
Análise Tomográfica de Redes Neurais aplicada a Problemas Quânticos de Muitos Corpos

Mardônio Jó Carvalho de França



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE FÍSICA
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

Uberlândia
2026

Mardônio Jó Carvalho de França

**Análise Tomográfica de
Redes Neurais aplicada a Problemas Quânticos
de Muitos Corpos**

Dissertação de mestrado apresentada ao
Programa de Pós-graduação de Física da
Universidade Federal de Uberlândia como
parte dos requisitos para a obtenção do
título de Mestre em Física.

Área de concentração: Física

Orientador: Edson Vernek

Uberlândia

2026

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

F814a
2026

França, Mardônio Jó Carvalho de, 1974-
Análise Tomográfica de Redes Neurais aplicada a Problemas
Quânticos de Muitos Corpos [recurso eletrônico] / Mardônio Jó
Carvalho de França. -- 2026.

Orientador: Edson Vernek.
Dissertação (Mestrado) -- Universidade Federal de Uberlândia,
Programa de Pós-graduação em Física.
Modo de acesso: Internet.
Disponível em: <http://doi.org/10.14393/ufu.di.2026.11035>
Inclui bibliografia.
Inclui ilustrações.

1. Física. I. Vernek, Edson, 1973-, (Orient.). II. Universidade
Federal de Uberlândia. Programa de Pós-graduação em Física. III.
Título.

CDU: 53

André Carlos Francisco
Bibliotecário(a)-Documentalista - CRB-6/3408



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Física				
Defesa de:	Dissertação de Mestrado				
Data:	05 de fevereiro de 2026	Hora de início:	14:00	Hora de encerramento:	16:20
Matrícula do Discente:	12322FIS005				
Nome do Discente:	Mardônio Jo Carvalho de França				
Título do Trabalho:	<i>Análise Tomográfica de Redes Neurais aplicada a Problemas Quânticos de Muitos Corpos</i>				
Área de concentração:	Física				
Linha de pesquisa:	Sistemas fortemente correlacionados				
Projeto de Pesquisa de vinculação:	N.A				

Reuniu-se, por meio de videoconferência, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Física, assim composta: Professores Doutores: Alexandre Marletta - INFIS/UFU, Antônio Carlos Ferreira Seridonio - UNESP, e Edson Vernek - INFIS/UFU, orientador do discente.

Iniciando os trabalhos o presidente da mesa, Prof. Edson Vernek, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao discente a palavra para a exposição do seu trabalho. A duração da apresentação do discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(às) examinadores(as), que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

Aprovado

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação

interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **ANTONIO CARLOS FERREIRA SERIDONIO, Usuário Externo**, em 05/02/2026, às 16:55, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Alexandre Marletta, Professor(a) do Magistério Superior**, em 05/02/2026, às 18:20, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Edson Vernek, Presidente**, em 06/02/2026, às 08:57, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **7040606** e o código CRC **76673BC7**.

*... conhecimento e imaginação: labirinto de multi-versos,
versos sobre pequenas pétalas, pequenos universos...
beija-flores.*

*Noite azul, pedra e chão.
Nada a temer, nada a conquistar.
Lô Borges*

(...) ode ao poema em linha reta de Fernando Pessoa.

*Não sei o que faço / A minha vida é uma luta sem fim.
Me sinto no espaço / O tempo todo à procura de mim.
Paulo Sérgio*

*A lagarta tem forma de serpente.
Cego Aderaldo*

Agradecimentos

Agradeço a Deus.

Agradeço à minha mãe, Francimar Carvalho de França, ela é bússola em minha vida.

Ao meu amor, Vláilda Rodrigues, que caminha comigo nesta jornada.

Aos meus filhos, Davi e Hilbert, com quem aprendi sobre o universo e o infinito.

Aos meus filhos-pássaros: Chicó, Filó e Blue, que enchem a casa de alegria e luz.

Aos meus amigos da pós-graduação em Física da UFU, Marcos, Caio, Gabriel, Olavo, Chris, Maysa e Márcio, que me ajudaram a construir este trabalho.

Aos meus amigos-poetas-irmãos Nuno, Cláudio Reis, Demétrios, Lucas Miranda, Misael e Nenas que partilham de sonhos e devaneios.

Ao meu orientador, Edson Vernek.

Aos professores Alexandre Marletta, Fabrício Macedo de Souza e Gerson J. Ferreira.

Ao Instituto de Física da Universidade Federal de Uberlândia.

Agradeço à CAPES pelo apoio financeiro.

À arte.

À música.

À poesia.

Aos físicos que inventaram esse parque de diversões conhecido por mecânica quântica.

I think I can safely say that nobody understands quantum mechanics.

Richard Feynmann

Resumo

O uso de redes neurais artificiais para representar estados quânticos de muitos corpos tem despertado grande interesse nos últimos anos, oferecendo uma abordagem promissora para lidar com problemas complexos que surgem neste contexto. Neste trabalho propomos uma tomografia dos pesos de uma rede neural aplicada ao problema para obter o estado fundamental de cadeias de spins 1D com interação entre vizinhos. Um dos objetivos é entender como métodos de aprendizado de máquina não supervisionado possa identificar regimes em cadeias de spins 1D, tais como ferromagnético, antiferromagnético e frustrado. Nosso trabalho busca contribuir para o entendimento como as redes neurais incorporam características físicas de modelos físicos como o emaranhamento em seus pesos sinápticos.

Palavras-chave: Redes neurais artificiais. Estados quânticos de muitos corpos.

Abstract

The use of artificial neural networks to represent many-body quantum states has sparked great interest in recent years, offering a promising approach to addressing complex problems arising in this context. In this work, we propose a tomography of the weights of a neural network applied to the problem to obtain the ground state of 1D spin chains with neighbor interaction. One of the objectives is to understand how unsupervised machine learning methods can identify regimes in 1D spin chains, such as ferromagnetic, antiferromagnetic, and frustrated regimes. Our work seeks to contribute to the understanding of how neural networks incorporate physical characteristics of physical models, such as entanglement, into their synaptic weights.

Keywords: Artificial neural networks. Quantum many-body states .

Lista de ilustrações

- Figura 1 – Exemplo ilustrativo de uma cadeia de spin com regime antiferromagnético com L sítios com d de distância em cada sítios. Para uma cadeia de spin é possível definir a condição de contorno periódica, ou seja, o último sítio é vizinho do primeiro sítio. 33
- Figura 2 – Função sigmoide logística $\sigma(x) = \frac{1}{1+e^{-x}}$, frequentemente utilizada como função de ativação em redes neurais. Essa função mapeia valores reais em um intervalo entre 0 e 1, sendo útil para introduzir não-linearidades no modelo e para representar probabilidades associadas à ativação de neurônios. Sua forma suave e derivável facilita o processo de retropropagação do erro durante o treinamento. 38
- Figura 3 – Estrutura básica de uma rede neural artificial com 3 camadas sendo 1 camada de entrada com 3 neurônios, 1 camada interna com 3 neurônios e uma camada de saída com 1 neurônio. Nesta arquitetura, todos os neurônios de entradas se conectam com os neurônios da camada interna. 39
- Figura 4 – O Aprendizado Supervisionado corresponde ao tipo de aprendizagem de máquina onde cada elemento do conjunto X possui uma correspondência unívoca (única e exclusiva) com um elemento do conjunto Y . Nesse caso, estabelece-se uma bijeção entre os conjuntos X e Y . A rede neural artificial ajusta seus pesos internos para aprender e reproduzir esse mapeamento. 40

Figura 5 – O aprendizado não-supervisionado possui diversas aplicações, como a identificação de agrupamentos (clusters) por meio de métodos como o K-Means. Além disso, pode ser empregado para a visualização de dados em duas ou três dimensões utilizando técnicas de redução de dimensionalidade, como o Principal Component Analysis (PCA). Neste exemplo, apresentamos uma simulação contendo três agrupamentos e seus respectivos centroides. A clusterização é realizada com base em métricas de distância geométrica, e os dados são representados nas duas componentes principais.	41
Figura 6 – Parametrização dos acomplamentos $J_1 - J_2$, em para $\theta \in [0, 2\pi)$. Em $T = 0$, conforme [3], o modelo $J_1 - J_2$ apresenta 3 fases: antiferromagnética, frustrada e ferromagnética.	43
Figura 7 – Estrutura geral para uma Rede Neural do tipo Máquina de Boltzmann Restrita (RBM), onde temos a representação da função de onda de um sistema de N spins. Cada configuração visível $\sigma = (\sigma_1^z, \sigma_2^z, \dots, \sigma_n^z)$ é conectada a neurônios na camada oculta $\mathbf{h} = (h_1, h_2, \dots, h_n)$ [9].	46
Figura 8 – Simulação com 100 <i>jobs</i> para uma cadeia de spins de tamanho $L = 8$, com 500 interações e acoplamento $\theta = \pi/4$, utilizando uma rede neural do tipo FFNN ($L = 8$, $\theta = \pi/4$, $IT = 500$). Nesta situação, temos por exemplo, alguns <i>jobs</i> apresentam um valor da energia muito distante do valor da energia exata.	52
Figura 9 – Na visualização gráfica do tipo <i>boxplot</i> para a simulação com 100 <i>jobs</i> ($NJ = 100$) para uma cadeia de spin de tamanho igual a 8, com 500 interações, e valor de acoplamento igual a $\pi/4$ e tipo de rede neural FFNN ($L = 8$, $\theta = \pi/4$, $IT = 500$). Observa-se a presença de alguns valores fora do intervalo de valores do <i>boxplot</i> , caracterizando a ocorrência de <i>outliers</i> no conjunto dos resultados.	52
Figura 10 – Simulação com 100 <i>jobs</i> ($NJ = 100$) para uma cadeia de spins de tamanho $L = 8$, com 500 interações e acoplamento $\theta = \pi/4$, utilizando uma rede neural do tipo FFNN ($L = 8$, $\theta = \pi/4$, $IT = 500$). Nesta simulação, os <i>outliers</i> foram removidos, resultando em valores de energia bastante próximos ao valor da energia exata. . . .	53
Figura 11 – Na visualização gráfica do tipo <i>box plot</i> referente à simulação de 100 <i>jobs</i> ($NJ = 100$) para uma cadeia de spins de tamanho $L = 8$, com 500 interações e acoplamento $\theta = \pi/4$, utilizando uma rede neural do tipo FFNN ($L = 8$, $\theta = \pi/4$, $IT = 500$) com os <i>outliers</i> removidos, observa-se que a distribuição dos valores concentra-se predominantemente entre o primeiro e o segundo quartil.	53

Figura 12 – Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações e acoplamento $\theta = \pi/4$, utilizando uma rede neural do tipo Feed-Forward Neural Network (FFNN) ($L = 8, \theta = \pi/4, IT = 500$). Observa-se que a partir do ciclo de interação próximo ao valor 180, o método converge para o valor da energia mínima. Em nossa simulação geral, esse <i>job</i> correspondeu à terceira execução dentre um total de 100 execuções.	55
Figura 13 – Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações e acoplamento $\theta = \pi/4$, utilizando uma rede neural do tipo FFNN ($L = 8, \theta = \pi/4, IT = 500$) com um <i>job</i> que na análise de dados foi caracterizado com um <i>outlier</i> sendo a simulação descartada. Em nossa simulação geral, esse <i>job</i> correspondeu à décima primeira execução dentre um total de 100 execuções.	55
Figura 14 – Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para os seguintes acoplamentos $\theta = [\pi/4, 7\pi/6, 5\pi/3]$, utilizando uma rede neural do tipo FFNN ($L = 8, IT = 500$). A execuções selecionados foram as seguintes : (i) para $\theta = \pi/4$, o 3 ^o <i>job</i> ; (ii) para $\theta = 7\pi/6$, o 70 ^o <i>job</i> e (iii) para $\theta = 5\pi/3$, o 72 ^o <i>job</i> de 100 execuções para cada simulação.	56
Figura 15 – Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para os seguintes acoplamentos $\theta = [\pi/6, \pi/4, \pi/3, 2\pi/3, 5\pi/3]$, utilizando uma rede neural do tipo FFNN ($L = 8, IT = 500$).	57
Figura 16 – Representação dos Pesos da Rede Neural para Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = \pi/4$, utilizando uma rede neural do tipo FFNN ($L = 8, IT = 500$). Temos algumas flutuações próximo à interação 50 ($IT = 50$) e à interação 160 ($IT = 160$) nos valores dos pesos dos neurônios que podem ser associadas a flutuações no valor da energia calculada.	58
Figura 17 – Representação dos Pesos da Rede Neural para Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = 7\pi/6$, utilizando uma rede neural do tipo FFNN ($L = 8, IT = 500$).	59
Figura 18 – Representação dos Pesos da Rede Neural para Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = 5\pi/3$, utilizando uma rede neural do tipo FFNN ($L = 8, IT = 500$).	59
Figura 19 – Representação dos Pesos da Rede Neural para Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = \pi/4$, utilizando uma rede neural do tipo FFNN no <i>job</i> para a execução 11, situação que foi detectada como outlier ($L = 8, IT = 500, NJ = 11$). Verifica-se que a mudança brusca no valor da energia reflete também ba mudança dos valores dos pesos da rede neural e nas também nas métricas estatísticas como log da dispersão, mediana, média, desvio-padrão e log da amplitude.	60

Figura 20 – Trajetória do PCA aplicado a representação dos Pesos da Rede Neural para Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = \pi/4$, o 11 ^o <i>job</i>	61
Figura 21 – Trajetória do PCA aplicado a representação dos Pesos da Rede Neural para Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = \pi/4$, o 3 ^o <i>job</i> ; (ii) para $\theta = 7\pi/6$, o 70 ^o <i>job</i> e (iii) para $\theta = 5\pi/3$, o 72 ^o <i>job</i>	62
Figura 22 – Trajetória do PCA aplicado a representação dos Pesos da Rede Neural para Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = \pi/4$ com a seguinte sequências de <i>job</i> $NJ = [3, 4, 11]$. Observa-se que as trajetórias dos <i>jobs</i> 3 e 4 são bastante semelhantes, ao passo que a trajetória do <i>job</i> 11 difere significativamente, apresentando uma quase descontinuidade após à iteração 200 ($IT = 200$).	63
Figura 23 – Correlação de Pearson para a representação dos Pesos da Rede Neural para Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = 5\pi/3$, o 72 ^o <i>job</i> de um total de 100 execuções.	64
Figura 24 – Histograma da Correlação de Pearson para a representação dos Pesos da Rede Neural para Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = 5\pi/3$, o 72 ^o <i>job</i> de um total de 100 execuções.	65
Figura 25 – Boxplot da Correlação de Pearson para a representação dos Pesos da Rede Neural para Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = 5\pi/3$, o 72 ^o <i>job</i> de um total de 100 execuções.	66
Figura 26 – Correlação de Pearson para cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = 7\pi/6$, o 70 ^o e $\theta = 5\pi/3$, o 72 ^o <i>job</i>	67
Figura 27 – Histograma da Correlação de Pearson para cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = 7\pi/6$, o 70 ^o e $\theta = 5\pi/3$, o 72 ^o <i>job</i>	67
Figura 28 – Boxplot da Correlação de Pearson para cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = 7\pi/6$, o 70 ^o e $\theta = 5\pi/3$, o 72 ^o <i>job</i>	68
Figura 29 – Boxplot da Correlação de Pearson para cadeia de spins de tamanho $L = 8$, com 500 interações para a fase frustrada para os seguintes acoplamentos $\theta = \pi/6, \pi/4, \pi/3, 2\pi/3$ e $5\pi/6$	68
Figura 30 – Determinação da energia do estado fundamental em uma cadeia de spins 1/2 para diferentes valores de L e números de interações I . O sistema é configurado para $L = 12$ e $L = 20$, com 80 interações ($I = 80$) com Rede Neural Tipo Restricted Boltzmann Machine (RBM).	81
Figura 31 – Determinação da energia do estado fundamental em uma cadeia de spins 1/2 para diferentes valores de L e números de interações I . O sistema é configurado para $L = 12$ e $L = 20$, com 80 interações ($I = 80$) com Rede Neural Tipo N-Layer Feed-Forward Neural Network (NFFNN).	82

Figura 32 – Determinação da energia do estado fundamental em uma cadeia de spins $1/2$ para L fixo e para o regime antiferromagnético. Temos variações no tipo da rede neural aplicada. O sistema é configurado para $L = 24$ com 300 interações ($I = 100$) para 4 tipos de Rede Neurais: RBM, SRBM, FFNN, NFFNN.	83
Figura 33 – Trajetória dos pesos sinápticos no espaço de PCA ao longo de $I = 60$ iterações para $L = 12$, utilizando uma Rede Neural Artificial (RNA) tipo RBM. A figura mostra a evolução dos vetores de pesos desde o início até a convergência para os regimes ferromagnético e antiferromagnético.	89
Figura 34 – Trajetória dos pesos sinápticos no espaço de PCA ao longo de $I = 40$ iterações para $L = 20$, utilizando uma RNA tipo RBM. A figura mostra a evolução dos vetores de pesos desde o início até a convergência. . . .	90
Figura 35 – 3 tipos de correlação de Pearson.(a) Correlação positiva; (b) Correlação negativa e (c) Correlação neutra.	91
Figura 36 – Matriz de correlação de Pearson para a tabela de matriz de dados exemplo 1. Por construção, as colunas $C1-C6$ apresentam correlação positiva entre si, enquanto $C9$ e $C10$ apresentam correlação negativa. A imagem (a) apresenta a matriz completa, a (b) mostra o subconjunto das correlações positivas ($C1-C6$), e a (c) exhibe o subconjunto das correlações negativas ($C9-C10$) e ($C8-C10$). A correlação entre ($C8-C9$) é quase nula (valor próximo de 0).	93

Lista de siglas

AM Aprendizagem de Máquina

AMP Aprendizagem de Máquina por Reforço

DMRG Density Matrix Renormalization Group Algorithm

DBSCAN Density-Based Spatial Clustering of Applications with Noise

FFNN Feed-Forward Neural Network

GMM Gaussian Mixture Models

MQ Mecânica Quântica

MPS Matrix Product States

NNQS Neural Network Quantum States

NFFNN N-Layer Feed-Forward Neural Network

PCA Principal Component Analysis

PEPS Projected Entangled Pair State

RNA Rede Neural Artificial

RNAP Redes Neurais Artificiais Profundas

RBM Restricted Boltzmann Machine

SRBM Symmetric Restricted Boltzmann Machine

TTN Tree Tensor Networks

VMC Variational Monte Carlo

Sumário

1	INTRODUÇÃO	25
1.1	A complexidade do Espaço de Hilbert	26
1.2	Estrutura da Dissertação	31
2	FUNDAMENTOS TEÓRICOS	33
2.1	Teoria de Cadeias de Spin	33
2.1.1	Dois sítios	34
2.1.2	N-Sítios	36
2.2	Redes Neurais Artificiais em Estados Quânticos	37
2.2.1	Método Supervisionado	39
2.2.2	Método Não-Supervisionado	41
2.2.3	Conexão entre o Aprendizado de Máquina e a Mecânica Quântica	42
3	APLICAÇÃO DO NNQS AO MODELO $J_1 - J_2$	43
3.1	O Problema	43
3.2	Matrix Product States	44
3.3	Neural Network Quantum States	46
3.4	Aspectos Caixa-Preta da Rede Neural	49
4	RESULTADOS	51
4.1	Preparação do método para Modelo $J_1 - J_2$	51
4.2	Avaliação para o Cálculo da Energia	56
4.3	Tomografia da Rede Neural Artificial	58
5	CONCLUSÃO	71
	Referências	73

APÊNDICES 79

APÊNDICE A	–	MODELO DE HEISENBERG	81
APÊNDICE B	–	ANÁLISE EXPLORATÓRIA DE DADOS . . .	85
B.1		Principal Component Analysis - PCA	85
B.2		Correlação de Pearson	91

Introdução

A mecânica quântica completa cerca de um século desde sua formulação moderna, consolidada a partir dos trabalhos seminais de Werner Heisenberg, Erwin Schrödinger, Paul Dirac, Max Born, Paul Jordan, entre outros. Ao longo desse período, essa teoria estabeleceu-se como um dos pilares da física contemporânea, fornecendo a base conceitual para inúmeros avanços científicos e tecnológicos[46, 13, 38].

Grande parte das transformações que moldaram a sociedade moderna como o desenvolvimento dos semicondutores, a invenção do transistor e, consequentemente, a revolução da computação tem origem direta nos princípios quânticos, evidenciando o profundo impacto dessa teoria para além do âmbito acadêmico [1].

Atualmente, a sociedade vivencia a era da inteligência artificial, um campo cuja história remonta ao século passado, em meados dos anos 1950, iniciando-se com os trabalhos fundadores de Alan Turing, John von Neumann e outros pioneiros [54, 33, 41]. Embora frequentemente associada apenas a avanços computacionais recentes, a inteligência artificial é fruto de um longo processo de evolução teórica e tecnológica, fortemente apoiado no progresso da física, da matemática e da ciência da computação.

Nesse contexto, a interação entre mecânica quântica e inteligência artificial surge como uma fronteira científica promissora, abrindo novas possibilidades para o entendimento e o desenvolvimento de tecnologias emergentes.

Recentemente, em 2024, o Prêmio Nobel de Física foi concedido a John Hopfield e Geoffrey Hinton, em reconhecimento a contribuições fundamentais para o desenvolvimento das redes neurais artificiais. Esses trabalhos possuem raízes conceituais profundas em modelos físicos associados a fenômenos coletivos, em particular aqueles relacionados ao comportamento de spins em sistemas estatísticos, como exemplificado pelo modelo de Hopfield e suas conexões com o modelo de Ising [50].

Tal reconhecimento evidencia a forte relação entre a física estatística e os fundamentos teóricos das redes neurais, ressaltando como conceitos originalmente desenvolvidos no contexto da física podem ser reinterpretados e aplicados de forma decisiva em áreas interdisciplinares.

Atualmente, as redes neurais artificiais configuram-se como ferramentas de grande potencial para o tratamento de problemas de elevada complexidade computacional. No contexto da mecânica quântica de sistemas de muitos corpos, essas metodologias têm se mostrado particularmente promissoras para o cálculo de grandezas relevantes, como a energia do estado fundamental. Nesse sentido, em face ao crescimento exponencial do espaço de Hilbert com o número de graus de liberdade, métodos numéricos tradicionais tornam-se frequentemente limitados, o que torna abordagens baseadas em aprendizado de máquina uma alternativa inovadora e eficiente. Além do cálculo de energias, as redes neurais têm sido empregadas na investigação de propriedades quânticas como fenômenos de emaranhamento. A capacidade dessas arquiteturas de representar funções altamente complexas e não lineares permite avanços significativos na descrição de estados quânticos fortemente correlacionados. Portanto a integração entre redes neurais artificiais e mecânica quântica estabelece um paradigma interdisciplinar emergente, com implicações relevantes tanto para a compreensão dos fundamentos da física quanto para o desenvolvimento de novas estratégias computacionais.

1.1 A complexidade do Espaço de Hilbert

A mecânica quântica não-relativística descreve os sistemas físicos em escalas microscópicas por meio da chamada *Equação de Schrödinger* [13, 42]. Essa equação determina a evolução temporal do vetor de estado $|\psi(\vec{r}, t)\rangle$, que contém toda a informação sobre o sistema. A forma geral é representada por

$$\hat{H} |\psi(\vec{r}, t)\rangle = i\hbar \frac{\partial}{\partial t} |\psi(\vec{r}, t)\rangle, \quad (1)$$

onde $\hat{H}$ é hamiltoniano dos sistema. $\hat{H}$ é um operador hermitiano que contém, em princípio, todas as informações do sistema. No caso particular de sistemas que não dependem explicitamente do tempo (sistemas estacionários), a *Equação de Schrödinger* assume a forma autovalor, conhecida como *Equação de Schrödinger Independente do Tempo*

$$\hat{H} |\psi_n(\vec{r})\rangle = E_n |\psi_n(\vec{r})\rangle. \quad (2)$$

Aqui E_n denota os níveis de energia permitidos (autovalores) e $|\psi_n(\vec{r})\rangle$ corresponde os estados quânticos estacionários (autovetores associados). Tal formulação é central na análise espectral de sistemas quânticos. Ressalte-se que os estados quânticos pertencem a um espaço vetorial complexo de dimensão infinita, completo em relação à norma induzida pelo produto interno, denominado *espaço de Hilbert*.

Com o intuito de ilustrar a complexidade envolvida na resolução de problemas associados à determinação dos níveis de energia permitidos em sistemas com muitas partículas, consideraremos, a título de exemplo, sistemas compostos por partículas com spin semi-inteiros (férmions), $S = \frac{1}{2}$.

Consideremos, primeiramente o caso de uma única partícula. Um estado quântico genérico de spin pertence ao *espaço de Hilbert* $\mathcal{H}$ e pode ser descrito como

$$|\psi\rangle = \{\alpha |\uparrow\rangle + \beta |\downarrow\rangle\}; \alpha, \beta \in \mathbb{C} \quad (3)$$

onde a base $B(\mathcal{H})$ é composta pelos kets $|\uparrow\rangle$ e $|\downarrow\rangle$.

Nesta situação, a dimensão do Espaço de Hilbert é 2. De uma forma geral, um estado arbitrário $|\psi\rangle$ poderá ser escrito de forma compacta,

$$|\psi\rangle = \sum_{a'} C_{a'} |a'\rangle, \quad (4)$$

onde $C_{a'}$ são os coeficientes complexos que representam as amplitudes de probabilidades associadas a cada estado da base ($C_{a'} \in \mathbb{C}$) e $|a'\rangle \in \{|\uparrow\rangle, |\downarrow\rangle\}$ são os vetores ortonormais que formam uma base no espaço de Hilbert como foi definido na equação (3) para um sistema de 1 partícula com spin $\frac{1}{2}$.

A ortononormalidade dos vetores de base é expressa por

$$\langle a'' | a' \rangle = \delta_{a'' a'}, \quad (5)$$

onde δ representa o delta de Kronecker. Os coeficiente $C_{a'}$ são obtidos explicitamente por

$$C_{a'} = \langle a' | \psi \rangle. \quad (6)$$

É possível representar de forma matricial um operador genérico V em uma base ortonormal $\{|a\rangle\}$ como

$$V = \sum_{a', a''} |a'\rangle \langle a' | V | a'' \rangle \langle a''|, \quad (7)$$

onde $\langle a' | V | a'' \rangle$ são os **elementos de matriz** de V na base escolhida. O somatório em a' e a'' percorre todos os estados da base e os elementos de matriz podem ser organizados explicitamente como

$$V \stackrel{\text{matricial}}{\equiv} \begin{pmatrix} \langle a' | V | a' \rangle & \langle a' | V | a'' \rangle \\ \langle a'' | V | a' \rangle & \langle a'' | V | a'' \rangle \end{pmatrix}, \quad (8)$$

que é a representação matricial do operador V .

Similarmente, para o caso de muitos spins $1/2$, temos o seguinte estado quântico genérico como sendo:

$$|\psi\rangle \in \mathcal{H}^{\otimes N}. \quad (9)$$

O espaço de Hilbert é representado pelo produto tensorial de N cópias do espaço local por

$$\mathcal{H}^{\otimes N} = \{\alpha_{\uparrow\uparrow\cdots\uparrow} |\uparrow\uparrow\cdots\uparrow\rangle + \alpha_{\uparrow\uparrow\cdots\downarrow} |\uparrow\uparrow\cdots\downarrow\rangle + \cdots + \alpha_{\downarrow\downarrow\cdots\downarrow} |\downarrow\downarrow\cdots\downarrow\rangle\}, \quad (10)$$

com base de vetores ortonormais representandos por

$$B(\mathcal{H}^{\otimes N}) = \{|\uparrow\uparrow\cdots\uparrow\rangle, |\uparrow\uparrow\cdots\downarrow\rangle, \dots, |\downarrow\downarrow\cdots\downarrow\rangle\}. \quad (11)$$

Nesse cenário, note a dimensão do espaço de Hilbert cresce exponencialmente com o número de partículas, ou seja, $\dim(B(\mathcal{H}^{\otimes N})) = 2^N$. Um estado quântico genérico $|\psi\rangle$ pode ser escrito como :

$$|\psi\rangle = \mathbb{1} |\psi\rangle = \sum_{x \in B(\mathcal{H})} |x\rangle \langle x|\psi\rangle = \sum_{x_1 \cdots x_N} \psi(x_1, \dots, x_N) |x_1 \cdots x_N\rangle = \sum_{\vec{x}} \psi(\vec{x}) |\vec{x}\rangle, \quad (12)$$

onde $\vec{x}$ representa um vetor que indexa os elementos do espaço de Hilbert. A função de onda $\psi(\vec{x})$ pode ser representada, por exemplo, como um vetor coluna

$$\psi(\vec{x}) = \begin{pmatrix} \psi(\uparrow \cdots \uparrow) \\ \psi(\uparrow \cdots \downarrow) \\ \vdots \\ \psi(\downarrow \cdots \downarrow) \end{pmatrix}. \quad (13)$$

Para representarmos, por exemplo, o operador Hamiltoniano H em forma matricial, teríamos :

$$H = \sum_{a''} \sum_{a'} \langle a''|H|a'\rangle, \quad (14)$$

ou seja, explicitamente como :

$$H^{\text{matricial}} \equiv \begin{pmatrix} \langle a'|H|a'\rangle & \langle a'|H|a''\rangle & \cdots \\ \langle a''|H|a'\rangle & \langle a''|H|a''\rangle & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}. \quad (15)$$

Desta forma, ao aumentar o número de partículas, a dimensão do espaço cresce exponencialmente como informado anteriormente. Para efeito ilustrativo, se tivéssemos um sistema com uma cadeia de spins com 50 partículas ($L = 50$) seria necessário ter um supercomputador com 18 PB de memória (1 petabyte (PB) = 10^{15} bytes), ou seja, para $L = 50$, temos $2^{50} = 1,125,899,906,842,624 \approx 1,126 \times 10^{15}$

Cada configuração é representada por uma amplitude complexa da função de onda. Considerando a precisão dupla complexa (16 bytes por número complexo, sendo 8 bytes para a parte real e 8 bytes para a imaginária), a memória necessária para armazenar o vetor de estado completo equivale a $2^{50} \times 16$ bytes = 16×2^{50} bytes = 2^{54} bytes. Como 1 PiB = 2^{50} bytes, obtemos 2^{54} bytes = 16×2^{50} bytes = 16 PiB. Convertendo para unidades decimais do SI (1 PiB $\approx 1,1259$ PB), 16 PiB $\approx 16 \times 1,1259$ PB ≈ 18 PB.

Desta forma, este crescimento exponencial é um dos principais desafios computacionais da mecânica quântica de muitos corpos. Desta forma, precisamos comprimir o espaço de Hilbert ou encontrar estratégias para descrever os estados do sistema sem explorar todo o espaço.

Dada a complexidade intrínseca dos sistemas quânticos de muitos corpos, diversas abordagens têm sido propostas com o intuito de contornar os desafios computacionais associados à sua descrição. Tais dificuldades estão relacionadas, principalmente, às demandas exponenciais de armazenamento e ao elevado número de operações requeridas

para a obtenção de quantidades físicas relevantes. Um exemplo notável é o cálculo do valor esperado de observáveis, como a energia do estado fundamental, cujo custo computacional cresce exponencialmente com o número de partículas do sistema, tornando inviável a aplicação de métodos exatos para sistemas com muitos graus de liberdade.

Na literatura, diversas abordagens foram desenvolvidas com o objetivo de mitigar a complexidade computacional associada à descrição de sistemas quânticos de muitos corpos [4, 14, 35, 49, 53, 55, 57]. No cerne destas abordagens, buscamos representar a função de onda total do sistema ψ como sendo uma ψ_θ de tal forma que a mesma consiga representar os estados quânticos com um número de coeficientes menor do que o máximo possível, ou seja, busca-se representar o estado em um subespaço reduzido.

Dentre essas abordagens, destacamos as seguintes:

1. Aproximação de Campo Médio (Mean-Field) - Estado Produto

Nesta abordagem, considera-se que o estado quântico total do sistema pode ser aproximado por um produto direto de estados locais, desconsiderando as correlações quânticas entre os diferentes sítios [14]. A função de onda assume, portanto, a forma fatorada

$$\begin{aligned}\psi_\theta(x_1, \dots, x_n) &= \psi_\theta(x_1) \psi_\theta(x_2) \cdots \psi_\theta(x_n) \\ &= \prod_{i=1}^n \psi_\theta(x_i).\end{aligned}\tag{16}$$

Essa aproximação reduz consideravelmente o custo computacional, porém não é capaz de capturar correlações quânticas não-locais, o que limita sua aplicabilidade em sistemas fortemente correlacionados.

2. Estados Produto de Matrizes (Matrix Product States - MPS)

Esta representação, amplamente utilizada no contexto de redes de tensores unidimensionais, descreve o estado quântico $|\psi_\theta\rangle$ como uma contração sequencial de tensores locais parametrizados, associados a cada sítio da rede, como representado a seguir

$$\psi_\theta(x_1, x_2, \dots, x_n) = \text{Tr} \left(A_{x_1}^{[1]} A_{x_2}^{[2]} \cdots A_{x_n}^{[n]} \right),\tag{17}$$

onde: $A_{x_i}^{[i]}$ é uma matriz (ou tensor) associada ao sítio i no estado local x_i ; θ denota o conjunto de parâmetros dos tensores, $\text{Tr}(\cdot)$ representa o traço sobre os graus de liberdade internos da rede.

Esta estrutura captura eficientemente o emaranhamento local em sistemas de muitos corpos sendo a base para métodos numéricos como o Density Matrix Renormalization Group Algorithm (DMRG). A técnica Matrix Product States (MPS) é especialmente eficaz para sistemas unidimensionais com interações locais e baixa

entropia de emaranhamento, sendo capazes de representar eficientemente tais estados [58]. Uma principais limitações da MPS está relacionada a estados altamente emaranhados e sistemas de alta dimensão (acima de 1D)[44].

3. Método Variacional com Monte Carlo (Variational Monte Carlo)

O método Variational Monte Carlo (VMC) baseia-se na definição de uma função de onda variacional parametrizada, $\psi(\vec{x})$, cuja energia esperada é minimizada por meio de técnicas estocásticas. A amostragem sobre o espaço de configurações é realizada com algoritmos como o de Metropolis-Hastings [55].

A flexibilidade do VMC permite empregar diversas formas funcionais para a função de onda, incluindo àquelas baseadas em arquiteturas de redes neurais ou representações em termos de produto de Jastrow, apresentado por

$$\psi(x_1, \dots, x_n) = \exp \left(\sum_{i < j} f(x_i, x_j) \right), \quad (18)$$

onde $f\theta(x_i, x_j)$ é um potencial de correlação de pares parametrizados que descrevem as interações entre as partículas.

Em sistemas de muitos corpos com emaranhamento de longo alcance, VMC implementados em termos de produtos de Jastrow falham em capturar correlações, subestimando drasticamente a energia do estado fundamental[12].

4. Estados Quânticos Variacionais com Redes Neurais (Net Neural Quantum States - NNQS)

Nesta abordagem utiliza-se redes neurais artificiais para representar a função de onda variacional do sistema [4, 5, 6]. Arquiteturas como RBM, Redes convolucionais (CNNs) ou Transformadores (*Transformers*) são treinadas para aproximar $\psi_\theta(\vec{x})$. As Neural Network Quantum States (NNQS) têm demonstrado capacidade de descrever estados fortemente emaranhados e sistemas de maior dimensionalidade. A função de onda é escrita em termos de uma rede neural artificial, de uma forma genérica

$$\psi(x_1, \dots, x_n) = \exp[NN(x_1 \cdots x_n)] = \exp \left[\sum_j a_j \sigma_j^z + \sum_i b_i h_i + \sum_{ij} W_{ij} h_i \sigma_j^z \right], \quad (19)$$

onde NN representa uma rede neural artificial; a, b, h são parâmetros da rede neural; $W_{i,j}$ representa os pesos de conexão e σ_j^z é o operador de spin- z do sítio j , escrito em termos da matriz de Pauli correspondente.

Neste trabalho, iremos nos aprofundar em NNQS. Esta abordagem combina técnicas de aprendizagem de máquina com métodos variacionais quânticos, permitindo: (i) a representação eficiente de estados quânticos em sistemas de muitos corpos, (ii) a otimização

de parâmetros usando Aprendizagem de Máquina por Reforço (AMP), e (iii) o estudo de propriedades físicas em regimes computacionalmente desafiadores como por exemplo, regimes com alta frustração.

1.2 Estrutura da Dissertação

Esta dissertação está organizada da seguinte forma: No [Capítulo 1](#), apresentamos uma introdução ao tema desta pesquisa, contextualizando a motivação e os desafios associados ao estudo de sistemas quânticos de muitos corpos. São discutidos brevemente os principais problemas relacionados à complexidade computacional desses sistemas, introduzindo o papel das redes neurais como ferramenta alternativa para representar estados quânticos. No [Capítulo 2](#), apresentamos a fundamentação teórica fundamental para o nosso trabalho. Apresentamos uma visão geral sobre cadeias de spins, enfatizando sua relevância como modelos para sistemas quânticos unidimensionais. Em seguida, introduzimos os conceitos centrais de redes neurais aplicadas à representação de estados quânticos. No [Capítulo 3](#), descrevemos a metodologia adotada, incluindo a formulação do problema físico considerado. Apresentamos MPS para a representação de Estados Quânticos e para a determinação de emaranhamento utilizando a medida da Entropia de Von Neumann. Por fim, detalhamos também como é a abordagem NNQS, como foi implementada e aplicada a diferentes regimes do sistema em estudo. No [Capítulo 4](#), apresentamos os resultados obtidos e respectivas comparações quantitativas entre os valores calculados pelas redes neurais e os obtidos via diagonalização exata. Além disso, analisamos a estrutura interna dos pesos das redes neurais de forma tomográfica relacionando a evolução dos pesos das redes neurais em relação ao regime físico do sistema. Por fim, no [Capítulo 5](#), apresentamos as conclusões deste trabalho, destacando os principais contribuições da abordagem proposta e possíveis extensões futuras que podem aprofundar ou expandir a pesquisa aqui desenvolvida.

Fundamentos Teóricos

Nesse capítulo iremos apresentar os conceitos fundamentais de nosso trabalho, a teoria das cadeias de spin e as redes neurais artificiais.

2.1 Teoria de Cadeias de Spin

Cadeias de spin são sistemas relativamente simples, porém altamente representativos, que capturam aspectos fundamentais do comportamento coletivo da matéria em escala quântica. Elas permitem descrever fases ordenadas, como os estados ferromagnéticos e antiferromagnéticos, além de fornecerem um cenário teórico ideal para o estudo de transições de fase quânticas. Esses modelos também desempenham um papel central na investigação do emaranhamento quântico e na análise da entropia de entrelaçamento, sendo, ao mesmo tempo, acessíveis a realizações experimentais por meio de sistemas como íons aprisionados, gases ultrafrios e materiais magnéticos de baixa dimensionalidade [2].

Um exemplo de um tipo de *cadeia de spins* pode ser representada como uma rede unidimensional composta por L sítios, igualmente espaçados por uma distância unitária $d = 1$.

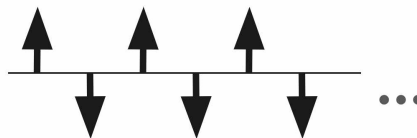


Figura 1 – Exemplo ilustrativo de uma cadeia de spin com regime antiferromagnético com L sítios com d de distância em cada sítios. Para uma cadeia de spin é possível definir a condição de contorno periódica, ou seja, o último sítio é vizinho do primeiro sítio.

Numa primeira aproximação, iremos considerar apenas a interação com o vizinho imediato (primeiro vizinho) e também vamos considerar condição de contorno periódica na cadeia, ou seja, o último sítio é vizinho do primeiro sítio.

A dinâmica e as interações entre os spins ao longo da cadeia são descritas, nestas condições, pelo *hamiltoniano de Heisenberg* [18], cuja forma geral é dada por

$$H = - \sum_{\langle i,j \rangle} J_{\langle i,j \rangle} \vec{S}_i \cdot \vec{S}_j, \quad (20)$$

onde S_i e S_j são operadores de spin, o símbolo $\langle i,j \rangle$ representa o somatório sobre vizinhos próximos, $J_{i,j}$ é a constante de acoplamento entre spins vizinhos (também conhecida como termo de troca), e o produto escalar $\vec{S}_i \cdot \vec{S}_j$ descreve a interação entre spins adjacentes.

No caso unidimensional, considerando apenas a interação entre o próximo vizinho, o hamiltoniano da equação (20) pode ser escrito da seguinte forma

$$H = - \sum_i J_{i,i+1} \vec{S}_i \cdot \vec{S}_{i+1}. \quad (21)$$

2.1.1 Dois sítios

Para fins de análise inicial, consideramos a resolução exata do hamiltoniano de Heisenberg em uma cadeia contendo dois sítios. Nosso objetivo é determinar o estado fundamental e investigar suas propriedades. Para isso, analisamos o sistema composto por dois spins identificados como 1 e 2. Definimos os subespaços de Hilbert correspondentes, juntamente com suas respectivas bases

$$B(\mathcal{H}_1) = |\uparrow\rangle_1, |\downarrow\rangle_1; B(\mathcal{H}_2) = |\uparrow\rangle_2, |\downarrow\rangle_2. \quad (22)$$

O espaço de Hilbert total é representado pelo produto tensorial entre o subespaço 1 e o subespaço 2., a seguir :

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2, \quad (23)$$

ou seja,

$$B(\mathcal{H}) = \{|\uparrow_1\rangle, |\downarrow_1\rangle\} \otimes \{|\uparrow_2\rangle, |\downarrow_2\rangle\}, \quad (24)$$

$$= \{|\uparrow_1\uparrow_2\rangle, |\uparrow_1\downarrow_2\rangle, |\downarrow_1\uparrow_2\rangle, |\downarrow_1\downarrow_2\rangle\} \quad (25)$$

$$= \{|\uparrow\uparrow\rangle, |\uparrow\downarrow\rangle, |\downarrow\uparrow\rangle, |\downarrow\downarrow\rangle\}. \quad (26)$$

Para resolver exatamente o Hamiltoniano de Heisenberg apresentado na equação (21), é necessário: (i) reescrevê-lo na base de estados vetoriais $B(\mathcal{H})$, utilizando os kets correspondentes às configurações de spin; (ii) expressar o hamiltoniano na forma matricial, com base nessa representação; e, por fim, (iii) diagonalizar essa matriz para obter seus autovalores e autofunções, que correspondem, respectivamente, às energias e aos autoestados do sistema.

Ao resolver analiticamente tal sistema, temos os seguintes autoestados com seus correspondentes autovalores. Os autovetores correspondentes a esses autovalores são :

1. Para $E_3 = +\frac{1}{4}(-J\hbar^2)$:

$$|\psi_2\rangle = \frac{1}{\sqrt{2}} (|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle).$$

2. Para $E_4 = -\frac{3}{4}(-J\hbar^2)$:

$$|\psi_1\rangle = \frac{1}{\sqrt{2}} (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle).$$

Com isso, os quatro autovalores do hamiltoniano completo são

$$E_1 = (-J\hbar^2)\left(\frac{1}{4}\right), \quad (\text{estado tripleto}) \quad (27)$$

$$E_2 = (-J\hbar^2)\left(\frac{1}{4}\right), \quad (\text{estado tripleto}) \quad (28)$$

$$E_3 = (-J\hbar^2)\left(\frac{1}{4}\right), \quad (\text{estado tripleto}) \quad (29)$$

$$E_4 = (+J\hbar^2)\left(\frac{3}{4}\right) \quad (\text{estado singlete}). \quad (30)$$

Os autoestados normalizados associados são

$$|S = 1, M = +1\rangle = |\uparrow\uparrow\rangle, \quad (31)$$

$$|S = 1, M = 0\rangle = \frac{1}{\sqrt{2}} (|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle), \quad (32)$$

$$|S = 1, M = -1\rangle = |\downarrow\downarrow\rangle, \quad (33)$$

$$|S = 0, M = 0\rangle = \frac{1}{\sqrt{2}} (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle). \quad (34)$$

O hamiltoniano tem dimensão 4 no entanto, como vimos, os autovalores de energia dependem apenas de S , que por sua vez é determinado pela simetria de troca dos spins (singlete ou tripleto). Portanto, apenas dois parâmetros são suficientes para definir o estado fundamental.

Analisando o termo de acoplamento J , podemos destacar as seguintes propriedades:

1. O sinal de J :

- a) Se $J < 0$ (antiferromagnético), o estado de menor energia é o singlete ($S = 0$) com valor correspondente a $\frac{3}{4}J\hbar^2$.
- b) Se $J > 0$ (ferromagnético), o estado de menor energia é o tripleto ($S = 1$) com valor correspondente a $-\frac{1}{4}J\hbar^2$.

2. A magnitude de J :

- a) Determina a separação de energia entre os estados, mas não afeta qual estado é o fundamental.

Desta forma, a simetria do hamiltoniano (invariância por rotação e troca de partículas) restringe fortemente as possibilidades. O sistema só pode estar no setor $S = 0$ ou $S = 1$, e o sinal de J decide qual setor tem menor energia. Os outros detalhes do Hamiltoniano (como anisotropias ou campos externos) não estão presentes neste caso, então não há mais graus de liberdade relevantes. A estrutura espectral do hamiltoniano de Heisenberg para dois spins $-\frac{1}{2}$ é completamente determinada pelo valor de J .

2.1.2 N-Sítios

Para um sistema de N partículas de spin $\frac{1}{2}$, o espaço de Hilbert cresce exponencialmente como 2^N , tornando a diagonalização completa do Hamiltoniano intratável para N grande. No entanto, em muitos casos, não é necessário acessar todos os 2^N estados para determinar o estado fundamental. Isso ocorre devido a simetrias e propriedades físicas do sistema. Assim como no caso de dois spins, onde apenas o setor de spin total S define o estado fundamental, para N spins, podemos explorar:

1. Simetrias do Hamiltoniano

- a) Invariância por rotação ($[H, \vec{S}_{\text{total}}] = 0$) permite a decomposição em setores de S_{total} .
- b) O estado fundamental geralmente reside no setor de S_{total} mínimo (para $J > 0$) ou máximo (para $J < 0$).

2. Métodos Numéricos Eficientes

- a) Algoritmos como DMRG (Density Matrix Renormalization Group) exploram a estrutura do emaranhamento do estado fundamental, evitando a diagonalização completa.
- b) Métodos de Monte Carlo quântico restringem a amostragem a setores relevantes.

Por exemplo, como apresentando em [14], para uma cadeia de spins antiferromagnética ($J > 0$) com N par:

1. A energia mínima pode ser encontrada sem diagonalizar toda a matriz $2^N \times 2^N$.

A determinação do estado fundamental em sistemas de muitos corpos não requer o acesso a todos os 2^N estados, devido a

1. Restrições impostas por simetrias.
2. Propriedades físicas dos emaranhamentos.

Isso justifica a eficácia de métodos modernos no estudo de sistemas quânticos macroscópicos, mesmo com a explosão combinatória do espaço de Hilbert, como por exemplo o uso de Redes Neurais Artificiais.

2.2 Redes Neurais Artificiais em Estados Quânticos

As RNAs têm desempenhado um papel central no desenvolvimento de tecnologias modernas, com impacto significativo em diversas áreas da ciência e engenharia [9, 15, 24, 31].

Devido à sua notável capacidade de extrair e representar estruturas complexas a partir de grandes volumes de dados, essas redes se tornaram ferramentas versáteis e eficazes para a resolução de uma ampla variedade de problemas, tanto teóricos quanto aplicados como por exemplo: i) reconhecimento de imagens e visão computacional (diagnósticos médicos por imagem e sistemas de vigilância); ii) processamento de linguagem natural (tradução automática, assistentes virtuais e sistemas de recomendação); iii) reconhecimento de voz e síntese de fala; iv) sistemas autônomos (veículos autônomos e robótica inteligente); v) previsão de séries temporais e dados financeiros; vi) modelagem e simulação de sistemas físicos (aplicações recentes em mecânica estatística, mecânica quântica e astronomia).

Esses avanços foram impulsionados principalmente por Redes Neurais Artificiais Profundas (RNAP) que permitem representar funções altamente não-lineares e também demonstram uma grande capacidade de resolver diversos problemas de alta dimensionalidade com extraordinária performance [15, 17, 24, 31].

As RNAs são inspiradas em sistemas naturais, em particular no sistema nervoso central, com base nos trabalhos seminais de McCulloch [29], Pitts [30], Rosenblatt [39], entre outros.

Em [20], Giuseppe Carleo destaca que as RNA possuem antecedentes estruturais vinculados à formulação do matemático David Hilbert para o problema 13 de sua célebre lista apresentada no início do século XX [19], bem como à solução proposta na forma do teorema de Kolmogorov-Arnold [22], o qual estabelece que toda função contínua de várias variáveis reais pode ser representada como uma composição de funções contínuas de uma variável e da operação de adição.

Carleo descreve as RNAs como funções em espaços de alta dimensão construídas por meio de uma composição hierárquica de funções elementares, em analogia à representação proposta por Kolmogorov-Arnold, nas quais cada função depende exclusivamente de parâmetros internos (neurônios) e de suas interconexões (sinapses). Essa estrutura funcional pode ser expressa de forma geral como

$$F(x; w; b) = A(B(C(x))) \quad (35)$$

onde A , B e C representam diferentes camadas da rede, com suas respectivas transformações; w denota as matrizes de pesos sinápticos; e b , os vetores de ajustes (bias) associados a cada camada.

De forma ilustrativa, vamos considerar A , a camada de entrada da rede neural; B , uma camada interna, e C a camada de saída. Inicialmente vamos considerar a definição da função ϕ como uma função sigmoide logística

$$\phi(x) = \frac{1}{1 + \exp(-x)}, \quad (36)$$

representada abaixo

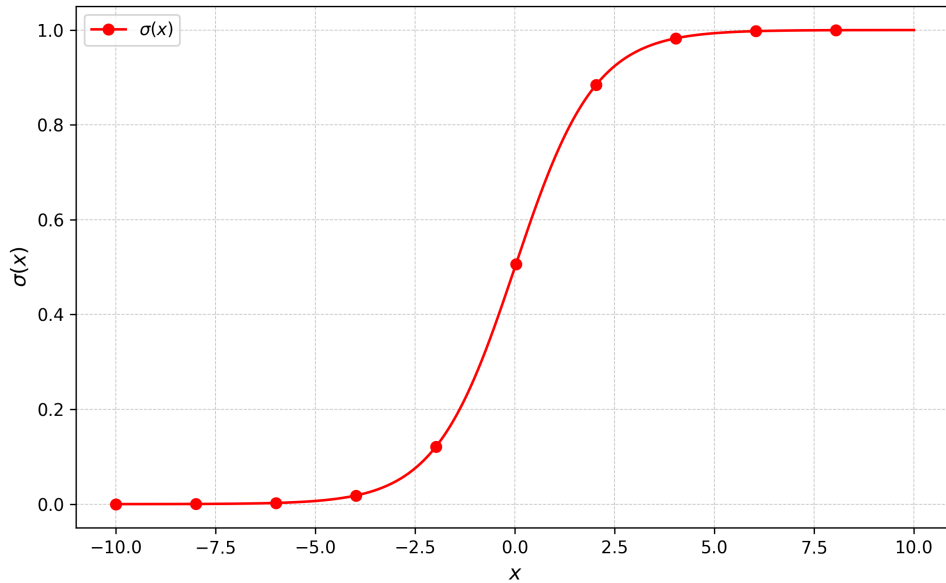


Figura 2 – Função sigmoide logística $\sigma(x) = \frac{1}{1+e^{-x}}$, frequentemente utilizada como função de ativação em redes neurais. Essa função mapeia valores reais em um intervalo entre 0 e 1, sendo útil para introduzir não-linearidades no modelo e para representar probabilidades associadas à ativação de neurônios. Sua forma suave e derivável facilita o processo de retropropagação do erro durante o treinamento.

Uma FFNN pode ser implementada, por exemplo, como uma combinação linear de funções a seguir

$$A : y_j = \phi \left(\sum w_{ij} x_i + b_j \right), \quad (37)$$

$$B : z_k = \phi \left(\sum w'_{jk} y_j + b'_k \right), \quad (38)$$

$$C : o = \phi \left(\sum w''_{k} z_k + b'' \right), \quad (39)$$

onde w_{ij}, w'_{jk}, w''_k representam matrizes de pesos sinápticos, b_j, b'_k, b'' são vetores de ajuste (bias), x é um vetor correspondente aos neurônios de entrada da primeira camada, representada abaixo:

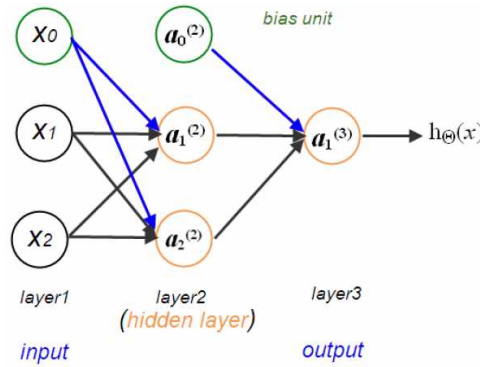


Figura 3 – Estrutura básica de uma rede neural artificial com 3 camadas sendo 1 camada de entrada com 3 neurônios, 1 camada interna com 3 neurônios e uma camada de saída com 1 neurônio. Nesta arquitetura, todos os neurônios de entradas se conectam com os neurônios da camada interna.

Os parâmetros da rede neural ($w_{ij}, w'_{jk}, w''_k, b_j, b'_k, b''$) são inicializados aleatoriamente sendo comum uma distribuição uniforme $\mathcal{U}$ para as matrizes de pesos ($a \approx 0.1$) e uma distribuição normal $\mathcal{N}$ para os vetores correspondentes aos ajustes (bias) ($\sigma \approx 0.01$).

2.2.1 Método Supervisionado

Quando temos dados previamente rotulados, ou seja, para cada conjunto de atributos existe uma classificação conhecida, estamos diante de um problema adequado para o *aprendizado supervisionado* [15, 31]. Por exemplo, considere um sistema físico em que, para diferentes conjuntos de características (como pressão, temperatura, volume, etc.), sabe-se a qual fase da matéria o sistema pertence: sólida, líquida ou gasosa. Nesse caso, trata-se de um problema de classificação supervisionada, pois a saída desejada é uma categoria discreta.

De forma semelhante, se tivermos um conjunto de parâmetros de entrada associados ao valor de uma função (como em problemas físicos, econômicos ou de engenharia), o objetivo pode ser estimar esse valor com base nos dados. Isso caracteriza um problema de regressão supervisionada, uma vez que a saída desejada é contínua. Em ambos os casos,

como conhecemos previamente os pares de entrada e saída desejada, o método apropriado para treinar o modelo é o aprendizado supervisionado.

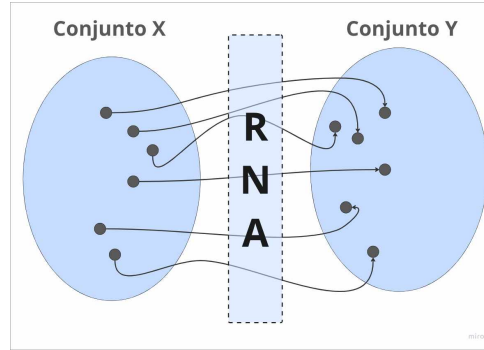


Figura 4 – O Aprendizado Supervisionado corresponde ao tipo de aprendizagem de máquina onde cada elemento do conjunto X possui uma correspondência unívoca (única e exclusiva) com um elemento do conjunto Y. Nesse caso, estabelece-se uma bijeção entre os conjuntos X e Y. A rede neural artificial ajusta seus pesos internos para aprender e reproduzir esse mapeamento.

Neste contexto podemos definir, com a arquitetura da máquina especificada por um conjunto de parâmetros $\mathbf{p} = (p_1, p_2, \dots, p_{N_p})$ por exemplo, os pesos e os ajustes (denotado como *bias*). A tarefa é encontrar uma estratégia eficaz para determinar numericamente os pesos e a estrutura da rede.

Nesse contexto, assume-se a existência de um grande conjunto de dados previamente disponíveis $x_1, x_2, \dots, x_{N_s}$, para os quais se conhece, de antemão, a saída desejada da máquina. Ou seja, também conhecemos os rótulos associados

$$y_i = F_o(x_i), \quad (40)$$

F_o denotada como função-alvo (função-ótima ou função-objetivo). O objetivo do aprendizado supervisionado é encontrar os melhores pesos da rede neural de forma que

$$F(x, \mathbf{p}) \approx F_o(x), \quad (41)$$

onde a função-alvo é conhecida apenas nos pontos do conjunto de dados.

Para determinar a qualidade do método de aprendizagem de máquina na realização de sua tarefa, podemos definir uma *função de perda* (*loss function*)

$$L(\mathbf{p}) = \frac{1}{N_s} \sum_{i=1}^{N_s} |F(x_i, \mathbf{p}) - y_i|, \quad (42)$$

a qual atinge o valor mínimo de zero em caso de reconstrução perfeita.

Dessa forma, o problema de aprendizado supervisionado pode ser interpretado, conceitualmente, como um problema simples de inferência ou como um problema de *ajuste não-linear*.

2.2.2 Método Não-Supervisionado

Em muitas das aplicações mais interessantes, é comum que os rótulos não estejam disponíveis durante o treinamento [15, 31, 34]. Por exemplo, considere o caso da classificação das fases da matéria: pode acontecer que não saibamos, de antemão, quais fases esperar. Nesta situação aplica-se um método conhecido como aprendizado não supervisionado (unsupervised learning).

Nesse contexto, é possível alcançar diversos objetivos mesmo na ausência de dados rotulados. Nesta situação, consideramos os dados $x_1, x_2, x_3, \dots, x_n$ sejam disponibilizados a partir de uma distribuição de probabilidade desconhecida $\rho(x)$, que desejamos estimar. Existem diversas abordagens de agrupamento de dados (*clusterização*) que são amplamente utilizadas na literatura, como os algoritmos K-Means [26], Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Gaussian Mixture Models (GMM) entre outros [21].

Essas técnicas visam identificar estruturas ou padrões em conjuntos de dados, agrupando amostras com características semelhantes de forma não-supervisionada, como representado na figura abaixo.

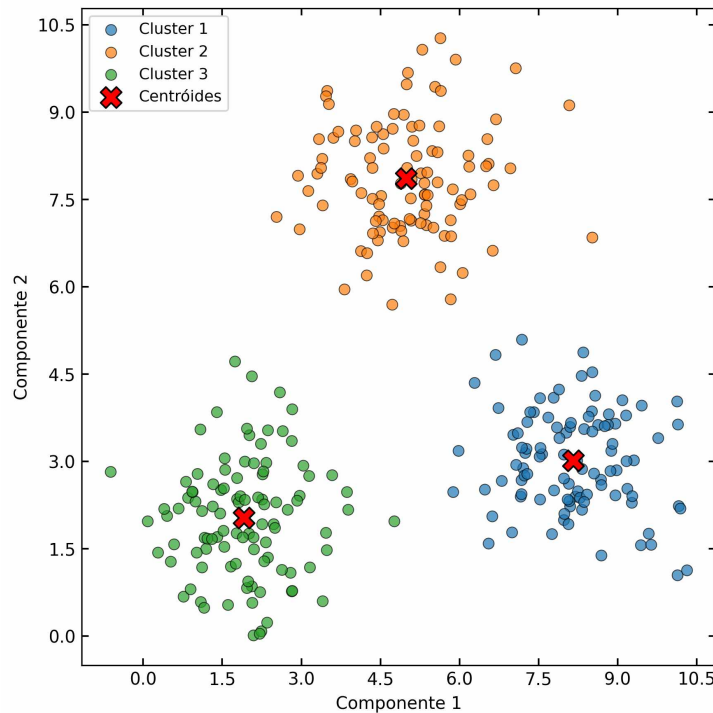


Figura 5 – O aprendizado não-supervisionado possui diversas aplicações, como a identificação de agrupamentos (clusters) por meio de métodos como o K-Means. Além disso, pode ser empregado para a visualização de dados em duas ou três dimensões utilizando técnicas de redução de dimensionalidade, como o PCA. Neste exemplo, apresentamos uma simulação contendo três agrupamentos e seus respectivos centroides. A clusterização é realizada com base em métricas de distância geométrica, e os dados são representados nas duas componentes principais.

2.2.3 Conexão entre o Aprendizado de Máquina e a Mecânica Quântica

No método de aprendizado de máquina precisamos de dois aspectos fundamentais:

1. **A máquina:** normalmente uma rede neural artificial, que é uma função altamente dimensional e não linear $F(x; p_1, \dots, p_{N_p})$, parametrizada por um conjunto de parâmetros $p_1, \dots, p_{N_p}$.
2. **O aprendizado:** os parâmetros $\mathbf{p}$ são ajustados com base em um processo de *otimização estocástica*, o qual visa minimizar alguma função de perda média $\langle L \rangle(\mathbf{p})$ sobre um conjunto de dados $x_1, x_2, \dots, x_{N_s}$. Por exemplo, em um cenário de aprendizado supervisionado com rótulos esperados y_i , uma possível função de perda seria

$$L(x_i) = |F(x_i; \mathbf{p}) - y_i|.$$

Por outro lado, o objetivo central da mecânica quântica é encontrar soluções para a equação de Schrödinger

$$H|\Psi_i\rangle = E_i|\Psi_i\rangle, \quad ; i = 0, 1, \dots; E_0 < E_1 < \dots \quad (43)$$

A pergunta que surge naturalmente é: *como podemos reinterpretar a mecânica quântica como um problema de aprendizado de máquina?*

Se considerarmos que a execução do nosso modelo de aprendizado de máquina possa identificar não apenas o estado fundamental (de menor energia), mas também os estados excitados do sistema, seria possível estabelecer uma correspondência formal entre o Aprendizado de Máquina (AM) e a Mecânica Quântica (MQ). No próximo capítulo iremos nos aprofundar nos detalhes do método NNQS que implementa esta tarefa.

Aplicação do NNQS ao modelo $J_1 - J_2$

3.1 O Problema

Iremos aplicar o método **NNQS** para encontrar o estado fundamental no hamiltoniano de Majumdar-Ghosh [28] (também conhecido por Modelo $J_1 J_2$ [7]), que descreve um sistema físico composto por uma cadeia unidimensional L spins interagindo entre si na ausência de campo externo. Explicitamente o modelo pode ser escrito como

$$H = J_1 \sum_{i=1}^{L-1} S_i \cdot S_{i+1} + J_2 \sum_{i=1}^{L-2} S_i \cdot S_{i+2}, \quad (44)$$

onde J_1 e J_2 correspondem, respectivamente, o acoplamento para os primeiros e os segundos vizinhos.

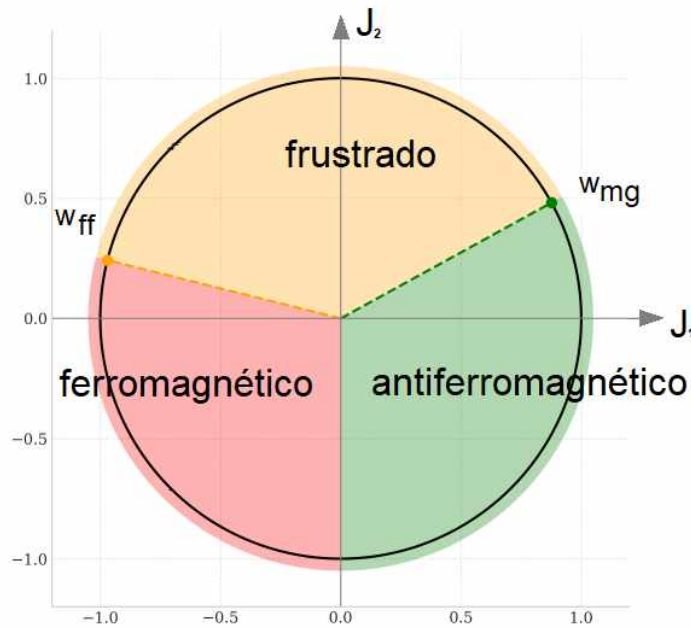


Figura 6 – Parametrização dos acoplamentos $J_1 - J_2$, em para $\theta \in [0, 2\pi)$. Em $T = 0$, conforme [3], o modelo $J_1 - J_2$ apresenta 3 fases: antiferromagnética, frustrada e ferromagnética.

A parametrização dos acoplamentos J_1 e J_2 é definida pelo vetor $\mathbf{J} = (J_1, J_2)$, com $J_1 = J_0 \cos \theta$ e $J_2 = J_0 \sin \theta$ para $\theta \in [0, 2\pi)$, conforme utilizado em [3, 56]. O trabalho [3] propõe um diagrama de fases em $T = 0$. Em [3], verifica-se que é possível identificar três regimes distintos: (i) regime ferromagnético, (ii) regime frustrado e (iii) regime antiferromagnético, como ilustrado na Figura 6, que mostra a transição entre esses regimes em 2 situações: no ângulo $w_{mg} = 0.2366$ radianos e no ângulo $w_{ff} = \tan^{-1}(\frac{-1}{4})$.

3.2 Matrix Product States

Matrix Product States (MPS) emergiu na década de 1990 a partir do trabalho seminal de Steven R. White [58] e posteriormente formalizada por Ulrich Schollwöck [43, 44]. Atualmente é uma ferramenta eficiente para representação de estados quânticos em sistemas de muitos corpos unidimensionais com baixo emaranhamento [37].

A essência do Matrix Product States (MPS) reside na identificação eficiente de estados quânticos através de uma estrutura de produto matricial que decompõe o vetor de estado de um sistema de muitos corpos em um produto contraído de tensores locais, permitindo a representação compacta de estados com emaranhamento limitado de acordo com o critério de área de Hastings [16], onde a dimensão de ligação (χ) controla a precisão da aproximação e está intrinsecamente relacionada ao emaranhamento do sistema. Considere um sistema de dois spins $-\frac{1}{2}$, como ilustração do uso de MPS, um estado quântico geral pode ser representado como

$$|\psi\rangle = \sum_{\sigma_1, \sigma_2} C_{\sigma_1 \sigma_2} |\sigma_1 \sigma_2\rangle, \quad (45)$$

onde $\sigma_i \in \{\uparrow, \downarrow\}$. Na forma MPS, os coeficientes $C_{\sigma_1 \sigma_2}$ decompõem-se como

$$C_{\sigma_1 \sigma_2} = \sum_{\alpha} A_{1\alpha}^{\sigma_1} B_{\alpha 1}^{\sigma_2}, \quad (46)$$

com: i) A^{σ_1} : matriz $1 \times \chi$ (tensor do sítio 1); ii) B^{σ_2} : matriz $\chi \times 1$ (tensor do sítio 2); iii) χ : dimensão de ligação ($\chi = 1$ para estado produto (estado puro ou estado não-emaranhado)).

Explicitamente para os estados temos i) estado singleto (anti-simétrico, $S = 0$)

$$|\psi^-\rangle = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle), \quad (47)$$

com as matrizes da seguinte forma :

$$A^{\uparrow} = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad A^{\downarrow} = \begin{pmatrix} 0 & 1 \end{pmatrix} \quad (48)$$

$$B^{\uparrow} = \begin{pmatrix} 0 \\ -\frac{1}{\sqrt{2}} \end{pmatrix}, \quad B^{\downarrow} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} \quad (49)$$

ii) estado tripleto simétrico ($S = 1$, $S_z = 0$ (emaranhado))

$$|\psi^+\rangle = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle), \quad (50)$$

$$A^\uparrow = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad A^\downarrow = \begin{pmatrix} 0 & 1 \end{pmatrix} \quad (51)$$

$$B^\uparrow = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \quad B^\downarrow = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \quad (52)$$

ii) estado tripleto simétrico ($S = 1, S_z = +1$ (produto))

$$|\psi\rangle = |\uparrow\uparrow\rangle, \quad (53)$$

$$A^\uparrow = \begin{pmatrix} 1 \end{pmatrix}, \quad A^\downarrow = \begin{pmatrix} 0 \end{pmatrix} \quad (54)$$

$$B^\uparrow = \begin{pmatrix} 1 \end{pmatrix}, \quad B^\downarrow = \begin{pmatrix} 0 \end{pmatrix} \quad (55)$$

ii) estado tripleto simétrico ($S = 1, S_z = -1$ (produto))

$$|\psi\rangle = |\downarrow\downarrow\rangle, \quad (56)$$

$$A^\uparrow = \begin{pmatrix} 0 \end{pmatrix}, \quad A^\downarrow = \begin{pmatrix} 1 \end{pmatrix} \quad (57)$$

$$B^\uparrow = \begin{pmatrix} 0 \end{pmatrix}, \quad B^\downarrow = \begin{pmatrix} 1 \end{pmatrix}. \quad (58)$$

A partir desta construção referente a [MPS](#), destacam-se as seguintes propriedades

1. dimensão de ligação :

a) $\chi = 2$ para estados emaranhados (singleto e tripleto $S_z = 0$)

b) $\chi = 1$ para estados produto (tripleto $S_z = \pm 1$)

2. normalização

$$\sum_{\alpha} (|B_{\alpha 1}^\uparrow|^2 + |B_{\alpha 1}^\downarrow|^2) = 1. \quad (59)$$

3. emaranhamento :

$$S = -\text{Tr}(\rho \log \rho) = \begin{cases} \ln 2 & \text{para } \chi = 2 \\ 0 & \text{para } \chi = 1. \end{cases} \quad (60)$$

O formalismo [MPS](#) sintetiza a relação entre emaranhamento e estrutura local via decomposição tensorial, onde a dimensão de ligação χ atua como parâmetro de controle para aproximações variacionais do estado fundamental. Sistemas 1D com gap são descritos exatamente com χ finito, enquanto sistemas críticos requerem $\chi \sim N^\alpha$, refletindo a violação logarítmica da lei de área . Neste trabalho utilizamos [MPS](#) com a técnica [DMRG](#) para identificar estados com alto emaranhamento principalmente no regime frustrado para analisar a performance das [RNAs](#).

3.3 Neural Network Quantum States

Neural Network Quantum States (NNQS) foi inicialmente proposta em 2017 por Carleo e Troyer [4] como uma nova abordagem para resolver problemas de física de muitos corpos, oferecendo uma alternativa promissora a métodos tradicionais como DMRG [58], Projected Entangled Pair State (PEPS) [57], Tree Tensor Networks (TTN) [49], entre outros. Essa proposta insere-se no contexto do uso de RNAs para lidar com problemas gerais da MQ, marcando o início de uma linha de pesquisa que vem sendo desenvolvida a partir de trabalhos seminais como os de [5, 6, 7, 52]. As primeiras aplicações em NNQS utilizaram RBMs como arquitetura de RNA [4]. RBMs são modelos rasos (*shallow models*) de RNA, consistindo em apenas duas camadas totalmente conectadas: uma camada visível, com N neurônios, e uma camada oculta, com M neurônios. A função de onda associada a uma RBM é dada por

$$\psi_\theta(s) = \sum_h e^{b_v^\dagger s + b_h^\dagger h + h^\dagger W s}, \quad (61)$$

onde s e h representam os estados dos neurônios nas camadas visível e oculta, respectivamente. Os parâmetros do modelo são dados por $\theta = \{b_v, b_h, W\}$, com b_v sendo o vetor de viés da camada visível, b_h o vetor de viés da camada oculta, e W a matriz de pesos que conecta as duas camadas. No contexto das RNAs, uma RBM pode ser interpretada como uma rede neural de propagação direta não linear de camada única, em que as unidades visíveis atuam como entradas e a função exponencial serve como função de ativação. Na RBM, por construção, o vetor viés (b_v) é adicionado nas camadas de entrada. Podemos reescrever a $\psi_\theta(s)$, assumindo que as unidades escondidas são binárias, $h_i \in \{-1, 1\}$, de tal forma

$$\psi_\theta(s) = e^{b_v^\dagger s} \prod_{i=1}^M 2 \cosh(b_{h,i} + W_i \cdot s), \quad (62)$$

onde $b_{h,i}$ e W_i representam, respectivamente, o i -ésimo viés oculto e a i -ésima linha da matriz de pesos. Cada neurônio visível v_i da RBM codifica o estado de um spin s_i em uma cadeia unidimensional, formando um espaço de configurações $\mathcal{V} = \{-1, +1\}^{\otimes N}$ como representado na figura 7.

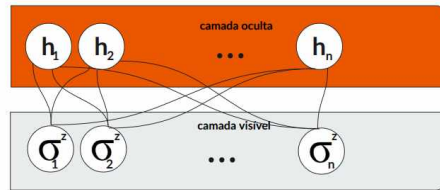


Figura 7 – Estrutura geral para uma Rede Neural do tipo Máquina de Boltzmann Restrita (RBM), onde temos a representação da função de onda de um sistema de N spins. Cada configuração visível $\sigma = (\sigma_1^z, \sigma_2^z, \dots, \sigma_n^z)$ é conectada a neurônios na camada oculta $h = (h_1, h_2, \dots, h_n)$ [9].

Determinando o estado fundamental

Nesta subseção, iremos apresentar como NNQS [5, 9, 20] pode ser utilizada para encontrar o estado fundamental de sistemas quânticos. De forma análoga a muitas tarefas em aprendizado de máquina, define-se uma função de perda L que depende dos parâmetros treináveis da rede neural. Nesta situação, essa função de perda corresponde à energia esperada do sistema, ou seja, o valor esperado do Hamiltoniano $\hat{H}$ em relação ao estado quântico parametrizado $|\Psi_\theta\rangle$

$$L(\theta) = E(\theta) = \langle \Psi_\theta | \hat{H} | \Psi_\theta \rangle. \quad (63)$$

A função de perda surge de forma natural neste contexto, sendo fundamentada no princípio variacional da mecânica quântica. Esse princípio estabelece que, para um dado Hamiltoniano $\hat{H}$, a energia esperada $E(\theta)$ associada a uma função de onda variacional $|\Psi_\theta\rangle$ fornece um limite superior para a energia do estado fundamental exato E_0 . Assim, ao minimizar $E(\theta)$, buscamos aproximar o estado de menor energia do sistema. A expressão da energia variacional é dada por

$$E(\theta) = \frac{\langle \Psi_\theta | \hat{H} | \Psi_\theta \rangle}{\langle \Psi_\theta | \Psi_\theta \rangle} \geq E_0. \quad (64)$$

A energia variacional é uma função de perda adequada para esta situação para determinação do estado fundamental pois, de acordo com o princípio variacional [8, 12], quanto menor for o valor esperado da energia, melhor será a aproximação do estado fundamental. Na prática, essa energia é estimada estocasticamente a partir de amostras geradas segundo a distribuição associada à função de onda $|\Psi_\theta\rangle$, o que pode levar a flutuações estatísticas. Em alguns casos, a média estimada pode ficar abaixo da energia do estado fundamental exato, mas isso não viola o princípio variacional, pois decorre da natureza estatística da amostragem.

O objetivo, portanto, é aumentar o número de amostras M de modo a reduzir sistematicamente essas flutuações. Seguindo a formulação geral de valores esperados de operadores [9, 20], a energia variacional pode ser aproximada por uma média sobre o estimador local E_{loc} , conforme

$$E(\theta) \approx \frac{1}{M} \sum_{i=1}^M E_{\text{loc}}(s^{(i)}), \quad (65)$$

onde o estimador local é definido como

$$E_{\text{loc}}(s) = \sum_{s'} \frac{\langle s | \hat{H} | s' \rangle \langle s' | \Psi \rangle}{\langle s | \Psi \rangle}. \quad (66)$$

Desta forma temos como meta minimizar essa função de perda por meio de algoritmos de otimização baseados em gradientes. Os gradientes da energia também podem ser escritos em termos de valores esperados

$$\frac{\partial E(\theta)}{\partial \theta_p} = 2 \operatorname{Re} \left[\langle E_{\text{loc}}(s) O_p^*(s) \rangle - \langle E_{\text{loc}}(s) \rangle \langle O_p^*(s) \rangle \right], \quad (67)$$

onde assumimos que os parâmetros são reais e que θ_p é o p -ésimo parâmetro da Rede Neural Quântica (NNQS).

O operador diagonal $\hat{O}_p$ é definido como

$$O_p(s) = \frac{\partial}{\partial \theta_p} \log \langle s | \Psi_\theta \rangle = \langle s | \hat{O}_p | s \rangle.$$

Notavelmente, quando a função de onda está próxima do estado fundamental exato, as flutuações estatísticas na energia local são suprimidas, implicando que as flutuações estatísticas dos gradientes também são pequenas.

O algoritmo de aprendizado [9, 20] tem os seguintes passos : i) inicializamos os pesos $\theta^{(0)}$. Em seguida, ii) implementamos uma sequência de M configurações de acordo com a distribuição de Born utilizando uma cadeia de Markov

$$P(s; \theta^{(s)}) \sim s^{(1)}, \dots, s^{(M)}.$$

O próximo passo é iii) calcular a média da energia local $E(\theta)$, que nos fornece a estimativa do valor esperado do Hamiltoniano. Para a última etapa, iv) utilizar-se um otimizador baseado em gradiente para atualizar os parâmetros para a próxima etapa, ou seja,

$$\theta_p^{(s+1)} = \theta_p^{(s)} - \eta \frac{\partial E(\theta)}{\partial \theta_p},$$

no caso do gradiente descendente simples, onde η é a taxa de aprendizado.

O procedimento é repetido até que realize a convergência para um mínimo de valores de energia. Neste método utilizado não existe um conjunto de dados de treinamento.. A tarefa apresentada tem como objetivo determinar a função de onda ótima (desconhecida) por meio da amostragem da distribuição de Born associada e utilizando uma rede neural para modelar o próprio estado. Esses passos estão apresentados de forma simplificada no Algoritmo [9, 20] abaixo:

Algorithm 1 Busca do estado fundamental

- 1: Inicialize θ aleatoriamente
- 2: **for** $i = 1$ até n_{steps} **do**
- 3: Gere M amostras tipo cadeia de Markov)
- 4: Calcule o gradiente da energia $\frac{\partial E(\theta)}{\partial \theta_p}$
- 5: Atualize os parâmetros

$$\theta_j \leftarrow \theta_j - \eta \frac{\partial E(\theta)}{\partial \theta_j}$$

- 6: **end for**
 - 7: **return** Parâmetros otimizados θ
-

3.4 Aspectos Caixa-Preta da Rede Neural

Apesar do sucesso das Rede Neural Artificiais (RNAs) em diversas aplicações, uma das principais críticas a esses modelos está relacionada à sua natureza de caixa-preta (black box) [48], o que significa que seu funcionamento interno é, em grande parte, opaco e difícil de interpretar. Diferentemente de modelos mais simples, como regressões lineares ou árvores de decisão, as redes neurais possuem um grande número de parâmetros, tornando desafiadora a tarefa de compreender como uma determinada entrada leva a uma saída específica. Neste sentido tem-se intensificado a pesquisa em métodos de interpretabilidade e explicabilidade em aprendizado de máquina, embora esse campo ainda esteja em desenvolvimento [25, 40].

Neste trabalho, investigamos, para diferentes problemas envolvendo cadeias de spin, em que medida a RNA é capaz de incorporar as interações entre os spins e capturar aspectos fundamentais da função de onda quântica. Especificamente, buscamos entender se e como a rede armazena informações sobre as propriedades de emaranhamento e outras características relevantes do estado quântico. Nosso objetivo é analisar se essas propriedades emergem naturalmente durante o processo de otimização e aprendizado, ou se exigem arquiteturas ou estratégias específicas de modelagem para serem corretamente representadas.

Em nossa análise inicial, investigamos a evolução dos pesos associados a cada neurônio da camada interna da RNA ao longo do processo de minimização da energia. Buscamos compreender como os valores dos pesos se ajustam durante o treinamento, qual o padrão de variação observado e de que forma essas mudanças contribuem para a representação eficiente do estado quântico alvo. Essa análise nos permite inferir o papel funcional de diferentes neurônios na modelagem da função de onda, e seus impactos serão discutidos em maior detalhe na seção de resultados.

Resultados

4.1 Preparação do método para Modelo $J_1 - J_2$

Iniciamos a análise da aplicação do método [NNQS](#) ao modelo $J_1 - J_2$ considerando um sistema com $L = 8$. Investigamos diferentes regimes de acoplamento definidos pelo vetor $\mathbf{J} = (J_1, J_2)$, onde os parâmetros são dados por $J_1 = J_0 \cos \theta$ e $J_2 = J_0 \sin \theta$, com $\theta \in [0, 2\pi)$. Os valores de θ considerados foram $\theta = [0, \pi/6, \pi/4, \pi/3, 2\pi/3, 5\pi/6, 7\pi/6, 5\pi/4, 49\pi/36, 5\pi/3, 11\pi/6]$ em graus radianos¹.

A escolha do θ está relacionado com a investigação dos 3 tipos de regimes: (i) regime frustrado; (ii) regime ferromagnético e (iii) regime anti-ferromagnético.

Para estes valores de acoplamentos $J_1 - J_2$, utilizamos o valor de 500 interações ($IT = 500$) em relação ao processamento da rede neural, ou seja, cada modelo roda 500 ciclos para determinar o valor da energia para o estado fundamental. Nesta simulação o tipo de rede neural escolhido foi o tipo [FFNN](#) devido a sua melhor performance comparada ao tipo de rede neural [RBM](#)².

Cada simulação o valor da energia é comparado em relação ao cálculo realizado por 2 métodos: (i) diagonalização exata via algoritmo de Lanczos [\[23\]](#) e (ii) abordagem variacional de Jastrow [\[11\]](#) com objetivo de realizar a validação cruzada. No processo de simulação, foi encontrado flutuações em relação ao valor calculado da energia em seu estado fundamental.

Estas flutuações exigiram um tratamento estatístico nos dados para a escolha da simulação adequada. Segue a ilustração a seguir, com os seguintes parâmetros $\theta = \pi/4$; $IT = 500$; $L = 8$. Realizamos um total de 100 *jobs* ($NJ = 100$) afim de encontrar o *job* ótimo, ou seja, o *job* que tem a melhor acurácia em relação ao cálculo da energia para o estado fundamental e que será utilizado para a análise dos pesos sinápticos das redes neurais para a caracterização da tomografia da rede neural proposta.

¹ $\theta = [0^\circ, 30^\circ, 45^\circ, 60^\circ, 120^\circ, 150^\circ, 210^\circ, 225^\circ, 245^\circ, 300^\circ, 330^\circ]$ em graus.

² No apêndice, no final deste trabalho, apresentamos este comparativo no contexto do Modelo de Heisenberg.

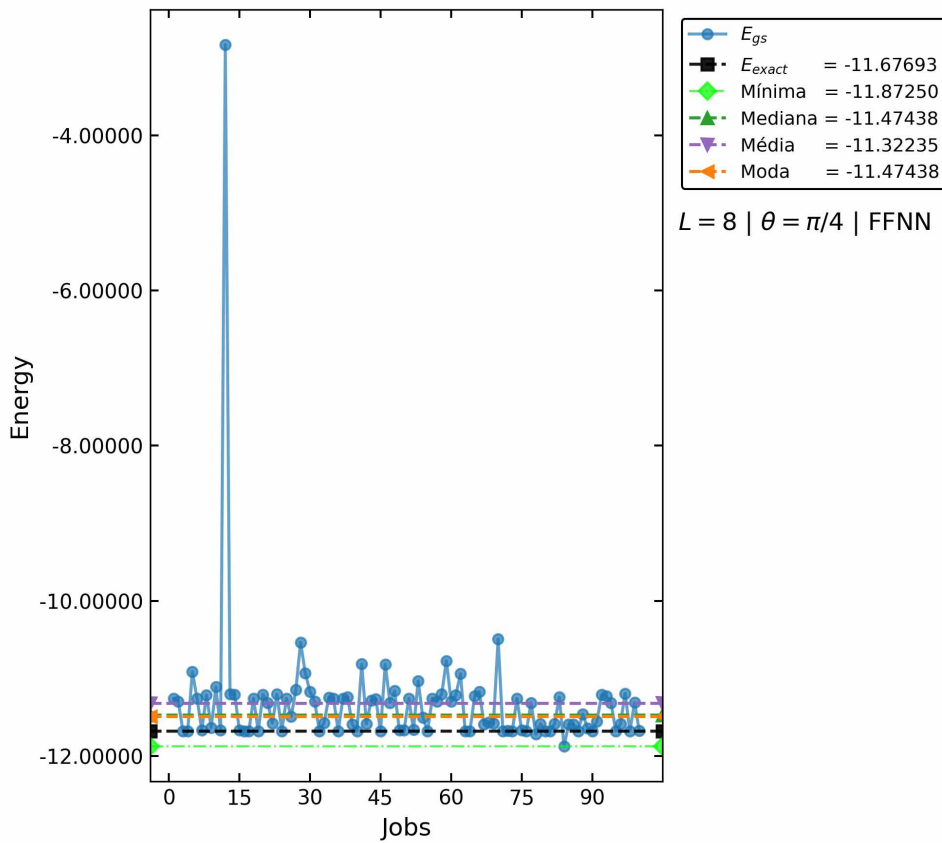


Figura 8 – Simulação com 100 *jobs* para uma cadeia de spins de tamanho $L = 8$, com 500 interações e acoplamento $\theta = \pi/4$, utilizando uma rede neural do tipo FFNN ($L = 8$, $\theta = \pi/4$, $IT = 500$). Nesta situação, temos por exemplo, alguns *jobs* apresentam um valor da energia muito distante do valor da energia exata.

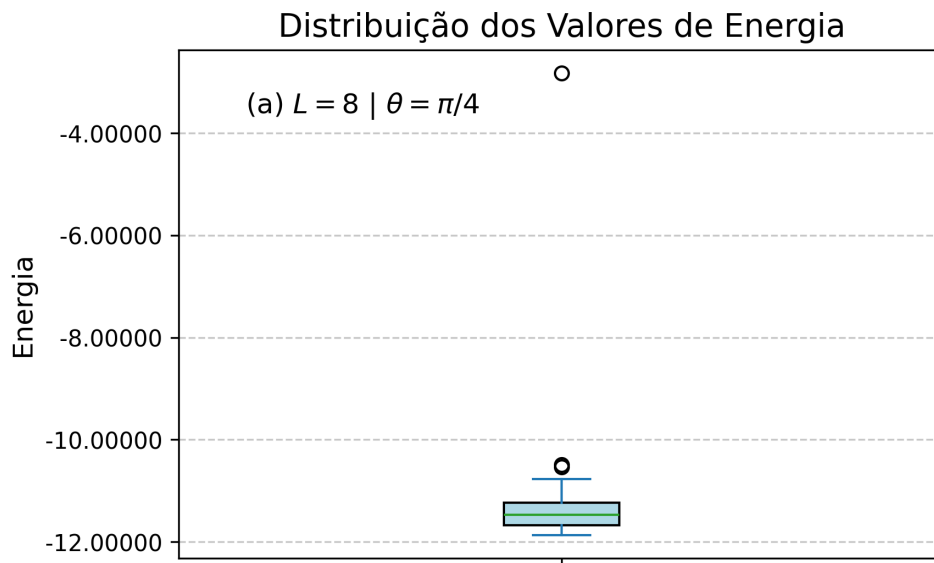


Figura 9 – Na visualização gráfica do tipo *boxplot* para a simulação com 100 *jobs* ($NJ = 100$) para uma cadeia de spin de tamanho igual a 8, com 500 interações, e valor de acoplamento igual a $\pi/4$ e tipo de rede neural FFNN ($L = 8$, $\theta = \pi/4$, $IT = 500$). Observa-se a presença de alguns valores fora do intervalo de valores do *boxplot*, caracterizando a ocorrência de *outliers* no conjunto dos resultados.

Implementamos um método de ajuste de dados para selecionar quais *jobs* são relevantes para o cálculo da energia. Utilizamos 2 técnicas para a remoção de *outliers*: (i) métodos dos quartis [10] e (ii) o método da distância de Mahalanobis [27]. A partir da aplicação destes 2 métodos, verificamos que encontramos um envelope de dados com simulações otimizadas.

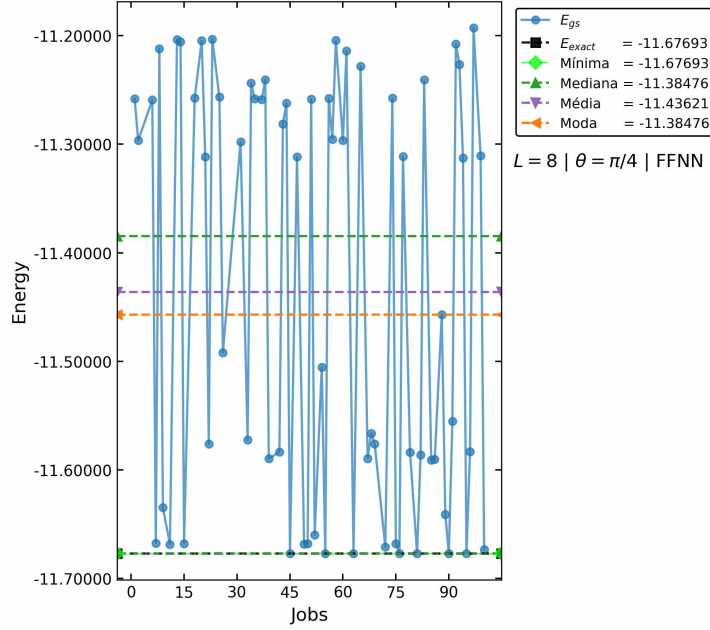


Figura 10 – Simulação com 100 *jobs* ($NJ = 100$) para uma cadeia de spins de tamanho $L = 8$, com 500 interações e acoplamento $\theta = \pi/4$, utilizando uma rede neural do tipo FFNN ($L = 8$, $\theta = \pi/4$, $IT = 500$). Nesta simulação, os *outliers* foram removidos, resultando em valores de energia bastante próximos ao valor da energia exata.

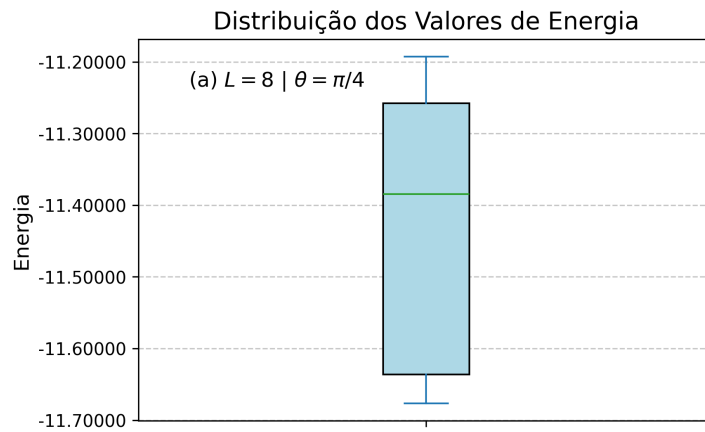


Figura 11 – Na visualização gráfica do tipo *box plot* referente à simulação de 100 *jobs* ($NJ = 100$) para uma cadeia de spins de tamanho $L = 8$, com 500 interações e acoplamento $\theta = \pi/4$, utilizando uma rede neural do tipo FFNN ($L = 8$, $\theta = \pi/4$, $IT = 500$) com os *outliers* removidos, observa-se que a distribuição dos valores concentra-se predominantemente entre o primeiro e o segundo quartil.

Uma vez determinado o envelope de dados dos *jobs*, escolhe-se o *job* que melhor representa a simulação para a realização da tomografia da rede. Para essa seleção, consideram-se três possibilidades: (i) o *job* cuja energia corresponde ao valor da moda; (ii) aquele cuja energia é igual à mediana; ou (iii) aquele que apresenta o valor mínimo. A escolha pelo valor mínimo baseia-se na comparação com a energia calculada de forma exata, ou seja, na opção que mais se aproxima desse valor. Para uma abordagem mais genérica para o tratamento deste problema, podemos determinar esta opção como sendo um hiperparâmetro de configuração da aplicação.

Para nossa abordagem em particular, temos o seguinte algoritmo da seleção do *job* ótimo:

Algorithm 2 seleção do *job* ótimo

```

1: Inicializa Lista de Jobs  $JList = []$ 
2: for  $i = 1$  até  $n_{\text{jobs}}$  do
3:   Calcule a Energia  $E_i$ ,
4:   Adição (Configurações  $C_i$ , Energia  $E_i$ ) para Lista de Jobs  $JList$ 
5: end for
6: Remove Jobs com Energia  $E_i$  outlier
7: return Job Energia Mínima : (Configurações  $C_i$   $\min(E_i)$ )
```

Com base no algoritmo apresentado, avaliamos a seguinte configuração: parâmetros $L = 8$, 500 iterações e acoplamento $\theta = \pi/4$, utilizando uma rede neural do tipo [FFNN](#). Após a execução de 100 *jobs*, selecionou-se aquele com o critério definido no algoritmo [2](#). Como seleção temos que o *job* executado na terceira chamada atendeu plenamente aos critérios definidos pelo algoritmo.

Para esse *job* específico, observou-se que a evolução do cálculo da energia apresenta convergência para o valor da energia E em torno do ciclo $IT = 180$, conforme ilustrado na Figura [12](#). Nesta simulação, verificou-se também que o método baseado na abordagem variacional de Jastrow não apresentou boa acurácia, resultando em um valor de energia diferente ao obtido pela diagonalização exata via algoritmo de Lanczos (na ordem de 10 percentuais) .

Para validar o método de ajuste de dados proposto anteriormente, apresentamos um *job* que fora descartada por representar um *outlier* representando na Figura [13](#). É fácil notar que o valor da energia encontra-se completamente fora do intervalo típico dos dados e que através da implementação da determinação do envelope de dados fora eliminado. A simulação quando se aproxima do ciclo 240 ($IT \approx 238$) sofre um desvio brusco no que se refere ao valor da Energia E ($E_{i=238} \approx -11.0$, $E_{i=250} \approx -3.0$).

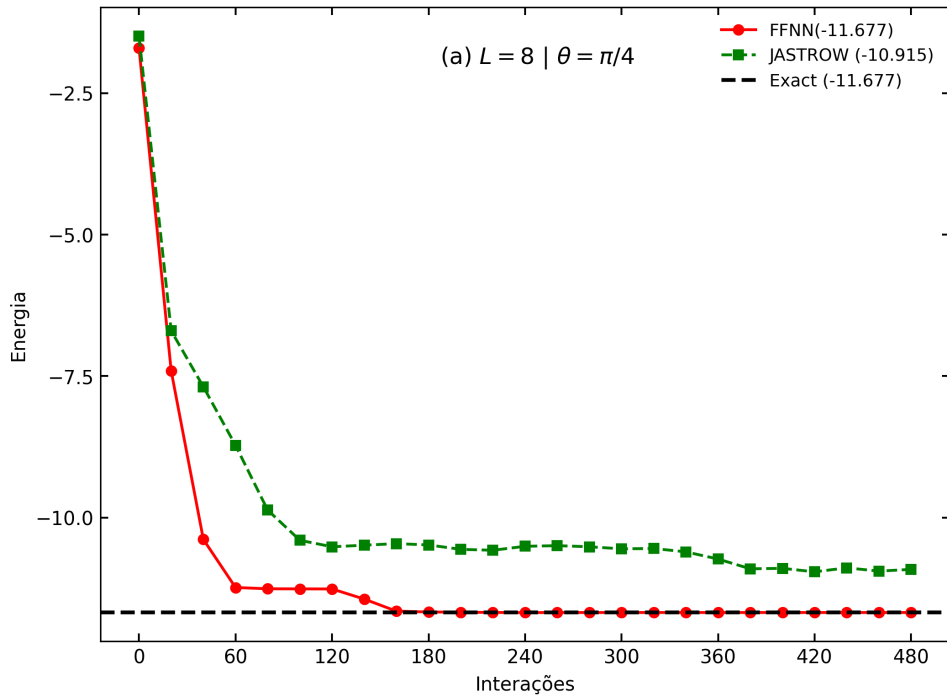


Figura 12 – Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações e acoplamento $\theta = \pi/4$, utilizando uma rede neural do tipo **FFNN** ($L = 8$, $\theta = \pi/4$, $IT = 500$). Observe-se que a partir do ciclo de interação próximo ao valor 180, o método converge para o valor da energia mínima. Em nossa simulação geral, esse *job* correspondeu à terceira execução dentre um total de 100 execuções.

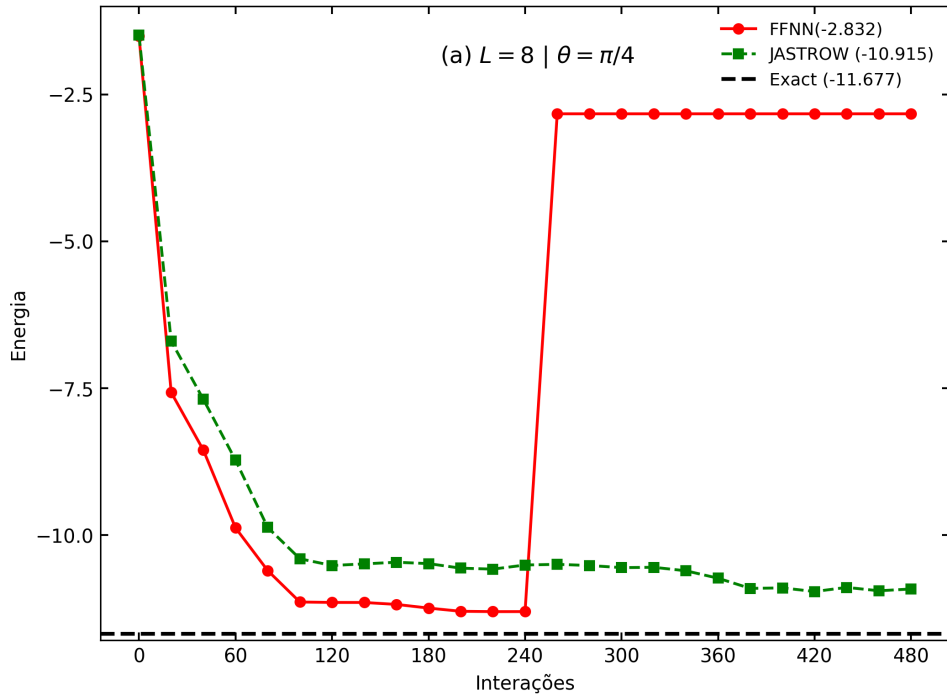


Figura 13 – Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações e acoplamento $\theta = \pi/4$, utilizando uma rede neural do tipo **FFNN** ($L = 8$, $\theta = \pi/4$, $IT = 500$) com um *job* que na análise de dados foi caracterizado com um *outlier* sendo a simulação descartada. Em nossa simulação geral, esse *job* correspondeu à décima primeira execução dentre um total de 100 execuções.

4.2 Avaliação para o Cálculo da Energia

Para avaliar o algoritmo de seleção do *job* ótimo e validar o cálculo da energia do estado fundamental implementado pelo modelo de RNA, foram considerados três valores de acoplamento definidos pelo parâmetro $\theta = [\pi/4, 7\pi/6, 5\pi/3]$, correspondentes, respectivamente, aos regimes frustrado, ferromagnético e antiferromagnético. Para cada caso considerado, aplicou-se o algoritmo de seleção do *job* ótimo. Como exemplo, os *jobs* selecionados corresponderam às seguintes execuções: $NJ = [3, 70, 72]$. Assim, o *job* da 3ª execução está associado a $\pi/4$, o da 70ª execução a $7\pi/6$ e o da 72ª execução a $5\pi/3$ representada na Figura 14.

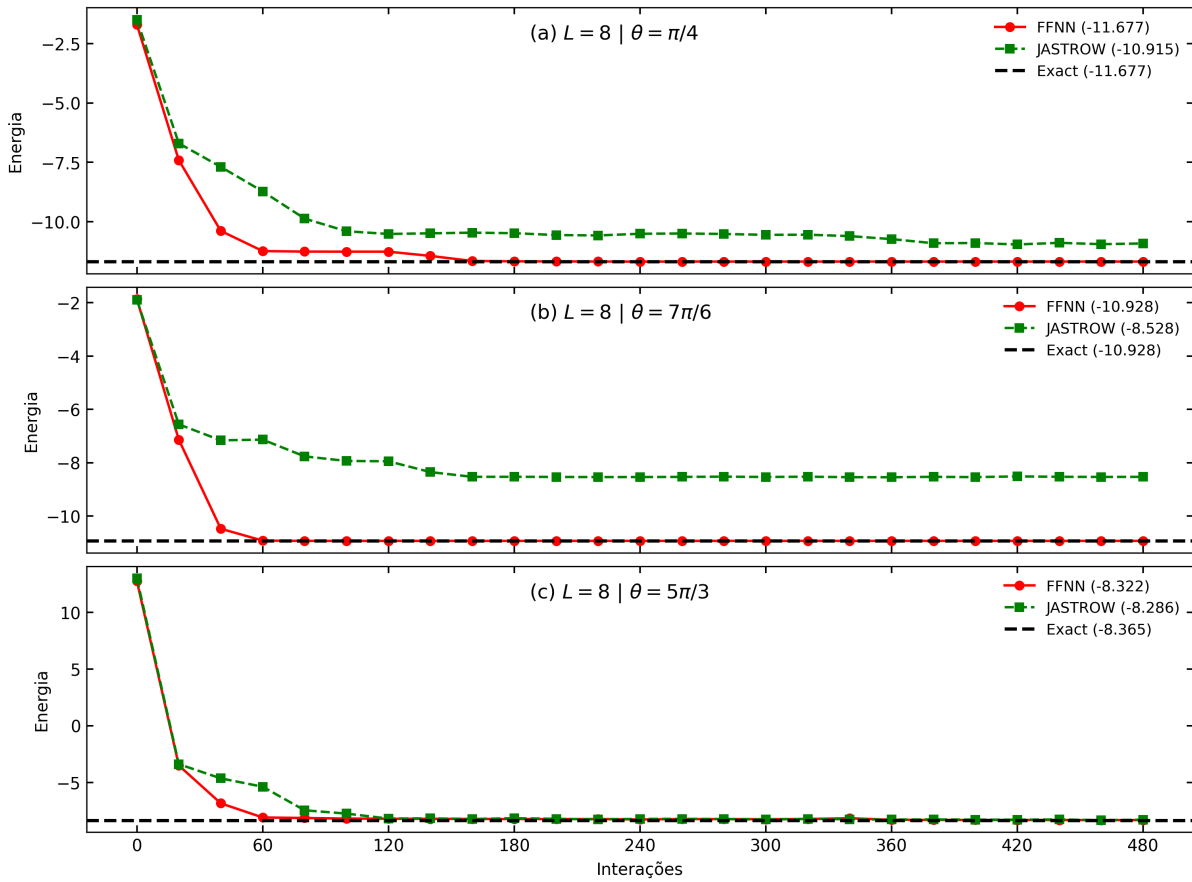


Figura 14 – Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 iterações para os seguintes acoplamentos $\theta = [\pi/4, 7\pi/6, 5\pi/3]$, utilizando uma rede neural do tipo FFNN ($L = 8$, $IT = 500$). As execuções selecionadas foram as seguintes : (i) para $\theta = \pi/4$, o 3º *job*; (ii) para $\theta = 7\pi/6$, o 70º *job* e (iii) para $\theta = 5\pi/3$, o 72º *job* de 100 execuções para cada simulação.

Nas simulações realizadas neste trabalho, observou-se, de modo geral, que a RNA converge mais rapidamente para o regime antiferromagnético em comparação ao regime ferromagnético, e nesta comparação o regime frustrado é aquele que converge mais lento dos três regimes. Esse comportamento identificado entre os regimes antiferromagnético e ferromagnético para a aplicação utilizando RNA também foi obtido a partir do modelo de Heisenberg, discutido no Apêndice, ao final deste trabalho, como material complementar.

Em relação ao regime frustrado, foram identificadas duas sub-regiões: a primeira situada no primeiro quadrante e a segunda no segundo. De forma geral, as simulações correspondentes ao primeiro quadrante - isto é, para $\theta = [\pi/6, \pi/4, \pi/3]$ - demandaram um número maior de iterações para o cálculo da energia do que aquelas associadas a $\theta = [2\pi/3, 5\pi/6]$ ilustrado na Figura 15. Tais comportamentos serão analisados posteriormente no contexto do que propomos como tomografia da RNA.

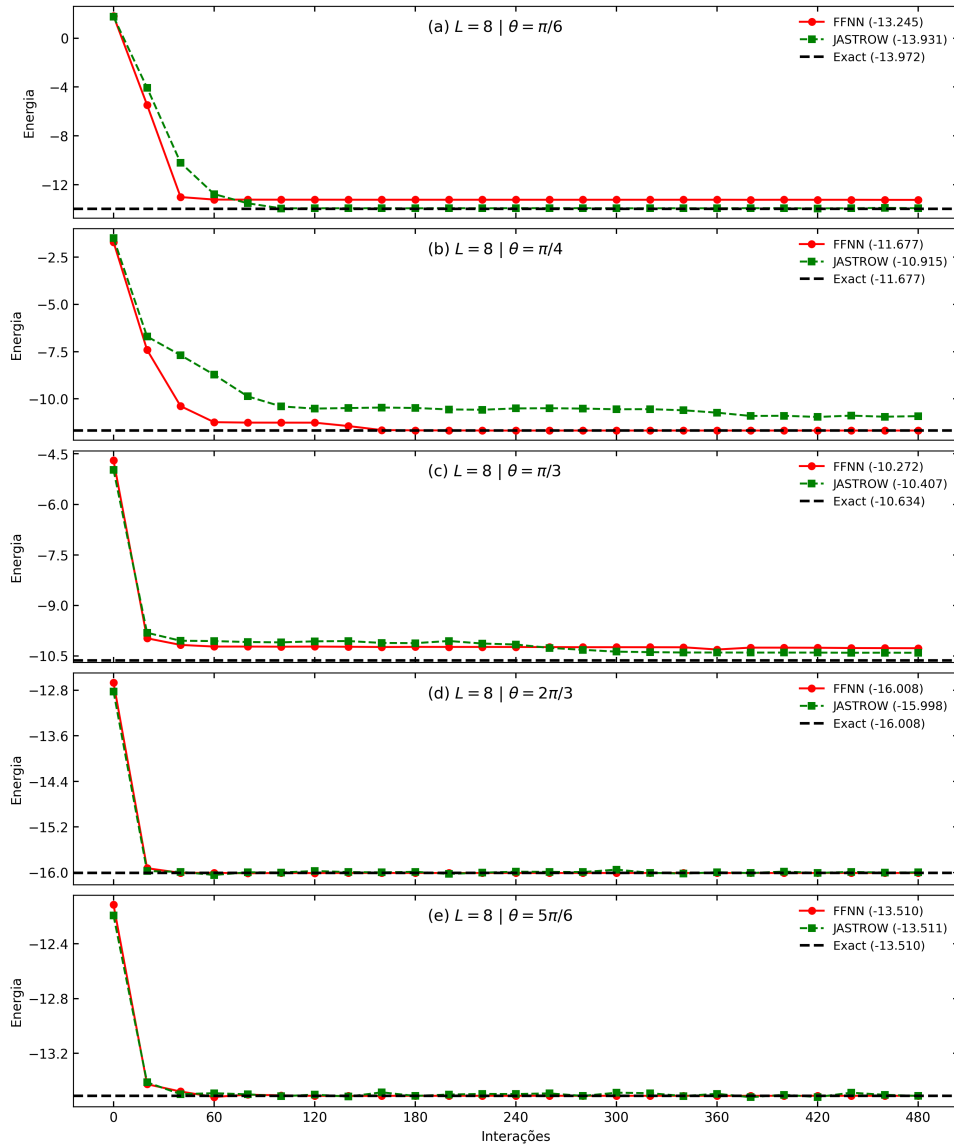


Figura 15 – Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 iterações para os seguintes acoplamentos $\theta = [\pi/6, \pi/4, \pi/3, 2\pi/3, 5\pi/6]$, utilizando uma rede neural do tipo FFNN ($L = 8$, $IT = 500$).

4.3 Tomografia da Rede Neural Artificial

No contexto da determinação da energia do estado fundamental através da RNA utilizando o método NNQS, nesta abordagem, analisamos para cada simulação a evolução dos pesos sinápticos da RNA com objetivo de relacionar os valores destes pesos com o carácter físico do problema da cadeia de spin.

Iremos a seguir apresentar para a configuração das simulações apresentada anteriormente para os parâmetro $\theta = [\pi/4, 7\pi/6, 5\pi/3]$, uma amostragem, representando um *sample*³ de neurônios do *kernel* para cada simulação.

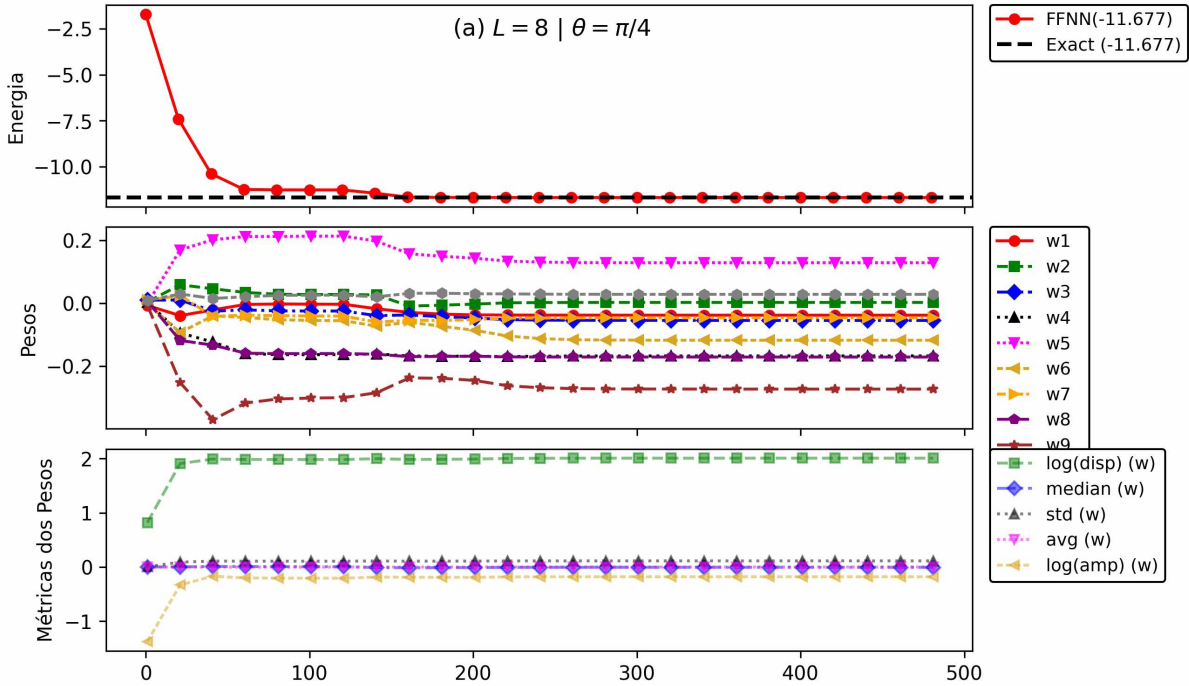


Figura 16 – Representação dos Pesos da Rede Neural para Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = \pi/4$, utilizando uma rede neural do tipo FFNN ($L = 8$, $IT = 500$). Temos algumas flutuações próximo à interação 50 ($IT = 50$) e à interação 160 ($IT = 160$) nos valores dos pesos dos neurônios que podem ser associadas a flutuações no valor da energia calculada.

Também foram calculadas diversas métricas estatísticas dos pesos de todos os neurônios, incluindo o logaritmo decimal da dispersão, a mediana, o desvio-padrão, a média e o logaritmo decimal da amplitude, identificadas na figura, respectivamente, como $\log(\text{disp})$, median , std , avg e $\log(\text{amp})$. Para a simulação representada na Figura 16, observam-se flutuações nos valores dos pesos dos neurônios em torno das iterações 50 ($IT = 50$) e 160 ($IT = 160$), as quais podem estar associadas a variações no valor da energia calculada.

³ Nesta simulação, temos um total de 128 neurônios e para efeito de avaliação sorteamos 9 neurônios.

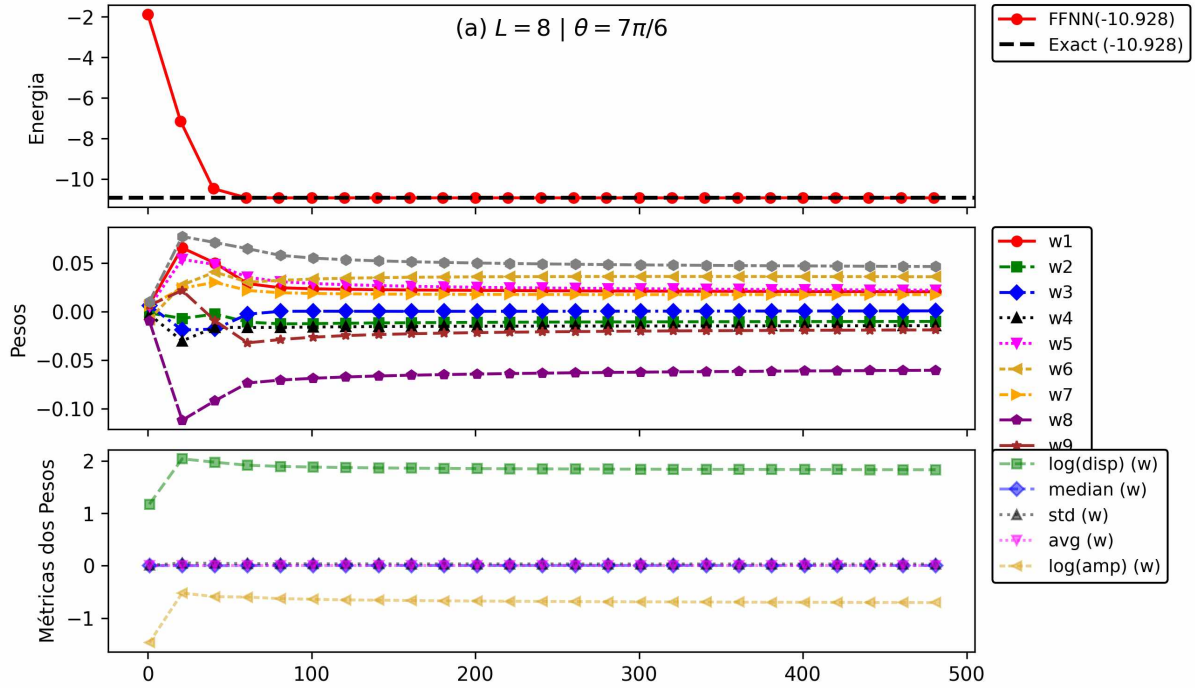


Figura 17 – Representação dos Pesos da Rede Neural para Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = 7\pi/6$, utilizando uma rede neural do tipo FFNN ($L = 8$, $IT = 500$).

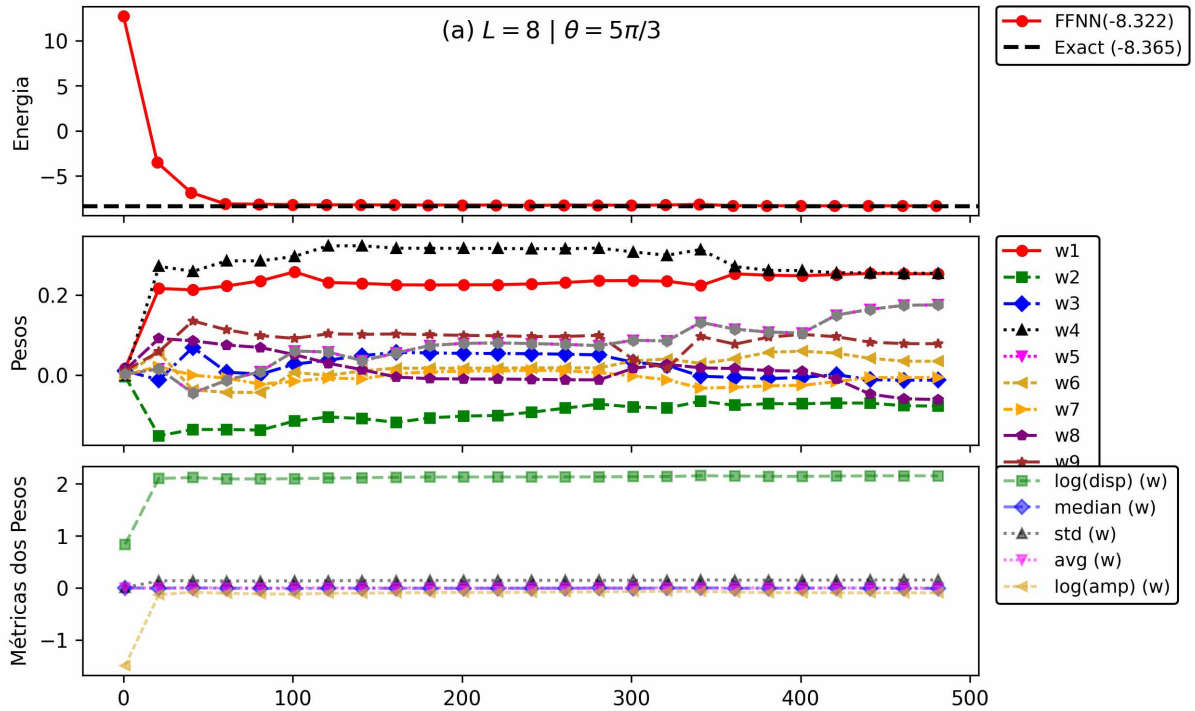


Figura 18 – Representação dos Pesos da Rede Neural para Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = 5\pi/3$, utilizando uma rede neural do tipo FFNN ($L = 8$, $IT = 500$).

Para a simulação representada na Figura 17, observa-se que, após a iteração 100 ($IT = 100$), os pesos dos neurônios selecionados na amostra de análise, bem como as respectivas métricas estatísticas, convergem para um valor constante, caracterizando um platô conforme o cálculo da energia. Na simulação representada na Figura 18, a energia é calculada até a iteração 100 ($IT = 100$) alcançando um valor platô; entretanto, no conjunto amostral dos pesos dos neurônios observam-se algumas flutuações que, contudo, não são refletidas nas métricas estatísticas, as quais permanecem praticamente constantes.

As análises tanto do conjunto amostral dos pesos dos neurônios quanto das métricas estatísticas nas simulações anteriores, nas quais a variação da energia é pequena, não fornecem evidências conclusivas sobre a relação entre os pesos dos neurônios e o valor da energia. No entanto, na Figura 19, observa-se que uma mudança brusca no valor da energia reflete-se igualmente em alterações significativas nos pesos da rede neural e nas métricas estatísticas correspondentes, como o logaritmo da dispersão, a mediana, a média, o desvio-padrão e o logaritmo da amplitude.

Com o objetivo de ampliar a análise dos pesos neuronais, serão incluídas duas abordagens adicionais: uma baseada na PCA e outra na correlação de Pearson.

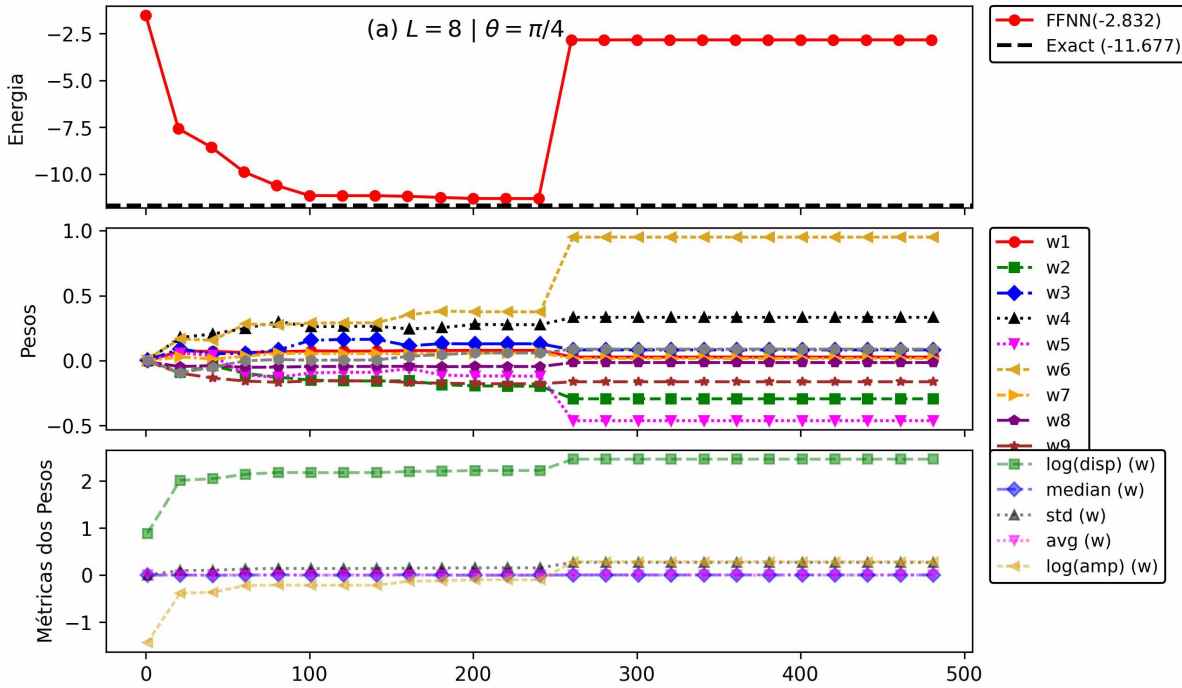


Figura 19 – Representação dos Pesos da Rede Neural para Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = \pi/4$, utilizando uma rede neural do tipo FFNN no job para a execução 11, situação que foi detectada como outlier ($L = 8$, $IT = 500$, $NJ = 11$). Verifica-se que a mudança brusca no valor da energia reflete também na mudança dos valores dos pesos da rede neural e nas também nas métricas estatísticas como log da dispersão, mediana, média, desvio-padrão e log da amplitude.

Análise baseada no PCA

Aplicou-se o método PCA, descrito em detalhe no apêndice ao final deste trabalho, para analisar a trajetória dos pontos iniciais e finais nas diversas realizações. Como ilustração, temos na Figura 20 a representação da trajetória do PCA aplicado a representação dos Pesos da Rede Neural para a simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = 11\pi/6$ para o 11º *job*. Verifica-se uma convergência próximo a interação 100 ($IT = 100$), o que ocorre também no cálculo da energia.

Na Figura 21, referente à simulação de uma cadeia de spins de tamanho $L = 8$ com 500 interações, observam-se os seguintes casos: (i) para $\theta = \pi/4$, 3º *job*; (ii) para $\theta = 7\pi/6$, 70º *job*; e (iii) para $\theta = 5\pi/3$, 72º *job*. O comportamento identificado é o seguinte: (a) a trajetória correspondente a $\theta = 7\pi/6$ percorre a maior distância; (b) a trajetória associada a $\theta = 5\pi/3$ é a que mais se aproxima de uma linha reta; e (c) à medida que a trajetória se aproxima do ponto final, a densidade de pontos aumenta, indicando um processo de convergência.

Na Figura 22, apresenta-se uma simulação em que o parâmetro $\theta = \pi/4$ é mantido fixo, enquanto o *job* é variado. Foram selecionados o 3º e o 4º *jobs*, considerados similares, e o 11º *job*, que foi descartado da simulação por corresponder a um valor *outlier*. Observa-se que as trajetórias dos *jobs* 3 e 4 são bastante semelhantes, ao passo que a trajetória do *job* 11 difere significativamente, apresentando uma quase descontinuidade após à interação 200 ($IT = 200$).

(a) $L = 8 \mid \theta = 11\pi/6 \mid \text{JOB}=16$

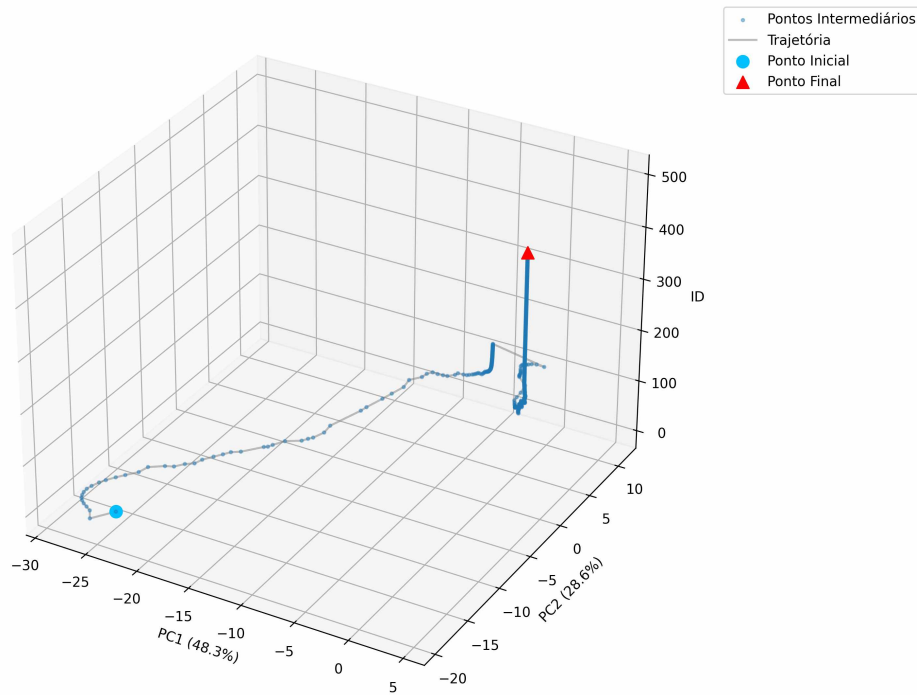


Figura 20 – Trajetória do PCA aplicado a representação dos Pesos da Rede Neural para Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = \pi/4$, o 11º *job*.

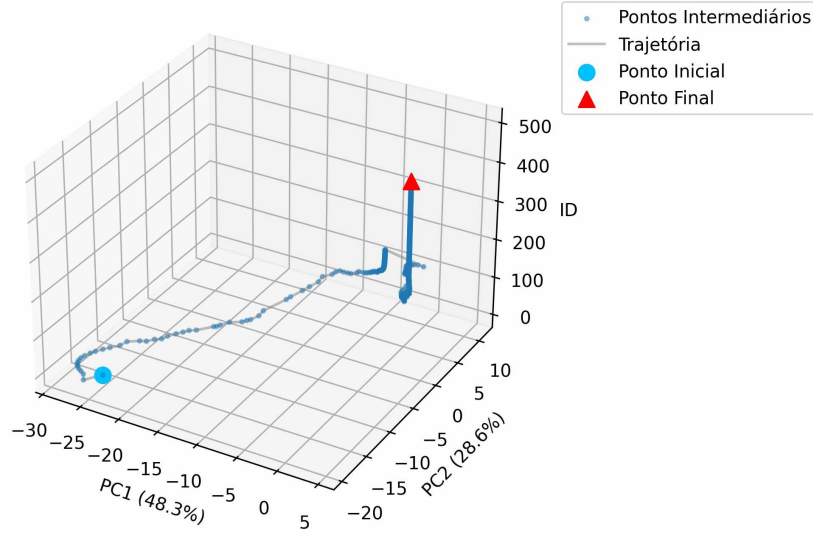
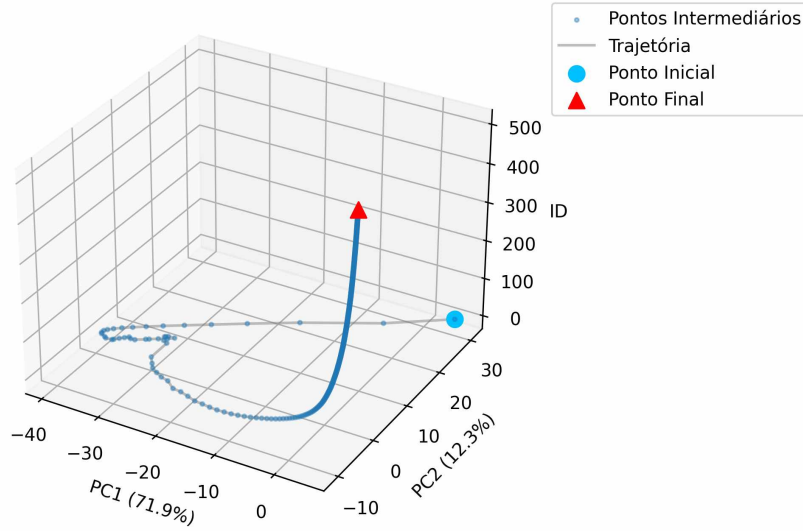
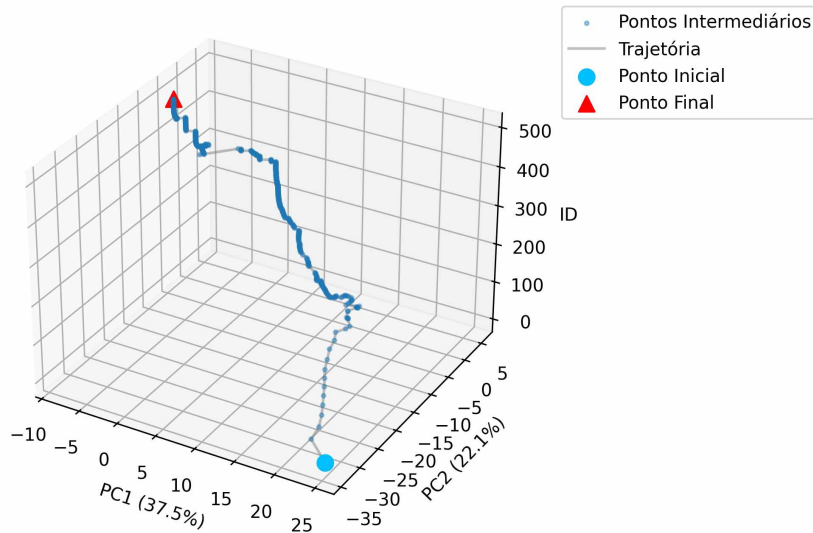
(a) $L = 8 \mid \theta = \pi/4 \mid \text{JOB}=3$ (b) $L = 8 \mid \theta = 7\pi/6 \mid \text{JOB}=70$ (c) $L = 8 \mid \theta = 5\pi/3 \mid \text{JOB}=72$ 

Figura 21 – Trajetória do PCA aplicado a representação dos Pesos da Rede Neural para Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = \pi/4$, o 3º job; (ii) para $\theta = 7\pi/6$, o 70º job e (iii) para $\theta = 5\pi/3$, o 72º job.

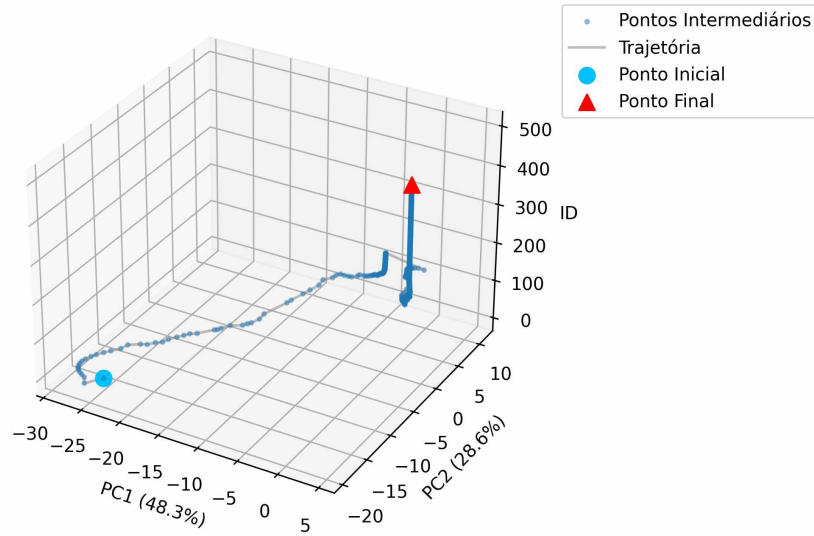
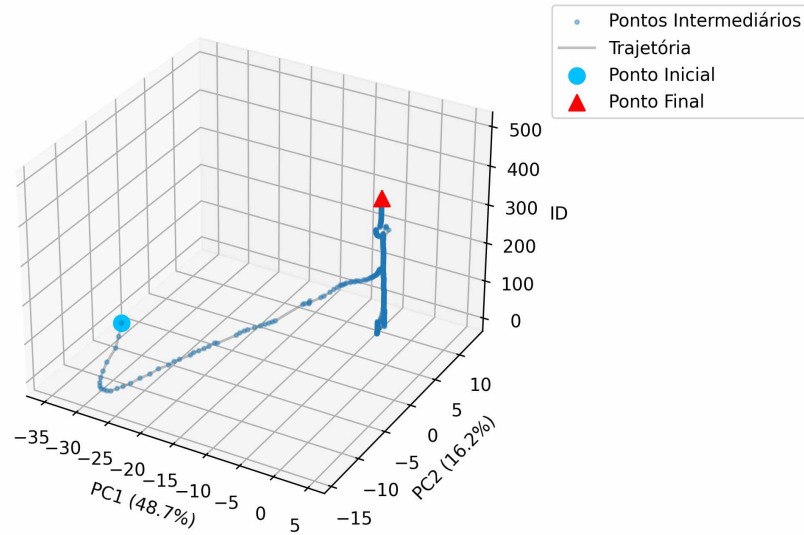
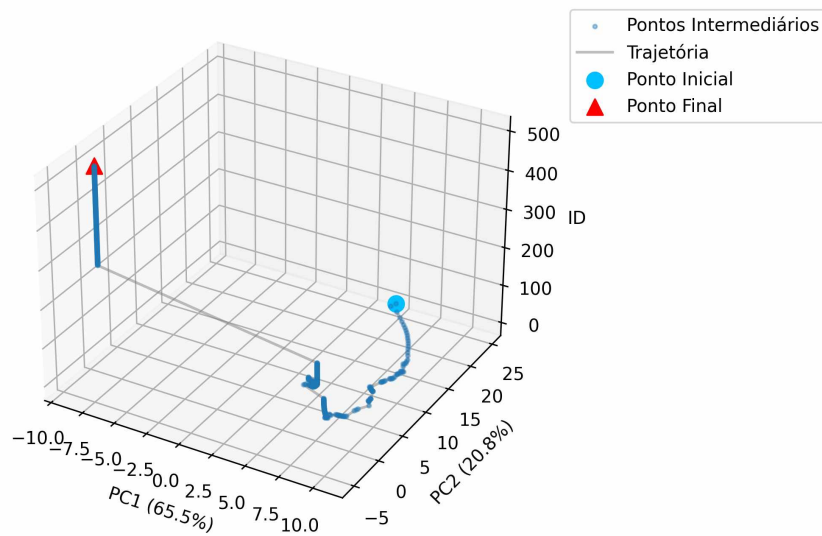
(a) $L = 8 \mid \theta = \pi/4 \mid \text{JOB}=3$ (b) $L = 8 \mid \theta = \pi/4 \mid \text{JOB}=4$ (c) $L = 8 \mid \theta = \pi/4 \mid \text{JOB}=11$ 

Figura 22 – Trajetória do PCA aplicado a representação dos Pesos da Rede Neural para Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = \pi/4$ com a seguinte seqüências de $\text{job } NJ = [3, 4, 11]$. Observa-se que as trajetórias dos $\text{jobs } 3$ e 4 são bastante semelhantes, ao passo que a trajetória do $\text{job } 11$ difere significativamente, apresentando uma quase descontinuidade após à iteração 200 ($IT = 200$).

Análise baseada em Correlação de Pearson

Em nossa tomografia de redes neurais, o objetivo é investigar se os pesos internos da rede são capazes de adquirir as características dos sistemas físicos aos quais são aplicadas, bem como compreender de que forma esses pesos aprendem e representam tais propriedades.

Um dos aspectos fundamentais em sistemas de cadeias de spin é o emaranhamento. A quantificação do emaranhamento constitui uma área em franca expansão dentro da área da informação quântica. Embora este trabalho não tenha como foco propor ou desenvolver novos quantificadores, buscamos verificar se as redes neurais são capazes, de algum modo, de representar essa característica intrínseca dos sistemas quânticos.

Utilizaremos o conceito de entropia de Von Neumann para representar o emaranhamento em nossas cadeias de spin. Para o cálculo efetivo dessa quantidade, empregaremos a ferramenta [DMRG](#) e relacionaremos seus resultados com o método de cálculo da correlação de Pearson.

Inicialmente iremos investigar a correlação de Pearson⁴ para a configuração para $L = 8$, com 500 interações para $\theta = 5\pi/3$ para o 72º *job* de um total de 100 execuções.

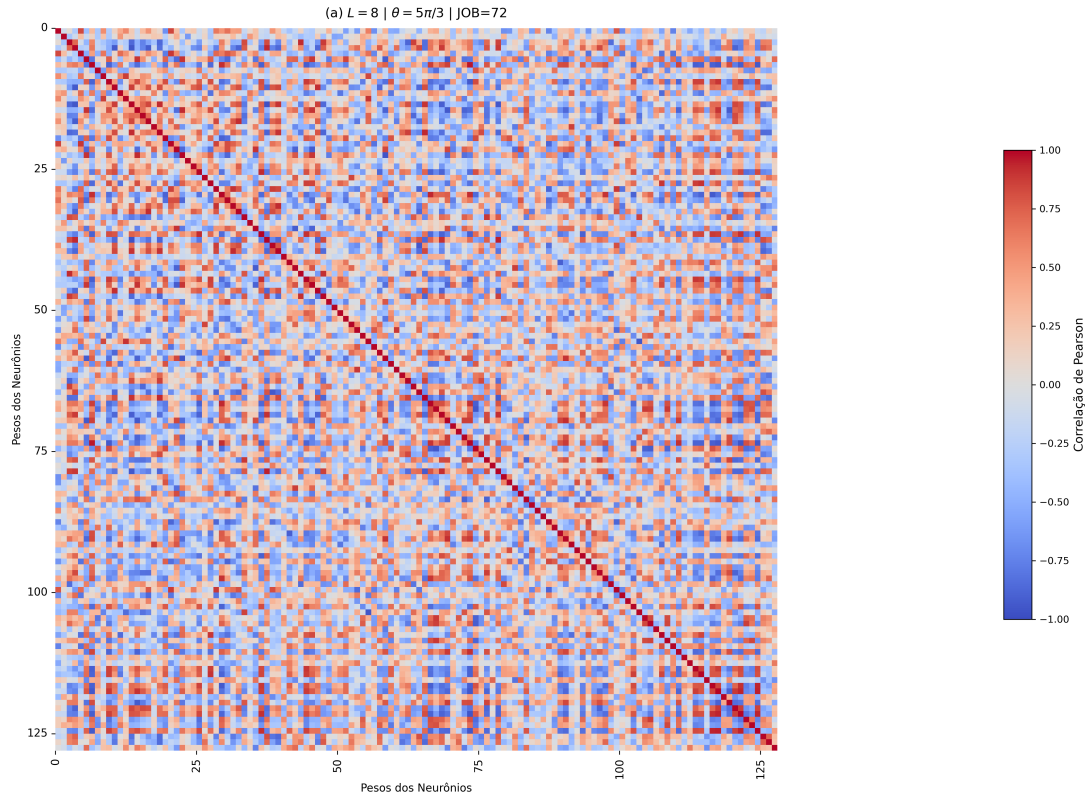


Figura 23 – Correlação de Pearson para a representação dos Pesos da Rede Neural para Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = 5\pi/3$, o 72º *job* de um total de 100 execuções.

⁴ Nos apêndices, ao final deste trabalho, apresentamos uma descrição detalhada a Medida de Correlação de Pearson.

Em uma primeira análise, observa-se que a Figura 23 apresenta alta complexidade visual e dificuldade de interpretação, assemelhando-se mais às telas de Jackson Pollock [32] do que propriamente a uma análise tomográfica de uma RNA.

Em geral, análises para Correlação de Pearson envolvem uma quantidade pequena de atributos ($\ll 30$) que nesta situação, temos um conjunto de 128 neurônios, gerando uma matriz quadrada de correlação com dimensão de 128 x 128.

A partir da dificuldade de interpretação das correlações em conjuntos de alta dimensionalidade, propomos dois diagramas para auxiliar a análise e a compreensão dos resultados. O primeiro diagrama, apresentado na Figura 24, mostra o histograma dos valores de correlação, os quais estão situados no intervalo $[-1, 1]$, juntamente com suas respectivas frequências.

Observa-se nesse diagrama que a distribuição dos valores de correlação se aproxima de uma curva gaussiana, concentrando um número considerável de correlações próximas de zero. Esse comportamento indica que o sistema apresenta um grande conjunto de vetores linearmente independentes.

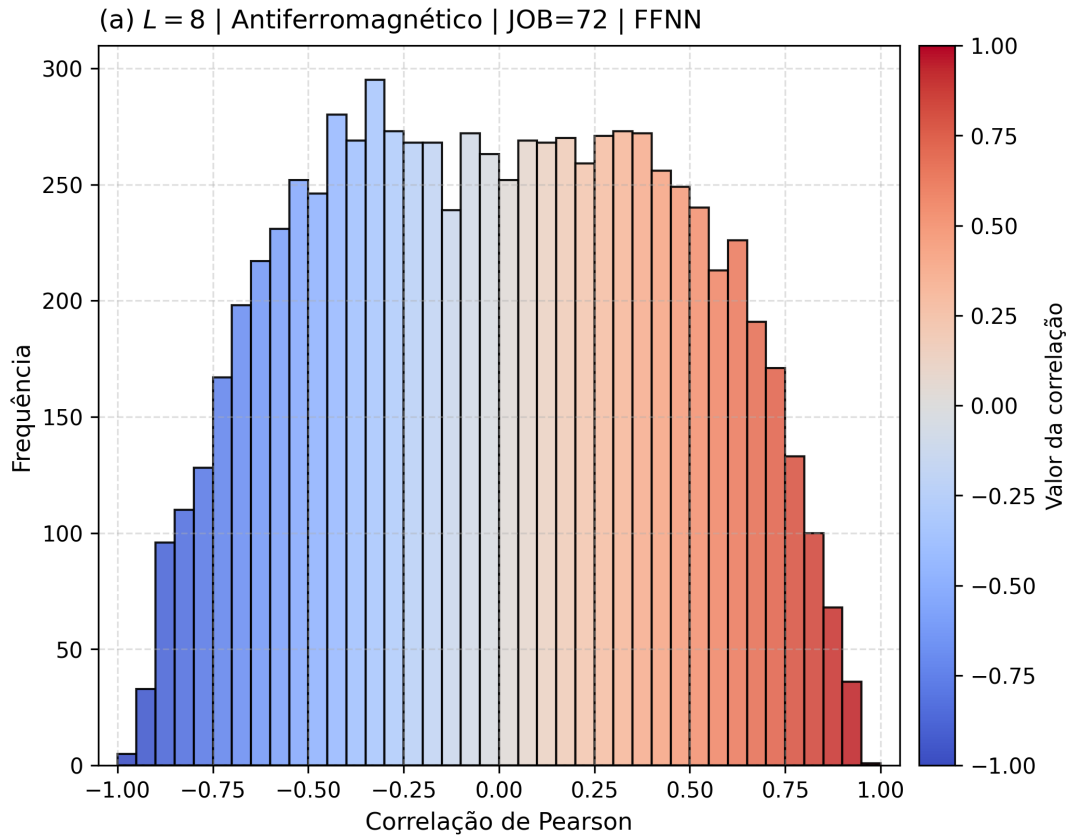


Figura 24 – Histograma da Correlação de Pearson para a representação dos Pesos da Rede Neural para Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = 5\pi/3$, o 72º job de um total de 100 execuções.

Para ratificar as observações obtidas a partir do primeiro diagrama, apresentamos agora um segundo diagrama, exibido na Figura 25, que mostra o boxplot dos valores de correlação, situados no intervalo $[-1, 1]$, juntamente com suas respectivas frequências.

Observa-se que a maioria dos valores de correlação se concentra na faixa aproximada entre -0.4 e 0.4 , o que indica que a correlação de Pearson global para esta configuração é relativamente baixa.

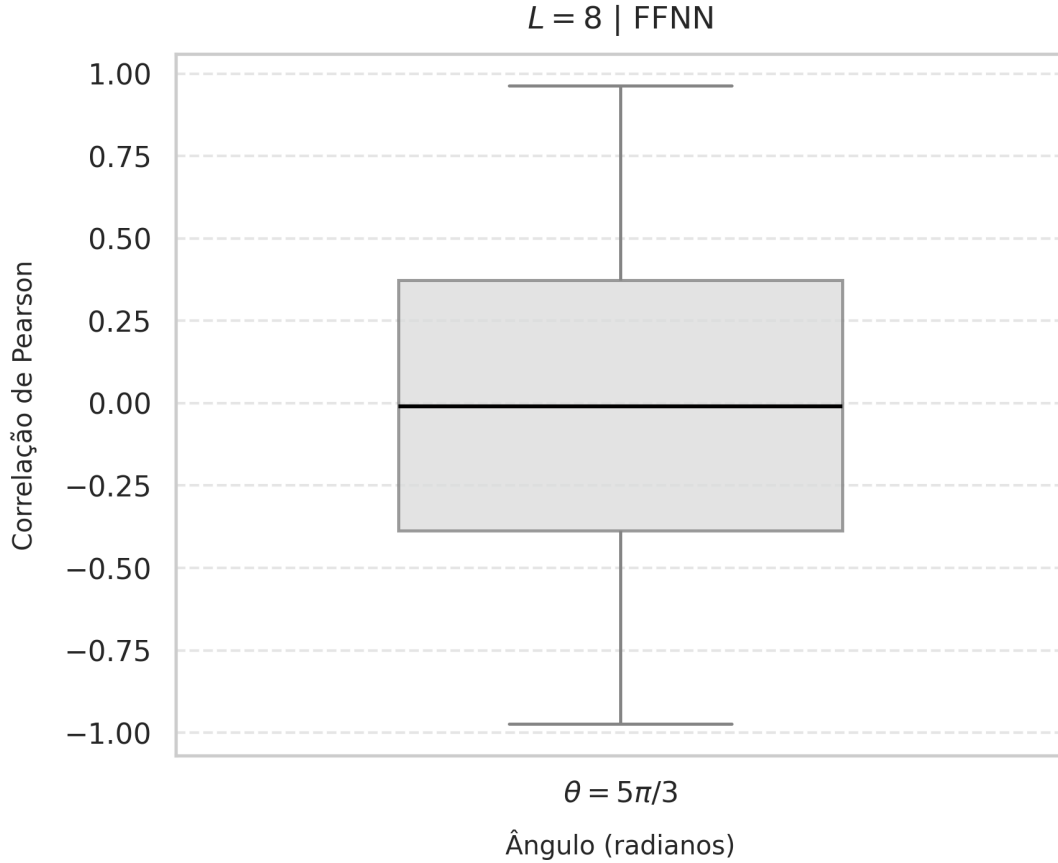


Figura 25 – Boxplot da Correlação de Pearson para a representação dos Pesos da Rede Neural para Simulação para uma cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = 5\pi/3$, o 72º job de um total de 100 execuções.

Para efeitos comparativos iremos, analisar a medida de correlação de Pearson com dois regimes: ferromagnético ($\theta = 7\pi/3$) e antiferromagnético ($\theta = 5\pi/3$).

Ângulo (°)	S_{vN}
$7\pi/6$	1.27
$5\pi/3$	1.74

Tabela 1 – Entropia de von Neumann para diferentes regimes do modelo $J_1 - J_2$.

Ao analisar apenas a matriz de correlação apresentada na Figura 26, observa-se que, no regime ferromagnético, os neurônios apresentam correlações mais intensas em comparação ao regime antiferromagnético. Essa observação está em consonância com o comportamento da entropia de von Neumann, mostrado na Tabela 1, em que se verifica que o regime antiferromagnético possui maior entropia. Essas evidências também podem ser visualizadas nos diagramas a seguir, correspondentes ao histograma e ao boxplot.

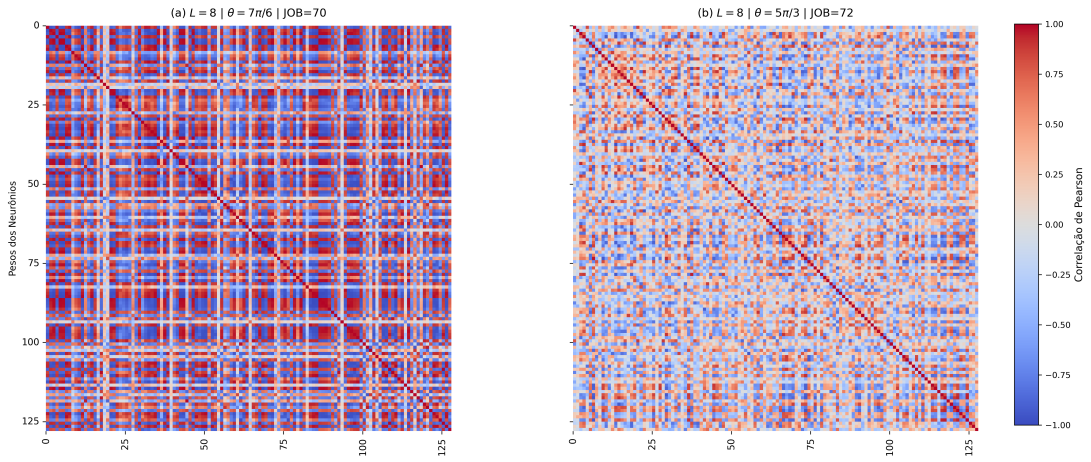


Figura 26 – Correlação de Pearson para cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = 7\pi/6$, o 70° e $\theta = 5\pi/3$, o 72° job .

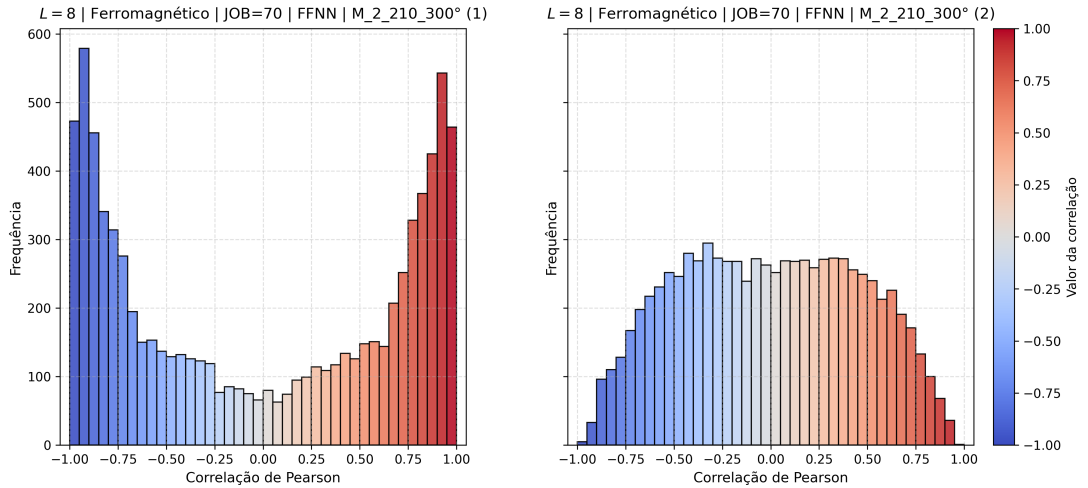


Figura 27 – Histograma da Correlação de Pearson para cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = 7\pi/6$, o 70° e $\theta = 5\pi/3$, o 72° job .

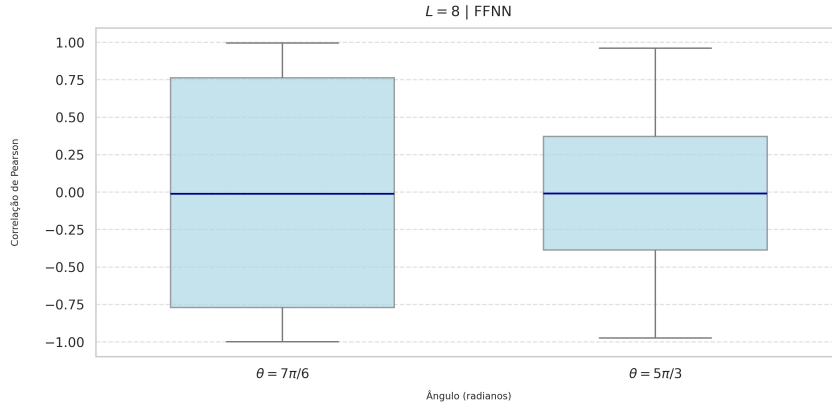


Figura 28 – Boxplot da Correlação de Pearson para cadeia de spins de tamanho $L = 8$, com 500 interações para $\theta = 7\pi/6$, o 70° e $\theta = 5\pi/3$, o 72° job .

Por fim, apresentamos uma análise referente ao regime frustrado, considerando seus respectivos valores de entropia de Von Neuman (Tabela 2), seus valores de correlação, representado no diagrama do tipo *boxplot*. Foram investigados os ângulos de acoplamento $\theta = \pi/6$, $\pi/4$, $\pi/3$, $2\pi/3$ e $5\pi/6$.

Ângulo ($^\circ$)	S_{vN}
$\pi/6$	1.963997
$\pi/4$	1.7917599
$\pi/3$	1.85275345
$2\pi/3$	2.25791149
$5\pi/6$	1.76432207

Tabela 2 – Entropia de von Neumann para diferentes acoplamentos referente ao regime frustrado do modelo $J_1 - J_2$.

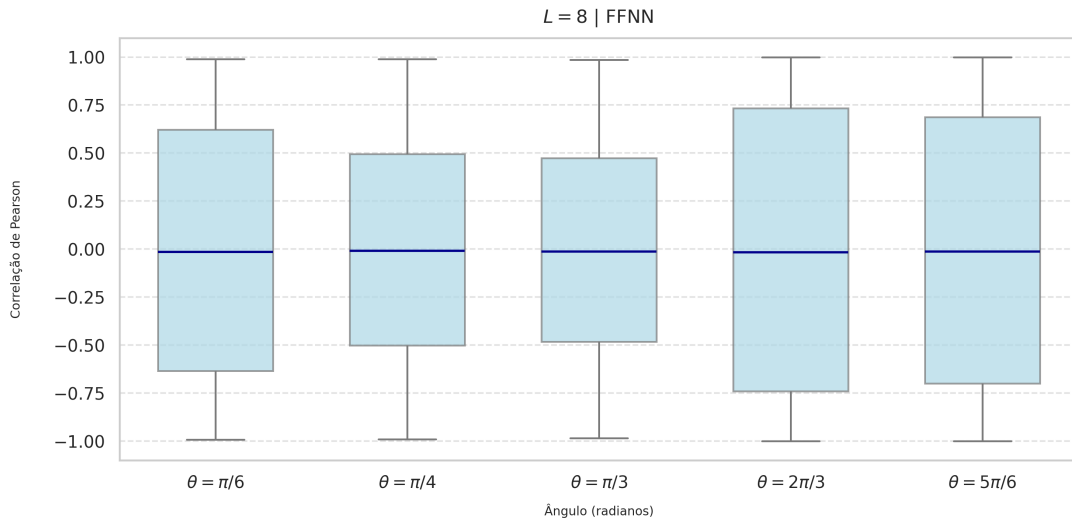


Figura 29 – Boxplot da Correlação de Pearson para cadeia de spins de tamanho $L = 8$, com 500 interações para a fase frustrada para os seguintes acoplamentos $\theta = \pi/6$, $\pi/4$, $\pi/3$, $2\pi/3$ e $5\pi/6$.

Discussão

Analisando os aspectos gerais da tomografia, podemos destacar que: (a) os pesos sinápticos das redes neurais incorporam as características intrínsecas do modelo físico estudado; (b) como o modelo de rede neural é fundamentado em cálculos numérico-computacionais, torna-se evidente a importância de realizar uma etapa de limpeza de dados, removendo *outliers* após a obtenção dos resultados iniciais; (c) é possível implementar uma métrica baseada na distância percorrida pela trajetória no método [PCA](#), de modo a avaliar o desempenho e a convergência do processo de aprendizado; (d) para os regimes ferromagnético e antiferromagnético, o método da correlação de Pearson, em primeira análise, consegue diferenciar cada regime, para o regime ferromagnético, os vetores de base que compõe a rede neural tem uma maior correlação entre si, ou seja, temos menos vetores linearmente independente, já no regime antiferromagnéticos, os vetores são mais linearmente independente, ou seja, o regime antiferromagnético apresenta mais configurações, requisitando mais vetores para representá-los, resultados consistentes com a propriedade da entropia de Von Neumann, onde o regime antiferromagnético apresenta uma maior entropia do que a entropia no regime ferromagnético; (e) por fim, observa-se a necessidade de realizar testes adicionais no regime frustrado, especialmente quanto ao cálculo da entropia de von Neumann via método [DMRG](#) e sua relação com a correlação de Pearson.

Conclusão

Em resumo, este trabalho investigou a determinação da energia do estado fundamental em cadeias unidimensionais de spins utilizando redes neurais artificiais. Foram considerado o modelo J_1 - J_2 , que incorpora interações entre primeiros e segundos vizinhos. Os resultados obtidos demonstram o potencial dessas arquiteturas neurais como ferramentas alternativas para o estudo de sistemas quânticos fortemente correlacionados.

Também avaliamos a evolução dos pesos sinápticos nas camadas internas das redes neurais ao longo do processo de otimização, observando como tais parâmetros passam a refletir, de forma emergente, características físicas intrínsecas ao sistema quântico modelado. Essa análise revela que a rede, ao minimizar a energia variacional, não apenas se ajusta numericamente, mas também tende a codificar propriedades fundamentais da Hamiltoniana, como simetrias e padrões de correlação entre spins. Esses resultados reforçam o potencial interpretativo das redes neurais como representações eficazes de estados quânticos, não apenas do ponto de vista computacional, mas também conceitual, aproximando inteligência artificial e mecânica quântica de sistemas de muitos corpos.

Para trabalhos futuros, seria interessante estender a aplicação desses métodos a sistemas de maior dimensionalidade, como redes bidimensionais (2D) e tridimensionais (3D), onde os desafios computacionais e a complexidade física se intensificam significativamente.

Além disso, explorar outros modelos de interesse em física da matéria condensada, como o modelo de Hubbard e suas extensões, pode oferecer novas perspectivas sobre a capacidade das redes neurais em representar estados fortemente correlacionados. Tais investigações contribuiriam não apenas para a validação da abordagem em cenários mais realistas, mas também para o desenvolvimento de arquiteturas de aprendizado de máquina cada vez mais adaptadas à descrição de sistemas quânticos complexos.

Referências

- [1] John Bardeen e Walter H. Brattain. “The Transistor, A Semiconductor Triode”. Em: *Physical Review* 74.2 (1948), pp. 230–231. DOI: [10.1103/PhysRev.74.230](https://doi.org/10.1103/PhysRev.74.230).
- [2] Hans Bethe. “Zur Theorie der Metalle. I. Eigenwerte und Eigenfunktionen der linearen Atomkette”. German. Em: *Zeitschrift für Physik* 71.3-4 (1931). Introdução do método de Bethe ansatz para o modelo de Heisenberg, pp. 205–226. DOI: [10.1007/BF01341708](https://doi.org/10.1007/BF01341708).
- [3] R. Bursill et al. “Numerical and approximate analytical results for the frustrated spin-1/2 quantum spin chain”. Em: *Journal of Physics: Condensed Matter* 7.45 (1995), pp. 8605–8620. DOI: [10.1088/0953-8984/7/45/012](https://doi.org/10.1088/0953-8984/7/45/012).
- [4] Giuseppe Carleo e Matthias Troyer. “Solving the quantum many-body problem with artificial neural networks”. Em: *Science* 355 (fev. de 2017), pp. 602–606. DOI: [10.1126/science.aag2302](https://doi.org/10.1126/science.aag2302). arXiv: [1606.02318](https://arxiv.org/abs/1606.02318).
- [5] Giuseppe Carleo et al. “Machine learning and the physical sciences”. Em: *Reviews of Modern Physics* 91.4 (2019), p. 045002. DOI: [10.1103/RevModPhys.91.045002](https://doi.org/10.1103/RevModPhys.91.045002).
- [6] Juan Carrasquilla e Roger G. Melko. “Machine learning phases of matter”. Em: *Nature Physics* 13.5 (2017), pp. 431–434. DOI: [10.1038/nphys4035](https://doi.org/10.1038/nphys4035).
- [7] Kelvin Ch’ng et al. “Machine Learning Phases of Strongly Correlated Fermions”. Em: *Phys. Rev. X* 7 (3 2017), p. 031038. DOI: [10.1103/PhysRevX.7.031038](https://doi.org/10.1103/PhysRevX.7.031038). URL: <https://link.aps.org/doi/10.1103/PhysRevX.7.031038>.
- [8] Claude Cohen-Tannoudji, Bernard Diu e Franck Laloë. *Quantum Mechanics*. Vol. 2. Original edition in French: 1977. Wiley-VCH, 2006. ISBN: 9780471164333.
- [9] Anna Dawid et al. “Modern applications of machine learning in quantum sciences”. Em: (2025). Acesso: Junho, 2025. DOI: [10.48550/arXiv.2204.04198](https://doi.org/10.48550/arXiv.2204.04198). arXiv: [2204.04198](https://arxiv.org/abs/2204.04198) [quant-ph]. URL: <https://arxiv.org/abs/2204.04198>.

- [10] Michel Dekking et al. *A Modern Introduction to Probability and Statistics: Understanding Why and How*. London: Springer, 2005, pp. 236–238. ISBN: 978-1-85233-896-1.
- [11] N. D. Drummond e R. J. Needs. “Variance-minimization scheme for optimizing Jastrow factors”. Em: *Physical Review B* 72.8 (ago. de 2005), p. 085124. DOI: [10.1103/PhysRevB.72.085124](https://doi.org/10.1103/PhysRevB.72.085124). arXiv: [physics/0505072](https://arxiv.org/abs/physics/0505072). URL: <https://doi.org/10.1103/PhysRevB.72.085124>.
- [12] W. M. C. Foulkes et al. “Quantum Monte Carlo simulations of solids”. Em: *Reviews of Modern Physics* 73.1 (2001), pp. 33–83. DOI: [10.1103/RevModPhys.73.33](https://doi.org/10.1103/RevModPhys.73.33).
- [13] Paulo de Tarso Cavalcante Freire. *Primeiros Passos na Mecânica Quântica*. isbn = 978854215713. Sigillum Scientia, Editora Corsário, 2020.
- [14] Nigel Goldenfeld. *Lectures on Phase Transitions and the Renormalization Group*. Abordagem pedagógica sobre a aproximação de campo médio. Addison-Wesley, 1992. Cap. 3, pp. 45–67.
- [15] Ian Goodfellow, Yoshua Bengio e Aaron Courville. *Deep Learning*. Acesso: Junho, 2025. MIT Press, 2016. URL: <https://www.deeplearningbook.org>.
- [16] Matthew B. Hastings. “An area law for one-dimensional quantum systems”. Em: *Journal of Statistical Mechanics: Theory and Experiment* 2007.08 (2007). Prova fundamental da lei de área para sistemas 1D, P08024. DOI: [10.1088/1742-5468/2007/08/P08024](https://doi.org/10.1088/1742-5468/2007/08/P08024). arXiv: [0705.2024](https://arxiv.org/abs/0705.2024) [quant-ph]. URL: <https://doi.org/10.1088/1742-5468/2007/08/P08024>.
- [17] Simon Haykin. *Neural Networks: A Comprehensive Foundation*. 2nd. Upper Saddle River, NJ, USA: Prentice Hall, 1998.
- [18] Werner Heisenberg. “Zur Theorie des Ferromagnetismus”. German. Em: *Zeitschrift für Physik* 49.9-10 (1928), pp. 619–636. DOI: [10.1007/BF01328601](https://doi.org/10.1007/BF01328601).
- [19] David Hilbert. “Mathematical Problems”. Em: *Bulletin of the American Mathematical Society* 8.10 (1902). Translation of the lecture delivered at the International Congress of Mathematicians, Paris, 1900. Translated by Mary Winton Newson, pp. 437–479.
- [20] ICTP School on Advanced Techniques for Scientific Programming and Artificial Intelligence for Quantum Many-Body Physics, 2020. *Neural Quantum States Lecture Notes*. https://gitlab.com/nqs/ictp_school/-/blob/master/notes.pdf. Acesso: Junho, 2025.
- [21] Anil K Jain, M Narasimha Murty e Patrick J Flynn. “Data clustering: a review”. Em: *ACM computing surveys (CSUR)* 31.3 (1999). Acesso: Junho, 2025, pp. 264–323. DOI: [10.1145/331499.331504](https://doi.org/10.1145/331499.331504). URL: http://users.eecs.northwestern.edu/~yingliu/datamining_papers/survey.pdf.

- [22] A. N. Kolmogorov. “On the representation of continuous functions of many variables by superpositions of continuous functions of one variable and addition”. English translation available. Em: *Doklady Akademii Nauk SSSR* 114 (1957). In Russian; English translation in *AMS Translations – Series 2*, vol. 28 (1963), pp. 55–59, pp. 953–956.
- [23] Cornelius Lanczos. “Um método iterativo para a solução do problema de autovalores de operadores diferenciais e integrais lineares”. Português. Em: *Journal of Research of the National Bureau of Standards* 45.4 (1950). (PDF), pp. 255–282. DOI: [10.6028/jres.045.026](https://doi.org/10.6028/jres.045.026). URL: <https://doi.org/10.6028/jres.045.026>.
- [24] Yann LeCun, Yoshua Bengio e Geoffrey Hinton. “Deep learning”. Em: *Nature* 521.7553 (2015), pp. 436–444. DOI: [10.1038/nature14539](https://doi.org/10.1038/nature14539).
- [25] Zachary C. Lipton. “The mythos of model interpretability”. Em: *Queue* 16.3 (2018), pp. 31–57. DOI: [10.1145/3236386.3241340](https://doi.org/10.1145/3236386.3241340).
- [26] J. MacQueen. “Some Methods for Classification and Analysis of Multivariate Observations”. Em: 1 (1967), pp. 281–297.
- [27] P. C. Mahalanobis. “Reprint of: On the Generalised Distance in Statistics”. Em: *Sankhya A* 80.1 (dez. de 2018). Reprint of the original 1936 paper, pp. 1–7. ISSN: 0976-8378. DOI: [10.1007/s13171-019-00164-5](https://doi.org/10.1007/s13171-019-00164-5). URL: <https://doi.org/10.1007/s13171-019-00164-5>.
- [28] C. K. Majumdar e D. K. Ghosh. “On next-nearest-neighbor interaction in linear chain. I”. Em: *Journal of Mathematical Physics* 10.8 (1969), pp. 1388–1398. ISSN: 1089-7658. DOI: [10.1063/1.1664978](https://doi.org/10.1063/1.1664978). URL: <http://dx.doi.org/10.1063/1.1664978>.
- [29] Warren S. McCulloch e Walter Pitts. “A Logical Calculus of the Ideas Immanent in Nervous Activity”. Em: *The Bulletin of Mathematical Biophysics* 5.4 (1943), pp. 115–133. DOI: [10.1007/BF02478259](https://doi.org/10.1007/BF02478259).
- [30] Marvin Minsky e Seymour Papert. *Perceptrons: An Introduction to Computational Geometry*. MIT Press, 1969.
- [31] Tom M. Mitchell. *Machine Learning*. McGraw-Hill, 1997.
- [32] Steven Naifeh e Gregory White Smith. *Jackson Pollock: An American Saga*. New York: Clarkson N. Potter, Inc., 1989, p. 503. ISBN: 0-517-56084-4.
- [33] John von Neumann. *The Computer and the Brain*. New Haven: Yale University Press, 1958.
- [34] A. Nigrin. *Neural Networks for Pattern Recognition*. MIT Press, 1993. ISBN: 978-0262140546.

- [35] Román Orús. “A practical introduction to tensor networks: Matrix product states and projected entangled pair states”. Em: *Annals of Physics* 349 (2014), pp. 117–158. DOI: [10.1016/j.aop.2014.06.013](https://doi.org/10.1016/j.aop.2014.06.013).
- [36] Pennsylvania State University. *STAT 462: Regression Analysis — 9.2.3 - Example of Interpreting R^2* . <https://online.stat.psu.edu/stat462/node/96/>. Accessed: 2025-09-15. 2025.
- [37] Shi-Ju Ran et al. *Lecture Notes of Tensor Network Contractions*. Vol. 964. Lecture Notes in Physics. Available as arXiv:1708.09213 [physics.comp-ph]. Springer, 2020. DOI: [10.1007/978-3-030-35462-6](https://doi.org/10.1007/978-3-030-35462-6). URL: <https://doi.org/10.48550/arXiv.1708.09213>.
- [38] Revista Corsário. “Cem anos de mecânica quântica”. Português. Em: *Corsário* (2024). Acesso em: jan. 2026. URL: https://www.corsario.art.br/prosa_1.
- [39] Frank Rosenblatt. “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain”. Em: *Psychological Review* 65.6 (1958), pp. 386–408. DOI: [10.1037/h0042519](https://doi.org/10.1037/h0042519).
- [40] Cynthia Rudin. “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead”. Em: *Nature Machine Intelligence* 1.5 (2019), pp. 206–215. DOI: [10.1038/s42256-019-0048-x](https://doi.org/10.1038/s42256-019-0048-x).
- [41] Stuart Russell e Peter Norvig. *Artificial Intelligence: A Modern Approach*. 3^a ed. Prentice Hall, 2010.
- [42] J. J. Sakurai e Jim Napolitano. *Modern Quantum Mechanics*. 2nd. Cambridge University Press, 2017. ISBN: 9781108422413.
- [43] Ulrich Schollwöck. “The density-matrix renormalization group”. Em: *Reviews of Modern Physics* 77 (2005). Artigo de revisão seminal sobre DMRG, pp. 259–315. ISSN: 0034-6861. DOI: [10.1103/RevModPhys.77.259](https://doi.org/10.1103/RevModPhys.77.259). URL: <https://link.aps.org/doi/10.1103/RevModPhys.77.259>.
- [44] Ulrich Schollwöck. “The density-matrix renormalization group in the age of matrix product states”. Em: *Annals of Physics* 326.1 (2011), pp. 96–192. DOI: [10.1016/j.aop.2010.09.012](https://doi.org/10.1016/j.aop.2010.09.012).
- [45] scikit-learn developers. *sklearn.preprocessing.MinMaxScaler*. <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>. Accessed: 2025-09-15. 2025.
- [46] Emilio Segrè. *Dos Raios X aos Quarks*. Português. Brasília: Editora Universidade de Brasília, 1980.
- [47] Shai Shalev-Shwartz e Shai Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014. URL: <http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning>.

- [48] Ian Sommerville. *Software Engineering*. 9th. Addison-Wesley, 2011. ISBN: 9780137035151.
- [49] L. Tagliacozzo, G. Evenbly e G. Vidal. “Simulation of two-dimensional quantum systems using a tree tensor network”. Em: *Physical Review B* 80.23 (2009), p. 235127. DOI: [10.1103/PhysRevB.80.235127](https://doi.org/10.1103/PhysRevB.80.235127).
- [50] The Nobel Foundation. *The Nobel Prize in Physics 2024*. Inglês. <https://www.nobelprize.org/prizes/physics/2024/summary/>. Prêmio concedido a John J. Hopfield e Geoffrey E. Hinton por contribuições fundamentais ao desenvolvimento das redes neurais artificiais. 2024.
- [51] Michael E. Tipping e Christopher M. Bishop. “Probabilistic Principal Component Analysis”. Em: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61.3 (1999), pp. 611–622. DOI: [10.1111/1467-9868.00196](https://doi.org/10.1111/1467-9868.00196).
- [52] Giacomo Torlai et al. “Neural-network quantum state tomography”. Em: *Nature Physics* 14.5 (2018), pp. 447–450. DOI: [10.1038/s41567-018-0048-5](https://doi.org/10.1038/s41567-018-0048-5).
- [53] Matthias Troyer e Uwe-Jens Wiese. “Computational complexity and fundamental limitations to fermionic quantum Monte Carlo simulations”. Em: *Physical Review Letters* 94 (2005), p. 170201. DOI: [10.1103/PhysRevLett.94.170201](https://doi.org/10.1103/PhysRevLett.94.170201).
- [54] Alan M. Turing. “Computing Machinery and Intelligence”. Em: *Mind* 59.236 (1950), pp. 433–460. DOI: [10.1093/mind/LIX.236.433](https://doi.org/10.1093/mind/LIX.236.433).
- [55] C. J. Umrigar, K. G. Wilson e J. W. Wilkins. “Optimized trial wave functions for quantum Monte Carlo calculations”. Em: *Physical Review Letters* 60.17 (1988), pp. 1719–1722. DOI: [10.1103/PhysRevLett.60.1719](https://doi.org/10.1103/PhysRevLett.60.1719).
- [56] E. Vernek. “Robustness of Stark many-body localization in the J_1 - J_2 Heisenberg model”. Em: *Physical Review B* 105 (2022), p. 075124. DOI: [10.1103/PhysRevB.105.075124](https://doi.org/10.1103/PhysRevB.105.075124). URL: <https://link.aps.org/doi/10.1103/PhysRevB.105.075124>.
- [57] F. Verstraete e J. I. Cirac. “Renormalization algorithms for Quantum-Many Body Systems in two and higher dimensions”. Em: *arXiv preprint cond-mat/0407066* (2004). arXiv: [cond-mat/0407066](https://arxiv.org/abs/cond-mat/0407066) [[cond-mat.str-el](https://arxiv.org/abs/cond-mat/0407066)].
- [58] Steven R. White. “Density matrix formulation for quantum renormalization groups”. Em: *Physical Review Letters* 69.19 (1992). Introdução ao método DMRG, crucial para estudos numéricos de cadeias de spin frustradas., pp. 2863–2866. DOI: [10.1103/PhysRevLett.69.2863](https://doi.org/10.1103/PhysRevLett.69.2863).

Apêndices

Modelo de Heisenberg

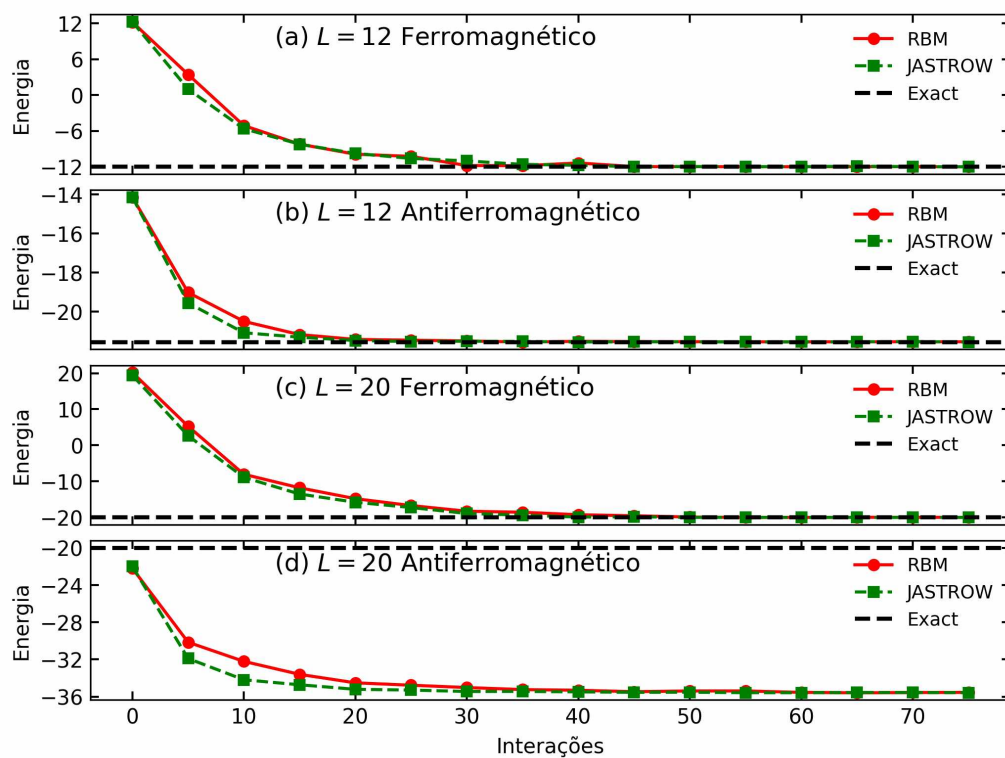


Figura 30 – Determinação da energia do estado fundamental em uma cadeia de spins $1/2$ para diferentes valores de L e números de interações I . O sistema é configurado para $L = 12$ e $L = 20$, com 80 interações ($I = 80$) com Rede Neural Tipo RBM.

Neste apêndice iremos apresentar resultados referente ao Modelo de Heisenberg. Esta implementação reproduz os resultados para uma cadeia de spin $1/2$ cujos fundamentos físicos estão detalhados em [4]. A energia do estado fundamental foi calculada utilizando uma RNA tipo RBM para os regimes ferromagnético e antiferromagnético, com validação cruzada realizada por meio de dois métodos independentes: (i) abordagem variacional de Jastrow e (ii) diagonalização exata via algoritmo de Lanczos. Os resultados comparativos para sistemas com $L = 12$ e $L = 20$ sítios, considerando $I = 80$ interações é apresentado

na Figura 30.

A análise comparativa revelou comportamentos similares entre os regimes magnéticos: tanto para o sistema menor ($L = 12$), considerando a RNA tipo RBM, quanto para o sistema maior ($L = 20$), o regime antiferromagnético convergiu de forma mais acelerada para o valor da energia fundamental. No caso como foi realizado um número limitado de simulações, ainda não pode ser discutido se o comportamento entre as escalas posso surgir uma transição na dinâmica do sistema relacionada ao tamanho da cadeia onde efeitos de tamanho tornam-se significativos na determinação das propriedades de convergência em relação a esse tipo de RNA.

Também foi realizado uma nova simulação utilizando uma RNA do tipo FFNN onde obtivemos um desempenho melhor e convergência para o regime antiferromagnético.

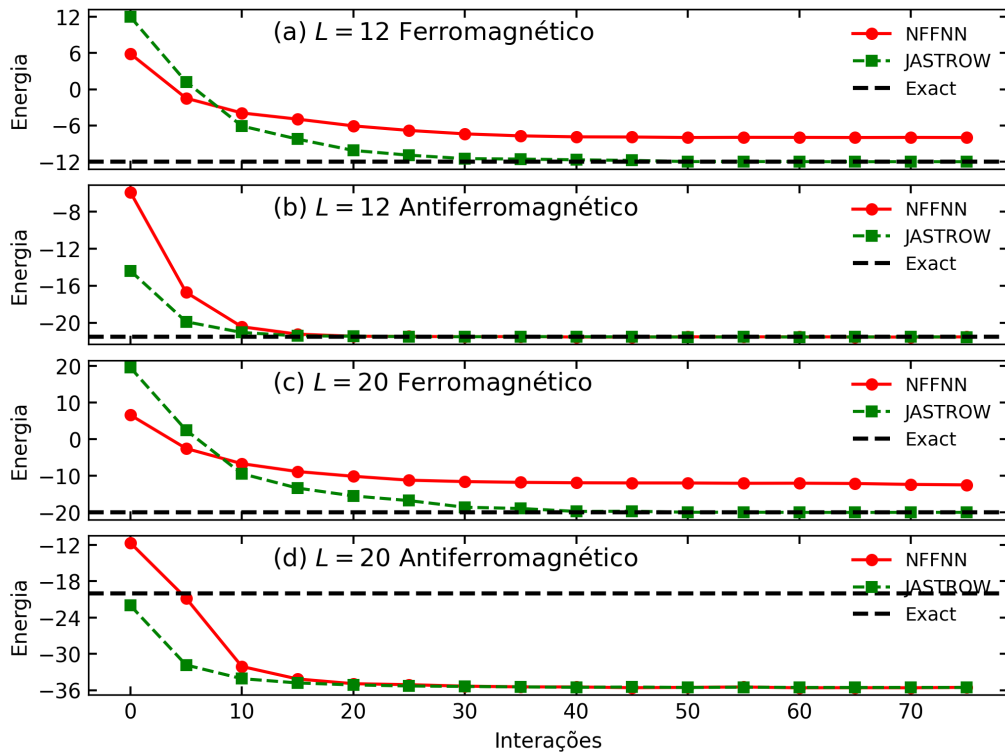


Figura 31 – Determinação da energia do estado fundamental em uma cadeia de spins 1/2 para diferentes valores de L e números de interações I . O sistema é configurado para $L = 12$ e $L = 20$, com 80 interações ($I = 80$) com Rede Neural Tipo NFFNN

Tipos de redes neurais aplicadas ao modelo de Heisenberg

Nesta simulação a seguir, iremos consideramos uma cadeia unidimensional com $L = 24$ sítios para analisar sistematicamente a influência da arquitetura de RNA na determinação do estado fundamental do sistema. Especificamente, compararemos o desempenho de diferentes topologias de redes neurais na estimativa da energia do estado fundamental, avaliando tanto a precisão quanto a eficiência computacional de cada abordagem.

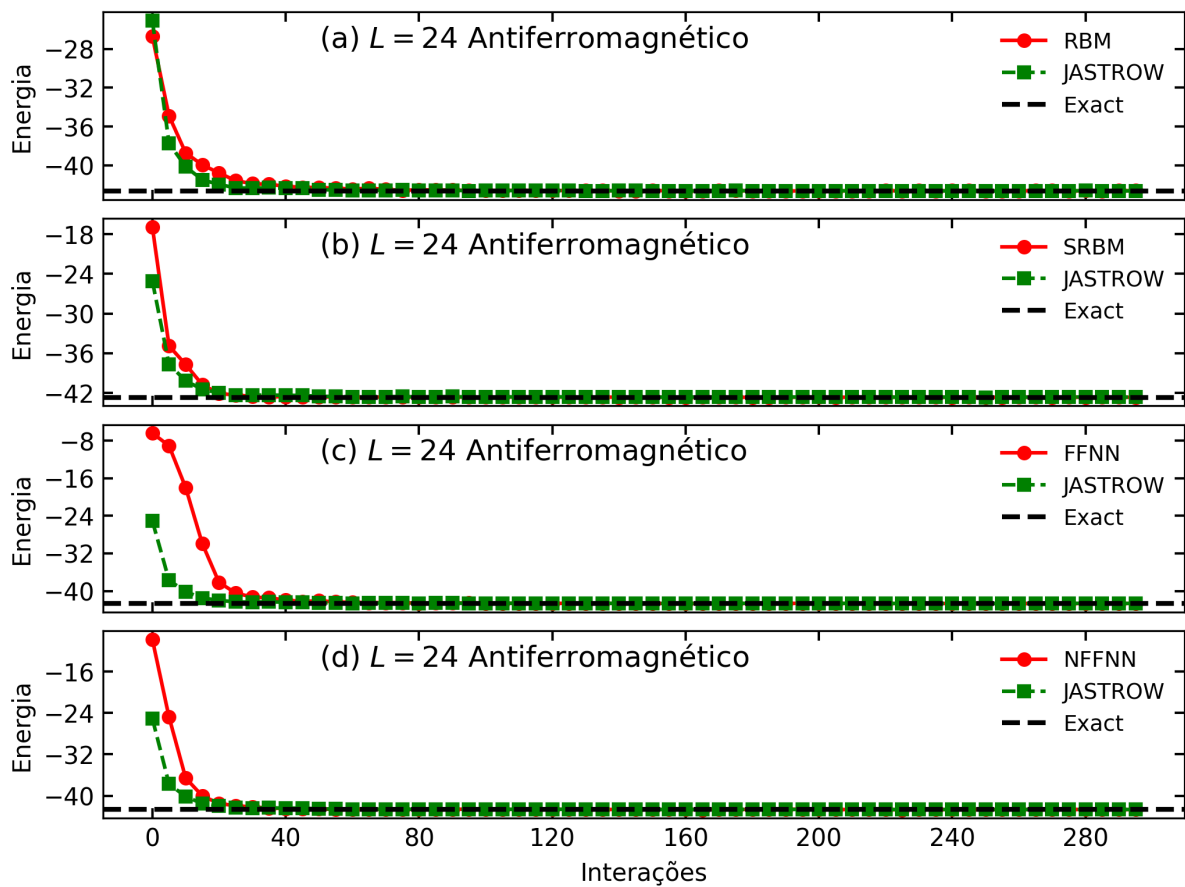


Figura 32 – Determinação da energia do estado fundamental em uma cadeia de spins 1/2 para L fixo e para o regime antiferromagnético. Temos variações no tipo da rede neural aplicada. O sistema é configurado para $L = 24$ com 300 interações ($I = 100$) para 4 tipos de Rede Neurais: RBM, SRBM, FFNN, NFFNN.

A Figura 32 apresenta os resultados para uma cadeia com $L = 24$, onde comparamos quatro arquiteturas de redes neurais: (i) RBM, (ii) Symmetric Restricted Boltzmann Machine (SRBM), (iii) FFNN de camada única e (iv) NFFNN multicamada ($N = 2$). Destas, a SRBM demonstrou convergência mais rápida, enquanto as variações FFNN e NFFNN apresentaram desempenho similares em termos de convergência, com uma melhor performance referente a convergência para a NFFNN.

Ressalta-se que a [NFFNN](#) exibe um comportamento anômalo na fase final da simulação, caracterizado por flutuações atípicas nos valores de convergência, o que sugere possíveis instabilidades numéricas ou dificuldades de otimização no regime assintótico.

Análise Exploratória de Dados

No capítulo de [Resultados](#), apresentamos a metodologia utilizada para investigar a evolução dos pesos sinápticos em [RNA](#), recorrendo a técnicas consolidadas de análise exploratória de dados, em particular a [PCA](#) e o coeficiente de correlação de Pearson, cujos fundamentos são detalhados a seguir.

B.1 Principal Component Analysis - PCA

O método [PCA](#), ou seja, a análise de componentes principais, é uma técnica estatística multivariada amplamente utilizada para redução de dimensionalidade e visualização de dados, em termos matemático, [PCA](#) é uma transformação linear ortogonal que transforma os dados para um novo sistema de coordenadas com uma dimensão menor do que o sistema de coordenadas do sistema inicial [47],[51].

Seu objetivo principal é transformar um conjunto de variáveis arbitrárias em um novo conjunto de variáveis ortogonais. Tal conjunto de variáveis ortogonais são definidos como componentes principais e são escolhidos numa ordem decrescente de variância em relação aos dados originais.

Ao projetar os dados em um espaço de dimensão reduzida, formado pelos componentes mais significativos, o [PCA](#) permite identificar padrões, tendências e agrupamentos, melhorando a interpretação de estruturas complexas inerentes a conjuntos de dados de alta dimensionalidade. Iremos ilustrar a partir de dados numéricos, por exemplo, como é o algoritmo do método [PCA](#). Dado uma matriz com os seguintes valores

$$A = \begin{bmatrix} 4 & 3 & 2 & 1 \\ 8 & 2 & 9 & 1 \\ 4 & 1 & 2 & 8 \\ 3 & 3 & 3 & 1 \end{bmatrix},$$

neste exemplo, cada linha representar um objeto qualquer representado em 4 dimensões.

Iremos construir a matriz de covariância. Inicialmente realizaremos a transformação de mudança de escala para que os dados fiquem entre o valor de -1 a 1 , fazendo com que a média dos valores fique próxima de 0 . Um método apropriado para esta finalidade é o método *min-max scaling* [45] dado por

$$x' = 2 \frac{x - \min(x)}{\max(x) - \min(x)} - 1, \quad (68)$$

onde $\min(x)$ e $\max(x)$ representam, respectivamente, o valor mínimo e o valor máximo do conjunto total dos dados. Após a transformação de escala obtemos a seguinte matriz

$$A' = \begin{bmatrix} -0.25 & -0.5 & -0.75 & -1 \\ 0.75 & -0.75 & 1 & -1 \\ -0.25 & -1 & -0.75 & 0.75 \\ -0.5 & -0.5 & -0.5 & -1 \end{bmatrix}.$$

O próximo passo é encontrar a matriz de covariância Σ , cujos elementos são definidos por

$$\Sigma_{ij} = \text{cov}(i, j) = \frac{1}{n-1} \sum_{k=1}^n (x_{k,i} - \bar{x}_i)(x_{k,j} - \bar{x}_j) \quad (69)$$

onde n é o número de observações; $x_{k,i}$ e $x_{k,j}$ são os valores das variáveis i e j na k -ésima observação; $\bar{x}_i$ e $\bar{x}_j$ são as médias calculada para as colunas i e j , respectivamente.

Esta matriz é simétrica e captura as relações lineares entre todos os pares de variáveis do conjunto de dados. Abaixo segue a matriz de covariância Σ calculada a partir da matriz A :

$$\Sigma = \begin{bmatrix} 0.3073 & -0.0365 & 0.4375 & -0.1094 \\ -0.0365 & 0.0573 & -0.0885 & 0.0365 \\ 0.4375 & -0.0885 & 0.6875 & -0.1875 \\ -0.1094 & 0.0365 & -0.1875 & 0.5729 \end{bmatrix}.$$

Para efeito ilustrativo, os valores foram arredondados para 4 casas decimais. Esta matriz é simétrica, como esperado para uma matriz de covariância.

Em seguida, a partir da matriz de covariância Σ , calculam-se seus **autovetores** e **autovalores**. Ordenam-se os autovetores em ordem decrescente de magnitude de seus autovalores correspondentes, formando assim uma base ortogonal onde os primeiros vetores capturam a máxima variância dos dados. Esses autovetores ordenados constituem os componentes principais da matriz original, permitindo a representação dos dados em um novo espaço de características, otimizado para reduzir dimensionalidade e destacar padrões estatisticamente significativos.

Para a matriz Σ em questão, temos os seguintes autovalores (ordenados de forma decrescente) e seus respectivos autovetores

$$\begin{aligned}
\lambda_1 &= 1.1272, & \mathbf{v}_1 &= \begin{bmatrix} 0.4567 \\ -0.1023 \\ 0.7845 \\ -0.4121 \end{bmatrix}, \\
\lambda_2 &= 0.4201, & \mathbf{v}_2 &= \begin{bmatrix} 0.5121 \\ 0.6234 \\ -0.2310 \\ 0.5417 \end{bmatrix}, \\
\lambda_3 &= 0.0732, & \mathbf{v}_3 &= \begin{bmatrix} -0.7012 \\ 0.3115 \\ 0.4023 \\ 0.4987 \end{bmatrix}, \\
\lambda_4 &= 0.0235, & \mathbf{v}_4 &= \begin{bmatrix} 0.1987 \\ 0.7054 \\ 0.4012 \\ -0.5523 \end{bmatrix}.
\end{aligned}$$

Para determinar a variância total, calcularemos a soma de todos os autovalores

$$\lambda_{\text{total}} = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1.1272 + 0.4201 + 0.0732 + 0.0235 = 1.6440.$$

Uma definição utilizada pelo [PCA](#) é a variância explicada, sendo calculada como o somatório dos autovalores respectivos, ou seja, para dois primeiros componentes temos a seguinte variância explicada como

$$\lambda_1 + \lambda_2 = 1.1272 + 0.4201 = 1.5473$$

A porcentagem de variância explicada por λ_1 e λ_2 vale

$$\frac{\lambda_1 + \lambda_2}{\lambda_{\text{total}}} = \frac{1.5473}{1.6440} \times 100\% \approx 94.12\%$$

Portanto, os dois primeiros componentes principais capturam **94.12%** da variância total dos dados. A projeção de cada vetor (linha) da matriz original A no novo espaço bidimensional é realizada mediante o produto escalar com os autovetores correspondentes a λ_1 e λ_2 . Formalmente, a matriz reduzida $A'' \in \mathbb{R}^{4 \times 2}$ é calculada por

$$A'' = A \cdot \mathbf{V}_2,$$

onde $\mathbf{V}_2 = [\mathbf{v}_1 \ \mathbf{v}_2]$ é a matriz cujas colunas são os autovetores dos dois principais autovalores. Abaixo, apresenta-se a matriz A'' resultante, que representa os dados originais no espaço dos componentes principais

$$A'' = \begin{bmatrix} [A'']_{11} & [A'']_{12} \\ [A'']_{21} & [A'']_{22} \\ [A'']_{31} & [A'']_{32} \\ [A'']_{41} & [A'']_{42} \end{bmatrix},$$

ou seja, a matriz original A projetada no espaço dos dois primeiros componentes principais vale

$$A'' = \begin{bmatrix} 2.6768 & 3.9983 \\ 10.0974 & 3.8063 \\ -0.0033 & 6.5434 \\ 3.0046 & 3.2552 \end{bmatrix},$$

onde a matriz A'' contém $\approx 94.12\%$ da variância dos dados originais.

Representação dos dados na rede neural **NNQS**

Em nosso problema, temos a seguinte representação de dados, através da matriz

$$M = \begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,k} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,k} \\ v_{3,1} & v_{3,2} & \cdots & v_{3,k} \\ \vdots & \vdots & \cdots & \vdots \\ v_{m,1} & v_{m,2} & \cdots & v_{m,k} \end{bmatrix},$$

onde cada linha i representa a i -ésima interação referente ao aprendizado da **RNA**; as colunas representam os neurônios com seus respectivos valores onde elemento $v_{i,j}$ representa o valor do neurônio j na interação i .

Para a representação da matriz M em termos de componentes principais, ou seja, utilizando o método **PCA**, realizamos os seguintes passos:

1. Cálculo da matriz M' a partir da transformação de escala $[-1, 1]$ com o algoritmo min-max scalling;
2. Cálculo da matriz de covariância;
3. Cálculo dos autovetores e autovalores da matriz de covariância;
4. Representação dos autovetores em ordem decrescente referente aos autovalores;
5. Escolha dos 2 autovetores mais representativos na ordem decrescente dos respectivos autovalores;
6. Projeção (produto escalar) de cada linha sobre os 2 autovetores definidos;
7. Geração da nova matriz de dados M''' ;

Desta forma, a matriz de dados M aplicada ao método acima transforma-se na matriz M''' com as seguinte formulação

$$M''' = \begin{bmatrix} p_{1,1} & p_{1,2} \\ p_{2,1} & p_{2,2} \\ p_{3,1} & p_{3,2} \\ \vdots & \vdots \\ p_{m,1} & p_{m,2} \end{bmatrix},$$

onde a cada linha representa as interações ; os elementos $p_{1,i}$ e $p_{i,2}$ representam as componentes principais para cada interação i . Com os elementos $p_{i,1}$ e $p_{i,2}$ e o valor da i -ésima interação realizamos a representação gráfica do PCA em 3 dimensões. Em seguida, é possível realizar as análises referente aos comportamentos de acordo com cada regime em questão, ilustrado graficamente nas figuras a seguir. Nas figuras 33, 34, temos 2 si-

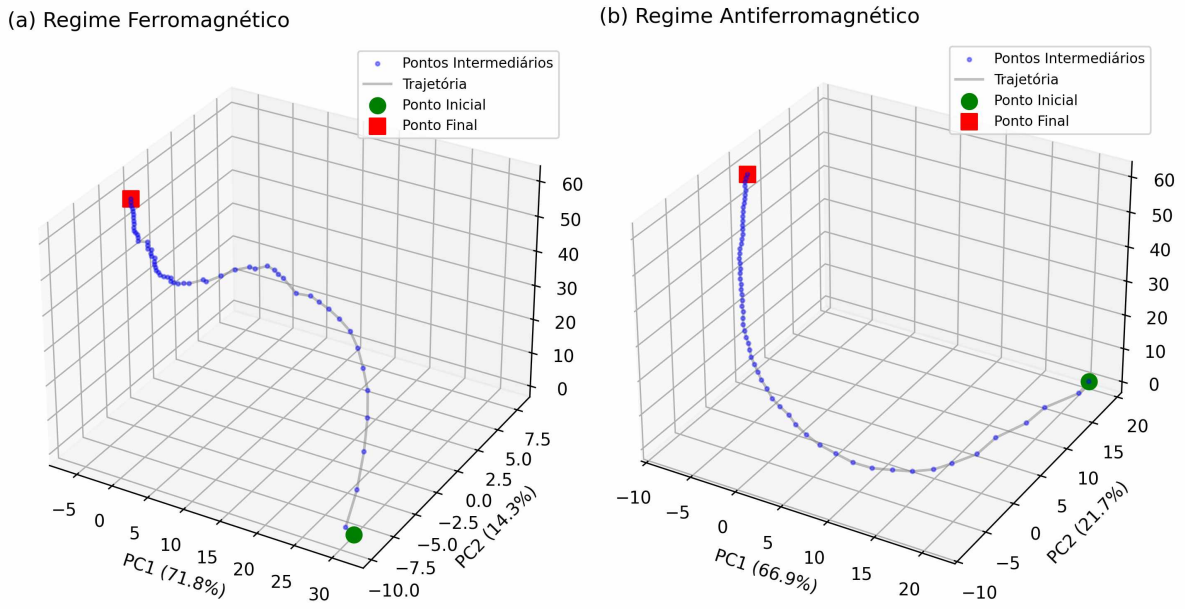


Figura 33 – Trajetória dos pesos sinápticos no espaço de PCA ao longo de $I = 60$ iterações para $L = 12$, utilizando uma RNA tipo RBM. A figura mostra a evolução dos vetores de pesos desde o início até a convergência para os regimes ferromagnético e antiferromagnético.

mulações para o modelo de Heisenberg para cadeias de spins para $L = 12$ e $L = 20$ com 60 interações ($IT = 60$) para os regimes ferromagnético e antiferromagnético, respectivamente. Verificamos 2 comportamentos diferentes: para o regime ferromagnético encontramos uma maior descontinuidade nos passos em relação ao regime antiferromagnético, em outras palavras, cada interação i e $i + 1$ no modelo ferromagnético tem uma maior distância euclidiana comparada ao regime antiferromagnético. Por exemplo, para a figura 34, para o regime ferromagnético entre a interação $IT = 5$ e $IT = 10$ temos um salto, o que não ocorre em toda a trajetória do regime antiferromagnético.

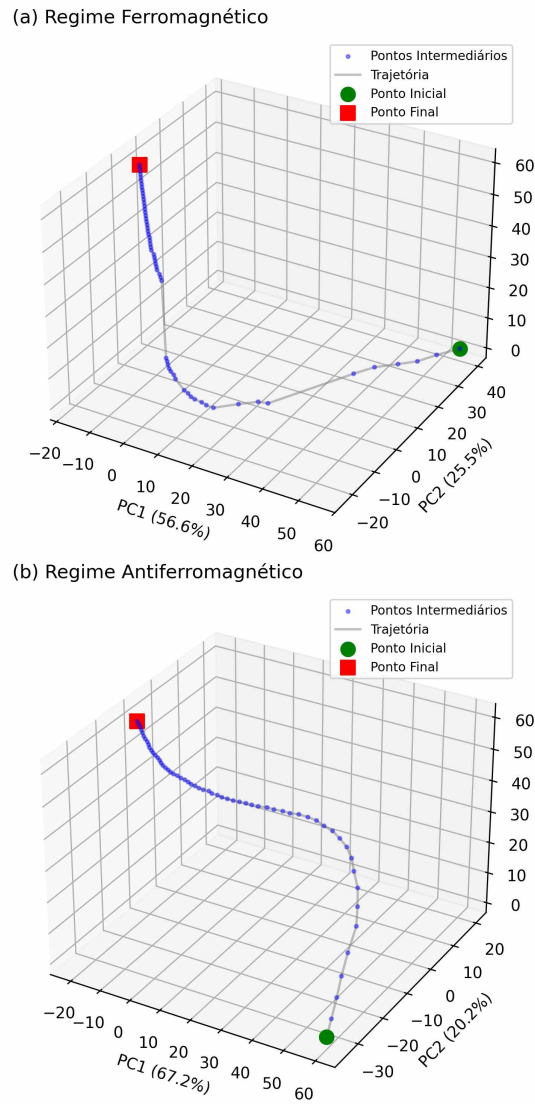


Figura 34 – Trajetória dos pesos sinápticos no espaço de **PCA** ao longo de $I = 40$ iterações para $L = 20$, utilizando uma **RNA** tipo **RBM**. A figura mostra a evolução dos vetores de pesos desde o início até a convergência.

B.2 Correlação de Pearson

A correlação de Pearson é uma técnica estatística para testar se há relação linear entre 2 variáveis quantitativas [36]. A correlação de Pearson é uma análise bivariada, ou seja, uma análise bidirecional, por exemplo se temos 3 variáveis A , B e C , as possíveis análises são entre a variável A e B ; B e C ; A e C . De forma qualitativa, temos 3 tipos de correlação: correlação positiva, correlação negativa e correlação neutra, ilustrado graficamente na figura 35 abaixo.

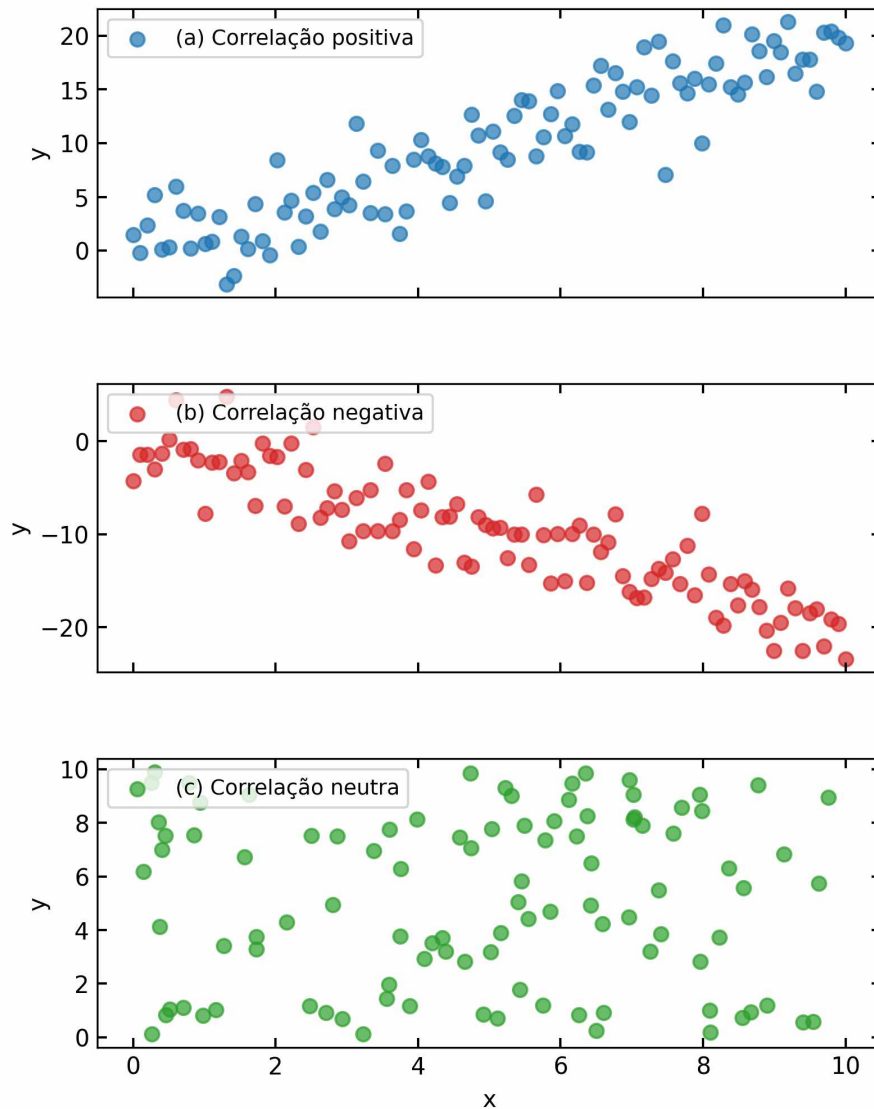


Figura 35 – 3 tipos de correlação de Pearson. (a) Correlação positiva; (b) Correlação negativa e (c) Correlação neutra.

O coeficiente de correlação de Pearson, denotado por r , mede a intensidade e a direção da relação linear entre duas variáveis quantitativas. Matematicamente, ele é definido como

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (70)$$

em que x_i e y_i representam os valores observados das variáveis X e Y , enquanto $\bar{x}$ e $\bar{y}$ correspondem às respectivas médias amostrais. Essa fórmula evidencia que o coeficiente de correlação é obtido pela razão entre a covariância de X e Y e o produto dos seus desvios padrão, tornando a medida adimensional.

O coeficiente de correlação de Pearson é altamente sensível a *outliers*, pois seu cálculo, fundamentado nas médias ($\bar{x}$, $\bar{y}$) e nos desvios padrão das variáveis. Estas duas medidas podem ser drasticamente influenciadas por um único valor atípico. Esses valores extremos inflam indevidamente as medidas de tendência central e dispersão, podendo alterar a magnitude e até mesmo o sinal da correlação, o que compromete a interpretação da real relação linear entre os dados.

Matriz de correlação de Pearson

A partir do coeficiente de correlação de Pearson é possível construir a matriz de correlação, que representa os valores do coeficiente calculados para todas as combinações possíveis entre as colunas (ou variáveis) de um conjunto de dados. Vamos ilustrar a partir da tabela abaixo.

Tabela 3 – Matriz de dados exemplo 1

Linha	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
1	1	2	1	5	3	7	10	-9	4	8
2	2	4	2	6	5	6	9	-8	7	5
3	3	6	2	7	6	5	12	-7	2	9
4	4	8	3	8	8	7	11	-6	5	6
5	5	10	4	9	7	6	13	-5	8	7
6	6	12	5	10	6	8	12	-4	3	4
7	7	14	5	11	9	7	14	-3	6	5
8	8	16	6	12	10	9	15	-2	5	6
9	9	18	6	13	8	10	13	-1	7	5
10	10	20	7	14	11	9	16	0	4	8

A partir da tabela de dados, calcula-se a matriz M de correlação, representada na Figura 36. Observa-se que a diagonal principal da matriz M apresenta correlação positiva perfeita ($r = +1$), pois cada variável é comparada consigo mesma, e que a matriz é simétrica, isto é, o elemento M_{ij} é igual ao elemento M_{ji} . Nesta construção, as colunas C1, C2, C3 e C4 apresentam correlação positiva entre si, enquanto as colunas C9 e C10 exibem uma correlação negativa entre si. Neste conjunto de dados gerados, observamos que as correlação entre (C1–C9), (C2–C9), (C3–C9), (C4–C9), (C5–C9), (C6–C9), (C7–C9) e (C8–C9) são correlações quase nulas pois os valores estão próximo de 0, ou seja, tais colunas não tem correlação entre si para estes dados apresentados na tabela anterior.

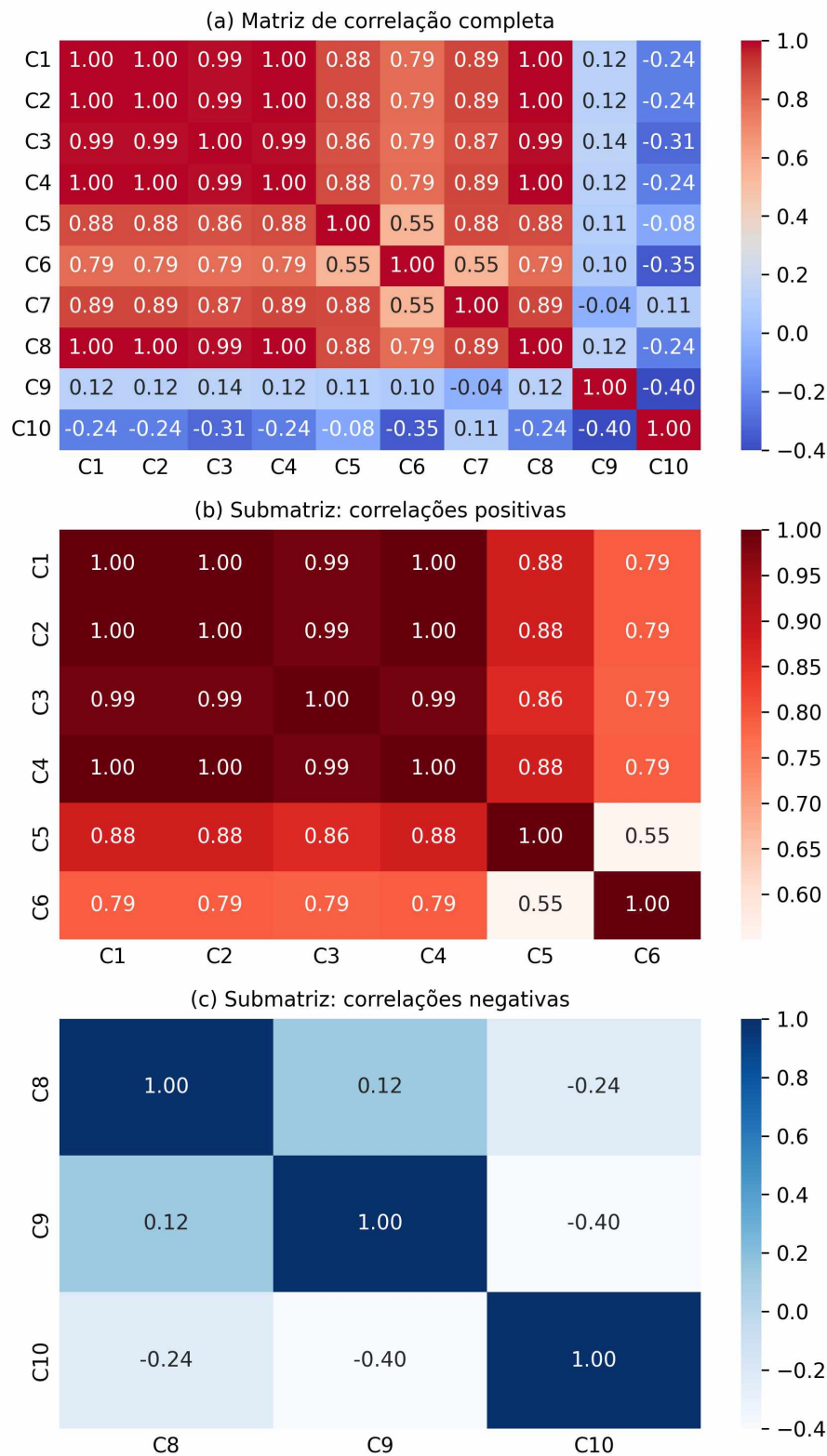


Figura 36 – Matriz de correlação de Pearson para a tabela de matriz de dados exemplo 1. Por construção, as colunas $C1$ – $C6$ apresentam correlação positiva entre si, enquanto $C9$ e $C10$ apresentam correlação negativa. A imagem (a) apresenta a matriz completa, a (b) mostra o subconjunto das correlações positivas ($C1$ – $C6$), e a (c) exibe o subconjunto das correlações negativas ($C9$ – $C10$) e ($C8$ – $C10$). A correlação entre ($C8$ – $C9$) é quase nula (valor próximo de 0).

