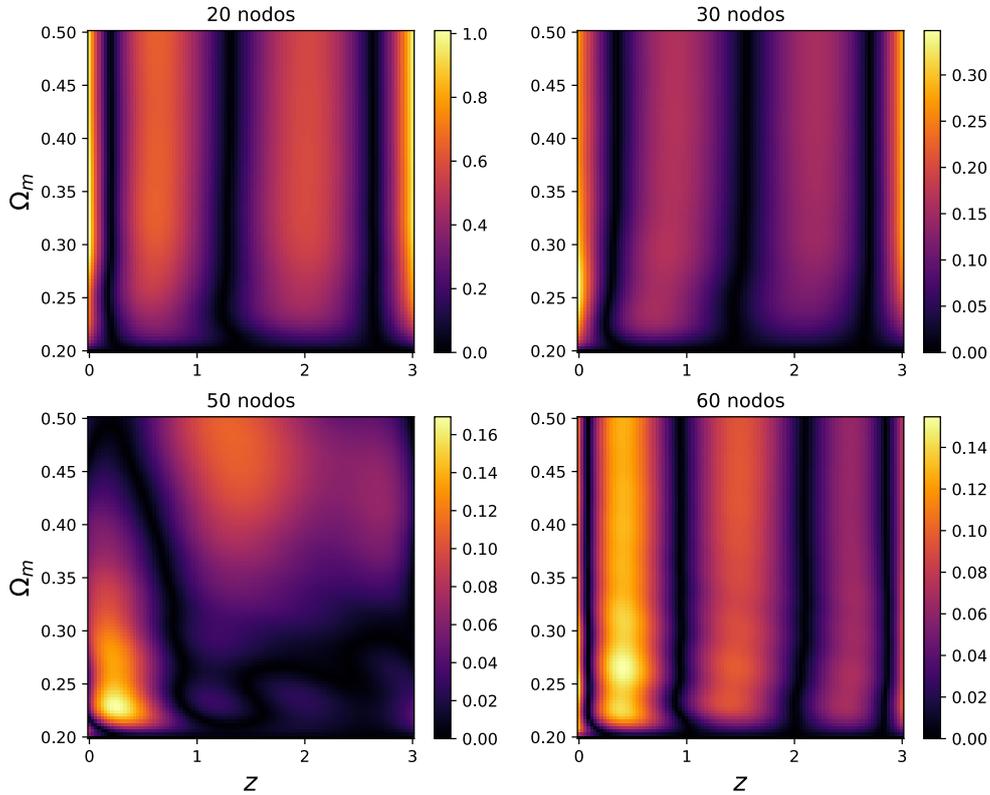
Error porcentual: Λ – CDM


FIGURA 6.2: Mapas de densidad del error porcentual por cada modelo, bajo el enfoque de parametrización de la PINN. Cada una con dos capas ocultas, pero diferente número de nodos por capa. Cada punto del mapa es una combinación del corrimiento al rojo z y la condición inicial $\Omega_{m,0}$.

6.2. Distancias cosmológicas en el modelo CPL: PINN para acelerar inferencia de parámetros

En este problema de investigación se muestran varias de las capacidades de las PINN para abordar problemas computacionales cosmológicos. Lo anterior se hizo resolviendo la integral 2.36 que define la distancia luminosa, pero bajo el modelo de energía oscura paramétrica CPL, para así probar cómo una PINN puede ser entrenada de tal manera que los parámetros libres de la ecuación original sean ahora variables del modelo provisto por la red neuronal, en este caso los parámetros de CPL ω_0, ω_a . Finalmente, dado que en cosmología la inferencia de parámetros es una tarea crucial y en la mayoría de ocasiones muy costosa computacionalmente, se procedió a hacer una inferencia de los parámetros antes mencionados utilizando el algoritmo de Metropolis–Hastings.

Para comenzar, se tomó el ejemplo ya mencionado en la sección 5.8 donde la

integral para evaluar la distancia cosmológica queda definida en la ecuación 3.17, por lo que al hacer lo que se explicó en la sección 5.5 la función de costo que resulta de expresar la integral como una ecuación diferencial está dada por la ecuación 5.20

$$\frac{du}{dz} - \frac{1}{E_{CPL}(z)} = 0, \quad (6.7)$$

$$\text{donde } E_{CPL}(z) = \sqrt{\Omega_{m,0}(1+z)^3 + (1 - \Omega_{m,0})(1+z)^{3(1+\omega_0+\omega_a)} \exp\left(-\frac{3\omega_a z}{1+z}\right)}.$$

Como se espera que la distancia para un objeto que se encuentra a un $z = 0$ sea 0, entonces la condición inicial para esta ecuación diferencial es $u(0) = 0$. Ahora, con el fin de que la solución provista por la red cumpla con ese problema de valor inicial, se realizará el enfoque de parametrización de la red neuronal. Puesto que se trata de un problema de valor inicial, la parametrización de la red será con base en la ecuación 5.15.

Con el fin de mostrar que la red neuronal puede recibir como variables a los parámetros libres de la ecuación diferencial, w_0, w_a , estos serán introducidos junto con el corrimiento al rojo en la entrada de la red, como se describió en la sección 5.8. Por lo que, la parametrización de la red queda definida por la ecuación 6.8.

$$\Psi_{d_L}(z, \omega_0, \omega_a) = u(z_0) + \left(1 - e^{-(z-z_0)}\right) ANN(z, \omega_0, \omega_a), \quad z_0 = 0. \quad (6.8)$$

Una vez definida la parametrización, se define la función de costo \mathcal{L}_{d_L} basada en la ecuación diferencial 6.7 y que está definida por la ecuación 6.9.

$$\mathcal{L}_{d_L} = \left(\frac{\partial}{\partial z} \Psi_{d_L}(z, \omega_0, \omega_a) - \frac{1}{E_{CPL}} \right)^2. \quad (6.9)$$

Para este problema en particular se escogió la arquitectura con 3 nodos en la entrada, 3 capas ocultas con 20 nodos cada una y \tanh como activación y 1 nodo en la capa de salida.

El conjunto de entrenamiento par este modelo se tomó con los siguientes rangos: $z \in [0, 3]$, $\omega_0 \in [-2, 0]$ y $\omega_a \in [-2, 2]$. Donde se tomaron 80 puntos para z , y 10 puntos para ω_0, ω_a cada uno. Para que la red neuronal pueda hacer predicciones sobre cualquier combinación de la variable independiente y los parámetros libres, el conjunto total de entrenamiento se tomó como el conjunto de todas las posibles combinaciones de estos elementos, usando el producto cartesiano de cada intervalo $x = [0.0, 3.0] \times [-2.0, 0.0] \times [-2.0, 2.0]$. Finalmente se hizo una permutación aleatoria de x para evitar sesgos por el orden de sus elementos en el entrenamiento.

Para el proceso de entrenamiento se decidió utilizar una tasa de aprendizaje variable al igual que con el entrenamiento de Λ -CDM, en este caso sólo se utilizaron

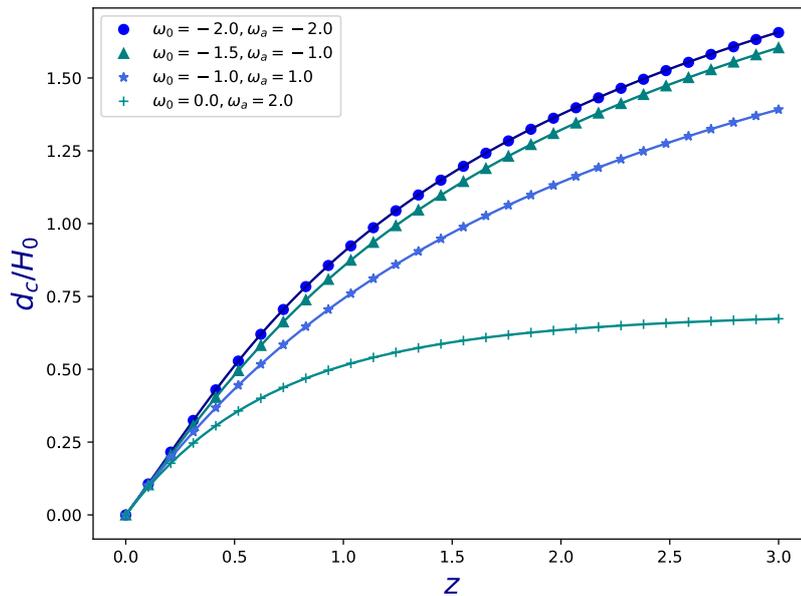


FIGURA 6.3: Algunas soluciones a la integral que define la distancia comóvil dadas por el modelo de la PINN para diferentes combinaciones de los parámetros libres ω_0, ω_a .

dos valores diferentes: $l_r = 0.01$ para las primeras 10,000 épocas y $l_r = 0.001$ para las siguientes 2000. Una vez que el modelo fue entrenado, se tomaron los mapas del error porcentual de z contra ω_a para algunos ω_0 fijos, como se puede observar en la figura 6.4.

Cuando el modelo mostró alcanzar un error porcentual aceptable, este puede ser evaluado en cualquier combinación de los parámetros libres, como se puede ver en la Figura 6.3. Así pues, se procedió a mostrar que la PINN tiene una ventaja sobre un modelo numérico convencional, en cuanto al desempeño computacional se refiere. La prueba también tiene como objetivo mostrar que las PINN tienen un potencial enorme en la astrofísica y la cosmología, por lo que se decidió usar la solución de la ecuación 6.7 para calcular la distancia luminosa cosmológica y posteriormente hacer la inferencia de los parámetros ω_0, ω_a utilizando el algoritmo de metropolis-hastings descrito en el algoritmo 2.

Error porcentual: CPL 20 nodos

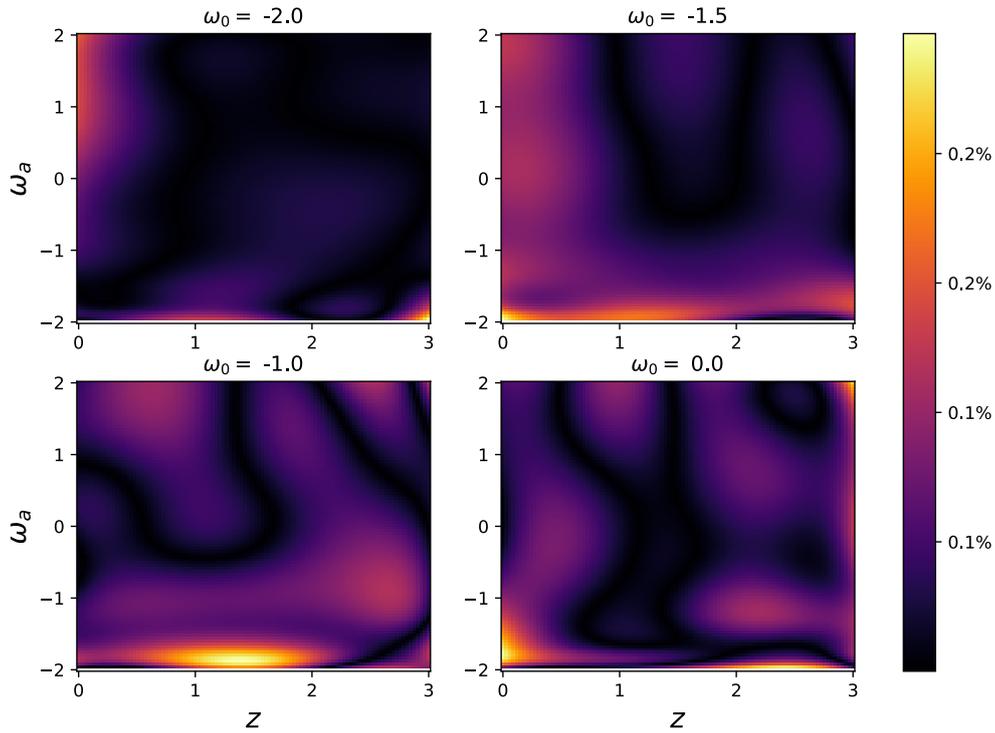


FIGURA 6.4: Mapa de densidad de error para distintos valores del corrimiento al rojo z y los parámetros ω_0, ω_a .

Algorithm 2 Metropolis-Hastings para Inferencia Cosmológica

- 1: **Entrada:** $ANN_{dL}, \omega_0, \omega_a, steps, \sigma$
 - 2: **Salida:** mejores_omegas
 - 3: **Inicializar:**
 - 4: $\omega_{actual} \leftarrow [\omega_0, \omega_a]$
 - 5: mejores_omegas $\leftarrow \emptyset$
 - 6: **for** $i = 1$ **to** $steps$ **do**
 - 7: $\omega_{propuesto} \leftarrow \omega_{actual} + \mathcal{N}(0, \sigma^2)$
 - 8: $ratio \leftarrow ANN_{dL}(\omega_{propuesto}) / ANN_{dL}(\omega_{actual})$
 - 9: **if** $ratio \leq 1$ **then**
 - 10: $\omega_{actual} \leftarrow \omega_{propuesto}$
 - 11: **else**
 - 12: $u \leftarrow \text{Uniforme}(0, 1)$
 - 13: **if** $u > ratio$ **then**
 - 14: $\omega_{actual} \leftarrow \omega_{propuesto}$
 - 15: **end if**
 - 16: **end if**
 - 17: mejores_omegas $\leftarrow mejores_omegas \cup \{\omega_{actual}\}$
 - 18: **end for**
 - 19: **return** mejores_omegas
-

Lo primero que se hizo fue definir la función *distancia cosmológica* a partir del modelo provisto por la PINN: $d_L(z, \omega_0, \omega_a)$, definida por la ecuación 6.10, donde es necesario regresar a las unidades que incluyen c , para ajustar con los datos observacionales.

$$d_L(z, \omega_0, \omega_a) = \frac{c(1+z)}{H_0} \Psi_{d_L}(z, \omega_0, \omega_a). \quad (6.10)$$

Los datos del módulo de distancia de supernovas mencionados en la sección 2.5, representan la relación entre el módulo de distancia μ y el corrimiento al rojo z , por lo que es necesario definir el módulo de distancia provisto por el modelo de la PINN, y eso está dada por la ecuación 6.11.

$$\mu_\Psi(z, \omega_0, \omega_a) = 25 + 5 \log_{10}(d_L(z, \omega_0, \omega_a)). \quad (6.11)$$

Y es la ecuación μ_Ψ la que se utiliza para hacer procesos de inferencia con los datos observacionales. Como se mencionó arriba, el proceso de inferencia se realizó utilizando el algoritmo 2 dos veces: una obteniendo las distancias cosmológicas utilizando la red al evaluar la función 6.11 y el otro, integrando numéricamente la ecuación 6.12.

$$(1+z) \int_0^z \frac{1}{E_{CPL}(z')} dz', \quad (6.12)$$

y utilizando su resultado para calcular el módulo de distancia μ de forma numérica. El proceso de inferencia consideró como función a optimizar la función de máxima verosimilitud asociada a los datos de *SHOES* (Riess et al., 2022), definida por la ecuación 6.13.

$$L(\Theta) \propto \exp\left(-\frac{1}{2} \mu^T C^{-1} \mu\right) \quad (6.13)$$

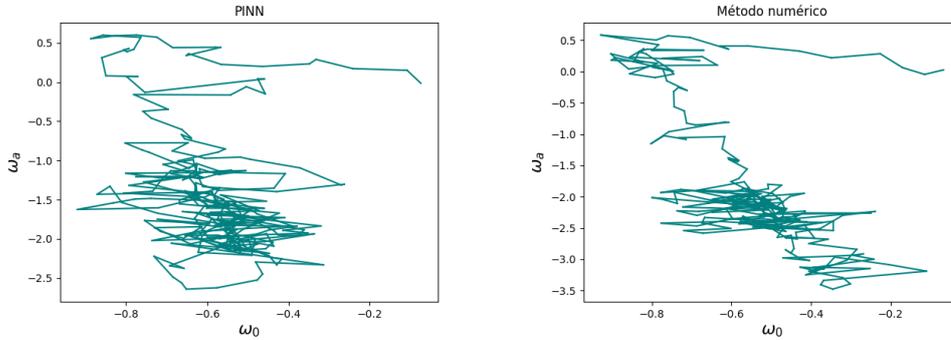
donde μ es la diferencia entre el módulo de distancia del modelo; μ_Ψ y el observacional μ_{obs} , $\mu = \mu_{obs} - \mu_\Psi$; C^{-1} es la inversa de la matriz de covarianza asociada a los datos; y Θ son los parámetros a inferir, en este caso $\Theta = \{\omega_0, \omega_a\}$. Para simplificar las operaciones, usualmente a la expresión 6.13 se le aplica el logaritmo natural y se multiplica por -1 .

$$-\ln\{L(\omega_0, \omega_a)\} = \frac{1}{2} \mu^T C^{-1} \mu. \quad (6.14)$$

Al minimizar la función dada por 6.14 se alcanza el óptimo para la función de máxima verosimilitud 6.13. Por lo que el algoritmo de inferencia se concentró en minimizar la función 6.14. El resultado en tiempo computacional para este proceso de inferencia se puede ver en la tabla 6.1.

Método	Tiempo promedio \pm std (ms)	Punto encontrado	Órdenes de magnitud
PINN	1.33 s \pm 21.4	$[-0.516, -2.081]$	$\mathcal{O}(0.93)$
Método numérico	11 s \pm 173	$[-0.504, -2.231]$	-

TABLA 6.1: Comparación del tiempo de cómputo



(A) Cadena resultante de aplicar el algoritmo de Metropolis-Hastings obteniendo la distancia luminosa a utilizando la PINN.

(B) Cadena que resultó de hacer el algoritmo metrópolis-Hastings obteniendo la distancia luminosa a partir de una integral numérica.

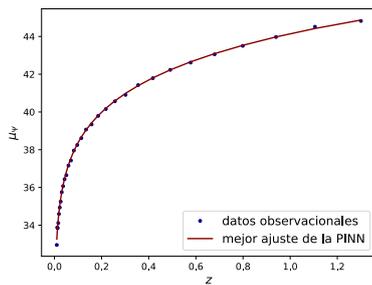
FIGURA 6.5: Cadenas resultantes de aplicar el algoritmo metrópolis hastings utilizando los datos de SH0ES, para inferir los parámetros ω_0, ω_a . Ambas cadenas convergieron a puntos mínimos similares, pero en diferentes tiempos.

La PINN propuesta logró una mejora significativa en el tiempo de cómputo, con un tiempo promedio de $1.341s \pm 0.011s$, en comparación con $11.107s \pm 0.107s$ del método numérico, lo que representa una mejora del 88.23 % al considerar el tiempo del método numérico menos el tiempo de la PINN, y tomando el tiempo numérico como el 100 %. Lo anterior significa que se redujo el tiempo computacional requerido para hacer la inferencia en un orden de magnitud. Las cadenas obtenidas al hacer la inferencia en ambos casos puede observarse en la figura 6.5

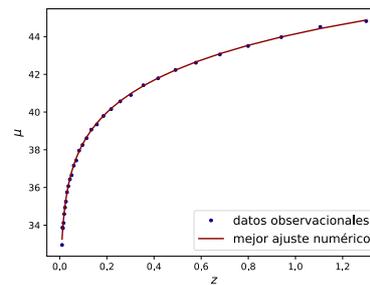
Ambas cadenas convergieron de forma exitosa a un punto mínimo $[\omega_0, \omega_a] = [-0.516, -2.081]$ cuando se utilizó la PINN y $[\omega_0, \omega_a] = [-0.504, -2.231]$ para el método numérico. Dado que ambos puntos óptimos son muy similares, y ambos hacen un buen ajuste a los datos observacionales del módulo de distancia, como se puede observar en la figura 6.6. La diferencia entre ambos procesos de inferencia radica casi por completo en la diferencia de tiempo computacional entre un método numérico tradicional y el desempeño mejorado por la PINN.

La arquitectura designada mostró emular de gran manera las soluciones de la integral en el rango del espacio de parámetros, con errores inferiores al 0.2 % verificando así la capacidad de la PINN en la resolución de Integrales y en la capacidad para evaluar en diferentes combinaciones de parámetros.

Dado lo anterior, una de las aplicaciones fundamentales de una red que puede



(A) Mejor ajuste con los parámetros encontrados usando integración numérica.



(B) Mejor ajuste con los parámetros encontrados usando integración numérica.

FIGURA 6.6: comparación del ajuste utilizando ambos métodos a los datos del módulo de distancia frente al corrimiento al rojo.

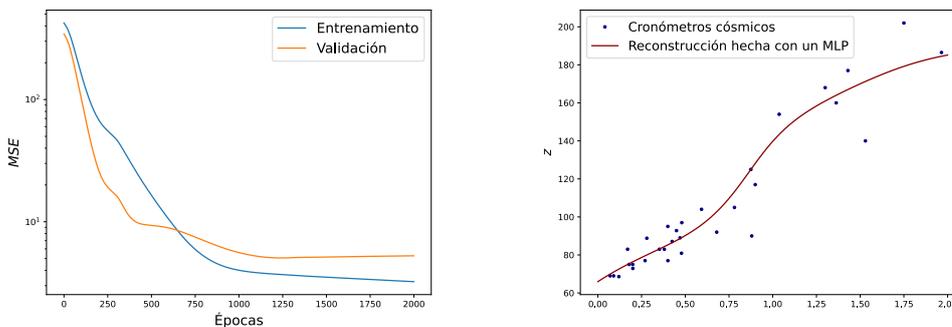
ser evaluada en diferentes elementos en el espacio de parámetros es ser el sustituto de métodos numéricos tradicionales para ahorrar tiempo de cómputo. Como en astrofísica y cosmología la inferencia de parámetros es una tarea crucial.

La prueba de desempeño de la red mostró que fue capaz de encontrar una combinación de parámetros consistente con la literatura, pero haciendo que el tiempo de cómputo se redujese en un orden de magnitud.

6.3. Ecuación de Klein-Gordon en Quintaesencia usando datos de cronómetros cósmicos

En este problema de investigación se abordará la ecuación de Klein-Gordon, en el marco cosmológico de Energía Oscura como campo escalar, mencionado en la sección 3.3 denominado *Quintaesencia*. Para comenzar, es necesario presentar una reconstrucción de una ecuación cosmológica hecha por un MLP estándar.

Para lo anterior se tomaron los datos (Tabla 2.1) de cronómetros cósmicos, que describen la tasa de expansión del universo en función del corrimiento al rojo. Esta reconstrucción se hizo utilizando un MLP convencional de una capa oculta con 50 nodos, con activación \tanh . La reconstrucción llegó a un error MSE de 3.0 y la evolución de la función del costo durante el entrenamiento y la reconstrucción final se puede ver en la Figura 6.7.



(A) Evolución del error durante el entrenamiento. El comportamiento de las curvas muestran un buen aprendizaje, aunque con tendencia al sobreajuste.

(B) Modelo provisto por el MLP al hacer la reconstrucción. Es evidente que la red tiende a replicar el comportamiento ruidoso de los datos.

FIGURA 6.7: Muestra de cómo una reconstrucción provista por un MLP, donde la red tiende a aprender también el comportamiento ruidoso de los datos.

Las redes neuronales tienden a aprender mucho del comportamiento ruidoso o sesgado de los datos con los que son entrenadas, es por eso que son tan propensas al sobre-ajuste. Lo anterior es uno de los principales problemas por el que las reconstrucciones hechas con redes neuronales convencionales (MLP) aún no son de las primeras opciones al generar modelos físicos a partir de datos.

Sin embargo, como se mencionó en el capítulo 1 al contar con una directriz matemática que guíe el entrenamiento, una PINN es capaz de ser robusta ante los datos ruidosos. Más aún, al combinar una PINN con datos observacionales como se describió en la sección 5.4 los beneficios son dobles, pues se soluciona el problema del sobre-ajuste que tienen las redes neuronales convencionales, gracias a la base matemática de la PINN, y los datos observacionales permiten fijar una solución a la ecuación diferencial utilizada.

En este caso se tomó las ecuaciones asociadas a la evolución del campo escalar de Quintaesencia. La ecuación a resolver es la ecuación de Klein-Gordon 3.22

$$\ddot{\phi} + 3H\dot{\phi} + \frac{dV}{d\phi} = 0. \quad (6.15)$$

El potencial utilizado para este problema es tomado de las propuestas de (Copeland, Liddle y Wands, 1998), siendo este uno de perfil exponencial, definido por la ecuación 6.16.

$$V(\phi) = V_0 e^{-\lambda\sqrt{\kappa_0}\phi}, \quad (6.16)$$

donde λ es un parámetro libre del problema que en esta ecuación será fijado al valor $\lambda = 0.008$ propuesto por (Chantada et al., 2023) al ser encontrado en un proceso de inferencia estadística.

Como se mencionó en la sección 3.3, la ecuación de Klein-Gordon puede resolverse como un sistema de dos por dos y acoplado descrito por el cambio de variable 3.24. Así que el sistema quedará dado en términos de las ecuaciones 3.25:

$$\begin{cases} \frac{dx}{dN} = -3x + \frac{\sqrt{6}}{2}\lambda y^2 + \frac{3}{2}x(1 + x^2 - y^2) \\ \frac{dy}{dN} = -\frac{\sqrt{6}}{2}xy\lambda + \frac{3}{2}y(1 + x^2 - y^2) \end{cases} \quad (6.17)$$

Como se explicó en el capítulo respectivo, este sistema tiene un par de condiciones iniciales x_0, y_0 asociado. Sin embargo para este problema se utilizó el método descrito en 5.4, donde se combina la directriz matemática provista por el sistema 3.25 con los datos observacionales de los cronómetros cósmicos nuevamente.

Los datos de cronómetros cósmicos deben seguir la expansión del universo, por lo que la solución al sistema 3.25 provisto por la red debe ser tal que al calcular la ecuación de Friedmann de Quintaesencia 6.18 derivada de la ecuación de restricción para este sistema dada por 3.26.

$$H = H_0 \sqrt{\frac{\Omega_{m,0}(1+z)^3}{1-x^2-y^2}}. \quad (6.18)$$

Por razones de simplicidad se fijaron ambos parámetros libres de la ecuación de Friedmann, donde se asumió que sus valores son $\Omega_{m,0} = 0.3$ y $H_0 = 67.0$. Por lo que la función que describe al sistema de ecuaciones es:

$$\mathcal{F} = \left(\frac{dx}{dN}, \frac{dy}{dN} \right), \quad (6.19)$$

donde $\frac{dx}{dN}, \frac{dy}{dN}$ se sustituyen por sus valores dados en la ecuación 3.25. Entonces la contribución debido a la ecuación diferencial se define por la ecuación 6.20.

$$\mathcal{L}_{\mathcal{F}} = \|\mathcal{F}(ANN_x, ANN_y)\|^2. \quad (6.20)$$

Por otro lado, la contribución que harán los datos observacionales estará dado por el error cuadrático medio (MSE), entre el modelo y los datos: \mathcal{L}_{MSE} . Para calcularlo es necesario que al mismo tiempo que se generan unos primeros parámetros aleatorios para la red neuronal, las salidas (ANN_x, ANN_y) son evaluadas en la expresión 6.18, y usadas para calcular el valor de H evaluado en el z observacional y evaluar el MSE teórico obtenido con el observado de los datos.

Finalmente la función total del costo para este problema de investigación está dada por la relación 6.21. Donde se escogió un coeficiente de regulación muy pequeño para la contribución de los datos observacionales para minimizar lo más posible la inyección del ruido en la solución del sistema de ecuaciones diferenciales y en el propio modelado de la ecuación 6.18. Se vuelve a recalcar que sólo esas dos cosas se le dieron a la PINN para su entrenamiento, NO hubo contribución de condiciones de frontera o de valores iniciales.

$$\mathcal{L} = \mathcal{L}_{\mathcal{F}} + 5 \times 10^{-6} \mathcal{L}_{MSE}. \quad (6.21)$$

La red neuronal designada para esta tarea fue una con dos capas ocultas con 50 nodos cada una y activación \tanh . Esta PINN se entrenó para un corrimiento al rojo en el intervalo $z \in [0.0, 2.4]$, utilizando 1000 elementos para el entrenamiento. El número de épocas fue de 5000 con una tasa de aprendizaje de $lr = 0.01$.

Una vez entrenado el modelo, éste llegó a un valor de la función de costo de $\mathcal{L} = 0.000786$. Para mostrar el desempeño del modelo provisto por la PINN, se muestra la comparación con la solución numérica, obtenida al tomar como condiciones iniciales los valores de $x_0, y_0 = ANN(z = 10) = x_0 = 0.00063$ y $y_0 = 0.04679$.

Las soluciones individuales para x, y se cambian de variable para poder analizarlas en función del corrimiento al rojo z , pues es la observable en cuestión y la variable en la que están dados los datos observacionales usados aquí. Ambas soluciones al sistema se pueden observar en la figura 6.8.

Una vez obtenido un modelo proveniente de una red neuronal es posible hacer cualquier tipo de manipulación matemática. Por lo que una vez que se tienen las soluciones adimensionales x, y , es posible invertir el cambio de variable dado por 3.25 regresando a las soluciones originales ϕ, V , y así obtener un modelo para la densidad ρ_ϕ y la presión p_ϕ , definidas por las ecuaciones 6.22 y 6.23.

$$\rho_\phi = \dot{\phi}^2 + V(\phi), \quad (6.22)$$

$$p_\phi = \dot{\phi}^2 - V(\phi). \quad (6.23)$$

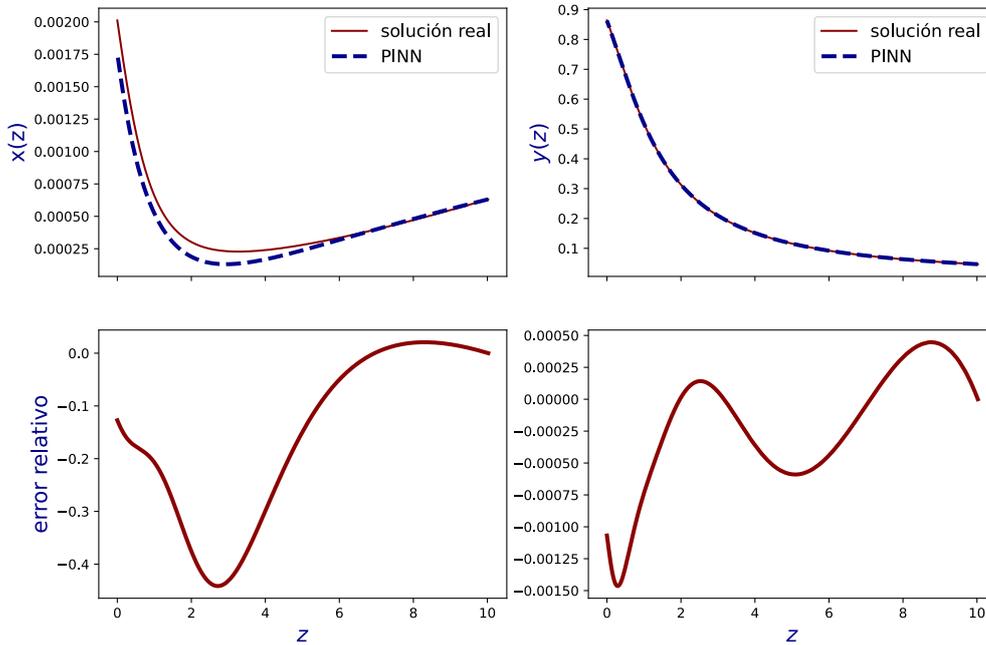


FIGURA 6.8: Solución al sistema de ecuaciones 3.25 dado por la PINN en comparación con la solución numérica. También se muestra el error relativo del modelo dado por la red.

Y un resultado muy importante que se espera de cualquier modelo que describa la evolución del campo escalar es poder modelar de igual manera la ecuación de estado ω , que se define a partir de las ecuaciones 6.22 y 6.23 y está dada por la relación 6.24.

$$\omega = \frac{2x^2}{x^2 + y^2} - 1. \tag{6.24}$$

Por lo tanto, la ecuación de estado también fue obtenida gracias a la PINN y su comportamiento así como su error relativo dado por la relación $\frac{\text{real} - \text{predicho}}{\text{real}}$. La comparación entre el modelo provisto por la red neuronal y la solución numérica se aprecia en la Figura 6.9.

Hasta aquí se obtuvo un modelo de red que describe a $x(z), y(z)$ y a partir del cual se pueden expresar ρ, p, ω , sin haberle asignado un conjunto de condiciones iniciales x_0, y_0 . También es posible evaluar el modelo en cualquier valor del intervalo de z , para el que fue entrenado, obteniendo así el valor de las condiciones iniciales con base en los datos observacionales de cronómetros cósmicos. Utilizando los valores obtenidos al evaluar el propio modelo provisto por la PINN, se tiene que en $z = 10$, el valor del conjunto de condiciones iniciales debería ser: $x_0 = 0.00063$ y $y_0 = 0.04679$, valores consistentes con los encontrados por (Chantada et al., 2023).

Este problema de investigación mostró que es posible utilizar una PINN no sólo

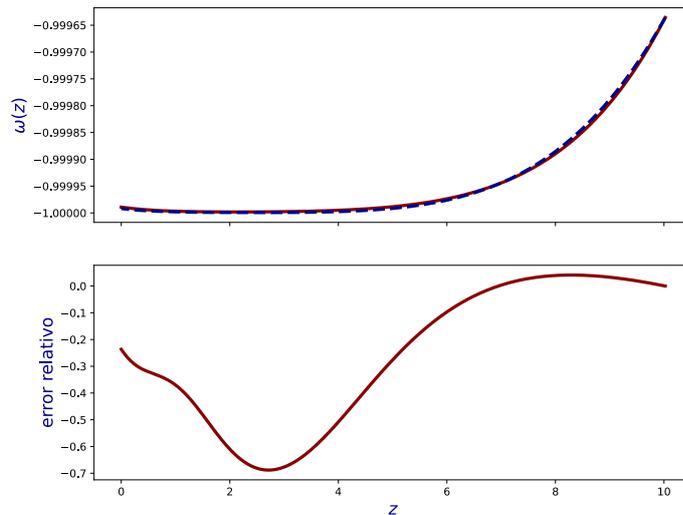
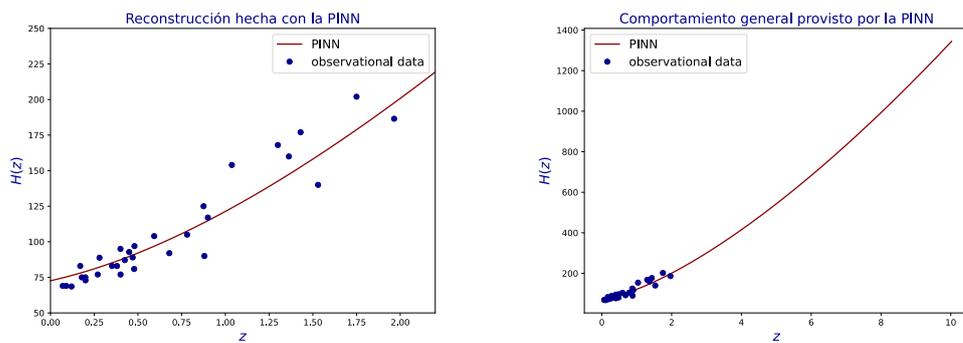


FIGURA 6.9: Evolución de la ecuación de estado asociada al campo escalar de quintaesencia, obtenida gracias al modelo provisto por la PINN.

para resolver una ecuación de orden superior, sino que también se puede prescindir de las condiciones iniciales del sistema que se esté abordando siempre y cuando se cuenten con datos observacionales y una expresión matemática que los describa y que dependa de las soluciones del sistema, como la ecuación de Friedmann, para este caso.



(A) Reconstrucción de la ecuación de Hubble de la PINN en el rango de los datos observacionales.

(B) Generalización de la reconstrucción de la ecuación de Friedmann más allá del rango de los datos.

FIGURA 6.10: Reconstrucción de la ecuación de Friedmann con base en los datos observacionales de cronómetros cósmicos y la directriz matemática de la ecuación de Klein-Gordon en el marco de Quintaesencia.

La PINN pudo lograr dos cosas, resolver la ecuación de Klein-Gordon, dando soluciones consistentes y de las cuales, se pudo inferir que las condiciones iniciales obtenidas son consistentes con la literatura. Además de poder realizar al mismo

tiempo una reconstrucción de la ecuación de Friedmann a partir de los datos de cronómetros cósmicos que no sólo no sufre de sobreajuste, sino que es capaz de hacer predicciones para grandes valores de corrimiento al rojo, mucho más allá del rango de los datos que le fueron provistos.

Finalmente una vez que se obtuvieron las soluciones x, y del sistema, fue posible obtener el comportamiento de la densidad del campo ρ y su presión p , y la más importante, un modelo para poder estudiar el comportamiento de la ecuación de estado $\omega(z)$. Para este problema se fijaron los parámetros libres $\Omega_{m,0}$ y H_0 , pero como se mostró en el capítulo de CPL, estos pueden ser establecidos como variables y que entren a la red, para luego ser determinados en un proceso de inferencia estadística.

Capítulo 7

Discusión y conclusiones

En este trabajo se abordó el uso de un paradigma de las Redes Neuronales Artificiales que incluye una modificación en el proceso de entrenamiento de un MLP, al incluir una expresión matemática en la función de costo, algo que no es posible en los perceptrones multicapa tradicionales.

Lo anterior se llevó a cabo a través de la implementación de redes neuronales físicamente informadas en tres problemas de investigación, con el fin de que fueran tres escenarios diferentes dentro de problemas relevantes en la astrofísica y la cosmología actuales. Con los resultados de los tres escenarios se obtuvieron muestras de las mejoras sustanciales que este tipo de inteligencia artificial puede aportar en dichas áreas y también varias de las consideraciones que se deben tener al utilizarlas.

En primer lugar se mostró cómo una PINN puede resolver la dinámica de un modelo cosmológico, al resolver el sistema de ecuaciones en el que se asienta, el modelo para este caso fue Λ -CDM, al mismo tiempo se mostró que cuando al abordar la resolución de un sistema de EDO, la manera más eficiente y precisa para cumplir con condiciones iniciales o de frontera es a través del enfoque de parametrización de la PINN. Lo anterior no se había puesto a prueba en la literatura anteriormente en un mismo trabajo.

Cuando se trabajó con el modelo CPL, se abordó un problema latente en la astrofísica y cosmología modernas, la complejidad computacional que conlleva realizar ciertas tareas, tales como la inferencia de parámetros. Al resolver la integral que define la distancia luminosa en el marco CPL, se demostró que una PINN también es capaz de resolver cualquier integral, más aún, es capaz de hacerlo incorporando como variables los parámetros libres del sistema de ecuaciones a resolver: ω_0, ω_a . Más aún, al obtener la solución a la distancia luminosa se hizo una inferencia de los parámetros estadística comparando el tiempo computacional la PINN contra un método numérico, donde la red mostró reducir el tiempo de cómputo en un orden de magnitud $\sim \mathcal{O}(10)$.

Finalmente se presentó el modelo alternativo *QC*DM de quintaesencia, cuya complejidad de tratamiento consiste en obtener un modelo que muestre el comportamiento de las soluciones al sistema dinámico y entregando un modelo para $x(z)$ y $y(z)$, como el de la ecuación de estado $\omega(z)$. Las complicaciones radican en resolver la ecuación de Klein-Gordon sin asumir hipótesis varias sobre la naturaleza y pasado del universo, o con métodos numéricos costosos computacionalmente. Al abordar este sistema de ecuaciones por medio de una PINN combinando la expresión matemática con datos observacionales, se obtuvo un modelo capaz de describir ρ, p, ω al mismo tiempo, sin embargo el mayor mérito de la PINN fue poder hacer lo anterior prescindiendo de las condiciones iniciales del sistema y por ende, prescindiendo de los problemas que implica establecerlas.

Capítulo 8

Perspectiva a futuro

A pesar de los resultados positivos obtenidos en este trabajo, aún hay ventanas de oportunidad a considerar para mejorar lo encontrado y explorar nuevas rutas de trabajo si se desea trabajar con Redes Neuronales Físicamente Informadas.

El primer siguiente paso en esta línea de trabajo es el abordar sistemas de ecuaciones diferenciales parciales, o sistemas de ecuaciones ordinarias pero cuya solución sea muy compleja de obtener, o bien se encuentre en más de dos o tres dimensiones.

Otro punto importante dentro del panorama cosmológico es incluir en el entrenamiento de PINN más datos observacionales y de diferentes fuentes, tal como se mostró en el Problema de Q -CDM.

Uno de los primeros puntos a mencionar es que las PINN al tener la arquitectura de un perceptrón multicapa, aún hereda algunas de sus desventajas. Quizá la más importante es que no ofrece información alguna sobre el error estadístico asociado a los datos que está modelando. Por lo anterior han surgido híbridos de PINN con otras redes neuronales que ayuden a generar un modelo más completo, como las *Redes Neuronales Bayesianas Físicamente Informadas*.

En virtud de prescindir cada vez más de parámetros libres o condiciones iniciales fijas, es necesario contar con más recursos computacionales que los utilizados en este trabajo para entrenar modelos más complejos y poderosos. Lo anterior también ayudará a mejorar el desempeño de dichos modelos al poder entrenarlos durante más épocas o al permitir buscar más combinaciones de hiperparámetros que más se acerquen a los valores esperados.

Finalmente en cuanto al código desarrollado, este aún tiene mucho por mejorar y optimizar en cuanto a eficiencia. A la comunidad científica le beneficiaría enormemente contar con una librería construida a partir del código generado durante esta tesis para poder crear sus propios modelos siempre que lo necesiten.

Referencias

- «2006 John C. Mather and George F. Smoot» (s.f.). En: *Nobel Lectures in Physics (2006 – 2010)*, págs. 1-66. DOI: [10.1142/9789814612692_0001](https://doi.org/10.1142/9789814612692_0001). eprint: https://www.worldscientific.com/doi/pdf/10.1142/9789814612692_0001. URL: https://www.worldscientific.com/doi/abs/10.1142/9789814612692_0001.
- Apunevych, S. et al. (2008). *WMAP2006: Cosmological Parameters and Large-scale Structure of the Universe*. arXiv: [0802.0599](https://arxiv.org/abs/0802.0599) [astro-ph].
- Bartelmann, Matthias (nov. de 2010). «Gravitational lensing». En: *Classical and Quantum Gravity* 27.23, pág. 233001. ISSN: 1361-6382. DOI: [10.1088/0264-9381/27/23/233001](https://doi.org/10.1088/0264-9381/27/23/233001). URL: <http://dx.doi.org/10.1088/0264-9381/27/23/233001>.
- Bonometto, Silvio, Vittorio Gorini y Ugo Moschella (ene. de 2002). «Modern Cosmology». En: *Modern Cosmology. Series: Series in High Energy Physics, Cosmology and Gravitation*, ISBN: 978-0-7503-0810-6. Taylor and Francis, Edited by Ugo Moschella, Vittorio Gorini and Silvio Bonometto. DOI: [10.1201/9781420033908](https://doi.org/10.1201/9781420033908).
- Borghini, Nicola, Michele Moresco y Andrea Cimatti (2022). «Toward a Better Understanding of Cosmic Chronometers: A New Measurement of H(z) at z = 0.7». En: *The Astrophysical Journal Letters* 928.1, pág. L4. DOI: [10.3847/2041-8213/ac3fb2](https://doi.org/10.3847/2041-8213/ac3fb2). URL: <https://dx.doi.org/10.3847/2041-8213/ac3fb2>.
- Brout, Dillon et al. (2022). «The Pantheon+ Analysis: Cosmological Constraints». En: *The Astrophysical Journal* 938.2, pág. 110. DOI: [10.3847/1538-4357/ac8e04](https://doi.org/10.3847/1538-4357/ac8e04). URL: <https://dx.doi.org/10.3847/1538-4357/ac8e04>.
- Buhrmester, Vanessa, David Münch y Michael Arens (2019). *Analysis of Explainers of Black Box Deep Neural Networks for Computer Vision: A Survey*. arXiv: [1911.12116](https://arxiv.org/abs/1911.12116) [cs.AI]. URL: <https://arxiv.org/abs/1911.12116>.
- Caldwell, R. R. (jun. de 2000). «An introduction to quintessence.» En: *Brazilian Journal of Physics* 30.2, págs. 215-229.
- Caldwell, R. R. (2000). «An introduction to quintessence.» En: *Brazilian Journal of Physics* 30.2, 215–229. ISSN: 0103-9733. DOI: [10.1590/S0103-97332000000200002](https://doi.org/10.1590/S0103-97332000000200002). URL: <https://doi.org/10.1590/S0103-97332000000200002>.
- Chacón, Jazhiel et al. (jun. de 2023). «Analysis of dark matter halo structure formation in N-body simulations with machine learning». En: *Physical Review D* 107.12. ISSN: 2470-0029. DOI: [10.1103/physrevd.107.123515](https://doi.org/10.1103/physrevd.107.123515). URL: <http://dx.doi.org/10.1103/PhysRevD.107.123515>.

- Chantada, Augusto T. et al. (2023). «Cosmology-informed neural networks to solve the background dynamics of the Universe». En: *Phys. Rev. D* 107 (6), pág. 063523. DOI: [10.1103/PhysRevD.107.063523](https://doi.org/10.1103/PhysRevD.107.063523). URL: <https://link.aps.org/doi/10.1103/PhysRevD.107.063523>.
- Chen, Chang et al. (2020a). *Learning the Evolution of the Universe in N-body Simulations*. arXiv: 2012.05472 [astro-ph.CO].
- Chen, Feiyu et al. (2020b). «NeuroDiffEq: A Python package for solving differential equations with neural networks». En: *Journal of Open Source Software* 5.46, pág. 1931. DOI: [10.21105/joss.01931](https://doi.org/10.21105/joss.01931). URL: <https://doi.org/10.21105/joss.01931>.
- Chevallier, Michel y David Polarski (abr. de 2001). «ACCELERATING UNIVERSES WITH SCALING DARK MATTER». En: *International Journal of Modern Physics D* 10.02, 213–223. ISSN: 1793-6594. DOI: [10.1142/S0218271801000822](https://doi.org/10.1142/S0218271801000822). URL: <http://dx.doi.org/10.1142/S0218271801000822>.
- Conroy, Charlie (2013). «Modeling the Panchromatic Spectral Energy Distributions of Galaxies». En: *Annual Review of Astronomy and Astrophysics* 51. Volume 51, 2013, págs. 393-455. ISSN: 1545-4282. DOI: <https://doi.org/10.1146/annurev-astro-082812-141017>. URL: <https://www.annualreviews.org/content/journals/10.1146/annurev-astro-082812-141017>.
- Copeland, Edmund J., Andrew R. Liddle y David Wands (abr. de 1998). «Exponential potentials and cosmological scaling solutions». En: *Physical Review D* 57.8, 4686–4690. ISSN: 1089-4918. DOI: [10.1103/physrevd.57.4686](https://doi.org/10.1103/physrevd.57.4686). URL: <http://dx.doi.org/10.1103/PhysRevD.57.4686>.
- Dai, Zhenyu et al. (nov. de 2023). «Physics-informed neural networks in the recreation of hydrodynamic simulations from dark matter». En: *Monthly Notices of the Royal Astronomical Society* 527.2, págs. 3381-3394. ISSN: 0035-8711. DOI: [10.1093/mnras/stad3394](https://doi.org/10.1093/mnras/stad3394). eprint: <https://academic.oup.com/mnras/article-pdf/527/2/3381/53668171/stad3394.pdf>. URL: <https://doi.org/10.1093/mnras/stad3394>.
- Dios Rojas Olvera, Juan de, Isidro Gómez-Vargas y Jose Alberto Vázquez (2022). «Observational Cosmology with Artificial Neural Networks». En: *Universe* 8.2. ISSN: 2218-1997. DOI: [10.3390/universe8020120](https://doi.org/10.3390/universe8020120). URL: <https://www.mdpi.com/2218-1997/8/2/120>.
- Einstein, A. (1916). «Die Grundlage der allgemeinen Relativitätstheorie». En: *Annalen der Physik* 354.7, págs. 769-822. DOI: <https://doi.org/10.1002/andp.19163540702>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.19163540702>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19163540702>.

- Fan, Chunling et al. (2008). «Bayesian Neural Networks and Its Application». En: *2008 Fourth International Conference on Natural Computation*. Vol. 3, págs. 446-450. DOI: [10.1109/ICNC.2008.624](https://doi.org/10.1109/ICNC.2008.624).
- Forsythe, George E. (1970). «Pitfalls in Computation, or Why a Math Book isn't Enough». En: *The American Mathematical Monthly* 77.9, págs. 931-956. DOI: [10.1080/00029890.1970.11992636](https://doi.org/10.1080/00029890.1970.11992636). eprint: <https://doi.org/10.1080/00029890.1970.11992636>. URL: <https://doi.org/10.1080/00029890.1970.11992636>.
- Friedman, A. (1922). «On the Curvature of Space». En: *General Relativity and Gravitation* 31 (12), págs. 1991-2000. DOI: [10.1023/A:1026751225741](https://doi.org/10.1023/A:1026751225741).
- Gharat, Sarvesh y Yogesh Dandawate (feb. de 2022). «Galaxy classification: a deep learning approach for classifying Sloan Digital Sky Survey images». En: *Monthly Notices of the Royal Astronomical Society* 511.4, págs. 5120-5124. ISSN: 0035-8711. DOI: [10.1093/mnras/stac457](https://doi.org/10.1093/mnras/stac457). eprint: <https://academic.oup.com/mnras/article-pdf/511/4/5120/42688228/stac457.pdf>. URL: <https://doi.org/10.1093/mnras/stac457>.
- Holdship, J. et al. (sep. de 2021). «Chemulator: Fast, accurate thermochemistry for dynamical models through emulation». En: *Astronomy and Astrophysics* 653, A76. ISSN: 1432-0746. DOI: [10.1051/0004-6361/202140357](https://doi.org/10.1051/0004-6361/202140357). URL: <http://dx.doi.org/10.1051/0004-6361/202140357>.
- Hubble, Edwin (mar. de 1929). «A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae». En: *Proceedings of the National Academy of Science* 15.3, págs. 168-173. DOI: [10.1073/pnas.15.3.168](https://doi.org/10.1073/pnas.15.3.168).
- Jimenez, Raul y Abraham Loeb (jul. de 2002). «Constraining Cosmological Parameters Based on Relative Galaxy Ages». En: 573.1, págs. 37-42. DOI: [10.1086/340549](https://doi.org/10.1086/340549). arXiv: [astro-ph/0106145](https://arxiv.org/abs/astro-ph/0106145) [astro-ph].
- Jimenez-Mesa, Carmen et al. (2023). «A non-parametric statistical inference framework for Deep Learning in current neuroimaging». En: *Information Fusion* 91, págs. 598-611. ISSN: 1566-2535. DOI: <https://doi.org/10.1016/j.inffus.2022.11.007>. URL: <https://www.sciencedirect.com/science/article/pii/S1566253522002172>.
- Jones, Bernard (ago. de 1997). «A Brief History of Cosmology». En: 126, págs. 1-10.
- Jospin, Laurent Valentin et al. (2022). «Hands-On Bayesian Neural Networks—A Tutorial for Deep Learning Users». En: *IEEE Computational Intelligence Magazine* 17.2, págs. 29-48. DOI: [10.1109/MCI.2022.3155327](https://doi.org/10.1109/MCI.2022.3155327).
- Kang, Zihan et al. (2023). «Periodic Variable Star Classification with Deep Learning: Handling Data Imbalance in an Ensemble Augmentation Way». En: *Publications of the Astronomical Society of the Pacific* 135.1051, pág. 094501. DOI: [10.1088/1538-3873/acf15e](https://doi.org/10.1088/1538-3873/acf15e). URL: <https://dx.doi.org/10.1088/1538-3873/acf15e>.

- Kujat, Jens, Robert J. Scherrer y A. A. Sen (2006). «Phantom dark energy models with negative kinetic term». En: *Phys. Rev. D* 74 (8), pág. 083501. DOI: [10.1103/PhysRevD.74.083501](https://doi.org/10.1103/PhysRevD.74.083501). URL: <https://link.aps.org/doi/10.1103/PhysRevD.74.083501>.
- Lagaris, I.E., A. Likas y D.I. Fotiadis (1998). «Artificial neural networks for solving ordinary and partial differential equations». En: *IEEE Transactions on Neural Networks* 9.5, págs. 987-1000. DOI: [10.1109/72.712178](https://doi.org/10.1109/72.712178).
- Lahav, Ofer y Andrew R Liddle (2019). *The Cosmological Parameters* (2019). arXiv: [1912.03687](https://arxiv.org/abs/1912.03687) [astro-ph.CO].
- Lawal, Zaharaddeen Karami et al. (2022). «Physics-Informed Neural Network (PINN) Evolution and Beyond: A Systematic Literature Review and Bibliometric Analysis». En: *Big Data and Cognitive Computing* 6.4. ISSN: 2504-2289. DOI: [10.3390/bdcc6040140](https://doi.org/10.3390/bdcc6040140). URL: <https://www.mdpi.com/2504-2289/6/4/140>.
- Lemaître, Georges (1927). «Un univers homogène de faible densité statistique et l'expansion de l'univers». En: *Annales de la Société Scientifique de Bruxelles* 47, págs. 49-80.
- Li, Xiaolei et al. (dic. de 2019). «Revisiting Metastable Dark Energy and Tensions in the Estimation of Cosmological Parameters». En: *The Astrophysical Journal* 887.2, pág. 153. ISSN: 1538-4357. DOI: [10.3847/1538-4357/ab535d](https://doi.org/10.3847/1538-4357/ab535d). URL: <http://dx.doi.org/10.3847/1538-4357/ab535d>.
- Liddle, Andrew (2015). *An introduction to modern cosmology*. John Wiley y Sons.
- Linder, Eric V. (2003). «Exploring the Expansion History of the Universe». En: *Phys. Rev. Lett.* 90 (9), pág. 091301. DOI: [10.1103/PhysRevLett.90.091301](https://doi.org/10.1103/PhysRevLett.90.091301). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.90.091301>.
- (dic. de 2007). «The dynamics of quintessence, the quintessence of dynamics». En: *General Relativity and Gravitation* 40.2-3, 329-356. ISSN: 1572-9532. DOI: [10.1007/s10714-007-0550-z](https://doi.org/10.1007/s10714-007-0550-z). URL: <http://dx.doi.org/10.1007/s10714-007-0550-z>.
- Liu, Dao-Jun et al. (jul. de 2008). «Revisiting the parametrization of equation of state of dark energy via SNIa data». En: *Monthly Notices of the Royal Astronomical Society* 388.1, págs. 275-281. ISSN: 0035-8711. DOI: [10.1111/j.1365-2966.2008.13380.x](https://doi.org/10.1111/j.1365-2966.2008.13380.x). eprint: <https://academic.oup.com/mnras/article-pdf/388/1/275/18721496/mnras0388-0275.pdf>. URL: <https://doi.org/10.1111/j.1365-2966.2008.13380.x>.
- Lu, Yulong y Jianfeng Lu (2020). «A Universal Approximation Theorem of Deep Neural Networks for Expressing Probability Distributions». En: *Advances in Neural Information Processing Systems*. Ed. por H. Larochelle et al. Vol. 33. Curran Associates, Inc., págs. 3094-3105. URL: https://proceedings.neurips.cc/paper_files/paper/2020/file/2000f6325dfc4fc3201fc45ed01c7a5d-Paper.pdf.

- Martin, John y Hanspeter Schaub (abr. de 2022). «Physics-informed neural networks for gravity field modeling of the Earth and Moon». En: *Celestial Mechanics and Dynamical Astronomy* 134.2, 13, pág. 13. DOI: [10.1007/s10569-022-10069-5](https://doi.org/10.1007/s10569-022-10069-5).
- Mattheakis, Marios et al. (2022). «Hamiltonian neural networks for solving equations of motion». En: *Phys. Rev. E* 105 (6), pág. 065305. DOI: [10.1103/PhysRevE.105.065305](https://doi.org/10.1103/PhysRevE.105.065305). URL: <https://link.aps.org/doi/10.1103/PhysRevE.105.065305>.
- Melia, Fulvio y Robert S. Maier (mayo de 2013). «Cosmic chronometers in the $R_h = ct$ Universe». En: *Monthly Notices of the Royal Astronomical Society* 432.4, págs. 2669-2675. ISSN: 0035-8711. DOI: [10.1093/mnras/stt596](https://doi.org/10.1093/mnras/stt596). eprint: <https://academic.oup.com/mnras/article-pdf/432/4/2669/18611775/stt596.pdf>. URL: <https://doi.org/10.1093/mnras/stt596>.
- Milne, E.A. (1935). *Relativity, Gravitation and World-structure*. International series of monographs on physics. Clarendon Press. ISBN: 9780598424150. URL: <https://books.google.com.mx/books?id=8pULAQAIAAJ>.
- Moresco, Michele (2015). «Raising the bar: new constraints on the Hubble parameter with cosmic chronometers at $z \approx 2$ ». En: *Monthly Notices of the Royal Astronomical Society: Letters* 450.1, págs. L16-L20.
- Moresco, Michele et al. (2012). «New constraints on cosmological parameters and neutrino properties using the expansion rate of the Universe to $z \approx 1.75$ ». En: *Journal of Cosmology and Astroparticle Physics* 2012.07, pág. 053.
- Moresco, Michele et al. (2016). «A 6 % measurement of the Hubble parameter at $z \approx 0.45$: direct evidence of the epoch of cosmic re-acceleration». En: *Journal of Cosmology and Astroparticle Physics* 2016.05, pág. 014.
- Moschou, S P et al. (2023). «Physics-informed neural networks for modeling astrophysical shocks». En: *Machine Learning: Science and Technology* 4.3, pág. 035032. DOI: [10.1088/2632-2153/acf116](https://doi.org/10.1088/2632-2153/acf116). URL: <https://dx.doi.org/10.1088/2632-2153/acf116>.
- Newton, Isaac (1687). *Philosophiæ Naturalis Principia Mathematica*. England.
- Perlmutter, S. et al. (1999). «Measurements of H_0 and Ω_m from 42 High-Redshift Supernovae». En: *The Astrophysical Journal* 517.2, pág. 565. DOI: [10.1086/307221](https://doi.org/10.1086/307221). URL: <https://dx.doi.org/10.1086/307221>.
- Peskin, Michael Edward y Daniel V. Schroeder (1995). *An Introduction to Quantum Field Theory*. Reading, USA: Addison-Wesley (1995) 842 p. Westview Press.
- Planck Collaboration et al. (2016). «Planck 2015 results - XIII. Cosmological parameters». En: *AandA* 594, A13. DOI: [10.1051/0004-6361/201525830](https://doi.org/10.1051/0004-6361/201525830). URL: <https://doi.org/10.1051/0004-6361/201525830>.
- Riess, Adam G. et al. (1998). «Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant». En: *The Astronomical Journal* 116.3, pág. 1009. DOI: [10.1086/300499](https://doi.org/10.1086/300499). URL: <https://dx.doi.org/10.1086/300499>.

- Riess, Adam G. et al. (jul. de 2022). «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». En: *The Astrophysical Journal Letters* 934.1, pág. L7. ISSN: 2041-8213. DOI: [10.3847/2041-8213/ac5c5b](https://doi.org/10.3847/2041-8213/ac5c5b). URL: <http://dx.doi.org/10.3847/2041-8213/ac5c5b>.
- Ruiter, Ashley J. (oct. de 2019). «Type Ia supernova sub-classes and progenitor origin». En: *Proceedings of the International Astronomical Union* 15.S357, 1–15. ISSN: 1743-9221. DOI: [10.1017/s1743921320000587](https://doi.org/10.1017/s1743921320000587). URL: <http://dx.doi.org/10.1017/S1743921320000587>.
- Scolnic, D. M. et al. (mayo de 2018). «The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample». En: *The Astrophysical Journal* 859.2, pág. 101. ISSN: 1538-4357. DOI: [10.3847/1538-4357/aab9bb](https://doi.org/10.3847/1538-4357/aab9bb). URL: <http://dx.doi.org/10.3847/1538-4357/aab9bb>.
- Signore, Monique y Denis Puy (1999). «Primordial nucleosynthesis». En: *New Astronomy Reviews* 43.2, págs. 185-200. ISSN: 1387-6473. DOI: [https://doi.org/10.1016/S1387-6473\(99\)00011-1](https://doi.org/10.1016/S1387-6473(99)00011-1). URL: <https://www.sciencedirect.com/science/article/pii/S1387647399000111>.
- Spurio Mancini, Alessio et al. (ene. de 2022). «<scp>CosmoPower</scp>: emulating cosmological power spectra for accelerated Bayesian inference from next-generation surveys». En: *Monthly Notices of the Royal Astronomical Society* 511.2, 1771–1788. ISSN: 1365-2966. DOI: [10.1093/mnras/stac064](https://doi.org/10.1093/mnras/stac064). URL: <http://dx.doi.org/10.1093/mnras/stac064>.
- Stern, Daniel et al. (2010). «Cosmic chronometers: constraining the equation of state of dark energy. I: H(z) measurements». En: *Journal of Cosmology and Astroparticle Physics* 2010.02, pág. 008.
- Tripp, Robert (mar. de 1998). «A two-parameter luminosity correction for Type IA supernovae». En: 331, págs. 815-820.
- Vázquez, J. Alberto et al. (2021). «Bayesian model selection on scalar ϵ -field dark energy». En: *Phys. Rev. D* 103 (4), pág. 043506. DOI: [10.1103/PhysRevD.103.043506](https://doi.org/10.1103/PhysRevD.103.043506). URL: <https://link.aps.org/doi/10.1103/PhysRevD.103.043506>.
- Vuik, Kees et al. (2023). *Numerical Methods for Ordinary Differential Equations*. TU Delft. ISBN: 9789463666657. DOI: <https://doi.org/10.5074/t.2023.001>. URL: <https://doi.org/10.5074/t.2023.001>.
- Zhang, Cong et al. (2014). «Four new observational H(z) data from luminous red galaxies in the Sloan Digital Sky Survey data release seven». En: *Research in Astronomy and Astrophysics* 14.10, pág. 1221.