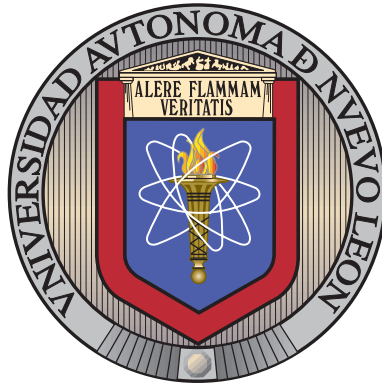


UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

POSGRADO EN INGENIERÍA FÍSICA INDUSTRIAL



PREDICCIONES POR REDES NEURONALES DE LA
CONTRIBUCIÓN DE UN MODELO SUPERSIMÉTRICO A LA
ANOMALÍA DEL MUON

POR

JORDÁN DANIEL SANTILLÁN MORALES

COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE
MAESTRÍA EN INGENIERÍA FÍSICA INDUSTRIAL
CON ORIENTACIÓN A FÍSICA DE PARTÍCULAS Y CAMPOS

SAN NICOLÁS DE LOS GARZA, NUEVO LEÓN

ABRIL 2026

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

POSGRADO EN INGENIERÍA FÍSICA INDUSTRIAL



PREDICCIONES POR REDES NEURONALES DE LA
CONTRIBUCIÓN DE UN MODELO SUPERSIMÉTRICO A LA
ANOMALÍA DEL MUON

POR

JORDÁN DANIEL SANTILLÁN MORALES

COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE
MAESTRÍA EN INGENIERÍA FÍSICA INDUSTRIAL
CON ORIENTACIÓN A FÍSICA DE PARTÍCULAS Y CAMPOS

SAN NICOLÁS DE LOS GARZA, NUEVO LEÓN

ABRIL 2026



UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS
POSGRADO EN INGENIERÍA FÍSICA INDUSTRIAL

Los miembros del Comité de Tesis recomendamos que la Tesis *Predicciones por redes neuronales de la contribución de un modelo supersimétrico a la anomalía del muon*, realizada por el alumno Jordán Daniel Santillán Morales, con número de matrícula 1740631, sea aceptada para su defensa como requisito parcial para obtener el grado de Maestría en Ingeniería Física Industrial con orientación a Física de Partículas y Campos.

El Comité de Tesis

Dr. Francisco Vicente Flores Baez
Asesor

Dra. Melina Gómez Bock
Coasesor

Dr. José Rubén Morones Ibarra
Revisor

Vo. Bo.

Dr. Álvaro Eduardo Cordero Franco
Subdirector de Estudios de Posgrado

San Nicolás de los Garza, Nuevo León, Abril 2026

Dedicatoria

*A mis familia, mi pareja, mis seres queridos por siempre apoyarme.
A quienes me inspiraron a seguir adelante; A Graciela, Diego y Elena.*

Agradecimientos

Agradezco a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SE-CIHTI) por el apoyo económico brindado estos dos años de maestría que fueron altamente fundamentales para poder desarrollar mis habilidades, capacidades e investigación de forma integral. Al Centro de Investigación en Ciencias Físico Matemáticas (CICFIM) por recibirme en su institución y ayudarme a desempeñar mi desarrollo en la investigación. A los profesores por su gran apoyo y disposición a ayudar; el Dr. Baez, la Dra. Melina que bajo su asesoría y coasesoría tuve una gran ayuda como físicos de partículas que son así como personas, un gran agradecimiento al Dr. Almaguer por las múltiples consultas que tuvimos, charlas y aceptarme de oyente en sus clases para reforzar mi conocimiento y apoyarme con mis numerosas dudas sobre la parte de optimización y estadística, a todos los profesores que tuve el gusto de conocer por todo lo que contribuyeron para mi formación, sin duda sin ustedes mi desempeño, motivación y aprendizaje no habrían sido los mismos, les agradezco todas las atenciones brindadas. A mis compañeros y todas las personas que fueron parte de mi proceso en el posgrado, por su apoyo, charlas constructivas, académicas que tuvimos por los pasillos del CICFIM, una gran retroalimentación para mí, siempre tratando de ayudar, enriquecer ideas o cuestionarlas, así como motivar el trabajo para finalmente ayudarnos los unos a los otros, pues la investigación consiste en ayudarse, trabajar en equipo y argumentar o cuestionar las ideas. Por último pero no menos importante agradezco a mi madre, mi hermano y mi pareja por siempre estar apoyando y alentando mis estudios, siempre motivándome a hacer lo que más me gusta.

Prólogo

El ser humano, en su avidez por comprender todo lo ignoto a lo que está sujeto en su entorno, ha sometido su razonamiento sobre la naturaleza a un devenir de ideas y conjeturas, constantemente estas van con una avidez acendrada en su afán por contraponer la misma existencia en su plano ideal, con algunas teorías interesantes, dignas de ser pasadas al conocimiento histórico de la humanidad. Después de todo, es menester comprender lo que nos rodea porque somos la naturaleza en su propia manifestación, un ciclo de cómo aprende de ella misma, parte de una misma sustancia que es extrapolada por todo lo que existe, lo que existió y lo que existirá en sus diferentes representaciones. A pesar de esta desdicha o fortuna, los seres humanos somos finitos en nuestra consciencia y razonamiento, esto ha conducido a pensamientos con oquedad, una de ellas es suponer que nosotros (la naturaleza) debe vislumbrar por una explicación basada en nuestros principios lógicos y matemáticos, buscamos la belleza en nuestras teorías pero no entendemos la belleza, tampoco comprendemos la estética de la misma, un meandro de ideas que debemos arribar. En su defecto hemos creado sistemas alternativos a nuestro desconocimiento de todas estas conjeturas. El falsacionismo es el método por el cuál los desarrollos de la ciencia moderna ha fungido como un papel crítico para nuestro avance en este camino, al ser el sendero más seguro para nosotros (por el momento y sujeto a cambios) tomaremos el mismo para llegar a este estudio.

Resumen

En este trabajo buscamos analizar la bien conocida anomalía del momento magnético del muon desde una perspectiva diferente. La discrepancia entre teoría y experimento se encuentra actualmente en 5.15σ . Para lograr esto, empleamos el Modelo Estándar Mínimo Supersimétrico (MSSM) con una extensión de Sabor que incluye una estructura familiar jerárquica en los términos escalares trilineales suaves supersimétricos del Lagrangiano, definido en el rompimiento de escala de Supersimetría (SUSY). Este marco de trabajo introduce 5 parámetros libres dentro de rangos específicos. Proponemos reducir la discrepancia entre teoría y experimento a el rango de $[3.15\sigma, 4.15\sigma]$. Usando Redes Neuronales (NN), predecimos la contribución a_μ^{SUSY} y examinamos la influencia de cada parámetro. Realizamos esta optimización al generar distribuciones aleatorias para los parámetros de SUSY y seleccionar aquellos que nos conducen el rango de desviación deseado, i.e., $a_\mu^{SUSY} \in [3.15\sigma, 4.15\sigma]$. Este conjunto de parámetros seleccionado es usado como entrada para nuestra red neuronal, mientras que los valores correspondientes de a_μ^{SUSY} , explícitamente computados del modelo de SUSY, son usados como salidas objetivo para el entrenamiento. El análisis final del modelo entrenado nos revela como los pesos de la red neuronal fueron asignados a cada uno de los 5 parámetros. Esta examinación descubre una fuerte relación entre ciertas constantes libres y sus correspondientes distribuciones. El objetivo que busca este trabajo es proporcionar un acercamiento alternativo para obtener predicciones fenomenológicas, tales como observables y contribuciones teóricas, evitando la carga computacional de los métodos analíticos computacionales tradicionales, sirviendo como una rápida alternativa.

ÍNDICE GENERAL

Dedicatoria	I
Agradecimientos	II
Prólogo	III
Resumen	IV
Lista de Figuras	VIII
Lista de Tablas	XII
Introducción	1
I Fundamentos Físicos	6
1. Teoría Cuántica de Campos	7
1.1. Notación Relativista	8
1.2. La Ecuación de Klein-Gordon	9
1.3. La Ecuación de Dirac	11
1.4. El Oscilador Armónico	16
1.5. Densidad Lagrangiana Clásica de Campos	18
1.6. Cuantización de la Teoría de Campos Lagrangiana	21
1.7. Simetrías y Leyes de Conservación	22
1.8. El Campo Real de Klein-Gordon	27
1.9. Diagramas de Feynman	30

1.10. Integrales de Camino	35
1.11. Electrodinámica Cuántica	39
1.11.1. Función del vértice del electrón	41
2. Teoría de Supersimetría	46
2.1. El Modelo Estándar	47
2.2. Supersimetría - Un Corto Resumen	49
2.3. Super Álgebra de Poincaré	50
2.4. Superespacios	51
2.5. Supercampos	52
2.6. Supercampos Escalares y Vectoriales	52
2.7. Lagrangiano Supersimétrico	54
2.8. Intensidad del Campo	55
2.9. El Modelo Estándar Mínimo Supersimétrico	57
2.10. MEMS Rompimiento Suave de Supersimetría	58
2.11. Paridad R	59
2.12. Ruptura de Simetría Electrodébil Radiativa	60
2.13. Electroweakinos	61
2.14. Anomalía del muon en términos suaves de supersimetría	63
2.14.1. Estructura de sabor en el rompimiento suave de supersimetría	64
2.14.2. Matriz de masa para los sfermiones	64
2.14.3. Términos Trilineales Suaves	65
2.14.4. Interacción Neutralino-lepton-slepton	67
2.14.5. La contribución del loop a la anomalía del muon	71

II	Machine Learning	73
3.	Conceptos básicos de Machine Learning	74
3.1.	Algunas nociones de Machine Learning	75
3.2.	Entrenamiento de la red	79
3.2.1.	Optimización de parámetros	80
3.3.	Retropropagación de Error	84
3.4.	Métricas de Error	88
3.5.	Funciones de Activación	90
3.6.	Conjuntos de Entrenamiento, Prueba y Validación(Train, Test and Validation Sets)	91
3.6.1.	Conjuntos de Prueba y Entrenamiento	91
3.6.2.	Hiperparámetros	92
3.7.	Sobreajuste, Subajuste y Regularización	93
3.7.1.	Detención Temprana o Early Stopping	94
3.7.2.	Dropout o regularización por apagado	95
3.8.	Inicialización de parámetros (Glorot Uniform)	97
3.9.	Escaladores	99
3.10.	Aprendizaje por Transferencia o Transfer Learning	100
3.11.	Distribuciones de Probabilidad Truncadas en el modelo	102
III	Experimento y Resultados	107
4.	Experimento	108
4.1.	Generación de Datos	108
4.2.	Selección de distribución	111
4.3.	Escalamiento	112

4.4. División de los datos	112
5. Resultados	114
5.1. Resultados de validación	114
5.2. Resultados de Prueba	122
5.3. Tiempo de computo	124
Conclusiones	126
Referencias	128

ÍNDICE DE FIGURAS

1.	En este diagrama (a) representa una partícula con una línea con una flecha en dirección del tiempo y una antipartícula con la flecha en dirección opuesta. (b) representa la creación de pares en el tiempo $t = t_0$. (c) Es una aniquilación de pares al tiempo $t = t_1$. (d) La trayectoria de una partícula interpretada como: (e) Una creación de pares en t_2 , la antipartícula se aniquila con una tercer partícula en $t = t_3$. Adaptado de [1, p. 175].	31
2.	Algunos diagramas espacio-tiempo de las interacciones más comunes. Adaptado de [1, p. 176].	33
3.	La anatomía de un diagrama de 3er orden desconectado. Adaptado de [1, p. 181].	33
4.	Reglas de Feynman para el espacio de momentos en la teoría ϕ^4 . Adaptado de [1, p. 185].	34
5.	Corte tiempo de una trayectoria de (t_a, q_a) a (t_b, q_b) . Adaptado de [1, p. 211].	37
6.	Dispersión de un electrón en un potencial externo (\times) al más bajo orden. Adaptado de [2, p. 78].	40
7.	Dispersión de electrones con un fotón virtual (loop) a primer orden.	41
8.	Diagrama de dispersión de electrones con un fotón virtual (loop) con momento $p - k$, fermiones entrantes p, k, p', k' , $k' = k + q$ y fotón saliente q	41
9.	Correcciones a 1-loop para la masa cuadrada del Higgs m_h^2 , debido a un fermión de dirac (a) f , y (b) un escalar S (Imagen obtenida de [3]).	48
10.	Contribución del slepton a a_μ . Diagrama obtenido de [4].	70

11.	Diagrama de la NN de dos capas correspondiente a (3.7). Las variables de entrada, escondidas, y de salida son representadas por nodos, los parámetros de peso se representan por ligas entre los nodos, en los cuales los parámetros sesgo son denotados por ligas llegando de entradas adicionales y variables escondidas x_0 y z_0 . Las flechas muestran la dirección del flujo de información en la red durante la propagación hacia adelante. Adaptado de [5, p. 228].	77
12.	Vista geométrica de la función de error $E(\mathbf{w})$ como una superficie sobre el espacio de pesos. El punto \mathbf{w}_A representa un mínimo local, mientras que \mathbf{w}_B es el mínimo global. En cualquier punto \mathbf{w}_C , el gradiente local de la superficie de error está dado por el vector ∇E . Imagen adaptada de [5]. . .	81
13.	Ilustración del cálculo de δ_j para una unidad oculta j mediante backpropagation desde las unidades k a las que j envía conexiones. Tomado de [5].	86
14.	Ejemplo de una red neuronal con una topología general de tipo feed-forward. Tomado de [5].	87
15.	Ejemplo de underfitting, capacidad apropiada y overfitting, adaptado de [6].	94
16.	Curvas de aprendizaje que muestran cómo cambia la función de pérdida de log-verosimilitud negativa a lo largo de las épocas de entrenamiento. Adaptado de [6].	95
17.	Dropout entrena un conjunto de subredes formadas al eliminar unidades que no pertenecen a la capa de salida de una red base subyacente. En este ejemplo, se muestran todas las dieciséis subredes que se pueden formar eliminando diferentes subconjuntos de unidades de la red original. Este método es efectivo para entrenar y evaluar una red neuronal más pequeña y eficiente. Adaptado de [6].	96

18. Ejemplo de propagación hacia adelante a través de una red de retroalimentación utilizando dropout. (Arriba) En este ejemplo, usamos una red de retroalimentación con dos unidades de entrada, una capa oculta con dos unidades ocultas y una unidad de salida. (Abajo) Para realizar la propagación hacia adelante con dropout, muestreamos aleatoriamente un vector μ con una entrada para cada unidad de entrada u oculta en la red. Las entradas de μ son binarias y se muestrean independientemente entre sí. La probabilidad de que cada entrada sea 1 es un hiperparámetro, generalmente 0.5 para las capas ocultas y 0.8 para la entrada. Cada unidad en la red es multiplicada por la máscara correspondiente y luego la propagación hacia adelante continúa a través del resto de la red. Esto es equivalente a seleccionar aleatoriamente una de las subredes como se muestra en la fig. 17 y ejecutar la propagación hacia adelante a través de ella. Adaptado de [6]. 98

19. Ilustración de un proceso de aprendizaje por transferencia. Adaptado de [7]. 101

20. Conjunto de las 4 distribuciones truncadas usadas para el modelo analítico. 109

21. Valores de M_1 y \tilde{m}_s que resuelven la anomalía de a_μ reduciendo a 3.15σ y 4.15σ , lo cual conlleva a filtrar los parámetros que cumplan tal condición, en el código con la distribución deseada. 110

22. Conjunto de los parámetros $\beta, A_0, M_1/\tilde{m}_s, \tilde{m}_s$ que cumplen la condición con respecto a μ_{SUSY} 111

23. Primera Red neuronal: valores obtenidos vs predicciones del modelo. 115

24. Evaluación de la primera NN utilizando el $\eta = 1 \times 10^{-3}$. con ReLu con valores hasta $\sim \times 10^4$ mostrando la validación de la pérdida con el train set y el validation set. 117

26. Mejor evaluación de la NN con η en escala log 117

25. Evaluación de la NN con diferentes η 118

27. Valores obtenidos con diferentes funciones de activación: ReLu, Swish, ELU, SoftPlus se llegan a diferentes validaciones de aprendizaje, en el cuál ReLu actúa como el más adecuado para usarse en nuestro modelo al tener el error más bajo. 119

28. Segunda Red neuronal: valores obtenidos vs predicciones del modelo. 120

29.	Evaluación de la segunda NN utilizando el $\eta = 1 \times 10^{-3}$. con ReLu con valores hasta $\sim \times 10^4$ mostrando la validación de la pérdida con el train set y el validation set.	121
30.	Gráfico de violín de la distribución de los pesos en cada parámetro del MSSM utilizado, de lado izquierdo vemos la primera NN con sus parámetros siendo ajustados para proseguir a crear las demás capas. En el lado derecho se observan los parámetros de la 2da NN.	122
31.	Modelo final expuesto a los datos de prueba, donde $y = x$ nos indica la forma en que la dispersión real y las predicciones coinciden, la barra de color nos muestra que mayormente los puntos de predicción tienen un error muy bajo y están cerca de la dispersión real.	123
32.	En el panel izquierdo tenemos como se observa el error absoluto con una línea vertical que nos indica donde se encuentra el promedio, en el panel derecho vemos la distribución de valores de nuestras predicciones junto con los valores reales.	124

ÍNDICE DE TABLAS

1.	Componentes del supercampo.[8]	52
2.	Acoplamientos escalares y pseudoscalares del bino–leptón–slepton con una mezcla de violación de sabor leptónico. Adaptada de [4].	69
3.	El espacio de parámetros donde se realizará el modelo para la red neuronal más adelante. Adaptado de [4]	71
4.	Funciones de activación que serán usadas en este trabajo: ReLU, ELU, Swish y SoftPlus con sus respectivas derivadas.	90
5.	Resultados de filtrado por distribución, coeficiente R^2 y tiempo de entrenamiento de la NN, el formato de tiempo es en horas:minutos:segundos. . .	112
6.	Fracción del data set total $\mathcal{D} = (X, y)$ usados para entrenamiento, validación y testing.	113
7.	Tiempos de cómputo promedio para cada etapa. El tiempo de código SUSY corresponde a la generación de datos de entrada. El entrenamiento se realizó con dos arquitecturas iguales para cada una [5, 128, 128, 128, 1] (primera y segunda NN). Dado que se generan valores aleatorios para la entrada entonces trabajamos $\sim 10^6$ que nos da el modelo SUSY, después las muestras se reducen a $\sim 2 \times 10^4$. El tiempo de prueba es el necesario para evaluar el modelo final sobre los datos. El formato de los tiempos es horas:minutos:segundos.	124

INTRODUCCIÓN

Para comenzar a entender el contexto, iniciamos con uno de nuestros temas centrales: el Modelo Estándar (SM) de partículas elementales. Este modelo es un formalismo matemático que describe a un conjunto de partículas consideradas como elementales y sus interacciones. En específico, se incluyen 3 de las 4 fuerzas fundamentales; nuclear débil, nuclear fuerte y electromagnética, se describe cada interacción mediante el intercambio de partículas, llamadas bosones de Gauge. Por otra parte, se incorpora el mecanismo de Higgs, para dotar de masa a cada partícula del modelo y diferente valor de acuerdo a la evidencia experimental. Derivado del diseño experimental para confirmar las predicciones de este modelo, se ha generado de manera secundaria, tecnología que hoy en día beneficia a la humanidad. Desde su construcción, el SM ha sido considerado uno de los mejores modelos teóricos que ha hecho el ser humano para describir la naturaleza a escalas pequeñas, mucho menores al tamaño del núcleo que por mucho tiempo atrás se idealizó como indivisible. Una de las partículas fundamentales del SM que nos interesa, el muon, fue descubierto en 1937 través de un experimento en el que se estudiaba la lluvia de rayos cósmicos, usando como detector la cámara de niebla y campos magnéticos. La evidencia en la cámara consistió de trazas, líneas visibles en el medio que se ionizaba con facilidad, y se interpretó como una señal de una partícula cargada con una masa mucho mayor que la del electrón pero menor que la del protón, dándole la categoría de “electrón pesado”. La clave para identificarlo fue que, a energías entre $300 - 400 MeV$, los muones presentaban una pérdida de energía diferente a la de otras partículas conocidas (electrón y protón) este comportamiento tan singular, observado inicialmente en estos rayos cósmico, fue lo que permitió reconocerlos como una nueva clase de partícula. Una vez que se identificaron sus características como masa, spin, carga eléctrica, se concluyó que era un nuevo tipo de partícula y se le asignó un nombre: "muon" [9].

Posteriormente, se midió lo que se conoce como anomalía del momento magnético del muon. Toda partícula con spin tiene asignado un momento magnético de spin, del mismo carácter que el momento magnético orbital asociado a una partícula cargada que sigue una trayectoria circular uniforme. Este concepto, el momento magnético, nos permite describir la interacción de este tipo de partículas con un campo magnético externo. El

momento magnético es proporcional al momento angular, y el factor de proporcionalidad se llama razón giromagnética. El factor g cuantifica esa proporcionalidad. El límite no relativista de la ecuación de Dirac para una partícula en un campo magnético externo, conocido como la ecuación de Pauli, predice el valor $g = 2$. Así, la anomalía se refiere a la desviación del valor canónico, es decir, $a_\mu = (g - 2)/2$. Hoy en día, la medición de esta anomalía es fundamental en la física de partículas, debido a que el SM predice la anomalía con un nivel de precisión similar al nivel de precisión del experimento, lo cual brinda información crucial sobre las interacciones entre el muon y las partículas virtuales producto de las fluctuaciones del vacío. La medición precisa de $g - 2$ del muon ha sido objeto de una intensa investigación experimental y teórica durante varias décadas, ya que cualquier discrepancia entre el valor predicho y el observado podría indicar la presencia de nueva física más allá del Modelo Estándar. Su importancia deriva de que es una de las pocas cantidades que se pueden medir con mucha precisión y, al mismo tiempo, son calculables desde los primeros principios. La sensibilidad relativa de la anomalía del muon a distancias cortas lo convierte en una herramienta útil para estudiar física que involucra partículas de gran masa. Esto está en fuerte contraste con el momento magnético anómalo del electrón, a_e , el cuál es bastante insensible a interacciones fuertes y débiles.

La primera persona en calcular la anomalía, para el electrón, fue Julian Schwinger, en 1948, dando como resultado una desviación del orden de $a_e^{QED} = \alpha/(2\pi)$, usando la recién formulada electrodinámica cuántica (QED), lo cuál fue considerado un éxito de la teoría cuántica de campos. De acuerdo a QED, las fluctuaciones del vacío genera fotones virtuales, es decir, fotones que se crean y que se absorben en tiempos muy cortos, impidiendo que puedan ser detectados en el experimento y que corresponden, dentro del formalismo de la teoría de campos cuánticos, a una corrección al siguiente orden en perturbaciones. La primera medición de $g - 2$ del muon fue realizada en 1957, en el ciclotrón de Nevis (Columbia university) por Garwin, Lederman y Weinrich. En 1961, el CERN publicó su primer resultado importante sobre el muon [10], que comparado con las predicción esto es; $a_\mu^{th} = 0.001165$, se escribe como $a_\mu^{exp} = a_\mu^{th}(0.983 \pm 0.019)$. Se observa una incertidumbre relativa de $0.019 \sim 2\%$. Es notable que el valor teórico era más grande que la medición, tomando en cuenta que solo se calculó la contribución de QED. Esto significa que el valor central experimental está 1.7% por debajo del valor teórico y que la incertidumbre relativa es $1.9\%(1\sigma)$. La significancia estadística está definida como $Z = \Delta/\sigma_\Delta = 1.7/1.9 \sim 0.9\sigma$ y se usa para ver qué tan improbable es que el dato sea puro azar. En este caso se obtuvo $\approx 1\sigma$ que significa una compatibilidad de la medición experimental con la teoría QED. Estos hallazgos reforzaron la idea de que el muon parecía ser un electrón pesado sin más interacciones que las electromagnéticas y las débiles. Este concepto no permitía dar explicación para la diferencia de masa muon-electrón, pero a su vez permitía calcular el momento magnético del muon a partir de la ecuación de Dirac y la Electrodinámica

Cuántica (QED).

El momento magnético anómalo del muon, a_μ , es una observable de alta precisión y una prueba sensible del Modelo Estándar. Sus contribuciones provienen de tres sectores: contribuciones hadrónicas (Cromodinámica Cuántica (QCD)) que es una teoría no abeliana $SU(3)_C$ cuyas partículas mediadoras de la interacción son los 8 gluones, las correcciones Electrodébiles (EW) que unifica a altas energías a los campos con simetrías $SU(2)_L \times U(1)_Y$ y adicionalmente el mecanismo de Higgs se genera un rompimiento espontáneo de simetría dando lugar a $U(1)_{EM}$, a los bosones W^\pm y Z^0 masivos y al fotón γ sin masa, por último la Electrodinámica Cuántica (QED) corresponde al campo electromagnético y precisamente al $U(1)_{EM}$ generado por el rompimiento de simetría del Higgs, con el fotón como la partícula mediadora del campo. La incertidumbre teórica está dominada principalmente por los términos hadrónicos, en particular la polarización del vacío (HVP) y la dispersión luz-por-luz hadrónica (HLbL). Históricamente, las primeras evaluaciones modernas combinaron datos para HVP y modelos para HLbL [11]. La precisión experimental ha sido mejorada significativamente desde comienzos de este siglo; por ejemplo, $a_{\mu^+} = 11659191(59) \times 10^{-10} (\pm 5 \text{ppm})$ en 2000 [12]. En paralelo, han realizado avances en los cálculos teóricos: correcciones de HVP en lattice con electromagnetismo e isospin explícitos, $\delta a_\mu^{HVP}(udsc) = 7.1(2.9) \times 10^{-10}$ [13], y reevaluaciones de HLbL con resonancias escalares efectivas $\sigma/f_0(500)$ que concentran gran parte de la incertidumbre [14].

Las contribuciones del SM a la anomalía del muon pueden separarse como sigue:

$$a_\mu = a_\mu^{QED} + a_\mu^{QCD} + a_\mu^{EW}$$

donde a_μ^{QED} es puramente electromagnético, a_μ^{QCD} es la parte de QCD y a_μ^{EW} corresponde a la interacción débil.

Al compararse la medición experimental con la predicción teórica, se obtiene

$$\Delta a_\mu = a_\mu^{exp} - a_\mu^{SM} = 249(22)(43) \times 10^{-11}$$

lo cual representa una discrepancia entre teoría y experimento de $\approx 5\sigma$, lo cual es alarmante en la física de partículas.

Para resolver esta discrepancia, se han propuesto diversos modelos, entre ellos uno del 2018 [15], en el cual se introducen un modelo auto interactuante de materia oscura que atacaría simultáneamente la explicación del $g - 2$.

Y la lista de propuestas continúa, un extenso trabajo que se ha desarrollado desde el 2018 sobre este tema, puede revisarse en: [16] Actualmente tenemos una de las recientes contribuciones experimentales a esta anomalía [17], desarrollada en el Fermilab National Accelerator Laboratory (FNAL). La anomalía a_μ se mide comparando dos frecuencias

en un anillo de almacenamiento magnético, muones polarizados donde su spin precesa a frecuencia ω_s y orbitan a ω_c . La frecuencia anómala se observa como una oscilación en el número de positrones de alta energía tras los decaimientos $\mu^+ \rightarrow e^+\nu\nu$: aquí el spin se alinea periódicamente con el momento, favoreciendo positrones más energéticos en ciertas fases. El campo magnético \mathbf{B} se calibra mediante la frecuencia de Larmor de protones ω_p medida con sondas de Resonancia Magnética Nuclear(NMR). Usando $\omega_a = \omega_s - \omega_c$, ω_p y el cociente magnético

$$\lambda = \frac{\mu_\mu}{\mu_p},$$

se deduce

$$a_\mu = \frac{R}{(\lambda - R)}, \quad R = \frac{\omega_a}{\omega_p}$$

Para suprimir los efectos del campo eléctrico se opera al impulso que anula el término eléctrico(conocido como *magic momentum*, $\gamma \approx 29.3$, $p \approx 3.094\text{GeV}/c$).

La medición más precisa hasta el momento fue reportada en febrero 23 del 2024 [18] que fue de $a_\mu^{exp} = 116592059(22) \times 10^{-11}(0.19\text{ppm})$, mediante los resultados obtenidos se ha llegado a que la discrepancia ha aumentado a una cifra de $(a_\mu^{exp} - a_\mu^{wp}) = 5.15\sigma \times 10^{-11}$. Esto da lugar a pensar en nueva física, partículas e interacciones que aún desconocemos y reconsiderar el Modelo Estándar(ME).

Sin embargo recientemente en Junio de 2025 [19] se ha presentado un nuevo resultado teórico y una actualización al white paper [20] sobre este problema, pero ahora agregan un progreso significativo en la contribución hadrónica luz-por-luz usando un método dispersivo basado en datos así como cálculos en lattice-QCD, este acercamiento reduce enormemente la discrepancia, dando por resultado, $a_\mu^{exp} - a_\mu^{SM} = 38(63) \times 10^{-11}$, lo cual supone que ya no hay tensión entre el Modelo Estándar y el experimento.

No obstante nosotros seguiremos este trabajo a pesar de este resultado, donde asumiremos que persiste una discrepancia entre el Modelo Estándar y la medición en el rango de $[3.15\sigma, 4.15\sigma]$. Por lo tanto, tomamos como objeto de estudio la discrepancia del ME con los resultados experimentales y se aborda la solución desde una teoría más allá del ME; la teoría de supersimetría que es un candidato prometedor para resolución de este y otros problemas que actualmente el modelo estándar no puede resolver. Esta teoría, que es una extensión del SM, introduce una relación simétrica entre los bosones y los fermiones. Así, cada partícula del ME tendría un compañero supersimétrico llamado “scompañero”, con un espín que difiere en una unidad semi-entera. Adicionalmente, esta propuesta incorpora otros operadores, otras contribuciones por nuevas partículas y por lo tanto, nuevas reglas de Feynman. En concreto vamos a utilizar el Modelo Estándar Mínimo Supersimétrico (MSSM) reportado en [4] para construir un modelo de regresión por medio de redes neuronales (NN) o redes neuronales artificiales las cuales son altamente utilizadas hoy en día

por su gran capacidad para aproximar funciones del tipo $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, estas redes neuronales son modelos paramétricos no lineales que aprenden a partir de datos. El modelo básico es el perceptrón, que aplica una transformación afín seguida de una activación no lineal; la composición de múltiples capas da lugar al perceptrón multicapa (MLP). Bajo condiciones suaves, los MLP actúan como aproximadores universales, lo que permite capturar relaciones complejas en tareas de regresión y clasificación. Esto las convierte en un candidato poderoso para aplicar a nuestro problema, los datos para optimizar serán 5 parámetros libres del MSSM como entrada a nuestra red, una función de pérdida (que se optimizará) y los valores de la predicción a_μ^{SUSY} como salidas (o valores objetivos) con el fin de desarrollar una forma alternativa para calcular y obtener esta anomalía. En pocas palabras buscamos predecir los valores de a_μ^{SUSY} a los que contribuye el MSSM solamente con el espacio de parámetros. Para poder cumplir con este propósito hay que abordar ciertos temas, que en esta tesis se organizarán como sigue; en la **parte I** vamos a revisar brevemente, en el capítulo 1, los fundamentos físicos de la teoría cuántica de campos (QFT) necesaria para establecer las bases del SM y cómo se calcula el $g - 2$. Siguiendo la línea de QFT en el capítulo 2 se darán unas nociones básicas de Supersimetría (SUSY) para nuevamente llegar a calcular la anomalía del muon ya que este resultado será utilizado en nuestra red neuronal. Para la **parte II** comprendido únicamente por el capítulo 3, revisaremos las bases fundamentales del Machine Learning (ML) hasta llegar a establecer qué es una red neuronal (NN) de regresión para poder construir nuestro modelo junto con algunas técnicas de ajuste fino. En la **parte III**, vamos a desglosar en el capítulo 4 nuestro experimento que generará los datos, escalamientos y detalles para preparar la red y construirla. En el capítulo 5 tendremos la sección de resultados y evaluaremos nuestro modelo. Finalmente reportaremos nuestras conclusiones sobre el estudio realizado.

Parte I

Fundamentos Físicos

CAPÍTULO 1

TEORÍA CUÁNTICA DE CAMPOS

En este capítulo vamos a dar una revisión de las bases y conocimientos necesarios para entender el modelo estándar (ME) de la física. Vamos a empezar hablando de la ecuación de Klein-Gordon y la ecuación de Dirac partiendo de partículas libres, estableceremos el tensor de 4-corriente y mencionaremos de las diferencias entre las diferentes densidades de probabilidad de cada una, todo esto más adelante nos ayudará a describir las interacciones y partículas mediadoras de los bosones y fermiones en el ME. Las convenciones y desarrollo que seguiremos en este capítulo son las establecidas en [21].

Daremos una breve revisión del Oscilador Armónico construyendo el Hamiltoniano, sus eigenfunciones y sus eigenvalores de energía para ilustrar la idea de los operadores de creación y aniquilación. Después estableceremos las bases para la notación relativista que se va a tomar en este trabajo.

Daremos las bases de cómo se construye una teoría de campos clásica y cuántica ya que aunque hayamos visto las ecuaciones de K-G y Dirac, este acercamiento no es suficiente, en la Mecánica Cuántica tradicional estamos limitados por un número fijo de partículas, pues para esta primera cuantización utilizamos solo los operadores de posición \hat{x}_i y momento \hat{p}_i . Surgen problemas como el principio de causalidad, energías negativas, vacío dinámico, entre otros. Estas dificultades se resuelven con la construcción de la Teoría Cuántica de Campos, en donde los campos clásicos son promovidos a operadores en el espacio-tiempo, con relaciones de conmutación y anticonmutación, esto nos garantiza que sigamos ya sea la estadística de Bose-Einstein o Fermi-Dirac, que nos permite describir y acoplar interacciones entre campos con una densidad Lagrangiana. Mediante esto es posible definir unas cantidades que se le conoce como propagadores para las diferentes partículas. También daremos una revisión de las simetrías y leyes de conservación provenientes de grupos de simetría. Esta información es fundamental debido a que nos permitirá calcular las contribuciones del $g - 2$ con QED, tema que se abordará al final de este capítulo.

1.1 NOTACIÓN RELATIVISTA

Antes de continuar esta revisión es conveniente introducir una notación compacta y relativista basada en cuatro vectores. Empezamos escribiendo x^μ ($\mu = 0, 1, 2, 3$) para el 4-vector de espacio-tiempo con la componente del tiempo $x^0 = ct$ y las coordenadas espaciales x^j ($j = 1, 2, 3$), $x^\mu = (ct, \mathbf{x})$. Usaremos el tensor métrico $g_{\mu\nu}$, con sus respectivas componentes

$$g_{00} = -g_{11} = -g_{22} = -g_{33} = \pm 1, \quad g_{\mu\nu} = 0 \quad \text{si} \quad \mu \neq \nu, \quad (1.1)$$

establecemos el vector covariante x_μ y el contravariante x^μ :

$$x_\mu = \sum_{\nu=0}^3 g_{\mu\nu} x^\nu = g_{\mu\nu} x^\nu. \quad (1.2)$$

Mediante (1.1) y (1.2) tenemos $x_\mu = (ct, -\mathbf{x})$. Ahora definimos el tensor métrico contravariante $g^{\lambda\mu}$

$$g^{\lambda\mu} g_{\mu\nu} = g_\nu^\lambda = \delta_\nu^\lambda \quad (1.3)$$

con δ_ν^λ como la delta de Kronecker. Tomando (1.1) y (1.3) nos queda $g^{\mu\nu} = g_{\mu\nu}$. Establezcamos una transformación de Lorentz

$$x^\mu \rightarrow x'^\mu = \Lambda_\nu^\mu x^\nu, \quad (1.4)$$

haciendo que

$$x^\mu x_\mu = (x^0)^2 - \mathbf{x}^2 \quad (1.5)$$

sea invariante ($x^\mu x_\mu = x'^\mu x'_\mu$) y una cantidad escalar. Esto nos conduce a

$$\Lambda^{\lambda\mu} \Lambda_{\lambda\mu} = \delta_\nu^\mu. \quad (1.6)$$

Un objeto de cuatro componentes s^μ (s_μ) se transforma como x^μ (x_μ) bajo transformaciones de Lorentz, donde es fácil de ver que $s^\mu s_\mu$, es un cuatro vector contravariante.

Generalizamos el operador gradiente ∇ transformándose como un cuatro-vector. Sea un campo escalar $\phi(x)$ tal que su variación debido a una transformación de Lorentz infinitesimal es

$$\delta\phi = \frac{\partial\phi}{\partial x^\mu} \delta x^\mu,$$

y

$$\frac{\partial\phi}{\partial x^\mu} \equiv \partial_\mu \phi \equiv \phi_{,\mu} \quad (1.7)$$

es un cuatro-vector covariante. De la misma manera

$$\frac{\partial\phi}{\partial x_\mu} \equiv \partial^\mu\phi \equiv \phi_{,\mu} \quad (1.8)$$

es un cuatro-vector contravariante. Por último notamos que el operador \square es un escalar ¹

$$\partial^\mu\partial_\mu = \frac{\partial^2}{c^2\partial t^2} - \nabla^2 \equiv \square. \quad (1.9)$$

1.2 LA ECUACIÓN DE KLEIN-GORDON

Mediante la mecánica cuántica conocemos la ecuación de Schrödinger, que es fundamental ya que describe el estado cuántico de un sistema a lo largo del espacio y tiempo. En el mundo cuántico la evolución del sistema se describe como amplitudes de probabilidad contenidas en una función de onda, a partir de esto es que podemos calcular probabilidades, las cuales no son una probabilidad en el sentido tradicional (i.e., que pertenece al espacio L^1) debido a que los postulados de la mecánica cuántica conducen formalmente a un espacio de Hilbert complejo que es donde emergen nuestras amplitudes de probabilidad. Esto como es bien sabido nos da consistencia con la linealidad, superposición, conservación de la norma. Solo por mencionar un poco de dicha necesidad, la clave de utilizar este espacio vectorial completo es que, debido a que está dotado de un producto interno positivo por definición, nos garantiza que las probabilidades sean positivas por su naturaleza compleja. Dicho todo esto, comencemos ilustrando la ecuación

$$i\hbar\frac{\partial\psi}{\partial x_0} = i\left[\frac{-\hbar^2}{2m_0}\nabla^2 + V(x)\right]\psi(x,t) \quad (1.10)$$

que nos conduce a la energía no relativista en forma de operador,

$$\hat{E} = \frac{\hat{p}^2}{2m_0} + V(x), \quad (1.11)$$

donde los operadores de energía y momento son

$$\hat{E} = i\hbar\frac{\partial}{\partial t}, \quad \hat{p} = -i\hbar\nabla. \quad (1.12)$$

Para obtener una ecuación de onda relativista comenzamos considerando partículas libres con una relación relativista.

¹Más sobre esta notación puede consultarse en [22].

$$p^\mu p_\mu = \frac{E^2}{c^2} - \vec{p} \cdot \vec{p} = m_0 c^2. \quad (1.13)$$

Sustituyendo el cuatro-momento p^μ por el operador cuatro-momento obtenemos

$$\begin{aligned} \hat{p}^\mu &= i\hbar \left\{ \frac{\partial}{\partial x_0}, -\nabla \right\} = i\hbar \left\{ \frac{\partial}{\partial(ct)}, -\nabla \right\} \\ &= i\hbar \left\{ \frac{\partial}{\partial(ct)}, -\nabla \right\} = \left\{ \hat{p}_0, \hat{\vec{p}} \right\}, \end{aligned} \quad (1.14)$$

entonces se obtiene la ecuación de Klein-Gordon para partículas libres,

$$\hat{p}^\mu \hat{p}_\mu \psi = m_0^2 c^2 \psi, \quad (1.15)$$

donde m_0 es la masa en reposo de la partícula. Usando el operador D'Alembert es posible reescribir (1.15) en la siguiente forma

$$\left(\square + \frac{m_0^2 c^2}{\hbar^2} \right) \psi = \left(\frac{\partial^2}{c^2 \partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2} + \frac{m_0^2 c^2}{\hbar^2} \right) \psi = 0. \quad (1.16)$$

La ecuación de Klein-Gordon es invariante de Lorentz. Las soluciones a esta ecuación son de la siguiente manera

$$\begin{aligned} \psi &= \exp \left(-\frac{1}{\hbar} p_\mu x^\mu \right) = \exp \left[-\frac{i}{\hbar} (p_0 x^0 - \vec{p} \cdot \vec{x}) \right] \\ &= \exp \left[+\frac{i}{\hbar} (\vec{p} \cdot \vec{x}) - Et \right]. \end{aligned} \quad (1.17)$$

Sustituyendo la solución (1.17) en (1.15) nos conduce a las siguientes energías permitidas

$$E = \pm \sqrt{m_0^2 c^2 + p^2}. \quad (1.18)$$

Las soluciones negativas de la energía están relacionadas con las antipartículas. Ahora es necesario construir el cuatro-tensor de corriente j_μ relacionado con (1.15), con la cual se construye una ley de conservación. Comenzamos con

$$(\hat{p}^\mu \hat{p}_\mu - m_0^2 c^2) \psi = 0,$$

tomamos el complejo conjugado,

$$(\hat{p}^\mu \hat{p}_\mu - m_0^2 c^2) \psi^* = 0.$$

A la solución se le multiplica por la izquierda por ψ^* y al complejo conjugado por ψ , tomamos la diferencia de ambos y nos queda lo siguiente,

$$\begin{aligned} \psi^*(\hat{p}^\mu \hat{p}_\mu - m_0^2 c^2)\psi - \psi(\hat{p}^\mu \hat{p}_\mu - m_0^2 c^2)\psi^* &= 0, \\ -\psi^*(\hbar^2 \nabla_\mu \nabla^\mu + m_0^2 c^2)\psi + \psi(\hbar^2 \nabla_\mu \nabla^\mu + m_0^2 c^2)\psi^* &= 0, \\ \rightarrow \nabla_\mu(\psi^* \nabla^\mu \psi - \psi \nabla^\mu \psi^*) &\equiv \nabla_\mu j^\mu = 0. \end{aligned} \quad (1.19)$$

Entonces el cuatro-tensor de densidad de corriente es

$$j_\mu = \frac{i\hbar}{2m_0}(\psi^* \nabla_\mu \psi - \psi \nabla_\mu \psi^*). \quad (1.20)$$

Desarrollando la expresión (1.19),

$$\frac{\partial}{\partial t} \left[\frac{i\hbar}{2m_0 c^2} \left(\psi^* \frac{\partial \psi}{\partial t} - \psi \frac{\partial \psi^*}{\partial t} \right) \right] + \nabla \left[\frac{i\hbar}{2m_0} (\psi^* \nabla \psi - \psi \nabla \psi^*) \right] = 0. \quad (1.21)$$

La expresión anterior puede escribirse como una ecuación de continuidad

$$\frac{\partial \rho}{\partial t} + \nabla j = 0. \quad (1.22)$$

Es natural intentar interpretar

$$\rho = \frac{i\hbar}{2m_0 c^2} \left(\psi^* \frac{\partial \psi}{\partial t} - \psi \frac{\partial \psi^*}{\partial t} \right) \quad (1.23)$$

como una densidad de probabilidad. Esto significa un problema, ya que en contraste con la mecánica cuántica no relativista, $\rho(x, t)$ no está definida positiva debido a que en dados valores de t , ψ y $\partial\psi/\partial t$ pueden ser arbitrarios. Aunado a esto, también existen soluciones para energía negativa. Para enfrentar esta situación se buscó una ecuación de onda relativista de primer orden con probabilidad definida positiva. Dirac más tarde formuló dicha ecuación, y a su vez también tiene soluciones de energía negativa, esto dio lugar a describir la existencia de antipartículas como ya hemos mencionado.

1.3 LA ECUACIÓN DE DIRAC

En mecánica cuántica no relativista, el electrón se describe utilizando la ecuación de Schrödinger, pero no incorpora de manera adecuada la definición de spin ni es compatible con la relatividad especial. Para resolver estas cuestiones, en 1928 Paul Dirac propuso una nueva ecuación incluyendo el marco relativista entre energía y momento, y que al

mismo tiempo nos arroja una interpretación probabilística consistente. El resultado fue una ecuación cuyas soluciones incorporan el concepto de spin (en específico spin 1/2), y matrices de Dirac. Comencemos con la formulación básica de la siguiente forma

$$i\hbar \frac{\partial \psi}{\partial t} = \hat{H} \psi \quad (1.24)$$

con densidad de probabilidad definida positiva. La ecuación (1.24) es lineal en la derivada del tiempo, por lo que podemos construir un Hamiltoniano que lo es también en las derivadas espaciales. Entonces la ecuación (1.24) se propone sea de la siguiente manera

$$i\hbar \frac{\partial \psi}{\partial t} = \left[\frac{\hbar c}{i} \left(\hat{\alpha}_1 \frac{\partial}{\partial x^1} + \hat{\alpha}_2 \frac{\partial}{\partial x^2} + \hat{\alpha}_3 \frac{\partial}{\partial x^3} \right) + \hat{\beta} m_0 c^2 \right] \equiv \hat{H}_f \psi \quad (1.25)$$

Los coeficientes $\hat{\alpha}_i$ tienen que ser matrices para que (1.25) sea invariante respecto a rotaciones espaciales. De aquí deducimos que ψ no puede ser un simple escalar si no un vector columna

$$\psi = \begin{pmatrix} \psi_1(x, t) \\ \psi_2(x, t) \\ \vdots \\ \psi_N(x, t) \end{pmatrix}, \quad (1.26)$$

con una densidad definida positiva de la forma

$$\rho(x) = \psi^\dagger \psi(x) = (\psi_1^*, \psi_2^*, \dots, \psi_N^*) \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_N \end{pmatrix} = \sum_{i=1}^N \psi_i^* \psi_i(x) \quad (1.27)$$

que es posible construir. Cabe aclarar que $\rho(x)$ debe ser la componente temporal de un cuatro-vector (corriente) de tal forma que, una ecuación de continuidad existe, así mismo su integral espacial $\int \rho(x) d^3x$ sea constante en el tiempo. Al establecer este hecho podemos decir que tenemos una interpretación de probabilidad de $\rho(x)$. La ecuación de Schrödinger de la forma (1.24) y (1.26) nos muestra un sistema de N ecuaciones diferenciales de primer orden acopladas de las componentes del espinor ψ_i , $i = 1, \dots, N$. Especificando y reescribiendo (1.25) de la siguiente manera

$$i\hbar \frac{\partial \psi_\sigma}{\partial t} = \frac{\hbar c}{i} \sum_{\tau=1}^N \left(\hat{\alpha}_1 \frac{\partial}{\partial x^1} + \hat{\alpha}_2 \frac{\partial}{\partial x^2} + \hat{\alpha}_3 \frac{\partial}{\partial x^3} \right)_{\sigma\tau} \psi_\tau + m_0 c^2 \sum_{\tau=1}^N \hat{\beta}_{\sigma\tau} \psi_\tau \equiv \sum_{\tau=1}^N (\hat{H}_f)_{\sigma\tau} \psi_\tau \quad (1.28)$$

Ahora es necesario exigir las propiedades:

1. La relación de energía-momento para una partícula libre relativista

$$E^2 = \mathbf{p}^2 c^2 + m_0^2 c^4, \quad (1.29)$$

2. La ecuación de continuidad para la densidad en (1.27)

3. La covarianza de Lorentz para (1.25) y (1.28), respectivamente.

Para cumplir (a), cada componente ψ_0 del spinor ψ debe satisfacer la ecuación K-G,

$$-\hbar^2 \frac{\partial^2 \psi_\sigma}{\partial t^2} = (-\hbar^2 c^2 \nabla^2 + m_0^2 c^4) \psi_\sigma \quad (1.30)$$

Y de (1.25) se sigue que

$$\begin{aligned} -\hbar^2 \frac{\partial^2 \psi}{\partial t^2} &= \hbar^2 c^2 \sum_{i,j=1}^3 \frac{\hat{\alpha}_i \hat{\alpha}_j + \hat{\alpha}_j \hat{\alpha}_i}{2} \frac{\partial^2 \psi}{\partial x^i \partial x^j} \\ &+ \frac{\hbar m_0 c^3}{i} \sum_{i=1}^3 (\hat{\alpha}_i \hat{\beta} + \hat{\alpha}_i \hat{\beta}) \frac{\partial \psi}{\partial x^i} + \hat{\beta}^2 m_0^2 c^4 \psi \end{aligned}$$

Al comparar con (1.30) encontraremos algunas características a cumplir para las matrices $\hat{\alpha}_i, \hat{\beta}$:

$$\begin{aligned} \hat{\alpha}_i \hat{\alpha}_j + \hat{\alpha}_j \hat{\alpha}_i &= 2\delta_{ij} \mathbb{I}, \\ \hat{\alpha}_i \hat{\beta} + \hat{\alpha}_i \hat{\beta} &= 0, \\ \hat{\alpha}_i^2 &= \hat{\beta}^2 = \mathbb{I}. \end{aligned} \quad (1.31)$$

Vemos que estas propiedades de anticonmutación dan lugar a un álgebra para las matrices de ψ . Para establecer la hermiticidad del hamiltoniano \hat{H}_f , las matrices $\hat{\alpha}_i, \hat{\beta}$ también deben ser hermitianas; esto es,

$$\hat{\alpha}_i^\dagger = \hat{\alpha}_i, \quad \hat{\beta}^\dagger = \hat{\beta}. \quad (1.32)$$

Con esto tenemos que sus eigenvalores de las matrices son reales. Dado (1.31) observamos que los eigenvalores solo pueden tomar las cantidades ± 1 . Debido a su naturaleza podemos expresarlos de manera diagonal en matrices. Para $\hat{\alpha}_i$ tendría la forma

$$\hat{\alpha}_i = \begin{pmatrix} A_1 & 0 & 0 & \cdots & 0 \\ 0 & A_2 & 0 & \cdots & 0 \\ 0 & 1 & A_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & A_N \end{pmatrix},$$

con eigenvalores A_1, \dots, A_N . Al tener $\hat{\alpha}_i^2 = \mathbb{I}$ es fácil ver que

$$A_k^2 = 1, \quad \text{i.e.} \quad A_k = \pm 1. \quad (1.33)$$

Entonces, debido a las relaciones de conmutación (1.31) esto nos da como resultado que la traza de cada $\hat{\alpha}_i$ y de $\hat{\beta}$ tienen que ser cero. Con (1.31) podemos obtener el siguiente resultado

$$\hat{\alpha}_i = -\hat{\beta}\hat{\alpha}_i\hat{\beta}.$$

Con la identidad

$$tr(\hat{A}\hat{B}) = tr(\hat{B}\hat{A})$$

llegamos a que

$$tr(\hat{\beta}^2\hat{\alpha}_i) = tr(\hat{\beta}\hat{\alpha}_i\hat{\beta}) = -tr(\hat{\alpha}_i), \quad \rightarrow tr(\hat{\alpha}_i) = 0. \quad (1.34)$$

Esto nos dice que la traza de la matriz es igual a la suma de sus eigenvalores, lo podemos interpretar como que \hat{U} convierte la matriz $\hat{\alpha}_i$ a su forma diagonal partiendo de la forma análoga a como con (1.31) llegamos al resultado (1.34). Entonces,

$$tr \begin{pmatrix} A_1 & 0 & \cdots \cdots \cdots \\ 0 & A_2 & \cdots \cdots \cdots \\ \vdots & \vdots & \ddots \\ \cdots & \cdots & \cdots A_N \end{pmatrix} = \sum_{k=1}^N A_k = tr(\hat{U}\hat{\alpha}_i\hat{U}^{-i}) = tr(\hat{\alpha}_i),$$

dado que sus eigenvalores son iguales a ± 1 , cada matriz debe ser de dimensión par. La dimensión más pequeña que pueden tomar para cumplir (1.31) es $N = 4$. Ahora busquemos una representación más exacta de las matrices de Dirac, esto es

$$\hat{\alpha}_i = \begin{pmatrix} 0 & \hat{\sigma}_i \\ \hat{\sigma}_i & 0 \end{pmatrix}, \quad \hat{\beta} = \begin{pmatrix} \mathbb{I} & 0 \\ 0 & \mathbb{I} \end{pmatrix}, \quad (1.35)$$

donde $\hat{\alpha}_i$ son las matrices de Pauli 2×2 y \mathbb{I} la matriz unitaria 2×2 . Con esta forma ya establecida es fácil ver por ejemplo de (1.31) que

$$\hat{\alpha}_i\hat{\alpha}_j + \hat{\alpha}_j\hat{\alpha}_i = 2\delta_{ij}\mathbb{I}.$$

También se observa que (1.35) muestra solamente una posible selección de las matrices de Dirac. Ahora necesitamos construir la 4-densidad de corriente y la ecuación de continuidad. Para eso multiplicamos (1.25) por $\psi^\dagger = (\psi_1^*, \psi_2^*, \psi_3^*, \psi_4^*)$

$$i\hbar\psi^\dagger \frac{\partial}{\partial t} \psi = \frac{\hbar c}{i} \sum_{k=1}^3 \psi^\dagger \hat{\alpha}_k \frac{\partial}{\partial x^k} \psi + m_0 c^2 \psi^\dagger \hat{\beta} \psi. \quad (1.36)$$

Formamos el Hermitiano conjugado de (1.25), esto es

$$-i\hbar\psi^\dagger \frac{\partial \psi^\dagger}{\partial t} = -\frac{\hbar c}{i} \sum_{k=1}^3 \frac{\partial \psi^\dagger}{\partial x^k} \hat{\alpha}_k + m_0 c^2 \psi^\dagger \hat{\beta}^\dagger,$$

multiplicando esta última ecuación por ψ , haciendo uso de las propiedades hermitianas de las matrices de Dirac ($\hat{\alpha}_i^\dagger = \hat{\alpha}_i, \hat{\beta}^\dagger = \hat{\beta}$) y restando ambas ecuaciones nos queda lo siguiente

$$i\hbar \frac{\partial}{\partial t} (\psi^\dagger \psi) = \frac{\hbar c}{i} \sum_{k=1}^3 \frac{\partial}{\partial x^k} (\psi^\dagger \hat{\alpha}_k \psi) \quad (1.37)$$

Lo que nos da la forma de

$$\frac{\partial \rho}{\partial t} + \text{div}(\mathbf{j}) = 0, \quad (1.38)$$

donde la densidad de probabilidad positiva de (1.27) es

$$\rho = \psi^\dagger \psi = \sum_{i=1}^4 \psi_i^* \psi_i \quad (1.39)$$

y la densidad de corriente es

$$j^k = c\psi^\dagger \hat{\alpha}^k \psi. \quad (1.40)$$

De (1.38) podemos determinar fácilmente una ley de conservación

$$\frac{\partial}{\partial t} \int_V d^3x \psi^\dagger \psi = - \int_V \text{div}(\mathbf{j}) d^3x = - \int_F \mathbf{j} \cdot d\mathbf{f} = 0 \quad (1.41)$$

Donde V denota un volumen y F su superficie, este tema puede ser abordado con profundidad en [23] y para analizar los principios y reglas de conservación [24]. En general podemos decir que la ecuación de Dirac se construyó para ser invariante ante las transformaciones de Lorentz, aplicando el operador de la representación espinorial del grupo de Lorentz, dichas transformaciones son las que garantizan que las leyes físicas están en el marco de la relatividad especial. Más adelante vamos a establecer las convenciones de la notación relativista para desarrollar de manera más conveniente la teoría cuántica de campos.

1.4 EL OSCILADOR ARMÓNICO

Ahora es relevante revisar brevemente el oscilador armónico cuántico, ya que es bien conocido que funciona como un ejemplo sencillo y adecuado de un sistema resoluble, e introducir de manera natural el formalismo de operadores de creación y aniquilación. Este formalismo es importante pues más adelante se reutilizan estas ideas para los campos cuánticos y su interpretación de los grados de libertad de una partícula en términos de cuantos que pueden crearse y destruirse. Como sabemos el oscilador armónico está dado por²,

$$H_{osc} = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 q^2,$$

donde p y q tiene la relación de conmutación $[q, p] = i\hbar$. Es necesario establecer los siguientes operadores

$$a = \frac{1}{(2\hbar m\omega)^{1/2}}(m\omega q + ip), \quad a^\dagger = \frac{1}{(2\hbar m\omega)^{1/2}}(m\omega q - ip).$$

Con la relación de conmutación

$$[a, a^\dagger] = 1, \tag{1.42}$$

con estas definiciones podemos reescribir el Hamiltoniano como sigue:

$$H_{osc} = \frac{1}{2}\hbar\omega(a^\dagger a + aa^\dagger) = \hbar\omega\left(a^\dagger a + \frac{1}{2}\right) \tag{1.43}$$

Ahora definimos el operador

$$N \equiv a^\dagger a, \tag{1.44}$$

y es definido positivo para cada estado $|\Psi\rangle$

$$\langle\Psi|N|\Psi\rangle = \langle\Psi|a^\dagger a|\Psi\rangle = \langle a\Psi|a\Psi\rangle \geq 0.$$

Esto nos dice que N tiene el eigenvalor más bajo no negativo

$$\alpha_0 \geq 0.$$

Mediante la ecuación de eigenvalores

$$N|\alpha\rangle = \alpha|\alpha\rangle$$

²Este y otros temas relacionados pueden ser revisados en [25] y [26]

y la relación de conmutación (1.42)

$$Na|\alpha\rangle = (\alpha - 1)a|\alpha\rangle, \quad Na^\dagger|\alpha\rangle = (\alpha + 1)a^\dagger|\alpha\rangle, \quad (1.45)$$

Se muestra que $a|\alpha\rangle$ y $a^\dagger|\alpha\rangle$ son eigenfunciones de N correspondientes a los eigenvalores $(\alpha - 1)$ y $(\alpha + 1)$. Como α_0 es el eigenvalor más bajo se debe cumplir que

$$a|\alpha_0\rangle = 0, \quad (1.46)$$

y

$$a^\dagger a|\alpha_0\rangle = \alpha_0|\alpha_0\rangle.$$

La ecuación (1.46) implica que $\alpha_0 = 0$. Con (1.42) y (1.45) vemos que los eigenvalores de N son enteros $n = 0, 1, \dots$, y que si $\langle n| = 1$, los estados $|n \pm 1\rangle$ definidos mediante

$$a|n\rangle = n^{1/2}|n-1\rangle, \quad a^\dagger|n\rangle = (n+1)^{1/2}|n+1\rangle, \quad (1.47)$$

están normalizados a la unidad. Si $\langle 0| = 1$, Las autofunciones normalizadas de N vienen dadas por

$$|N\rangle = \frac{(a^\dagger)^n}{\sqrt{n!}}|0\rangle, \quad n = 0, 1, 2, \dots \quad (1.48)$$

Así mismo son eigenfunciones del Hamiltoniano del oscilador armónico (1.43) con los siguientes valores de energía

$$E_n = \hbar\omega \left(n + \frac{1}{2} \right), \quad n = 0, 1, 2, \dots \quad (1.49)$$

Los operadores a y a^\dagger se les llama operadores de creación y aniquilación por las propiedades de (1.47). Más adelante se mostrará que en la teoría de campo cuantizada $|n\rangle$ representa un estado con n quanta. Como es de esperarse el operador a (transformando $|n\rangle$ a $|n-1\rangle$) aniquilará un cuanto; a su vez, a^\dagger , creará un cuanto.

En la forma de Heisenberg los operadores son funciones del tiempo

$$i\hbar \frac{da(t)}{dt} = [a(t), H_{osc}] \quad (1.50)$$

con la condición inicial $a(0) = a$. como H_{osc} es independiente del tiempo, y $a(t)^\dagger$ y $a(t)$ satisfacen (1.42), la ecuación queda como

$$\begin{aligned} \frac{da(t)}{dt} &= i\omega a(t), \\ a(t) &= ae^{-i\omega t}. \end{aligned} \quad (1.51)$$

Nuestro siguiente acercamiento será cuantizar directamente los campos en cada punto del espacio así como sus variables dinámicas. De esta forma podremos generalizar la mecánica clásica a un sistema de partículas, y su cuantización hacia un sistema continuo; los campos. Introduciremos una densidad Lagrangiana, usando el principio Hamiltoniano e introduciendo el conjugado del momento a los campos, dando como resultado relaciones de conmutación canónica. Primeramente necesitaremos desarrollar la teoría en la forma covariante relativista la cuál será mostrada a continuación.

1.5 DENSIDAD LAGRANGIANA CLÁSICA DE CAMPOS

Ahora llevaremos la mecánica Lagrangiana a campos que dependen del espacio y del tiempo. Vamos a considerar un sistema con varios campos $\phi_r(x)$, $r = 1, \dots, N$. El índice r puede ser para componente del mismo campo o campos independientes. Se abordará la teoría desde el principio variacional de una integral de acción involucrando una densidad Lagrangiana³

$$\mathcal{L} = \mathcal{L}(\phi_r, \phi_{r,\alpha}) \quad (1.52)$$

vemos que la densidad Lagrangiana depende de los campos y sus primeras derivadas. Ahora establecemos la integral de acción $S(\Omega)$ para una región arbitraria Ω del continuo espacio-tiempo 4-dimensional

$$S(\Omega) = \int_{\Omega} d^4x \mathcal{L}(\phi_r, \phi_{r,\alpha}), \quad (1.53)$$

donde d^4x es para el elemento cuatro-dimensional $dx^0 dx^3 \mathbf{x}$. Formularemos las ecuaciones de movimiento, esto es las de campo. Siguiendo el principio variacional, para una región arbitraria Ω , consideramos variaciones en los campos,

$$\phi_r(x) \rightarrow \phi_r(x) + \delta\phi_r(x), \quad (1.54)$$

el cual se hace muy pequeño en la superficie $\Gamma(\Omega)$ acotando la región Ω

$$\delta\phi_r(x) = 0 \text{ en } \Gamma(\Omega). \quad (1.55)$$

³Esta información está basada en [24, 27, 28]

Calculando la variación $\delta S(\Omega)$ usando la forma de (1.53)

$$\begin{aligned} \delta S(\Omega) &= \int_{\Omega} d^4x \left\{ \frac{\partial \mathcal{L}}{\partial \phi_r} \delta \phi_r + \frac{\partial \mathcal{L}}{\partial \phi_{r,\alpha}} \delta \phi_{r,\alpha} \right\} \\ &= \int_{\Omega} d^4x \left\{ \frac{\partial \mathcal{L}}{\partial \phi_r} + \frac{\partial}{\partial x^\alpha} \left(\frac{\partial \mathcal{L}}{\partial \phi_{r,\alpha}} \right) \right\} \delta \phi_r + \int_{\Omega} d^4x \frac{\partial}{\partial x^\alpha} \left(\frac{\partial \mathcal{L}}{\partial \phi_{r,\alpha}} \delta \phi_r \right), \end{aligned} \quad (1.56)$$

el último desarrollo fue hecho con integración por partes, ya que

$$\delta \phi_{r,\alpha} = \frac{\partial}{\partial x^\alpha} \delta \phi_r.$$

Usando el teorema de la divergencia en cuatro dimensiones podemos convertir el último término de (1.56) a una integral de superficie sobre $\Gamma(\Omega)$. Como $\delta \phi_r = 0$ en Γ , esta integral se hace cero. $\delta S(\Omega)$ se acerca a cero para regiones arbitrarias Ω y variaciones arbitrarias $\delta \phi_r$, la ec. (1.56) nos lleva a las ecuaciones de Euler-Lagrange

$$\frac{\partial \mathcal{L}}{\partial \phi_r} - \frac{\partial}{\partial x^\alpha} \left(\frac{\partial \mathcal{L}}{\partial \phi_{r,\alpha}} \right) = 0, \quad r = 1, \dots, N. \quad (1.57)$$

Y así hemos obtenido las ecuaciones de movimiento de los campos. Como es de esperarse para cuantizar la teoría clásica necesitamos introducir variables conjugadas. Vamos a aproximar el sistema teniendo un número contable de grados de libertad para así dirigirnos al límite continuo.

Consideremos un sistema fijo en un tiempo t donde vamos a descomponer su espacio tridimensional, en pequeñas celdas de volumen $\delta \mathbf{x}_i$; $i = 1, 2, \dots$. Realizaremos esta aproximación de los valores de los campos dentro de cada celda en, por ejemplo, el centro de la celda $\mathbf{x} = \mathbf{x}_i$. Nuestro sistema estará descrito por el conjunto discreto de coordenadas generalizadas:

$$q_{ri}(t) \equiv \phi_r(i, t) \equiv \phi_r(\mathbf{x}_i, t), \quad r = 1, \dots, N, \quad i = 1, \dots \quad (1.58)$$

Necesitamos conocer los campos evaluados en la vecindad de los puntos de la red i, i' , así la función Lagrangiana queda como sigue

$$L(t) = L \left(\phi_r(i, t), \dot{\phi}_r(i, t), \phi_r(i', t) \right) \quad (1.59)$$

Decimos que la densidad Lagrangiana en la i -ésima celda, \mathcal{L}_i , depende de los campos en las redes i' vecinas debido a la aproximación de las derivadas espaciales. Ahora establecemos nuestro momento conjugado (campo adjunto) a $q_{ri}(t)$ como

$$p_{ri}(t) = \frac{\partial L}{\partial \dot{q}_{ri}} \equiv \frac{\partial L}{\partial \dot{\phi}_r(i, t)} \equiv \pi_r(i, t) \delta \mathbf{x}_i \quad (1.60)$$

con

$$\pi_r(i, t) \equiv \frac{\partial \mathcal{L}_i}{\partial \dot{\phi}_r(i, t)}. \quad (1.61)$$

El Hamiltoniano del sistema discreto está dado por

$$H = \sum_i p_{ri} \dot{q}_{ri} - L = \sum_i \delta \mathbf{x}_i \left\{ \pi_{ri}(i, t) \dot{\phi}_r(i, t) - \mathcal{L}_i \right\}. \quad (1.62)$$

Donde $\delta \mathbf{x}_i$ es el elemento de volumen. Encogiendo el elemento de volumen $\delta \mathbf{x}_i \rightarrow 0$, se define

$$\pi_r(x) = \frac{\partial \mathcal{L}}{\partial \dot{\phi}_r}. \quad (1.63)$$

Conforme $\delta \mathbf{x}_i \rightarrow 0$, $\pi_r(i, t)$ tiende a $\pi_r(\mathbf{x}_i, t)$, por lo que el Lagrangiano y Hamiltoniano quedan de la siguiente manera

$$L(t) = \int d^3 \mathbf{x} \mathcal{L}(\phi_r, \phi_{r,\alpha}) \quad (1.64)$$

y

$$H = \int d^3 \mathbf{x} \mathcal{H}(x), \quad (1.65)$$

con la densidad hamiltoneana,

$$\mathcal{H}(x) = \pi_r(x) \dot{\phi}_r(x) - \mathcal{L}(\phi_r, \phi_{r,\alpha}). \quad (1.66)$$

Como ejemplo, consideremos la siguiente densidad Lagrangiana

$$\mathcal{L} = \frac{1}{2} (\phi_{,\alpha} \phi^\alpha - \mu^2 \phi^2) \quad (1.67)$$

para el campo real $\phi(x)$, con μ^2 constante. La ecuación de movimiento (1.57) para este caso es la ecuación de Klein-Gordon

$$(\square + \mu^2) \phi(x) = 0, \quad (1.68)$$

Con su respectivo campo conjugado (1.63)

$$\pi(x) = \frac{1}{c^2} \dot{\phi}(x) \quad (1.69)$$

por último su densidad Hamiltoniana (1.66) queda como

$$\mathcal{H}(x) = \frac{1}{2} (c^2 \pi^2(x) + (\nabla \phi)^2 + \mu^2 \phi^2). \quad (1.70)$$

1.6 CUANTIZACIÓN DE LA TEORÍA DE CAMPOS LAGRANGIANA

Ahora que se ha identificado el campo y su adjunto, se tomarán las ecs. (1.58) y (1.60) como operadores de Heisenberg (nuestra convención es mantener la notación de la misma manera pero de ahora en adelante considerando a estos campos como operadores) que es lo que nos traslada de una teoría clásica de campos a una teoría cuántica donde los campos se vuelven operadores que crean y destruyen partículas, y sujetos a las relaciones de conmutación⁴

$$[\phi_r(j, t), \pi_s(j', t)] = i\hbar \frac{\delta_{rs} \delta_{jj'}}{\delta \mathbf{x}_j}, \quad (1.71)$$

este resultado surge de

$$\begin{aligned} [q_{rj}, p_{r',j'}] &= i\hbar \delta_{rr'} \delta_{jj'}, \\ [q_{rj}, \pi_r'(j', t) \delta \mathbf{x}_j] &= i\hbar \delta_{rr'} \delta_{jj'}, \\ [\phi_r(j, t), \pi_r'(j', t) \delta \mathbf{x}_j] &= i\hbar \delta_{rr'} \delta_{jj'}, \end{aligned}$$

donde $\delta_{rs}, \delta_{jj'}$ son las delta de Kronecker y el $\delta \mathbf{x}_j$ representa el elemento de volumen.

Cuando $\delta \mathbf{x}_j \rightarrow 0$,

$$[\phi_r(\mathbf{x}, t), \pi_s(\mathbf{x}', t)] = \lim_{\delta \mathbf{x}_j \rightarrow 0} \frac{i\hbar \delta_{rs} \delta_{jj'}}{\delta \mathbf{x}_j},$$

lo cual implica las siguientes relaciones de conmutación

$$\begin{aligned} [\phi_r(\mathbf{x}, t), \pi_s(\mathbf{x}', t)] &= i\hbar \delta_{rs} \delta(\mathbf{x} - \mathbf{x}'), \\ [\phi_r(\mathbf{x}, t), \phi_s(\mathbf{x}', t)] &= 0, \\ [\pi_r(\mathbf{x}, t), \pi_s(\mathbf{x}', t)] &= 0, \end{aligned} \quad (1.72)$$

donde $\delta(\mathbf{x} - \mathbf{x}')$ es la delta de Dirac tridimensional. El reemplazo de los corchetes de Poisson por conmutadores que hemos hecho se utiliza para campos bosónicos que tienen como característica tener spin 1, un trato análogo se hace para campos fermiónicos pero a estas se les conoce como relaciones de anticonmutación y corresponden a partículas de spin 1/2. A continuación veremos un detalle importante que son las simetrías de este formalismo, la física se rige por las simetrías que encontramos en nuestros propios modelos y construcciones matemáticas, se busca para generalizar las ideas establecidas y hacer más sencillo los casos de estudio.

⁴Para más detalles esta información está basada en [27-29]

1.7 SIMETRÍAS Y LEYES DE CONSERVACIÓN

Las constantes de movimiento surgen de sus propiedades de invarianza bajo grupos de transformaciones; en el caso de las invarianzas traslacional y rotacional nos llevan a la conservación del momento lineal o angular por ejemplo, como ya mencionamos, es necesario establecer estas constantes provenientes de su naturaleza matemática, pues tener una física que siempre está cambiando en nuestros modelos nos llevaría a procedimientos más exhaustivos. Para establecer una teoría de campos, derivada de la densidad Lagrangiana \mathcal{L} , se construyen cantidades conservadas de la invarianza de \mathcal{L} bajo transformaciones de simetría. Dicha invarianza nos debe llevar a ecuaciones de la forma⁵

$$\partial_\alpha f^\alpha = 0; \quad f^\alpha = \text{corriente conservada.} \quad (1.73)$$

Definimos

$$F^\alpha(t) = \int d^3\mathbf{x} f^\alpha(\mathbf{x}, t), \quad (1.74)$$

con el dominio de integración en todo el espacio, la ecuación de continuidad nos da

$$\frac{1}{c} \frac{dF^0(t)}{dt} = - \int d^3\mathbf{x} \sum_{j=1}^3 \frac{\partial}{\partial x^j} f^j(\mathbf{x}, t) = 0 \quad (1.75)$$

Utilizando el teorema de la divergencia y asumiendo que los campos, f^j , tiende a cero rápidamente en el infinito, nos lleva a lo siguiente

$$F^0 = \int d^3\mathbf{x} f^0(\mathbf{x}, t) \quad (1.76)$$

esto nos dice que hay una constante de movimiento, es decir, una cantidad conservada. En este orden proponemos la siguiente transformación para los campos:

$$\phi_r(x) \rightarrow \phi'_r(x) = \phi_r(x) + \delta\phi_r(x). \quad (1.77)$$

Los cambios en \mathcal{L} son

$$\delta\mathcal{L} = \frac{\partial\mathcal{L}}{\partial\phi_r} \delta\phi_r + \frac{\partial\mathcal{L}}{\partial\phi_{r,\alpha}} \delta\phi_{r,\alpha} = \frac{\partial}{\partial x^\alpha} \left(\frac{\partial\mathcal{L}}{\partial\phi_{r,\alpha}} \delta\phi_r \right).$$

⁵En esta subsección se sigue la presentación estándar de simetrías continuas y cantidades conservadas dada en [24, 27, 28].

Pero si \mathcal{L} es invariante bajo transformaciones del tipo propuesto tal que $\delta\mathcal{L} = 0$, se reduce a la ecuación de continuidad

$$f^\alpha = \frac{\partial\mathcal{L}}{\partial\phi_{r,\alpha}}\delta\phi$$

con esto la constante de movimiento nos da,

$$F^0 = c \int d^3\mathbf{x}\pi_r(x)\delta\phi_r(x). \quad (1.78)$$

Ahora ilustraremos un caso importante para operadores no hermitianos (campos complejos ϕ_r). Suponemos que \mathcal{L} es invariante bajo las siguientes transformaciones

$$\phi_r \rightarrow \phi'_r = e^{i\epsilon}\phi_r \approx (1 + i\epsilon)\phi_r, \quad \phi_r^\dagger \rightarrow \phi'^\dagger_r = e^{-i\epsilon}\phi_r^\dagger \approx (1 - i\epsilon)\phi_r^\dagger \quad (1.79)$$

donde $\mathbb{R} \ni \epsilon \ll 1$. Luego tenemos $\delta\phi_r = i\epsilon\phi_r$, $\delta\phi_r^\dagger = -i\epsilon\phi_r^\dagger$.

La ecuación (1.78) se convierte en

$$F^0 = i\epsilon c \int d^3\mathbf{x} [\pi_r(x)\phi_r(x) - \pi_r^\dagger(x)\phi_r^\dagger(x)].$$

Sea $\epsilon = -q/\hbar c$, donde q resultará ser la carga eléctrica.

Como F^0 es constante, debe cumplir esta propiedad al multiplicarlo por otra constante, consideremos entonces

$$Q = -\frac{iq}{\hbar} \int d^3\mathbf{x} [\pi_r(x)\phi_r(x) - \pi_r^\dagger(x)\phi_r^\dagger(x)]. \quad (1.80)$$

Aquí hay algunas observaciones que podemos concebir:

Si $\phi_r(x)$ es un campo real, entonces $\phi_r^* = \phi_r$, a su vez $\pi_r^* = \pi_r$, quiere decir que $F^0 = 0$ (no hay carga eléctrica).

Si $\phi_r(x) \in \mathbb{C}$, entonces $\phi_r^* \neq \phi_r$, nos lleva a $F^0 \neq 0$, $F^0 >$, $F^0 <$ 0. En otras palabras; un campo complejo representará una partícula cargada.

Ahora evaluamos el conmutador $[Q, \phi_r(x)]$. Recordando que ϕ_r y ϕ_r^\dagger son campos independientes y sus relaciones de conmutación,

$$[Q, \phi_r(x)] = -\frac{iq}{\hbar} \int d^3\mathbf{x}' [\pi_s(x'), \phi_r(x)]\phi_s(x'),$$

finalmente nos queda

$$[Q, \phi_r(x)] = -q\phi_r(x). \quad (1.81)$$

Mediante la última ecuación podemos comprobar que si $|Q'\rangle$ es un eigenestado de Q con eigenvalor Q' , de manera análoga $\phi_r(x)|Q'\rangle$ es un eigenestado de Q con eigenvalor $(Q' - q)$ y $\phi_r^\dagger(x)|Q'\rangle$ con eigenvalor $(Q' + q)$.

La conservación de energía, momento angular y momento se pueden obtener con la propiedad de invarianza de la densidad Lagrangiana bajo transformaciones de rotación y traslación. Considere la siguiente transformación infinitesimal en 4-dimensiones

$$x_\alpha \rightarrow x'_\alpha \equiv x_\alpha + \delta x_\alpha \equiv x_\alpha + \epsilon_{\alpha\beta} x^\beta + \delta_\alpha, \quad (1.82)$$

donde δ_α es un desplazamiento infinitesimal y $\epsilon_{\alpha\beta}$ un tensor antisimétrico infinitesimal, $\epsilon_{\alpha\beta} = -\epsilon_{\beta\alpha}$.

La transformación de la ec. (1.82) va a realizar una transformación en los campos de la siguiente manera

$$\phi_r(x) \rightarrow \phi'_r(x') = \phi_r(x) + \frac{1}{2} \epsilon_{\alpha\beta} S_{rs}^{\alpha\beta} \phi_s(x). \quad (1.83)$$

Para esta sección se usa el convenio de la suma bajo índices repetidos donde r, s es el número de campos y α, β como índices de Lorentz, x y x' representan el mismo punto en el espacio-tiempo vistos desde dos marcos de referencia ϕ_r, ϕ'_r . Los coeficientes $S_{rs}^{\alpha\beta} \phi_s$ son antisimétricos en α y β , los cuales se determinan por las propiedades de transformación de los campos. La invarianza bajo las transformaciones de la ec. (1.82) y (1.83) significa que la densidad Lagrangiana tiene la misma forma de funcional cuando se utilizan las nuevas coordenadas y campos como si estas fuesen las originales:

$$\mathcal{L}(\phi_r(x), \phi_{r,\alpha}(x)) = \mathcal{L}(\phi'_r(x'), \phi'_{r,\alpha}(x')). \quad (1.84)$$

Recordando que $\phi_{r,\alpha}(x') \equiv \partial \phi'_r(x') / \partial x'^\alpha$. Las leyes de conservación se obtienen al expresar el lado derecho de la ec. (1.84) en términos de las coordenadas originales y campos mediante las transformaciones propuestas (ec. (1.82) y (1.83)). Para la traslación ($\epsilon_{\alpha\beta} = 0$), obtenemos las cuatro ecuaciones de continuidad

$$\frac{\partial \mathcal{F}^{\alpha\beta}}{\partial x^\alpha} = 0, \quad (1.85)$$

con

$$\mathcal{F}^{\alpha\beta} = \frac{\partial \mathcal{L}}{\partial \phi_{r,\alpha}} \frac{\phi_r}{\partial x_\beta} - \mathcal{L} g^{\alpha\beta}, \quad (1.86)$$

las cuatro cantidades quedan como

$$cP^\alpha \equiv \int d^3\mathbf{x} \mathcal{F}^{0\alpha} = \int d^3\mathbf{x} \left[c\pi_r(x) \frac{\partial \phi_r(x)}{\partial x_\alpha} - \mathcal{L} g^{0\alpha} \right]. \quad (1.87)$$

Donde $P^{0\alpha}$ es el cuatro-vector energía-momento

$$cP^0 = \int d^3\mathbf{x} [\pi_r(x) \dot{\phi}_r(x) - \mathcal{L}(\phi_r, \phi_{r,\alpha})]$$

$$= \int d^3\mathbf{x} \mathcal{H} = H \quad (1.88)$$

siendo este el Hamiltoniano, y

$$P^j = \int d^3\mathbf{x} \pi_r(x) \frac{\phi_r(x)}{\partial x_j} \quad (1.89)$$

siendo las componentes del momento de los campos.

Ahora vamos a derivar las ecuaciones de continuidad. La variación de la función $\phi_r(x)$ de acuerdo a la ec. (1.77) es

$$\delta\phi_r(x) \equiv \phi'_r(x) - \phi_r(x). \quad (1.90)$$

también definiremos la variación

$$\delta_T\phi_r(x) \equiv \phi'_r(x') - \phi_r(x). \quad (1.91)$$

Podemos reescribir entonces de la siguiente manera

$$\begin{aligned} \delta_T\phi_r(x) &= [\phi'_r(x') - \phi_r(x)] + [\phi'_r(x) - \phi_r(x)] \\ &= \delta\phi_r(x') + \frac{\partial\phi_r(x)}{\partial x_\beta} \delta x_\beta, \end{aligned} \quad (1.92)$$

mediante ec. (1.82) se muestra δx_β . A primer orden podemos aproximar para cantidades pequeñas de la siguiente manera

$$\delta_T\phi_r(x) = \delta\phi_r(x) + \frac{\partial\phi_r}{\partial x_\beta} \delta x_\beta. \quad (1.93)$$

Similarmente para la ec. (1.84)

$$\begin{aligned} 0 &= \mathcal{L}(\phi'_r(x'), \phi_{r,\alpha}(x')) - \mathcal{L}(\phi_r(x), \phi_{r,\alpha}(x)) \\ &= \delta\mathcal{L} + \frac{\partial\mathcal{L}}{\partial x^\alpha} \delta x^\alpha. \end{aligned} \quad (1.94)$$

Para $\delta\mathcal{L}$ obtenemos lo siguiente,

$$\begin{aligned} \delta\mathcal{L} &= \frac{\partial\mathcal{L}}{\partial\phi_r} \delta\phi_r + \frac{\partial\mathcal{L}}{\partial\phi_{r,\alpha}} \delta\phi_{r,\alpha} \\ &= \frac{\partial}{\partial x^\alpha} \left[\frac{\partial\mathcal{L}}{\partial\phi_{r,\alpha}} \delta\phi_r \right] = \frac{\partial}{\partial x^\alpha} \left[\frac{\partial\mathcal{L}}{\partial\phi_{r,\alpha}} \left(\delta_T\phi_r - \frac{\partial\phi_r}{\partial x_\beta} \delta x_\beta \right) \right], \end{aligned} \quad (1.95)$$

Combinando ec. (1.94) y (1.95) nos queda

$$\frac{\partial f^\alpha}{\partial x^\alpha} = 0, \quad (1.96)$$

donde

$$f^\alpha = \frac{\partial \mathcal{L}}{\partial \phi_{r,\alpha}} \delta_T \phi_r - \mathcal{F}^{\alpha\beta} \delta x_\beta \quad (1.97)$$

donde $\mathcal{F}^{\alpha\beta}$ es el dado en la ec. (1.86), como siguiente paso vamos a considerar traslaciones de la forma $\epsilon_{\alpha\beta} = 0$ tal que de (1.82) y (1.83), $\delta x_\beta = \delta_\beta$ y $\delta_T \phi_r = 0$, así la ec. (1.97) se convierte en $f^\alpha = -\mathcal{F}^{\alpha\beta} \delta x_\beta$, considerando la independendencia de los cuatro desplazamiento, llegamos a la ecuación de continuidad para la conservación de energía y momento[27].

Supongamos una rotación $\delta_\alpha = 0$

$$\begin{aligned} x'_\alpha &= x_\alpha + \epsilon_{\alpha\beta} x^\beta, \\ \delta x_\alpha &= x'_\alpha - x_\alpha = \epsilon_{\alpha\beta} x^\beta, \\ \phi'_r(x') &= \phi_r(x) + \frac{1}{2} \epsilon_{\alpha\beta} S_{rs}^{\alpha\beta}, \\ \rightarrow \delta_T \phi_r &= \phi'_r(x') - \phi_r(x), \\ &= \frac{1}{2} \epsilon_{\alpha\beta} S_{rs}^{\alpha\beta} \phi_s(x), \end{aligned}$$

luego

$$\begin{aligned} f^\alpha &= \frac{\partial \mathcal{L}}{\partial \phi_{r,\alpha}} \delta_T \phi_r - \mathcal{J}^{\alpha\beta} \delta x_\beta, \\ &= \frac{\partial \mathcal{L}}{\partial \phi_{r,\alpha}} \frac{1}{2} \epsilon_{\mu\nu} S_{rs}^{\mu\nu} - \mathcal{J}^{\alpha\beta} \epsilon_{\beta\nu} x^\nu, \\ &\rightarrow \partial_\alpha \mathcal{J}^{\alpha\beta} = 0, \end{aligned}$$

$$\int \mathcal{J}^{0\beta} dx^3 = cte; \quad \mathcal{J}^{\alpha\beta} = \left(\frac{\partial \mathcal{L}}{\partial \phi_{r,\alpha}} \right) \left(\frac{\partial \phi_r}{\partial x_\beta} \right) - \mathcal{L} g^{\alpha\beta}.$$

Así

$$\begin{aligned} f^\alpha &= \frac{1}{2} \epsilon_{\beta\nu} \mathcal{M}^{\alpha\beta\nu} \quad \text{con} \quad \mathcal{M}^{\alpha\beta\nu} = \frac{\partial \mathcal{L}}{\partial \phi_{r,\alpha}} S_{rs}^{\beta\nu} \phi_s + (x^\beta \mathcal{J}^{\alpha\nu} - x^\nu \mathcal{J}^{\alpha\beta}), \\ \partial_\alpha f^\alpha &= 0 \quad \Rightarrow \quad \partial_\alpha \mathcal{M}^{\alpha\beta\nu} = 0. \end{aligned}$$

Integrando en todo el volumen obtenemos

$$c \mathcal{M}^{\alpha\beta} = \int dx^3 M^{0\alpha\beta} = 0 = \text{cantidad conservada.}$$

En la siguiente sección utilizaremos el análisis de Fourier de algunos operadores de campo que hemos estado observando en algunas secciones, esto nos llevará a obtener que los coeficientes de Fourier que serán tratados como operadores, satisfacen algunas relaciones de conmutación, por ejemplo: como los de absorción y creación. Lo que nos lleva a que la interpretación en el campo cuantizado está garantizada.

1.8 EL CAMPO REAL DE KLEIN-GORDON

En esta subsección se sigue la presentación estándar del campo escalar real de Klein-Gordon y su formulación en teoría cuántica de campos dada en [27–29]. Vamos a considerar partículas relativistas de spin 0. A diferencia de la ecuación de K-G vista en la sección 1.2 aquí nuestro objeto es un campo real. Para partículas con masa en reposo m , la energía y momento están relacionados por

$$E^2 = m^2 c^4 + c^2 \mathbf{p}^2 \quad (1.98)$$

Si las partículas son descritas por una función de onda escalar $\phi(x)$, tenemos la noción no relativista de la mecánica cuántica

$$\mathbf{p} \rightarrow -i\hbar\nabla, \quad E \rightarrow i\hbar\partial/\partial t \quad (1.99)$$

que nos conduce a la ecuación de Klein-Gordon (1.68)

$$(\square + \mu^2)\phi(x) = 0. \quad (1.100)$$

Un campo escalar describe partículas con spin cero; por tanto, cualquier momento angular es puramente orbital. En consecuencia, la ecuación de Klein-Gordon(K-G) es un gran candidato para la descripción de π -mesones (piones) y K -mesones, ambos con spin 0. Tomemos un campo escalar real $\phi(x)$, que satisface la ecuación de K-G. Este campo está relacionado a partículas eléctricamente neutras. Por secciones anteriores se sabe que la ecuación de K-G puede obtenerse mediante la densidad Lagrangiana

$$\mathcal{L} = \frac{1}{2} (\phi_{,\alpha}\phi^{,\alpha} - \mu^2\phi^2), \quad (1.101)$$

con el campo conjugado a ϕ

$$\pi(x) = \frac{\partial\mathcal{L}}{\partial\dot{\phi}} = \frac{1}{c^2}\dot{\phi}(x). \quad (1.102)$$

Para la cuantización, el campo real ϕ se convierte en un operador Hermitiano, esto es; $\phi^\dagger = \phi$, que satisface la relación de conmutación de igual-tiempo:

$$\begin{cases} [\phi(\mathbf{x}, t), \dot{\phi}(\mathbf{x}', t)] = i\hbar c^2 \delta(\mathbf{x} - \mathbf{x}'), \\ [\phi(\mathbf{x}, t), \phi(\mathbf{x}', t)] = [\dot{\phi}(\mathbf{x}, t), \dot{\phi}(\mathbf{x}', t)] = 0. \end{cases} \quad (1.103)$$

Identificamos la interacción con las partículas al expandir $\phi(x)$ en un conjunto completo de soluciones de la ecuación de K-G:

$$\phi(x) = \phi^+(x) + \phi^-(x) \quad (1.104)$$

donde

$$\phi^+(x) = \sum_{\mathbf{k}} \left(\frac{\hbar c^2}{2V\omega_{\mathbf{k}}} \right)^{1/2} a(\mathbf{k}) e^{-i\mathbf{k}x} \quad (1.105)$$

y

$$\phi^-(x) = \sum_{\mathbf{k}} \left(\frac{\hbar c^2}{2V\omega_{\mathbf{k}}} \right)^{1/2} a^\dagger(\mathbf{k}) e^{i\mathbf{k}x}. \quad (1.106)$$

Las sumas son sobre los valores del vector de onda \mathbf{k} permitidos por las condiciones de frontera periodicas, sin embargo k^0 y $\omega_{\mathbf{k}}$ están dadas por

$$k^0 = \frac{1}{c} \omega_{\mathbf{k}} = +(\mu^2 + \mathbf{k}^2)^{1/2}, \quad (1.107)$$

esto es el 4-vector de onda de una partícula con masa $m = \mu\hbar/c$, momento $\hbar\mathbf{k}$ y energía

$$E = \hbar\omega_{\mathbf{k}} = + [m^2 c^4 + c^2 (\hbar\mathbf{k})^2]^{1/2}. \quad (1.108)$$

Cada operador $a(\mathbf{k})$ sucede emparejado con su adjunto $a^\dagger(\mathbf{k})$ en la ec (1.105) y (1.106) debido a que ϕ es Hermitiano.

Con las ecs. (1.103)-(1.106) podemos obtener las relaciones de conmutación $a(\mathbf{k})$ y $a^\dagger(\mathbf{k})$

$$\begin{cases} [a(\mathbf{k}), a^\dagger(\mathbf{k}')] = \delta_{\mathbf{k}\mathbf{k}'}, \\ [a(\mathbf{k}), a(\mathbf{k}')] = [a^\dagger(\mathbf{k}), a^\dagger(\mathbf{k}')] = 0. \end{cases} \quad (1.109)$$

Estas relaciones son casualmente las del oscilador armónico, así como todos los resultados que le conciernen, por ejemplo el operador número

$$N(\mathbf{k}) = a^\dagger(\mathbf{k})a(\mathbf{k}) \quad (1.110)$$

donde sus eigenvalores tienen los valores

$$n(\mathbf{k}) = 0, 1, 2, \dots, \quad (1.111)$$

es fácil ver que los operadores $a(\mathbf{k})$ y $a^\dagger(\mathbf{k})$ corresponden a los operadores de creación y aniquilación de partículas con momento $\hbar\mathbf{k}$ y energía $\hbar\omega_{\mathbf{k}}$. El Hamiltoniano y operador momento del campo de K-G son

$$H = \int d^3\mathbf{x} \frac{1}{2} \left[\frac{1}{c^2} \dot{\phi}^2 + (\nabla\phi)^2 + \mu^2\phi^2 \right] \quad (1.112)$$

y

$$\mathbf{P} = - \int d^3\mathbf{x} \frac{1}{c^2} \dot{\phi} \nabla\phi. \quad (1.113)$$

Sustituyendo la expansión de ϕ (ecuaciones (1.104)-(1.106)) en las últimas dos ecuaciones obtenemos

$$H = \sum_{\mathbf{k}} \hbar\omega_{\mathbf{k}} \left(a^\dagger(\mathbf{k})a(\mathbf{k}) + \frac{1}{2} \right), \quad (1.114)$$

$$\mathbf{P} = \sum_{\mathbf{k}} \mathbf{k} \left(a^\dagger(\mathbf{k})a(\mathbf{k}) + \frac{1}{2} \right), \quad (1.115)$$

corroborando nuestra interpretación de $[a^\dagger(\mathbf{k}), a(\mathbf{k})]$ como el operador número para partículas con el vector de onda \mathbf{k} .

De la ecuación (1.114) obtenemos el estado de menor energía, el estado fundamental, del campo de K-G; el estado de vacío $|0\rangle$, donde no hay partículas presentes dados por

$$a(\mathbf{k}) |0\rangle = 0, \quad \forall \mathbf{k}, \quad (1.116)$$

y en términos de los operadores del campo,

$$\phi^+(x) |0\rangle = 0, \quad \forall x. \quad (1.117)$$

Esto nos lleva al resultado de que el vacío tiene energía infinita $\frac{1}{2} \sum_{\mathbf{k}} \hbar\omega_{\mathbf{k}}$. Pero solo las diferencias de energía son observables. Estas constantes infinitas por lo tanto no son de mucha preocupación y se pueden tratar mediante la medición de las energías relativas a el estado de vacío. Es posible evadir estas constantes infinitas ordenando los operadores. En un **producto normal** los operadores de absorción están en la derecha de los de creación en cada producto de operadores. Representando el producto normal con $N(\dots)$, tenemos

$$N(a(\mathbf{k}_1)a(\mathbf{k}_2)a^\dagger(\mathbf{k}_3)) = a^\dagger(\mathbf{k}_3)a(\mathbf{k}_1)a(\mathbf{k}_2), \quad (1.118)$$

y

$$\begin{aligned} N[\phi(x)\phi(y)] &= N[(\phi^+(x) + \phi^-(x))(\phi^+(y) + \phi^-(y))] \\ &= \phi^+(x)\phi^+(y) + \phi^-(y)\phi^+(x) + \phi^-(x)\phi^+(y) + \phi^-(x)\phi^-(y) \end{aligned} \quad (1.119)$$

Al organizar los operadores como se muestra en la ecuación (1.119), se puede concebir como que todos los conmutadores se anulan. El valor esperado del vacío de cualquier producto normal se cancela. Mediante observables definidas como productos normales, con el valor esperado del vacío establecido, las ecs. (1.114) y (1.115) se convierten en

$$P^\alpha = (H/c, \mathbf{P}) = \sum_{\mathbf{k}} \hbar k^\alpha a^\dagger(\mathbf{k})a(\mathbf{k}). \quad (1.120)$$

Con los estados del vacío $|0\rangle$, se construyen los estados con partículas, donde por ejemplo los de una partícula son superposiciones lineales de

$$a^\dagger(\mathbf{k})|0\rangle = |\mathbf{k}\rangle = |1_{\mathbf{k}}\rangle \quad (1 \text{ partícula}). \quad (1.121)$$

El de 2 partículas son superposiciones lineales de

$$a^\dagger(\mathbf{k})a^\dagger(\mathbf{k}')|0\rangle \approx |1_{\mathbf{k}}1_{\mathbf{k}'}\rangle, \quad \forall \mathbf{k} \neq \mathbf{k}'. \quad (1.122)$$

y

$$\frac{1}{\sqrt{2}}a^\dagger(\mathbf{k})a^\dagger(\mathbf{k}') = \frac{1}{\sqrt{2}}[a^\dagger(\mathbf{k})]^2 \Rightarrow \frac{1}{\sqrt{2}}[a^\dagger(\mathbf{k})]^2|0\rangle = |2_{\mathbf{k}}\rangle. \quad (1.123)$$

Claramente si el vacío está normalizado, esto es $\langle 0|0\rangle = 1$ entonces los ejemplos mostrados de 1 y 2 partículas también están normalizados. Para más de dos partículas se sigue el mismo razonamiento.

Otro aspecto destacable es que las partículas del campo de Klein Gordon son bosones; pueden tomar cualquiera de los números de ocupación $n(\mathbf{k}) = 0, 1, 2, \dots$, también son simétricos bajo intercambio de etiquetas ya que los operadores de creación conmutan entre sí.

1.9 DIAGRAMAS DE FEYNMAN

Richard Feynman hizo grandes contribuciones a la QED, una de sus grandes creaciones más ingeniosas fue la introducción de los diagramas que llevan su nombre. Estos representan términos de la expansión de perturbación de la matriz S, vistos de forma directa, pueden interpretarse como diagramas de espacio-tiempo que representan la propagación e interacción entre partículas. Aclarando que hay varias formas de relacionar estas ilustraciones, nosotros tomaremos la siguiente: Imaginemos al tiempo corriendo en

sentido vertical hacia arriba y representando partículas con líneas con una flecha en la dirección del tiempo (Observe 1 para ver algunos casos) dónde las antipartículas se representan con flechas en la dirección opuesta⁶.

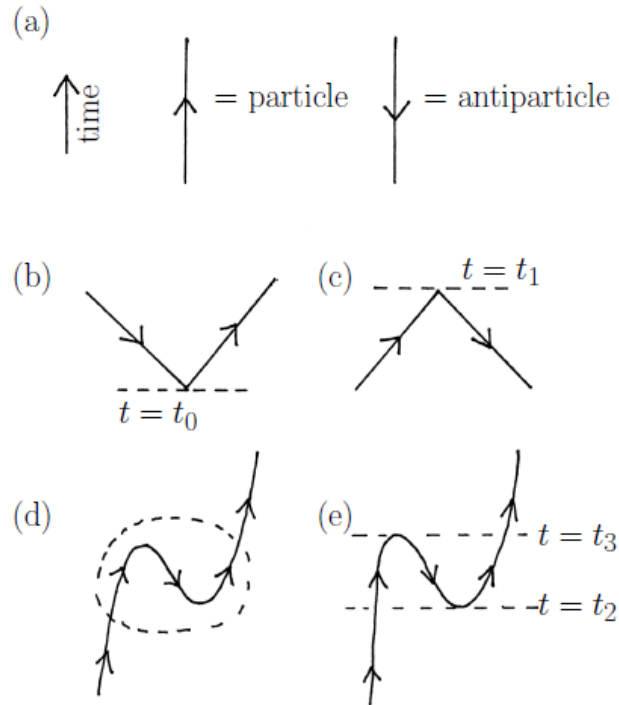


Figura 1: En este diagrama (a) representa una partícula con una línea con una flecha en dirección del tiempo y una antipartícula con la flecha en dirección opuesta. (b) representa la creación de pares en el tiempo $t = t_0$. (c) Es una aniquilación de pares al tiempo $t = t_1$. (d) La trayectoria de una partícula interpretada como: (e) Una creación de pares en t_2 , la antipartícula se aniquila con una tercer partícula en $t = t_3$. Adaptado de [1, p. 175].

Hay un número de modelos de interacciones que podemos probar. Buscamos ilustrar el proceso de interacción, las interpretaciones serán posibles cuando los operadores del campo actúen en estados. Veamos algunos ejemplos:

1. La interacción más simple del Hamiltoniano $\mathcal{H}_1(z)$ involucra un campo escalar $\phi(z)$ que interactúa con una fuente de campo externa $J(z)$ en un punto z del espacio-tiempo:

$$\mathcal{H}_1(z) = J(z)\phi(z). \quad (1.124)$$

Dibujamos una fuente $J(z)$ como una bola en el punto z . El campo sería unas líneas que están atadas a la fuente que es la bola (Observe la fig. 2(a)).

⁶En esta subsección se sigue la presentación estándar de los diagramas de Feynman en teoría cuántica de campos dada en [1, 28, 30].

2. La más simple auto-interacción de los campos es la teoría de ϕ^4 descrita por

$$\mathcal{H}_1(z) = \frac{\lambda}{4!} \phi(z)^4. \quad (1.125)$$

La interacción provoca cuatro campos escalares que se encuentran en un punto z del espacio-tiempo (2(b)).

3. Otra interacción típica es

$$\mathcal{H}_1(z) = g\psi^\dagger(z)\psi(z)\phi(z), \quad (1.126)$$

donde ϕ es el campo escalar y ψ es un campo escalar complejo. Esta interacción da lugar a un campo ψ adjunto (ψ^\dagger), un campo ψ y un campo escalar ϕ que se encuentran en z . Para ver el movimiento de los campos los representamos con una línea y una flecha; el adjunto apunta en dirección opuesta a la del vértice de interacción (ϕ^\dagger), en dónde el campo escalar es simplemente una línea, todos se encuentran en el punto z (2(c)).

Un acercamiento necesario para este tema es introducir la expansión perturbativa de la S -matriz y su representación en términos de los diagramas de Feynman. En resumen mostramos **las reglas de Feynman para la teoría ϕ^4 en el espacio posición** para calcular una amplitud en la expansión de la matriz S , traducimos las ecuaciones a diagramas de la siguiente manera:

- Cada vértice contribuye a un factor $-i\lambda$.
- Cada línea da un factor propagador $\Delta(x - y)$, donde x y y son el punto inicial y final de la línea.
- Líneas externas contribuyen a las entrantes ($-ip \cdot x$) o salientes ($+ip \cdot x$) ondas $e^{\pm ip \cdot x}$.
- Integra la posición de los vértices sobre todo el espacio.
- Buscando obtener el coeficiente correcto frente a el término divide por el factor de simetría D .

Algunas propiedades se muestran en la fig. 3.

Ahora presentamos el conjunto de las **reglas de Feynman para la teoría de ϕ^4 en el espacio de momentos**

- Cada vértice contribuye un factor $-i\lambda$ (fig. 4 (a)).

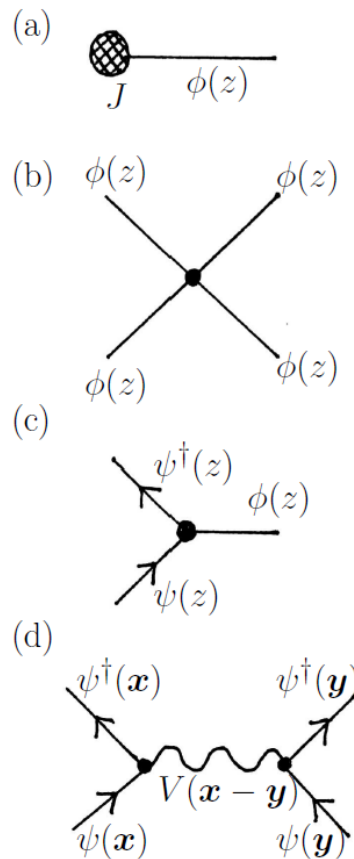


Figura 2: Algunos diagramas espacio-tiempo de las interacciones más comunes. Adaptado de [1, p. 176].

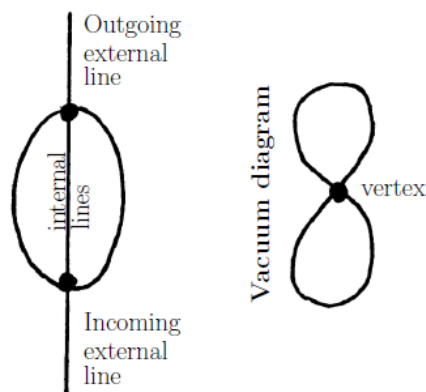


Figura 3: La anatomía de un diagrama de 3er orden desconectado. Adaptado de [1, p. 181].

- Etiqueta cada línea interna con un momento q fluyendo sobre ella y descrita por un propagador $\frac{i}{q^2 - m^2 + i\epsilon}$ 4(b)).

- Fuerza la suma de cada momento llegando a los vértices a ser igual al momento saliendo.
- Integra sobre el momento interno sin restricciones con una medida $\frac{d^4q}{(2\pi)^4}$.
- Líneas externas contribuyen a un factor 1 (fig. 4(c)).
- Divide por el factor de simetría.
- Incluye un total de energía-momento conservando la función delta para cada diagrama.

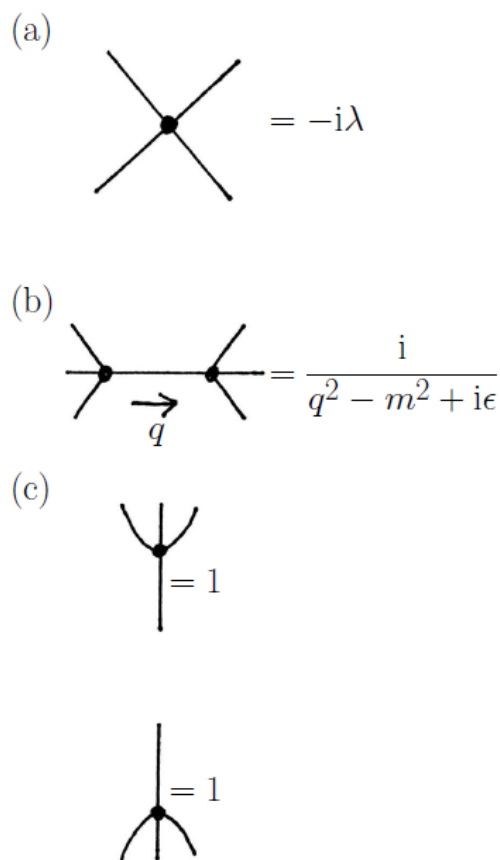


Figura 4: Reglas de Feynman para el espacio de momentos en la teoría ϕ^4 . Adaptado de [1, p. 185].

Es común ver las **reglas de Feynman para QED en el espacio de momentos** ya que suele ser más útil y sencillo de comprender, dicho esto las presentaremos a continuación:

1. Para cada vértice, escribe un factor $ie\gamma^\alpha$.

2. Para cada línea interna de un fotón (con momento k):

$$iD_{F\alpha\beta}(k) = \frac{-ig_{\alpha\beta}}{k^2 + i\epsilon}$$

3. Para cada línea interna de un fermión (momento p):

$$iS_F(p) = \frac{i}{p} - m + i\epsilon$$

4. Para cada línea externa, se colocan los siguientes factores:

a) Electrón inicial : $u_r(p)$

b) Electrón final: $\bar{u}_r(p)$

c) Positrón inicial: $\bar{v}_r(p)$

d) Positrón final: $v_r(p)$

e) Fotón inicial: $\varepsilon_{r\alpha}(k)$

f) Fotón final: $\varepsilon_{r\alpha}^*(k)$

5. Multiplica por los factores de spin (γ -matrices, propagadores, espinores) en secuencia, de derecha a izquierda a lo largo de las líneas de fermiones, congeniando con la dirección de las flechas.

6. Para cada loop de fermiones, ingresa la traza del producto y multiplica por -1 .

7. El 4-momento asociado al vértice debe satisfacer la conservación de momento. Para los casos que esto no se cumpla, integra $(2\pi)^{-4} \int d^4q$ en cada loop cerrado.

8. Multiplica por un factor de fase δ_p igual a $+1(-1)$ ya sea par o impar el intercambio de operadores fermiónicos necesario para obtener el correcto orden normal.

1.10 INTEGRALES DE CAMINO

En esta subsección se sigue la formulación estándar de las integrales de camino presentada en [28–30]. Entraremos en una reformulación de la Teoría Cuántica de Campos (QFT) también propuesta por Richard Feynman, las llamadas integrales de camino. Cantidades importantes en QFT como el propagador, se pueden calcular usando una integral funcional, dónde esta integra todos los posibles caminos de la partícula de un punto a otro en el espacio-tiempo.

La Mecánica Cuántica (QM) se trata de amplitudes de probabilidad, podemos pensar en

una forma de ejemplificar esto en preguntarnos, ¿Cuál es la amplitud de probabilidad de que una partícula que empiece en el punto q_a al tiempo t_a termine en el punto q_b al tiempo t_b ? Este proceso o amplitud del mismo es llamada propagador. La trayectoria debe ser tal que minimice la acción

$$S = \int_{t_a}^{t_b} dt L[q(t)], \quad (1.127)$$

Para ver cómo funciona esto en QM se establece que una partícula que va de A hasta B tomará todas las trayectorias posibles. En otras palabras, la partícula al toma cada camino hacia adelante y atrás en el tiempo, en bucle, etc. Feynman propuso que para obtener dicha amplitud cada trayectoria contribuye con un factor $e^{iS/\hbar}$ que debemos sumar. Este razonamiento de sumar cada factor asociado a cada trayectoria para obtener la amplitud se le llama “**suma sobre historias**” por el hecho de que la onda presente es obtenida mediante la suma de todas las posibles trayectorias pasadas. el propagador de amplitud G a obtener sobre el punto en el espacio-tiempo $A \equiv [(t_a, q_a)]$ a el punto $B \equiv [(t_b, q_b)]$ los escribimos como $\langle q_b | \hat{U}(t_b, t_a) | q_a \rangle$, donde $|q\rangle$ es un eigenestado de posición y $\hat{U}(t_b, t_a)$ es el operador de la evolución en el tiempo $\hat{U}(t_b, t_a) = e^{-i\hat{H}(t_b-t_a)}$ (donde hemos establecido $\hbar = 1$). Tenemos entonces

$$G = \langle q_b | e^{-i\hat{H}(t_b-t_a)} | q_a \rangle. \quad (1.128)$$

Ahora separamos la trayectoria en N pasos infinitesimales de longitud Δt . Esto también se le conoce como “corte-tiempo”. el hecho de que \hat{U} es un operador unitario nos permite hacer este corte de tiempo. Digamos que $\hat{U}(t_b - t_a) = \hat{U}(t_b - t_x) \hat{U}(t_x - t_a)$ y entonces el corte-tiempo $\hat{U}(t) = e^{-i\hat{H}(t_b-t_a)}$ nos lleva a

$$\begin{aligned} G &= \langle q_b | (e^{-i\hat{H}\Delta t})^N | q_a \rangle \\ &= \langle q_b | e^{-i\hat{H}\Delta t} \dots e^{-i\hat{H}\Delta t} \dots e^{-i\hat{H}\Delta t} | q_a \rangle. \end{aligned} \quad (1.129)$$

Es notable ver que cada uno de estos operadores pequeños de evoluciones-tiempo mueven a la partícula por el tiempo Δt entre dos posiciones espaciados estrechamente q_n y q_{n+1} (véase la fig. 5).

Veamos un ejemplo, recordando un poco de QM se suele hacer uso frecuente de un conjunto de estados de completitud. Para los estados de posición q_n tenemos $\int dq_n |q_n\rangle \langle q_n| = 1$. Necesitamos insertar una resolución de identidad para llegar más allá con G entre cada mini operador de evolución-tiempo. Haremos uso de $N - 1$ unidades ampliadas, con esto obtendremos

$$G = \langle q_b | e^{-i\hat{H}\Delta t} \left[\int dq_{N-1} |q_{N-1}\rangle \langle q_{N-1}| \right] e^{-i\hat{H}\Delta t} \dots$$

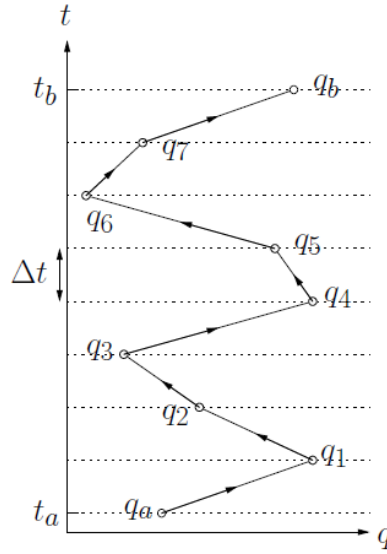


Figura 5: Corte tiempo de una trayectoria de (t_a, q_a) a (t_b, q_b) .
Adaptado de [1, p. 211].

$$\begin{aligned} \dots e^{-i\hat{H}\Delta t} \left[\int dq_{n+1} |q_{n+1}\rangle \langle q_{n+1}| \right] e^{-i\hat{H}\Delta t} \left[\int dq_n |q_n\rangle \langle q_n| \right] e^{-i\hat{H}\Delta t} \dots \\ \dots e^{-i\hat{H}\Delta t} \left[\int dq_1 |q_1\rangle \langle q_1| \right] e^{-i\hat{H}\Delta t} |q_a\rangle. \end{aligned} \quad (1.130)$$

Lo que está sucediendo aquí es que en cada resolución estamos variando los q_n independientemente y haciendo nuevas trayectorias mientras lo hacemos. Reordenando términos, nos queda lo siguiente

$$G = \int dq_q dq_2 \dots dq_{N-1} \langle q_b | e^{-i\hat{H}\Delta t} |q_{N-1}\rangle \dots \langle q_{n+1} | e^{-i\hat{H}\Delta t} |q_n\rangle \dots \langle q_1 | e^{-i\hat{H}\Delta t} |q_a\rangle. \quad (1.131)$$

Entre ambas amplitudes al final tenemos un arreglo de mini-propagadores $G_n = \langle q_{n+1} | e^{i\hat{H}\Delta t} |q_n\rangle$. Observemos un típico mini-propagador:

$$\begin{aligned} G_n &= \langle q_{n+1} | e^{-i\hat{H}\Delta t} |q_n\rangle \\ &= \langle q_{n+1} | e^{-i \left[\frac{\hat{p}^2}{2m} + \hat{V}(q) \right] \Delta t} |q_n\rangle, \end{aligned} \quad (1.132)$$

ahora necesitamos reemplazar los operadores por los eigenvalores. los términos potencia actúan en el estado de posición: $e^{-i\hat{V}(q)\Delta t} |q_n\rangle = |q_n\rangle e^{-iV(q_n)\Delta t}$. Tenemos entonces

$$G_n = \langle q_{n+1} | e^{-i\frac{\hat{p}^2}{2m}\Delta t} |q_n\rangle e^{-iV(q_n)\Delta t}. \quad (1.133)$$

El siguiente paso es analizar como actúa \hat{p} en un estado de posición ya que $|q_n\rangle$ no es su eigenestado ($[\hat{q}, \hat{p}] = i$), entonces vamos a expandir $|q_n\rangle$ en términos de los eigenestados de momento $|q_n\rangle = \int \frac{dp}{(2\pi)} |p\rangle \langle p| q_n\rangle = \int dp |p\rangle e^{-ipq_n} / (2\pi)^{\frac{1}{2}}$. Sustituyendo nos da

$$\begin{aligned} G_n &= \int \frac{dp}{(2\pi)^{\frac{1}{2}}} \langle q_{n+1}| e^{-i\frac{p^2}{2m}\Delta t} |p\rangle e^{-ipq_n} e^{-iV(q_n)\Delta t} \\ &= \int \frac{dp}{(2\pi)^{\frac{1}{2}}} \langle q_{n+1}| |p\rangle e^{-i\frac{p^2}{2m}\Delta t} e^{-ipq_n} e^{-iV(q_n)\Delta t} \\ &= \int \frac{dp}{(2\pi)^{\frac{1}{2}}} e^{ipq_{n+1}} e^{-i\frac{p^2}{2m}\Delta t} e^{-ipq_n} e^{-iV(q_n)\Delta t} \\ &= \int \frac{dp}{(2\pi)^{\frac{1}{2}}} e^{-i\frac{p^2}{2m}\Delta t + ip(q_{n+1} - q_n)} e^{-iV(q_n)\Delta t}. \end{aligned} \quad (1.134)$$

Ya que se han removido todos los operadores es fácil ver que esta es una integral Gaussiana que podemos resolver y nos queda como sigue

$$G_n = \left(\frac{-im}{2\pi\Delta t} \right)^{\frac{1}{2}} e^{\frac{im}{2} \frac{(q_{n+1} - q_n)^2}{\Delta t}} e^{-iV(q_n)\Delta t}. \quad (1.135)$$

Introduciendo este valor en (1.131) y escribiendo $(-im/2\pi\Delta t)^{\frac{1}{2}} = \xi^{-1}$, encontramos que el propagador de amplitud G se puede escribir como un producto de estos mini-propagadores

$$G = \prod_{n=1}^{N-1} \int \frac{dq_n}{\xi} e^{\frac{im}{2} \frac{(q_{n+1} - q_n)^2}{\Delta t}} e^{-iV(q_n)\Delta t}. \quad (1.136)$$

Por último tomamos el límite $N \rightarrow \infty, \Delta t \rightarrow 0$. Esto nos hace convertir nuestras trayectorias irregulares en curvas suaves, resultando $\frac{(q_{n+1} - q_n)^2}{\Delta t}$ en \dot{q}^2 y $\sum_n \Delta t \rightarrow \int dt$.

Así nuestra amplitud total es

$$G = \int \mathcal{D}[q(t)] e^{i \int dt \left[\frac{m\dot{q}^2}{2} - V(q) \right]}, \quad (1.137)$$

Dónde se define la **medida de integración** $\int \mathcal{D}[q(t)]$ como

$$\int \mathcal{D}[q(t)] = \lim_{N \rightarrow \infty} \prod_{n=1}^{N-1} \int \frac{dq_n}{\xi}. \quad (1.138)$$

De manera general podemos dar un resumen a las ideas de las integrales de camino:

- Las integrales de camino de Feynman involucran encontrar amplitudes al agregar contribuciones de todas las trayectorias posibles. Cada una hace una contribución $e^{iS/\hbar}$.
- Un tipo de integral puede ser evaluada y resultar $\int \mathcal{D}[f] e^{-\frac{1}{2} \int f A f + \int b f} = \left(\frac{(2\pi)^N}{\det A} \right)^{\frac{1}{2}} e^{\frac{1}{2} \int b A^{-1} b}$.
- El oscilador armónico simple proporciona un ejemplo mostrando las infinitudes que tenemos que tomar en cuenta en un cálculo típico (en orden de abordar estos temas con mayor profundidad, estas ideas se pueden consultar en [1]).

1.11 ELECTRODINÁMICA CUÁNTICA

En este tema abordaremos algunas construcciones importantes en ciertos procesos que involucran a la QED. Vamos a analizar la **dispersión de electrones de Coulomb**. Calculemos la dispersión de Rutherford de un electrón en un potencial de Coulomb fijo. Para $f \neq i$ obtenemos mediante elementos apropiados de la matriz S

$$S_{fi} = -ie \int d^4x \bar{\psi}_f(x) A(x) \Psi_i(x). \quad (1.139)$$

Donde $e < 0$ es obviamente la carga del electrón, para hacer que esta solución sea posible en práctica tomaremos el camino de la teoría de perturbaciones en el orden más bajo. $\Psi_i(x)$ es aproximado por la onda entrante $\psi_i(x)$ de un electrón con momento p_i y spin s_i :

$$\psi_i(x) = \sqrt{\frac{m_0}{E_i V}} u(p_i, s_i) e^{-ip \cdot x}. \quad (1.140)$$

Donde V es el volumen de normalización, lo que nos dice que la función ψ_i está normalizada en una caja con volumen V . De manera similar tenemos que

$$\psi_f(x) = \sqrt{\frac{m_0}{E_f V}} \bar{u}(p_f, s_f) e^{ip \cdot x}. \quad (1.141)$$

El potencial de Coulomb $A_0(x)$ se genera por un punto cargado estático $-Ze$, entonces

$$A_0(x) = A_0(\mathbf{x}) = -\frac{Ze}{|\mathbf{x}|}; \quad \mathbf{A}(x) = 0. \quad (1.142)$$

Véase la (fig. 6), con estas premisas podemos (1.139) nos queda como

$$S_{fi} = -iZe^2 \frac{1}{V} \sqrt{\frac{m_0}{E_f E_i}} \bar{u}(p_f, s_f) \gamma^0 u(p_i, s_i) \int d^4x e^{i(p_f - p_i) \cdot x} \frac{1}{|\mathbf{x}|}. \quad (1.143)$$

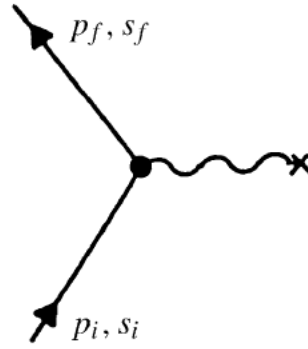


Figura 6: Dispersión de un electrón en un potencial externo (\times) al más bajo orden. Adaptado de [2, p. 78].

La integral puede ser separada

$$\int_{-\infty}^{\infty} dx_0 e^{i(E_f - E_i)x_0} = 2\pi\delta(E_f - E_i). \quad (1.144)$$

Esto quiere decir que la energía es conservada en la dispersión en un potencial independiente del tiempo. Nos queda proseguir con la transformada de Fourier de el potencial de Coulomb (1.142)

$$A_0(\mathbf{q}) = -Ze \int d^3x \frac{1}{|\mathbf{x}|} e^{-i\mathbf{q}\cdot\mathbf{x}},$$

con el momento transferido $\mathbf{q} = \mathbf{p}_f - \mathbf{p}_i$, podemos resolver con el siguiente truco:

$$\begin{aligned} \int d^3x \frac{1}{|\mathbf{x}|} e^{-i\mathbf{q}\cdot\mathbf{x}} &= -\frac{1}{\mathbf{q}^2} \int d^3x \frac{1}{|\mathbf{x}|} \Delta e^{-i\mathbf{q}\cdot\mathbf{x}}, \\ &= -\frac{1}{\mathbf{q}^2} \int d^3x \left(\Delta \frac{1}{|\mathbf{x}|} \right) e^{-i\mathbf{q}\cdot\mathbf{x}} \\ &= -\frac{1}{\mathbf{q}^2} \int d^3x (-4\pi\delta^3(\mathbf{x})) e^{-i\mathbf{q}\cdot\mathbf{x}} = \frac{4\pi}{\mathbf{q}^2}. \end{aligned} \quad (1.145)$$

Entonces el elemento de la matriz S (1.143) queda como

$$S_{fi} = -iZe^2 \frac{1}{V} \sqrt{\frac{m_0}{E_f E_i}} \bar{u}(p_f, s_f) \gamma^0 u(p_i, s_i) \frac{4\pi}{\mathbf{q}^2} 2\pi\delta(E_f - E_i). \quad (1.146)$$

Por último el número de estados finales dN_f dentro del rango del momento d^3p_f está dado por

$$dN_f = V \frac{d^3p_f}{(2\pi)^3}. \quad (1.147)$$

(Para una comprensión más detallada y profunda consulte [2] cap 3.1). Es importante observar los procesos de dispersión porque el cálculo de las contribuciones al a_μ se basa en las dispersiones de diferentes partículas interactuando mediante los campos mediadores por lo que veremos un proceso de suma importancia para nuestro estudio a continuación.

1.11.1 FUNCIÓN DEL VÉRTICE DEL ELECTRÓN

Hemos visto que podemos calcular dispersión de procesos de partículas. Ahora veremos un breve cálculo de correcciones a la dispersión por la presencia de fotones virtuales.

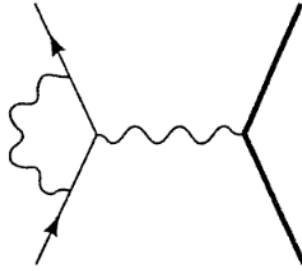


Figura 7: Dispersión de electrones con un fotón virtual (loop) a primer orden.

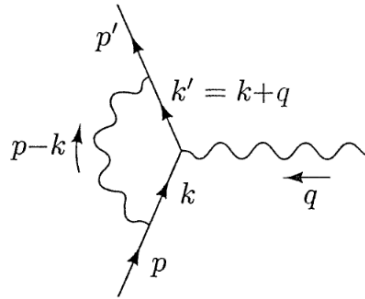


Figura 8: Diagrama de dispersión de electrones con un fotón virtual (loop) con momento $p - k$, fermiones entrantes $p, k, p', k', k' = k + q$ y fotón saliente q .

Este procedimiento (8) con un loop y generalmente cualquier proceso conteniendo un loop o más suele ser bastante extenso, por lo que haremos algunas simplificaciones para llegar al acercamiento necesario que necesitamos. Aplicando las reglas de Feynman tenemos que, al orden α , $\Gamma^\mu = \gamma^\mu + \delta\Gamma^\mu$, donde

$$\delta\Gamma^\mu(p', p) = \int \frac{d^4k}{(2\pi)^4} \frac{-ig_{\nu\rho}}{(k-p)^2 + i\epsilon} \bar{u}(p')(-ie\gamma^\nu) \frac{i(k'+m)}{k'^2 - m^2 + i\epsilon} \gamma^\mu \frac{i(k+m)}{k^2 - m^2 + i\epsilon} (-ie\gamma^\rho) u(p)$$

$$= 2ie^2 \int \frac{d^4 k}{(2\pi)^4} \frac{\bar{u}(p') [k' \gamma^\mu k' + m^2 \gamma^\mu - 2m(k + k')^\mu] u(p)}{((k - p)^2 + i\epsilon)(k'^2 - m^2 + i\epsilon)(k^2 - m^2 + i\epsilon)}. \quad (1.148)$$

Utilizando la técnica de parámetros de Feynman para el denominador (revisar capítulo 6.3 de [31])

$$\frac{1}{((k - p)^2 + i\epsilon)(k'^2 - m^2 + i\epsilon)(k^2 - m^2 + i\epsilon)} = \int_0^1 dx dy dz \delta(x + y + z - 1) \frac{2}{D^3}, \quad (1.149)$$

donde ahora D es

$$\begin{aligned} D &= x(k^2 - m^2) + y(k'^2 - m^2) + z(k - p)^2 + (x + y + z)i\epsilon \\ &= k^2 + 2k(yq - zp) + yq^2 + zp^2 - (x + y)m^2 + i\epsilon. \end{aligned} \quad (1.150)$$

Usando

$$\ell \equiv k + yq - zp.$$

Con algo de álgebra podemos simplificar a

$$D = \ell^2 - \Delta + i\epsilon,$$

donde

$$\Delta = -xyq^2 + (1 - z)^2 m^2. \quad (1.151)$$

Hemos trabajado el denominador, ahora debemos expresar el numerador en términos de ℓ . Vemos que D solo depende de ℓ ,

$$\int \frac{d^4 \ell}{(2\pi)^4} \frac{\ell^\mu}{D^3} = 0; \quad (1.152)$$

$$\int \frac{d^4 \ell}{(2\pi)^4} \frac{\ell^\mu \ell^\nu}{D^3} = \int \frac{d^4 \ell}{(2\pi)^4} \frac{\frac{1}{4} g^{\mu\nu} \ell^2}{D^3}. \quad (1.153)$$

Ahora reescribimos el numerador. Queremos agrupar todo en dos términos, proporcional a γ y $i\sigma^{\mu\nu} q_\nu$. Lo más adecuado es buscar una expresión como

$$\gamma^\mu \cdot A + (p'^\mu + p^\mu) \cdot B + q^\mu \cdot C,$$

haciendo un poco de tediosa álgebra de Dirac y usando la identidad de Gordon para eliminar $(p' + p)$. Nuestra expresión queda como sigue

$$\begin{aligned} \bar{u}(p')\delta\Gamma^\mu(p', p)u(p) &= 2ie^2 \int \frac{d^4\ell}{(2\pi)^4} \int_0^1 dx dy dz \delta(x + y + z - 1) \frac{2}{D^3} \\ &\times \bar{u}(p') \left[\gamma^\mu \cdot \left(-\frac{1}{2}\ell^2 + (1-x)(1-y)q^2 + (1-4z+z^2)m^2 \right) \right. \\ &\left. + \frac{i\sigma^{\mu\nu}q_\nu}{2m} (2m^2z(1-z)) \right] u(p), \end{aligned} \quad (1.154)$$

donde $q^2(0) < 0$ para un proceso de dispersión, entonces $\Delta > 0$.

Ahora usaremos la técnica de rotación de Wick para las integrales. Así usamos las siguientes relaciones

$$\int \frac{d^4\ell}{(2\pi)^4} \frac{1}{[\ell^2 - \Delta]^m} = \frac{i(-1)^m}{(4\pi)^2} \frac{1}{(m-1)(m-2)} \frac{1}{\Delta^{m-2}} \quad (1.155)$$

$$\begin{aligned} \int \frac{d^4\ell}{(2\pi)^4} \left(\frac{\ell^2}{[\ell^2 - \Delta]^3} - \frac{\ell^2}{[\ell^2 - \Delta_\Lambda]^3} \right) &= \frac{i}{(4\pi)^2} \int_0^\infty d\ell_E^2 \left(\frac{\ell_E^4}{[\ell_E^2 + \Delta]^3} - \frac{\ell_E^4}{[\ell_E^2 + \Delta_\Lambda]^3} \right) \\ &= \frac{i}{(4\pi)^2} \log \left(\frac{\Delta_\Lambda}{\Delta} \right). \end{aligned} \quad (1.156)$$

Nuestra expresión explícita para la corrección del vértice a un loop queda como

$$\begin{aligned} \bar{u}(p')\delta\Gamma^\mu(p', p)u(p) &= \frac{\alpha}{2\pi} \int_0^1 dx dy dz \delta(x + y + z - 1) \\ &\times \bar{u}(p') \left(\gamma^\mu \left[\log \frac{z\Lambda^2}{\Delta} + \frac{1}{\Delta} \left((1-x)(1-y)q^2 + (1-4z+z^2)m^2 \right) \right] \right. \\ &\left. + \frac{i\sigma^{\mu\nu}q_\nu}{2m} \left[\frac{1}{\Delta} 2m^2z(1-z) \right] \right) u(p). \end{aligned} \quad (1.157)$$

Se observa en la ec. (1.157) que el vértice puede ser factorizado como sigue

$$\Gamma^\mu(p', p) = \gamma F_1(q^2) + \frac{i\sigma^{\mu\nu}q_\nu}{2m} F_2(q^2), \quad (1.158)$$

donde las funciones de q^2 son llamadas factores de forma. al más bajo orden (tree level) $F_1 = 1$ y $F_2 = 0$ pero como estamos al siguiente orden (one loop) vemos que podemos relacionar estos resultados. Antes de hacer esto recordemos que mediante la matriz S nos y trabajando algunas representaciones de los espinores podemos llegar a

$$i\mathcal{M} = -i(2m) \cdot e \xi^{\prime\dagger} \left(-\frac{1}{2m} \sigma^k [F_1(0) + F_2(0)] \right) \xi \tilde{B}^k(\mathbf{q}),$$

donde

$$\tilde{B}^k(\mathbf{q}) = -i\epsilon^{ijk} q^i \tilde{A}^j(\mathbf{q})$$

es la transformada de Fourier del campo magnético producido por $\mathbf{A}(x)$. Podemos interpretar \mathcal{M} como la aproximación de Born a la dispersión de un potencial. Siendo el potencial de la interacción del momento magnético,

$$V(x) = -\langle \mu \rangle \cdot \mathbf{B}(x),$$

donde

$$\langle \mu \rangle = \frac{e}{m} \left[F_1(0) + F_2(0) \right] \xi^{\prime\dagger} \frac{\sigma}{2} \xi.$$

La expresión del momento magnético también se puede reescribir como

$$\mu = g \left(\frac{e}{2m} \right) \mathbf{S},$$

donde \mathbf{S} es el spin. Ahora solo nos queda calcular el factor de forma $F_2(0)$ para obtener la anomalía del electrón

$$\begin{aligned} F_2(q^2 = 0) &= \frac{\alpha}{2\pi} \int_0^1 dx dy dz \delta(x + y + z - 1) \frac{2m^2 z(1-z)}{m^2(1-z)^2} \\ &= \frac{\alpha}{\pi} \int_0^1 dz \int_0^{1-z} dy \frac{z}{1-z} = \frac{\alpha}{2\pi}. \end{aligned} \tag{1.159}$$

Entonces, obtenemos la corrección al factor g del electrón:

$$a_e \equiv \frac{g - 2}{2} = \frac{\alpha}{2\pi} \approx 0.0011614. \quad (1.160)$$

Este resultado fue obtenido por Schwinger en 1948. El cuál fue inequívocamente cercano al experimental $a_e = 0.0011597$, a órdenes superiores en α , es más cercano aún. Más adelante se usará un proceso análogo para identificar las contribuciones de SUSY pero en el muon, por eso la importancia de identificar las piezas que conforman este proceso de dispersión y el cálculo.

TEORÍA DE SUPERSIMETRÍA

En este capítulo se sigue principalmente la exposición general de la supersimetría y del modelo estándar mínimo supersimétrico presentada en [3, 32, 33]. Vamos a dar una breve introducción a conocimientos básicos de supersimetría, primero vamos a hablar sobre el panorama general e histórico en el que se encuentra el ME así como diversos problemas que enfrenta con el campo del Higgs con una definición sencilla de sus operadores y sus características. Después pasaremos a hablar de qué es una super-álgebra partiendo del grupo de Poincaré. Después Tocaremos el tema de los superespacios, para darle una interpretación geométrica a SUSY ya que estos viven en otro tipo de variedad diferente al espacio 4-dimensional de Minkowski que usamos en QFT. Siguiendo esta línea podremos empezar con los supercampos escalares y vectoriales como funciones de las coordenadas del superespacio, la importancia de la derivada covariante y algunas restricciones para seguir trabajando en una teoría gauge. Después con todas estas bases formularemos la densidad lagrangiana supersimétrica, el superpotencial y finalmente el lagrangiano general supersimétrico. Hasta ese punto notaremos que la construcción del ME y de SUSY comparten la misma base: ambas se formulan a partir de simetrías de gauge, siguiendo una estructura y requerimientos análogos. Con todo lo anterior, seremos capaces de llegar a un ME construido bajo supersimetría, el Modelo Estándar Mínimo Supersimétrico (MEMS) dando un listado de las nuevas partículas que se añadirán para este Modelo, y a su vez garantizar que podemos reproducir una teoría que sea consistente y pueda derivar, a bajas energías, en el ME. Hablaremos del cómo se rompe la super-simetría en de manera suave del MEMS a través de introducir ciertos términos en el lagrangiano de la teoría, porque al igual que el ME, esto es necesario para generar propiedades en las partículas. Debido a que la extensión supersimétrica más general del Modelo Estándar permite términos en el superpotencial que pueden violar el número bariónico y leptónico, podrían aparecer procesos no observados experimentalmente, como el decaimiento rápido del protón. Para evitar estas interacciones y restringir la teoría al sector fenomenológicamente aceptable, se impone una simetría discreta llamada paridad-R. Este modelo también contiene su parte electrodébil, de tal forma que la ruptura de simetría electrodébil radiativa pueda reprodu-

cir al rompimiento espontáneo de simetría EW en el ME. Casi llegando a las partes finales de los requerimiento para nuestra contribución a la anomalía del muon se explicarán los campos asociados a las partículas supersimétricas asociadas a la teoría EW: los electro-weakinos, comprendidas por el photino, zino y wino, que son los estados supersimetricos compañeros fermiónicos del fotón, y los bosones de gauge Z^0 y W^\pm , provenientes de la parte electrodébil una vez que se rompe la simetría EW. Los neutrinos y charginos vienen una posible mezcla entre estos y los grados fermiónicos supercompañeros de los campos de Higgs. Como tienen los mismos números cuánticos, estos se pueden mezclar, obteniendo los neutralinos como mezcla de cuatro grados de libertad neutros y los charginos como mezcla de 2 grados de libertad cargados. Con esto será posible hablar del $g - 2$ en términos suaves de supersimetría, la interacción que tomará lugar a esta contribución será la de neutralino-lepton-slepton para poder calcularla a un loop y obtener el a_μ^{SUSY} , de forma similar a la establecida en la sección 1.11.1, nuevamente con su amplitud invariante y sus factores de forma a priori agregando unas reglas de Feynman nuevas para estas interacciones y supercampos.

El trabajo hecho hasta ahora será crucial ya que tomará el rol de datos objetivo o target en nuestro modelo de redes neuronales, al que se intentará optimizar basado en los parámetros libres como inputs o datos de entrada.

2.1 EL MODELO ESTÁNDAR

Gran parte de los desarrollos en la física de partículas elementales han sido concebidos a partir del estudio de las simetrías presentes en dichos sistemas. Alrededor de 1974 se propuso una nueva teoría, distinta de las previamente establecidas, llamada "supersimetría"(SUSY). Algo destacable es que los fermiones y bosones se pueden obtener uno a otro mediante un operador, es decir; con un operador que representa un bosón se puede obtener un fermión, dentro de los supercampos. Si la Supersimetría se conserva implicaría que las partículas del ME tendrían unas partículas supercompañeras asociadas con la misma masa. Aunque en la naturaleza sabemos que no hay fermiones y bosones con la misma masa, surge la cuestión de cuáles son los motivos para estudiarla. La teoría electrodébil constituye un ejemplo de cómo ciertas simetrías fundamentales pueden permanecer ocultas a bajas energías debido a su ruptura espontánea, mientras que a escalas de energía más altas dichas simetrías se restauran y se manifiestan de manera explícita. Es claro que este modelo debe ser extendido para describir física a altas energías, en particular en el régimen asociado a la escala de Planck $M_p = (8_{Newton})^{-1/2} = 2.4 \times 10^{18} GeV$, donde los efectos cuánticos gravitacionales comienzan a manifestarse. Al considerar el cociente M_p/M_W , donde M_W es la masa del bosón débil(Z o W) surge el problema de

la jerarquía: las correcciones cuánticas tienden a hacer que la masa del Higgs sea muy sensible a física a escalas altas de energía, lo que implica una fuerte necesidad de ajuste fino.

Tomando solo la parte electricamente neutra del doblete de Higgs como H , el potencial para este campo tomará la forma

$$V = m_H^2 |H|^2 + \lambda |H|^4. \quad (2.1)$$

El ME necesita de un valor esperado en el vacío (VEV) distinto de cero para el campo escalar de Higgs neutro, H en el mínimo del potencial. Sabemos experimentalmente que el $\langle H \rangle \approx 125 GeV$. Tomando el ME como correcto en una teoría de campos efectiva nos conduce a que m_H^2 recibe enormes correcciones cuánticas de las partículas virtuales de cualquier interacción que se acople directa o indirectamente, al campo de Higgs.

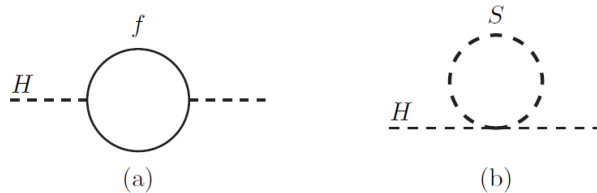


Figura 9: Correcciones a 1-loop para la masa cuadrada del Higgs m_h^2 , debido a un fermión de Dirac (a) f , y (b) un escalar S (Imagen obtenida de [3]).

Para ilustrar estas ideas veamos la fig. 9, para el diagrama (a) el campo de Higgs se acopla a un fermión de masa m_f con un término en el Lagrangiano de $-\lambda_f H \bar{f} f$, esto nos conduce a una corrección

$$\Delta m_H^2 = \frac{|\lambda_f^2|}{8\pi^2} \Lambda_{UV}^2 + \dots \quad (2.2)$$

El *cutoff* ultravioleta que regula el loop integral es Λ_{UV} . Los puntos suspensivos representan términos subdominantes proporcionales a m_f^2 con dependencia logarítmica en el *cutoff*. La contribución dominante, sin embargo, crece cuadráticamente con el *cutoff*, $\propto \Lambda_{UV}^2$. Las correcciones más grandes se presentan cuando f es un quark top con $\lambda_f \approx 0.94$. Esto representa un problema si Λ_{UV} es del orden de M_P , entonces esta corrección a m_H^2 es 30 órdenes de magnitud mayores a los requeridos de $m_H^2 \approx (92.9 GeV)^2$ a nivel árbol (sin correcciones radiativas), tales implicaciones conllevan a ver afectado en su totalidad al espectro de masas del ME directa o indirectamente por la sensibilidad al *cutoff* Λ_{UV} .

Podemos suponer una partícula escalar con masa m_s , como vemos en la el diagrama de la fig. 9 (b), que se acopla al Higgs mediante la interacción $\mathcal{L}_I = -\lambda_s |H|^2 |S|^2$, nos daría una corrección

$$\Delta m_H^2 = \frac{\lambda_s}{16\pi^2} [\Lambda_{UV}^2 - 2m_s^2 \ln(\Lambda_{UV}/m_s) + \dots]. \quad (2.3)$$

Si usamos regularización dimensional en el loop, no habría divergencia UV. Sin embargo, queda un término m_s^2 que no es posible eliminar sin un injustificado ajuste de un contratérmino específico. Como solución a este problema se propuso la idea de que el Higgs no es fundamental, si no el resultado de un fenómeno colectivo, esto con el fin de dar una solución alternativa, cuyo corte ultravioleta estaría en la escala de Planck, sin embargo esta idea no prosperó.

Para evitar el problema de correcciones enormes, notamos que se requiere una cancelación sistemática del tipo

$$\Delta m_H^2 = -\frac{|\lambda_f|^2}{8\pi^2} \Lambda_{UV}^2 + \dots, \quad (2.4)$$

$$\Delta m_H^2 = +\frac{\lambda_s}{16\pi^2} [\Lambda_{UV}^2 - 2m_s^2 \ln(\Lambda_{UV}/m_s) + \dots]. \quad (2.5)$$

Si cada quark y leptón del ME está acompañado por dos escalares complejos entonces la contribución UV se cancela. Esto sucede si asumimos que existe una simetría entre fermiones y bosones (más información en [34]).

2.2 SUPERSIMETRÍA - UN CORTO RESUMEN

Como se mencionó anteriormente, la supersimetría es una propuesta para abordar diversos problemas presentes en los modelos actuales. Esta teoría postula que a cada partícula le corresponde un supercompañero cuya estadística difiere en un valor de espín de 1/2. En este marco, las transformaciones supersimétricas permiten relacionar estados bosónicos con estados fermiónicos. Sea Q un operador de spin que transforma un estado bosónico a uno fermiónico y viceversa

$$Q |Boson\rangle = |Fermion\rangle, \quad Q |Fermion\rangle = |Boson\rangle. \quad (2.6)$$

Para los fermiones sus compañeros supersimétricos son llamados *sfermiones* porque son escalares y se reduce en un medio el spin($1/2 - 1/2 = 0$), para los bosones son llamados

bosinos. La regla es añadir el prefijo “s” para supercompañeros fermiónicos y el sufijo “ino” para supercompañeros bosónicos¹

2.3 SUPER ÁLGEBRA DE POINCARÉ

El grupo de Poincaré es el grupo de simetrías del espacio-tiempo de Minkowski propio de la relatividad espacial; incluye traslaciones y transformaciones de Lorentz (rotaciones y boosts), debido a esto funciona muy bien para cualquier teoría relativista de campos por su simetría completa (de Poincaré), es decir; sus ecuaciones son invariantes bajo todas esas transformaciones. Afortunadamente el ME puede ser descrito como una representación irreducible del grupo de Poincaré [38]. Pero esto no basta, en nuestro caso es necesario extender estas simetrías que, para este contexto, significa a ampliar el conjunto de transformaciones permitidas añadiendo nuevos generadores que sean compatibles con la estructura relativista. Posteriormente a [38] se hizo una revisión y se llegó a un teorema excluyente de Coleman-Mandula [36] al extender el concepto de álgebras de Lie, el cual es la estructura algebraica que describe las simetrías continuas a nivel infinitesimal: están formadas por generadores y una operación de conmutador $[\cdot, \cdot]$ que codifica cómo se combinan esas simetrías. En [32] mencionan que “Todos los generadores de supersimetría deben ser fermiónicos, i.e, ellos deben cambiar el spin por una mitad semi-impar y cambiar la estadística del estado”.

Haciendo uso de este álgebra podemos formalizar la supersimetría extendiendo el álgebra de Poincaré a lo que se conoce como *Súper-Álgebra de Poincaré*. Una forma de proporcionar la extensión es agregar al grupo de Poincaré un espinor de Majorana con las componentes Q_a , donde $a = 1, \dots, 4$, a estos operadores se les llama *cargas de espinor*. La forma que usaremos nosotros es en dos componentes de los espinores de Weyl Q_A y \bar{Q}_A . La relación entre estas dos extensiones es $Q_a = \begin{pmatrix} Q_A \\ \bar{Q}_A \end{pmatrix}$. El álgebra resultante en términos de los espinores de Weyl, es la siguiente

$$\{Q_A, Q_B\} = \{\bar{Q}_A, \bar{Q}_B\} = 0, \quad (2.7)$$

$$\{Q_A, \bar{Q}_B\} = 2\sigma_{AB}^\mu P_\mu, \quad (2.8)$$

$$[Q_A, P_\mu] = [\bar{Q}_A, P_\mu] = 0, \quad (2.9)$$

$$[Q_A, M^{\mu\nu}] = \sigma_A^{\mu\nu B} Q_B, \quad (2.10)$$

¹La información que se revisará sobre supersimetría particularmente en las secciones 2.2-2.13 es obtenida y puede consultarse con más profundidad en [8, 33, 35–37].

Donde $\sigma^{\mu\nu}$ está relacionado con las matrices de Pauli, $M^{\mu\nu}$ representan los generadores del grupo de Lorentz. Los índices con punto y sin punto van de $A = 1, 2$ y $\dot{A} = \dot{1}, \dot{2}$, respectivamente. El punto sobre los índices es para notar que el operador se transforma de acuerdo a su representación conjugada. La representación del espinor de Weyl izquierdo es $(\frac{1}{2}, 0)$, con spin $\frac{1}{2}$ y de dimensión dos. De manera análoga, el punto corresponde a la representación del espinor de Weyl derecho $(0, \frac{1}{2})$. Ambas son representaciones del grupo $SL(2, \mathbb{C})$. Por otro lado, P_μ denota los generadores de las traslaciones, mientras que M_μ representa los generadores de las rotaciones y los boost del álgebra de Poincaré. Las ecuaciones (2.7) a (2.10) no incluye ninguna simetría interna de grupos. Podemos asumir generadores adicionales, como cargas de espinor $Q_A^\alpha = 1, \dots, N$. El álgebra correspondiente a $N = 1$ es llamada *álgebra de supersimetría* mientras que las mayores a 1 son álgebras de supersimetría extendidas. Tomaremos el caso de $N = 1$ por ser la extensión mínima del álgebra de Poincaré (supersimetría mínima).

2.4 SUPERESPACIOS

La supersimetría puede tener una interpretación geométrica usando superespacios, estos dividen a las coordenadas en dos tipos y las combina: las usuales de espacio tiempo x^μ y adicionando coordenadas θ^μ de dimensiones que anti-conmutan. Estas supercoordenadas permiten que los supercampos (funciones definidas en el superespacio) puedan expandirse en un número finito de términos (debido a la propiedad de que las variables fermiónicas al cuadrado son cero). En pocas palabras el superespacio es una variedad que se obtiene agregando 4 coordenadas fermiónicas a las de los bosones, parametrizan el espacio-tiempo usual t, x, y, z ; los campos asociados son bosónicos. Los campos del ME viven en el 4-dimensional espacio de Minkowski, mientras que los supercampos viven en la variedad 8-dimensional llamada *superespacio*,

$$z^\pi = (x^\mu, \theta^A, \bar{\theta}^{\dot{A}}), \quad (2.11)$$

donde x^μ corresponde a las coordenadas bosónicas del espacio de Minkowski, mientras que θ^A y $\bar{\theta}^{\dot{A}}$ son coordenadas fermiónicas (números de Grassmann). En particular, θ^A se toma como un espinor de Weyl complejo (izquierdo) y $\bar{\theta}^{\dot{A}}$ denota su conjugado complejo, es decir $\bar{\theta}^{\dot{A}} \equiv (\theta^A)^*$; el punto en el índice indica que transforma en la representación conjugada. Sus relaciones de anticonmutación están dadas por

$$\{\theta_A, \theta_B\} = 0, \quad \{\bar{\theta}_{\dot{A}}, \bar{\theta}_{\dot{B}}\} = 0, \quad \{\theta_A, \bar{\theta}_{\dot{B}}\} = 0. \quad (2.12)$$

La naturaleza de los *números* (o variables) de Grassmann es muy distinta de la de los números ordinarios: son generadores de un *álgebra de Grassmann* (o *álgebra exterior*) y, por definición, *anti-conmutan*. En particular, a partir de (2.12) se cumple que $\theta_A\theta_B = -\theta_B\theta_A$ y $\theta_A^2 = 0$. Por ello, al considerar (2.12), cualquier serie de potencias de una función $f(\theta_A)$ se trunca a un orden finito; para una sola variable, la expansión general queda como

$$f(\theta_A) = a + b\theta_A. \quad (2.13)$$

2.5 SUPERCAMPOS

Hasta ahora hemos visto como un supercampo Φ es una función de las coordenadas del superespacio. Ahora vamos a expandir tal supercampo alrededor de las variables que anticonmutan usando las ecuaciones (2.12) y (2.13), generalmente llegamos a

$$\begin{aligned} \Phi(x, \theta, \bar{\theta}) = & f(x) + \theta^A \phi_A + \bar{\theta}_{\dot{A}} \bar{\chi}^{\dot{A}}(x) + \theta\theta m(x) + \bar{\theta}\bar{\theta} n(x) \\ & + \theta\sigma^\mu \bar{\theta} V_\mu(x) + \theta\theta \bar{\theta}_{\dot{A}} \bar{\lambda}^{\dot{A}} + \bar{\theta}\bar{\theta} \theta^A \psi_A(x) + \theta\theta \bar{\theta}\bar{\theta} d(x). \end{aligned} \quad (2.14)$$

Las componentes están listados en la Tabla 1. La ecuación (2.14) es muy general, para nuestro caso, en un supercampo, tendremos que imponer algunas restricciones, definiendo así el *supercampo escalar* y *supercampo vectorial*, los cuales ya funcionan como representaciones irreducibles de la súper álgebra de Poincaré.

Campos Componentes	Tipo	Grados de libertad
$f(x), m(x), n(x)$	(Pseudo) escalar complejo	2
$\psi_A(x), \phi_A(x)$	Espinor izquierdo de Weyl	4
$\bar{\chi}^{\dot{A}}(x), \bar{\lambda}^{\dot{A}}(x)$	Espinor derecho de Weyl	4
$V_\mu(x)$	4-vector de Lorentz	8
$d(x)$	Escalar complejo	2

Tabla 1: Componentes del supercampo.[8]

2.6 SUPERCAMPOS ESCALARES Y VECTORIALES

Para imponer las restricciones asociadas que determinan un supercampo escalar, necesitamos respetar las simetrías gauge. Como se discute en [39], esto requiere de una conexión en un fibrado principal $\pi : P \rightarrow M$ con grupo de estructura G , es decir, un espacio total P que localmente se ve como $M \times G$ y sobre el cual actúa G (a la derecha). Esta estructura permite comparar elementos de fibras sobre puntos cercanos de la base

M , donde M representa el espacio-tiempo y cada fibra $\pi^{-1}(x)$ es isomorfa a G (el grupo de simetría interna). La conexión induce, en los fibrados asociados, a lo que llamamos *derivada covariante*, con esta garantizaremos la invariancia gauge. Las derivadas covariantes que conmutan con una transformación supersimétrica son

$$D_A \equiv \partial_A + i(\sigma^\mu \bar{\theta})_A \partial_\mu, \quad \bar{D}_{\dot{A}} \equiv -\partial_{\dot{A}} - i(\theta \sigma^\mu)_{\dot{A}} \partial_\mu. \quad (2.15)$$

Con esto, los supercampos escalares izquierdo y derecho se definen imponiendo las restricciones

$$\bar{D}_{\dot{A}} \Phi(x, \theta, \bar{\theta}) = 0, \quad (2.16)$$

$$D_A \Phi^\dagger(x, \theta, \bar{\theta}) = 0, \quad (2.17)$$

donde las ecuaciones (2.16) y (2.17) corresponden a las restricciones para el supercampo escalar izquierdo y derecho, respectivamente. Vamos a derivar el supercampo escalar izquierdo usando la ecuación (2.16). Con un cambio de variable $y^\mu = x^\mu + i\theta\sigma^\mu\bar{\theta}$ a la segunda ecuación de (2.15) entonces la derivada covariante supersimétrica queda de la siguiente forma

$$\begin{aligned} y^\mu &= x^\mu + i\theta\sigma^\mu\bar{\theta}, \\ \partial_{\dot{A}} &= \frac{\partial}{\partial\bar{\theta}^{\dot{A}}} \Big|_x = \frac{\partial}{\partial\bar{\theta}^{\dot{A}}} \Big|_y + \frac{\partial y^\mu}{\partial\bar{\theta}^{\dot{A}}} \frac{\partial}{\partial y^\mu}, \\ \text{pero } \frac{\partial y^\mu}{\partial\bar{\theta}^{\dot{A}}} &= i\theta\sigma^\mu, \\ \Rightarrow \bar{D}_{\dot{A}} &\equiv -\partial_{\dot{A}} \Big|_y - i(\theta\sigma^\mu)_{\dot{A}} \frac{\partial}{\partial y^\mu} + i(\theta\sigma^\mu)_{\dot{A}} \frac{\partial}{\partial y^\mu} = -\partial_{\dot{A}} \Big|_y, \\ \partial_\mu &= \frac{\partial}{\partial y^\mu} \Rightarrow \bar{D}_{\dot{A}} \equiv -\partial_{\dot{A}} \Big|_y = 0, \end{aligned}$$

ahora sí, derivando el campo escalar:

$$\bar{D}_{\dot{A}} \Phi(y, \theta) = -\partial_{\dot{A}} \Phi(y, \theta) = -\frac{\partial}{\partial\bar{\theta}^{\dot{A}}} \Phi(y, \theta) = 0. \quad (2.18)$$

Debe notarse que el resultado nulo en (2.18) no significa que $\bar{D}_{\dot{A}}$ sea idénticamente cero, sino que corresponde a la restricción de quiralidad impuesta en (2.16). En coordenadas quirales,

$$\bar{D}_{\dot{A}} = -\frac{\partial}{\partial\bar{\theta}^{\dot{A}}} \Big|_y,$$

de modo que $\bar{D}_{\dot{A}}\Phi(y, \theta) = 0$ implica que el supercampo Φ no depende explícitamente de $\bar{\theta}^{\dot{A}}$.

En concordancia con lo anterior el campo escalar lo escribimos como

$$\Phi(y, \theta) = A(y) + \sqrt{2}\theta\psi(y) + \theta\theta F(y), \quad (2.19)$$

donde $A(y)$ y $F(y)$ son escalares complejos y $\psi(y)$ es el espinor de Weyl izquierdo. para el supercampo escalar derecho es un proceso similar pero con signo opuesto.

Un *supercampo vectorial* está definido por la restricción de que su conjugado es el mismo supercampo escalar

$$\Phi^\dagger(x, \theta, \bar{\theta}) = \Phi(x, \theta, \bar{\theta}). \quad (2.20)$$

Usando esta condición llegamos a la forma general del supercampo vectorial

$$\begin{aligned} \Phi(x, \theta, \bar{\theta}) = & f(x) + \theta\varphi(x) + \bar{\theta}\bar{\varphi}(x) + \theta\theta m(x) + \bar{\theta}\bar{\theta} m^*(x) \\ & + \theta\sigma^\mu\bar{\theta}V_\mu(x) + \theta\theta\bar{\theta}\bar{\lambda}(x) + \bar{\theta}\bar{\theta}\theta\lambda(x) + \theta\theta\bar{\theta}\bar{\theta}d(x). \end{aligned} \quad (2.21)$$

En el Modelo Estándar, las condiciones gauge resultantes al pedir invariancia bajo las transformaciones gauge, reducen el número de grados de libertad. De la misma manera, introducimos una generalización supersimétrica de dicha transformación, llamada *transformación supergauge*. Implementamos el *gauge de Wess-Zumino*. Entonces el supercampo vectorial queda de la siguiente manera

$$\begin{aligned} V_{WZ}(x, \theta, \bar{\theta}) = & (\theta\sigma^\mu\bar{\theta}) [V_\mu(x) + i\partial_\mu(A(x) - A^*(x))] \\ & + \theta\theta\bar{\theta}\bar{\lambda}(x) + \bar{\theta}\bar{\theta}\theta\lambda(x) + \theta\theta\bar{\theta}\bar{\theta}d(x), \end{aligned} \quad (2.22)$$

donde $A(x) + A^*(x) = f(x)$ representa una restricción, tal que $[V_\mu(x) + i\partial_\mu(A(x) - A^*(x))]$ nos proporciona tres grados bosónicos de libertad. Asimismo, los espinores de Weyl $\bar{\lambda}(x)$ y $\lambda(x)$ nos dan cuatro grados de libertad fermiónica, por último el $d(x)$ nos da uno.

2.7 LAGRANGIANO SUPERSIMÉTRICO

Como se discutió en la sección 1.5, se busca una acción que sea invariante bajo transformaciones supersimétricas. Esto se cumple si el lagrangiano se transforma como

$$\mathcal{L} \rightarrow \mathcal{L}' = \mathcal{L} + \partial^\mu f(x),$$

es decir, que su variación difiera solo en un término de derivada total. Las componentes de más alto orden θ y $\bar{\theta}$ de la expansión de un supercampo siempre se transformarán mediante una derivada total como se muestra arriba, al aplicar al supercampo escalar así como al supercampo vectorial, y la integración sobre los números de Grassmann proyecta esas componentes (del supercampo). Debido a esto podemos redefinir la integración para obtener lo siguiente

$$S = \int d^4x \int d^4\theta \mathcal{L} = \int d^4x \int d^2\theta d^2\bar{\theta} \mathcal{L}, \quad (2.23)$$

la Lagrangiana es una función de N supercampos $\mathcal{L} = f(\Phi_i, \Phi_j, \Phi_i^\dagger, \Phi_j^\dagger)$, esto nos garantiza que sea un invariante supersimétrico. Ahora, podemos escribir una forma convencional la Lagrangiana de los supercampos escalares,

$$\mathcal{L} = \mathcal{L}_{\theta\theta\bar{\theta}\bar{\theta}} + \theta\theta \mathcal{L}_{\bar{\theta}\bar{\theta}} + \bar{\theta}\bar{\theta} \mathcal{L}_{\theta\theta}, \quad (2.24)$$

los índices de Grassmann nos indican la mayor potencia de los componentes lagrangianos. Ya que también se busca que la Lagrangiana sea renormalizable, obtenemos una restricción más, limitando a campos escalares a no ser mayores que la tercera potencia. Esto nos quedará como

$$\mathcal{L} = \Phi_i^\dagger \Phi_i + \bar{\theta}\bar{\theta} W[\Phi] + \theta\theta W[\Phi^\dagger], \quad (2.25)$$

donde el término $\Phi_i^\dagger \Phi_i$ genera los términos cinéticos canónicos de los supercampos quirales, mientras que $W[\Phi]$ es el *superpotencial*, es decir, una función holomorfa de los supercampos quirales que determina los términos de masa y las posibles interacciones entre ellos. En el caso renormalizable, el superpotencial puede escribirse como

$$W[\Phi] = g_i \Phi_i + m_{ij} \Phi_i \Phi_j + \lambda_{ijk} \Phi_i \Phi_j \Phi_k, \quad (2.26)$$

donde los coeficientes g_i , m_{ij} y λ_{ijk} describen, respectivamente, términos lineales, términos de masa bilineales y acoplamientos trilineales.

2.8 INTENSIDAD DEL CAMPO

Aunque la invariancia supergauge desempeña un papel fundamental en la construcción del modelo, su desarrollo formal no se abordará aquí, dado que excede los objetivos principales de esta tesis y ha sido tratado extensamente en la literatura.

Consideremos la siguiente transformación supercampo escalar izquierdo bajo un grupo G

$$\Phi \rightarrow \Phi' = e^{-iq\Lambda^a t_a} \Phi. \quad (2.27)$$

En este caso t_a es parte de los generadores de grupo para el caso no Abelian, q es la carga del supercampo Φ y Λ el parámetro de la transformación gauge. Finalmente para el término cinético debemos introducir un *supercampo vectorial compensador de gauge* V tal que se cumpla que es invariante bajo la transformación gauge

$$\Phi^\dagger e^{qV^a t_a} \Phi \rightarrow \Phi^\dagger e^{qV'^a t_a} \Phi' = \Phi^\dagger e^{iq\Lambda^a t_a} e^{qV'^a t_a} e^{-iq\Lambda^a t_a} \Phi, \quad (2.28)$$

que satisface si

$$e^{qV'^a t_a} = e^{-iq\Lambda^a t_a} e^{qV^a t_a} e^{iq\Lambda^a t_a}.$$

Ahora vamos a expresar la intensidad de campo mediante el tensor de la Electrodinámica Cuántica $F_{\mu\nu}^a = \partial_\mu A_\nu - \partial_\nu A_\mu$. Definimos la intensidad de campo supersimétrico haciendo uso de la matriz espinorial del supercampo vectorial

$$\begin{aligned} \mathcal{W}_A &\equiv -\frac{1}{4} \bar{D} \bar{D} e^{-V^a t_a} D_A e^{V^a t_a}, \\ \bar{\mathcal{W}}_{\dot{A}} &\equiv -\frac{1}{4} D D e^{-V^a t_a} \bar{D}_{\dot{A}} e^{V^a t_a}. \end{aligned} \quad (2.29)$$

La primera ecuación representa el supercampo izquierdo, asimismo la segunda el derecho, respectivamente, y la traza de los productos consigo mismos son los términos invariantes del supergauge, tales que aparecen como potenciales en el lagrangiano supersimétrico. Se puede mostrar que al expandir W_A contiene el tensor de intensidad de campo ordinario, proveniente del Modelo Estándar.

Al tomar la Lagrangiana de la ecuación (2.25) con los grupos no abelianos mencionados, nuestro **Lagrangiana general supersimétrica** es

$$\mathcal{L} = \Phi^\dagger e^{qV^a t_a} \Phi + \bar{\theta} \bar{\theta} W[\Phi] + \theta \theta W[\Phi^\dagger] + \frac{1}{2T(R)} \bar{\theta} \bar{\theta} \text{Tr} [\mathcal{W}^a \mathcal{W}_a], \quad (2.30)$$

$T(R)$ es una constante de normalización que surge de la representación del grupo gauge.

2.9 EL MODELO ESTÁNDAR MÍNIMO SUPERSIMÉTRICO

Ya hemos mencionado que el Modelo Estándar Mínimo Supersimétrico (MEMS) corresponde al álgebra supersimétrica con $N = 1$. Se le denomina “mínimo” porque introduce el menor contenido adicional de supercampos necesario para construir una teoría supersimétrica renormalizable, compatible con la simetría gauge del Modelo Estándar y libre de anomalías. En particular, esta extensión conserva la estructura $SU(3)_C \times SU(2)_L \times U(1)_Y$ e incorpora el contenido de campos requerido para reproducir la fenomenología del Modelo Estándar, incluyendo dos dobletes de Higgs. Ahora vamos a listar el contenido del MEMS

■ Leptones y Sleptones

En la sección 2.6 vimos como construir en el ME un par partícula-antipartícula de Dirac a partir de dos supercampos quirales. Este procedimiento introduce también dos campos escalares complejos; cada uno describe una partícula escalar y su antipartícula, de modo que en conjunto se tienen cuatro grados de libertad escalares reales. En el sector leptónico, estos corresponden a los sleptones. Para el campo de los leptones introducimos los supercampos

$$L_i = (\nu_i, l_i) \quad \text{y} \quad \tilde{E}_i, \quad (2.31)$$

donde l_i y E_i son los supercampos para los leptones y sleptones cargados, a su vez ν_i para los neutrinos y sneutrinos izquierdos. Los índices $i = 1, \dots, 3$ denotan la generación de partículas. En el doblete $SU(2)_L$ tenemos a los supercampos l_i, ν_i y para el singlete $SU(2)_L$ el \tilde{E}_i .

■ Quarks y Squarks

De manera análoga a los leptones y sleptones introducimos

$$Q_i = (u_i, d_i), \quad \tilde{U}_i \quad \text{y} \quad \tilde{D}_i, \quad (2.32)$$

cabe aclarar que aquí se han omitido los índices de color, los supercampos izquierdos están en el doblete Q_i de $SU(2)_L$, el resto se transforma como un singlete.

■ Bosones Gauge y Gauginos

En la ecuación (2.30) tenemos el término $qV^a t_a$, esto significa que necesitamos un supercampo vectorial V^a para cada generador t_a de los tres grupos de simetría del ME $SU(3)_c, SU(2)_L, U(1)_Y$. Se denotarán los supercampos que pertenecen a estos

grupos como

$$C^a, \quad W^a \quad y, \quad B^0, \quad (2.33)$$

respectivamente. C^a nos da los 8 gluones, mientras que los restantes nos dan los 3+1 bosones vectoriales antes del rompimiento espontáneo de simetría electrodébil. Los supercompañeros fermiónicos del ME bosones gauge, los gauginos, son denotados \tilde{g} , \tilde{W}^a y \tilde{B}^0 , son llamados el gluino, wino, y el bino, respectivamente.

■ Higgs y Higgsinos

Requerimos de supercampos de Higgs. Para que este supercampo otorgue masa para los quarks tipo up y down necesitaremos definir dos dobletes escalares complejos de Higgs

$$H_u = (H_0^+, H_u^0), \quad H_d = (H_d^0, H_d^-), \quad (2.34)$$

donde la carga está representada por lo signos. En la ecuación (2.34) tenemos lo que significa los cuatro supercampos quirales del sector de Higgs, cuyas componentes fermiónicas corresponden a cuatro espinores de Weyl que se combinan para formar los Higgsinos, y ocho grados de libertad bosónicos; de los cuales tres son bosones de Goldstone que son tomados por Z y W^\pm mediante el mecanismo de Higgs y les da masa, por otro lado los cinco grados de libertad restantes incorporan los eigenestados de masa h^0 , H^0 , A^0 y H^\pm . El Higgs más ligero es el Higgs del ME, $m_h \approx 125.7$.

Aplicando todos estos conceptos **podemos reproducir el Modelo Estándar con el MEMS**. Su forma es

$$\mathcal{L}_{MSSM} = \mathcal{L}_{kin} + \mathcal{L}_W + \mathcal{L}_V + \mathcal{L}_{soft}, \quad (2.35)$$

donde \mathcal{L}_{kin} contiene los términos cinéticos de materia y gauge, \mathcal{L}_W los términos derivados del superpotencial, \mathcal{L}_V las contribuciones del sector gauge y de intensidad de campo supersimétrica, y \mathcal{L}_{soft} los términos de rompimiento suave de supersimetría. En particular, el potencial escalar del sector de Higgs se obtiene de la suma de las contribuciones F , D y de rompimiento suave, es decir,

$$V_H = V_F + V_D + V_{soft}.$$

2.10 MEMS ROMPIMIENTO SUAVE DE SUPERSIMETRÍA

El rompimiento suave adecuado debe ser el que a altas energías mantenga la simetría intacta, mientras que a bajas energías la simetría se rompe por términos suaves. Estos términos deben preservar las cancelaciones en las contribuciones de los loops para masas

escalares, como la masa del Higgs. Entonces los términos para el rompimiento suave del MEMS son²

$$\begin{aligned}
\mathcal{L}_{\text{soft}}^{\text{MSSM}} = & -\frac{1}{2} \left(M_3 \tilde{g}g + M_2 \tilde{W}\tilde{W} + M_1 \tilde{B}\tilde{B} + \text{c.c.} \right) \\
& - \left(\tilde{u}A_{\tilde{u}_i}^H QH_u - \tilde{d}A_{\tilde{d}_i}^H QH_d - \tilde{e}A_{\tilde{e}_i}^H LH_d + \text{c.c.} \right) \\
& - \tilde{Q}^\dagger \mathbf{m}_{\tilde{Q}}^2 \tilde{Q} - \tilde{L}^\dagger \mathbf{m}_{\tilde{L}}^2 \tilde{L} - \tilde{u} \mathbf{m}_{\tilde{u}}^2 \tilde{u}^\dagger - \tilde{d} \mathbf{m}_{\tilde{d}}^2 \tilde{d}^\dagger - \tilde{e} \mathbf{m}_{\tilde{e}}^2 \tilde{e}^\dagger \\
& - m_{H_u}^2 H_u^* H_u - m_{H_d}^2 H_d^* H_d - (B_{ij} H_u H_d + \text{c.c.}), \tag{2.36}
\end{aligned}$$

donde M_3, M_2, M_1 son los términos de masa de los gluinos, wino y bino. La segunda línea es un acoplamiento escalar trilineal a los campos escalares, donde cada \mathbf{A} es una matriz 3×3 , donde dichos elementos están en correspondencia con los acoplamientos de Yukawa del superpotencial. La tercera línea contiene las masas de los quarks y squarks, cada \mathbf{m}^2 es una matriz hermitiana 3×3 . La última línea consiste en términos de masas escalares, términos de masa de Higgs $m_{H_u}^2, m_{H_d}^2$, y B_{ij} el término que mezcla los dos dobletes contribuyentes del potencial de Higgs.

2.11 PARIDAD R

En la lagrangiana supersimétrica del MEMS el número bariónico B y el leptónico L ya no son cantidades conservadas, términos como LH_u y $\bar{U}\bar{D}\bar{D}$ en el superpotencial del MEMS violan el número bariónico y conservación del número leptónico, respectivamente. Aunque experimentalmente se ha probado no encontrar dichas violaciones en estas leyes de conservación, de forma perceptible podrían existir, pero muy pequeñas. Por ejemplo la medida en el límite más bajo en el tiempo de vida de protones: $\tau_{p \rightarrow e + \pi^0} \leq 5.0 \cdot 10^{40} s$, para justificar la poca frecuencia del decaimiento del protón debiésemos ajustar la constante de acoplamiento λ de los términos de violación a ser extremadamente pequeños. De forma alternativa, podemos usar una ley de conservación para evadir dichos acoplamientos, nombrada Paridad-R³

$$R = (-1)^{2s+3B+L}, \tag{2.37}$$

donde s es el spin de la partícula. Todas las partículas del ME tendrán $R = 1$, y todos los supercompañeros $R = -1$. Imponer paridad-R tiene implicaciones importante: Las spartículas son siempre producidas en números pares. Las spartículas deben decaer, con excepción de la *partícula supersimétrica más ligera* (LSP), la cual es completamente es-

²Las subsecciones 2.10 a 2.13 se desarrollan principalmente con base en [3, 8], y de manera complementaria en [33].

³La definición y discusión de la paridad R se siguen de [3, 8].

table y constituye el estado final de cualquier cadena de decaimiento. Esto implica que la cadena de decaimiento consiste en un número impar de spartículas.

2.12 RUPTURA DE SIMETRÍA ELECTRODÉBIL RADIATIVA

En supersimetría necesitamos un mecanismo de rompimiento de simetría electrodébil (EWSB), pero es llamado radiativo-EWSB (REWSB). El potencial del MEMS para los componente de campo del escalar de Higgs es⁴

$$\begin{aligned}
V(H_u, H_d) = & |\mu|^2 (|H_u^0|^2 + |H_u^+|^2 + |H_d^0|^2 + |H_d^-|^2) \\
& + \frac{1}{8} (g^2 + g'^2) (|H_u^0|^2 + |H_u^+|^2 - |H_d^0|^2 - |H_d^-|^2)^2 \\
& + \frac{1}{2} g^2 |H_u^+ H_d^{0*} + H_u^0 H_d^{-*}|^2 \\
& + m_{H_u}^2 (|H_u^0|^2 + |H_u^+|^2) + m_{H_d}^2 (|H_d^0|^2 + |H_d^-|^2) \\
& + [b (H_u^+ H_d^- - H_u^0 H_d^0) + \text{c.c.}], \tag{2.38}
\end{aligned}$$

donde μ es la variante supersimétrica del parámetro encontrado en el ME de la masa del bosón de Higgs. Los términos g^2 y g'^2 son los términos de acoplamiento gauge, provenientes del grupo $SU(2)_L, U(1)_Y$. Los tres coeficientes restantes m_{H_u}, m_{H_d} y b son los parámetros que corresponden a los términos de la ruptura suave del lagrangiano. Para los cuatro campos escalares complejos, H_u^+, H_u^0, H_d^0 y H_d^- , tenemos que el potencial de Higgs tiene ocho grados de libertad. Como sabemos para que las partículas del ME adquieran masa se requiere romper la simetría $SU(2)_L \times U(1)_Y \rightarrow U(1)_{em}$. Esto se cumple bajo las siguientes condiciones: 1. El mínimo debe preservar la simetría $U(1)$; 2. Debe tener valores distintos de cero en el mínimo; 3. Debe estar acotado por abajo. En el ME estas condiciones se satisfacen por las desigualdades $\lambda > 0$ y $\mu^2 < 0$. Para el caso del MEMS, vemos que es posible que la libertad para producir transformaciones gauge $SU(2)_L$ nos da la posibilidad de eliminar un VEV mediante una rotación para un componente débil de isoespín, tal que en el mínimo del potencial, $V(H_u^0, H_d^0)$, podemos establecer $H_u^+ = 0$. Tenemos que en el mínimo debe cumplirse que $\partial V / \partial H_u^+ = 0$. Esto significa que los componentes cargados del escalar de Higgs son ambos cero $H_d^+ = H_d^- = 0$. Entonces no pueden alcanzar VEVs y $U(1)_{em}$ permanece no roto, cumpliendo la primera condición. Al hacer cero los términos antes mencionados en la ecuación (2.38) sigue al potencial

$$V(H_u^0, H_d^0) = (|\mu|^2 + m_{H_u}^2) |H_u^0|^2 + (|\mu|^2 + m_{H_d}^2) |H_d^0|^2$$

⁴La presentación de esta subsección se apoya en [3, 8].

$$+ \frac{1}{8} (g^2 + g'^2) (|H_u^0|^2 - |H_d^0|^2)^2 - (bH_u^0 H_d^0 + \text{c.c.}) \quad (2.39)$$

El rompimiento de simetría resultando de REWSB debe ser capaz de reproducir los resultados de EWSB, por lo tanto está sujeto a la condición

$$\langle H_u^0 \rangle + \langle H_d^0 \rangle = v_u^2 + v_d^2 \equiv v^2 = \frac{2m_Z^2}{g^2 + g'^2} \approx (246 \text{ GeV})^2, \quad (2.40)$$

La condición en los VEVs nos permite formular la expresión en términos de un parámetro

$$\tan\beta \equiv \frac{v_u}{v_d}, \quad 0 < \beta < \pi/2. \quad (2.41)$$

La condición del mínimo debe satisfacer

$$\frac{\partial V}{\partial H_u^0} = \frac{\partial V}{\partial H_d^0} = 0, \quad (2.42)$$

nos permite eliminar los parámetros b y $|\mu|$ en (2.39), pero no el signo de μ , en su lugar podríamos eliminar $m_{H_u}^2$ y $m_{H_d}^2$.

2.13 ELECTROWEAKINOS

En el rango de energías que se rompe la simetría electrodébil tendremos estados de partículas que son mezclas de campos fermiónicos⁵. Para esto solo tomaremos las cargas de $U(1)_{em}$, tales son las únicas con relevancia para este caso. Los campos supersimétricos fermiónicos, los bosinos, que son los candidatos para la mezcla son los Higgsinos $\tilde{H}_u^+, \tilde{H}_u^0, \tilde{H}_d^-, \tilde{H}_d^0$ y los gauginos electrodébiles: el bino \tilde{B}^0 y los tres winos $\tilde{W}^0, \tilde{W}^\pm$.

Para mezclar los campos la única condición necesaria es que deben tener la misma carga. Entonces tenemos un eigenestado base consistiendo de campos neutrales

$$\tilde{\chi}^0 = \left(\tilde{B}, \tilde{W}^0, \tilde{H}_d^0, \tilde{H}_u^0 \right), \quad (2.43)$$

y para los campos cargados

$$\tilde{\chi}^\pm = \left(\tilde{W}^\pm, \tilde{H}_u^\pm, \tilde{W}^\mp, \tilde{H}_d^\mp \right). \quad (2.44)$$

⁵La descripción del sector electroweakino se basa principalmente en [3, 8].

En estas bases el lagrangiano del electroweakino queda como

$$\mathcal{L}_{\text{EWino}} = -\frac{1}{2} (\tilde{\chi}^0)^T M_N \tilde{\chi}^0 - \frac{1}{2} (\tilde{\chi}^\pm)^T M_C \tilde{\chi}^\pm + \text{c.c.}, \quad (2.45)$$

donde $M_{\tilde{\chi}}$ y $M_{\tilde{C}}$ son las matrices de masa.

Photino y Zino. El fotino $\tilde{\gamma}$, y el zino \tilde{Z} , son los compañeros supersimétricos del fotón y el bosón Z , respectivamente, y provienen como una mezcla de gauginos neutros

$$\begin{aligned} \tilde{\gamma} &= N'_{11} \tilde{B}^0 + N'_{12} \tilde{W}^0 \\ \tilde{Z} &= N'_{21} \tilde{B}^0 + N'_{22} \tilde{W}^0, \end{aligned} \quad (2.46)$$

donde N' son las matrices 2×2 de la mezcla de bosones gauge del ME.

Neutralinos. Mezclando los cuatro bosinos neutros EW de la ecuación (2.43) obtenemos un *neutralino*, con los siguientes eigenestados de masa:

$$\tilde{\chi}_i^0 = N_{i1} \tilde{B}^0 + N_{i2} \tilde{W}^0 + N_{i3} \tilde{H}_d^0 + N_{i4} \tilde{H}_u^0, \quad (2.47)$$

hay cuatro neutralinos que denota el índice i . Se ordenan de manera ascendente por sus masas $m_{\tilde{\chi}_1^0} < m_{\tilde{\chi}_2^0} < m_{\tilde{\chi}_3^0} < m_{\tilde{\chi}_4^0}$, donde el índice j recorre los autoestados de gauge ($\tilde{B}^0, \tilde{W}^0, \tilde{H}_d^0, \tilde{H}_u^0$), de modo que N_{ij} describe la composición del neutralino más ligero $\tilde{\chi}_1^0$ (LSP). primero se considera el LSP $\tilde{\chi}_1^0$. La matriz \mathbf{N} en una matriz unitaria 4×4 de mezcla donde los elementos j indican el tamaño del correspondiente campo en el i -ésimo neutralino.

La matriz de masa para los neutralinos en la ecuación (2.45) puede escribirse, a nivel árbol, de la siguiente manera

$$M_{\tilde{\chi}^0} = \begin{bmatrix} M_1 & 0 & -c_\beta s_W m_Z & s_\beta s_W m_Z \\ 0 & M_2 & c_\beta c_W m_Z & -s_\beta c_W m_Z \\ -c_\beta s_W m_Z & c_\beta c_W m_Z & 0 & -\mu \\ s_\beta s_W m_Z & -s_\beta c_W m_Z & -\mu & 0 \end{bmatrix} \quad (2.48)$$

M_1 y M_2 son las masas del bino y vino neutro \tilde{B}^0, \tilde{W}^0 , del MEMS, mientras que $-\mu$ es el parámetro de masa de los higgsinos $\mu H_u H_d$ en el superpotencial. por último, las variables $s_\beta = \sin \beta$, $c_\beta = \cos \beta$, $s_W = \sin \theta_W$ y $c_W = \cos \theta_W$, donde β es el ángulo de mezcla que relaciona los valores de expectativa de vacío de los dos dobletes de Higgs, $\tan \beta = v_u/v_d$, y θ_W es el ángulo de mezcla débil de Weinberg, definido por $\cos \theta_W = m_W/m_Z$. Estos términos provienen del acoplamiento Higgs-higgsino-gaugino de la parte cinética del superpotencial. Aquí tenemos estas cuatro expresiones en términos del ángulo de mezcla

débil de Weinberg θ_W , definido mediante $\cos\theta_W = m_W/m_Z$ (donde m_W y m_Z son las masas del bosón W y Z respectivamente), y la masa m_Z como sigue

$$\frac{1}{\sqrt{2}}g'v_{d(u)} = c_\beta/s_\beta s_W m_Z, \quad \frac{1}{\sqrt{2}}gv_{d(u)} = (c_\beta/s_\beta) c_W m_Z.$$

Se pueden encontrar las masas $\tilde{\chi}_i^0$ al diagonalizar la matriz (2.48).

Charginos. Análogamente si mezclamos los campos cargados obtenemos los *charginos*, donde los eigenestados de masa se relacionan a los eigenestados gauge de la ecuación (2.44) como

$$\tilde{\chi}_i^+ = V_{i1}\widetilde{W}^+ + V_{i2}\widetilde{H}_u^+, \quad \tilde{\chi}_i^- = U_{i1}\widetilde{W}^- + U_{i2}\widetilde{H}_d^-, \quad (2.49)$$

donde U y V son las matrices de mezcla 2×2 . La matriz de masa de la parte cargada del lagrangiano es

$$M_{\tilde{\chi}_i^\pm} = \begin{bmatrix} 0 & X^T \\ X & 0 \end{bmatrix}, \quad X = \begin{bmatrix} M_2 & \sqrt{2}s_\beta m_W \\ \sqrt{2}c_\beta m_W & \mu \end{bmatrix}, \quad (2.50)$$

los términos nuevamente vienen del lagrangiano del MEMS, donde M_2 viene de los términos suaves del wino cargado W^\pm , μ viene del superpotencial, y el resto es la parte cinética, donde similar a los neutralinos, se usó $gv_{d(u)} = \sqrt{2}(s_\beta/c_\beta)m_W \dots$. Similar a los neutralinos, las matrices de mezcla U , V se seleccionan de tal manera que diagonalicen la matriz de masa

$$UXV^{-1} = D, \quad (2.51)$$

también para la entrada (1, 2) de la matriz $M_{\tilde{\chi}_i^\pm}$, dada por X^T .

2.14 ANOMALÍA DEL MUON EN TÉRMINOS SUAVES DE SUPERSIMETRÍA

Para abordar el problema del $g - 2$, se plantea una extensión de sabor del MEMS, introduciendo una estructura jerárquica en los términos trilineales suaves del lagrangiano. Estos términos emergen en el contexto de la ruptura de escala de la SUSY. En esta sección, y en particular en las subsecciones siguientes, se seguirá el desarrollo presentado en [4] para determinar la contribución supersimétrica a un loop.

2.14.1 ESTRUCTURA DE SABOR EN EL ROMPIMIENTO SUAVE DE SUPERSIMETRÍA

El rompimiento suave del lagrangiano de SUSY incluye acoplamientos trilineales escalares A_{ij}^H , bilineales B_{ij} , términos escalares de masa \tilde{M}_f^2 , y masa para los gauginos M_i^2 . En ausencia de la mezcla de sabor, la parte escalar fermiónica, en los términos suaves de SUSY, tendrá la siguiente forma

$$\mathcal{L}_{\text{soft}}^f = - \sum_{\tilde{f}_i} M_{\tilde{f}_i}^2 \tilde{f}_i \tilde{f}_i - \left(A_{\tilde{f},i} \tilde{f}_L^i H_1 \tilde{f}_R^i + h.c. \right), \quad (2.52)$$

donde \tilde{f} son los campos escalares en el supermultiplete. Los primeros términos de (2.52) contribuyen a los diagonales de la matriz de masa de los sfermiones 2×2 , los segundos son los acoplamientos del Higgs con diferentes sfermiones. En este caso se considera una mezcla de sabor en los términos trilineales, A_f^{ij} una matriz 3×3 , ya que se consideran tres sabores, con dos campos escalares para cada uno. Los términos trilineales fermiónicos son

$$\mathcal{L}_{H\tilde{f}_i\tilde{f}_j}^{\text{soft}} = -A_u^{ij} \tilde{Q}_i H_2 \tilde{U}_j - A_d^{ij} \tilde{Q}_i H_1 \tilde{D}_j - A_l^{ij} \tilde{L}_i H_1 \tilde{E}_j + c.c. \quad (2.53)$$

donde \tilde{Q}_i denota la componente escalar del supercampo Q_i definido en la ec. 2.32, mientras que \tilde{U}_j , \tilde{D}_j , \tilde{L}_i y \tilde{E}_j son directamente las componentes escalares de los supercampos \tilde{U}_j , \tilde{D}_j , \tilde{L}_i y \tilde{E}_j respectivamente. \tilde{Q}_i representa el doblete del squark compañero del ME $SU(2)$ doblete izquierdo y \tilde{U}_j , \tilde{D}_j son los singletes de los squarks, \tilde{L}_i es el doblete del sleptón y \tilde{E} el singlete. Para este análisis solo se toma en cuenta la parte sleptónica. Cuando se considera la ruptura de simetría electrodébil, el lagrangiano para los sleptones queda como

$$\mathcal{L}_{H\tilde{f}_i\tilde{f}_j} = \frac{A_l^{ij}}{\sqrt{2}} \left[(\phi_1^0 - i\chi_1^0) \tilde{l}_{iR}^* \tilde{l}_{jL} - \sqrt{2} \phi_1^- \tilde{l}_{iR}^* \tilde{\nu}_{jL} + v_1 \tilde{l}_{iR}^* \tilde{l}_L \right] + h.c. \quad (2.54)$$

2.14.2 MATRIZ DE MASA PARA LOS SFERMIONES

La contribución a la matriz de masa de los sfermiones proviene de la interacción de los escalares del Higgs con ellos, donde aparece en el superpotencial y los términos de ruptura suave de SUSY. En nuestro caso, las contribuciones provenientes de los términos de masa suave son $\tilde{M}_{l,LL}^2$, $\tilde{M}_{l,RR}^2$, del acoplamiento trilineal después de romper la simetría electrodébil A_{ij}^l y los términos F , D . Siendo los siguientes

$$\tilde{M}_l^2 = \begin{pmatrix} \tilde{M}_{l,LL}^2 + F_{l,LL} + D_{l,LL} & A_{ij}^l + F_{l,LR} \\ (A_{ij}^l + F_{l,LR})^\dagger & \tilde{M}_{l,RR}^2 + F_{l,RR} + D_{l,RR} \end{pmatrix}. \quad (2.55)$$

Aquí F_f , y D_f son los campos auxiliares en el supermultiplete, tales que tengan el mismo grado de libertad bosónico y fermiónico. Los campos auxiliares F vienen del supercampos quiral del Higgs y contribuyen como sigue⁶:

$$\begin{aligned} F_{l,LL,RR} &= m_l^2 \left(\tilde{l}_L \tilde{l}_L + \tilde{l}_R \tilde{l}_R \right), \\ F_{l,LR} &= m_l \mu \tan \beta \left(\tilde{l}_R^* \tilde{l}_L + \tilde{l}_L^* \tilde{l}_R \right). \end{aligned} \quad (2.56)$$

Para los campos auxiliares D de los supercampos escalares fermiónicos tenemos lo siguiente:

$$D_{l,LL,RR} = -M_Z^2 \cos 2\beta \left[(T_{3l} - s_W^2 Q_l) \tilde{l}_L \tilde{l}_L + s_W^2 Q_l \tilde{l}_R \tilde{l}_R \right], \quad (2.57)$$

donde $l = e, \mu, \tau$. Los elementos para la matriz de masa de los sleptones en (2.55), para los diferentes sabores (i.e. $i, j = e, \mu, \tau$) son

$$m_{LL,l}^2 = \tilde{M}_{L,l}^2 + m_l^2 + \frac{1}{2} \cos 2\beta (2M_W^2 - M_Z^2), \quad (2.58)$$

$$m_{RR,l}^2 = \tilde{M}_{E,l}^2 + m_l^2 - \cos 2\beta \sin^2 \theta_W M_Z^2, \quad (2.59)$$

$$m_{LR,l}^2 = \frac{A_l v \cos \beta}{\sqrt{2}} - m_l \mu \tan \beta. \quad (2.60)$$

2.14.3 TÉRMINOS TRILINEALES SUAVES

Se asume una escala de rompimiento que $\tilde{m}_{RR,l}^2 \approx \tilde{m}_{LL,l}^2 = \tilde{m}_S^2$ pero se propone una mezcla de dos familias escalares de leptones en los términos de masa. Se va a considerar las contribuciones totales de las dos familias de compañeros supersimétricos escalares de leptones en el mismo orden de magnitud considerando la tercera familia asociado directamente al valor del trilineal A_0 , mientras que los demás términos podrán llegar al mismo orden de magnitud o menor: $w, y \in [0, 1]$ resultando en la forma de los términos trilineales

$$A_l = \begin{pmatrix} 0 & 0 & 0 \\ 0 & w & y \\ 0 & y & 1 \end{pmatrix} A_0. \quad (2.61)$$

Como se ha asumido que la mezcla es en los smuones y staus solamente y los selectrones están desacoplados, la matriz de masa del smuon-stau 4×4 asociados originalmente a smuones y staus. En la mezcla los nuevos estados se pierde la etiqueta de sabor y tendrá

⁶En este trabajo μ_{SUSY} corresponde al parámetro μ del superpotencial del MEMS, definido en la (2.39). Esta aclaración es importante por el uso que se le da en secciones posteriores para el planteamiento del código y los parámetros en la red neuronal, por lo que se usa de manera indistinta una u otra a lo largo de este texto

la forma

$$\tilde{M}_{\mu-\tau}^2 = \begin{pmatrix} m_{LL,\mu}^2 & X_m & 0 & A_y \\ X_m & m_{RR,\mu}^2 & A_y & 0 \\ 0 & A_y & m_{LL,\tau}^2 & X_t \\ A_y & 0 & X_t & m_{RR,\tau}^2 \end{pmatrix}, \quad (2.62)$$

donde

$$\begin{aligned} A_y &= \frac{1}{\sqrt{2}} y A_0 v \cos \beta, \\ X_m &= \frac{1}{\sqrt{2}} w A_0 v \cos \beta - \mu_{\text{SUSY}} m_\mu \tan \beta, \\ X_t &= \frac{1}{\sqrt{2}} A_0 v \cos \beta - \mu_{\text{SUSY}} m_\tau \tan \beta. \end{aligned} \quad (2.63)$$

Mediante este camino tendremos masas físicas de sleptones no degeneradas.

$$\begin{aligned} m_{\tilde{\tau}_{1,2}}^2 &= \frac{1}{2} (2\tilde{m}_S^2 - X_m - X_t \pm R), \\ m_{\tilde{\mu}_{1,2}}^2 &= \frac{1}{2} (2\tilde{m}_S^2 + X_m + X_t \pm R), \end{aligned} \quad (2.64)$$

donde $R = \sqrt{4A_y^2 + (X_t - X_m)^2}$

La transformación que diagonaliza la matriz de masas para los sleptones $Z_{\tilde{i}}$ que es una matriz 2×2 , $Z_{\tilde{i}}^\dagger \tilde{M}_\mu^2 Z_{\tilde{i}} = \tilde{M}_{\tilde{i},diag}^2$, tiene la forma de

$$Z_{\tilde{i}} = \frac{1}{\sqrt{2}} \begin{pmatrix} \Phi & -\Phi \\ \Phi\sigma^3 & \Phi\sigma^3 \end{pmatrix}, \quad (2.65)$$

σ_3 representa la matriz de Pauli y

$$\Phi = \begin{pmatrix} -\sin \frac{\varphi}{2} & -\cos \frac{\varphi}{2} \\ \cos \frac{\varphi}{2} & -\sin \frac{\varphi}{2} \end{pmatrix}. \quad (2.66)$$

Esta es la transformación de la base de estados físicos a la base de estados de sabor. Transformamos hacia los eigenestados físicos mediante

$$\begin{pmatrix} \tilde{\mu}_L \\ \tilde{\tau}_L \\ \tilde{\mu}_R \\ \tilde{\tau}_R \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} -\sin \frac{\varphi}{2} & -\cos \frac{\varphi}{2} & \sin \frac{\varphi}{2} & \cos \frac{\varphi}{2} \\ \cos \frac{\varphi}{2} & -\sin \frac{\varphi}{2} & -\cos \frac{\varphi}{2} & \sin \frac{\varphi}{2} \\ -\sin \frac{\varphi}{2} & \cos \frac{\varphi}{2} & -\sin \frac{\varphi}{2} & \cos \frac{\varphi}{2} \\ \cos \frac{\varphi}{2} & \sin \frac{\varphi}{2} & \cos \frac{\varphi}{2} & \sin \frac{\varphi}{2} \end{pmatrix} \begin{pmatrix} \tilde{l}_1 \\ \tilde{l}_2 \\ \tilde{l}_3 \\ \tilde{l}_4 \end{pmatrix}$$

$$= Z_{\tilde{l}} \begin{pmatrix} \tilde{\mu}_1 \\ \tilde{\tau}_1 \\ \tilde{\mu}_2 \\ \tilde{\tau}_2 \end{pmatrix}, \quad (2.67)$$

donde

$$\tan\psi = \frac{2A_y}{X_m - X_t}. \quad (2.68)$$

Además \tilde{l}_i con $i = 1, 2, 3, 4$ son los autoestados de sabor de los sleptones, relacionados con los autoestados físicos $(\tilde{\mu}_L, \tilde{\tau}_L, \tilde{\mu}_R, \tilde{\tau}_R)$ mediante la matriz de mezcla $Z_{\tilde{l}}$ de la (2.65).

2.14.4 INTERACCIÓN NEUTRALINO-LEPTON-SLEPTON

Asumimos con base en el MEMS que los neutralinos $\tilde{\chi}_i^0$, con $i = 1, 2, 3, 4$ consisten en una mezcla de partes fermiónicas de supercampos vectoriales, i.e. gauginos y Higgsinos. La matriz de masa de los neutralinos es

$$M_N = \begin{pmatrix} M_1 & 0 & -M_Z \sin\theta_W \cos\beta & M_Z \sin\theta_W \sin\beta \\ * & M_2 & M_Z \cos\theta_W \cos\beta & -M_Z \cos\theta_W \sin\beta \\ * & * & 0 & -\mu_{\text{SUSY}} \\ * & * & * & 0 \end{pmatrix}. \quad (2.69)$$

La diagonalización de esta matriz implica transformar los neutralinos como sigue

$$\begin{pmatrix} \tilde{\chi}_1^0 \\ \tilde{\chi}_2^0 \\ \tilde{\chi}_3^0 \\ \tilde{\chi}_4^0 \end{pmatrix} = \begin{pmatrix} \eta_1 & 0 & 0 & 0 \\ 0 & \eta_2 & 0 & 0 \\ 0 & 0 & \eta_3 & 0 \\ 0 & 0 & 0 & \eta_4 \end{pmatrix} (\Theta_N) \begin{pmatrix} \tilde{B}^0 \\ \tilde{W}^0 \\ \tilde{H}_1^0 \\ \tilde{H}_2^0 \end{pmatrix}. \quad (2.70)$$

El lagrangiano de interacción para neutralino-fermion-sfermion en el MEMS es

$$\begin{aligned} \mathcal{L}_{\tilde{\chi}^0 \tilde{f} f} = & -\frac{g}{\sqrt{2}} \sum_{n=1}^4 \left\{ \sum_{X=1}^3 \tilde{\chi}_n^0 [l_n^{NfL} P_L + r_n^{NfL} P_R] f \tilde{f}_X^* \right. \\ & \left. + \sum_{X=4}^6 \tilde{\chi}_n^0 [l_n^{NfR} P_L + r_n^{NfR} P_R] f \tilde{f}_X^* \right\} + h.c. \end{aligned} \quad (2.71)$$

donde (l_n) y (r_n) son los acoplamientos fermion-neutralino izquierdo y derecho, respectivamente. $P_{L,R}$ son los operadores quirales ordinarios, las etiquetas para los supercompañeros

escalares de los fermiones son L para sfermiones $X = 1, 2, 3$ y $R = 4, 5, 6$ en la interacción base, y g es la constante de acoplamiento $U(1)$. Los acoplamientos del neutralino-fermion-sfermion están dados por

$$l_n^{NeL} = -\eta_n^* \left[(\Theta_N)_{n2} + \frac{s_W}{c_W} (\Theta_N)_{n1} \right], \quad (2.72)$$

$$r_n^{NeL} = \eta_n \frac{m_e}{M_W \cos \beta} (\Theta_N)_{n3}, \quad (2.73)$$

$$l_n^{NeR} = \eta_n^* \frac{m_e}{M_W \cos \beta} (\Theta_N)_{n3}, \quad (2.74)$$

$$r_n^{NeR} = 2\eta_n \frac{s_W}{c_W} (\Theta_N)_{n1}, \quad (2.75)$$

donde $\eta\Theta_N$ es la matriz de rotación que diagonaliza a la matriz de masa del neutralino. Ahora mediante los eigenestados de masa de los sleptonos (2.67) podemos reescribir el lagrangiano de interacción en la base de los estados físicos de sleptonos, neutralino-lepton-slepton como

$$\begin{aligned} \mathcal{L}_{\tilde{\chi}^0 \tilde{l} \tilde{u}} = & -\frac{g}{\sqrt{2}} \sum_{n=1}^4 \tilde{\chi}_n^0 \left\{ C_{n+}^{NeLR} \left[\sin \frac{\varphi}{2} (\tau \tilde{\tau}_2^* - \mu \tilde{\mu}_1^*) + \cos \frac{\varphi}{2} (\mu \tilde{\tau}_2^* + \tau \tilde{\mu}_1^*) \right] \right. \\ & \left. - C_{n-}^{NeLR} \left[\sin \frac{\varphi}{2} (\tau \tilde{\tau}_1^* - \mu \tilde{\mu}_2^*) + \cos \frac{\varphi}{2} (\mu \tilde{\tau}_1^* + \tau \tilde{\mu}_2^*) \right] \right\}, \end{aligned} \quad (2.76)$$

donde

$$C_{n\pm}^{NeLR} = C_n^{NeL} \pm C_n^{NeR}, \quad \text{and} \quad C_n^{NeL(R)} = l_n^{NeL(R)} P_L + r_n^{NeL(R)} P_R.$$

Para introducir la violación de sabor leptónico y calcular los loops supersimétricos se trabajó en una base física al diagonalizar la matriz de masa obteniendo sabores de leptones mezclados, al incluir sólo dos parámetros libres, que son reducidos cuando hacemos $w = 1$, asumiendo (2.61). Usando los acoplamientos obtenidos para calcular el proceso de violación de sabor. En específico, para calcular la contribución a un loop supersimétrico del sfermion-neutralino para el proceso de violación de sabor leptónico $\tau \rightarrow \mu + \gamma$ (este proceso no se desarrolló de explícita en este trabajo), para simplificar, se asumió que el neutralino más ligero es principalmente un bino (\tilde{B}).

Considerando el límite $M_1, M_2, \mu_{SUSY} \gg m_Z$, el neutralino más ligero se considera un bino $\tilde{\chi}_1^0 \approx \tilde{B}$, esto nos permite hacer $(\Theta_N)_{1i} \approx \delta_{1i}$ en (2.72). El eigenvalor de la masa para el neutralino más ligero es [34]

$$m_{\tilde{N}_1} = M_1 - \frac{m_Z^2 s_W^2 (M_1 + \mu \sin 2\beta)}{\mu^2 - M_1^2} + \dots \quad (2.77)$$

Tabla 2: Acoplamientos escalares y pseudosacalares del bino–leptón–slepton con una mezcla de violación de sabor leptónico. Adaptada de [4].

\tilde{l}	$\tilde{\mu}_1$	$\tilde{\mu}_2$	$\tilde{\tau}_1$	$\tilde{\tau}_2$
$S_{\tilde{B}\tilde{\tau}\tilde{l}}$	$3 \sin \frac{\varphi}{2}$	$\sin \frac{\varphi}{2}$	$\cos \frac{\varphi}{2}$	$3 \cos \frac{\varphi}{2}$
$P_{\tilde{B}\tilde{\tau}\tilde{l}}$	$\sin \frac{\varphi}{2}$	$3 \sin \frac{\varphi}{2}$	$3 \cos \frac{\varphi}{2}$	$\cos \frac{\varphi}{2}$
$S_{\tilde{B}\tilde{\mu}\tilde{l}}$	$3 \cos \frac{\varphi}{2}$	$\cos \frac{\varphi}{2}$	$-\sin \frac{\varphi}{2}$	$-3 \sin \frac{\varphi}{2}$
$P_{\tilde{B}\tilde{\mu}\tilde{l}}$	$\cos \frac{\varphi}{2}$	$3 \cos \frac{\varphi}{2}$	$-3 \sin \frac{\varphi}{2}$	$-\sin \frac{\varphi}{2}$

Lo que nos permite tomarlo como neutralino tipo bino en el límite para $M_1 < M_2 \ll |\mu_{SUSY}|$. Podemos escribir en este caso el acoplamiento bino-lepton-slepton como

$$g_{\tilde{B}l\tilde{l}} = -\frac{g \tan \theta_W}{4} [S_{\tilde{B}l\tilde{l}} + P_{\tilde{B}l\tilde{l}}\gamma^5], \quad (2.78)$$

donde \tilde{l} corre sobre todos los eigenestados dados en (2.67). Para el decaimiento de $\tau \rightarrow \mu + \gamma$ los acoplamientos escalares y pseudoescalares están dados en la tabla 2. La amplitud total es invariante gauge y libre de divergencias ultravioleta, la podemos escribir como sigue [4]

$$\mathcal{M}_T = \bar{u}(p_1) i\sigma^{\mu\nu} k_\nu \epsilon_\mu (E + F\gamma^5) u(p_2), \quad (2.79)$$

donde las funciones de los loops E y F contienen la contribución de los sleptones $\tilde{l}_{1,2,3,4}$ corriendo dentro del loop.

$$E = 5C \sum_{\tilde{l}} E_{\tilde{l}}, \quad F = -3C \sum_{\tilde{l}} F_{\tilde{l}},$$

$$C = \frac{ieg^2 \tan^2 \theta_W \sin \varphi}{2(16\pi)^2 (m_\tau^2 - m_\mu^2)}. \quad (2.80)$$

Las funciones $E_{\tilde{l}}, F_{\tilde{l}}$ están escritas en términos de las funciones de Passarino-Veltman Obtenidas de [40].

Ahora para continuar con el problema del muon con una contribución del MEMS mediante una violación de sabor leptónico(LFV) en el sector sleptónico. Buscamos el proceso de LFV $\tau \rightarrow \mu + \gamma$ y calcular el $g - 2$ mediante una mezcla de familias smuones y staus, $a_\mu^{\tilde{I}\tilde{B}}$,

Con las proposiciones que se hicieron se evitan contribuciones extras $\mu \rightarrow e + \gamma$. Para establecer una restricción se considera una condición tal que, $|a_\mu^{\text{Exp}} - a_\mu^{\text{TH}}| \leq 3.3\sigma$, donde $a_\mu^{\text{TH}} = a_\mu^{\text{th}} + a_\mu^{\tilde{I}\tilde{B}}$ nos indica que el loop supersimétrico de la violación de sabor leptónico en leptones cargados no es la única contribución que resuelve la discrepancia como proponen

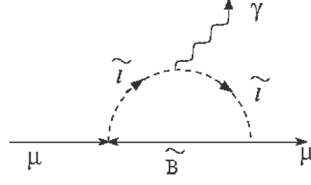


Figura 10: Contribución del slepton a a_μ . Diagrama obtenido de [4].

en [4]. Usando los términos de LFV que se han mencionado hasta ahora podemos obtener la contribución del momento magnético anómalo del muon $a_\mu^{\tilde{l}\tilde{B}}$. Definiendo el radio $r = m_\mu/M_1$ y tomando los términos dominantes cuando $r^2 \rightarrow 0$, y $M_1 = m_{\tilde{B}}$ como la masa del bino. Siguiendo el método de [28], aislamos el término $(p_1 + p_2)^\mu$, calculando la contribución a un loop, se puede escribir

$$\begin{aligned}\bar{u}(p_1)\Gamma^\mu u(p_2) &= \bar{u}(p_1) [B(p_1 + p_2)^\mu + \dots] u(p_2) \\ &= \bar{u}(p_1) \left[\frac{i\sigma^{\mu\nu}}{2m_\mu} q_\nu \right] F_2(q^2) u(p_2) + \dots\end{aligned}\quad (2.81)$$

donde solo estamos mostrando los términos que son proporcionales a $(p_1 + p_2)^\mu$ ya que son estos términos los que extraen la interacción magnética para $g - 2$, el resto que no lo es, se distingue con los puntos suspensivos. Ahora podemos definir la anomalía como $\delta_{a_\mu} = \frac{g-2}{2} = F(0)$ con $q_1 = p_2 - p_1$. Ahora escribimos la interacción magnética en términos del proceso del loop proporcionales a la suma de los momentos como

$$\begin{aligned}B(p_1 + p_2)^\mu, \\ \frac{g-2}{2} = F_2(q^2 \rightarrow 0) = -2m_\mu B.\end{aligned}\quad (2.82)$$

Al considerar estos términos y juntándolos, la contribución de la violación de sabor a la anomalía $g - 2$ dado a un sleptón \tilde{l} se escribe como

$$\begin{aligned}\frac{g-2}{2} &= \frac{g_c^2}{(4\pi)^2} \left(S_{\tilde{B}\mu,\tilde{l}}^2 - P_{\tilde{B}\mu,\tilde{l}}^2 \right) \frac{2m_{\tilde{B}}m_\mu}{\Delta_{\tilde{l}\tilde{B}}} \left[-\frac{1}{2} - \frac{m_{\tilde{B}}^2}{\Delta_{\tilde{l}\tilde{B}}} - \frac{m_{\tilde{B}}^2 m_{\tilde{l}}^2}{\Delta_{\tilde{l}\tilde{B}}^2} \ln \left(\frac{m_{\tilde{B}}^2}{m_{\tilde{l}}^2} \right) \right] \\ &+ \frac{g_c^2}{(4\pi)^2} \left(S_{\tilde{B}\mu,\tilde{l}}^2 + P_{\tilde{B}\mu,\tilde{l}}^2 \right) \frac{2m_\mu^2}{\Delta_{\tilde{l}\tilde{B}}} \left[\frac{1}{6} - \frac{m_{\tilde{B}}^2}{2\Delta_{\tilde{l}\tilde{B}}} - \frac{m_{\tilde{B}}^4}{\Delta_{\tilde{l}\tilde{B}}^2} - \frac{m_{\tilde{B}}^4 m_{\tilde{l}}^2}{\Delta_{\tilde{l}\tilde{B}}^3} \ln \left(\frac{m_{\tilde{B}}^2}{m_{\tilde{l}}^2} \right) \right],\end{aligned}\quad (2.83)$$

con $g_c^2 = \frac{\tan^2\theta_W g_1^2}{16}$, y $\Delta_{\tilde{l}\tilde{B}} = m_{\tilde{l}}^2 - M_1^2$, conteniendo las cuatro contribuciones con \tilde{l} desde 1, ..., 4 con los valores de los acoplamientos $S_{\tilde{B}\mu,\tilde{l}}, P_{\tilde{B}\mu,\tilde{l}}$ dados por la tabla 2.

En esta base de interacción los términos LR con una explícita dependencia de parámetros libres de SUSY ya que usan los elementos de la matriz de masas de los sleptones.

En la tabla 3, se muestra el espacio para realizar los cálculos analíticos y poder considerar una mezcla de violación de sabor leptónico en términos trilineares suaves del MEMS. Más adelante esto será usado en el sección 5

Tabla 3: El espacio de parámetros donde se realizará el modelo para la red neuronal más adelante. Adaptado de [4]

$\mu_{\text{SUSY}} \in [-15, 15] \text{ TeV}$	$A_0 \in [50, 5000] \text{ GeV}$
$\tilde{m}_S \in [50, 5000] \text{ GeV}$	$\frac{M_1}{\tilde{m}_S} \in [0.2, 5] \text{ GeV}$
$\tan \beta \in [1, 60]$	$w = -1, y = 1$

2.14.5 LA CONTRIBUCIÓN DEL LOOP A LA ANOMALÍA DEL MUON

Como ya hemos visto en la sección 1.11 vamos a mostrar como es la corrección al vértice desde el proceso ya mencionado en este caso, como se muestra en el Apéndice B de [4] dado por

$$\begin{aligned}
\bar{u}(p_1) \Gamma^\mu u(p_2) &= i g_c \bar{u}(p_1) \left[S_{\tilde{B}\mu, \tilde{l}} + P_{\tilde{B}\mu, \tilde{l}} \gamma^5 \right] \\
&\times \frac{1}{(2\pi)^4} \int d^4 k \frac{(\not{k} + m_{\tilde{B}}) \tan \theta_W g_1/4}{D_t} \left[S_{\tilde{B}\mu, \tilde{l}} - P_{\tilde{B}\mu, \tilde{l}} \gamma^5 \right] \\
&\times \frac{1}{D_2} \frac{1}{D_1} [2k + p_1 + p_2]^\mu u(p_2) \\
&= i g_c^2 (S_{\tilde{B}\mu, \tilde{l}}^2 - P_{\tilde{B}\mu, \tilde{l}}^2) m_{\tilde{B}} \bar{u}(p_1) \int \frac{d^4 k}{(2\pi)^4} \frac{(2k + p_1 + p_2)^\mu}{D_t D_1 D_2} u(p_2) \\
&+ g_c^2 \bar{u}(p_1) (S_{\tilde{B}\mu, \tilde{l}}^2 + P_{\tilde{B}\mu, \tilde{l}}^2) \int \frac{d^4 k}{(2\pi)^4} \frac{(2k + p_1 + p_2)^\mu \not{k}}{D_t D_1 D_2} u(p_2) + \dots \\
&= g_c^2 \bar{u}(p_1) \left[(S_{\tilde{B}\mu, \tilde{l}}^2 - P_{\tilde{B}\mu, \tilde{l}}^2) m_{\tilde{B}} B_1^\mu(q^2) + (S_{\tilde{B}\mu, \tilde{l}}^2 + P_{\tilde{B}\mu, \tilde{l}}^2) B_2^\mu(q^2) \right] u(p_2) \\
&+ \dots
\end{aligned} \tag{2.84}$$

donde $q^2 = (p_2 - p_1)^\mu$ y las elipsis representan términos que no contribuyen en la anomalía. Los propagadores son

$$D_t = \frac{1}{k^2 - m_{\tilde{B}}^2}, \tag{2.85}$$

$$D_1 = \frac{1}{(p_1 + k)^2 - m_{\tilde{l}}^2}, \tag{2.86}$$

$$D_2 = \frac{1}{(p_2 + k)^2 - m_l^2}. \quad (2.87)$$

Al hacer que $q^2 = 0$ y asumiendo que la masa del muon es despreciable en comparación con las masas de las partículas dentro del loop, las contribuciones de la anomalía están dadas por

$$B_1^\mu(0) = (p_1 + p_2)^\mu (b_1 + b_2), \quad (2.88)$$

$$B_2^\mu(0) = m_\mu (p_1 + p_2)^\mu (b_2 + b_4), \quad (2.89)$$

donde m_μ es la masa del muon y las funciones escalares $b_{1,2,4}$ son

$$b_1 = -i \frac{1}{(4\pi)^2} \frac{1}{\Delta_{lb}} \left[1 + \frac{m_{\tilde{B}}^2}{\Delta_{lb}} \ln \left(\frac{m_{\tilde{B}}^2}{m_l^2} \right) \right], \quad (2.90)$$

$$b_2 = -i \frac{1}{(4\pi)^2} \frac{1}{\Delta_{lb}} \left[-\frac{1}{2} + \frac{m_{\tilde{B}}^2}{\Delta_{lb}} - \frac{m_{\tilde{B}}^4}{\Delta_{lb}^2} \ln \left(\frac{m_l^2}{m_{\tilde{B}}^2} \right) \right], \quad (2.91)$$

$$b_4 = -i \frac{1}{(4\pi)^2} \frac{1}{\Delta_{lb}} \left[-\frac{m_{\tilde{B}}^6}{\Delta_{lb}^3} \ln \left(\frac{m_l^2}{m_{\tilde{B}}^2} \right) - \frac{1}{2\Delta_{lb}} m_{\tilde{B}}^2 - \frac{2}{3} \Delta_{lb} + \frac{m_{\tilde{B}}^4}{\Delta_{lb}^2} \right], \quad (2.92)$$

donde $\Delta_{lb} = m_l^2 - m_{\tilde{B}}^2$. Reuniendo todas estas piezas, finalmente nuestra contribución a la anomalía del muon se escribe como (véase el Apéndice B de [4])

$$a_\mu = \frac{g_c^2 m_\mu}{(4\pi)^2} \left[\left(S_{\tilde{B}\mu, \tilde{l}}^2 + P_{\tilde{B}\mu, \tilde{l}}^2 \right) \frac{m_\mu}{6m_l^2} F_1^N(x) - \left(S_{\tilde{B}\mu, \tilde{l}}^2 - P_{\tilde{B}\mu, \tilde{l}}^2 \right) \frac{m_{\tilde{B}}}{3m_l^2} F_2^N(x) \right] \quad (2.93)$$

siendo $x = m_{\tilde{B}}^2/m_l^2$ y, como ya hemos definido $g_c^2 = \frac{\tan^2 \theta_W g_1^2}{16}$. Para las funciones $F_{1,2}^N(x)$ se han usado las establecidas en [41]. La ecuación (2.93) es la que usaremos para nuestro modelo de redes neuronales como valores y objetivo.

Parte II

Machine Learning

CAPÍTULO 3

CONCEPTOS BÁSICOS DE MACHINE LEARNING

El problema de buscar patrones en los datos es fundamental y siempre lo ha sido a lo largo de la historia, en cada campo de estudio en la física y en otras ciencias. En la mayoría de aplicaciones, las variables ingresadas originales son preprocesadas para transformarlas en un nuevo espacio de variables, en donde el patrón de reconocimiento sea más fácil de obtener y resolver. El Machine Learning(ML) ha llegado a cambiar la forma en que resolvemos los problemas y encontramos correlaciones para diversidad de sucesos. En este capítulo daremos una introducción a los conocimientos básicos para su introducción al tema.

El *Machine Learning* es el estudio y aplicación de métodos automatizados donde el objetivo es encontrar patrones significativos en complejas estructuras de datos. En particular, el objetivo es ajustar parámetros a un modelo adaptativo para reconocer patrones adecuados para inferencia deductiva, permitiendo predicciones y proposiciones estadísticas más allá de los datos. Un modelo tiene la capacidad de generalizar si completó el objetivo de predecir nuevos ejemplos de datos correctamente.

Hay tres tipos de Machine Learning: El **predictivo** o **supervisado**, el objetivo de este es aprender a mapear desde las entradas \mathbf{x} a las salidas y , dado un par de entradas-salidas etiquetadas $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$, donde \mathcal{D} es llamado el conjunto de entrenamiento, y N el número de ejemplos de entrenamiento. El segundo es llamado **descriptivo** o **no supervisado**. En este solo proporcionamos datos de entrada $\mathcal{D} = \{(\mathbf{x}_i)\}_{i=1}^N$ y el objetivo es encontrar patrones interesantes en los datos. Este es un problema menor bien-definido, ya que no se nos menciona qué clase de patrones hay que encontrar, claramente este no tiene una métrica de error (a diferencia del supervisado que podemos comparar nuestras predicciones \hat{y} para un dado \mathbf{x} al valor observado y). Ya que el contenido de este estudio abarca solo el ML supervisado, nos vamos a centrar en los fundamentos básicos para el mismo. Y por último el **aprendizaje por refuerzo**. Este tipo de aprendizaje, se podría decir que está entre los dos ya mencionados. Aquí el algoritmo no tiene objetivos predefi-

nidos y a los que llegar, si no que debe encontrar las metas óptimas mediante un proceso de prueba y error, donde el algoritmo se premia o penaliza según sea el caso. Su principal objetivo es maximizar la ganancia. Un ejemplo típico es jugar un juego de computadora, tetris, por ejemplo. Podemos usar machine learning supervisado, con esto el algoritmo aprende a jugar con partidas grabadas de los mejores jugadores humanos. En este caso estaríamos proporcionándole un objetivo predefinido y, pero esto no permitiría al algoritmo convertirse en mejor jugador que los mejores jugadores humanos. Si permitimos que el algoritmo descubra objetivos por sí mismo, dentro de las reglas de juego, tendrá la posibilidad de descubrir estrategias de juego nunca jamás usadas por los humanos. La ganancia sería buscar maximizar el puntaje de juego. La búsqueda de objetivos en el aprendizaje por refuerzo es normalmente muy costosa computacionalmente¹.

3.1 ALGUNAS NOCIONES DE MACHINE LEARNING

Feed Forward (propagación hacia adelante). Vamos a empezar con algunos conceptos necesarios para entender un poco sobre **Redes Neuronales o Neural Network (NN)**. Los modelos lineales para regresión y clasificación se basan en una combinación lineal de funciones base no lineales ajustadas $\phi_j(\mathbf{x})$ y toman la forma (3.1). Para este caso vamos a usar formas paramétricas para las funciones base en los cuales los valores de sus parámetros puedan adaptarse durante el entrenamiento. A este tipo de modelos y de los más efectivos en este contexto es el llamado *red neuronal de propagación hacia adelante* (*feed-forward neural network*), también conocida como *perceptrón de multicapa*

$$y(\mathbf{x}, \mathbf{w}) = f \left(\sum_{j=1}^M w_j \phi_j(\mathbf{x}) \right) \quad (3.1)$$

donde \mathbf{x} representa el vector de entrada y \mathbf{w} el conjunto de parámetros adaptativos del modelo. La función $f(\cdot)$ es una función de activación no lineal en problemas de clasificación y la identidad en problemas de regresión. En este caso, se busca extender el modelo introduciendo parámetros ajustables, de modo que estos, junto con los coeficientes w_j , puedan optimizarse durante el entrenamiento. Las redes neuronales emplean funciones base que siguen la forma de (3.1) (i.e., ϕ_j), de tal manera que cada una de ellas constituye una función no lineal de una combinación lineal de las entradas, cuyos coeficientes corresponden a parámetros adaptativos. Esto da lugar al modelo básico de red neuronal, el cual puede describirse como una sucesión de transformaciones funcionales. El procedimiento para ajustar dichos parámetros se presentará más adelante al discutir la retropropagación

¹Este capítulo se basa en textos fundamentales de aprendizaje automático y aprendizaje profundo [5, 6, 42].

del error y el gradiente descendente. Empezamos construyendo M combinaciones lineales de las variables de entrada x_1, \dots, x_D en la forma de

$$a_j = \sum_{i=1}^D w_{ji}^{(1)} x_i + w_{j0}^{(1)}; \quad j = 1, \dots, M. \quad (3.2)$$

El superíndice indica los parámetros correspondientes que están en la primer capa de la red. Les llamaremos “pesos” a los parámetros $w_{ji}^{(1)}$ y sesgos a los $w_{j0}^{(1)}$. Las cantidades a_j se les conocen como “activaciones”, cada una se transforma usando una función de activación no lineal y diferenciable $h(\cdot)$ dando

$$z_j = h(a_j). \quad (3.3)$$

Esto corresponde a las salidas de las funciones base en (3.1), que en el contexto de las NN son llamadas “unidades escondidas”. Las funciones no lineales $h(\cdot)$ son generalmente elegidas para ser funciones sigmoide como la función \tanh por ejemplo. Siguiendo (3.1) los valores vuelven a ser combinados linealmente para dar activaciones de unidades de salida.

$$a_k = \sum_{j=1}^M w_{kj}^{(2)} z_j + w_{k0}^{(2)}; \quad k = 1, \dots, K \text{ (número total de outputs)}. \quad (3.4)$$

y esta transformación corresponde a la segunda capa de la Red, donde $+w_{k0}^{(2)}$ son los parámetros de sesgo. Finalmente, las activaciones de unidades de salida son transformadas usando una función de activación para dar un conjunto de salidas de la Red y_k . La función de activación es determinada por la naturaleza de los datos y la distribución asumida de las variables objetivo. Para problemas estándar de regresión, la función de activación es la identidad, tal que $y_k = a_k$. De manera análoga, en problemas de clasificación binaria múltiple, cada activación de unidad de salida es transformada usando una función logística sigmoide tal que

$$y - k = \sigma(a_k) \quad (3.5)$$

donde

$$\sigma(a) = \frac{1}{1 + e^{(-a)}}. \quad (3.6)$$

Para problemas multiclase, se usa una función de activación softmax ($\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$). Combinamos varias etapas para dar la red en general que, para funciones de activación de unidades de salidas sigmoides, se sigue como

$$y(\mathbf{x}, \mathbf{w}) = \sigma \left(\sum_{j=1}^M w_{kj}^{(2)} h \left(\sum_{i=1}^D w_{ji}^{(1)} x_i + w_{j0}^{(1)} \right) + w_{k0}^{(2)} \right) \quad (3.7)$$

donde el conjunto de todos los pesos y parámetros sesgo han sido agrupados dentro de un vector \mathbf{w} . Entonces, el modelo NN es una función no lineal de un conjunto de variables de entrada $\{x_i\}$ a un conjunto de variables de salida $\{y_k\}$ controlados por un vector \mathbf{w} de parámetros ajustables. Podemos representar esta función con un diagrama de la red como se muestra en la fig. 11.

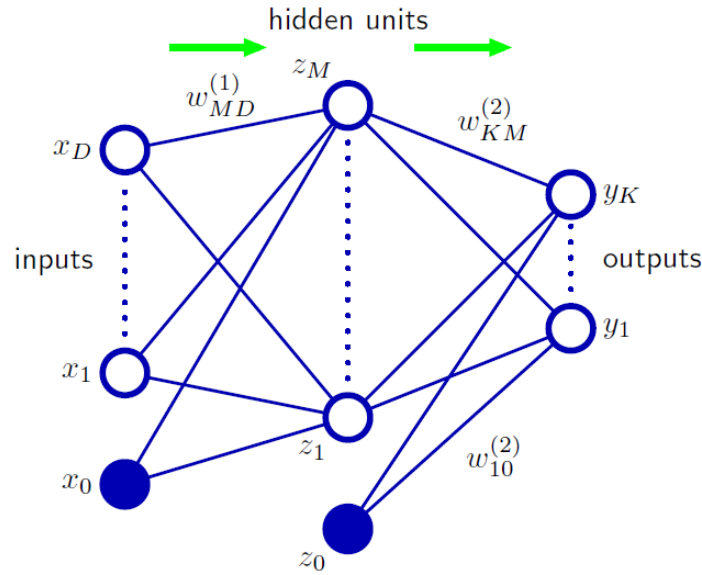


Figura 11: Diagrama de la NN de dos capas correspondiente a (3.7). Las variables de entrada, escondidas, y de salida son representadas por nodos, los parámetros de peso se representan por ligas entre los nodos, en los cuales los parámetros sesgo son denotados por ligas llegando de entradas adicionales y variables escondidas x_0 y z_0 . Las flechas muestran la dirección del flujo de información en la red durante la propagación hacia adelante. Adaptado de [5, p. 228].

El proceso de evaluar la ecuación (3.7) es lo que hemos llamado *feed forward* de la información en la red. Más adelante veremos como darle una interpretación probabilística a esta red neuronal. Una forma de absorber el parámetro de sesgos (bias) de la ecuación (3.2) es definir $x_0 = 1$ así toma la siguiente forma

$$a_j = \sum_{i=0}^D w_{ji}^{(1)} x_i. \quad (3.8)$$

Análogamente los sesgos de la segunda capa pueden ser absorbidos, la función de la red en general se transforma en

$$y_k(\mathbf{x}, \mathbf{w}) = \sigma \left(\sum_{j=0}^M w_{kj}^{(2)} h \left(\sum_{i=0}^D w_{ji}^{(1)} x_i \right) \right), \quad (3.9)$$

$$y_k^{(l)}(\mathbf{x}, \mathbf{w}) = \sigma \left(\sum_{j=0}^{n^{(l)}} w_{kj}^{(l)} h^{(l-1)} \left(\sum_{i=0}^{n^{(l-1)}} w_{ji}^{(l-1)} a_i^{(l-1)} \right) \right). \quad (3.10)$$

En la primera ecuación de (3.9) para el caso de dos capas, y (3.10) para el caso generalizado.

Es importante mencionar que las funciones de la red neuronal son diferenciables con respecto a los parámetros de la red, esta propiedad tomará un papel importante en el entrenamiento de la red. Para terminología de la red se recomienda el tipo del cual la red de la fig. 3 se considera de red de dos capas, ya que es el número de capas de los pesos adaptativos, la primera son solo los datos de entrada. Las propiedades de las feed forward networks han sido ampliamente estudiadas en [43–50]. Por lo tanto las redes neuronales son aproximadores universales. Para ilustrar, un ejemplo es, una red neuronal de dos capas con salidas lineales que puede aproximar uniformemente cualquier función continua en un dominio compacto de entrada, para precisión arbitraria, la red neuronal tendría un considerable número de unidades ocultas(hidden units). Entonces el problema reside en cómo encontrar el conjunto de parámetros adecuado dado unos datos de entrenamiento. Más adelante mostraremos soluciones basadas en máxima verosimilitud que se ajustan bien a esta problemática.

Haremos un pequeño espacio para hablar de **regresión lineal**, para dar un punto de vista más amplio, esto afirma que la respuesta de regresión es una función lineal de las entradas, totalmente familiar a (3.2)

$$y(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + \epsilon = \sum_{j=1}^D w_j x_j + \epsilon \quad (3.11)$$

donde $\mathbf{w}^T \mathbf{x}$ representa el producto escalar entre el vector de entrada \mathbf{x} y el vector de peso del modelo \mathbf{w}^T , ϵ representa el **error residual** entre nuestras predicciones y la respuesta verdadera. Se suele asumir que ϵ tiene una distribución Gaussiana o normal. Para hacer esta conexión más evidente, podemos reescribir el modelo como sigue:

$$p(y | \mathbf{x}, \boldsymbol{\theta}) = \mathcal{N}(y | \mu(\mathbf{x}), \sigma^2(\mathbf{x})), \quad (3.12)$$

donde $p(y | \mathbf{x}, \boldsymbol{\theta})$ denota la densidad condicional de la variable objetivo y dada la entrada \mathbf{x} y los parámetros del modelo $\boldsymbol{\theta}$; en el caso de la regresión lineal, $\boldsymbol{\theta} = (\mathbf{w}, \sigma^2)$, donde \mathbf{w} es el vector de pesos y σ^2 la varianza del ruido, si esta se considera explícitamente. Además, $\mathcal{N}(y | \mu(\mathbf{x}), \sigma^2(\mathbf{x}))$ corresponde a una distribución normal con media $\mu(\mathbf{x})$ y varianza $\sigma^2(\mathbf{x})$. De esta forma, se observa que $p(y | \mathbf{x}, \boldsymbol{\theta})$ es una densidad de probabilidad

condicional. En el caso más simple, $\mu(\mathbf{x})$ es una función lineal de \mathbf{x} , es decir, $\mu = \mathbf{w}^T \mathbf{x}$, y el ruido se considera fijo, por lo que $\sigma^2(\mathbf{x}) = \sigma^2$. Así, $\theta = (\mathbf{w}, \sigma^2)$ corresponde al conjunto de parámetros del modelo. Si se supone, por ejemplo, que la entrada es de una dimensión, la respuesta esperada es

$$\mu(\mathbf{x}) = w_0 + w_1 x = \mathbf{w}^T \mathbf{x}, \quad (3.13)$$

donde w_0 es el bias o sesgo, w_1 es la pendiente, con el vector $\mathbf{x} = (1, x)$. La regresión lineal puede modelar relaciones no lineales reemplazando \mathbf{x} por una función no lineal de las entradas, $\phi(\mathbf{x})$, lo que nos da

$$p(y | \mathbf{x}, \theta) = \mathcal{N}(y | \mathbf{w}^T \phi(\mathbf{x}), \sigma^2). \quad (3.14)$$

Esto es conocido como **función de expansión base**. Muchos métodos de ML se pueden ver como maneras diferentes de estimar funciones base desde los datos.

Otro punto importante sobre estos modelos es evitar algo llamado **sobreajuste** de los datos, esto significa tratar de evadir el modelar cada pequeña variación en la entrada, esto es más probable que termine en ruido que una señal verdadera.

3.2 ENTRENAMIENTO DE LA RED

Ahora necesitamos una forma de determinar los parámetros de la red, esto se puede hacer minimizando una función de error, en este caso el error cuadrático. Dado un conjunto de entrenamiento formado por un conjunto de vectores de entrada $\{\mathbf{x}_n\}$, donde $n = 1, \dots, N$, junto con un conjunto de vectores objetivo $\{\mathbf{t}_n\}$, minimizamos la función de error

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \|\mathbf{y}(\mathbf{x}_n, \mathbf{w}) - \mathbf{t}_n\|^2. \quad (3.15)$$

También podemos darle un vistazo más general al entrenamiento de la red al determinar una interpretación probabilística de las salidas de la red. Considerando una sola variable objetivo t que puede tomar valores reales. Asumiendo que tiene una distribución Gaussiana con un promedio x dependiente, dado por la salida de la red neuronal

$$p(t | \mathbf{x}, \mathbf{w}) = \mathcal{N}(t | y(\mathbf{x}, \mathbf{w}), \beta^{-1}), \quad (3.16)$$

donde β es la precisión (varianza inversa) del ruido Gaussiano. Para esta distribución condicional, podemos tomar la función de activación de salida a ser la identidad (lo usual en una red de regresión), ya que dicha red puede aproximarse a cualquier función conti-

nua de \mathbf{x} a y . Dado un un conjunto de datos independientes N , muestras idénticamente distribuidas $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, con sus correspondientes valores objetivo $\mathbf{t} = \{t_1, \dots, t_N\}$ podemos construir su función de distribución de probabilidad

$$p(\mathbf{t} | \mathbf{X}, \mathbf{w}, \beta) = \prod_{n=1}^N p(t_n | \mathbf{x}_n, \mathbf{w}, \beta). \quad (3.17)$$

Tomando el logaritmo negativo, obtenemos la función de error

$$\frac{\beta}{2} \sum_{n=1}^N \{y(\mathbf{x}_n, \mathbf{w}) - t_n\}^2 - \frac{N}{2} \ln \beta + \frac{N}{2} \ln(2\pi) \quad (3.18)$$

que se puede usar para aprender los parámetros \mathbf{w} y β . En la literatura se suele usar más la minimización de la función de error que la maximización de la verosimilitud, nosotros vamos a seguir esta convención. Primero vamos a determinar \mathbf{w} . Cabe mencionar que maximizar la función de verosimilitud es equivalente a minimizar la función de error de suma de cuadrados dada por

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(\mathbf{x}_n, \mathbf{w}) - t_n\}^2. \quad (3.19)$$

El valor que minimiza $E(\mathbf{w})$ será denotado como \mathbf{w}_{ML} que corresponde a la solución de máxima verosimilitud. En la práctica la no linealidad de la función de la red $y(\mathbf{x}_n)$, \mathbf{w} provoca que $E(\mathbf{w})$ sea no convexo, por lo que se puede encontrar un máximo local de la verosimilitud, lo que conlleva a un mínimo local de la función de error. Esta función tiene una propiedad que nos será útil más adelante²

$$\frac{\partial E}{\partial a_k} = y_k - t_k. \quad (3.20)$$

3.2.1 OPTIMIZACIÓN DE PARÁMETROS

Nuestra tarea es encontrar el vector \mathbf{w} que minimice la función $E(\mathbf{w})$. Podemos visualizar este objetivo en la fig. 12. Al hacer un desplazamiento infinitesimal en el espacio de pesos de \mathbf{w} a $\mathbf{w} + \delta\mathbf{w}$ el cambio en la función de error sería $\delta E \approx \delta\mathbf{w}^T \nabla E(\mathbf{w})$, donde el vector $\nabla E(\mathbf{w})$ apunta en la dirección de mayor tasa de incremento de la función de error. Debido a que el error $E(\mathbf{w})$ es una función suave y continua de \mathbf{w} , su valor más pequeño

²El entrenamiento de redes neuronales (y de la optimización de parámetros) supervisadas se describe ampliamente en [5, 6].

ocurre donde el gradiente de la función de error se hace cero, entonces

$$\nabla E(\mathbf{w}) = 0 \quad (3.21)$$

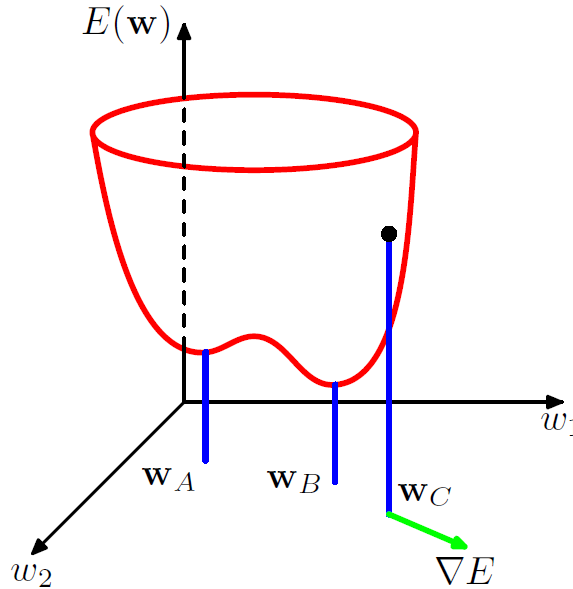


Figura 12: Vista geométrica de la función de error $E(\mathbf{w})$ como una superficie sobre el espacio de pesos. El punto \mathbf{w}_A representa un mínimo local, mientras que \mathbf{w}_B es el mínimo global. En cualquier punto \mathbf{w}_C , el gradiente local de la superficie de error está dado por el vector ∇E . Imagen adaptada de [5].

Para una aplicación aceptable de las redes neuronales no es necesario encontrar un mínimo global (y en general no se sabrá si se ha encontrado el mínimo global), pero es necesario comparar varios mínimos locales para encontrar una solución lo suficientemente buena. Por la naturaleza de nuestro problema vamos a recurrir a un procedimiento iterativo y numérico. La mayoría de técnicas involucran elegir algún valor inicial \mathbf{w}^0 para el vector de peso y moverse en una sucesión de pasos en la forma

$$\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} + \Delta \mathbf{w}^{(\tau)}, \quad (3.22)$$

donde τ representa el paso de iteración.

Optimización por Gradiente Descendiente. El acercamiento más simple es la información del gradiente, que mencionamos en (3.22) al elegir una actualización de los

pesos en un pequeño paso en la dirección del gradiente negativo, tal que

$$\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} - \eta \nabla E(\mathbf{w}^{(\tau)}) \quad (3.23)$$

donde $\eta > 0$ es conocido como *coeficiente de aprendizaje*. Después de cada actualización, el gradiente es re evaluado para el nuevo vector de peso y se repite el proceso. Las técnicas que utilizan todos los datos son llamados métodos por *lotes*. En cada paso el vector de peso es movido en la dirección de la mayor tasa de decrecimiento de la función de error, a esto se le conoce como *gradiente descendiente*. Este método resulta no ser el algoritmo más adecuado, para eso usaremos una versión mejorada que mostraremos más adelante. Pero antes de pasar a nuestro algoritmo principal, hay que mencionar que existe una versión en línea que ha probado ser muy útil para entrenar largo conjuntos de datos [51]. Las funciones de error basadas en máxima verosimilitud para un conjunto de muestras independientes comprenden una suma de términos, uno por cada punto de datos

$$E(\mathbf{w}) = \sum_{n=1}^N E_n(\mathbf{w}). \quad (3.24)$$

El descenso por gradiente en línea, conocido como *descenso por gradiente estocástico*, realiza una actualización al vector de pesos basada en un solo punto de datos a la vez en la ecuación (3.23). Esto se repite ciclando a través de los datos ya sea en secuencia o seleccionando punto aleatorios con reemplazo. Existen escenarios intermedios en los que estas actualizaciones se basan en lotes de puntos de datos.

Algoritmo de Gradiente Adaptativo(AdaGrad). Otra forma de mejorar la tasa de convergencia es adaptar η durante el entrenamiento. Para esto escribimos la ecuación (3.23) en forma de componentes,

$$\begin{bmatrix} w_{t+1}^{(1)} \\ w_{t+1}^{(2)} \\ \vdots \\ w_{t+1}^{(M)} \end{bmatrix} = \begin{bmatrix} w_t^{(1)} \\ w_t^{(2)} \\ \vdots \\ w_t^{(M)} \end{bmatrix} - \begin{bmatrix} \eta g_t^{(1)} \\ \eta g_t^{(2)} \\ \vdots \\ \eta g_t^{(M)} \end{bmatrix}, \quad (3.25)$$

la ecuación anterior muestra la actualización escrita por componentes, donde $g_t^{(i)}$ denota la i -ésima componente del gradiente en la iteración t , mientras que η corresponde a una única tasa de aprendizaje compartida por todos los parámetros. Se puede ver que cada elemento del gradiente comparte el mismo η . Digamos que por ejemplo si $g_t^{(1)} \gg g_t^{(2)}$, con $g_t^{(1)}$ siendo muy grande que sobre pasa del mínimo y diverge en la dimensión de $w^{(1)}$, mientras que $g_t^{(2)}$ es muy pequeño y no avanza mucho en la dimensión de $w^{(2)}$. Para resolver esto podemos agregar un escalado por parámetro a la tasa de aprendizaje que ajuste las

componentes del gradiente independientemente. Conocido como *AdaGrad* escalamos de la siguiente manera

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \left(\frac{\eta}{\sqrt{\text{diag}(\mathbf{G}_t + \varepsilon \mathbf{I})}} \right) \odot \mathbf{g}_t, \quad (3.26)$$

aquí \odot indica producto componente a componente (producto de Hadamard), G_t es una matriz diagonal en la que cada elemento es la suma de los cuadrados de los gradientes anteriores hasta el paso actual t , $G_t = \sum_{\tau=1}^t \mathbf{g}_\tau \mathbf{g}_\tau^T$. Agregamos un valor muy pequeño ε , para evitar dividir por cero. Para un peso dado w^i , g^i , el algoritmo queda como

$$w_{t+i}^{(i)} = w_t^{(i)} - \frac{\eta}{\sqrt{g_1^2 + g_2^2 \dots g_t^2 + \varepsilon}} g_t^{(i)}, \quad (3.27)$$

donde $g_t = \left(\frac{\partial E(\mathbf{w})}{\partial w^{(i)}} \right)_{\tau=1}^t$ es la derivada parcial calculada para la t -ésima iteración, que corresponde al peso w^i . Las desventajas de este método es que la tasa de aprendizaje decrece monotónicamente cuando el denominador acumula muchos términos positivos.

Propagación de la Raíz Cuadrada Media(RMSprop). Una solución a la desventaja de AdaGrad es proporcionada por el método RMSprop, que en lugar de agregar los cuadrados de los gradientes, usa promedio exponencial o media exponencial. Ya que uno utiliza el promedio para normalizar más que para agregar valor, el progreso no es reducido prematuramente por un factor de escalamiento, en pocas palabras usamos un factor de decaimiento $\beta \in (0, 1)$, y ponderamos las derivadas parciales al cuadrado que ocurrieron t actualizaciones atrás con β^t . Esto se puede lograr de manera sencilla al multiplicar el agregado cuadrado actual (la estimación en curso) por β y sumando $(\beta - 1)$ veces la derivada parcial (al cuadrado). La estimación actual se inicializa en cero. Esto causa sesgos en las primeras iteraciones, pero desaparecen a largo plazo. Entonces nuestra actualización queda como sigue

$$E[g^2]_{t+1} = \rho E[g^2]_t + (1 - \rho) g_t^2, \quad (3.28)$$

o de forma más compacta

$$v_{t+1} = \rho v_t + (1 - \rho) g_t^2, \quad (3.29)$$

y así el gradiente se actualiza como

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{v_{t+1} + \varepsilon}} g_t. \quad (3.30)$$

Estimación Adaptativa de Momentos(ADAM). Ya hemos visto algunos métodos del gradiente, todos estos tienen relación con el que veremos ahora, el cual también es el método que usaremos en esta tesis, llamado *ADAM* [52]. Este algoritmo utiliza una normalización “señal-ruido” muy similar a AdaGrad y RMSprop; No obstante también suaviza exponencialmente el gradiente de primer orden para incorporar el momento en la actualización. Además, aborda el sesgo en el suavizado exponencial cuando la estimación en curso de un valor suavizado se inicializa en cero (lo cual no es deseable en la práctica). Como en RMSprop usamos la media exponencial de la ecuación (3.29). Asimismo, se mantiene un valor suavizado exponencialmente del gradiente, cuyo componente t -ésimo se denota por m_t . Utilizamos un parámetro de suavizado diferente $\rho_f \in [0, 1)$

$$\mathbf{m}_{t+1} = \rho_f \mathbf{m}_t + (1 - \rho_f) \mathbf{g}_t, \quad (3.31)$$

además se introduce una corrección estadística de los sesgos en el primer y segundo momento

$$\hat{\mathbf{m}}_{t+1} = \frac{\mathbf{m}_{t+1}}{1 - \rho_f^{t+1}}, \quad (3.32)$$

$$\hat{\mathbf{v}}_{t+1} = \frac{\mathbf{v}_{t+1}}{1 - \rho^{t+1}}, \quad (3.33)$$

por último nuestro gradiente se actualiza de la siguiente manera

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \frac{\eta}{\sqrt{\hat{\mathbf{v}}_{t+1} + \epsilon}} \odot \hat{\mathbf{m}}_{t+1}. \quad (3.34)$$

3.3 RETROPROPAGACIÓN DE ERROR

Nuestro objetivo ahora es encontrar una técnica para evaluar el gradiente de una función de error $E(\mathbf{w})$ para una red de propagación hacia adelante. Veremos que esto puede lograrse con un esquema local de mensajes en el cual la información se envía alternadamente hacia adelante y hacia atrás a través de la red, y es conocida como **retropropagación de error**. Vamos a derivar el algoritmo de retropropagación para una red general con una topología feed-forward arbitraria, funciones no lineales arbitrarias y diferenciales, y una amplia clase de funciones de error. Muchas funciones de error de interés, comprenden una suma de términos como vimos en la ecuación (3.24). Aquí vamos

a evaluar $\nabla E_n(\mathbf{w})$ para un término en la función de error. Esto puede ser utilizado para una optimización secuencial, o acumulados sobre un conjunto de entrenamiento para el método por lotes. Considere el siguiente modelo lineal con salidas y_k , con combinaciones lineales de las variables de entrada x_i

$$y_k = \sum_i w_{ki} x_i, \quad (3.35)$$

con una función de error, para un patrón de entrada n , toma la siguiente forma

$$E_n = \frac{1}{2} \sum_k (y_{nk} - t_{nk})^2 \quad (3.36)$$

donde $y_{nk} = y_k(\mathbf{x}_n, \mathbf{w})$. El gradiente de la función de error con respecto a un peso w_{ji} es

$$\frac{\partial E_n}{\partial w_{ji}} = (y_{nj} - t_{nj}) x_{ni}. \quad (3.37)$$

En una red feed-forward en general, cada unidad calcula una suma ponderada de sus entradas de la forma

$$a_j = \sum_i w_{ji} z_i, \quad (3.38)$$

donde z_i es la activación de una unidad, o entrada, que envía una conexión a la unidad j , donde w_{ji} es el peso que asocia a esa conexión. La suma en (3.38) es transformada por una función de activación no lineal $h(\cdot)$ que nos da la activación z_j de la unidad j

$$z_j = h(a_j). \quad (3.39)$$

Todo esto es el proceso tradicional que ya vimos en la sección 3.1. Ahora consideremos la evaluación de la derivada de E_n con respecto al peso únicamente a través de la entrada sumada a_j a la unidad j . Aplicamos entonces la regla de la cadena

$$\frac{\partial E_n}{\partial w_{ji}} = \frac{\partial E_n}{\partial a_j} \frac{\partial a_j}{\partial w_{ji}}. \quad (3.40)$$

Introduciremos la notación

$$\delta_j = \frac{\partial E_n}{\partial a_j} \quad (3.41)$$

donde a los δ se les denomina como *errores*. Usando (3.38) escribimos

$$\frac{\partial a_j}{\partial w_{ji}} = z_i. \quad (3.42)$$

Substituyendo lo anterior llegamos a

$$\frac{\partial E_n}{\partial w_{ji}} = \delta_j z_i. \quad (3.43)$$

Esta última ecuación nos muestra que para obtener la derivada que buscamos solamente necesitamos multiplicar δ de la unidad en el extremo de la salida del peso por el valor de z de la unidad en el extremo de la entrada de peso (con $z = 1$ en el caso de un sesgo o bias). Entonces solo necesitamos calcular el valor de δ_j para cada unidad oculta y de salida de la red, y luego aplicar (3.42). Como ya hemos visto, para las unidades de salida tenemos

$$\delta_k = y_k - t_k \quad (3.44)$$

siempre que estemos utilizando el enlace canónico como la función de activación de salida. Para evaluar δ_j en las unidades ocultas, hacemos

$$\delta_j \equiv \frac{\partial E_n}{\partial a_j} = \sum_k \frac{\partial E_n}{\partial a_k} \frac{\partial a_k}{\partial a_j}, \quad (3.45)$$

donde la suma corre sobre todas las unidades k a las cuales la unidad j les envía conexiones.

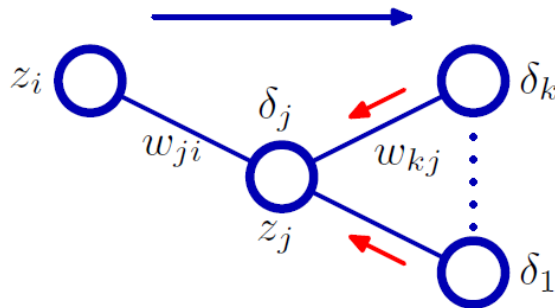


Figura 13: Ilustración del cálculo de δ_j para una unidad oculta j mediante backpropagation desde las unidades k a las que j envía conexiones. Tomado de [5].

El arreglo de unidades y pesos se muestra en la fig. 13. Cabe mencionar que las unidades k podrían incluir otras unidades ocultas o salidas. Al usar (3.45) contemplamos en hecho de que las variaciones de a_j dan lugar a variaciones en la función de error solo a través de variaciones de en las variables a_k . Sustituyendo la definición (3.41) en (3.45), y usando (3.38) y (3.39), obtenemos la fórmula para la retropropagación o backpropagation

$$\delta_j = h'(a_j) \sum_k w_{kj} \delta_k. \quad (3.46)$$

Obsérvese que la suma se toma sobre el primer índice w_{kj} (correspondiente a la retropropagación de la información a través de la red) a diferencia de una propagación feed-forward como se muestra en la ecuación (3.47) que se toma sobre el segundo índice y se muestra en la fig. 14

$$z_k = h\left(\sum_j w_{kj} z_j\right). \quad (3.47)$$

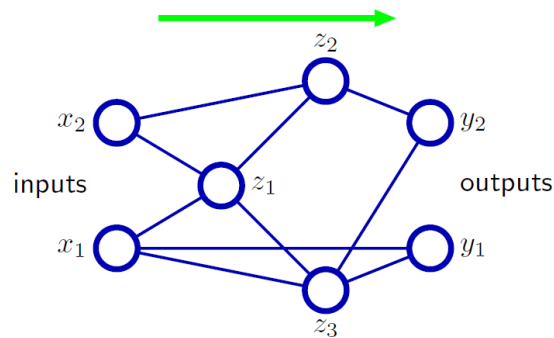


Figura 14: Ejemplo de una red neuronal con una topología general de tipo feed-forward. Tomado de [5].

Debido a que ya sabemos los valores de las δ para las unidades de salida, solo es cuestión de aplicar (3.46) recursivamente, esto puede ser evaluado para todas las unidades ocultas en una red feed-forward, sin importar su topología. Ahora un breve resumen del procedimiento de la retropropagación.

Retropropagación de error.

- Aplica un vector de entrada \mathbf{x}_n a la red y propaga hacia adelante por medio de la red usando (3.38) y (3.39) para encontrar las activaciones de todas las unidades ocultas y de salida.
- Evalúa el δ_k para todas las unidades ocultas usando (3.44).
- Retropropaga las δ usando (3.46) para obtener el δ_j en cada unidad oculta dentro de la red.
- Usa (3.43) para finalmente evaluar las derivadas.

Para un método por lotes, que es el que se usará en esta tesis, solo es cuestión de repetir los pasos anteriores para cada patrón en el conjunto de entrenamiento y sumar sobre todos los patrones³:

$$\frac{\partial E}{\partial w_{ji}} = \sum_n \frac{\partial E_n}{\partial w_{ji}}. \quad (3.48)$$

3.4 MÉTRICAS DE ERROR

Por la naturaleza del problema que hemos estado desarrollando es fácil intuir que necesitamos métricas adecuadas que puedan evaluar en general el rendimiento del modelo, esto es, medir la distancia entre una predicción \hat{y} y un valor real y con exactitud. Usaremos tres métricas para este proceso, una de ellas ya la hemos estado manejando (ecuación (3.19)), pero en su forma no promediada, esta nueva pero familiar métrica es el **error cuadrado medio(MSE)** dado como

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (3.49)$$

Mientras más bajo sea el MSE, mejor será nuestro modelo, claramente esto puede ser dependiente de la escala en la que estemos trabajando para ello hay que poner atención en escalar adecuadamente nuestros datos ya que no tendría mucho sentido comparar valores del MSE con conjuntos de datos de diferentes escalas, este es un tema que se abordará más adelante. Pero en cuestión un valor muy bajo indica un modelo favorable una vez tomado en cuenta el factor del orden de magnitud en el qué estamos trabajando.

Nuestra segunda métrica será la del **error absoluto medio(MAE)** y se define de la siguiente manera

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|. \quad (3.50)$$

A diferencia del MSE este no penaliza tanto los errores grandes, nos da una visión más intuitiva ya que simplemente nos da el error promedio del modelo, al igual que el MSE buscamos un valor muy cercano a cero para indicarnos que tenemos un buen modelo. Es necesario aclarar que usualmente en la literatura hacen referencia a *función de costo* o *función de error*, la función de costo representa la media de las pérdidas, mientras que

³La retropropagación del error constituye el mecanismo central para actualizar los parámetros de la red [6, 51].

la función de error está definida para un solo punto de datos, básicamente la función de costo es la ecuación (3.49) o (3.50), algo que se especifica previamente en cada análisis.

La última métrica que usaremos será el **coeficiente de determinación** (R^2), el cual compara la suma de cuadrados residuales por la suma total de cuadrados y se define como

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (3.51)$$

donde y_i es el valor observado de la variable objetivo para la i -ésima observación, \hat{y}_i es el valor predicho por el modelo, \bar{y} es la media de los valores observados y n es el número total de observaciones. Así, el numerador representa la suma de cuadrados residuales, mientras que el denominador corresponde a la suma total de cuadrados. Otra forma de interpretar este coeficiente es con el MSE y la varianza de los objetivos

$$R^2 = 1 - \frac{MSE}{\text{Var}(y)} = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}, \quad (3.52)$$

en un caso unidimensional, la varianza se puede interpretar como las distancias cuadráticas medias a una línea horizontal, donde $y = \bar{y}$ representa el promedio. El MSE sería la distancia cuadrática media a la línea del modelo \hat{y} . Con esto, podemos entender a esta métrica como la medida de qué tanto la distancia cuadrática desde los puntos de datos al promedio es menor o mayor que las distancias cuadráticas del modelo. Un R^2 cerca de zero indica un modelo muy deficiente. Lo podemos interpretar como la precisión de la línea horizontal promedio siendo similar a la precisión de tu modelo, $MSE \approx \text{Var}(y)$, y $R^2 \approx 0$. Por lo tanto un, R^2 cerca de uno es un buen indicador de un gran rendimiento, quiere decir que el modelo de regresión es un buen ajuste de los datos observados, $MSE \ll \text{Var}(y)$, entonces $R^2 \approx 1$. A pesar de esto el coeficiente de determinación tiene ciertas limitaciones. Por ejemplo si un MSE es más pequeño que la varianza, e.g. $MSE = 0.99\text{Var}(y)$, podría dar $R^2 = 0.01$. Esto no nos indica que es un mal modelo aunque a priori se podría pensar que sí. Aún cuando haya una pequeña diferencia en rendimiento entre la línea horizontal promedio y la línea del modelo, eso es completamente relativo al problema. El mismo tipo de limitaciones puede aplicar para un alto valor de R^2 . Con el fin de tomar precauciones entre las desventajas y ventajas de cada métrica, vamos a usar con precaución un conjunto de ellas para no caer en las desventajas de una en particular y tener un análisis más amplio de nuestros resultados.

3.5 FUNCIONES DE ACTIVACIÓN

Hay una diversidad enorme en el tema de las funciones de activación que pueden ser usadas ya sea para activar cada unidad oculta de la red para imponer no linealidad o transformar la salida final dentro de un rango sensible para motivos de clasificación. En esta tesis, aunque haremos uso de algunas de ellas para el primer caso, nos concentraremos en una sola, debido a que resulta ser la más efectiva para este objetivo de estudio. Cabe mencionar, que hay otro tipo de función de activación que se usa para las salidas finales en modelos de regresión, esta es obviamente como mencionamos en la sección 3.1; la identidad. En este punto es normal preguntarse, ¿Qué función de activación debo usar? Bueno este tema puede ser abordado de muchas maneras que hay en la literatura las cuales no cubriremos en este trabajo. Pero vamos a hablar brevemente de dos conceptos que suelen mencionarse sobre el tema: *explosión de gradientes* y el *problema del gradiente desaparecido*. El primero se refiere a un aprendizaje divergente ya que los pesos en cada paso del proceso iterativo(al actualizarlos) tienden al infinito. El segundo es lo opuesto, nos habla de que una vez cubramos la retropropagación nos arroje muchos gradientes cero, lo cual significa que la red no aprenderá. Elegir la función de activación correcta para las capas ocultas, es crucial para evitar estos problemas.

Tabla 4: Funciones de activación que serán usadas en este trabajo: ReLU, ELU, Swish y SoftPlus con sus respectivas derivadas.

Nombre	$f(x)$	$f'(x)$
ReLU	$\begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{si } x > 0 \end{cases}$	$\begin{cases} 0 & \text{si } x \leq 0 \\ 1 & \text{si } x > 0 \end{cases}$
S(ELU)	$\begin{cases} \lambda x & \text{si } x > 0 \\ \lambda \alpha (e^x - 1) & \text{si } x \leq 0 \end{cases}$	$\begin{cases} \lambda & \text{si } x > 0 \\ \lambda \alpha e^x & \text{si } x \leq 0 \end{cases}$
Swish	$\frac{x}{1+e^{-x}}$	$\frac{e^{-x}}{(1+e^{-x})^2}$
Softplus	$\ln(1 + e^x)$	$\frac{e^x}{1+e^x}$

En la tabla 4 podemos ver las diferentes funciones de activación, adaptadas de [53–56] que utilizaremos en este trabajo junto con sus derivadas, centrándonos en ReLU como nuestra principal función por motivos que veremos en la sección 5

$$\text{ReLU}(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{si } x > 0 \end{cases}. \quad (3.53)$$

3.6 CONJUNTOS DE ENTRENAMIENTO, PRUEBA Y VALIDACIÓN (TRAIN, TEST AND VALIDATION SETS)

Es momento de hablar de los datos, en cómo manejarlos a la hora de entrenar nuestro modelo, porque si usáramos todos los pares de entrada-salida $\mathcal{D} = (\mathbf{X}, \mathbf{y}) = \{(\mathbf{x}_i, y)\}_{i=1}^n$ que tenemos disponibles para entrenar el modelo no tendríamos datos para probar el modelo. Recordemos que en modelos así no nos sirve probar con los mismos datos que ya ha sido entrenado, esto puede significar sobreajuste, un tema que tocaremos más adelante, por lo tanto buscamos ver qué tan predictivo es el modelo con datos nuevos, en función de su buen rendimiento nos dirá si aprendió las abstracciones y patrones del objeto de estudio dentro de los datos o no.

3.6.1 CONJUNTOS DE PRUEBA Y ENTRENAMIENTO

Para determinar la habilidad de un modelo en predecir nuevos datos, se puede extraer una parte de ellos mediante un muestreo aleatorio. Esto se le conoce como *datos de prueba*, un subconjunto $\mathcal{D}_{test} \subset \mathcal{D}$ (donde \mathcal{D} es el conjunto de los datos) el cual no se usa durante el entrenamiento. La parte restante que sí es usada en el entrenamiento se le llama *datos de entrenamiento* $\mathcal{D}_{train} \subset \mathcal{D}$.

Daremos un ejemplo para ilustrar estas ideas, el objetivo será predecir lluvias: Digamos que quieres usar un conjunto de datos \mathbf{X} de n filas de datos, donde cada fila de datos, llamada *ejemplo*, contiene un conjunto de p características. Las cuales pueden ser presión atmosférica, precipitación, etc.

Cada ejemplo en \mathbf{X} tiene su correspondiente salida en \mathbf{y} , la cuál es la precipitación en el día siguiente después del día donde el ejemplo fue sampleado. Las medidas junto con las salidas forman el conjunto de datos $\mathcal{D} = (\mathbf{X}, \mathbf{y})$. Ahora, lo que se busca es usar estos datos para predecir el clima de mañana, *no vistos*. Claramente no tenemos acceso al clima de mañana, si queremos ver qué tan bien predice el modelo a un clima no visto antes de mañana, podemos intentar tratando algunos de tus datos como *no vistos*. Deberías entonces separar \mathcal{D} a lo largo de las filas, nuestra matriz \mathbf{X} queda así⁴

⁴Consultar [5, 6, 57] para más información ya que es de dónde está basada esta sección

$$\begin{array}{l} 1 \\ 2 \\ \vdots \\ n \end{array} \left[\begin{array}{cccc} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{array} \right] \left. \begin{array}{l} \\ \\ \\ \end{array} \right\} \begin{array}{l} X_{\text{train}} \\ \\ \\ X_{\text{test}} \end{array}$$

donde $\mathcal{D}_{\text{train}} = (\mathbf{X}_{\text{train}}, y_{\text{train}})$ y $\mathcal{D}_{\text{test}} = (\mathbf{X}_{\text{test}}, y_{\text{test}})$. Una opción usual que se usa normalmente es 80% los datos de entrenamiento, y 20% los datos de prueba. Pero este porcentaje puede ser relativo dependiendo del problema. Ahora podemos ajustar el modelo a los datos $\mathcal{D}_{\text{train}}$, siguiendo un conjunto de parámetros $\hat{\mathbf{w}}$. Por ejemplo podríamos usar los mínimos cuadrados ordinarios para estimar los parámetros $\hat{\mathbf{w}} = (\mathbf{X}_{\text{train}}^T \mathbf{X}_{\text{train}})^{-1} \mathbf{X}_{\text{train}}^T \mathbf{y}_{\text{train}}$ y así hemos ajustado el modelo a los datos de entrenamiento, ahora para predecir objetivos dentro y fuera de las muestras haciendo lo siguiente

$$\begin{aligned} \hat{y} &= \mathbf{w}^* \mathbf{X}_{\text{test}} \\ \tilde{y} &= \mathbf{w}^* \mathbf{X}_{\text{train}} \end{aligned} \quad (3.54)$$

podemos usar las métricas que ya hemos establecido para calcular el *error de prueba o error de generalización*

$$\text{MSE}_{\text{test}} = \frac{1}{n} \sum_{i=1}^n (y_i^{\text{test}} - \hat{y}_i)^2, \quad (3.55)$$

y el *de entrenamiento o error empírico*

$$\text{MSE}_{\text{train}} = \frac{1}{n} \sum_{i=1}^n (y_i^{\text{train}} - \tilde{y}_i)^2. \quad (3.56)$$

Comparar estos dos errores es crucial para el uso del Machine Learning, de otra forma sería muy complicado saber si construimos un buen modelo.

3.6.2 HIPERPARÁMETROS

Los *hiperparámetros* son cantidades que se ajustan manualmente a diferencia de los parámetros que los aprende el propio modelo. Puede haber un rango grande de hiperparámetros en un modelo, los más comunes son: parámetros de regularización, coeficiente de aprendizaje η para el proceso de optimización de pesos, por ejemplo, los *epochs* que

representan cuantas veces el modelo recorre todos los datos de entrenamiento, el *tamaño de lotes* o *batch size* que establece cuantos ejemplos procesa el modelo antes de actualizar los pesos, la propia estructura de la red; su número de capas y neuronas, por ejemplo un número M de neuronas, con esto podemos controlar, por ejemplo, el número de parámetros, entonces podríamos esperar una máxima verosimilitud ajustando un valor óptimo de M que nos dé el mejor rendimiento general, esto nos puede conducir a un sobreajuste o subajuste, este tema se cubrirá más adelante. Usualmente los hiperparámetros óptimos tienen que ser buscados por medio de fuerza bruta. Esto involucra entrenar varios modelos usando hiperparámetros diferentes y verificando cuál modelo nos arroja el error generalizado más bajo.

3.7 SOBREAJUSTE, SUBAJUSTE Y REGULARIZACIÓN

El reto central del machine learning es que nuestro algoritmo debe ejecutar adecuadamente nuevas “*entradas no vistas*”, no solo aquéllas con las que el modelo fue entrenado. A esta capacidad se le conoce como **generalización**. Lo que distingue al machine learning de la optimización es que buscamos que el error de generalización, también llamado **error de prueba**, sea bajo. El error de generalización se define como el valor esperado del error en una nueva entrada. Una relación importante entre el error de entrenamiento y el error de prueba es que el error de entrenamiento esperado de un modelo seleccionado aleatoriamente es igual al error de prueba esperado de ese mismo modelo. Sin embargo, cuando utilizamos un algoritmo de aprendizaje automático, los parámetros no se fijan con anticipación, sino que se estiman a partir del conjunto de entrenamiento. En otras palabras, se toma una muestra del conjunto de entrenamiento y se utiliza para seleccionar los parámetros que minimizan el error de entrenamiento; posteriormente, se evalúa el modelo con una muestra del conjunto de prueba. Bajo este procedimiento, el error de prueba esperado es mayor o igual que el error de entrenamiento esperado. Los factores que determinan qué tan bien se desempeñará un algoritmo de machine learning dependen de su habilidad para⁵

- Hacer el error de entrenamiento pequeño.
- Hacer que la brecha entre el error de entrenamiento y prueba sea pequeña.

Estos factores corresponden a los retos centrales del machine learning: **subajuste** y **sobreajuste**. El subajuste ocurre cuando el modelo no es capaz de conseguir un valor de error adecuado en el conjunto de entrenamiento. Sobreajuste ocurre cuando la brecha entre

⁵La información de esta sección está basada y puede consultarse con más profundidad en [6]

el error de entrenamiento y el error de prueba es muy grande. Podemos controlar ya sea si un modelo es más probable a subajuste o sobreajuste alterando su **capacidad**. Informalmente esta se define como la habilidad de ajustar una amplia variedad de funciones. Los modelos con baja capacidad pueden tener dificultades para ajustarse al conjunto de datos de entrenamiento. Los que tienen alta capacidad pueden sobreajustar memorizando propiedades del conjunto de entrenamiento que no les sirven bien en el conjunto de prueba. Una forma de controlar la capacidad de este algoritmo de aprendizaje es escogiendo su **espacio de hipótesis**, el conjunto de funciones que el algoritmo tiene permitido seleccionar como la solución. Por ejemplo, para la regresión lineal tiene el conjunto de todas las funciones lineales de sus entradas como el espacio de hipótesis. Podemos generalizar la regresión lineal para que incluya polinomios. Hacer esto incrementa la capacidad del modelo. Pero como sabemos la capacidad alta puede llevar a sobreajuste. En la fig. 15 se muestra más claramente como se manifiestan las capacidades.

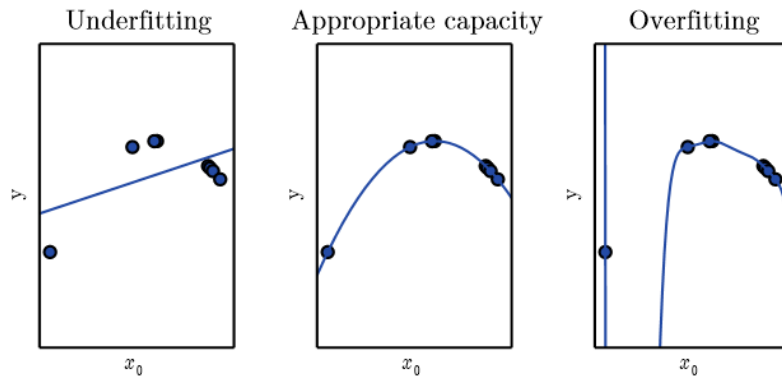


Figura 15: Ejemplo de underfitting, capacidad apropiada y overfitting, adaptado de [6].

Existen varias estrategias para arribar el problema que hemos estado planteando, en específico reducir el error de prueba, probablemente a expensas de incrementar el error de entrenamiento. Colectivamente las técnicas de las que hablaremos se llaman *regularización*, para este trabajo se utilizaron dos técnicas y son de las que hablaremos a continuación.

3.7.1 DETENCIÓN TEMPRANA O EARLY STOPPING

Cuando entrenamos largos modelos con suficiente capacidad representacional para sobreajustar el objetivo, solemos observar que el error de entrenamiento decrece de manera constante, pero el error de validación no, en la fig. 16 podemos ver un ejemplo de esto.

La estrategia que queremos mostrar con esto se le conoce como *detención temprana* o *early stopping*. El entrenamiento puede detenerse en el punto del error más pequeño

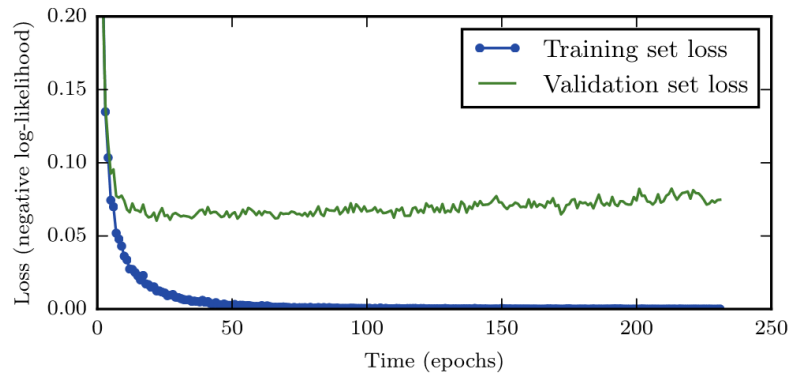


Figura 16: Curvas de aprendizaje que muestran cómo cambia la función de pérdida de log-verosimilitud negativa a lo largo de las épocas de entrenamiento. Adaptado de [6].

con respecto al conjunto de validación para que nuestra red pueda obtener un buen rendimiento de generalización. Detener el entrenamiento antes de que el error de entrenamiento alcance un mínimo representa entonces una manera de limitar efectivamente la complejidad de la red. En la fig. 16 podemos ver que este hiperparámetro tiene una curva de rendimiento de validación que se ve como un decremento exponencial. La mayoría de hiperparámetros que controlan la capacidad del modelo tienen esta curva de rendimiento. En el caso de *early stopping*, estamos controlando la capacidad efectiva del modelo al determinar cuántos pasos se necesitan para hacer el ajuste al conjunto de entrenamiento. La mayoría de los hiperparámetros deben ser elegidos mediante un costoso proceso de prueba y error, donde establecemos los hiperparámetros al inicio del entrenamiento, luego corremos el entrenamiento varias veces para ver su efecto. *Early stopping* es una forma de regularización muy discreta, que no requiere de grandes cambios en el proceso de entrenamiento, la función objetivo, o el conjunto de valores permitidos de los parámetros. Esto lo hace sencillo de usar sin dañar la dinámica de aprendizaje, y lo convierte en una de las formas más usadas de regularización debido a su simplicidad y efectividad.

3.7.2 DROPOUT O REGULARIZACIÓN POR APAGADO

La siguiente que usaremos se le conoce como *Dropout* o *regularización por apagado* es un método computacionalmente económico y poderoso para regularizar una amplia familia de modelos. En general se puede describir como un método para hacer que el *bagging* (método de aprendizaje por conjuntos) sea práctico para conjunto de redes neuronales grandes y numerosas. El *bagging* involucra entrenar múltiples modelos y evaluar múltiples modelos en cada ejemplo de prueba⁶. Esto parece poco práctico cuando cada modelo es una larga red neuronal. *Dropout* proporciona una aproximación menos costosa para entre-

⁶Esta sección está basada en [6]

nar y evaluar exponencialmente un ensamble bagging de muchas redes. Específicamente, dropout entrena el ensamble conformado de todas las subredes que pueden ser formadas al remover unidades o neuronas de no salidas de una red base, en la fig. 17 se ilustra este proceso.

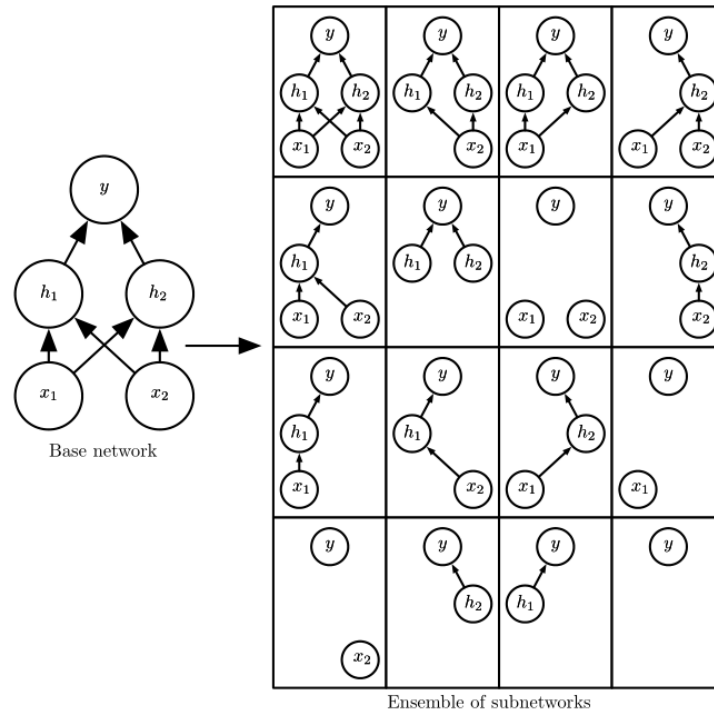


Figura 17: Dropout entrena un conjunto de subredes formadas al eliminar unidades que no pertenecen a la capa de salida de una red base subyacente. En este ejemplo, se muestran todas las dieciséis subredes que se pueden formar eliminando diferentes subconjuntos de unidades de la red original. Este método es efectivo para entrenar y evaluar una red neuronal más pequeña y eficiente. Adaptado de [6].

En la mayoría de redes neuronales actuales, podemos remover neuronas de manera efectiva al multiplicar su salida por cero. Esto requiere unas modificaciones menores tales como redes de función de base radial, las cuales toman la diferencia entre los estados de las neuronas y algún valor de referencia. Hay muchas formas de remover neuronas en una red, nosotros presentaremos uno de los tantos algoritmos con los que se puede implementar el dropout, por simplicidad en términos de una multiplicación por cero. Para el aprendizaje con bagging, definimos k modelos diferentes, construyendo k conjuntos de datos al samplear desde el conjunto de entrenamiento con reemplazo, y luego entrenar el modelo i en el conjunto de datos i . Dropout aspira aproximar este proceso, pero con un número exponencial de redes. Específicamente, para entrenar con dropout, usamos un algoritmo de aprendizaje basado en minibatches que hace pequeños pasos, tal como el gradiente descendiente estocástico. Cada vez que cargamos un ejemplo dentro de un

minibatch, muestreamos aleatoriamente una máscara binaria diferente para aplicar a todas las entradas y unidades o neuronas ocultas en la red. La máscara para cada neurona es muestreada independientemente de las otras. La probabilidad de muestrear un valor de uno en una máscara (causando que una neurona sea incluida) es un hiperparámetro ajustado antes del inicio del entrenamiento, donde la **máscara de dropout** es un vector o matriz binaria aleatoria cuyos elementos indican si una neurona permanece activa o si se apaga temporalmente en una iteración dada del entrenamiento. No es una función del valor actual del modelo o del ejemplo de entrada. Usualmente, una neurona de entrada es incluida con probabilidad 0.8, y una neurona oculta es incluida con probabilidad 0.5. Entonces realizamos la propagación hacia adelante, hacia atrás, después actualizamos el aprendizaje como normalmente se hace. En la fig. 18 se muestra como se realiza una propagación hacia adelante con dropout.

Formalmente, supóngase que un vector máscara μ indica qué neuronas de la red permanecen activas y cuáles se desactivan temporalmente durante el entrenamiento. Así, $J(\theta, \mu)$ define el costo del modelo determinado por los parámetros θ y por la máscara μ . Entonces, el entrenamiento con *dropout* consiste en minimizar $\mathbb{E}_{\mu}[J(\theta, \mu)]$. La expectativa contiene muchos términos exponenciales, pero podemos obtener una estimación no sesgada de su gradiente muestreando valores de μ . Aunque puede sonar parecido, el dropout no es igual al entrenamiento bagging. En el segundo se supone que todos los modelos son independientes. En el caso de dropout, los modelos comparten parámetros. Esto contiene implicaciones importantes. Por ejemplo se puede llegar a que la predicción en un entrenamiento bagging dada por la media aritmética de todas sus contribuciones es,

$$\frac{1}{k} \sum_{i=1}^k p^{(i)}(y|x). \quad (3.57)$$

Mientras que en el caso de dropout, cada submodelo definido por un vector máscara μ define una distribución de probabilidad $p(y|x, \mu)$. Su media aritmética sobre todas las máscaras está dada por

$$\sum_{\mu} p(\mu) p(y|x, \mu), \quad (3.58)$$

donde $p(\mu)$ es la distribución de probabilidad que fue usada para samplear μ en el tiempo de entrenamiento.

3.8 INICIALIZACIÓN DE PARÁMETROS (GLOROT UNIFORM)

Hemos visto casi todas las partes necesarias a tomar en cuenta en nuestra construcción de una red neuronal densa (deep neural network), ahora estamos en posición de

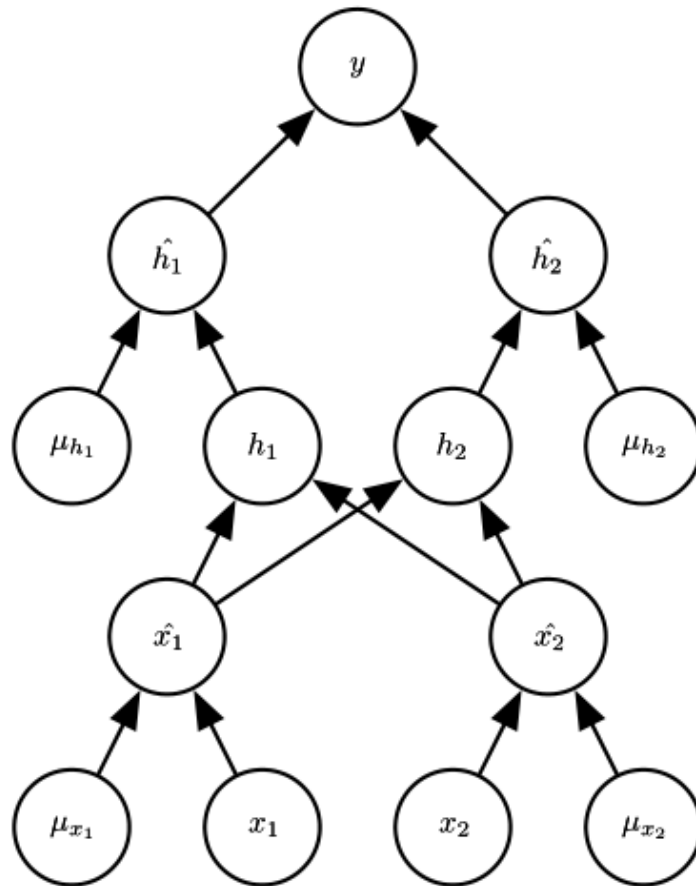


Figura 18: Ejemplo de propagación hacia adelante a través de una red de retroalimentación utilizando dropout. (Arriba) En este ejemplo, usamos una red de retroalimentación con dos unidades de entrada, una capa oculta con dos unidades ocultas y una unidad de salida. (Abajo) Para realizar la propagación hacia adelante con dropout, muestreamos aleatoriamente un vector μ con una entrada para cada unidad de entrada u oculta en la red. Las entradas de μ son binarias y se muestrean independientemente entre sí. La probabilidad de que cada entrada sea 1 es un hiperparámetro, generalmente 0.5 para las capas ocultas y 0.8 para la entrada. Cada unidad en la red es multiplicada por la máscara correspondiente y luego la propagación hacia adelante continúa a través del resto de la red. Esto es equivalente a seleccionar aleatoriamente una de las subredes como se muestra en la fig. 17 y ejecutar la propagación hacia adelante a través de ella. Adaptado de [6].

hablar de la inicialización de parámetros, por que no hemos mencionado más allá de cómo optimizar, pero hay que dar un valor inicial a los pesos en cuestión. El entrenamiento de algoritmos de aprendizaje profundo es usualmente iterativo y requiere de un punto inicial en dónde comenzar la iteración. Un buen punto inicial puede determinar si el algoritmo converge siquiera. Pero puede pasar lo opuesto si hacemos una mala elección. Cuando el aprendizaje converge, el punto inicial puede ahorrarnos tiempo de convergencia y a su

vez costo computacional, uno puede ser más rápido que otro. Asimismo, puntos de un costo comparable pueden tener una alta variación en su error generalizado, entonces el punto inicial puede afectar la generalización. Las estrategias modernas de inicialización son simples y heurísticas. Diseñar un método mejorado de inicialización es una tarea aún complicada y en desarrollo debido a nuestro escaso entendimiento de la optimización en redes neuronales. La única propiedad conocida con mayor certeza es que los parámetros iniciales deben romper la simetría entre diferentes neuronas. Curiosamente un aspecto en común y paralelismo altamente estudiado y fundamentado en la física de partículas como vimos en la parte I, donde esto relaciona desde como las partículas adquieren masa, entre otras cosas. Volviendo a nuestro tema, típicamente, para los sesgos o *bias* en cada neurona se condicionan a ser heurísticamente constantes (en nuestro caso se inicializan en cero), y solo la inicialización de los pesos se hace de manera aleatoria. Casi siempre se suelen inicializar todos los pesos en un modelo mediante una distribución Gaussiana o uniforme. La elección de estas dos parece no tener altas repercusiones pero aún falta estudio más exhaustivo sobre el tema. La escala es otro tema importante en la distribución inicial, esta tiene efectos a largo plazo tanto en la salida del proceso de optimización como en la habilidad de la red para generalizar. Hay varias heurísticas disponibles para elegir la escala inicial de los pesos. Una de ellas es inicializar los pesos de una capa totalmente conectada con m entradas y n salidas al samplear cada peso desde $U(-\frac{1}{\sqrt{m}}, \frac{1}{\sqrt{m}})$. Para este estudio usaremos este método pero con una pequeña modificación, llamada *glorot uniform*⁷, la cual es muy utilizada actualmente, una **inicialización normalizada**

$$W_{i,j} \sim U\left(-\sqrt{\frac{6}{m+n}}, \sqrt{\frac{6}{m+n}}\right). \quad (3.59)$$

3.9 ESCALADORES

En esta sección se presentan dos escaladores ampliamente utilizados en la literatura para el preprocesamiento de variables de entrada [5, 6, 57]. En la sección pasada mencionamos la importancia de la escala para la distribución de los pesos, resulta que en los vectores de entrada el escalamiento también se convierte en algo necesario, las variables de entrada se preprocesan para transformarlas en un nuevo espacio de variables, donde se espera, que el problema de reconocimiento sea más fácil de resolver. Esto reduce la variabilidad dentro de cada clase de dígitos. Es fácil ver la necesidad de escalar nuestros datos ya que acelera ampliamente el proceso de cómputo, al escalar solo estamos extrayendo las características de los datos, de otra forma se convierte en un tedioso y costoso

⁷Toda esta sección está basada en [58]

proceso porque a veces manejamos datos con rangos de valores muy distantes entre sí. Para este estudio utilizaremos dos escaladores, pero hay muchos más en la literatura, estos se eligieron ya que proporcionan buenos resultados.

El primero llamado *escalador máximos y mínimos*

$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \quad (3.60)$$

y el segundo es el *escalador estándar*

$$X_{\text{scaled}} = \frac{X - \mu}{\sigma}. \quad (3.61)$$

3.10 APRENDIZAJE POR TRANSFERENCIA O TRANSFER LEARNING

El último tema pero no menos importante, y que se utilizará en este estudio, es llamado *aprendizaje por transferencia* o *transfer learning* (ha sido ampliamente estudiado en [7] y es de dónde se basa esta sección), antes de comenzar a mencionarlo hagamos algunas definiciones: Un dominio \mathbb{D} consiste en dos componentes: un espacio de características \mathcal{X} y una distribución de probabilidad marginal $\mathbb{P}^{\mathcal{X}}$, donde cada instancia de entrada $\mathbf{x} \in \mathcal{X}$. En general, si dos dominios son diferentes, entonces pueden tener diferente espacio de características o diferente distribución de probabilidad marginal. Dado un dominio específico, $\mathbb{D} = \{\mathcal{X}, \mathbb{P}^{\mathcal{X}}\}$, un objetivo o tarea \mathbb{T} consiste en dos componentes: espacio de clases o etiquetas \mathcal{Y} y una función $f(\cdot)$ (denotado por $\mathbb{T} = \{\mathcal{Y}, f(\cdot)\}$). La función $f(\cdot)$ es una función predictiva que puede ser usada para hacer predicciones en instancias no vistas $\{\mathbf{x}^*\}_s$. En forma probabilística, $f(\mathbf{x})$ puede ser escrito como $P(y|\mathbf{x})$. Por simplicidad, ahora nos enfocamos en el caso donde hay una fuente de dominio \mathbb{D}_s y un dominio objetivo \mathbb{D}_t . En particular, denotamos $\mathcal{D}_s = \{(\mathbf{x}_{s_i}, y_{s_i})\}_{i=1}^{n_s}$ como datos etiquetados del dominio fuente, donde $\mathbf{x}_{s_i} \in \mathcal{X}_s$ es la instancia de datos y $y_{s_i} \in \mathcal{Y}_s$ su correspondiente etiqueta de clase. Similarmente, denotamos $\mathcal{D}_t = \{(\mathbf{x}_{t_i}, y_{t_i})\}_{i=1}^{n_t}$ como datos etiquetados del dominio objetivo, donde las entradas \mathbf{x}_t están en \mathcal{X}_t y $y_t \in \mathcal{Y}_t$ es el correspondiente output. En la mayoría de los casos $0 \leq n_t \ll n_s$. Basado en estas notaciones podemos definir transfer learning como sigue a continuación. Se define **Aprendizaje por transferencia**. Dado un dominio fuente \mathbb{D}_s y una tarea de aprendizaje \mathbb{T}_s , un dominio objetivo \mathbb{D}_t , y una tarea de aprendizaje \mathbb{T}_t , *el aprendizaje por transferencia* tiene como propósito ayudar a mejorar el aprendizaje de la función predictiva objetivo $f(\cdot)_t$ para el dominio objetivo usando el conocimiento en \mathbb{D}_s y \mathbb{T}_s , donde $\mathbb{D}_s \neq \mathbb{D}_t$ o $\mathbb{T}_s \neq \mathbb{T}_t$. En la fig. 19 se muestra como es el proceso de aprendizaje por transferencia. A la izquierda se muestra el

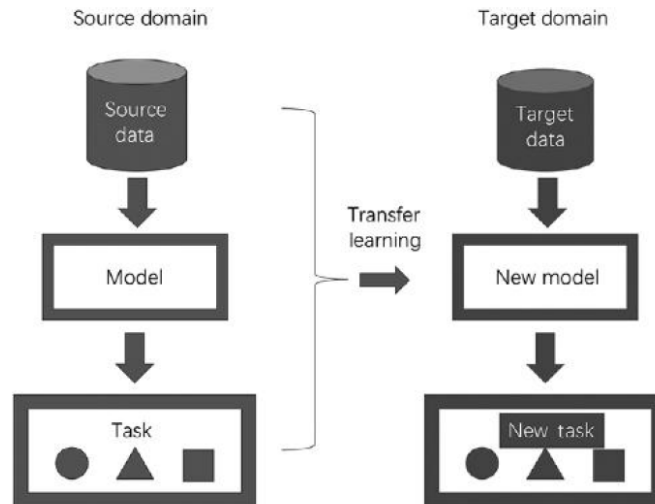


Figura 19: Ilustración de un proceso de aprendizaje por transferencia. Adaptado de [7].

proceso común de aprendizaje por machine learning. En la derecha corresponde al proceso de aprendizaje por transferencia. Aquí está el corazón de este método, el aprendizaje por transferencia hace uso no solo de los datos en el dominio objetivo de tarea como entradas para el algoritmo de aprendizaje, si no cualquiera del dominio fuente en el proceso de aprendizaje, incluyendo los datos de entrenamiento, modelos y tareas objetivo. Un concepto clave sobre el aprendizaje por transferencia que nos resume esto es: contrarresta el problema de la falta de datos de entrenamiento en el dominio objetivo con más conocimiento adquirido del dominio fuente. Hay una vasta literatura sobre este tema [7], como aprendizaje por transferencia heterogéneo y homogéneo donde clasifican con base en sus distribuciones marginales y la intersección de sus subespacios. En nuestro caso, utilizaremos un algoritmo basado en extracción de características. En particular, se optó por transferir dos capas de un modelo previamente entrenado y mantenerlas congeladas en un nuevo modelo. Esta elección busca aprovechar parte del conocimiento aprendido en el dominio fuente sin incrementar innecesariamente el número de parámetros ajustables en el nuevo entrenamiento. Buscamos realizar una extracción de dos capas de un modelo y agregar a un nuevo modelo y congelarlas. El conocimiento transferido corresponde a un subespacio generado por las características del dominio fuente y objetivo. Una idea intuitiva es aprender representaciones buenas de las características para el dominio fuente y el dominio objetivo, proyectando datos dentro de una nueva representación, los datos etiquetados del dominio fuente pueden reutilizarse para entrenar un clasificador preciso para el dominio objetivo. De este modo, el conocimiento que se transfiere entre dominios puede considerarse como la representación de características aprendidas. Aunque hemos mencionado a la parte de clasificación como ejemplo, este esquema es aplicable tanto para modelos de regresión como de clasificación. Podemos resumir nuestro proceso de aprendi-

zaje por transferencia basado en extracción de características en los siguientes pasos para ser implementado:

- Tomar una red preentrenada en el dominio fuente \mathcal{D}_s .
- Reutilizar algunas de sus capas como extractor de características.
- Congelar las capas.
- Agregar nuevas capas a una nueva red creando un grafo computacional ya sea igual o diferente al de la red preentrenada. Esto nos implica que $\mathbb{D}_s \neq \mathbb{D}_t$ y $\mathbb{T}_t = \mathbb{T}_s$ para nuestro propósito.
- Entrenar solo estas nuevas capas con los datos del dominio objetivo \mathbb{D}_t .

Aunque este proceso puede ser similar a agregar nuevas capas a nuestro modelo inicial congelando solo algunas capas, nuestra hipótesis fue realizar este proceso para observar si esto realmente contribuye a una mejora o no en la generalización, significativa en cuestión, en la sección 5 se muestran los resultados de esta idea. Claro que también es válido realizar el otro caso, y más sencillo, pero en este trabajo solo se abordó esta última opción.

3.11 DISTRIBUCIONES DE PROBABILIDAD TRUNCADAS EN EL MODELO

Ahora vamos con una breve información acerca de las distribuciones de probabilidad, se menciona hasta este punto ya que el dominio fuente de los datos que usaremos para nuestra red necesita una distribución de probabilidad truncada, esto debido a que tenemos una restricción de valores como se puede observar en la tabla 3 del capítulo 2, entonces hay que analizar la forma de hacer el truncamiento.

Sea $f(x; \theta_1, \dots, \theta_r)$ y $F(x; \theta_1, \dots, \theta_r)$ designadas como la función de densidad de probabilidad (PDF, por sus siglas en inglés) y la función de distribución acumulada (CDF por sus siglas en inglés) de una distribución sin restricciones (i.e., completa) con parámetros $\theta_1, \dots, \theta_r$. Las funciones de probabilidad de muestras truncadas están dadas de la siguiente manera⁸.

Truncada por la izquierda en $x = T$:

$$L() = \frac{1}{1 - F(T)} \prod_{i=1}^n f(x_i), \quad T \leq x_i. \quad (3.62)$$

⁸La presente sección sobre distribuciones de probabilidad truncadas se basa principalmente en [59].

Truncada por la derecha en $x = T$:

$$L() = \frac{1}{F(T)} \prod_{i=1}^n f(x_i), \quad x_i \leq T. \quad (3.63)$$

Doblemente truncada en T_1 y T_2 :

$$L() = \frac{1}{F(T_2) - F(T_1)} \prod_{i=1}^n f(x_i), \quad T_1 \leq x_i \leq T_2. \quad (3.64)$$

Ahora vamos a aplicar esto a las distribuciones usadas en este trabajo, empezaremos con la **distribución normal**, usualmente está dada como

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right), \quad (3.65)$$

en su forma truncada que es la que nos interesa y usando la ecuación (3.64), la **distribución normal truncada** es

$$f(x; \mu, \sigma) = \frac{1}{[F(T_2) - F(T_1)] \sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad (3.66)$$

$T_1 \leq x \leq T_2$, 0 fuera del intervalo.

En unidades estandarizadas de la distribución completa, los puntos de truncamiento son

$$\xi_1 = \frac{T_1 - \mu}{\sigma} \quad \text{y} \quad \xi_2 = \frac{T_2 - \mu}{\sigma}. \quad (3.67)$$

Se sigue que

$$F(T_1) = \Phi(\xi_1) \quad \text{y} \quad F(T_2) = \Phi(\xi_2), \quad (3.68)$$

donde Φ es la cdf de la distribución normal estándar (0, 1).

La función log-verosimilitud de una muestra aleatoria de tamaño n de una distribución normal truncada (ecuación (3.66)) es

$$\begin{aligned} \ln L = & -n \ln [\Phi(\xi_2) - \Phi(\xi_1)] - n \ln \sigma \\ & - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 + \text{const.} \end{aligned} \quad (3.69)$$

Las ecuaciones de estimación de máxima verosimilitud, obtenidas al igualar a cero las derivadas parciales de la función log-verosimilitud con respecto a μ y σ son

$$\frac{\partial \ln L}{\partial \mu} = -\frac{n}{\sigma} \left[\frac{\phi_1 - \phi_2}{\Phi_2 - \Phi_1} \right] + \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0, \quad (3.70)$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma^2} \left[\frac{\xi_1 \phi_1 - \xi_2 \phi_2}{\Phi_2 - \Phi_1} \right] - \frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \mu)^2 = 0,$$

se han abreviado la pdf y cdf a $\phi_{1,2}$ y $\Phi_{1,2}$. Para simplificar más, definimos

$$\bar{Q}_1 = \frac{\phi_1}{\Phi_2 - \Phi_1} \quad \text{y} \quad \bar{Q}_2 = \frac{\phi_2}{\Phi_2 - \Phi_1}. \quad (3.71)$$

Sustituyendo en (3.70) nos queda

$$\begin{aligned} (\bar{x}_n - \mu) &= \sigma (\bar{Q}_1 - \bar{Q}_2), \\ s_n^2 + (\bar{x}_n - \mu)^2 &= \sigma^2 (1 + \xi_1 \bar{Q}_1 - \xi_2 \bar{Q}_2). \end{aligned} \quad (3.72)$$

Ahora usamos la ecuación (3.68) para simplificar y obtener

$$\begin{aligned} s_n^2 &= \sigma^2 [1 + \xi_1 \bar{Q}_1 - \xi_2 \bar{Q}_2 - (\bar{Q}_1 - \bar{Q}_2)^2], \\ (\bar{x}_n - T) &= \sigma (\bar{Q}_1 - \bar{Q}_2 - \xi_1). \end{aligned} \quad (3.73)$$

Nuevamente usando la ecuación (3.68) pero con $w = T_2 - T_1$, reescribimos

$$\sigma = \frac{w}{\xi_2 - \xi_1}. \quad (3.74)$$

Sustituyendo todo esto en (3.73) tenemos las siguientes ecuaciones de estimación

$$\begin{aligned} H_1(\xi_1, \xi_2) &= \frac{\nu_1}{w}, \\ H_2(\xi_1, \xi_2) &= \frac{s_n^2}{w^2}, \end{aligned} \quad (3.75)$$

donde

$$\begin{aligned} H_1(\xi_1, \xi_2) &= \frac{\bar{Q}_1 - \bar{Q}_2 - \xi_1}{\xi_2 - \xi_1}, \\ H_2(\xi_1, \xi_2) &= \frac{1 + \xi_1 \bar{Q}_1 - \xi_2 \bar{Q}_2 - (\bar{Q}_1 - \bar{Q}_2)^2}{(\xi_2 - \xi_1)^2} \end{aligned} \quad (3.76)$$

y

$$\nu_1 = \bar{x}_n - T_1. \quad (3.77)$$

Se pueden resolver las ecuaciones de estimación para los estimadores $\hat{\xi}_1$ y $\hat{\xi}_2$. Se sigue que

$$\hat{\sigma} = \frac{w}{\hat{\xi}_2 - \hat{\xi}_1} \quad y \quad \hat{\mu} = T_1 - \hat{\sigma}\hat{\xi}_2. \quad (3.78)$$

Se ha prestado una gran atención a esta distribución ya que es la que desempeñará un papel crucial en nuestro trabajo, por dar los mejores resultados, aunque también se trataron con otras distribuciones pero con menor éxito, que veremos ahora.

La siguiente distribución que mostraremos es la **distribución weibull**, que está definida como

$$f(x; \beta, \delta) = \begin{cases} \frac{\delta}{\beta^\delta} x^{\delta-1} \exp\left[-\left(\frac{x}{\beta}\right)^\delta\right], & 0 \leq x < \infty, \\ 0, & \text{en otro caso.} \end{cases} \quad (3.79)$$

Para truncarla hacemos usamos la definición (3.64) para llegar a algo de este tipo

$$f_{\text{trunc}}(x; \beta, \delta, T_1, T_2) = \begin{cases} \frac{f(x; \beta, \delta)}{F(T_2; \beta, \delta) - F(T_1; \beta, \delta)}, & T_1 \leq x \leq T_2, \\ 0, & \text{en otro caso,} \end{cases} \quad (3.80)$$

donde la cdf no truncada es

$$F(x; \beta, \delta) = 1 - \exp\left[-\left(\frac{x}{\beta}\right)^\delta\right], \quad x \geq 0. \quad (3.81)$$

A veces y en nuestro caso se requiere un escalamiento de los valores de truncamiento, mostraremos la normalización Min-Max que es muy usada para quedarnos con valores en el rango $[c, d]$

$$T'_1 = \frac{T_1 - T_{\text{mín}}}{T_{\text{máx}} - T_{\text{mín}}} (d - c) + c, \quad (3.82)$$

$$T'_2 = \frac{T_2 - T_{\text{mín}}}{T_{\text{máx}} - T_{\text{mín}}} (d - c) + c.$$

La tercera distribución que mostraremos es la **distribución Gamma** dada por

$$f(x; \gamma, \beta, \rho) = \begin{cases} \frac{\beta^{-\rho}}{\Gamma(\rho)} (x - \gamma)^{\rho-1} \exp\left(-\frac{x - \gamma}{\beta}\right), & x \geq \gamma, \rho > 0, \\ 0, & \text{en otro caso.} \end{cases} \quad (3.83)$$

Para truncarla es el mismo caso que ya hemos visto, aplicar el cociente de la diferencia de los valores de truncamiento en la cdf con los mismos parámetros. Por último y la más sencilla la **distribución uniforme**

$$f(x; a, b) = \begin{cases} \frac{1}{b - a}, & a \leq x \leq b, \\ 0, & \text{en otro caso.} \end{cases} \quad (3.84)$$

Con su respectivo truncamiento similar a lo que ya hemos mencionado en las distribuciones anteriores.

Parte III

Experimento y Resultados

CAPÍTULO 4

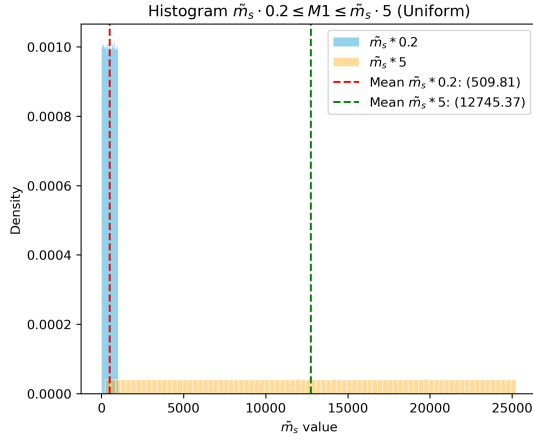
EXPERIMENTO

En este capítulo vamos a mostrar cómo es que se generaron los datos con el modelo teórico de SUSY que se ha estado planteando, cómo es que se distribuyen los datos y su preparación, el tiempo que tarda nuestro modelo en compilar, cómo tratamos ciertos problemas transcurridos en su implementación, entre otras cosas. Todos los scripts para usar el modelo, generar los datos, entrenarlos, analizarlos y graficarlos para toda la parte III en general, están disponibles en el siguiente repositorio: <https://github.com/Jsan2178/Data-science-analysis-project-por-MS.c.git>.

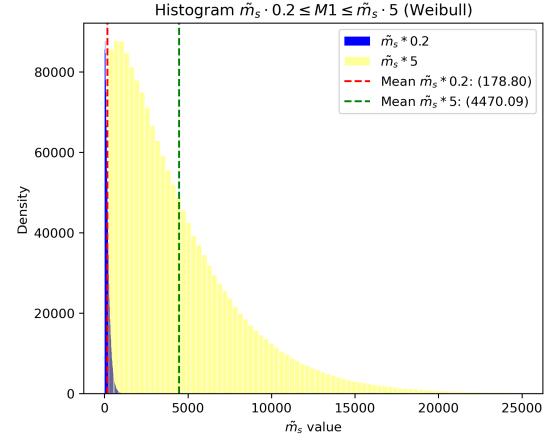
4.1 GENERACIÓN DE DATOS

Para obtener los datos que se van a usar en el modelo analítico como ya vimos en la tabla 3 necesitamos primero generar valores en esos rangos. Para realizar esto se propusieron 4 distribuciones: Uniforme, Weibull, Gamma y Gaussiana, todas estas doblemente truncadas obviamente. Un problema que se enfrentó en esta generación fue el cociente de M_1/\tilde{m}_s ya que al no saber exactamente el valor de M_1 si no en su forma acoplada, suponer que \tilde{m}_s sigue una distribución específica, afecta directamente a M_1 , y aunque podemos hacer particiones finitas de la forma $M_1 \in [\tilde{m}_s \cdot n_1, \tilde{m}_s \cdot n_i]$, donde $n_i = 0.2, \dots, 5$ esto puede ser costoso a largo plazo, por lo que nos concentramos en los extremos, es decir; tomar $M_1 \in [\tilde{m}_s \cdot 0.2, \tilde{m}_s \cdot 5]$. Una vez seleccionado como tratar este cociente para tomar solo una variable, se espera que la distribución tenga dos picos, uno muy fino y otro más pronunciado por los propios valores de los extremos, esto se puede observar en la fig. 20 donde se muestran las cuatro distribuciones

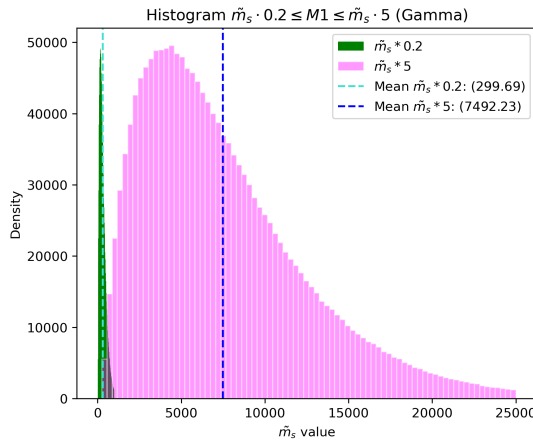
El siguiente paso es computar el modelo analítico que fue desarrollado en la sección 2.14, todo esto consiste en escribir un código que representa la ecuación (2.93) que al final del desarrollo es lo que nos da la contribución del muon en términos de los parámetros



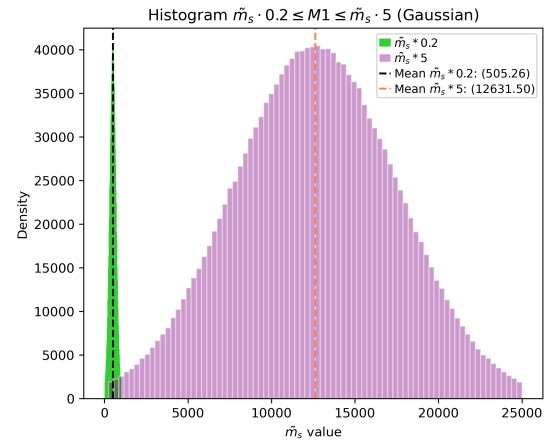
(a) Gráfica de distribución Uniforme para los valores que toma M_1 , donde $\tilde{m}_s \cdot 0.2 \leq M_1 \leq \tilde{m}_s \cdot 5$.



(b) Gráfica de distribución Weibull para los valores que toma M_1 , donde $\tilde{m}_s \cdot 0.2 \leq M_1 \leq \tilde{m}_s \cdot 5$.



(c) Gráfica de distribución Gamma para los valores que toma M_1 , donde $\tilde{m}_s \cdot 0.2 \leq M_1 \leq \tilde{m}_s \cdot 5$.



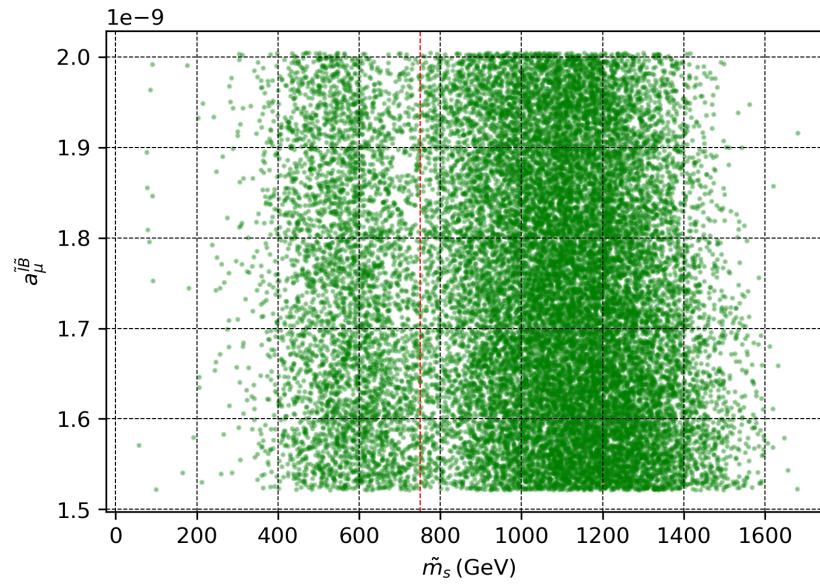
(d) Gráfica de distribución Gaussian para los valores que toma M_1 , donde $\tilde{m}_s \cdot 0.2 \leq M_1 \leq \tilde{m}_s \cdot 5$.

Figura 20: Conjunto de las 4 distribuciones truncadas usadas para el modelo analítico.

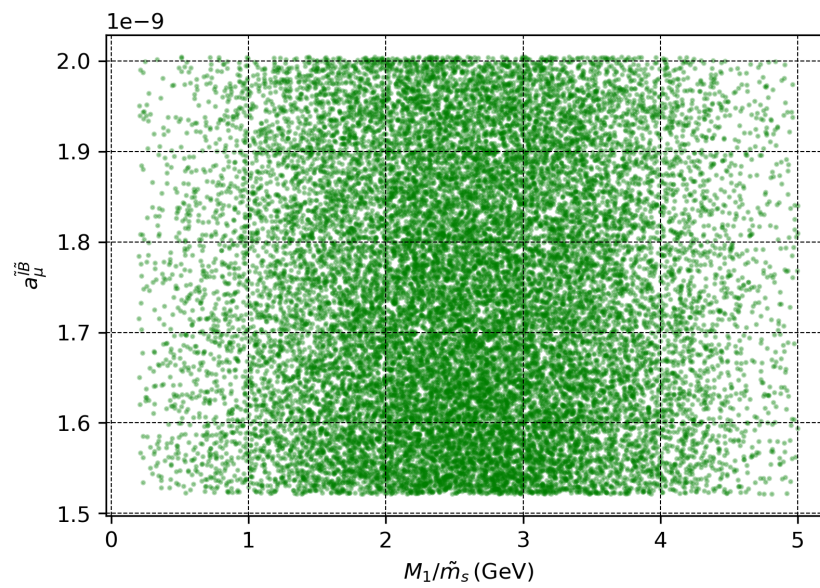
del sector soft de SUSY.¹

Nuestra propuesta es proponer que la contribución, en términos de la desviación estándar, se encuentre en $a_\mu^{SUSY} \in [3.15\sigma, 4.15\sigma]$, esto conlleva a filtrar los parámetros que cumplan la condición, al programar esto, en el caso de la distribución gamma (que es la que se usará, más adelante comentamos nuestra selección) se observó la dispersión de los parámetros que se muestra en la fig. 21 y 22.

¹(Todo esto se encuentra en el repositorio antes mencionado: Github).

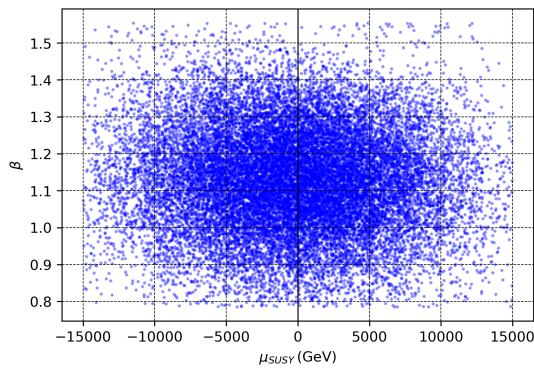


(a) Valores filtrados de \tilde{m}_s que cumplen la condición requerida con respecto a a_μ .

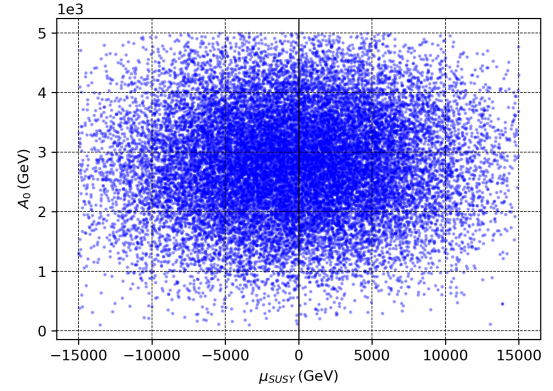


(b) Valores filtrados de M_1/\tilde{m}_s que cumplen la condición requerida con respecto a a_μ .

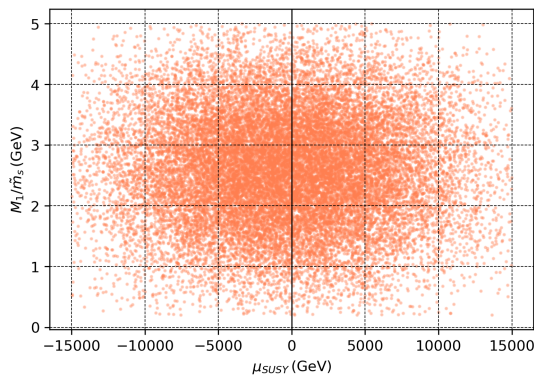
Figura 21: Valores de M_1 y \tilde{m}_s que resuelven la anomalía de a_μ reduciendo a 3.15σ y 4.15σ , lo cual conlleva a filtrar los parámetros que cumplan tal condición, en el código con la distribución deseada.



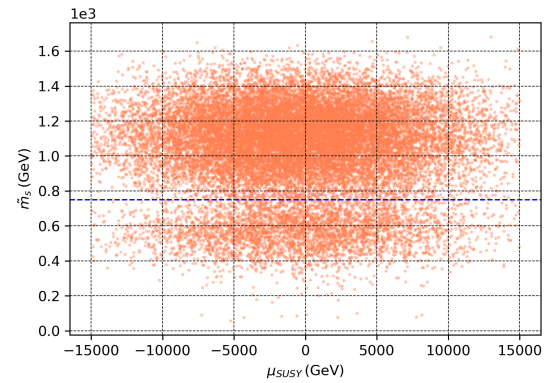
(a) Valores filtrados β y μ_{SUSY} que cumplen la condición.



(b) Valores filtrados A_0 y μ_{SUSY} que cumplen la condición.



(c) Valores filtrados M_1/\tilde{m}_s y μ_{SUSY} que cumplen la condición.



(d) Valores filtrados \tilde{m}_s y μ_{SUSY} que cumplen la condición.

Figura 22: Conjunto de los parámetros β , A_0 , M_1/\tilde{m}_s , \tilde{m}_s que cumplen la condición con respecto a μ_{SUSY} .

4.2 SELECCIÓN DE DISTRIBUCIÓN

En el siguiente capítulo se van a mostrar los resultados pero con una sola distribución seleccionada, esta fue elegida ya que al generar los datos y utilizar la red neuronal construida ² tuvimos algunas diferencias de rendimiento que es necesario mencionar aquí para dejar en claro qué se usó para determinar los resultados.

En la tabla 5 se muestra después de varias corridas, los resultados que arrojaron todas las distribuciones junto con su desempeño para la red, es interesante notar que a pesar de que algunas distribuciones arrojen más datos para nuestra red, al final estos no mejoran el entrenamiento de la misma, por esta razón los datos inherentemente son más importantes que su cantidad, la red converge más rápido con menos datos si estos datos son los

²(Esto se abordará con detalle en el siguiente capítulo).

Tabla 5: Resultados de filtrado por distribución, coeficiente R^2 y tiempo de entrenamiento de la NN, el formato de tiempo es en horas:minutos:segundos.

Distribución	Valores filtrados	Coeficiente R^2	Tiempo de cómputo
Normal	$\approx 2.3 \times 10^4$	≈ 0.93	0:02:08
Uniforme	$\approx 4.2 \times 10^4$	≈ 0.9	0:03:11
Weibull	$\approx 1.12 \times 10^5$	≈ 0.89	0:04:11
Gamma	$\approx 9.3 \times 10^4$	≈ 0.91	0:07:50

adecuados para el problema, claro que determinar esto analíticamente es muy complicado, por lo que debemos hacer diferentes prueba-error para analizar lo más conveniente en tiempo de computo y en generalización del problema, por esta razón seleccionamos la distribución Gaussiana o Normal, a pesar de que es la que nos arroja menos datos, es la que hace converger más rápido la red y obtenga una mejor generalización, sin mencionar que reduce altamente el tiempo de entrenamiento.

4.3 ESCALAMIENTO

Para el escalamiento de los datos usamos los métodos que listamos en la sección 3.9, el Min-Max y el Standard-Scaler pero mezclados, no por separado, es decir para el conjunto de entrenamiento y prueba; valores de entrada y objetivo, en cada uno de estos subconjuntos utilizó uno diferente.

4.4 DIVISIÓN DE LOS DATOS

Los datos fueron divididos de la siguiente manera: el 80 % para \mathcal{D}_{train} , y 20 % para el \mathcal{D}_{test} , uno puede hacer su conjunto y subconjuntos con cantidades diferentes, pero usualmente esta partición es funcional, también nos aseguramos de obtener la misma división con la misma semilla para no enfrentarnos a diferentes divisiones en cada corrida que quisiéramos volver a calcular la NN. Para todo esto se usaron las librerías típicas: *sklearn*, *tensorflow/keras*. La división se muestra de manera más explícita en la tabla 6.

Cabe mencionar que esta división se hizo debido a que en la construcción de la red, al hacer el ajuste o fit establecimos un conjunto de validación `validation_split` como parte del conjunto \mathcal{D}_{train} del 20 %, dejando menos valores para el entrenamiento por este motivo, el resto se queda intacto (el conjunto de prueba o test).

Tabla 6: Fracción del data set total $\mathcal{D} = (X, y)$ usados para entrenamiento, validación y testing.

Data set	$\mathcal{D}_{\text{train}}$	\mathcal{D}_{val}	$\mathcal{D}_{\text{test}}$
Fraction	64 %	16 %	20 %

RESULTADOS

En este capítulo vamos a presentar los resultados obtenidos por los modelos entrenado así como las predicciones finales y algunas consideraciones implicadas en los parámetros de SUSY que tomamos: 1) presentamos las redes construidas, 2) resultados predictivos y 3) los problemas asociados a el propio modelo de supersimetría para este caso.

5.1 RESULTADOS DE VALIDACIÓN

Para el desarrollo de la red neuronal se utilizaron las bibliotecas *TensorFlow* [60] y *Keras* [61]. Diferentes ajustes y configuraciones para la NN fueron configurados hasta llegar a un modelo aceptable, para la primera NN se llegó a 50 epochs con un *learning_rate* = 1×10^{-3} para un buen ajuste, tomando en cuenta que del modelo de SUSY tomamos un conjunto de parámetros de entrada de $\sim 2 \times 10^4$ filtrados de los $\sim 10^6$ que nos da el modelo. Llegando a un $R^2 = 0.92 \pm 1$, un $MSE = 1.2 \times 10^{-21}$ y un $MAE = 1.7 \times 10^{-11}$. A continuación podemos observar las predicciones obtenidas para esta primera red en la fig. 23.

Descripción de parámetros: vamos a dar un resumen de los hiperparámetros utilizados así como los rangos establecidos. Mostraré la configuración más conveniente a la se llegó, sujeto a que al ser algo tan variable este aspecto, nos centramos en la que mejor ha funcionado hasta ahora.

- **Función de Activación.** Se usó una función muy conocida la cuál es *ReLU* para un primer acercamiento y fue la que nos proporcionó buenos resultados, aunque también probamos otras que más adelante mostraremos. Solamente en la última capa del output \hat{y} donde se utilizó la *linear* con un esto por obvias razones ya que debemos obtener una sola salida y esta debe coincidir con el valor de la contribución de SUSY en la anomalía del muon.

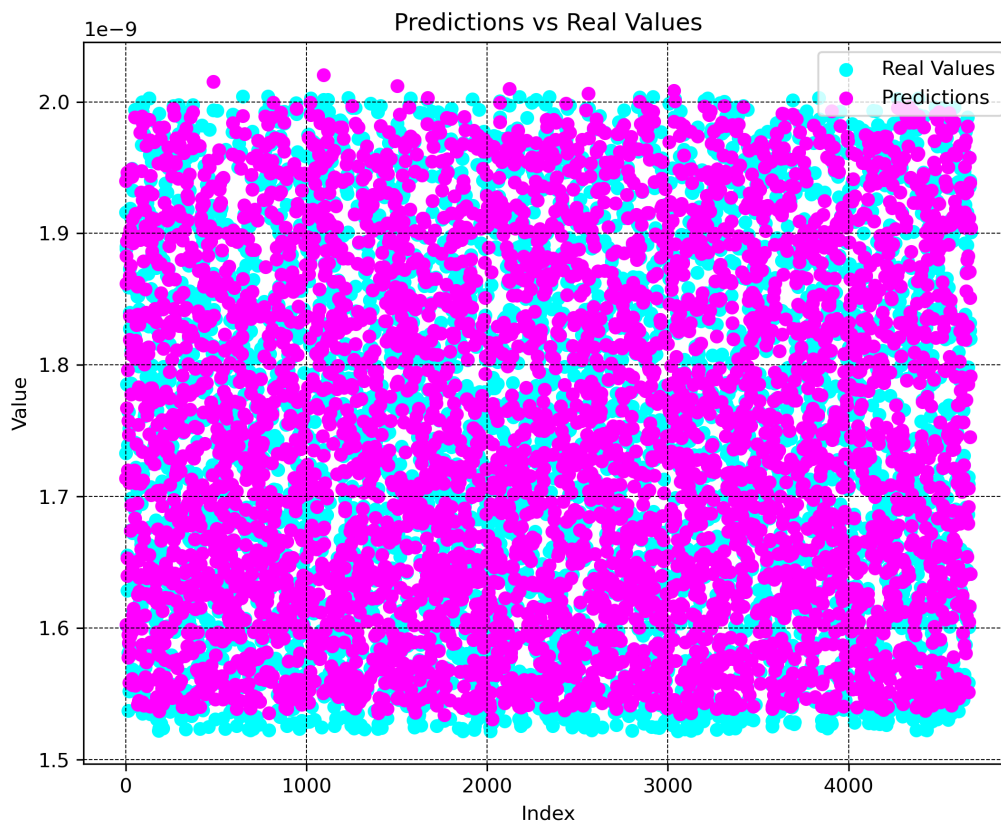


Figura 23: Primera Red neuronal: valores obtenidos vs predicciones del modelo.

- **Arquitectura.** Usamos un modelo *Sequential* con keras de la librería de tensorflow con una cantidad de 5 capas con 128 neuronas todas excepto la última que contiene 1 sola para el output, más 2 de regularización dropout para evitar el sobreajuste utilizando solamente un rango de 0.2, esto fue un arreglo sencillo y que funcionó bien.
- **Optimizador, Learning Rate y Batch size.** Se utilizó el optimizador Adam con probando valores de $\eta \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$ para la configuración más óptima llegamos al *learning rate* $\eta_{NN} = 10^{-3}$ por lo tanto solo haremos énfasis en resultados con este último aunque se mostrará un pequeño análisis que se obtuvo con los diferentes learning rates.
Utilizamos diferentes batch sizes $S \in \{8, 16, 32, 64\}$ donde el batch size de 32 fue visto como el más adecuado para nuestra configuración por último la función de error utilizada fue la métrica de *Mean Squared Error*.
- **Epochs, Patience, Sample size.** Utilizamos epochs desde 10 hasta 90 concluyendo en el mejor en 50. Para el valor de Patience utilizado en *EarlyStopping* fue 15 monitoreado en el *val_loss* para los mejores pesos. El tamaño de los datos iniciales

es 10^6 una vez filtrado obtenemos 10^4 en donde para nuestros datos de validación utilizamos el 0.2 de los filtrados y el resto para entrenamiento.

- **Distribuciones.** Para este caso partimos utilizando una distribución uniforme en la generación de parámetros, lo cuál nos dio un tiempo extenso para el entrenamiento y la necesidad de más datos, después de probar la distribución Weibull, Gamma y Normal Gaussiana pudimos concluir que al truncar esta última nos daba un mejor tiempo reducido (como se vio en la sección 4.2) para el entrenamiento y aprendizaje en la red por lo que la Distribución Normal Gaussiana Truncada es la que se usó en este análisis.

Performance: Función de Activación Una vez hecha la arquitectura de la red que fue [5, 128, 128, 128, 1] con la función de activación utilizada y ajustando los parámetros de la red y distribuciones el train set y el validation set obtuvimos (sin escalar) obtuvimos unos valores muy malos debido a que trabajamos con valores objetivo muy pequeños del orden de 10^{-9} , al realizar el escalamiento de los datos de entrada y con un patience de 15 obtuvimos el rendimiento listado en la fig. 24.

Posteriormente revisamos para el conjunto de η ya mencionado qué tan viable podría ser, comprobando lo que se redactó acerca del mejor ajuste para ese hiperparámetro como podemos ver en la fig. 25 y un vistazo más general del mínimo error encontrado con cada Learning Rate en la fig. 26.

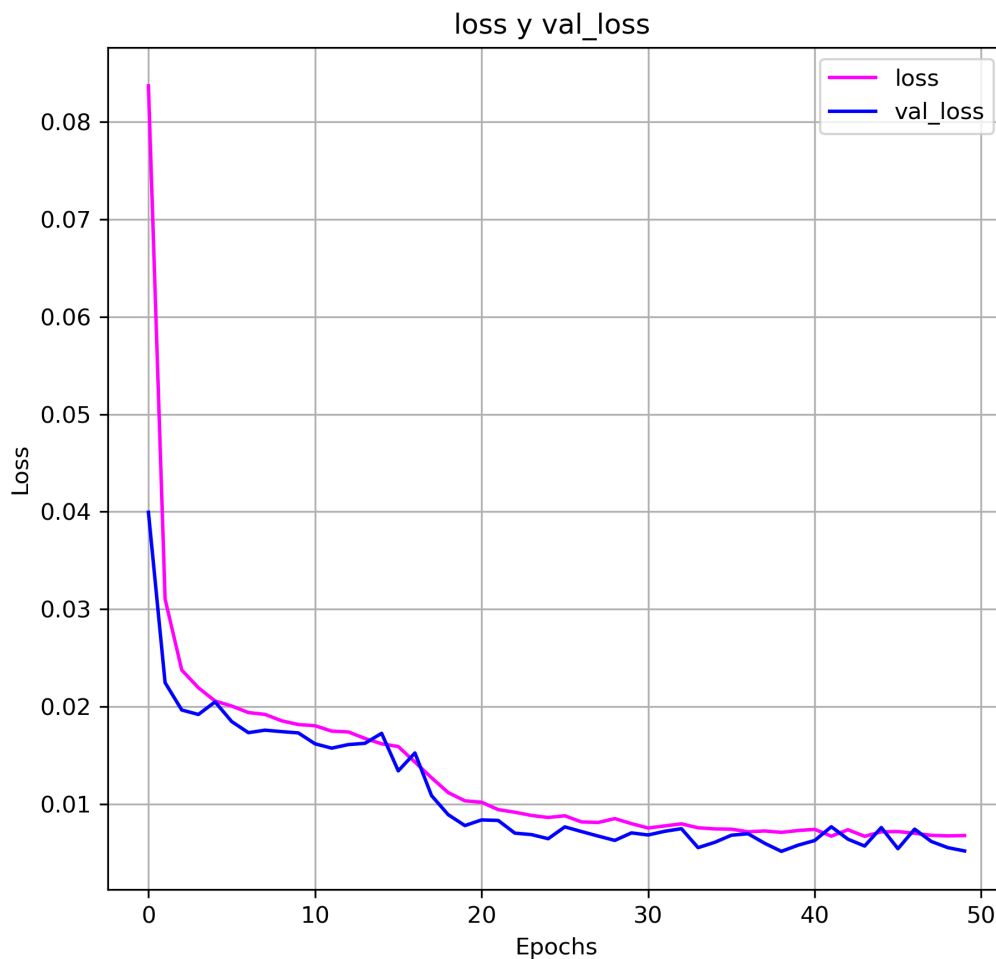


Figura 24: Evaluación de la primera NN utilizando el $\eta = 1 \times 10^{-3}$. con ReLu con valores hasta $\sim \times 10^4$ mostrando la validación de la pérdida con el train set y el validation set.

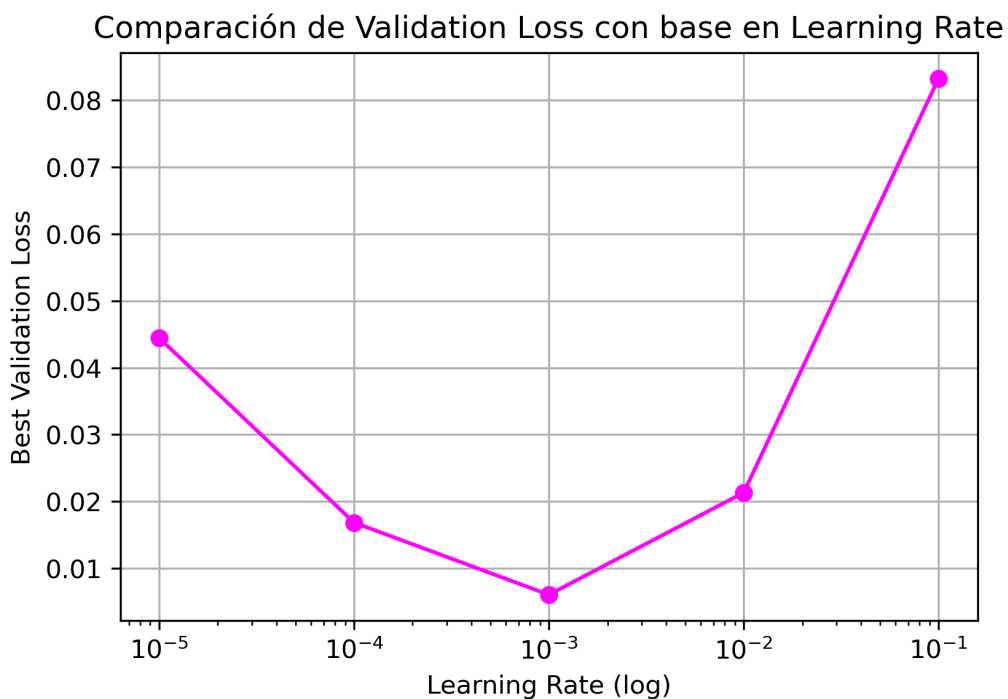


Figura 26: Mejor pérdida de validación correspondiente a los $\eta = \{1 \times 10^{-1}, 1 \times 10^{-2}, 1 \times 10^{-3}, 1 \times 10^{-4}, 1 \times 10^{-5}\}$ en una escala log.

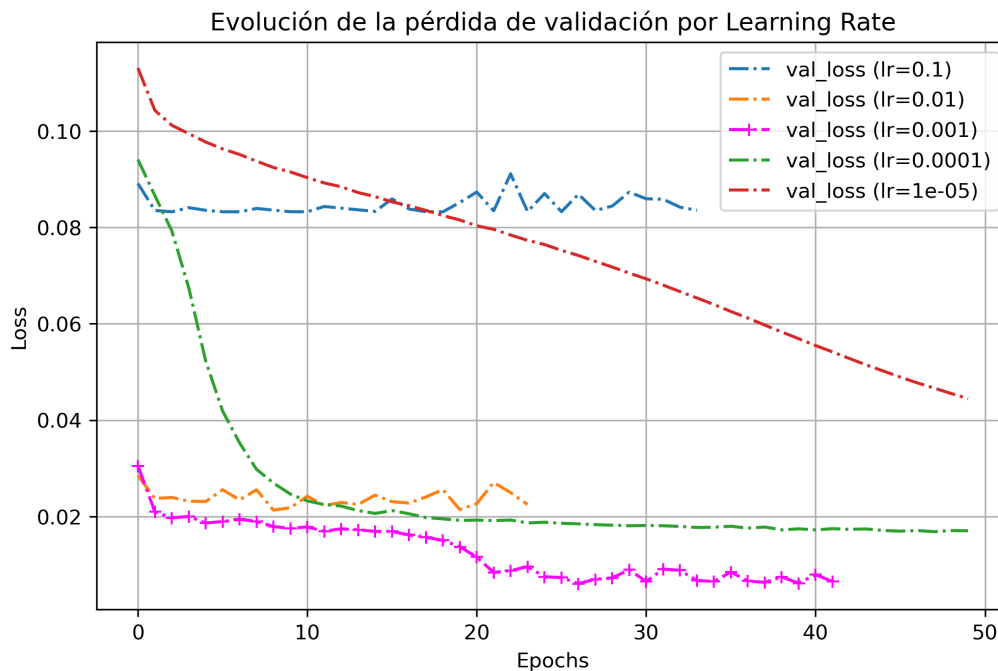


Figura 25: Evaluación de la primera NN utilizando $\eta = \{1 \times 10^{-1}, 1 \times 10^{-2}, 1 \times 10^{-3}, 1 \times 10^{-4}, 1 \times 10^{-5}\}$. con ReLu con valores hasta $\sim \times 10^4$ mostrando la validación de la pérdida con el validation set.

Adicionalmente establecimos una comparación de qué tan buen ajuste podría obtener el modelo si utilizáramos diferentes funciones de activación, concluyendo nuevamente que la función ReLu es la más adecuada, véase la fig. 27.

Distribución de Validation Loss por Epoch y Activación

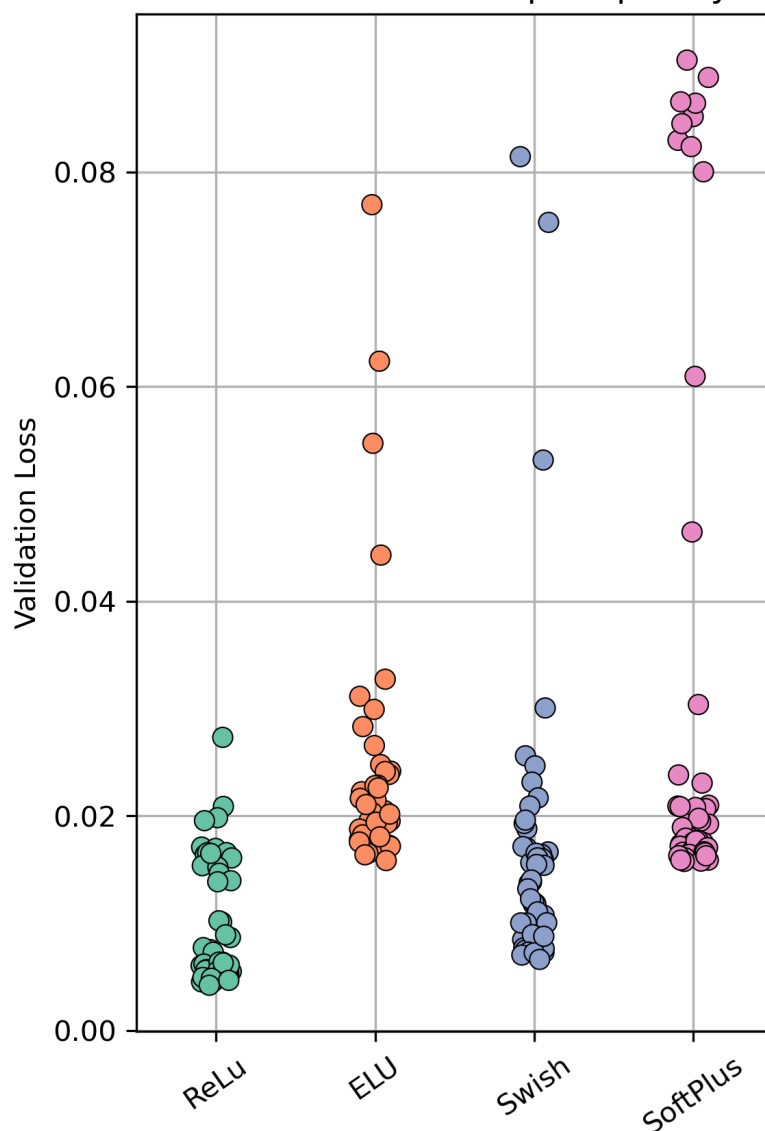


Figura 27: Valores obtenidos con diferentes funciones de activación: ReLu, Swish, ELU, SoftPlus se llegan a diferentes validaciones de aprendizaje, en el cuál ReLu actúa como el más adecuado para usarse en nuestro modelo al tener el error más bajo.

Segunda Red: Creamos una segunda red con la misma arquitectura de la primera pero haciendo uso de *transfer learning* pasando solo lo aprendido de las últimas 2 capas de nuestro primer modelo a esta última, esto con el fin de refinar más el ajuste que ya habíamos obtenido, se congelaron las capas heredadas del modelo anterior y solo se entrenaron las primeras tres.

Para entrenar la 2da red se volvió a correr el código del modelo analítico de SUSY para obtener nuevos datos y evitar un sobreajuste utilizando los mismos. Esto nos arrojó

una muy ligera mejora solamente de $R^2 = 0.93 \pm 1$ con un $MSE = 1.07 \times 10^{-21}$ y un $MAE = 1.6 \times 10^{-11}$, en las métricas es donde se observó un rendimiento más considerable véase la fig. 28 y 29.

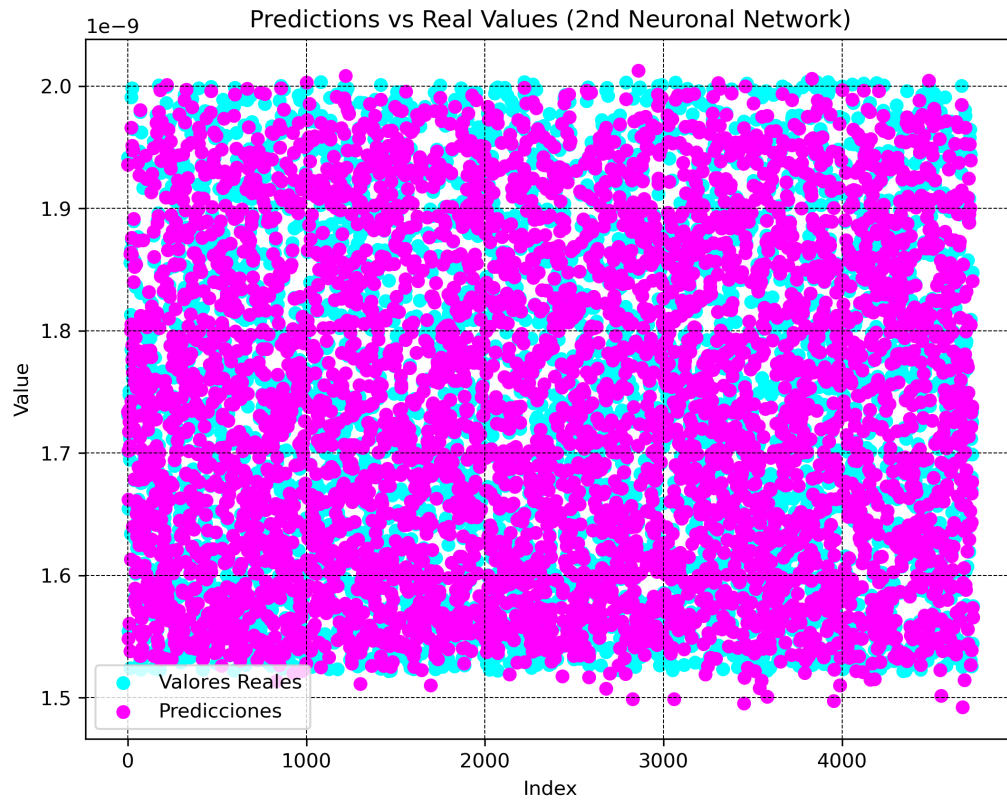


Figura 28: Segunda Red neuronal: valores obtenidos vs predicciones del modelo.

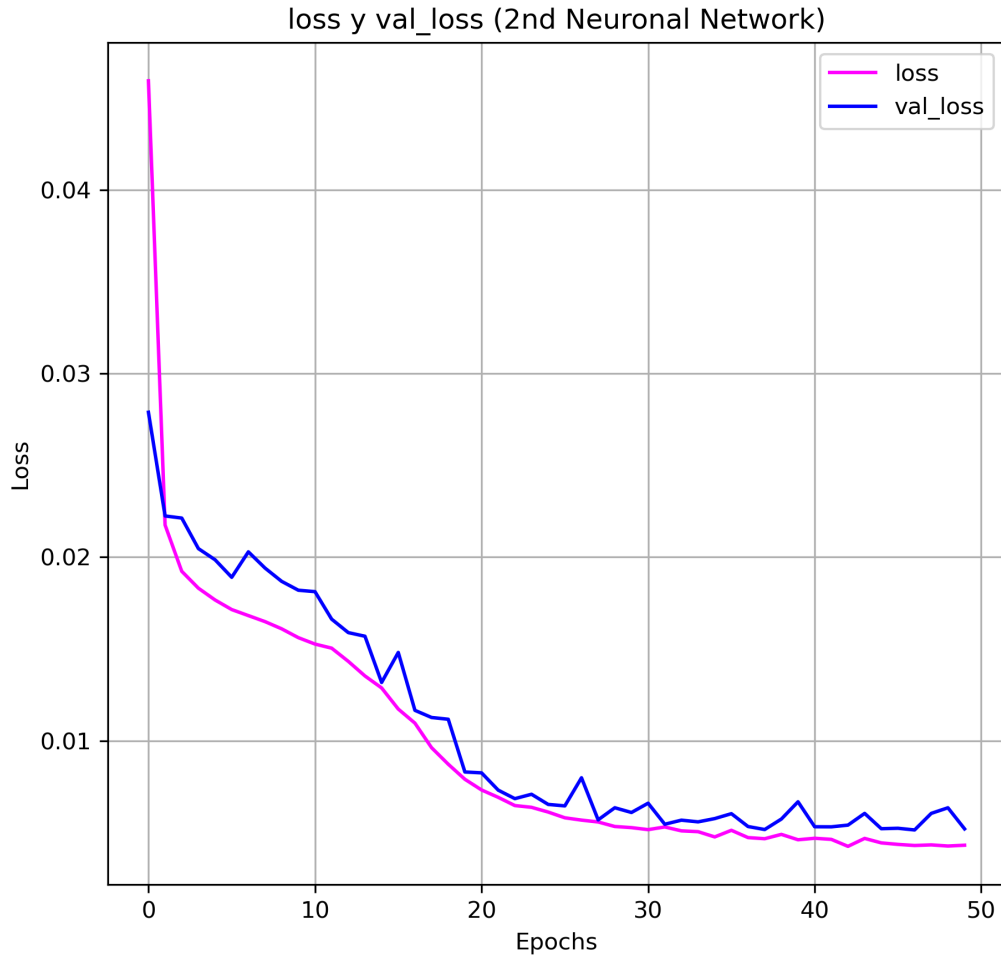


Figura 29: Evaluación de la segunda NN utilizando el $\eta = 1 \times 10^{-3}$. con ReLu con valores hasta $\sim \times 10^4$ mostrando la validación de la pérdida con el train set y el validation set.

Algunas características destacables en el entrenamiento en la primera capa fueron que para cada parámetro seleccionado del MSSM con LFV hubo un ajuste de pesos distintos (véase fig. 30), en la masa del gaugino \tilde{m}_s hay una gran dispersión, esto quiere decir que los pesos se ajustaron con mayor rango en esta parte y con una mayor influencia para la NN, aunque escalado, puede ser congruente por el propio rango de valores que tiene a priori, que podemos observar en la fig. 21a. Por otro lado el cociente entre la masa máxima de los términos de rompimiento de SUSY y la masa del gaugino M_1/\tilde{m}_s notamos que estos se concentran en un rango pequeño y no tiene un alto impacto en la NN, el mismo razonamiento aplicado a la masa del gaugino, los valores son más estrechos por el propio rango del que provienen como se muestra en la fig. 21b.

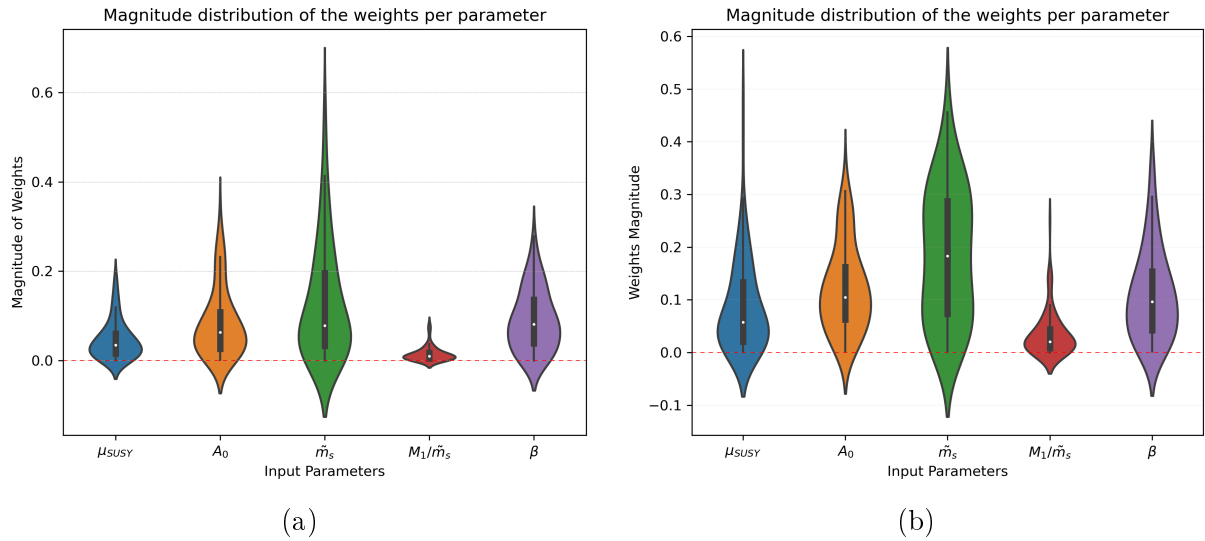


Figura 30: Gráfico de violín de la distribución de los pesos en cada parámetro del MSSM utilizado, de lado izquierdo vemos la primera NN con sus parámetros siendo ajustados para proseguir a crear las demás capas. En el lado derecho se observan los parámetros de la 2da NN.

5.2 RESULTADOS DE PRUEBA

En esta sección vamos a partir de la sección 5.1 anterior para usar el mejor modelo obtenido de la validación para la contribución del $g - 2$ mediante el proceso $\mu \rightarrow \mu + \gamma$ establecido en nuestra NN para analizar su rendimiento y performance con los datos de prueba.

En nuestro rendimiento obtuvimos un entrenamiento interesante, dando lo siguiente: $R_{1train}^2 = 0.92 \pm 1$, $R_{2train}^2 = 0.93 \pm 1$ y $R_{model}^2 = 0.91 \pm 1$, lo cuál quiere decir que $R_{2train}^2 > R_{1train}^2 > R_{model}^2$, esto nos podría dar a entender que el modelo mejora y después empeora un poco, pero ciertamente el entrenamiento se puede visualizar de una manera ajena a estas métricas ya que hay veces que este orden de mejora se invierte entre la 1era NN y la 2da NN, esto no lo podemos tomar como una medida determinista de que haya o no un sobreajuste ya que al ver como se distribuyen los puntos en la fig. 23 y 28 lo que observamos es que independientemente del coeficiente de determinación los valores de predicción abarcan un rango más amplio ya que se van expandiendo a los extremos, cosa que en la 1era NN siempre carece un poco de ello, en contraste con la 2da donde hay una mejora en esta característica, las métricas también nos muestran esta mejora. Otra forma de analizar estos resultados para nuestro modelo final es con el gráfico expuesto en la fig. 31, más adecuado que los anteriores porque muestra el margen ideal de entrenamiento de mejor manera, el cuál debe estar cerca de $y = x$, si utilizáramos los gráficos ante-

riores se observaría de manera saturada ya que las predicciones cubren más los valores reales.

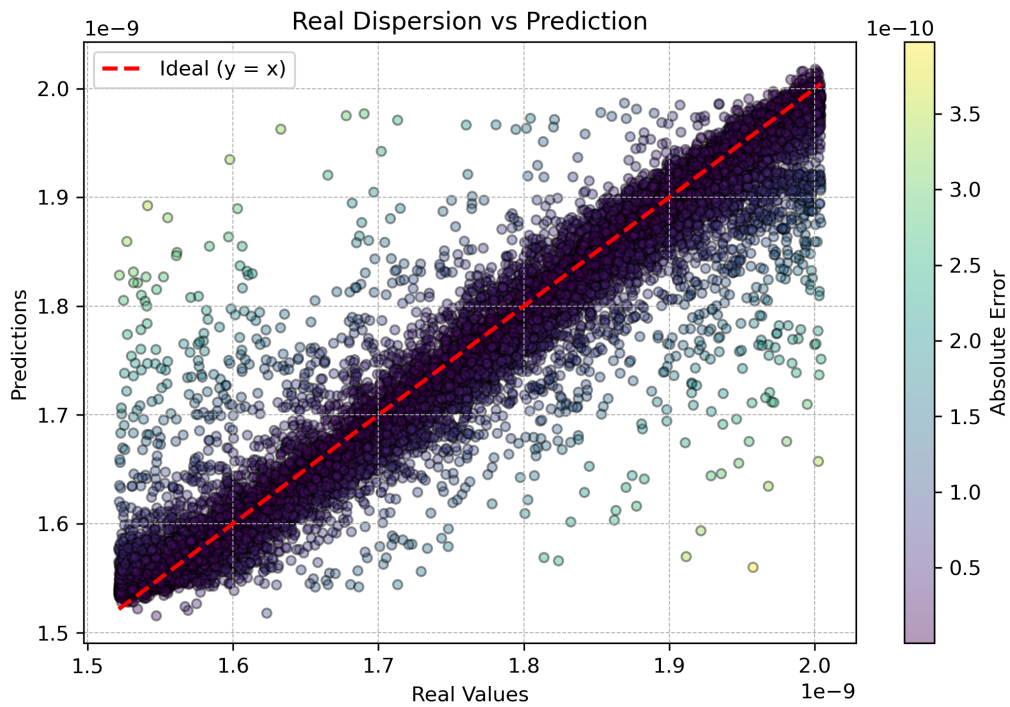


Figura 31: Modelo final expuesto a los datos de prueba, donde $y = x$ nos indica la forma en que la dispersión real y las predicciones coinciden, la barra de color nos muestra que mayormente los puntos de predicción tienen un error muy bajo y están cerca de la dispersión real.

El error absoluto es muy bajo y evidentemente todos los puntos están concentrados cerca y alrededor de la recta esperada, donde los más alejados es claro que tienen un error más grande debido a los colores que presentan, pero cabe mencionar que estos solo constituyen una cantidad muy pequeña de valores. Podemos entonces decir que nuestro modelo ha sido entrenado de una manera aceptable, probablemente podría ser mejorado con una NN más elaborada, ya sea una Red Bayesiana como se utiliza en [8] o alguna otra.

Por último en la fig. 32a podemos observar con más precisión el error absoluto, confirmando que este es muy pequeño y su promedio se concentra casi en 0 lo cuál es algo esperado y necesario para construir un buen modelo, en la fig. 32b se presenta como se distribuyen nuestros valores del modelo final junto con los reales, afortunadamente volvemos a confirmar que nuestras predicciones tienen una buena tendencia sin embargo tiene algunas zonas en los bordes que se podrían mejorar.

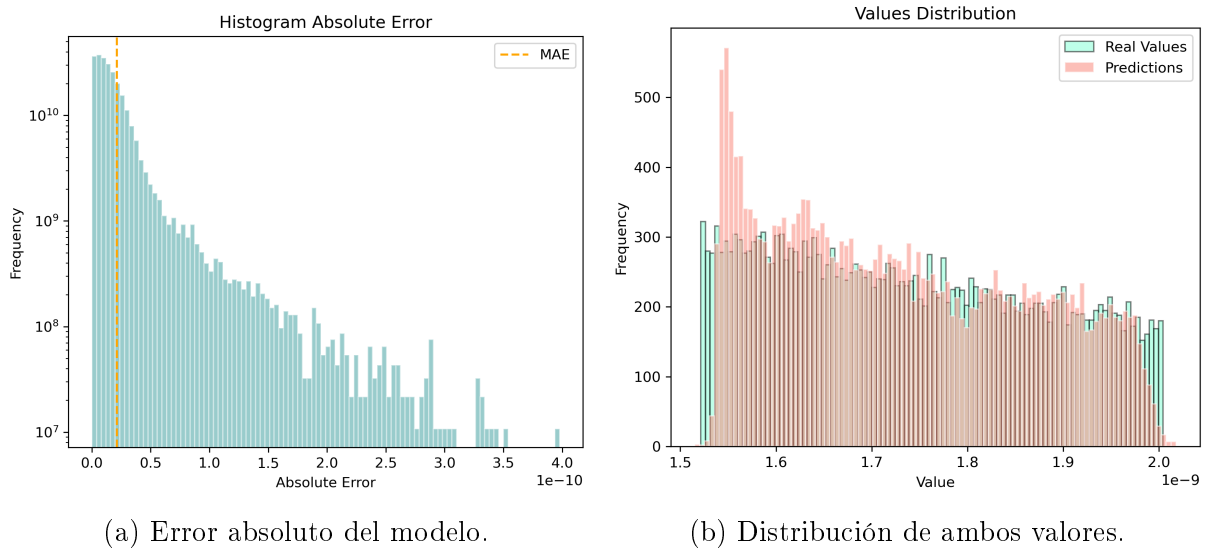


Figura 32: En el panel izquierdo tenemos como se observa el error absoluto con una línea vertical que nos indica donde se encuentra el promedio, en el panel derecho vemos la distribución de valores de nuestras predicciones junto con los valores reales.

5.3 TIEMPO DE COMPUTO

SUSY Code Time	1 ^{ra} NN Time	2 ^{da} NN Time	Test NN Time
0:01:08	0:02:05	0:01:49	0:00:03

Tabla 7: Tiempos de cómputo promedio para cada etapa. El tiempo de código SUSY corresponde a la generación de datos de entrada. El entrenamiento se realizó con dos arquitecturas iguales para cada una $[5, 128, 128, 128, 1]$ (primera y segunda NN). Dado que se generan valores aleatorios para la entrada entonces trabajamos $\sim 10^6$ que nos da el modelo SUSY, después las muestras se reducen a $\sim 2 \times 10^4$. El tiempo de prueba es el necesario para evaluar el modelo final sobre los datos. El formato de los tiempos es horas:minutos:segundos.

El tiempo tomado para entrenar nuestro modelo fue variando un poco, como se puede observar en la tabla 7 al principio obtuvimos un tiempo de 2min aproximadamente, difiriendo muy poco en nuestro segundo entrenamiento del modelo por tan solo unos segundos pero aún así reduciéndose. Terminado el modelo en las pruebas el rendimiento mejoró considerablemente, haciendo una gran diferencia con el modelo analítico de SUSY el cuál tomaba aproximadamente 1min contra nuestro modelo de 3s, lo cuál nos hace ver

que entrenar un modelo que obtenga estas predicciones de la contribución es útil en costo computacional.

CONCLUSIONES

El objetivo principal fue construir una red neuronal de regresión que pudiera arrojar buenas predicciones, con el propósito de encontrar una manera alternativa para calcular la contribución de la anomalía del muon mediante la interacción del neutralino-lepton-slepton, esto funcionó de una manera aceptable, como vimos nuestro tiempo de entrenamiento fue mayor que el propio cálculo analítico, pero esto es algo esperado en una red neuronal, lo realmente importante es el tiempo una vez que se ha entrenado el modelo, puesto que este modelo se puede guardar y cargar para ser reutilizado sin necesidad de más entrenamiento. Dicho lo anterior podemos decir que se cumplió el objetivo, ahora en nuestros objetivos secundarios buscamos mayor rapidez en la obtención de estas contribuciones, ya que es adecuado que al buscar un método alternativo a los cálculos también tenga una propia ventaja, como lo es la eficiencia y rapidez, cosa que también se cumplió con una diferencia significativa. Podemos decir que esos son los resultados más relevantes y aceptables que se pudieron obtener. Entre otras cosas a considerar sobre este acercamiento al problema es: contemplar el hecho de qué tan bueno o cuándo es más conveniente aplicar el transfer learning, porque en nuestro caso tuvo un ligero impacto, como vimos en el capítulo de resultados, la diferencia aunque es ciertamente notable, como el hecho de que parece expandir la predicción ya que con la primer red queda un cierto espacio en la frontera que no se cubre, cuando aplicamos la segunda red parece expandirse un poco más el radio de predicción, pero esto con el costo de no mejorar totalmente o significativamente el R^2 aunque es necesario recalcar que esta métrica no es definitiva y como hablamos en la sección de Machine Learning, se debe complementar las métricas, por su propia definición, puesto que si tenemos una varianza muy baja puede conllevarnos a un alto coeficiente R^2 pero a priori buscamos un MSE bajo también, esto significa que no podemos basarnos en una sola métrica, afortunadamente nuestro MSE y MAE fueron muy bajos, lo que nos da una gran certeza de nuestros resultados. También es posible examinar agregar otras métricas como RMSE(Root Mean Squared Error) ya que es más intuitiva y da las mismas unidades, a diferencia de MSE que penaliza los errores más grandes, o se puede examinar alguna otra métrica para complementar.

En consideraciones futuras para seguir este trabajo podemos pensar en alguna otra téc-

nica de regularización o una Red Neuronal Bayesiana, esta nos brindaría más libertad de ajustar los mejores pesos y así obtener una mejora en el rendimiento ya que aunque nuestro modelo generaliza bien, aún tiene mucho que mejorar puesto que no se pudo llegar a un $R^2 = 0.95$ o disminuir más todas las métricas contempladas, esto hace que ciertos valores se escapen un poco en los bordes nuevamente como ya vimos en la fig. 31, adicional a esto podemos agregar otra mejora al modelo que podría ser aumentar el número de capas y de datos, haciendo una simulación más numerosa. Partiendo de lo anterior algo que se considera muy relevante es emplear la misma distribución gaussiana que usamos nosotros pero tomar en cuenta las implicaciones que esta tiene ya que acota mucho la cantidad de parámetros, o buscar alguna otra que haga al modelo converger más rápido de una manera eficiente y equilibrada. Prestar atención al cociente M_1/\tilde{m}_s es algo que se puede tomar en cuenta para mejorar nuestro conjunto de datos, una alternativa a lo que se realizó en este trabajo podría ser hacer es una simulación Monte Carlo para obtener un M_1 y observar qué tan bien o mal repercute en la generalización del modelo. Otro punto importante es aclarar que las relaciones de este modelo son específicas para el mismo, es decir; no podemos dar argumentos concluyentes sobre la correlación de la fig. 30 con el modelo analítico ya que propiamente este modelo de regresión ajusta de manera numérica y óptima para darnos los resultados de la anomalía del muon, en un principio se pensó que tal vez el hecho de que \tilde{m}_s tuviese una alta dispersión y repercusión en la primera capa de la NN podría indicarnos algo acerca de la física del modelo de SUSY y de sus parámetros contemplados en este trabajo pero dichas ideas se contemplan solo a nivel del modelo numérico, no a un nivel global de la teoría, por lo que es necesario aclarar esta idea para no adentrarnos en preposiciones que por el momento y en este trabajo no se han demostrado con la información recavada. Como perspectiva de **trabajos futuros** algunas de las ideas que se mencionaron podrían ser consideradas e implementadas para seguir este trabajo.

REFERENCIAS

- [1] Tom Lancaster and Stephen J. Blundell. *Quantum Field Theory for the Gifted Amateur*. Oxford, 2014.
- [2] Walter Greiner and Joachin Reinhardt. *Quantum Electrodynamics*. Springer, 2009.
- [3] Stephen P. Martin. A supersymmetry primer. arXiv:hep-ph/9709356, 1997. Version 7, January 2016.
- [4] F. V.Flores-Baez, M Gómez Bock, and M. Mondragon. Muon $g - 2$ through a flavor structure on soft susy terms. *The European Physical Journal C*, 76(561), 2016. doi: 10.1140/epjc/s10052-016-4402-4.
- [5] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [6] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, Cambridge, Massachusetts, 2016. URL <http://www.deeplearningbook.org>.
- [7] Qiang Yang, Yu Zhang, Wenyuan Dai, and Sinno Jialin Pan. *Transfer Learning*. Cambridge University Press, Cambridge, United Kingdom, 2020. ISBN 9781107016903. doi: 10.1017/9781139061773. URL <https://www.cambridge.org/9781107016903>.
- [8] Sønderland and Per-Dimitri B. Supersymmetry cross section predictions with bayesian neural networks. Master’s thesis, University of Oslo, Oslo, Norway, June 2021.
- [9] Carl D. Anderson and Seth Neddermeyer. New evidence for the existence of a particle of mass intermediate between the proton and electron. *Physical Review*, 52(1003), 1937. doi: <https://doi.org/10.1103/PhysRev.52.1003>.
- [10] Charpak G., Farley F. J. M., Garwin R. L., Muller T., Sens J. C., Telegdi V. L., and Zichichi A. Measurement of the anomalous magnetic moment of the muon. *Physical Review D*, 6(3):128—132, 1961. doi: 10.1103/physrevlett.6.128.
- [11] Kinoshita, T., Niić, B., Okamoto, and Y. Hadronic contributions to the anomalous magnetic moment of the muon. *Physical Review D*, 31(8):2108—2119, 1985. doi: [doi:10.1103/physrevd.31.2108](https://doi.org/10.1103/physrevd.31.2108).

- [12] H. N. Brown, G. Bunce, R. M. Carey, et al. Improved measurement of the positive muon anomalous magnetic moment. *Physical Review D*, 62(091101(R)), 2000. doi: 10.1103/physrevd.62.091101.
- [13] D. Giusti, V. Lubicz, G. Martinelli, et al. Electromagnetic and strong isospin-breaking corrections to the muon $g-2$ from lattice qcd + qed. *Physical Review D*, 99(114502), 2019. doi: 10.1103/PhysRevD.99.114502.
- [14] M.Knecht, S.Narison, and D.Rabetiarivony A.Rabemananjara. Scalar meson contributions to a from hadronic light-by-light scattering. *Physics Letters B*, 787:111–123, 2018. doi: <https://doi.org/10.1016/j.physletb.2018.10.048>.
- [15] Ayuki Kamada, Kunio Kaneta, Keisuke Yanagi, and Hai-Bo Yu. Self-interacting dark matter and muon ($g-2$) in a gauged $u(1)_\mu$ - l - τ model. *High Energy Physics*, 2018 (117), 2018. doi: <https://doi.org/10.1007/JHEP06%282018%29117>.
- [16] T. Aoyama, N. Asmussen, M. Benayoun, et al. The anomalous magnetic moment of the muon in the standard model. *Physics Reports*, 887(3):1–166, 2020. doi: <https://doi.org/10.1016/j.physrep.2020.07.006>.
- [17] B.Abi, T Albahri, and S. Al-Kilani. Measurement of the positive muon anomalous magnetic moment to 0.46 ppm. *Physical Review Letters*, 126(141801), 2021. doi: 10.1103/PhysRevLett.126.141801.
- [18] D.P. Aguillard, T. Albahri, D. Allspach, et al. Detailed report on the measurement of the positive muon anomalous magnetic moment to 0.20 ppm. *arXiv:2402.15410v2*, 2024. doi: arXiv:2402.15410v2.
- [19] R. Aliberti, T. Aoyama, E. Balzani, et al. The anomalous magnetic moment of the muon in the standard model: an update. *arXiv preprint arXiv:2506.21476*, 2025. URL <https://arxiv.org/abs/2506.21476>. CERN-TH-2025-101, FERMILAB-PUB-25-0344-T, KEK Preprint 2025-22, IPARCOS-UCM-25-029, MITP-25-037, UWThPh 2025-15, ZU-TH 37/25.
- [20] T. Aoyama et al. The anomalous magnetic moment of the muon in the standard model. *Physics Reports*, 887:1–166, 2020. doi: 10.1016/j.physrep.2020.07.006. URL <https://www.sciencedirect.com/science/article/pii/S0370157320302556>.
- [21] Franz Mandl and Graham Shaw. *Quantum Field Theory*. Wiley, 2010.
- [22] Carroll and Sean M. *Spacetime and geometry*. Cambridge University Press, 2019.
- [23] Walter Greiner. *Relativistic Quantum Mechanics*. Springer, 2000.

-
- [24] Goldstein Herbert, Poole Charles, and Safko John. *Classical mechanics*. Pearson, 2002.
- [25] Sakurai, Jun John, Commins, and Eugene D. *Modern quantum mechanics, revised edition*. Cambridge University Press, 1995.
- [26] Griffiths, David J., Schroeter, and Darrell F. *Introduction to quantum mechanics*. Cambridge University Press, Cambridge ; New York, NY, third edition edition, 2018. ISBN 978-1-107-18963-8.
- [27] Mandl Franz and Shaw Graham. *Quantum field theory*. John Wiley & Sons, 2010.
- [28] Peskin, Michael E., Schroeder, and Daniel V. *An Introduction to Quantum Field Theory*. Addison-Wesley Publishing Company, Reading, Massachusetts, 1995.
- [29] Lewis H. Ryder. *Quantum Field Theory*. Cambridge University Press, 1996.
- [30] David Tong. Quantum field theory. *Department of Applied Mathematics and Theoretical Physics*, 2007.
- [31] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory*. Addison-Wesley, 1995.
- [32] Martin F. Sohnius. Introducing supersymmetry. *Physics Reports*, 128(2–3):39–204, 1985. Review Section of Physics Letters.
- [33] H.J.W. Müller-Kirsten and A. Wiedemann. *Supersymmetry. An introduction with conceptual and calculation details*. World Scientific Publishing, 1987.
- [34] S. P. Martin. A supersymmetry primer. *Advances in Series in Direct High Energy Physics*, 21:1, 2010. <https://arxiv.org/abs/hep-ph/9709356>.
- [35] Dominik Stöckinger. The muon magnetic moment and supersymmetry. *Journal of Physics G: Nuclear and Particle Physics*, 34(R45), 2006. doi: 10.1088/0954-3899/34/2/R01.
- [36] R. Haag, J. T. Lopuszanski, and M. Sohnius. All possible generators of supersymmetries of the s matrix. *Nucl. Phys. B*, 88:257, 1975. doi: 10.1016/0550-3213(75)90279-5.
- [37] M. F. Sohnius. *Introducing Supersymmetry*, volume 128. Elsevier B.V., 1985.
- [38] S. R. Coleman and J. Mandula. All possible symmetries of the s matrix. *Phys. Rev.*, 159:1251–1256, 1967. doi: 10.1103/PhysRev.159.1251.
- [39] Mikio Nakahara. *Geometry, Topology and Physics*. CRC Press, 2 edition, 2003.

-
- [40] G. Passarino and M. Veltman. One Loop Corrections for e^+e^- Annihilation Into $\mu^+\mu^-$ in the Weinberg Model. *Nucl. Phys. B*, 160:151, 1979. doi: 10.1016/0550-3213(79)90234-7.
- [41] Dominik Stockinger. The muon magnetic moment and supersymmetry. *Journal of Physics G: Nuclear and Particle Physics*, 34:R45, 2007. doi: 10.1088/0954-3899/34/2/R01.
- [42] Kevin P. Murphy. *Machine Learning A Probabilistic Perspective*. Massachusetts Institute of Technology, 2012.
- [43] K. Funahashi. On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 2:183–192, 1989.
- [44] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(3):303–314, 1989.
- [45] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.
- [46] N. E. Cotter. The stone–weierstrass theorem and its application to neural networks. *IEEE Transactions on Neural Networks*, 1(4):290–295, 1990.
- [47] Y. Itô. Approximation of functions on a compact set by finite sums of a sigmoid function without scaling. *Neural Networks*, 4(1):105–115, 1991.
- [48] Y. Itô. Approximation capabilities of layered neural networks with sigmoidal units on two layers. *Neural Computation*, 6(6):1233–1243, 1994.
- [49] K. Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991.
- [50] B. D. Ripley. Neural networks and related methods for classification. *Neural Networks*, 8(31):31–37, 1996.
- [51] LeCun, Yann, Boser, Bernhard, Denker, John S, Henderson, Donnie, Howard, Richard E, Hubbard, Wayne, Jackel, and Lawrence D. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, 1989.
- [52] Kingma, Diederik P, Ba, and Jimmy. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [53] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML)*, pages 807–814, 2010. doi: 10.5555/3104322.3104425.

-
- [54] Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-normalizing neural networks. In *Advances in Neural Information Processing Systems 30*, 2017.
- [55] Prajit Ramachandran, Barret Zoph, and Quoc V. Le. Searching for activation functions. In *International Conference on Learning Representations (ICLR)*, 2018. doi: 10.48550/arXiv.1710.05941.
- [56] Charles Dugas, Yoshua Bengio, François Bélisle, Claude Nadeau, and René Garcia. Incorporating second-order functional knowledge for better option pricing. In *Advances in Neural Information Processing Systems 13*, pages 472–478, 2000. doi: 10.5555/3008751.3008817.
- [57] Murphy and Kevin P. *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge, Massachusetts; London, England, 2012. ISBN 978-0-262-01802-9.
- [58] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feed-forward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9, pages 249–256. PMLR, 2010. URL <https://proceedings.mlr.press/v9/glorot10a.html>.
- [59] A. Clifford Cohen. *Truncated and Censored Samples: Theory and Applications*. CRC Press, Taylor & Francis Group, Boca Raton, FL, 1991. ISBN 978-0-8247-8447-8.
- [60] Martín Abadi et al. TensorFlow: Large-scale machine learning on heterogeneous systems. <https://www.tensorflow.org/>, 2015. URL <https://www.tensorflow.org/>.
- [61] François Chollet et al. Keras. <https://keras.io>, 2015.