

#### UNIVERSIDADE ESTADUAL DE CAMPINAS Instituto de Física Gleb Wataghin

### WILLIAM FREITAS E SILVA

# Abordagem de *deep learning* para sistemas quânticos de muitos corpos

## Deep learning approach for quantum many-body systems

CAMPINAS 2024

#### WILLIAM FREITAS E SILVA

# Abordagem de *deep learning* para sistemas quânticos de muitos corpos

Tese apresentada ao Instituto de Física Gleb Wataghin da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de DOUTOR EM CIÊN-CIAS, na Área de FÍSICA.

Orientador: SILVIO ANTONIO SACHETTO VITIELLO

ESTE TRABALHO CORRESPONDE À VER-SÃO FINAL DA TESE DEFENDIDA PELO ALUNO WILLIAM FREITAS E SILVA E ORI-ENTADA PELO PROF. DR. SILVIO ANTO-NIO SACHETTO VITIELLO.

> CAMPINAS 2024

Ficha catalográfica Universidade Estadual de Campinas (UNICAMP) Biblioteca do Instituto de Física Gleb Wataghin Lucimeire de Oliveira Silva da Rocha - CRB 8/9174

Silva, William Freitas e, 1996Si38a Abordagem de *deep learning* para sistemas quânticos de muitos corpos / William Freitas e Silva. – Campinas, SP : [s.n.], 2024.
Orientador: Silvio Antonio Sachetto Vitiello. Tese (doutorado) – Universidade Estadual de Campinas (UNICAMP), Instituto de Física Gleb Wataghin.
1. Aprendizado profundo. 2. Método variacional de Monte Carlo. 3. Sistemas finitos. I. Vitiello, Silvio Antonio Sachetto, 1950-. II. Universidade Estadual de Campinas (UNICAMP). Instituto de Física Gleb Wataghin. III. Título.

#### Informações Complementares

Título em outro idioma: Deep learning approach for quantum many-body systems Palavras-chave em inglês: Deep learning Variational Monte Carlo method Finite systems Área de concentração: Física Titulação: Doutor em Ciências Banca examinadora: Silvio Antonio Sachetto Vitiello [Orientador] Lucas Madeira Maurice de Koning Ricardo Luís Doretto Von Braun Nascimento Data de defesa: 30-08-2024 Programa de Pós-Graduação: Física

Identificação e informações acadêmicas do(a) aluno(a) - ORCID do autor: https://orcid.org/0000-0002-8020-2117 - Currículo Lattes do autor: http://lattes.cnpq.br/4775055720566003



MEMBROS DA COMISSÃO EXAMINADORA DA TESE DE DOUTORADO DO ALUNO WILLIAM FREITAS E SILVA - RA 157585 APRESENTADA E APROVADA AO INSTITUTO DE FÍSICA GLEB WATAGHIN, DA UNIVERSIDADE ESTADUAL DE CAMPINAS, EM 30/08/2024.

INSTITUTO DE FÍSICA

**GLEB WATAGHIN** 

COMISSÃO JULGADORA:

- Prof. Dr. Silvio Antonio Sachetto Vitiello (IFGW/ UNICAMP) Presidente e Orientador
- Prof. Dr. Lucas Madeira (Instituto de Física de São Carlos USP)
- Prof. Dr. Maurice de Koning (IFGW/ UNICAMP)
- Prof. Dr. Ricardo Luís Doretto (IFGW/ UNICAMP)

- Prof. Dr. Von Braun Nascimento (Universidade Federal de Minas Gerais, ICE, Departamento de Física)

**OBS.**: Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

Aos meus pais.

These have been the best years of my life, and they are mine. Tomorrow is promised to no one, Doctor, but I insist upon my past. I am entitled to that. It's mine.

(Clara Oswald)

## AGRADECIMENTOS

Agradeço aos meus pais Ivan e Ivone, aos quais dedico esta tese de doutorado. Agradeço também aos meus pais de consideração Cida e Tony, que me acolheram por muitos anos. Certamente não poderia deixar de agradecer novamente ao Ricardo Takamatsu e ao meu orientador Silvio Vitiello, os dois professores que mais me influenciaram nessa trajetória. Sob a segurança desses pilares que este trabalho foi construído.

Muitas pessoas foram importantes nesses anos, dentre elas agradeço aos meus primos e amigos de infância Jonathan, Marcus, Igor, Fabiano, Josiane e Beatriz; aos amigos de graduação Luis Henrique e Juliana; aos amigos de pós Luiz Guilherme, Luiz Filipe, Helder, Mateus, Valéria, Isabelle, Gabriel Reis, Gabriel Soares; aos amigos que a vida me presenteou como a Clara, a Júlia e o Caio. Agradeço a Larissa pelo apoio especial desses últimos meses. Agradeço também ao Randy por estar sempre aqui!

Agradeço especialmente meu amigo Lucas Menezes e a minha psicóloga Silvia por me ajudarem em momentos delicados, nos quais a saúde mental foi um fator crucial. Muitas vezes a saúde mental é um aspecto da vida que é deixado de lado ou dado pouca importância pelas pessoas, porém, eu não teria terminado esse doutorado se não houvesse procurado ajuda, dado a devida importância a este tópico e batalhado para me manter saudável mentalmente.

Por fim, o presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, processo nº 88887.514478/2020-00. Agradeço o apoio financeiro da CAPES pelos primeiros seis meses de auxílio no início do meu doutorado. Agradeço também pelo apoio financeiro durante o período restante de doutoramento da Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) através do processo nº 2020/10505-6 e ao processo vinculado nº 2016/17612-7. Agradeço a infraestrutura computacional utilizada para realizar as simulações deste trabalho que foram realizadas em parte no Centro Nacional de Processamento de Alto Desempenho em São Paulo (CENAPAD-SP) e no supercomputador Delta no National Center for Supercomputing Applications (NCSA).

## RESUMO

Dentro do contexto da mecânica quântica, a descrição de estados quânticos é fundamental para explicar e prever o comportamento de sistemas finitos. Uma estratégia para realizar isso é por meio do método variacional, onde os estados quânticos são descritos por funções de onda tentativa. Além disso, a semelhança impressionante entre algoritmos de aprendizado profundo e o método variacional sugere que funções tentativa podem ser representadas por redes neurais artificiais. De fato, essa representação foi empregada neste trabalho para analisar sistemas compostos de bósons, como aglomerados de hélio-4, e férmions, como pontos quânticos. Ao usar a integração de Monte Carlo para calcular as integrais relevantes e treinar a rede neural para encontrar a energia do estado fundamental desses sistemas, os resultados obtidos atingiram uma precisão sem precedentes em comparação com os métodos quânticos de Monte Carlo. Embora estes métodos produzam valores estatísticos robustos para a energia de sistemas bosônicos, eles exigem extrapolação para quantidades que não comutam com o hamiltoniano, potencialmente introduzindo viés. Portanto, com o objetivo de contornar extrapolações, a representação do estado fundamental através de redes neurais foi empregada para calcular várias propriedades, como perfis de densidade, funções de distribuição de pares e funções densidade de pares. Além disso, correlações de curto alcance no contexto de universalidade fraca para aglomerados de hélio foram investigadas. Também, para tentar entender o sucesso de redes neurais descrevendo sistemas quânticos, propriedades de escala do tamanho da rede e a estrutura nodal das funções tentativa otimizadas foram analisadas para pontos quânticos.

**Palavras-chave:** Aprendizado profundo; Método variacional de Monte Carlo; Sistemas finitos

## ABSTRACT

In the field of quantum mechanics, the description of quantum states is fundamental to explain and predict the behaviour of finite systems. One strategy to accomplish this is through the variational method, where the quantum states are described by trial wave functions. Moreover, the striking resemblance between deep learning algorithms and the variational method suggests that trial functions can be represented by artificial neural networks. Hence, this representation was employed in this work to analyse systems composed of bosons, such as helium-4 clusters, and fermions, such as quantum dots. By using Monte Carlo integration to compute the relevant integrals and by training the neural network to find the ground state energy of these systems, the yielded results reached unprecedented accuracy compared with quantum Monte Carlo methods. While those methods produce robust statistical values for the energy of bosonic systems, it requires extrapolation for quantities that do not commute with the Hamiltonian, potentially introducing bias. Therefore, aiming to circumvent extrapolations, the neural network representation of ground states was employed to compute several properties such as density profiles, pair distribution functions, and pair density functions. Moreover, short range correlations in the context of weak universality for helium clusters was investigated. Also, in order to understand the success of neural networks describing quantum systems, properties of scaling the size of the network and the nodal structure of the optimised trial functions were analysed for quantum dots.

Keywords: Deep learning; Variational Monte Carlo method; Finite systems

## LISTA DE FIGURAS

2.1	Representação esquemática do conjunto experimental usado para produção		
<u>?</u> ?	Representação comparativa dos experimentos de Andronikashvili "eléssico"	20	
2.2	(a) e microscópico (b).	26	
2.3	Comparação entre os potenciais de Lennard-Jones (vermelho) e HFD-B (azul) em função da distância para interações entre átomos de hélio. A unidade de distância está em Angstrom e os potenciais em unidades de Kel- vin.	28	
3.1	Representação de um ponto quântico circular de dupla heteroestrutura	36	
4.1	Ilustração esquemática de um neurônio artificial. Os sinais $a_i$ chegam ao neurônio com diferentes pesos $w_i$ e são combinados linearmente junto com o viés $b$ . O resultado desse processamento passa por um filtro não linear e é unitida para expréssion	4.77	
4.2	emitido para os proximos neuronios	47	
5.1	parâmetros $\theta$ pelas matrizes de pesos <b>W</b> Diagrama das etapas do processo de otimização empregadas no método va- riacional de Monte Carlo. Cada execução das 4 etapas é uma iteração e elas	49	
	são realizadas repetidamente até a minimização da energia.	68	
6.1	Modelo esquemático da rede neural artificial que compõe a função de onda tentativa. O canal de partícula única é representado nos quadrados azuis, enquanto o canal de duas partículas está em verde. A fase final de montagem	70	
7.1	da função e mostrada em amarelo	72	
	Lennard-Jones.	82	
7.2	Diversos processos de treinamento mostrando o progresso da energia (azul) em função das iterações de otimização. Os aglomerados considerados conti- nham de 2 até 10 átomos. Para comparação, resultados de DMC são exibidos por um linha tracejada laranja.	83	
	x 0 0	-	

7.3	Representação das distribuições $P(r)$ (Eq. 2.16) para o trímero obtidas	
	através de redes neurais e de cálculos de DMC em laranja e azul, respecti-	
	vamente.	85
7.4	Energias totais em função do número de partículas no aglomerado é exibido	
	em círculos pretos. Um ajuste desses dados ao modelo de gota é representado	
	por uma linha vermelha. Por razões de comparação, um ajuste similar feito	
	com base em resultados de DMC é apresentado em azul tracejado.	86
7.5	Dimensionamento do tempo computacional gasto $t_N$ em termos de <i>walltime</i>	
	para diversos tamanhos de <i>clusters</i> em relação ao tempo de treinamento do	
	dímero $t_2$ . A linha tracejada em laranja representa um ajuste polinomial de	
	segundo grau.	86
7.6	Função de onda do dímero projetada no plano $xy$ em função das coordenadas	
	após a otimização da energia. Um dos átomos está fixo na origem e o eixo $z$	
	mostra em uma escala arbitrária o valor da função quando o outro átomo está	
	nas coordenadas $(x, y, 0)$ . Para enfatizar as regiões de alta probabilidade,	
	uma mapa de cores foi utilizado.	88
7.7	Perfis de densidade em função da distância ao centro de massa para aglo-	
	merados compostos de 2 até 14 átomos, onde as cores vão do amarelo para	
	o azul com o aumento do número de partículas. São considerados apenas	
	<i>clusters</i> com números pares de átomos.	89
7.8	Função densidade de pares para aglomerados de <sup>4</sup> He <sub>N</sub> contendo $N = \{4, 6,$	
	8, 10, 12, 14 } átomos em função da distância e apresentadas nas cores do	
	amarelo para o azul, nessa ordem.	90
7.9	Valores para os contatos $C_2^{(N)}$ em função do número de partículas N em	
	escala logarítmica.	90
7.10	Colapso para curtas distâncias das funções densidade de pares $\rho_N$ normali-	
	zadas pelos seus respectivos contatos $C_2^{(N)}$ . As diferentes cores do amarelo	
	para o azul mostram resultados para aglomerados contendo de $N = 2$ até	
	N = 14.	91
7.11	Distribuições de Jacobi $P_N^{(3,jacobi)}$ de três corpos para diferentes tamanhos de	
	aglomerado em função da distância $\rho$ de um átomo até o centro de massa	
	de um par de outros dois átomos.	91
7.12	Colapso para curtas distâncias $\rho$ das distribuições de Jacobi $P_N^{(3,\mathrm{jacobi})}$ nor-	
	malizadas pelos seus respectivos contatos par-átomo $C_{2+1}^{(N)}$	92

7.13 Processo de otimização da energia para diferentes representações dos atributos da rede neural, isto é, para diferentes funções tentativa. O pontos azuis utilizaram a função  $\psi_{r_{ij}}$ , os verdes empregaram a função  $\psi_{\mathbf{r}_{ij}}$  e os laranjas a função completa  $\psi_{\mathbf{q}_i,\mathbf{r}_{ij}}.$ O destaque mostra as energias variacionais estimadas após o processo de otimização para as respectivas funções tentativa. 93 7.14 Módulo das diferenças de probabilidade das funções não radiais com relação à função radial, onde os pontos laranja são dados por  $|\psi_{r_{ij}} - \langle \psi_{\mathbf{q}_i,\mathbf{r}_{ij}} \rangle|$  e os pontos verde por  $|\psi_{r_{ij}} - \langle \psi_{\mathbf{r}_{ij}} \rangle|$ . O destaque mostra a densidade de probabilidade  $\psi_{r_{ij}}^2$ .... 93 Média móvel da energia total entre 100 simulações independentes em função 8.1 da evolução das iterações do processo de treinamento em azul escuro para um ponto quântico contendo 12 elétrons. O desvio padrão móvel também é mostrado nas faixas em tons de azul claro, bem como a energia calculada 96 por DMC em tracejado vermelho. Histogramas de frequência da distribuição de energia para diferentes momen-8.2 tos do processo de otimização, sendo que os histogramas são construídos a partir das médias móveis nas iterações  $2^{12}$ ,  $2^{13}$  e  $2^{15}$  do azul escuro para o claro, respectivamente. As médias de cada distribuição são representadas 97 por linhas tracejadas. Histograma da distribuição de energia ao final do processo de treinamento. 8.3 A linha tracejada preta representa a média da distribuição, enquanto a tracejada vermelha representa a energia calculada através do DMC. Note a escala diferente das energias no eixo x com respeito a da Fig. 8.2. . . . . 98 Evolução da energia durante o processo de otimização para pontos quânti-8.4 cos contendo 6 e 12 elétrons nos quadros superiores, 20 e 30 elétrons nos inferiores, ambos da direita para esquerda. A energia é mostrada em azul, enquanto os tracejados laranjas representam as energia calculadas pelo método DMC. 99 8.5 Histograma bidimensional normalizado das posições dos elétrons em um ponto quântico contendo 20 elétrons. O perfil de densidade foi construído utilizando  $100 \times 2^{10}$  configurações amostradas da função de onda tentativa após o processo de otimização. 100 Mapa de cores do logaritmo da função de onda não normalizada em função 8.6 das coordenadas e em escala arbitrária. Para compor o mapa, uma configuração foi amostrada da densidade de probabilidade gerada pela função de onda e a função foi computada ao varrer um dos elétrons por todo espaço bidimensional, com todos os outros elétrons fixos. 101

- 8.7 Comportamento da energia de um ponto quântico contendo 6 elétrons com o aumento da largura do canal de um elétron. As diferentes cores representam larguras diferentes para o canal de dois elétrons, sendo que o azul, laranja, verde e vermelho correspondem à 8, 12, 16 e 24 de largura.
- 8.8 Comportamento da energia de um ponto quântico contendo 12 elétrons com o aumento da largura do canal de um elétron. As diferentes cores representam larguras diferentes para o canal de dois elétrons, sendo que o azul, laranja, verde e vermelho correspondem à 8, 12, 16 e 24 de largura. . . . 103

102

## LISTA DE TABELAS

2.1	Parâmetros para o potencial HFD-B(HE). Os dígitos mostrados são sufici-	
	entes para não causar erros de arredondamento	27
3.1	Resumo das várias escalas de energia e comprimento envolvidas na modela-	
	gem dos elétrons de pontos quânticos circulares baseados em semicondutores.	38
7.1	Típicos valores dos hiper parâmetros empregados nas simulações de aglome-	
	rados de hélio	81
7.2	Em unidades de Kelvin, são apresentadas a energia cinética $\langle \mathcal{T} \rangle$ e a energia	
	total $\langle E \rangle$ obtidas com a função tentativa $\psi_{\rm BHC}$ para aglomerados de ${}^4{\rm He}$	
	contendo $N$ átomos. A quarta coluna mostra resultados de DMC da literatura.	84
8.1	Comparação entre valores de energia total em unidades de $H^\ast$ obtidas com	
	a função tentativa $\psi_{\rm FN}$ e com o método DMC para pontos quânticos de	
	camada fechada contendo ${\cal N}$ = 6, 12, 20 e 30 elétrons. A terceira coluna	
	mostra a diferença entre os dois métodos em $mH^*$	99
D.1	Em unidades de Kelvin, são apresentadas a energia cinética $\langle \mathcal{T} \rangle$ e a energia	
	total $\langle E \rangle$ obtidas com a função tentativa $\psi_{\rm BHC}$ para aglomerados de ${}^4{\rm He}$	
	contendo N átomos. $\ldots$	130
D.2	Coeficientes ajustados ao modelo de gota considerando estimativas de ener-	
	gia através do método DMC e através das estimativas variacionais obtidas	
	pela representação da função de onda tentativa $\psi_{\rm BHC}$	130
D.3	Estimativas dos valores de contato por partícula ${\cal C}_2^{(N)}/N$ para vários tamanhos	
	de aglomerados de ${}^{4}\text{He}_{N}$	130

# Conteúdo

## 1. INTRODUÇÃO E MOTIVAÇÃO

ТI
----

Ι	Co	onsiderações sobre os sistemas de interesse	23
2.	Сан	RACTERÍSTICAS DE AGLOMERADOS DE HÉLIO	<b>24</b>
	2.1	Propriedades de aglomerados ou <i>clusters</i> de hélio	24
	2.2	Apresentação e análise dos potenciais de interação	26
	2.3	Intuições físicas do comportamento da função de onda	28
	2.4	Propriedades universais e correlações de curto alcance	31
3.	ATF	RIBUTOS DO GÁS DE ELÉTRONS BIDIMENSIONAL CONFINADO	35
	3.1	Escalas de energia e comprimento característicos	35
	3.2	Modelo para os pontos quânticos	38
II	Р	anorama dos métodos computacionais	41
4.	Cor	NCEITOS DO APRENDIZADO DE MÁQUINA	42
	4.1	Estabelecendo definições de conceitos chave	43
	4.2	Modelos de inteligência artificial	45
	4.3	Descrição das redes neurais diretas	48
	4.4	Elementos de algoritmos de otimização	50
5.	ASP	PECTOS DO MÉTODO VARIACIONAL NA MECÂNICA QUÂNTICA	56
	5.1	Princípio variacional e o método de Ritz	56
	5.2	Minimização do valor esperado da energia	60
	5.3	Método de reconfiguração estocástica	64
	5.4	Método variacional de Monte Carlo	67
6.	Cor	NSTRUÇÃO DOS ESTADOS QUÂNTICOS VARIACIONAIS	69
	6.1	Explorando estados quânticos como inspiração $\ldots \ldots \ldots \ldots \ldots \ldots \ldots$	69
	6.2	Incorporação de simetrias físicas de aglomerados na rede neural $\ \ldots \ \ldots$	72
	6.3	A função tentativa para aglomerados e suas variações $\ldots$ $\ldots$ $\ldots$	76
	6.4	Comentários acerca da função para pontos quânticos	77

Π	I	Apresentação e discussão dos resultados	79
7.	Арі	LICAÇÕES EM AGLOMERADOS DE HÉLIO	80
	7.1	Validando a metodologia empregada	81
	7.2	Cálculo de propriedades e a unitariedade fraca	87
	7.3	Influência das características de entrada	92
8.	Арі	LICAÇÕES EM PONTOS QUÂNTICOS	95
	8.1	Analisando o caráter estocástico do otimizador	96
	8.2	Investigando os resultados e sucesso das redes neurais	98
	8.3	Estabelecendo critérios de acurácia	101
9.	Cor	NCLUSÕES E PERSPECTIVAS	105
Bı	BLIC	OGRAFIA	108
AI	PÊNI	DICE A. INTEGRAÇÃO DE MONTE CARLO E FERRAMENTAS ESTATÍSTI	-
	CAS	3	121
	A.1	Algoritmo de Metropolis	121
	A.2	Estatística por blocos e médias móveis	122
A	PÊNI	DICE B. SOLUÇÃO LIMITE PARA O POTENCIAL DE LENNARD-JONES	124
A	PÊNI	DICE C. ALGORITMO DE RETRO-PROPAGAÇÃO PARA REDES NEURAIS	126
	C.1	A matriz de covariâncias no contexto de <i>machine learning</i>	128
AI	PÊNI	DICE D. RESULTADOS COMPLEMENTARES	130

Em virtude da teoria quântica ser uma descrição em nível atômico e subatômico, a maioria dos fenômenos estudados através desses princípios possuem um caráter microscópico. Entretanto, alguns sistemas macroscópicos apresentam comportamentos quânticos, resultando em fenômenos como a superfluidez e a supercondutividade. Ambos efeitos tem origem no transporte sem dissipação de cargas (no sentido de teoria quântica de campos). No caso da supercondutividade, essa carga é a carga elétrica, enquanto nos superfluidos a carga é a massa. Apesar desses dois comportamentos quânticos serem muito similares, eles diferem fundamentalmente do ponto de vista teórico. No supercondutor tradicional, a origem desse comportamento vem da quebra espontânea de simetria local. Por outro lado, a superfluidez tem origem na quebra global de simetria de *qauqe* do sistema [1]. Isto é, não é a natureza fermiônica ou bosônica que define o fenômeno, ao contrário do que se pode pensar. Por causa da importância desses fenômenos físicos macroscópicos de natureza quântica, a investigação de aglomerados de partículas se torna importante. Entender como sistemas finitos se comportam pode trazer uma compreensão e um entendimento mais fundamental sobre a física envolvida, bem como aprofundar o conhecimento de como esses fenômenos ocorrem em sistemas finitos. De modo especial, aglomerados de hélio e pontos quânticos são dois desses sistemas finitos que exibem comportamentos interessantes.

Apesar da superfluidez em aglomerados de <sup>4</sup>He ser identificada através características diferentes do hélio líquido, é consensual que o mesmo fenômeno é observado [2]. Mesmo assim, o comportamento de impurezas em matrizes de hélio macroscópico e aglomerados é bastante distinto. Átomos e moléculas são expulsas para as fronteiras de amostras macroscópicas de hélio em seu estado super fluído. Contudo, nem sempre esse é o caso para aglomerados. Uma característica aplicável das nano gotas de hélio é sua habilidade de absorver qualquer espécie com as quais elas colidem, tornando-as uma ótima matriz de estudo de moléculas [3]. Dependendo da intensidade da interação entre o dopante e o hélio, o dopante pode ficar no interior da gota ou em sua superfície.

Um outro aspecto desses aglomerados, que satisfazem a estatística de Bose-Einstein, envolve o estudo das propriedades do estado fundamental, o qual exibe comportamentos peculiares devido a combinação das forças interatômicas fracas e a massa leve do hélio. A união dessas características torna o estado fundamental altamente deslocalizado. Além disso, esses sistemas também manifestam correlações universais de curto alcance [4], comportamento intimamente ligado com o alcance e baixa intensidade da interação entre átomos de hélio. Assim, aglomerados de hélio representam um sistema versátil para explorar propriedades quânticas na escala microscópica e suas conexões com fenômenos macroscópicos e universalidades. Com isto em vista, esse sistema foi escolhido para ser objeto de estudo deste trabalho. No Cap. 2 é apresentada uma modelagem de aglomerados de hélio e alguns fenômenos físicos relevantes que surgem nesse contexto.

Por sua vez, pontos quânticos podem ser realizados através de estruturas feitas de material semicondutor de tal maneira que os elétrons desses dispositivos sejam confinados em uma pequena região do espaço. Do ponto de vista de sistemas macroscópicos, o confinamento é feito em todas as dimensões. Essa restrição espacial exerce efeitos únicos para as propriedades óticas e eletrônicas dos pontos quânticos [5]. Da perspectiva de aplicações tecnológicas, isso é bastante interessante na medida em que é possível manipular essas propriedades através do tamanho, da forma e do material que são feitos. Em alguns regimes de densidade, além do confinamento, as correlações elétron-elétron são fortes e impactam na descrição do espectro [6]. Dessa forma, no Cap. 3 é exposto intuições físicas de como modelar os elétrons de um ponto quântico e os comportamentos aproximados que se espera da função de onda dessas cargas. Assim, este trabalho investiga tanto sistemas que satisfazem a estatística de Bose-Einstein, como aqueles que obedecem a estatística de Fermi-Dirac.

Além disso, o desejo de criar vida artificial permeia a humanidade pelo menos desde a Grécia antiga, onde lendas contavam figuras míticas como Hefesto, deus do fogo, dos ferreiros, dos artesãos e da metalurgia. Segundo a mitologia grega, Pandora, a primeira mulher mortal foi forjada por Hefesto, evidenciando as reflexões existentes na época acerca de inteligência humana criada artificialmente. Além disso, esse mito também menciona que características e qualidades como graça e persuasão foram dadas à Pandora pelos deuses. Um paralelo com computadores atuais pode ser traçado, onde um algoritmo, isto é, uma sequência de instruções adaptáveis de acordo com uma finalidade especifica é transmitida para uma máquina com intuito de automatizar um procedimento. As características programáveis dos computadores são como as qualidades ofertadas à Pandora e, de forma similar a Grécia antiga, a aspiração de criar computadores que pensam ainda intriga a sociedade atualmente.

Em particular, o jogo da imitação proposto por Alan Turing [7] visa investigar se máquinas são capazes de pensar. Devido ao caráter filosófico do significado de pensar, a questão é substituída por uma mais prática. Assim, a investigação se resume em descobrir se máquinas são capazes de imitar inteligência humana. Portanto, o teste de Turing envolve três participantes, um humano entrevistador e dois entrevistados, sendo um humano e uma máquina. O entrevistador conversa com ambos entrevistados separadamente por mensagens de texto e deve fazer perguntas com o objetivo de descobrir quem é o humano e quem é a máquina. Por sua vez, a máquina tentará imitar o comportamento humano para enganar o entrevistador. Se máquina conseguir convencer o entrevistador de que ela é o humano, então essa máquina seria capaz de pensar no sentido proposto por Turing.

Com o desenvolvimento dos grandes modelos generativos como o GPT-3, o teste de

Turing está cada vez mais próximo de ser alcançado da maneira como foi construído [8]. Para chegar nesses avanços, as aplicações de inteligência artificial atuais avançaram consideravelmente, resultando em tecnologias como realidade aumentada e virtual [9], veículos autônomos [10], assistentes virtuais [11] e gêmeos digitais [12]. Porém, para alcançar o estágio atual das tecnologias de inteligência artificial, foram necessários alguns importantes desenvolvimentos tanto intelectuais como tecnológicos. Em particular, fundamental para o campo da inteligência artificial, a origem dos computadores programáveis reside no desejo de automatizar cálculos numéricos. Historicamente, o início desse desejo possivelmente começou após Galileo iniciar o processo de formulação matemática das ciências naturais [13]. Isso teve uma influência direta na criação da máquina de adição de Pascal em 1642 e, posteriormente, na elaboração da roda de Leibniz em 1673. O dispositivo de Pascal era capaz de somar e subtrair, enquanto o de Leibniz acrescentava também a capacidade de multiplicar e dividir à máquina de adição. Mais tarde, no século XIX, esses dispositivos ajudaram Charles Babbage no desenvolvimento da máquina diferencial e da máquina analítica, as quais são consideradas precursores dos computadores programáveis. Na mesma época e baseada nos trabalhos de Babbage, Ada Lovelace propôs um método representado por uma sequência de passos para computar os números de Bernoulli utilizando a máquina analítica [14]. Frequentemente, esse método é referenciado como o primeiro algoritmo criado para ser processado por uma máquina, tornando Ada a primeira programadora.

Após o surgimento dos computadores programáveis e até chegar ao estágios atuais das tecnologias de inteligência artificial, um fator importante foi o avanço tecnológico dos computadores, especialmente fenômeno conhecido como lei de Moore [15], o qual enuncia que o número de transistores em um microchip dobra aproximadamente a cada dois anos. Paralelo e suportado por esse avanço, ocorreu o desenvolvimento de uma das principais ferramentas empregadas em inteligência artificial atualmente, o aprendizado profundo, ou também conhecido como deep learning. Porém, para chegar na forma atual dessa ferramenta, houveram três grandes ondas de desenvolvimento [16]. A primeira onda é conhecida como cibernética e teve seu pico de popularidade durante década de 1940 até aproximadamente 1960. Nesse período, os primeiros precursores do aprendizado profundo eram apenas modelos lineares, sendo o neurônio de McCulloch-Pitts [17] um deles. Esse modelo era capaz de reconhecer duas categorias diferentes de informações, entretanto, era necessário ajustar o modelo manualmente para que ele fosse capaz de separar corretamente as categorias. Alguns anos depois, um outro modelo linear conhecido como *perceptron* foi proposto e ele era capaz de aprender como separar as informações nas categorias a partir de exemplos [18]. Atualmente, esse modelo ainda é amplamente usado no contexto de machine learning, isto é, o aprendizado de máquina. Este termo também é uma ferramenta utilizada em aplicações de inteligência artificial e é mais abrangente que o aprendizado profundo, no sentido em que o campo de estudo do *deep learning* está contido no aprendizado de máquina.

A segunda onda de desenvolvimento ficou historicamente conhecida como conexionismo e teve seu auge nas década de 1980 e 1990. Diversos avanços aconteceram nesse período e muitos deles permanecem até hoje como base para construção das redes neurais artificiais atuais. Dentre os progressos dessa época, um modelo chamado de *cognitron* [19] introduziu maneiras não lineares de processar as informações fornecidas ao modelo. Até então, apenas um único neurônio artificial era utilizado para compor os modelos. Entretanto, a ideia central no conexionismo era alcançar comportamentos inteligentes através de conectar unidades computacionais simples entre si, de maneira semelhante aos neurônios biológicos. Por isso, os precursores dos modelos atuais chamados de redes neurais artificiais foram propostos nesse período.

Nesse contexto, surgiu o conceito de representação distribuída [20]. Para entender essa ideia, considere que certo sistema de visão é capaz de reconhecer armários, mesas e cadeiras, os quais podem ser feitos de madeira, metal ou plástico. Uma forma de organizar os neurônios do modelo seria atribuir um para cada combinação possível de objeto e material, resultando em nove neurônios. Outra possibilidade seria dividir em duas camadas de conceitos, onde uma camada diferenciaria o objeto e a outra o material. Conectando densamente essas camadas, seriam necessários somente três neurônios em cada camada, ou seja, um total de seis. Essa seria uma representação distribuída do modelo, o que reduz a quantidade total de neurônios necessários. Esse tipo de representação possibilita aos neurônios entender conceitos complexos em termos de uma hierarquia de conceitos mais simples. Ainda nesse período, uma importante contribuição foi o êxito no emprego do algoritmo de retro propagação, ou *back-propagation* [21], para treinar os modelos de *deep learning* e a popularização desse procedimento.

A partir de 2006, a terceira onda se inicia com um trabalho que mostrou a possibilidade de um modelo específico ser treinado eficientemente com uma estratégia chamada de pré-treinamento ganancioso camada a camada [22]. Logo na sequência, outros trabalhos mostraram que a mesma estratégia poderia ser usada para treinar outros tipos de modelos e torná-los melhores sistematicamente [23]. Foi nessa onda de trabalhos que o termo *deep learning* se popularizou. Eventualmente, o avanço dessas técnicas nas diferentes áreas do conhecimento atingiu à física nos seus mais diversos domínios, em especial, na física da matéria condensada [24].

Embora o uso extensivo das ferramentas de *machine learning* aplicadas à física seja recente, há fertilização cruzada entre a física estatística e o aprendizado de máquina desde a onda do conexionismo [25]. Um exemplo interessante é a transição de fase de primeira ordem que ocorre em *perceptrons* ao serem treinados por exemplos [26]. Conforme o número de exemplos aumenta, o modelo atinge o estado de referência. Outra conexão notável vem do fato de um modelo de aprendizado de máquina, chamado máquina restrita de Boltzmann, ser conhecido como o problema do modelo de Ising inverso [27] na literatura de física. Nesse problema, a ideia é encontrar valores para os acoplamentos entre *spins* a partir de um conjunto de configurações de Ising. Junto a isso, a recente ampla adoção e integração das ferramentas de inteligencia artificial na pesquisa científica promoveu uma considerável evolução sobre diversos domínios, em particular na física. Muitas aplicações demonstraram relevância considerável, abrangendo desde estudos de cosmologia [28] e física de partículas [29] até sistemas de muitos corpos quânticos [30], química quântica [31], mecânica estatística [32] e materiais [33].

A integração das redes neurais no estudo da física é em certa medida natural. Isso porque redes neurais são construídas com o propósito de reconhecer padrões em dados e adquirir conhecimento através desse processo. Da mesma forma, reconhecer padrões nos fenômenos físicos e descrevê-los em termos matemáticos permeia as ciências naturais possivelmente desde as primeiras civilizações que estudavam a astronomia. Portanto, empregar essas ferramentas de inteligência artificial para auxiliar o desenvolvimento de conhecimento é em certo aspecto imediato. Especialmente no contexto da física quântica, uma sinergia notável entre o método variacional e algoritmos de *deep learning* se manifesta através de um paralelo entre os dois métodos. Por exemplo, a função de onda tentativa do método variacional é uma função auxiliar cujo propósito é correspondente ao papel exercido pela rede neural nos algoritmos de *deep learning*.

Historicamente, funções variacionais já demonstraram serem extremamente efetivas para explicar alguns fenômenos como supercondutividade [34], líquidos de *spin* [35] e efeito Hall quântico fracionário [36]. Consequentemente, explorar redes neurais para construir funções tentativa é promissor por dois motivos. O primeiro deles se sustenta na afirmação de que redes neurais são capazes de representar uma classe bastante abrangente de funções [37]. O segundo motivo reside no sucesso já demonstrado de estados variacionais que empregam redes neurais em sua construção, isso para a investigação de sistemas quânticos abertos [38], magnetos quânticos frustrados [39], cadeias de *spin* [40], estimadores de forças inter atômicas [41], entre outros [31, 42].

O Cap. 4 aborda os conceitos do aprendizado de máquina com o intuito de formar uma base conceitual e aplicá-la em problemas de muitos corpos quânticos posteriormente. Além disso, o Cap. 5 apresenta o método variacional, técnicas de otimização e também explora outros aspectos da sinergia existente com algoritmos de aprendizado de máquina. A técnica empregada para computar estimativas das integrais envolvidas no método variacional e no processo de otimização foi a integração de Monte Carlo. Já para combinar as redes neurais com o método variacional, as funções tentativa empregadas para estudar os sistemas apresentados nos Cap. 2 e Cap. 3 são construídas no Cap. 6 utilizando redes neurais em sua construção. Assim, a forma dessas funções tentativas é intimamente ligada e dependente de redes neurais artificiais, estabelecendo a ponte que conecta a física com a inteligência artificial.

O objetivo deste trabalho foi verificar a capacidade de redes neurais artificiais representarem estados quânticos de sistemas de muitos corpos, especificamente funções de onda de aglomerados de hélio e de pontos quânticos em seus estados fundamentais. Os Cap. 7 e Cap. 8 apresentam resultados robustos que demonstram o poder de representação das funções tentativa baseadas em redes neurais. Adicionalmente, a capacidade de generalização dessas funções também é demonstrada através do estudo de propriedades que não comutam com o hamiltoniano e de uma análise dos resultados obtidos através das funções tentativas em função do tamanho das redes neurais. Por fim, as conclusões e perspectivas desse estudo são incluídas no Cap. 9.

# Parte I

# Considerações sobre os sistemas de interesse

# 2 CARACTERÍSTICAS DE AGLOMERADOS DE HÉLIO

Um dos fatores que geram interesse no estudo de sistemas de hélio na matéria condensada é a riqueza única desse elemento químico. O hélio é o menor átomo que possui todas as camadas eletrônicas com ocupação totalmente preenchidas quando neutro. Isso o torna um elemento praticamente inerte numa ampla faixa de temperatura e pressão. Na natureza ele é encontrado na forma de dois isótopos, o <sup>4</sup>He com dois neutrons e o <sup>3</sup>He, mais leve, com apenas um neutron. Em particular, o hélio-4 tem spin inteiro e, portanto, é um bóson e, por sua vez, o hélio-3 é um férmion. Apesar de terem massa e estatísticas de *spin* distintas, o potencial de interação entre pares de átomos de hélio é independente da combinação de isótopos envolvidos. Mesmo assim, as diferenças apontadas geram fenômenos físicos distintos. Essa flexibilidade atrai o interesse de teóricos e experimentais em diversas aplicações. Os estudos se estendem desde aglomerados sem e com impurezas [2, 43], sistemas em dimensões reduzidas, tanto 1D [44] quanto filmes em 2D [45, 46], até situações no limite termodinâmico, abrangendo efeitos como superfluidez [47] e estados metaestáveis [48].

#### 2.1 Propriedades de aglomerados ou *clusters* de hélio

Um ponto peculiar dos átomos de hélio é que ambos isótopos podem existir tanto no limite termodinâmico na forma de líquidos, quanto em estados autoligados, também conhecidos como *clusters* ou aglomerados atômicos. Em especial, o hélio 3 é o único férmion neutro conhecido que tem essa dupla característica [3], entretanto, é necessário um número mínimo de átomos para que a formação de um estado auto ligado seja possível. Cálculos teóricos indicam que esse número mínimo não é maior que  $N_{\rm min} = 32$  [49]. Por outro lado, o hélio 4 já forma estado ligado a partir do dímero. Além disso, estudos com partículas hipotéticas de <sup>4</sup>He fermiônico e <sup>3</sup>He bosônico sugerem que esses compostos seriam capazes de formar aglomerados fracamente ligados a partir de ao menos oito átomos [50]. Isso indicaria que, tanto a massa mais leve, quanto a estatística são importantes na determinação de  $N_{\rm min}$  e para formação desses *clusters*, sendo que a massa mais leve está diretamente associada com um valor mais alto de momento de ponto zero.

#### Processo de produção de aglomerados de hélio

As gotas de hélio são geradas através de uma expansão adiabática de gás hélio ultra purificado, pressurizado na faixa de 20 à 60 bars e previamente resfriado até por volta de 12 à 22 Kelvin [51]. Armazenado nessas condições, o gás passa por um bocal de 5  $\mu m$  e vai para uma câmara de vácuo, onde a expansão ocorre e, durante esse processo, as gotas são formadas. Após a formação das gotas, elas são colimadas em direção à câmara de coleta, onde elas perdem temperatura rapidamente através de resfriamento por evaporação, até chegarem na temperatura de equilíbrio de 0.37 K para o <sup>4</sup>He ou de 0.15 K para o <sup>3</sup>He. Esse processo ocorre eficientemente devido à fraca coesão entre os átomos do aglomerado. O arranjo experimental para produzir *clusters* de hélio é esquematizado na Fig. 2.1.



Figura 2.1: Representação esquemática do conjunto experimental usado para produção de aglomerados de hélio. Figura adaptada de [51].

Para sistemas macroscópicos, o que difere um líquido comum de um superfluido é a baixíssima viscosidade deste. Do ponto de vista prático, o superfluido é definido por como ele responde à movimentos lentos de suas bordas. Em particular, o experimento de Andronikashvili [52] investiga o comportamento da circulação do líquido sob ação de um tipo de pendulo de torção. Nele, um rotor com diversos discos finos e proximamente espaçados é colocado dentro de um recipiente com hélio líquido e girado suavemente. A partir do momento de inércia do rotor, a quantidade de hélio que é arrastado e passa a se movimentar com os discos é observada. Ao resfriar o sistema, o hélio líquido sofre uma transição de fase e passa a se comportar como um superfluido. A temperatura na qual a transição ocorre é chamada de ponto lambda, que para o hélio 4 é de aproximadamente 2.7 K. Abaixo do ponto lambda, verifica-se que cada vez menos átomos de hélio se movem junto com o rotor, sobretudo experimentos recentes indicam que apenas poucas camadas de hélio são arrastadas com os discos [53].

Contudo, no que diz respeito à sistemas microscópicos, a definição do fenômeno da superfluidez torna-se imprecisa, uma vez que viscosidade é uma propriedade essencialmente macroscópica. Ainda assim, o experimento de Andronikashvili pode ser estendido para compreender sistemas microscópicos. Uma representação comparativa das duas versões do experimento são mostradas na Fig. 2.2. Em especial, a versão microscópica desse experimento foi primeiramente conduzido por Grebenev em 1998 [54]. Nesse experimento, uma única molécula de sulfeto de carbonila (OCS) foi adicionada tanto em gotas de hélio 4 puro, como também em gotas de hélio 3 puro, ambas de mesmo tamanho. Esse tipo de dopagem é feita na câmara de coleta como esquematizado na Fig. 2.1. Posteriormente, as gotas dopadas passam para a câmara de espectroscopia onde o espectro rotacional infra vermelho dos dopantes é medido. Quando o dopante está na presença de hélio 4, as linhas



Figura 2.2: Representação comparativa dos experimentos de Andronikashvili "clássico" (a) e microscópico (b). Figura adaptada de [55]

observadas são estreitas e bem definidas, enquanto na presença de hélio 3 apenas um pico largo e esparso foi encontrado. No caso do hélio 4, o subsistema da molécula experimenta rotações livres e apresenta transições rotacionais bem resolvidas, como se o rotor estivesse girando e o hélio não, semelhantemente ao caso original. Fenomenologicamente, a molécula age como se fosse o rotor do experimento macroscópico e seu movimento é desacoplado do *cluster*. Uma vez que esse comportamento não aparece no caso do hélio 3, essa diferença nas observações indica a manifestação microscópica da superfluidez.

#### 2.2 Apresentação e análise dos potenciais de interação

No estudo de sistemas de muitos corpos, um fator fundamental são as interações entre as partículas. Modelar matematicamente de maneira acurada funções para representar interações reais é um desafio antigo na física. Mesmo interações entre gases nobres são complicadas de se modelar. Em particular, uma tentativa bem sucedida e amplamente estudada foi o potencial de Lennard-Jones [56]. Sua forma paramétrica é representada por:

$$V_{\rm LJ}(r) = 4\epsilon \left[ \left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right] , \qquad (2.1)$$

onde  $\sigma$  é uma escala de distância e  $\epsilon$  uma escala de energia. Ambas escalas dependem do par de objetos que está interagindo. Esse modelo de interação reflete características essenciais de interações de átomos e moléculas simples. Isso pode ser visto pelos dois termos da Eq. 2.1, onde o termo proporcional ao inverso da décima segunda potencia da distância é um termo repulsivo que atua principalmente em curtas distâncias e é proveniente da repulsão das nuvens eletrônicas dos objetos interagentes. Por outro lado, o termo negativo da mesma equação é proveniente de interações do tipo dipolo-dipolo, também conhecidas como as forças de Van der Waals do sistema, isto é, é um termo atrativo que atua principalmente à distâncias moderadas [57]. Em particular, para interações hélio-hélio, a escala de energia vale  $\epsilon = 10.22$  K, enquanto a de distância vale  $\sigma = 2.556$  Å [58]. Note que a constante de energia foi dividida pela constante de Boltzmann e, por isso, está em unidades de Kelvin. Entretanto, nem todos os resultados usando o potencial de Lennard-Jones são rigorosos o suficiente atualmente. Para cálculos de seção de choque de espalhamento, coeficientes da expansão do virial e propriedades de transporte mais precisos, já é possível utilizar potenciais mais acurados [59]. De fato, o potencial de Lennard-Jones se revela excessivamente repulsivo em seu termo  $r^{-12}$  e com o mínimo de energia razoavelmente maior do que outros modelos. Em particular, com o objetivo de alcançar resultados mais precisos e para caracterizar a interação entre átomos de hélio em uma larga escala de temperaturas, foi construído o potencial HFD-B(HE) [60], também conhecido como potencial Aziz-87 e sua forma é dada por:

$$V_{A87}(x) = \epsilon \left[ A^* \exp(-\alpha^* x + \beta^* x^2) - F(x) \sum_{j=0}^2 \frac{c_{2j+6}}{x^{2j+6}} \right] , \qquad (2.2)$$

onde  $x = r/r_m$  é a distância em unidades adimensionais, a função F(x) tem a seguinte expressão :

$$F(x) = \begin{cases} \exp\left[-\left(\frac{D}{x}-1\right)^2\right], & x < D, \\ 1, & x \ge D, \end{cases}$$
(2.3)

e as constantes envolvidas estão apresentadas na Tab. 2.1. Novamente, a constante de energia está em Kelvin, isto é, foi dividida pela constante de Boltzmann.

Tabela 2.1: Parâmetros para o potencial HFD-B(HE).	Os dígitos mostrados são suficientes para não causar
erros de arredondamento [60].	

$\epsilon/K$	10.948
A*	$1.8443101 \cdot 10^5$
$\alpha^*$	10.43329537
$\beta^*$	-2.27965105
D	1.4826
$c_6$	1.36745214
$c_8$	0.42123807
<i>c</i> <sub>10</sub>	0.17473318
$r_m/Å$	2.963

Esse potencial é consistente com cálculos *ab initio* tanto para curto quanto para longo alcance, bem como consistente com medidas de coeficientes da expansão virial em uma larga faixa de temperaturas e com medidas de propriedades de transporte. Também descreve bem medidas de seção de choque de espalhamento diferencial e integral [60].

Para comparação, ambos os potenciais de Lennard-Jones e Aziz são mostrados na Fig. 2.3, onde o comportamento em função da distância pode ser comparado visualmente. Para longas distâncias, ambos os potenciais se comportam semelhantemente, porém, as diferenças começam a aparecer a partir de distâncias moderadas. Como já pontuado, o mínimo do potencial de Aziz é inferior ao mínimo do potencial Lennard-Jones, resultando em interações mais fortes nessa região ao usar potenciais mais realistas como HFD-B. A imagem não consegue manifestar a repulsão excessiva do potencial Lennard-Jones, entretanto, é simples mostrar que, para curtas distâncias, este diverge com  $r^{-12}$ , enquanto o potencial Aziz-87 tende para uma constante.



Figura 2.3: Comparação entre os potenciais de Lennard-Jones (vermelho) e HFD-B (azul) em função da distância para interações entre átomos de hélio. A unidade de distância está em Angstrom e os potenciais em unidades de Kelvin.

# 2.3 Intuições físicas do comportamento da função de onda

Além de um potencial acurado, como o Aziz-87, estudar as propriedades quânticas de *clusters* de hélio requer também funções de onda acuradas. A análise dessas funções para diferentes tamanhos de aglomerados permite identificar características de efeitos de correlação para diferentes regimes de tamanho desses *clusters*. Além disso, é possível determinar suas estruturas geométricas e estados excitados. Dessa forma, nesta seção serão apresentadas algumas propostas de funções tentativa baseadas em intuições físicas sobre os aglomerados de hélio.

Uma forma geral de função de onda tentativa muito usada para estados com simetria par de troca de partículas é o fator generalizado de Jastrow usando funções de correlação até de três corpos da forma:

$$\Psi(\mathbf{R}) = f_1(\mathbf{R}) f_2(\mathbf{R}) f_3(\mathbf{R}) , \qquad (2.4)$$

onde  $\mathbf{R} = {\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N}$  é o conjunto de vetores posição das N partículas que compõem o

sistema e  $f_i$  são termos de 1, 2 e 3 corpos.

Para que as partículas se mantenham ligadas, imagine que elas estão sob o efeito de um campo harmônico médio, ou seja, uma função tipo gaussiana ligando-as à origem do sistema de coordenadas. Por consequência dos corpos estarem ligados entre si, e não exatamente à origem, esse campo médio pode ser pensado como uma componente de correlação associada ao fragmento de (N-1) átomos interagindo com o átomo restante. Desse ponto de vista, o termo  $f_1$  pode ser escrito como:

$$f_1(\mathbf{R}) = \exp\left[-\frac{1}{2}\sum_i \left(a_0 |\mathbf{r}_i - \mathbf{r}_{\rm cm}| + b_0 |\mathbf{r}_i - \mathbf{r}_{\rm cm}|^2\right)\right] , \qquad (2.5)$$

onde  $\mathbf{r}_{cm}$  é o vetor centro de massa,  $a_0 \in b_0$  são parâmetros variacionais. Uma segunda possibilidade seria substituir o  $\mathbf{r}_{cm}$  pelo centro de massa do fragmento de (N-1) átomos.

Para as correlações de dois corpos, a função de onda deve incluir os efeitos de correlação de curto e longo alcance. Para distâncias curtas, a função é dominada por efeitos repulsivos entre as partículas. Um termo que reflete muito bem esse comportamento foi sugerido por McMillan em 1965 para o estudo de líquidos de hélio e obteve resultados de energias de coesão em razoável acordo com dados experimentais [58]. Essa função pode ser expressa da seguinte maneira:

$$f_{\rm mcm} = \exp\left[-\frac{1}{2}\sum_{i$$

onde *b* e *m* são parâmetros variacionais, enquanto  $r_{ij}$  denota as distâncias relativas entre as partículas, isto é,  $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ . Como *b* e *m* são usualmente positivos, repare que para curtas separações entre qualquer par de partículas, o argumento da exponencial e altamente negativo e, portanto, o termo  $f_{\rm mcm}$  se anula rapidamente. Em particular, essa função é a solução da equação de Schrodinger de um corpo para o potencial de Lennard-Jones no limite em que as distancias são pequenas, ou seja,  $r \ll \sigma$  (veja o Ap. B). Entretanto, utilizando somente esse termo, algumas propriedades não se comportam devidamente, como por exemplo o fator de estrutura e a distribuição de momentos para pequenos valores de vetor de onda. Isso ocorre principalmente por conta dessa expressão não considerar correlações de longo alcance, como efeitos vibracionais. Em 1966, Chester e Reatto propõem uma extensão da função  $f_{\rm mcm}$  para incluir essas considerações [61], resultando no termo adaptado para *clusters* como se segue:

$$f_{\rm CR} = \exp\left[-\frac{1}{2}\sum_{i< j}\frac{\alpha_0}{r_{ij}^2}\right] , \qquad (2.7)$$

onde  $\alpha_0$  é um parâmetro variacional. Além disso, no estudo de aglomerados, em particular, existe uma correlação de longo alcance associada com uma partícula que é removida do *cluster* e levada para longe dele. Então, esse corpo pode ser representado por uma onda circular espalhada no infinito, isto é, representada por um decaimento do tipo  $\exp(ikr')/(kr')$ e que dependente da distância r' dessa partícula afastada com relação ao centro de massa das demais. Esse termo também será considerado em  $f_2$ , uma vez que esse comportamento assintótico pode ser representado através das distâncias relativas entre as partículas da seguinte maneira:

$$f_{2L'} = \exp\left[-\frac{1}{N-1}\sum_{i< j} \left(\alpha_1 r_{ij} + \alpha_2 \ln r_{ij}\right)\right] , \qquad (2.8)$$

onde o momento k foi substituído por dois parâmetros variacionais  $\alpha_1 \in \alpha_2$ . Isso confere à função de onda tentativa mais flexibilidade, aumentando o espaço de funções que ela pode representar.

Por fim, em 1979, O'Reilly e Tsang chamam atenção para a existência de regiões de alta densidade e baixa densidade em líquidos de hélio, além de um ordenamento espacial no estado superfluido [62]. Esse ordenamento é influenciado pelo relaxamento dos átomos de hélio em torno de vacâncias juntamente com a alta amplitude vibracional de ponto zero, criando qualitativamente uma modulação em termos de ondas de densidade de átomos num líquido no zero absoluto de temperatura. Essas proposições levaram à construção do seguinte termo:

$$f_{2L''} = \exp\left[-\frac{1}{2} \alpha_3 \sum_{i < j} \frac{\cos(k_0 r_{ij})}{k_0 r_{ij}}\right] , \qquad (2.9)$$

onde, igualmente,  $k_0 \in \alpha_3$  serão parâmetros variacionais. Juntando todas as correlações de dois corpos, a função  $f_2$  terá a seguinte expressão:

$$f_2(\mathbf{R}) = \exp\left\{-\frac{1}{2}\sum_{i< j} \left[\left(\frac{b}{r_{ij}}\right)^m + \frac{\alpha_0}{r_{ij}^2} + \alpha_1 r_{ij} + \alpha_2 \ln r_{ij} + \alpha_3 \frac{\cos(k_0 r_{ij})}{k_0 r_{ij}}\right]\right\} .$$
 (2.10)

Correlações de ordem superiores são importantes em sistemas como hélio líquido e sólido, onde correlações de três corpos correspondem à aproximadamente 10% da energia de ligação. Entretanto, para aglomerados que possuem entre três até por volta de uma dezena de partículas, as correlações de dois corpos dominam a energia de ligação que as mantém juntas, correspondendo entre 97% até 99% da energia de ligação [63]. Por consequência, é uma boa aproximação considerar a funções de correlação de três corpos como unitária. Portanto, para os *clusters* considerados pequenos, segue que  $f_3(\mathbf{R}) = 1$ . Dessa forma, esses são os principais comportamentos que uma função tentativa deve capturar e, portanto, que a rede neural deve aprender.

## 2.4 Propriedades universais e correlações de curto alcance

Diferentemente da mecânica clássica, um potencial atrativo nem sempre é capaz de formar estados ligados entre duas partículas na mecânica quântica. Isso tem origem na energia de ponto zero, isto é, nas flutuações quânticas da energia cinética no estado fundamental, que tendem a afastar as partículas através da movimentação, enquanto a parte atrativa do potencial de interação procura ligá-las. Devido a essa competição, um potencial atrativo pode ser fraco demais para formar estados ligados entre um par de partículas. Se a interação é intensa o suficiente para cancelar esse efeito da energia cinética e suportar um estado fracamente ligado, essa interação é dita ressonante, já que durante um processo de espalhamento à baixas energias esse par de partículas ficam muito perto de formarem um estado ligado durante a colisão [64]. Isso é caracterizado por um comprimento de espalhamento  $a_s$  de onda s muito maior que o alcance  $r_0$  do potencial.

Especificamente, o comprimento de espalhamento pode ser extraído da teoria de espalhamento [65] através do comportamento da função de onda para distâncias muito maiores que o alcance do potencial. Esse comportamento é dado por

$$\psi \propto \frac{1}{r} - \frac{1}{a_s} , \qquad (2.11)$$

onde r é a distância relativa do par de partículas. Entretanto, geralmente  $a_s$  e  $r_0$  tem a mesma ordem de grandeza e apenas em casos excepcionais o comprimento de espalhamento é muito maior que o alcance. Particularmente, a partir de potenciais de interação entre pares de átomos de hélio, o comprimento de espalhamento calculado é de  $a_s = 189 a_0$ , enquanto o alcance efetivo é de  $r_s = 14 a_0$  [66], onde  $a_0$  é o raio de Bohr. Além disso, em experimentos com gases ultra frios é possível ajustar a interação através de campos magnéticos, onde exatamente na ressonância o comprimento de espalhamento diverge [67].

Para espalhamentos entre duas partículas interagindo por um potencial central, a energia do sistema é conservada e há apenas transferência de momento. Dessa forma, observáveis como a seção de choque de espalhamento são determinadas essencialmente pelo comprimento de espalhamento. Por outro lado, se uma terceira partícula está perto do par que está colidindo, podem ocorrer processos de transferência de energia do par para a terceira partícula, de maneira que o comportamento da função de onda de dois corpos para distâncias da ordem do alcance do potencial influenciam no espalhamento [68]. Porém, se a interação é ressonante, o comprimento de espalhamento deixa de ser um comprimento característico do sistema, os detalhes da interação deixam de ser importantes. Quando observáveis físicos são insensitivos ao alcance e outros detalhes da interação, eles são ditos possuírem comportamentos universais. Efeitos interessantes da física de trímeros nessa situação foram estudados por Efimov [69]. Para potenciais como Lennard-Jones e Aziz, a interação possui um termo atrativo e um termo repulsivo. No caso de interações entre átomos de hélio, o comprimento de espalhamento é razoavelmente maior que o alcance efetivo potencial, entretanto, a distância interatômica média em aglomerados de hélio é da mesma ordem de grandeza que o alcance da calda de Van der Waals, isto é, o termo atrativo. Nesse caso, não é esperado uma universalidade forte. Porém, devido à intensa interação de repulsão a curtas distâncias, pares de partículas próximos praticamente não são influenciados por outras partículas ao redor [4]. Isso coloca os aglomerados de hélio em uma posição intermediária, na qual é esperado que haja comportamentos universais apenas no regime de curto alcance. Essa situação é chamada de unitariedade fraca.

Para enxergar essa característica de pares próximos não serem influenciados por outras partículas ao redor, primeiramente considere um sistema formado por férmions, cujo o alcance da interação é muito menor que outros comprimentos característicos. Estudos anteriores mostraram que esses sistemas podem ser analisados pela aproximação de alcance nulo. Isso significa que os detalhes da interação podem ser negligenciados, surgem características universais no sistema e elas podem ser descritas em termos de um pequeno conjunto de parâmetros, como o comprimento de espalhamento. Essa é uma aproximação plausível pois, em suficientemente baixas energias, o momento relativo  $p = \hbar k$  de dois átomos também é pequeno e, consequentemente, o comprimento de onda de de Broglie  $\lambda_B = \frac{2\pi\hbar}{p}$  é grande. Dessa maneira, o valor de  $\lambda_B$  é maior que a extensão espacial dos átomos, o que impossibilita resolver a estrutura interna do par interagente [66]. Em particular, várias propriedades são governadas apenas pelo parâmetro conhecido como contato C quando a interação é ressonante, o qual tem origem nas relações de Tan [70–72].

Várias relações podem ser definidas com o contato. Para pares de partículas, o contato de dois corpos é uma medida da probabilidade de um par de partículas estarem próximas. Para ver isso, considere que a função de onda para um sistema de N partículas pode ser fatorada em um produto de uma função universal de dois corpos  $\phi_2$  e uma função de estado dependente  $A_2^{(N)}$  que descreve o sistema residual, quando um par de partículas *ij* está próximo. Novamente, o curto alcance da interação possibilita essa aproximação, resultando na expressão [70]

$$\Psi(\mathbf{r}_1,...,\mathbf{r}_N) \xrightarrow[r_{ij}\to 0]{} \phi_2(r_{ij}) A_2^{(N)}(\mathbf{r}_{\mathrm{cm}}^{(ij)},\mathbf{r}_{k\neq\{i,j\}}) , \qquad (2.12)$$

onde  $\mathbf{r}_{cm}^{(ij)} = \mathbf{r}_i + \mathbf{r}_j/2$  é o centro de massa do par. Na aproximação de alcance nulo, a função de dois corpos  $\phi_2$  assume a forma da Eq. 2.11. Então, o contato é definido como [4]

$$C_2^{(N)} = \binom{N}{2} \left\langle A_2^{(N)} \middle| A_2^{(N)} \right\rangle , \qquad (2.13)$$

onde o primeiro fator é um coeficiente binomial. Usando essas expressões, é possível conectar uma propriedade para um sistema de N corpos com o dímero. Dessa maneira, é interessante utilizar os contatos para estudar as correlações de curto alcance em aglomerados de hélio-4, já que é esperado que esses sistemas apresentem comportamentos universais de curto alcance.

Por exemplo, para conectar propriedades de um aglomerado composto por N átomos com uma propriedade do dímero, examine o seguinte operador:

$$\hat{\rho}_N(r) = \sum_{i < j}^N \frac{\delta(r - r_{ij})}{r^2} .$$
(2.14)

O valor esperado desse operador define a chamada função densidade de pares  $\rho_N(r)$  para um sistema composto de N partículas [4]. Devido à definição do operador, a integral em distâncias r dessa função resultará no número de pares presentes no sistema, isto é, a densidade é normalizada tal que  $\int \rho_N(r)r^2 dr = \frac{N(N-1)}{2}$ . Quando um sistema está no regime da unitariedade fraca, é esperado que o comportamento da função densidade de pares para pequenas distâncias r seja semelhante nos diferentes tamanhos de *clusters* [73]. Especificamente, a função densidade de N partículas deve ser proporcional à função densidade do dímero, isto é

$$\rho_N(r) \xrightarrow[r \ll r_0]{} C_2^{(N)} \rho_2(r) .$$
(2.15)

Devido ao princípio de exclusão de Pauli, férmions idênticos são impedidos de se aproximarem em estados relativos de onda s. Isso faz com que interações de três corpos sejam negligíveis em misturas com férmions de duas componentes [4]. Contudo, é esperado que isso seja relevante em sistemas compostos por bósons. Em particular, como já mencionado, o comprimento de espalhamento é mais de dez vezes o alcance efetivo da interação [66] para a interação entre átomos de hélio, tornando esse sistema um ótimo candidato para observar os efeitos da coalescência de três corpos em um sistema que exibe características de sistemas com comportamentos universais.

Um modo diferente de olhar as correlações de pares de partículas consiste no operador dado por

$$\hat{P}(r) = \sum_{i < j} \delta(r - r_{ij}) , \qquad (2.16)$$

onde a função distribuição de pares P(r) é normalizada tal que  $\int P(r) dr = 1$ . Outro operador de interesse é expresso por

$$\hat{n}(r) = \sum_{i} \frac{\delta(q_i - r)}{r^2} ,$$
(2.17)

onde  $q_i = |\mathbf{r}_i - \mathbf{r}_{cm}|$  é a distância ao centro de massa. Nesse caso, o valor esperado desse operador resulta no perfil de densidade n(r) em relação ao centro de massa do aglomerado. Por definição, a integral sob a variável r terá como resultado o número total de partículas, ou seja, o perfil de densidade é normalizado tal que  $\int r^2 n(r) dr = N$ . Tanto P(r) quanto n(r) são utilizados para estudar propriedades estruturais dos aglomerados de hélio.

Já para correlações de três corpos, a função que deve ser monitorada é a distribuição de Jacobi de três corpos  $P_N^{(3,\text{jacobi})}$  [4], a qual é computada através do operador

$$\hat{P}_{N}^{(3,\text{jacobi})}(\rho) = \frac{2}{N(N-1)(N-2)} \sum_{j=1}^{N-2} \sum_{k>j}^{N-1} \sum_{l>k}^{N} \left[ \frac{\delta(\rho_{jk,l}-\rho)}{\rho^2} + \frac{\delta(\rho_{jl,k}-\rho)}{\rho^2} + \frac{\delta(\rho_{kl,j}-\rho)}{\rho^2} \right] ,$$
(2.18)

onde a distribuição de Jacobi é normalizada tal que  $\int \rho^2 P_N^{(3,\text{jacobi})}(\rho) d\rho = 1$  e as distâncias  $\rho_{jk,l}$  correspondem à distância da partícula l ao centro de massa do par jk, isto é,

$$\rho_{jk,l} = \left| \mathbf{r}_l - \frac{\mathbf{r}_j + \mathbf{r}_k}{2} \right| . \tag{2.19}$$

Assim, a quantidade  $\rho^2 P_N^{(3,\text{jacobi})}(\rho)$  é uma medida da probabilidade de encontrar uma partícula a uma distância  $\rho$  do centro de massa de um par de outras duas partículas. Analogamente ao contato  $C_2^{(N)}$  de dois corpos, a definição do contato par-átomo  $C_{2+1}^{(N)}$  para sistemas compostos por  $N \leq 3$  é dada através da expressão

$$P_N^{(3,\text{jacobi})}(r) \xrightarrow[r \ll r_0]{} C_{2+1}^{(N)} P_3^{(3,\text{jacobi})}(r) .$$

$$(2.20)$$

Portanto, através dessas propriedades, a unitariedade fraca presente em sistemas de aglomerados de hélio pode ser investigada. O estudo de propriedades estruturais, como a distância média entre átomos, a função de distribuição radial e correlações de dois corpos, entre outras características destes sistemas permite a exploração de conceitos que não são apenas válidos para aglomerados de átomos de hélio, mas são de aplicação geral para outros sistemas que exibem comprimentos de espalhamento grandes e fracas energias de ligação. Assim, os aglomerados com mais de dois átomos de <sup>4</sup>He constituem-se, portanto, num sistema modelo paradigmático para o estudo de conceitos fundamentais da mecânica quântica e da física da matéria condensada.

## 3 Atributos do gás de elétrons bidimensional confinado

Em diversas situações, a natureza apresenta estruturas nas quais os elétrons estão confinados, como por exemplo em átomos, onde essas partículas ficam presas ao núcleo através de forças coulombianas atrativas. Outros sistemas como *clusters* metálicos e gases atômicos aprisionados também compõem essa categoria de sistemas que definiram paradigmas no contexto do estudo da física de muitos corpos [74,75]. Em particular, os avanços tecnológicos permitiram a fabricação de dispositivos nanométricos chamados pontos quânticos ou *quantum dots*, os quais são capazes de confinar elétrons com alto grau de controle e ajuste de propriedades [76]. Seu sucesso e ampla gama de aplicações tecnológicas conduziu à concessão do prêmio Nobel de química de 2023 à Moungi G. Bawendi, Louis E. Brus e Aleksey Yekimov pela descoberta e desenvolvimento desses dispositivos. Pontos quânticos também compartilham diversas características com os sistemas confinados existentes na natureza. Adicionalmente, por serem manufaturados cuidadosamente, explorar essas analogias de maneira controlada torna-se uma ótima opção para se estudar aspectos pertinentes da mecânica quântica.

Além disso, as aplicações são múltiplas, impactando desde a pesquisa científica até dispositivos eletrônicos usados no dia a dia. As propriedades eletrônicas e ópticas características desses objetos nanométricos são exploradas no contexto da informação quântica [77], na implementação de q-bits [78], além de também estarem presentes em aplicações médicas [79], dispositivos fotovoltaicos [80] e em aparelhos de televisão na forma de QLEDs [81]. De modo especial, um atributo de interesse reside na sensibilidade das propriedades ao tamanho físico do quantum dot. Por exemplo, a emissão e absorção óptica de pontos feitos de seleneto de cadimo (CdSe) podem ser ajustados por todo o intervalo visível do espectro óptico dependendo do seu tamanho. Outra questão importante é o confinamento dos elétrons, em um átomo eles são ligados ao núcleo pelo potencial de Coulomb, enquanto os elétrons em pontos quânticos semicondutores são confinados por potenciais eletrostáticos artificialmente criados. Detalhes associados a essas questões são explorados nas próximas seções com intuito de chegar em um modelo para descrever esses sistemas.

#### 3.1 Escalas de energia e comprimento característicos

Apesar de existirem diversos tipos de pontos quânticos, neste trabalho foi explorado aqueles cuja a construção é baseada em semicondutores, como já brevemente indicado. Uma representação esquemática da montagem experimental é apresentada na Fig. 3.1.



Figura 3.1: Representação de um ponto quântico circular de dupla heteroestrutura [82].

Com o objetivo de construir um modelo hamiltoniano para descrever os elétrons presentes nesses pontos quânticos, inicialmente uma análise qualitativa utilizando um potencial de confinamento simplificado é apresentada. Na sequência, considerações sobre um potencial mais condizente com os pontos quânticos circulares são apresentadas.

Tipicamente, em um sólido composto de um material semicondutor e especificamente no seu estado fundamental, os elétrons ocupam sua banda de valência, enquanto a banda de condução fica vazia. Entre as duas bandas existe uma região de energia inacessível ao sistema chamada de *gap*. Excitações térmicas podem levar elétrons da banda de valência para a banda de condução, situação na qual a teoria quântica de sólidos explica que o comportamento destes elétrons é aproximadamente equivalente ao de elétrons livres [83]. Entretanto, pela influência da rede cristalina, o comportamento é modificado de tal forma que os elétrons agem como se tivessem uma massa efetiva  $m^*$ .

Por outro lado, uma maneira de olhar a estrutura eletrônica de pontos quânticos reside em analisar os elétrons da banda de condução, que neste caso estão confinados espacialmente nas três dimensões. Nesse caso, os elétrons podem ser considerados não interagentes e aprisionados em uma caixa de lado a, cujo os níveis de energia são dados por

$$E_{n_x,n_y,n_z} = \frac{\hbar^2 \pi^2}{2m^* a^2} (n_x^2 + n_y^2 + n_z^2) , \qquad (3.1)$$

onde  $n_x$ ,  $n_y$  e  $n_z$  são números quânticos. Esse modelo é extremamente simplificado, porém uma característica importante é capturada, os níveis de energia são essencialmente discretos por causa do confinamento. Além disso, é possível extrair uma noção da escala de energia dos espaçamentos entre os níveis, ou seja,

$$\Delta E \propto \frac{\hbar^2}{m^* a^2} . \tag{3.2}$$

Uma componente chave na etapa de fabricação de pontos quânticos é a criação de um gás de elétrons bidimensional (2DEG), visto que uma forma de produzir os *quantum dots* é
através de hétero-estruturas resultantes do empilhamento de dois tipos de semicondutores dissimilares. Como consequência, os elétrons ficam aprisionados em um poço de potencial aproximadamente simétrico na direção de empilhamento dos materiais. Em virtude da largura desse poço ser pequena (~ 10 nm) e em baixas temperaturas (~ 0.2 K), somente o estado de menor energia desse poço é ocupado, gerando um confinamento perpendicular à direção de empilhamento [84]. Dessa forma, considerando que os materiais estão empilhados na direção do eixo z, pode-se supor que o movimento dos elétrons se limitam ao plano z = 0 em boa aproximação, isto é, ocorre a formação de um 2DEG.

Por outro lado, o confinamento lateral efetivo, ou seja, o potencial V real que atua nos elétrons no plano em que eles podem se mover não é o de uma caixa bidimensional. Além disso, diversos fatores podem contribuir para a forma de V, inclusive a preparação experimental. Uma descrição simples, porém eficiente, se resume em adotar um potencial parabólico isotrópico [5] da forma

$$V(r) = \frac{m^* \omega^2 r^2}{2} , \qquad (3.3)$$

onde  $\omega$  é a frequência de confinamento. Escrevendo esse potencial em unidades adimensionais para o oscilador harmônico, surge uma escala de comprimento, onde o inverso desse comprimento é dado por

$$\alpha_{\rm HO} = \sqrt{\frac{m^*\omega}{\hbar}} \,. \tag{3.4}$$

Entretanto, uma descrição mais precisa também deve levar em conta as interações repulsivas presentes entre os elétrons. Especificamente, a interação de Coulomb entre dois elétrons em um semicondutor também é blindada pela constante dielétrica  $\epsilon$ . Então, o potencial fica escrito como:

$$v(r_{ij}) = \frac{k_e e^2}{\epsilon} \frac{1}{r_{ij}} .$$

$$(3.5)$$

onde  $k_e$  é a constante de Coulomb e e é a carga eletrônica. Portanto, a energia coulombiana de repulsão  $v_{ee}$  entre os elétrons para um *quantum dot* cujas dimensões tem tamanho a é da ordem de

$$v_{ee} \propto \frac{k_e e^2}{\epsilon a}$$
, (3.6)

Como as diferenças entre níveis de energia diminuem com o inverso quadrado do tamanho do sistema, enquanto a repulsão diminui somente com o inverso de a, isto implica que os efeitos da interação eletrônica são relativamente mais importantes quanto maior é o ponto quântico [85]. Uma distância recorrente no estudo desse potencial é o raio de Bohr, porém, é necessário empregar o raio de Bohr efetivo  $a_B^*$  com intuito de levar em conta a influência do fundo semicondutor, assim  $a_B^*$  pode ser definido como:

$$a_B^* = \frac{\epsilon \hbar^2}{k_e m^* e^2} . \tag{3.7}$$

Ainda nesse contexto, outra escala de energia é o Hartree efetivo  $E_h^*$ , que é uma adaptação equivalente considerando o material semicondutor para a energia de Hartree, a qual comumente aparece na solução do átomo de hidrogênio. Dessa forma,  $E_h^*$  é escrito como:

$$E_{h}^{*} = \frac{\hbar^{2}}{m^{*} \left(a_{B}^{*}\right)^{2}} = m^{*} \left(\frac{k_{e}e^{2}}{\epsilon\hbar}\right)^{2} .$$
(3.8)

Com isso, as principais escalas de energia e comprimento característicos são resumidamente apresentadas na Tab. 3.1. Se um ponto quântico 2D semicondutor possui qualquer semelhança com um 2DEG, é esperado que a energia cinética seja dominante no regime de alta densidade, enquanto no regime de baixa densidade as interações coulombianas sejam dominantes [86]. Naturalmente, se a distância inter eletrônica média for alta, isto é, baixas densidades, como a energia cinética decresce com o quadrado da distância característica, isso significa que a interação terá mais importância, uma vez que seu decréscimo é apenas com o inverso da distância. Entretanto, se a densidade for alta e a distância inter eletrônica média for baixa, a energia cinética será dominante.

Tabela 3.1: Resumo das várias escalas de energia e comprimento envolvidas na modelagem dos elétrons de pontos quânticos circulares baseados em semicondutores.

Quantidade	Escala
Energia $[E_h^*]$	$m^* \left(\frac{k_e e^2}{\epsilon \hbar}\right)^2$
Comprimento $[a_B^*]$	$\frac{\epsilon\hbar^2}{k_em^*e^2}$
Interação $[v_{ee}]$	$rac{k_e e^2}{\epsilon a_B^*}$
Confinamento $[V]$	$\hbar\omega$

### 3.2 Modelo para os pontos quânticos

Com estas considerações em vista, um modelo adequado que descreve o comportamento dos N elétrons compondo o ponto quântico, confinados por um potencial parabólico e se movendo no plano z = 0 é determinado pelo hamiltoniano:

$$\mathcal{H} = -\frac{\hbar^2}{2m^*} \sum_i \nabla_i^2 + \frac{m^* \omega^2}{2} \sum_i r_i^2 + \frac{k_e e^2}{\epsilon} \sum_{i < j} \frac{1}{r_{ij}} \,. \tag{3.9}$$

Novamente, os efeitos do fundo semicondutor são condensados na massa efetiva e na constante dielétrica do material. Uma forma interessante de reescrever a Eq. 3.9 consiste em dividir ambos lados pela energia de Hartree efetiva. Ao fazer isso, obtêm-se um hamiltoniano adimensional dado por

$$\hat{\mathcal{H}} = -\sum_{i} \frac{\left(a_B^* \boldsymbol{\nabla}_i\right)^2}{2} + \left(\frac{\hbar\omega}{E_h^*}\right)^2 \sum_{i} \frac{1}{2} \left(\frac{r_i}{a_B^*}\right)^2 + \sum_{i < j} \left(\frac{a_B^*}{r_{ij}}\right)$$
(3.10a)

$$= -\sum_{i} \frac{\hat{\nabla}_{i}^{2}}{2} + \lambda^{-2} \sum_{i} \frac{\hat{r}_{i}^{2}}{2} + \sum_{i < j} \frac{1}{\hat{r}_{ij}} , \qquad (3.10b)$$

onde a constante adimensional  $\lambda = E_{h}^{*}/\hbar\omega$  reflete a importância relativa entre o potencial de confinamento e a interação coulombiana. Em outras palavras, se a configuração experimental é tal que a razão  $\lambda$  está no intervalo entre zero e um, o confinamento é dominante. Um típico semicondutor usado em pontos quânticos é o arsenieto de gálio (GaAs), o qual possui elétrons com uma massa efetiva de  $m^{*} = 0.067 \ m_{e}$ , sendo  $m_{e}$  a massa de repouso do elétron, e uma constante dielétrica de  $\epsilon = 12.7$  [6]. Já para valores típicos do confinamento, a energia  $\hbar\omega$  é da ordem meV, em particular, o confinamento considerado nos cálculos realizados foi de  $\hbar\omega = 0.28 \ E_{h}^{*}$ , que é equivalente à aproximadamente 1.1 meV. Esse é um valor tipicamente encontrado na literatura [87]. Essa breve descrição é suficiente para os propósitos deste trabalho, uma vez que o maior objetivo consiste em analisar como a aplicação de redes neurais pode amenizar o problema do sinal em métodos de Monte Carlo quântico.

#### Comportamentos esperados para função de onda

No estudo de férmions de *spin* <sup>1</sup>/<sub>2</sub>, uma abordagem muito comum utilizada para garantir o caráter antissimétrico da função de onda é o uso de determinantes de Slater [88], os quais podem ser escritos da seguinte forma

$$\Phi_{S}(\mathbf{R}) = \det\left[\phi_{i}\left(\mathbf{r}_{j}^{\uparrow}\right)\right] \det\left[\phi_{i}\left(\mathbf{r}_{j}^{\downarrow}\right)\right] , \qquad (3.11)$$

onde nesse caso  $\mathbf{R} = {\{\mathbf{r}_{1}^{\uparrow}, ..., \mathbf{r}_{n^{\uparrow}}^{\uparrow}, \mathbf{r}_{1}^{\downarrow}, ..., \mathbf{r}_{n^{\downarrow}}^{\downarrow}\}}, \mathbf{r}_{j}^{\uparrow}$  são as coordenadas do elétron j com spin para cima,  $\mathbf{r}_{j}^{\downarrow}$  são as coordenadas do elétron j com spin para baixo,  $n^{\uparrow} e n^{\downarrow}$  são o número de elétrons com spin para cima e para baixo, respectivamente. O determinante é computado considerando a matriz cujo o elemento da linha i e coluna j é dado por  $\phi_{i}(\mathbf{r}_{j}^{\alpha})$ , onde  $\alpha = {\uparrow, \downarrow}$  pode assumir os dois possíveis valores de projeção do spin. Usualmente, essas funções  $\phi_{i}$  são soluções do sistema não interagente, isto é, soluções do oscilador harmônico quântico nesse caso.

Para incluir as correlações entre os elétrons, uma maneira comum é multiplicar os determinantes de Slater por fatores de Jastrow [6], isto é,

$$\psi_{SJ}(\mathbf{R}) = \Phi_S(\mathbf{R})\Psi_J(\mathbf{R}) , \qquad (3.12)$$

onde o fator de Jastrow pode ser escrito em termos de um produtório sobre os possíveis pares de partículas,

$$\Psi_J(\mathbf{R}) = \prod_{i < j} J(r_{ij}) .$$
(3.13)

Nesse caso, esse termo inclui apenas correlações de dois corpos. Uma forma comum para função Jé dada por

$$J(r) = \exp\left(\frac{r}{a+br}\right) , \qquad (3.14)$$

onde as constantes  $a \in b$  são parâmetros variacionais. Outros detalhes podem ser encontrados em [89].

## Parte II

Panorama dos métodos computacionais

No contexto da inteligência artificial, um dos desafios é fazer com que computadores resolvam tarefas que são complicadas de descrever formalmente em termos computacionais, como, por exemplo, reconhecimento facial em imagens. Nesse sentido, existem dois modos distintos em que a expressão "algoritmo de aprendizado de máquina" é empregada. O algoritmo pode ser utilizado para treinamento ou para predição, isto é, ou ele está sendo treinado para executar uma tarefa, ou ele está sendo usado para realizar uma predição sobre uma tarefa. Em particular, esses algoritmos aprendem a partir de conjuntos de dados ao serem treinados. Entretanto, para que esse treinamento seja efetivo, os conjuntos de dados precisam ter tamanhos consideráveis. Felizmente, o aumento exponencial do poder computacional visto nas últimas décadas proporcionou um aumento na capacidade de gerar e analisar conjuntos enormes de dados [15]. Junto com a tendência de aumento no conjunto de dados e do poder computacional, novas técnicas de análise e teoria do aprendizado foram desenvolvidas, as quais foram amplamente inspiradas em estatística, neurociência, ciências da computação e também na física [90]. Esses avanços possibilitaram aplicações abrangendo visão computacional [91], reconhecimento de fala [92], processamento de linguagem natural [93], entre outros [16].

Por outro lado, uma característica intrínseca da mecânica quântica é sua natureza probabilística. Isso torna sistemas físicos fontes praticamente inesgotáveis de dados, seja através de experimentos, seja através de dados sintéticos provenientes de simulações [25]. Tais circunstâncias tornam esses sistemas um ótimo campo para aplicações de algoritmos de aprendizado de máquina. Especificamente, o problema de muitos corpos pode ser atacado utilizando os conceitos do aprendizado de máquina através de algumas perspectivas distintas. Dentre as possíveis aplicações estão a identificação de fases quânticas [24], aceleração de simulações [32], construção de potenciais para simulações atomísticas [33] e representação de estados quânticos [24,31].

Para entender particularmente como algoritmos de aprendizado de máquina são capazes de representar estados quânticos, este capítulo discute os conceitos chave de algoritmos de *machine learning*, como construir seus elementos fundamentais, como treinar esses algoritmos para realizar uma determinada tarefa e quais cuidados necessários devem ser tomados para que um treinamento eficiente seja executado. Nessa etapa, o foco reside em entender as ferramentas matemáticas empregadas no contexto da inteligência artificial, enquanto a conexão dos conceitos de aprendizado de máquina com o problema de muitos corpos quântico e o método de Monte Carlo é feita nos Cap. 5 e Cap. 6.

#### 4.1 Estabelecendo definições de conceitos chave

O aprendizado de máquina é um campo de estudo dedicado ao desenvolvimento de algoritmos e modelos estatísticos que permitem ao computador realizar tarefas sem ser explicitamente programado para isso. Para que isso seja possível, apresenta-se ao computador um conjunto de informações chamados de dados de treinamento. Em particular, o *machine learning* pode ser definido como uma forma de estatística aplicada com maior ênfase ao uso de computadores para estimar funções complicadas e menor ênfase aos intervalos de confiança dessas funções [16]. Em outras palavras, supondo que a tarefa em questão possa ser representada por uma função, a ênfase reside em estimar com acurácia essa função independente de quando novos dados são apresentados, diferentemente de um ajuste de função, procedimento no qual existe uma preocupação maior em torno da região na qual o ajuste é confiável. Adicionalmente, o **modelo** é essa função capaz de executar a tarefa.

Portanto, os algoritmos de aprendizado tem como objetivo genérico reconhecer padrões em dados e aprender a partir destes. Entretanto, aprender é um conceito subjetivo que pode ser interpretado de diversas maneiras. Por essa razão, surge a seguinte definição sugerida por Mitchell [94]: dado uma experiência E com respeito à uma classe de tarefas T e uma medida de performance P, o conceito de **aprender** para um programa computacional consiste em aprimorar sua performance P nas tarefas T com a experiência E. Assim sendo, o conjunto de dados de treinamento utilizado para treinar o algoritmo é a experiência E apresentada a ele.

Com isto em vista, aprender significa adquirir a habilidade de desempenhar bem em uma determinada tarefa. Em geral, essas tarefas são descritas em termos de como um exemplo específico deve ser processado pelo algoritmo. Nesse contexto, um exemplo é entendido como um conjunto de **atributos**, isto é, um conjunto de dados numéricos que podem ser representados por um vetor  $\mathbf{x} \in \mathbb{R}^n$ , onde  $x_i$  são os elementos desse vetor. Para exemplificar, uma imagem de um animal doméstico é um exemplo cujos atributos são os valores de seus píxels. Para esse caso, uma possível tarefa seria identificar qual tipo de animal doméstico está presente na foto. Há diversos tipos de tarefa que o aprendizado de máquina é capaz de enfrentar, porém, dois dos tipos mais importantes são a classificação e a regressão.

A classificação é um tipo de tarefa onde o objetivo é especificar a qual categoria os exemplos pertencem, ou seja, dado um conjunto de exemplos e um de categorias identificadas por um código numérico y, a ideia é encontrar uma função que conecte os exemplos  $\mathbf{x}$  as suas devidas classificações y. Reconhecimento de faces em fotos usa esse tipo de ideia para marcar perfis de redes sociais de maneira automática [95]. Note que, neste caso y pode assumir apenas valores discretos. Regressão, por outro lado, tem como objetivo prever um valor numérico y a partir de um exemplo  $\mathbf{x}$ . Novamente, a intenção é achar a função que conecte os dois, ou seja, ambas tarefas são semelhantes, com a diferença que, para a regressão, os valores de y pertencem ao contínuo. Um exemplo de regressão é a análise do preço de ações em bolsas de valores [96]. Outros exemplos de tarefa são a transcrição, usada para reconhecimento de texto e fala, a detecção de anomalia, usada para reconhecimento de fraudes em transações bancárias, e a síntese e amostragem, usada para geração de novos dados a partir dos exemplos usados para treinamento [16].

Além disso, os exemplos processados pelo algoritmo geralmente podem ser classificados em três tipos, aqueles que pertencem aos dados de treinamento, ao conjunto de teste ou aos novos dados. Usualmente, o conjunto de dados teste é a coleção de informações usadas para computar a performance do algoritmo. Porém, para avaliar o desempenho do algoritmo ao executar uma tarefa, torna-se necessário a elaboração de uma medida quantitativa de performance. O tipo de medida depende diretamente de qual tarefa se quer realizar. Para a classificação, por exemplo, a proporção de exemplos para os quais o modelo produziu o resultado correto, isto é, a acurácia é uma boa medida de performance. De maneira geral, a performance pode ser entendida como um retorno ou *feedback* para o processamento de um exemplo pelo modelo. Ao obter uma boa pontuação na performance, isso é equivalente à ter um retorno positivo, indicando ao modelo para continuar aprendendo nessa direção. De maneira semelhante, uma baixa pontuação mostra um *feedback* negativo, levando à uma mudança de caminho no aprendizado.

Entretanto, a forma como a experiência é apresentada ao algoritmo também pode variar. Por consequência, os algoritmos são classificados majoritariamente em três maneiras dependendo de qual tipo de experiência é apresentada. Os algoritmos de **aprendizado não supervisionado** são aqueles que experienciam um conjunto de dados contendo diversos atributos e são capazes de aprender as estruturas e padrões internos ao conjunto. Já para os algoritmos de **aprendizado supervisionado**, além dos atributos de entrada, a cada exemplo do conjunto é atribuída uma categoria de saída ou valor alvo de saída que o modelo tem como objetivo prever. Por último, há também o **aprendizado reforçado**, no qual o algoritmo interage com um ambiente, de maneira que existe um ciclo de *feedbacks* entre o sistema e suas experiências. Um exemplo desse tipo de aprendizado é o treinamento de modelos para jogar xadrez.

Com base no conceito de aprender, usualmente, um modelo antes de entrar em contato com uma experiência tem uma baixa performance. O processo de aprendizado, isto é, a melhoria da performance com a experiência está intimamente ligado à algoritmos de otimização. Frequentemente, a performance é descrita em termos de uma função, cujo s pontos de máximo ou mínimo correspondem às melhores performances que o modelo pode ter. Também, como mencionado anteriormente, um desafio central no processo de aprendizagem reside em obter algoritmos que tenham um bom desempenho no processamento de exemplos não vistos na etapa de treinamento, ou seja, novos dados. A demonstração de um bom desempenho em dados não observados previamente é conhecida como **generalização**. Exatamente isso é o que separa uma otimização simples de um algoritmo de aprendizado de máquina, neste último também se deseja que a medida de performance P seja boa em dados não utilizados no treinamento.

### 4.2 Modelos de inteligência artificial

Para compreender com mais detalhes estes conceitos, dois exemplos serão explorados a seguir. O primeiro é amplamente conhecido nas ciências exatas, a regressão linear, enquanto os neurônios artificias são o segundo, estrutura base para diversos modelos utilizados no contexto de inteligência artificial. Com estes, pode-se definir os modelos conhecidos como redes neurais diretas, as quais foram amplamente empregadas no estudo e aplicações em *machine learning* apresentados anteriormente.

#### Regressão linear

Muitas vezes, as tecnologias de inteligência artificial são vistas como obscuras e incompreensíveis. Entretanto, analisando a definição de aprender apresentada, um exemplo simples de algoritmo de aprendizado de máquina é a regressão linear. Por exemplo, suponha que a tarefa considerada seja encontrar uma lei para um determinado fenômeno cujo experimento nos fornece um certo conjunto de dados. Considere também que cada exemplo nesse conjunto seja composto de um vetor  $\mathbf{x}$  com n atributos e um valor alvo y. Modelos nada mais são do que funções de muitas variáveis. Essas variáveis são separadas em dois tipos, os já mencionados atributos  $\mathbf{x}$  e os parâmetros  $\boldsymbol{\theta}$ . Assim, denotando essa função por  $\phi$ , pode-se escrever um modelo linear como

$$\phi(\mathbf{x}, \boldsymbol{\theta}) = \sum_{i} w_{i} x_{i} + b = \mathbf{w}^{\top} \mathbf{x} + b , \qquad (4.1)$$

onde  $\mathbf{w} = \{w_1, ..., w_n\}$  é um conjunto de parâmetros pertencentes à  $\boldsymbol{\theta}$  chamados de pesos e, semelhantemente, *b* também pertence à  $\boldsymbol{\theta}$  e é chamado de viés. Nessa equação, foi utilizada uma notação matricial, onde considera-se que um vetor é uma matriz coluna. Em vista disso, a tarefa pode ser reinterpretada como predizer o valor alvo *y* a partir do exemplo **x** através do cálculo de  $\phi(\mathbf{x}, \boldsymbol{\theta})$ .

Uma maneira de representar o conjunto total de dados é através da matriz design  $\mathbf{X}$ , onde os elementos compondo uma coluna formam um exemplo e elementos de linhas diferentes correspondem a atributos distintos. Essa matriz armazena dois tipos de exemplos, onde no processo de avaliação da performance é empregado o conjunto teste  $\mathbf{X}^{\text{test}}$  e na etapa de aprendizagem é utilizado o conjunto de treinamento  $\mathbf{X}^{\text{train}}$ , ambos contidos na matriz design e totalmente excludentes, isto é, não possuem exemplos iguais entre si. De maneira semelhante, os valores alvo de cada exemplo são armazenados no vetor  $\mathbf{y}$  na ordem correspondente, além de também estar separado entre valores alvo de teste  $\mathbf{y}^{\text{test}}$  e de treinamento  $\mathbf{y}^{\text{train}}$ .

Nesse caso, uma boa escolha de medida de performance é o quanto a previsão dada pelo modelo desvia do valor alvo esperado, em outras palavras, o erro médio do modelo. Especificamente, o erro médio quadrático é uma ótima medida para esse exemplo, portanto, a performance será avaliada sob esta ótica. Como mencionado anteriormente, essa medida é feita no conjunto de teste, consequentemente, o desempenho do modelo será dada por

$$P = \frac{1}{m} \sum_{i} \left[ \phi(\mathbf{x}_{i}^{\text{test}}, \boldsymbol{\theta}) - y_{i}^{\text{test}} \right]^{2} , \qquad (4.2)$$

onde m é o número de amostra no conjunto teste,  $\mathbf{x}_i^{\text{test}}$  é um exemplo de  $\mathbf{X}^{\text{test}}$  e  $y_i^{\text{test}}$  é seu valor alvo correspondente. Note que nessa situação, quanto menor é o valor de P, melhor é o desempenho.

Para que o algoritmo possa aprender, isto é, melhorar o desempenho, é necessário construir um algoritmo que melhore os pesos  $\mathbf{w}$  e viés b de forma que reduza P quando o algoritmo é apresentado à experiência, ou seja, ao conjunto de dados de treinamento. Uma forma intuitiva de fazer isso é minimizar o erro quadrático médio para o conjunto de treinamento. Com esse propósito, o objetivo torna-se encontrar os valores de  $\boldsymbol{\theta}$  que anulam o gradiente dessa função. Para isso, basta encontrar a solução de

$$\boldsymbol{\nabla}_{\boldsymbol{\theta}} \sum_{i} \left[ \phi(\mathbf{x}_{i}^{\text{train}}, \boldsymbol{\theta}) - y_{i}^{\text{train}} \right]^{2} = \mathbf{0} .$$
(4.3)

Particularmente para a regressão linear, esse sistema tem solução exata conhecida. De posse dos parâmetros  $\boldsymbol{\theta}$  ótimos que minimizam o erro médio, o algoritmo tem seu desempenho melhorado e, como consequência, ele aprendeu. É claro que a regressão linear é um algoritmo de aprendizado de máquina muito simples e tem muitas limitações, mas mostra de maneira simplificada como podem funcionar algoritmos mais complexos.

#### Neurônios artificiais

Uma das deficiências com a regressão linear é a necessidade de correlações não lineares entre os atributos e as previsões do modelo, situação presente em diversas tipos de tarefas. Para transpor esse problema surgiram modelos chamados de neurônios artificiais. Semelhantemente aos neurônios de animais, esses modelos recebem sinais primários ou proveniente de outros neurônios, processam de modo não linear e transmitem o sinal processado, seja para outros neurônios artificiais ou para composição da resposta do modelo. Os sinais primários são os atributos e correspondem aos sentidos, como a visão por exemplo. De maneira geral, um neurônio artificial vai receber um vetor, denotado pela letra **a**, com as informações dos sinais que ele está recebendo.

Após receber as informações, o neurônio artificial junta todas elas através de uma combinação linear, semelhante ao procedimento da Eq. 4.1. Já o tratamento não linear do sinal é executado através de uma função de ativação  $\sigma$ , que é associada ao neurônio. Na Fig. 4.1 é apresentado o esquema de um neurônio artificial, onde é possível ver os sinais

chegando ao neurônio, sendo processados e, posteriormente, sendo emitido. A expressão desse processo pode ser escrita como

$$h = \sigma \left( \mathbf{w}^{\top} \mathbf{a} + b \right) \ . \tag{4.4}$$



Figura 4.1: Ilustração esquemática de um neurônio artificial. Os sinais  $a_i$  chegam ao neurônio com diferentes pesos  $w_i$  e são combinados linearmente junto com o viés b. O resultado desse processamento passa por um filtro não linear e é emitido para os próximos neurônios.

Note que a figura representa a combinação linear pela letra grega  $\Sigma$  e a operação não linear pelo gráfico da função *sigmoid*, a qual tem a seguinte expressão

$$\sigma(t) = \frac{1}{1 + e^{-t}} \ . \tag{4.5}$$

Essa função pode ser vista como um filtro, onde até uma certa amplitude de sinal não há transmissão e para amplitudes maiores há transmissão. Existem diversas funções de ativação empregadas nesses modelos, dentre elas estão a função *sigmoid* citada anteriormente, como também estão a tangente hiperbólica, a função sinal e a unidade linear retificada (ReLU). Esta última é definida como zero para valores negativos e o próprio argumento para valores positivos, o que esclarece o nome atribuído.

A partir dessa estrutura é possível compor diversos neurônios artificiais para criar as chamadas redes neurais. Como a maioria dos neurônios artificiais não interagem diretamente com os atributos de entrada ou com a composição do resultado final do modelo, essas estruturas são chamadas de unidades ocultas (*hidden units*), por isso são representados pela letra h na Eq. 4.4.

### 4.3 Descrição das redes neurais diretas

Os modelos empregados nos algoritmos de aprendizado de máquina com frequência se fundamentam nas chamadas redes neurais diretas, as quais utilizam neurônios artificiais organizados em formato de camadas sucessivas e inter conectadas, com cada camada sendo uma função da camada anterior. Essas redes são alimentadas com os atributos, também chamados de elementos de entrada ou camada de entrada, processam essa informação através dos neurônios presentes e realizam, por exemplo, uma classificação ou uma previsão relacionada com a tarefa que se quer cumprir.

O número total de camadas encadeadas é chamado de **profundidade** do modelo, onde a última delas é conhecida como camada de saída. Adicionalmente, cada camada contém uma determinada quantidade de neurônios, propriedade conhecida como **largura**. Pelo fato das camadas serem conectadas sequencialmente, pode-se escrever relações de recorrência entre elementos de uma dada camada  $\ell$  e os elementos da camada anterior  $\ell - 1$ . Como mencionado anteriormente, essas camadas intermediarias são chamadas de camadas ocultas.

Da mesma forma que no neurônio artificial, uma camada  $\ell$  recebe um vetor  $\mathbf{a}^{(\ell-1)}$  da camada anterior  $\ell-1$  para ser processado. Entretanto, ao contrário de um único neurônio, cada camada geralmente possui diversos neurônios, ou seja, o resultado do processamento também será um vetor. Portanto, para realizar a combinação linear é necessário utilizar uma matriz de pesos  $\mathbf{W}^{(\ell)}$  e um vetor de vieses  $\mathbf{b}^{(\ell)}$  para conectar camadas subsequentes. Para tanto, uma quantidade intermediária chamada de pré-ativação é definida como

$$\mathbf{z}^{(\ell)} = \mathbf{W}^{(\ell)} \mathbf{a}^{(\ell-1)} + \mathbf{b}^{(\ell)} .$$

$$(4.6)$$

Note que cada linha das matrizes  $\mathbf{W}^{(\ell)}$  representa os pesos de um dos neurônios da camada, bem como os elementos de  $\mathbf{b}^{(\ell)}$  representam os vieses de cada neurônio. De forma mais explicita, essa representação matricial pode ser reescrita em termos dos elementos. Assim, o elemento *i* da pré-ativação da camada  $\ell$  é dado por

$$z_i^{(\ell)} = \sum_j W_{ij}^{(\ell)} \cdot a_j^{(\ell-1)} + b_i^{(\ell)} .$$
(4.7)

Na etapa seguinte do processamento, a operação não linear é realizada, isto é, a função de ativação  $\sigma_{\ell}$  é aplicada a cada elemento da pré-ativação. Consequentemente, os sinais que serão transmitidos à próxima camada são o resultado desse processamento e podem ser escritos vetorialmente como

$$\mathbf{a}^{(\ell)} = \sigma_{\ell} \left( \mathbf{z}^{(\ell)} \right) \,, \tag{4.8}$$

onde a função de ativação está sendo aplicada elemento a elemento do vetor de pré-ativação

 $\mathbf{z}^{(\ell)}$ . Para uma rede neural cuja a profundidade é de L camadas, a expressão final do modelo é dada por

$$\phi_{\rm nn}(\mathbf{x},\boldsymbol{\theta}) = \sigma_L \left( \mathbf{W}^{(L)} \mathbf{a}^{(L-1)} + \mathbf{b}^{(L)} \right) , \qquad (4.9)$$

onde  $\mathbf{x} = \mathbf{a}^{(0)}$  representa a camada de entrada e o vetor  $\boldsymbol{\theta} = {\mathbf{W}^{(\ell)}} \cup {\mathbf{b}^{(\ell)}}$  é o conjunto de todos os parâmetros da rede neural. Note que esse cálculo pode retornar também um vetor de resultados, além de valores escalares. Essa flexibilidade reside na liberdade de escolha do tamanho da largura de cada camada, em particular, da largura da última camada.

Um fator importante para que o modelo tenha êxito na tarefa é a maneira como as informações disponíveis são codificadas nos atributos  $\mathbf{a}^{(0)}$ . Em uma tarefa relacionada com o processamento de imagens, como por exemplo reconhecimento de objetos em imagens, é preciso representar a imagem em termos computacionalmente legíveis, isto é, números. Uma forma comum é transformar cada pixel da imagem em uma trinca de valores, onde cada valor pode assumir valores reais entre 0 e 1 e representa a quantidade de vermelho, verde e azul daquele pixel. Essa é uma maneira de construir os atributos  $\mathbf{a}^{(0)}$ , outras formas também existem e algumas delas são mais eficientes por serem pensadas para uma tarefa especifica. Portanto, vale pontuar que essa é uma etapa importante do processo e dependente do problema em questão. A forma como os atributos são apresentados à rede neural é chamada de **representação**. Um ponto no plano pode ser representado em coordenadas cartesianas ou polares, por exemplo.

Para fins ilustrativos, uma representação esquemática de uma rede neural é apresentada na Fig. 4.2, a qual é detalhada a seguir. A rede representada no esquema apresenta



Figura 4.2: Ilustração esquemática de uma rede neural direta. A rede possui duas camadas ocultas, uma com 4 neurônios e outra com 2. Nesse caso, a camada de saída é um único número representado pela função  $\phi_{nn}(\mathbf{x}, \boldsymbol{\theta})$ , onde os atributos  $\mathbf{x}$  são representados pelos sinais da camada de entrada  $a_i^{(0)}$  e os parâmetros  $\boldsymbol{\theta}$  pelas matrizes de pesos  $\mathbf{W}$ .

uma camada de entrada com 3 atributos, duas camadas ocultas, sendo uma de largura 4 e outra de largura 2. A camada de saída nesse exemplo é um escalar. Por motivos de clareza, os vieses não foram representados na figura. De posse do modelo, ideias semelhantes às introduzidas no exemplo da regressão linear podem ser empregadas para atacar diversos tipos de tarefa, em particular, tarefas nas quais os valores alvo tem correlações não lineares com os atributos. Sobretudo, considerar que a tarefa pode ser representada por uma função introduz um questionamento relacionado à capacidade do modelo apresentado de representar a função em questão. De fato, o teorema de Cybenko [37] garante que a classe de funções multivariáveis da Eq. 4.9 é capaz de aproximar tão bem quanto se queira qualquer função real com suporte no hiper-cubo unitário, isto é, funções compactas em um subconjunto do  $\mathbb{R}^n$ . Geralmente, esse é o caso para problemas de aprendizado de máquina.

Com modelos desse tipo, desempenhos avaliados por quantidades similares ao erro médio quadrático e dados representados adequadamente e em quantidades suficientes para treinar essas redes neurais, quase todas as componentes do algoritmo de aprendizado de máquina estão descritas. A componente que completa os algoritmos de aprendizado de máquina são os algoritmos de otimização.

#### 4.4 Elementos de algoritmos de otimização

Métricas de performance são sempre construídas tal que seus máximos, ou mínimos dependendo da construção, resultem em um melhor desempenho do algoritmo. Por esse motivo, é necessário discutir algoritmos de otimização para redes neurais, os quais são responsáveis por encontrar os parâmetros  $\boldsymbol{\theta}$  do modelo tal que este seja capaz de realizar a tarefa com um bom desempenho. Usualmente, a função que mede a performance também é chamada de função perda e é denotada por  $\mathcal{L}$ . Como já mencionado, um exemplo de função perda é o erro quadrático médio. Nesse caso, a função perda depende do resultado das redes neurais, ou seja, é uma função composta  $\mathcal{L}(\phi_{nn}(\mathbf{x}, \boldsymbol{\theta}))$ . Por simplificação, considere que esse seja o caso geral. Portanto, para o cálculo de  $\mathcal{L}$ , emprega-se a rede neural direta utilizando os dados de entrada  $\mathbf{x}$  para produzir valores de saída  $\phi_{nn}(\mathbf{x}, \boldsymbol{\theta})$ , onde a informação flui pela rede progressivamente, isto é, ela avança de camada para camada. Esse processo é chamado de propagação progressiva.

Para encontrar os parâmetros que otimizam o desempenho do modelo, métodos baseados em gradiente estão entre os mais comuns. Assim, com o objetivo de calcular os gradientes da função perda, o algoritmo de retro-propagação foi desenvolvido [21]. Esse procedimento também é conhecido como *back-propagation* e permite que a informação flua da função perda para a rede neural, ou seja, de maneira retrógrada. Nesse sentido, a forma analítica dos gradientes de  $\mathcal{L}$  são calculados através de uma série de regras da cadeia. Na primeira etapa desse cálculo obtém-se:

$$\nabla_{\boldsymbol{\theta}} \mathcal{L}(\phi_{nn}(\mathbf{x}, \boldsymbol{\theta})) = \mathcal{L}'(\phi_{nn}(\mathbf{x}, \boldsymbol{\theta})) \nabla_{\boldsymbol{\theta}} \phi_{nn}(\mathbf{x}, \boldsymbol{\theta}) .$$
(4.10)

Note que a função perda é um escalar e, por isso, sua derivada pode ser denotada simplesmente como  $\mathcal{L}'$ . A partir desse ponto, é preciso dividir o gradiente em duas categorias, uma para as derivadas com relação aos vieses  $\mathbf{b}^{(\ell)}$  e outra para as derivadas com relação aos pesos  $\mathbf{W}^{(\ell)}$ . Para a camada de saída L, as expressões para as derivadas são dadas por

$$\frac{\partial \mathbf{a}^{(L)}}{\partial \mathbf{b}^{(L)}} = \frac{\partial \mathbf{a}^{(L)}}{\partial \mathbf{z}^{(L)}} = \mathbb{1} \odot \sigma'_L \left( \mathbf{z}^{(L)} \right) , \qquad (4.11)$$

$$\frac{\partial \mathbf{a}^{(L)}}{\partial \mathbf{W}^{(L)}} = \frac{\partial \mathbf{a}^{(L)}}{\partial \mathbf{b}^{(L)}} \otimes \mathbf{a}^{(L-1)} .$$
(4.12)

A operação  $\odot$  denota um produto onde cada elemento de um vetor multiplica cada elemento de uma coluna da matriz, isto é,  $(\mathbf{M} \odot \mathbf{v})_{ij} = M_{ij}v_j$ . Enquanto a operação  $\otimes$  indica produto tensorial. Note a necessidade de primeiro computar as derivadas da última camada para depois fluir as informações de forma retrógrada pela rede até chegar ao início, sendo assim, uma retro-propagação. As quantidades das Eq. 4.11 e Eq. 4.12 são retroativamente empregadas no cálculo das derivadas com relação aos pesos e vieses de uma camada  $\ell < L$  da rede neural, cujas expressões são escritas como

$$\frac{\partial \mathbf{a}^{(L)}}{\partial \mathbf{b}^{(\ell)}} = \frac{\partial \mathbf{a}^{(L)}}{\partial \mathbf{z}^{(\ell+1)}} \mathbf{W}^{(\ell+1)} , \qquad (4.13)$$

$$\frac{\partial \mathbf{a}^{(L)}}{\partial \mathbf{W}^{(\ell)}} = \frac{\partial \mathbf{a}^{(L)}}{\partial \mathbf{b}^{(\ell)}} \otimes \mathbf{a}^{(\ell-1)} .$$
(4.14)

Matematicamente, a Eq. 4.13 também são derivadas com relação as pré-ativações, além de serem chamadas de sensibilidades e denotadas por  $e^{(\ell)}$ . Detalhes adicionais da derivação dessas expressões podem ser encontrados no Ap. C.

Diferentemente da regressão linear, encontrar os parâmetros que anulam o gradiente de redes neurais não se resume a um conjunto de equações lineares exatamente solúvel porque a não linearidade do modelo torna a função perda não convexa, isto é, ela pode possuir múltiplos mínimos locais. Então, torna-se necessário explorar algoritmos de otimização para achar parâmetros do modelo que aumentem seu desempenho. Um método bastante empregado para otimização nesse contexto é o gradiente estocástico descendente [97]. As possibilidades de método e detalhes variam conforme a tarefa.

Costumeiramente, os métodos de otimização são iterativos e os parâmetros são atualizados durante as iterações do algoritmo de acordo regras do tipo

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \boldsymbol{v}_t \;, \tag{4.15}$$

onde t é um índice inteiro que monitora as iterações e  $v_t$  é uma variação potencialmente benéfica para o desempenho. Geralmente, essa quantidade é proporcional ao gradiente da função perda, ou seja,  $v_t = -\eta \nabla_{\theta} \mathcal{L}_t$ , em que o coeficiente de proporcionalidade  $\eta$  é conhecido como **taxa de aprendizagem**. Essa escolha tem origem na propriedade do gradiente de apontar para a direção de crescimento da função no ponto. Uma justificativa com mais detalhes é apresentada no Cap. 5. Entretanto, devido aos mínimos locais, pontos de sela e possíveis ruídos nos dados, basear-se somente em gradientes pode levar a resultados sub-otimizados. Além disso, parâmetros diferentes frequentemente estão a distâncias distintas dos seus valores ótimos e, portanto, uma taxa de aprendizagem global também pode atrapalhar o processo.

Por causa de problemas desse tipo, as atualizações v dos parâmetros precisam utilizar estratégias para transpor essas dificuldades. Dentre elas está o algoritmo do **momento** [98], o qual introduz um termo em v que faz o papel de uma velocidade no espaço de parâmetros. Essencialmente, esse termo acumula durante o processo de otimização informações das derivadas da função perda de iterações passadas, isto é

$$\boldsymbol{v}_{t+1} = \alpha \boldsymbol{v}_t - \eta \nabla_{\boldsymbol{\theta}} \mathcal{L}_t , \qquad (4.16)$$

onde a constante  $\alpha$  é usualmente referenciada como momento. Esse procedimento acumula uma média móvel com decaimento exponencial do gradiente da função perda e é geralmente realizado logo antes da atualização dos parâmetros. Como  $\alpha$  é escolhido entre 0 e 1, os termos de iterações anteriores decaem exponencialmente, uma vez que, a cada iteração  $\boldsymbol{v}$ é multiplicado por  $\alpha$  antes de acumular o novo gradiente.

Em diversos casos, devido aos gradientes serem estimados estocasticamente, o cálculo dos gradientes contém ruídos provenientes das amostras aleatórias. Assim, outra prática necessária é a adaptação da taxa de aprendizagem como função das iterações do algoritmo. Uma forma de atualizar essa taxa  $\eta$  na iteração t é dado por

$$\eta(t) = \eta_0 \left(\frac{1}{1 + \frac{t}{t_0}}\right)^p , \qquad (4.17)$$

onde  $\eta_0$  é a taxa de aprendizagem inicial,  $t_0$  é a **taxa de retardamento** e p é a **taxa de decaimento**. Isso é necessário, pois o ruído introduzido por métodos estocásticos na estimativa não se anula nos mínimos. Ao utilizar este tipo de decaimento, visa-se diminuir as variações de parâmetro durante suas atualizações, de forma que os ruídos dos dados influenciem menos ao fim do processo de treinamento.

Como mencionado anteriormente, o que diferencia aprendizado de máquina de uma simples otimização é como construir um algoritmo que desempenhe bem não só no conjunto de dados de treinamento, como também nos dados de teste. Existem diversas estratégias utilizadas com esse objetivo que possibilitam uma diminuição do erro de teste, mesmo que em detrimento do erro de treinamento em alguns casos. Esse conjunto de estratégias é geralmente referido como **regularização**. Usualmente, essas táticas se baseiam em restrições ou penalidades incluídas na função perda ou no modelo. Uma maneira comum de implementar isso é modificar a função perda para acrescentar a constrição da norma dos parâmetros do modelo como penalidade. A função perda modificada fica igual à

$$\mathcal{L}' = \mathcal{L} + \lambda \boldsymbol{\theta}^{\top} \boldsymbol{\theta} , \qquad (4.18)$$

onde  $\lambda$  é chamado de **constrição da norma** e é um valor escolhido previamente que define quanto é a preferência por valores menores de  $\boldsymbol{\theta}$ . Isso garante que os valores ótimos de  $\boldsymbol{\theta}$ sejam os menores possíveis.

Neste trabalho, foi utilizado uma modificação do método de aproximação da curvatura com fatorização de Kronecker (KFAC) [99], que é uma aproximação para o gradiente natural descendente [100] apropriado para redes neurais. Do ponto de vista da mecânica quântica, esse método é uma variação do método de reconfiguração estocástica, que é explorado no Cap. 5. Fundamentalmente, o gradiente natural é um método que considera a quantidade de informação que uma variável aleatória carrega sobre um parâmetro desconhecido, isto é, a informação que um determinado parâmetro do modelo carrega sobre os outros parâmetros. A medida dessa quantidade de informação é feita através da matriz de informação de Fisher (FIM), definida como

$$\mathcal{F} = \mathbb{E}_{\rho} \left[ \left( \frac{\partial \log \mathcal{L}}{\partial \boldsymbol{\theta}} \right) \left( \frac{\partial \log \mathcal{L}}{\partial \boldsymbol{\theta}} \right)^{\mathsf{T}} \right] , \qquad (4.19)$$

onde  $\mathbb{E}_{\rho}$  representa o valor esperado com respeito a densidade de probabilidade  $\rho$  da função log-verossimilhança log  $\mathcal{L}$ . Para os casos em que o modelo pode ser interpretado como uma densidade de probabilidade, o logaritmo da função perda pode ser visto como a função log-verossimilhança. Note que o valor esperado é computado sobre o conjunto de dados de treinamento e para cada par de parâmetros do modelo, formando uma matriz simétrica e, portanto, diagonalizável. Além disso, essa matriz fornece uma medida da curvatura da função log-verossimilhança. Com base nisso, a proposta de variação  $\boldsymbol{v}$  nos parâmetros  $\boldsymbol{\theta}$ na iteração t + 1 do processo de treinamento é dada por

$$\boldsymbol{v}_{t+1} = \alpha \boldsymbol{v}_t - \eta(t) \mathcal{F}_t^{-1} \nabla_{\theta} \mathcal{L}_t' , \qquad (4.20)$$

onde  $\mathcal{F}_t \in \mathcal{L}'_t$  são a matriz de Fisher e a função perda calculadas na iteração t, respectivamente. Esse procedimento leva em conta a geometria do espaço de parâmetros ao considerar a curvatura. Isso também torna o processo de otimização mais robusto e eficiente, uma vez que a convergência é comumente atingida mais rapidamente do que usando gradiente descendente simples e problemas relacionados com divergência e desaparecimento de gradientes são evitados ao considerar a geometria do espaço. Contudo, para modelos de larga escala, os quais possuem de milhares até milhões de parâmetros, computar a inversa da matriz de Fisher toda iteração pode ser impraticável. O método KFAC faz duas aproximações com o intuito de amenizar esse problema. A primeira consiste em considerar que os elementos  $\mathcal{F}_{ij}$  são nulos se  $\theta_i \in \theta_j$  pertencem a camadas distintas da rede neural. Isso torna a matriz de Fisher bloco diagonal entre camadas e razoavelmente mais eficiente de se inverter. A segunda aproximação decorre da fatoração de Kronecker da inversa desses blocos, que pode ser escrita em termos das derivadas do modelo com relação aos parâmetros como

$$\mathbb{E}_{\rho}\left[\left(\frac{\partial \log \rho}{\partial \operatorname{vec} \mathbf{W}^{(\ell)}}\right)\left(\frac{\partial \log \rho}{\partial \operatorname{vec} \mathbf{W}^{(\ell)}}\right)^{\mathsf{T}}\right]^{-1} = \mathbb{E}_{\rho}\left[\left(\mathbf{a}^{(\ell)} \otimes \mathbf{e}^{(\ell)}\right)\left(\mathbf{a}^{(\ell)} \otimes \mathbf{e}^{(\ell)}\right)^{\mathsf{T}}\right]^{-1}$$
(4.21a)

$$\approx \mathbb{E}_{\rho} \left[ \mathbf{a}^{(\ell)} \mathbf{a}^{(\ell)T} \right]^{-1} \otimes \mathbb{E}_{\rho} \left[ \mathbf{e}^{(\ell)} \mathbf{e}^{(\ell)T} \right]^{-1} , \quad (4.21b)$$

onde vec $\mathbf{W}^{(\ell)}$  são os parâmetros da camada  $\ell$  dispostos em um vetor,  $\mathbf{a}^{(\ell)}$  são as ativações diretas e  $\mathbf{e}^{(\ell)}$  são as sensibilidades retroativas da camada  $\ell$ . Dessa forma, as informações transmitidas diretamente são desacopladas das informações fluindo retroativamente. Outros detalhes dessa aproximação podem ser encontrados em [99]. Inicialmente, o método KFAC foi pensando para densidades de probabilidades normalizadas, porém, pequenas alterações no procedimento podem estende-lo para casos em que apenas se tem acesso a densidades de probabilidade não normalizadas [31].

Entretanto, mesmo com esses cuidados, inverter a matriz de Fisher pode apresentar problemas numéricos. Uma maneira de melhorar a estabilidade da inversão da matriz é adicionando um termo proporcional à matriz identidade antes de computar a inversa da matriz de Fisher. Dessa forma, a regra de atualização dos parâmetros é escrita como

$$\boldsymbol{v}_{t+1} = \alpha \boldsymbol{v}_t - \eta(t) \left( \mathcal{F}_t + \lambda_d \mathbb{1} \right)^{-1} \nabla \mathcal{L}'_t , \qquad (4.22)$$

onde a constante  $\lambda_d$  é chamada de **amortecimento**. Essa técnica estabiliza o processo de treinamento, reduzindo oscilações e prevenindo divergências. Outro método de regularização utilizado foi o decaimento de médias móveis da matriz de covariância. Esse método introduz uma penalização baseada na covariância das ativações da rede neural, o que encoraja ativações descorrelacionadas.

$$\boldsymbol{v}_{t+1} = \alpha \boldsymbol{v}_t - \eta(t) \left[ \left( \mathcal{F}_t + \lambda_d \mathbb{1} \right)^{-1} \nabla \mathcal{L}'_t + \lambda_c \operatorname{cov}(\mathbf{A}_t) \boldsymbol{\theta}_t \right] , \qquad (4.23)$$

onde o parâmetro  $\lambda_c$  é chamado de **decaimento da matriz de covariâncias** e  $\mathbf{A}_t$  é a matriz de covariância das ativações. Neste trabalho, a atualização dos parâmetros da rede neural foram conduzidas utilizando a biblioteca KFAC-JAX [101, 102].

A quantidade total de iterações é chamada de iterações de treinamento ou itera-

ções de otimização. Além disso, o número de exemplos considerado na matriz design de treinamento é chamado de tamanho do *batch*. Os valores da profundidade e largura do modelo, bem como o número de iterações de treinamento, o tamanho do *batch*, a taxa de aprendizagem inicial, o momento, as taxas de decaimento e retardamento, a constrição de norma e o amortecimento são quantidades que controlam o comportamento do algoritmo durante o processo de treinamento. Para diferenciar dos parâmetros  $\boldsymbol{\theta}$  do modelo, eles são

durante o processo de treinamento. Para diferenciar dos parametros  $\theta$  do modelo, eles sao chamados de hiper parâmetros do algoritmo. De posse desses elementos, da rede neural e do conjunto de dados, a função perda e as atualizações dos parâmetros são computados, os parâmetros são atualizados de acordo com regras do tipo da Eq. 4.22 e esse procedimento é repetido até o desempenho atingir seu máximo, ou mínimo, dependendo de como a função perda foi construída.

## 5 ASPECTOS DO MÉTODO VARIACIONAL NA MECÂNICA QUÂNTICA

A partir do início do século XX, a equação de Schrödinger tem sido uma das principais fontes de pesquisa para desenvolver o entendimento científico das leis da física na escala microscópica. Entretanto, resolvê-la se mostrou tão problemático, que somente poucos sistemas tem solução conhecida como, por exemplo, a partícula livre, o oscilador harmônico e o potencial coulombiano. No que diz respeito a modelos realistas, os quais geralmente possuem interações entre muitos corpos, torna-se ainda mais complicado, ou até impossível, resolver a equação analiticamente. De fato, esses problemas são ditos pertencer a classe NP-*hard* [103]. Todavia, existem dois métodos bastante genéricos e que são ferramentas úteis para determinar aproximadamente os níveis de energia e funções de onda de espectros discretos, nomeadamente, a teoria de pertubação e o método variacional. Em particular, este último será o objeto de estudo deste capítulo.

#### 5.1 Princípio variacional e o método de Ritz

O princípio por trás do método variacional pode ser exposto como a seguir. Considere uma equação a ser estudada e um espaço de funções  $\mathfrak{F}$  completo. Além disto, seja uma função arbitrária  $\psi$  pertencente ao espaço  $\mathfrak{F}$ . Agora, suponha que as soluções da equação sob estudo são aquelas funções tais que, para um dado funcional  $Q[\psi]$ , ele é estacionário, isto é:

$$\delta Q[\psi] = 0 . \tag{5.1}$$

O objetivo do método é encontrar soluções da equação acima em um subespaço de funções  $\mathfrak{F}'$  mais restrito, ou seja, contido no espaço original completo  $\mathfrak{F}$ . Esse procedimento é chamado método variacional de Ritz [65], onde a Eq. 5.1 representa um cálculo de variações, além de ser conhecido como o princípio variacional. Particularmente, na mecânica quântica, o funcional apropriado é o valor esperado do hamiltoniano, como será mostrado a seguir. Vale apontar também que, fundamentalmente, o método de Ritz é um problema de otimização do funcional Q. Nesse cenário, a escolha do subespaço é importante por dois motivos, o problema de otimização deve ser resolvível ao menos numericamente nesse subespaço, ou não faria sentido usar o método, e a escolha do subespaço deve ser feita cuidadosamente para que uma solução aproximada esteja em  $\mathfrak{F}'$  e seja qualitativamente próxima à solução exata. Para mostrar que o valor esperado do hamiltoniano respeita o princípio variacional, assuma um sistema genérico, governado pelo hamiltoniano  $\mathcal{H}$ , o qual possui níveis de energia discretos e não degenerados, sem perda de generalidade. Os níveis de energia podem ser escritos como  $E_n$  e os respectivos autoestados como  $|\varphi_n\rangle$ , onde *n* pertence aos números naturais incluindo o zero. Em especial, n = 0 define o estado fundamental, em outras palavras, o nível de menor energia. É claro que, como apontado anteriormente, essas soluções não são conhecidas, entretanto, existem algumas propriedades importantes que podem ser derivadas sem seu conhecimento prévio. Seja  $|\psi\rangle$  um estado arbitrário do espaço de estados total gerado pelos autoestados  $|\varphi_n\rangle$ , então, ele pode ser expandido em termos desses autoestados da seguinte maneira:

$$|\psi\rangle = \sum_{n} c_n |\varphi_n\rangle \quad , \tag{5.2}$$

onde  $c_n$  são coeficientes complexos que determinam a superposição com o *n*-ésimo estado. Essa combinação linear de estados compõe o cálculo da energia média do sistema, podendo ser escrita usando a notação de Dirac como:

$$\langle \mathcal{H} \rangle \left[ \psi \right] = \frac{\langle \psi \left| \mathcal{H} \right| \psi \rangle}{\langle \psi \left| \psi \right\rangle} = \frac{\sum\limits_{n,m} c_m^* c_n E_n \left\langle \varphi_m \left| \varphi_n \right\rangle}{\sum\limits_{n,m} c_m^* c_n \left\langle \varphi_m \left| \varphi_n \right\rangle} = \frac{\sum\limits_n |c_n|^2 E_n}{\sum\limits_n |c_n|^2} .$$
(5.3)

Uma vez que a energia do estado fundamental  $E_0$  é sempre menor ou igual a energia  $E_n$ para qualquer estado n, então, o quociente da fração acima é sempre maior ou igual a  $E_0 \langle \psi | \psi \rangle$ , sendo igual se, e somente se, o estado  $|\psi\rangle$  for exatamente o estado fundamental. Portanto, segue que a desigualdade a seguir é sempre válida,

$$\langle \mathcal{H} \rangle \left[ \psi \right] \ge E_0 \ . \tag{5.4}$$

Por consequência, é garantido que para qualquer função de onda escolhida, a energia média associada nunca terá um valor inferior ao valor da energia do estado fundamental, implicando necessariamente que o funcional  $\langle \mathcal{H} \rangle [\psi]$  tem um mínimo global no espaço de estados. Como resultado, se existe um mínimo, também há fortes evidências de existir propriedades estacionárias para esse funcional. Considerando que o objetivo é encontrar uma aproximação para as soluções, em particular, o estado fundamental, o valor esperado do hamiltoniano parece se encaixar perfeitamente no método variacional de Ritz. De fato, o teorema de Ritz garante que [88]:

Para que o valor médio do hamiltoniano seja estacionário, é uma condição necessária e suficiente que o estado  $|\psi\rangle$  seja um auto estado discreto de  $\mathcal{H}$ , isto é, o valor esperado  $\langle \mathcal{H} \rangle [\psi]$  é estacionário nas vizinhanças de seus auto estados discretos:

$$|\psi\rangle \ \acute{e} \ auto \ estado \ de \ \mathcal{H} \iff \delta \langle \mathcal{H} \rangle = 0 \ .$$
 (5.5)

Para provar o teorema, considere uma versão ligeiramente modificada da Eq. 5.3, mas totalmente equivalente, que pode ser escrita como

$$\langle \psi | \psi \rangle \langle \mathcal{H} \rangle = \langle \psi | \mathcal{H} | \psi \rangle \quad . \tag{5.6}$$

Neste momento, para computar uma variação infinitesimal do valor médio, será tomado o diferencial dos dois lados da equação acima. Em especial, o vetor de estado  $|\psi\rangle$  sofrerá uma variação arbitrária infinitamente pequena de  $|\delta\psi\rangle$ , resultando em:

$$\langle \psi | \psi \rangle \,\delta \,\langle \mathcal{H} \rangle + \left( \langle \psi | \delta \psi \rangle + \langle \delta \psi | \psi \rangle \right) \,\langle \mathcal{H} \rangle = \langle \psi | \mathcal{H} | \,\delta \psi \rangle + \langle \delta \psi | \mathcal{H} | \,\psi \rangle \quad. \tag{5.7}$$

Isolando o primeiro termo do lado esquerdo da igualdade e definindo o vetor de estado  $|\varphi\rangle = (\mathcal{H} - \langle \mathcal{H} \rangle) |\psi\rangle$ , é imediato ver que a Eq. 5.7 pode ser reescrita em termos desse novo estado, alcançando a expressão simplificada:

$$\delta \langle \mathcal{H} \rangle \langle \psi | \psi \rangle = \langle \varphi | \delta \psi \rangle + \langle \delta \psi | \varphi \rangle \quad . \tag{5.8}$$

Quando o vetor de estado arbitrário  $|\psi\rangle$  é um auto estado do hamiltoniano, o valor esperado da energia é o seu próprio autovalor, consequentemente, o estado  $|\varphi\rangle$  é identicamente nulo. Portanto, é imediato que, se o estado  $|\psi\rangle$  é um auto estado de  $\mathcal{H}$ , então o valor esperado da energia é estacionário, ou  $\delta \langle \mathcal{H} \rangle = 0$ . Assim, fica provado uma parte do teorema de Ritz. Não obstante, a Eq. 5.8 deve ser satisfeita para qualquer vetor de estado infinitesimal  $|\delta\psi\rangle$ .

Para provar a segunda parte do teorema, o estado arbitrário infinitamente pequeno pode ser escolhido como  $|\delta\psi\rangle = \delta\lambda |\varphi\rangle$ , onde  $\delta\lambda$  é um número real infinitesimal. Considerando que na segunda parte a hipótese é que o valor médio da energia é estacionário, isto é,  $\delta \langle \mathcal{H} \rangle = 0$ , isso leva a conclusão de que, independente do número infinitesimal arbitrário  $\delta\lambda$ , a norma de  $|\varphi\rangle$  é sempre zero, logo, o próprio estado só pode ser o estado nulo e, consequentemente,

$$\mathcal{H} \left| \psi \right\rangle = \left\langle \mathcal{H} \right\rangle \left| \psi \right\rangle \ , \tag{5.9}$$

que é uma equação de autovalores e, portanto, prova que  $|\psi\rangle$  é um auto estado do hamiltoniano.

Antes de deixar este tópico, há alguns detalhes subjacentes importantes de se mencionar. Primeiramente, para estados degenerados, o teorema se mantém, porém com um aspecto de diferença, o estado  $|\psi\rangle$  poderá ser uma combinação linear de estados com o mesmo valor de energia, ou seja, ele vai pertencer ao subespaço gerado pelos autovetores de um particular nível de energia. Em segundo lugar, procurar uma aproximação do estado fundamental na mecânica quântica pelo método variacional de Ritz vai consistir em, resumidamente, escolher uma família de estados  $|\psi_{\theta}\rangle$ , os quais dependem de um conjunto de parâmetros variacionais  $\theta$ . Em seguida, encontrar as soluções para o funcional  $\delta \langle \mathcal{H} \rangle [|\psi_{\theta} \rangle] = 0$ , ou seja, minimizar o valor esperado da energia com relação aos parâmetros variacionais. Note que, como dito anteriormente, essa família de funções pertencem à um espaço de estados mais restrito do que aquele que contém as soluções da equação de Schrödinger. Por consequência, encontrar esses estados ótimos que minimizam o valor do funcional da energia irá fornecer a melhor aproximação para a solução dentro desse subespaço de acordo com o princípio variacional.

Esses vetores  $|\psi_{\theta}\rangle$  são chamados de estados tentativos ou, quando representado no espaço de coordenadas, funções de onda tentativas. Elas são construídas com base em visões e percepções de propriedades do sistema e do hamiltoniano. Por essa razão, a sobreposição com o estado desejado é usualmente apreciável. Mesmo assim, o método nunca alcança a solução exata, porque o espaço de funções gerado por esses estados tentativos não é completo, diferentemente do espaço gerado pelas autofunções do sistema, que é completo. Para exemplificar, considere um sistema de três níveis de energia discretos e não degenerados. Agora, suponha que o estado tentativo escolhido possua um parâmetro variacional real  $\theta$  e a expansão desse vetor em termos dos auto estados do sistema é escrito como:

$$|\psi_{\theta}\rangle \propto (1+|\theta|) |\varphi_{0}\rangle + \cos\theta |\varphi_{1}\rangle + \sin\theta |\varphi_{2}\rangle .$$
(5.10)

Observe que, independente do valor do parâmetro, o estado tentativo nunca será o estado fundamental exatamente. Ao propor uma função tentativa, geralmente não se sabe como os parâmetros vão influenciar nos coeficientes da expansão, podendo inclusive correlacionar os coeficientes, como neste exemplo. Dessa maneira, fica mais tangível o fato de o subespaço gerado pelas funções tentativas ser incompleto. Um outro exemplo interessante é o caso da partícula no poço de potencial infinito. O estado fundamental exato é dado por  $\sin(\pi x)$ para um poço de tamanho unitário. Uma boa escolha de função de onda tentativa seria o polinômio x(1-x), que pode ser reescrito em termos de polinômios de Legendre como  $-2/_{3}P_{2}(x) + P_{1}(x) - 1/_{3}P_{0}(x)$ . É pertinente lembrar que a função seno pode ser expandida em uma soma infinita de polinômios de Legendre, isto é, o conjunto total desses polinômios consegue gerar as soluções exatas. Entretanto, a função tentativa proposta usa apenas três polinômios do conjunto completo e, mais uma vez, o subespaço escolhido demonstra ser insuficiente para alcançar a solução exata. Vale ressaltar também que os estados tentativos devem satisfazer as condições de contorno do problema, do contrário, a energia poderia violar o princípio variacional. Um exemplo desse caso seria considerar a função constante como função tentativa do poço de potencial infinito. Dessa forma, a energia calculada seria nula e, portanto, menor que a do estado fundamental. Porém, por não se anular continuamente nas paredes do poço infinito, a função não satisfaz as condições de contorno do problema.

## 5.2 Minimização do valor esperado da energia

Com o funcional da energia em mãos, a finalidade do método variacional no contexto da mecânica quântica é obter a função de onda tentativa que minimiza o valor esperado da energia. Essa tarefa é um problema de otimização, assunto recorrente no campo da matemática. Tanto na natureza, como nos estágios iniciais da civilização humana, o princípio de otimização se mostrou fundamental [104]. Uma aranha, ao sentir um mosquito ficar preso em sua teia, procura o fio tensionado para encontrar o trajeto mais curto até sua presa. Ao soltar bolas do alto de uma montanha, elas tendem a descer pelos caminhos mais rápidos até a base. Na matemática, o problema de otimização consiste em selecionar o melhor elemento de um conjunto de alternativas, que satisfaça um particular critério. Formalmente, otimização matemática é o processo de [105]:

- i. definição da função fa ser otimizada, e
- ii. determinação da solução para

$$\min_{\text{c.r.a} \boldsymbol{\theta}} \left[ f(\boldsymbol{\theta}) \right], \ \boldsymbol{\theta} = \left[ \theta_1, \theta_2, ..., \theta_n \right]^T \in \mathbb{R}^n \ , n \in \mathbb{N} \ , \tag{5.11}$$

onde c.r.a significa "com respeito à", e sujeito aos vínculos

$$g_i(\boldsymbol{\theta}) \leq 0, \ i = 1, 2, ..., q, \ q \in \mathbb{N} ,$$
  
$$h_j(\boldsymbol{\theta}) = 0, \ j = 1, 2, ..., p, \ p \in \mathbb{N} .$$

Dentre as diversas formas de programação não linear, outro nome pelo qual a otimização matemática é conhecida, os procedimentos baseados em gradiente são um dos mais utilizados. Dentre eles, o método do gradiente descendente é um dos mais intuitivos e foi primeiramente proposto por Cauchy [106]. Dado que o gradiente de uma função pode ser interpretado como a direção de crescimento naquele determinado ponto avaliado, intuitivamente, basta ir no sentido contrário para procurar o mínimo. Assim sendo, o método do gradiente descendente é mais formalmente descrito como a seguir.

#### Método do gradiente descendente

Seja  $u = f(\boldsymbol{\theta})$  uma função multivariada contínua e não negativa em seu domínio. Perceba que a condição da função não ser negativa é equivalente a garantir que ela possui pelo menos um mínimo global, em particular, considere que esse mínimo global é em u = 0, sem perda de generalidade. Então, para encontrar os valores de  $\boldsymbol{\theta}$  que satisfazem a equação u = 0, basta seguir uma sequência tal que a função decresça até que ela se anule. Tome um particular valor  $\boldsymbol{\theta}_i$  de parâmetros como inicial e chame o valor da função avaliada nesse ponto de  $u_i$ . Aqui, como o gradiente descendente é um processo iterativo, *i* vai representar o número da iteração do método. Então, realizando pequenos incrementos nesses parâmetros, o valor atualizado da função pode ser escrito aproximadamente como:

$$f(\boldsymbol{\theta}_i + \boldsymbol{\delta}) \simeq u_i + \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}_i) \cdot \boldsymbol{\delta}$$
 (5.12)

Tomando essa pequena variação  $\boldsymbol{\delta}$  proporcional ao gradiente da função no ponto original, isto é,  $\boldsymbol{\delta} = -\eta \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}_i)$ , a função avaliada nesse novo ponto será aproximadamente:

$$u_{i+1} = f(\boldsymbol{\theta}_{i+1}) = f(\boldsymbol{\theta}_i - \eta \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}_i)) \simeq u_i - \eta |\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}_i)|^2 .$$
(5.13)

Isso implica que, para um valor  $\eta$  pequeno o suficiente, cada iteração do método reduz o valor do modelo f, gerando uma sequência monotônica decrescente de valores  $u_i$ . Note que  $\eta$  é a taxa de aprendizagem mencionada no Cap. 4. Assim, esse procedimento é repetido até uma certa condição de precisão ser satisfeita, como por exemplo  $f(\boldsymbol{\theta}_i) < \varepsilon$ , onde  $\varepsilon$  é um número que determina o quão numericamente próximo ao resultado exato se deseja alcançar. Dessa forma, o procedimento de otimização se resume ao seguinte algoritmo:

Algoritmo 5.1 Pseudo algoritmo para o método do gradiente descendente.
$oldsymbol{ heta} oldsymbol{ heta} oldsymbol{ heta}_0$
while $f(\boldsymbol{\theta}) > \varepsilon  \operatorname{\mathbf{do}}$
$\boldsymbol{\delta} \leftarrow -\eta \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta})$
$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \boldsymbol{\delta}$
end while

Usualmente, o hiper parâmetro  $\eta$  é conhecido como a taxa de aprendizado no contexto do aprendizado de máquina. Valores muito pequenos tornam o processo lento e ineficiente, além de haver riscos maiores do algoritmo ficar preso em mínimos locais. Por outro lado, valores muito altos tornam a aproximação da Eq. 5.12 inadequada, impedindo que a sequência decrescente seja gerada de maneira consistente. Contudo, o cálculo do gradiente do valor esperado do hamiltoniano com relação aos parâmetros variacionais da função de onda pode parecer, em alguma medida, inocente. Entretanto, é necessário muito cuidado para computar essas quantidades.

#### Cálculo do gradiente da energia

Para computar essas derivadas, considere um sistema composto de N partículas modelado pelo hamiltoniano  $\mathcal{H}$  e um conjunto de estados tentativos  $\psi_{\theta}$ , onde  $\theta$  são parâmetros variacionais. A energia escrita na representação de coordenadas é dada por

$$E(\boldsymbol{\theta}) = \frac{\int d\tau \ \psi_{\boldsymbol{\theta}}^*(R) \mathcal{H} \psi_{\boldsymbol{\theta}}(R)}{\int d\tau \ |\psi_{\boldsymbol{\theta}}(R)|^2} , \qquad (5.14)$$

onde  $R = {\mathbf{r}_1, ..., \mathbf{r}_N}$  é o conjunto de todas as coordenadas cartesianas das partículas e  $d\tau$  é o diferencial de volume das coordenadas desses corpos. Para simplificar a notação, a dependência em  $\boldsymbol{\theta}$  da função de onda será omitida daqui em diante. Perceba que, tanto o quociente, como o dividendo dependem dos parâmetros variacionais e, por consequência, ambos devem ser derivados. Além disso, é conveniente definir a seguinte densidade de probabilidade normalizada:

$$p(R) = \frac{|\psi(R)|^2}{\int d\tau \ |\psi(R)|^2} \ . \tag{5.15}$$

Reescrever as equações em termos dessa densidade de probabilidade vai se mostrar essencial, uma vez que o método para estimar essas integrais multidimensionais será a integração de Monte Carlo. Outra definição importante é a quantidade dada por:

$$E_L(R) = \frac{\mathcal{H}\psi(R)}{\psi(R)} , \qquad (5.16)$$

também conhecida como energia local. Com base nessas definições, os primeiros passos para o cálculo do gradiente se resumem a aplicar a regra da divisão para derivadas e, na sequência, inverter a ordem das integrais com a derivada, já que as duas operações atuam em variáveis distintas. Ao realizar essas operações, chega-se em:

$$\nabla_{\boldsymbol{\theta}} E(\boldsymbol{\theta}) = \frac{\int d\tau \ \nabla_{\boldsymbol{\theta}} \left(\psi^* \mathcal{H}\psi\right)}{\int d\tau \ |\psi|^2} - \frac{\int d\tau \ \nabla_{\boldsymbol{\theta}} |\psi|^2 \int d\tau \ \psi^* \mathcal{H}\psi}{\left(\int d\tau \ |\psi|^2\right)^2} , \qquad (5.17)$$

onde a dependência em R da função de onda também foi omitida para aumentar a clareza na leitura. Pretendendo chegar em uma expressão mais eficaz do ponto de vista computacional, os dois termos do gradiente serão calculados separadamente.

Então, o cálculo do primeiro termo se dá como a seguir:

$$\frac{\int d\tau \ \nabla_{\boldsymbol{\theta}} \left(\psi^* \mathcal{H}\psi\right)}{\int d\tau \ |\psi|^2} = \frac{\int d\tau \ \left(\nabla_{\boldsymbol{\theta}}\psi^*\right) \mathcal{H}\psi}{\int d\tau \ |\psi|^2} + \frac{\int d\tau \ \psi^* \mathcal{H} \left(\nabla_{\boldsymbol{\theta}}\psi\right)}{\int d\tau \ |\psi|^2}$$
(5.18a)

$$= \frac{\int d\tau \ |\psi|^2 \left(\frac{\nabla_{\theta}\psi^*}{\psi^*}\right) \left(\frac{\mathcal{H}\psi}{\psi}\right)}{\int d\tau \ |\psi|^2} + \frac{\int d\tau \ |\psi|^2 \left(\frac{\psi^*\mathcal{H}}{\psi^*}\right) \left(\frac{\nabla_{\theta}\psi}{\psi}\right)}{\int d\tau \ |\psi|^2}$$
(5.18b)

$$= 2 \int d\tau \ p(R) E_L(R) \left( \nabla_{\boldsymbol{\theta}} \ln |\psi| \right)$$
(5.18c)

$$= \int d\tau \ p(R) E_L(R) \left( \nabla_{\boldsymbol{\theta}} \ln |\psi|^2 \right) \ . \tag{5.18d}$$

Um aspecto interessante da mecânica quântica reside na aplicação de um operador hermitiano na função de onda, podendo ser aplicado tanto à direita como à esquerda, lembrando que neste último caso é preciso aplicar o operador conjugado. Como o hamiltoniano e seu conjugado hermitiano são iguais para os sistemas considerados, então é equivalente aplicar  $\mathcal{H}$  em ambos os lados. Essa propriedade foi usada na passagem da Eq. 5.18a para a Eq. 5.18b, isto é, no segundo termo da Eq. 5.18b, o hamiltoniano é aplicado à esquerda. Isso evita o acúmulo de operações em uma única função onda, caso contrário, seria necessário aplicar o hamiltoniano no gradiente da função de onda com relação aos parâmetros variacionais. Outro ponto é que, na passagem da Eq. 5.18b para a Eq. 5.18c, foi considerado que a função de onda é real. Essa suposição é razoável para os objetivos deste trabalho porque os hamiltonianos examinados tem simetria de reversão temporal.

Já o segundo termo da Eq. 5.17 pode ser expresso como:

$$\frac{\int d\tau \ \nabla_{\boldsymbol{\theta}} |\psi|^2 \int d\tau \ \psi^* \mathcal{H} \psi}{\left(\int d\tau \ |\psi|^2\right)^2} = \frac{\int d\tau \ \left[\psi^* \left(\nabla_{\boldsymbol{\theta}} \psi\right) + \left(\nabla_{\boldsymbol{\theta}} \psi^*\right) \psi\right]}{\int d\tau \ |\psi|^2} \times \frac{\int d\tau \ \psi^* \mathcal{H} \psi}{\int d\tau \ |\psi|^2} \tag{5.19a}$$

$$= \frac{\int d\tau \ 2 \ |\psi|^2 \left(\nabla_{\theta} \ln |\psi|\right)}{\int d\tau \ |\psi|^2} \times \frac{\int d\tau \ \psi^* \mathcal{H} \psi}{\int d\tau \ |\psi|^2}$$
(5.19b)

$$= \int d\tau \ p(R) \left( \nabla_{\boldsymbol{\theta}} \ln |\psi|^2 \right) \int d\tau \ p(R) E_L(R) \ . \tag{5.19c}$$

Por fim, a expressão final para o gradiente do valor esperado do hamiltoniano com relação aos parâmetros variacionais é escrita como [31]:

$$\nabla_{\boldsymbol{\theta}} E(\boldsymbol{\theta}) = \int d\tau \ p(R) \left[ E_L(R) - \int d\tau' p(R') E_L(R') \right] \nabla_{\boldsymbol{\theta}} \ln |\psi_{\boldsymbol{\theta}}(R)|^2$$
(5.20a)

$$= \operatorname{Cov}_{p(R)} \left[ E_L(R), \nabla_{\boldsymbol{\theta}} \ln |\psi_{\boldsymbol{\theta}}(R)|^2 \right] .$$
(5.20b)

Avaliando do ponto de vista da teoria de probabilidades, é seguro dizer que o gradiente do funcional da energia é a covariância amostral entre a energia local e o gradiente do logaritmo natural da função de onda ao quadrado. Especificamente, a densidade de probabilidade utilizada na avaliação dessa quantidade é a densidade de probabilidade p(R) previamente definida.

Com essa quantidade em mãos, pode-ser aplicar um dos métodos de otimização baseado em gradientes para minimizar a energia. Também, como declarado anteriormente, as integrais envolvidas no cálculo do gradiente do valor esperado do hamiltoniano serão estimadas através do método de integração de Monte Carlo. Como esse método de integração estima os resultados das integrais através de médias de amostras geradas estocasticamente da densidade de probabilidade, esses gradientes terão, similarmente, um aspecto estocástico. Entretanto, como já comentado no Cap. 4, alguns métodos de minimização mais rebuscados também utilizam quantidades relacionadas com a correlação de parâmetros. Em particular, dentre eles está o método KFAC, o qual detêm diversas similaridades com o método de reconfiguração estocástica [107], que por sua vez é amplamente empregado no contexto de Monte Carlo variacional, o qual será explorado na Sec. 5.4.

## 5.3 Método de reconfiguração estocástica

Com o intuito de entender como, a partir da mecânica quântica, pode-se chegar em um método de minimização de segunda ordem semelhante ao KFAC, considere o seguinte raciocínio. Primeiro, uma maneira de projetar o estado de menor energia é analisar a equação de Schrödinger sob a ótica do tempo imaginário, em outras palavras, efetuar a mudança de variável  $\beta = it$ . Isto posto, a equação transforma-se em:

$$-\hbar \frac{\partial |\psi(\beta)\rangle}{\partial \beta} = \mathcal{H} |\psi(\beta)\rangle \quad . \tag{5.21}$$

A suposição de que o hamiltoniano independe da variável temporal é razoável para diversos sistemas, particularmente, isso é verdade para os modelos estudados. Então, uma maneira de expressar a solução formal da igualdade acima é dada pelo estado

$$|\psi(\beta)\rangle = e^{-\frac{\mathcal{H}\beta}{\hbar}} |\psi(0)\rangle \quad , \tag{5.22}$$

onde  $|\psi(0)\rangle$  é o estado no tempo imaginário igual à zero. Uma vez que o estado fundamental tem a menor energia do sistema, é possível realizar um deslocamento no espectro de energias, tal que o estado fundamental torne-se o zero de energia, isto é,  $E_0 = 0$ . Fazendo isto e supondo que as soluções da equação independente do tempo sejam os auto estados  $|\varphi_n\rangle$  com autoenergias  $E_n$ , o estado inicial pode ser expandido e, ao tomar o limite de longos tempos imaginários, obtêm-se:

$$\lim_{\beta \to \infty} |\psi(\beta)\rangle = \lim_{\beta \to \infty} \sum_{n} c_n \ e^{-\frac{E_n \beta}{\hbar}} |\varphi_n\rangle = c_0 |\varphi_0\rangle \ .$$
(5.23)

Como o estado no tempo zero é arbitrário, considere que ele faça parte de uma família de estados  $|\psi_{\theta}\rangle$ , onde  $\theta$  é um conjunto de parâmetros variacionais. A coleção de estados gerados por todos os possíveis valores de  $\theta$  gera um subespaço de estados  $\mathfrak{F}'$ , que está contido no espaço completo  $\mathfrak{F}$  gerado pelos auto estados  $|\varphi_n\rangle$ . Perceba que a projeção do estado fundamental, no limite de longos tempos imaginários, ocorre no espaço completo de estados, ou seja, executar esta estratégia num subespaço como  $\mathfrak{F}'$  irá projetar o estado mais próximo do estado fundamental nesse subespaço, similarmente ao objetivo do método variacional.

Com isto em mente, considere o propagador  $e^{-\delta\beta\hat{\mathcal{H}}}$  para tempos  $\delta\beta$  curtos, onde, tanto o intervalo de tempo imaginário, quanto o hamiltoniano  $\hat{\mathcal{H}}$  estão em unidades adimensionais. Esse propagador pode ser aproximado em primeira ordem por  $(\mathbb{1} - \delta\beta \hat{\mathcal{H}})$ . Dessa forma, considere a equação iterativa a seguir baseada na Eq. 5.22:

$$|\psi_{k+1}\rangle = \left(\mathbb{1} - \delta\beta \ \hat{\mathcal{H}}\right) |\psi_k\rangle \ , \tag{5.24}$$

onde  $|\psi_k\rangle = |\psi_{\theta_k}\rangle$  é o estado tentativo da iteração k, cujos parâmetros variacionais são  $\theta_k$ .

Vale ressaltar que, ao efetuar um grande número de iterações, isso é equivalente a aplicar o propagador repetidas vezes, de modo que, ao final do processo, isso é equivalente a ter aplicado o propagador para tempos longos no estado inicial. Portanto, esse procedimento iterativo vai conduzir o estado inicial para a melhor solução para o estado fundamental no subespaço em questão.

É razoável expandir o estado  $|\psi_{k+1}\rangle$  em uma série de Taylor multivariada até primeira ordem, uma vez que, para um valor suficientemente pequeno de  $\delta\beta$ , não é esperada uma modificação severa no estado entre iterações subsequentes. Assim, a Eq. 5.24 também pode ser escrita como:

$$|\psi_{k+1}\rangle = \sum_{j=0}^{n_{\theta}} \mathcal{O}^{j} |\psi_{k}\rangle \,\Delta\theta_{j} , \qquad (5.25)$$

onde  $\mathcal{O}^{j} = \partial/\partial\theta_{j}$  é o operador que calcula a derivada do estado com relação ao parâmetro  $\theta_{j}$ ,  $n_{\theta}$  é o número total de parâmetros e  $\Delta\theta_{j}$  é uma pequena variação. Juntamente, o operador para j = 0, isto é,  $\mathcal{O}^{0} = 1$  será a matriz identidade. Na sequência, é interessante calcular o produto interno entre o estado  $\langle \psi_{k} | \mathcal{O}^{i}$  e as duas versões do estado  $|\psi_{k+1}\rangle$ . Consequentemente, essas quantidades são resumidas nas seguintes expressões:

$$\left\langle \psi_{k} \left| \mathcal{O}^{i} \right| \psi_{k+1} \right\rangle = \left\langle \psi_{k} \left| \mathcal{O}^{i} (\mathbb{1} - \delta \beta \ \hat{\mathcal{H}}) \right| \psi_{k} \right\rangle$$
(5.26a)

$$=\sum_{j=0}^{n} \left\langle \psi_k \left| \mathcal{O}^i \mathcal{O}^j \right| \psi_k \right\rangle \Delta \theta_j \quad .$$
 (5.26b)

Por construção, os estados  $|\psi_k\rangle$  não são normalizados, mesmo que o estado inicial seja. Por isso, é importante dividir as equações acima pela norma de  $|\psi_k\rangle$ . Além disso, dado que o objetivo é encontrar o valor dos parâmetros variacionais para a próxima iteração, isso equivale a encontrar o valor das variações  $\Delta \theta_i$ . Por consequência, as equações acima formam um sistema linear com  $n_{\theta} + 1$  equações, cujas incógnitas a serem determinadas são os valores de  $\Delta \theta_i$ . Assim, para a equação i = 0, a variação fica:

$$\Delta \theta_0 = \langle \mathbb{1} - \delta \beta \cdot \mathcal{H} \rangle - \sum_{j=1}^{n_\theta} \left\langle \mathcal{O}^j \right\rangle \Delta \theta_j , \qquad (5.27)$$

onde  $\langle \cdot \rangle$  é uma simplificação da notação  $\langle \psi_k | \cdot | \psi_k \rangle / \langle \psi_k | \psi_k \rangle$ . Em seguida, substituindo o valor encontrado para  $\Delta \theta_0$  nas equações onde  $i \neq 0$ , chega-se em:

$$\left\langle \mathcal{O}^{i} \right\rangle \left( \left\langle \mathbb{1} - \delta\beta \cdot \mathcal{H} \right\rangle - \sum_{j=1}^{n_{\theta}} \left\langle \mathcal{O}^{j} \right\rangle \Delta\theta_{j} \right) + \sum_{j=1}^{n_{\theta}} \left\langle \mathcal{O}^{i} \mathcal{O}^{j} \right\rangle \Delta\theta_{j} = \left\langle \mathcal{O}^{i} (\mathbb{1} - \delta\beta \cdot \mathcal{H}) \right\rangle .$$
(5.28)

Por último, agrupando cuidadosamente os valores esperados e separando de um lado da equação as quantidades que multiplicam as incógnitas  $\Delta \theta_i$  e do outro as demais, o sistema linear adquire o aspecto:

$$\sum_{j=1}^{n_{\theta}} \left( \left\langle \mathcal{O}^{i} \mathcal{O}^{j} \right\rangle - \left\langle \mathcal{O}^{i} \right\rangle \left\langle \mathcal{O}^{j} \right\rangle \right) \Delta \theta_{j} = -\delta \beta \left( \left\langle \mathcal{O}^{i} \mathcal{H} \right\rangle - \left\langle \mathcal{O}^{i} \right\rangle \left\langle \mathcal{H} \right\rangle \right)$$
(5.29)

Não é difícil notar a semelhança do lado esquerdo da equação com uma multiplicação de matrizes, em particular, entre uma matriz quadrada e uma matriz coluna. De fato, a matriz das covariâncias das derivadas com relação aos parâmetros variacionais, definida por:

$$\mathcal{M}_{ij} = \left\langle \left( \mathcal{O}^i - \left\langle \mathcal{O}^i \right\rangle \right) \left( \mathcal{O}^j - \left\langle \mathcal{O}^j \right\rangle \right) \right\rangle , \qquad (5.30)$$

pode ser usada para expressar os coeficientes desse sistema linear. Em contrapartida, o lado direito da Eq. 5.29 é proporcional ao gradiente do valor esperado da energia, que está presente na Eq. 5.20b. Logo, reescrevendo a Eq. 5.29 em termos dessas quantidades, a nova expressão assume a forma:

$$\sum_{j} \mathcal{M}_{ij} \Delta \theta_j = -\frac{1}{2} \delta \beta \frac{\partial E(\boldsymbol{\theta})}{\partial \theta_i} .$$
 (5.31)

Agora, de posse desse sistema linear com  $n_{\theta}$  equações e  $n_{\theta}$  incógnitas, basta resolvê-lo para encontrar as variações nos parâmetros e, dessa forma, achar qual será o estado  $|\psi_{k+1}\rangle$ da iteração seguinte. Já que a solução de um sistema linear é a multiplicação da matriz inversa dos coeficientes pela matriz coluna dos termos constantes, a equação iterativa para os parâmetros variacionais pode ser escrita como:

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \frac{1}{2} \delta \beta \ \boldsymbol{\mathcal{M}}^{-1} \cdot \nabla_{\boldsymbol{\theta}} E \ . \tag{5.32}$$

Existem formas de estimar o valor de  $\delta\beta$ , entretanto, a performance do método é pouco afetada ao manter essa quantidade constante [108], o que torna essa abordagem mais eficaz, já que o custo computacional para estimá-lo é considerável. Finalmente, assim como para o cálculo de  $\nabla_{\theta} E$ , a matriz das covariâncias será calculada usando integração de Monte Carlo. Para isso, é necessário escrever o seguinte valor esperado na representação de coordenadas:

$$\langle \mathcal{O}^i \rangle = \int d\tau \ p(\mathbf{x}) \ \frac{\partial \ln \psi_k(\mathbf{x})}{\partial \theta_i} \ .$$
 (5.33)

Ainda na computação dos elementos da matriz, nos termos que correlacionam duas derivadas, os operadores  $\mathcal{O}^i \in \mathcal{O}^j$  serão aplicados à esquerda e à direita, respectivamente. Novamente, isso é feito para evitar o acúmulo de operações em um única função de onda, levando à seguinte forma desses termos:

$$\left\langle \mathcal{O}^{i}\mathcal{O}^{j}\right\rangle = \int d\tau \ p(\mathbf{x}) \ \frac{\partial \ln \psi_{k}(\mathbf{x})}{\partial \theta_{i}} \frac{\partial \ln \psi_{k}(\mathbf{x})}{\partial \theta_{j}} \ .$$
 (5.34)

Um detalhe computacional pertinente é o cancelamento catastrófico [109]. Esse problema ocorre quando as operações multiplicativas estão sujeitas a erros de arredondamento. Ao subtrair quantidades com valores próximos, isso pode causar erros relativos tão altos que todos os dígitos do resultado podem ser insignificantes, em razão dos dígitos significativos se cancelarem, sobrando apenas erros de arredondamento. Ao calcular a matriz de covariâncias da forma expressa na Eq. 5.29, ela está suscetível à esse cancelamento. Ao contrário, no formato da Eq. 5.30, os elementos são livres desse problema, porém exigem mais armazenamento de memória. De maneira geral, a primeira abordagem é suficiente para maioria dos casos tratados.

A conexão do método de reconfiguração estocástica com o método de aproximação da curvatura com fatorização de Kronecker se dá na identificação da matriz de covariâncias com a matriz de Fisher. Essencialmente, a diferença entre elas reside no fato de que uma é calculada empregando uma densidade de probabilidade normalizada e a outra não. Dessa forma, é possível provar que elas são iguais a menos de um fator constante, como demonstrado no Ap. C.

#### 5.4 Método variacional de Monte Carlo

Com o intuito de unir os aspectos do princípio variacional, dos métodos de otimização apresentados na seções anteriores e encontrar uma representação para o estado fundamental de sistemas de muitos corpos quânticos, o método variacional de Monte Carlo se apresenta como um candidato para alcançar essa tarefa. Através da utilização de cadeias de Markov, a integração de Monte Carlo detalhada no Ap. A é aplicada para estimar as integrais [110] relevantes no processo de determinação dos parâmetros da função tentativa que minimizam o valor esperado da energia.

Como um todo, o método é condensado em quatro etapas repetidas iterativamente, sendo elas a amostragem, a estimativa da energia, a estimativa dos gradientes e a atualização dos parâmetros. Um diagrama com a sequência das etapas é mostrado na Fig. 5.1. Na amostragem, um conjunto de M configurações  $\{R_i\}$  são amostradas da densidade de probabilidade  $|\psi|^2$ , sendo  $\psi$  a função tentativa em questão. Vale ressaltar que essas configurações  $R_i$  encapsulam as informações de todas as coordenadas do sistema. Além disso, a amostragem é feita através do algoritmo de Metropolis [111]. Essencialmente, o algoritmo propõe novas amostras baseadas nas atuais e aceita ou rejeita a nova amostra baseado na razão das distribuições de probabilidade da nova e da amostra atual. Isso cria uma cadeia de Markov, para *i* grande, cuja sequência de configurações  $R_i$  pode ser usada como a distribuição de probabilidade  $|\psi|^2$ .



Figura 5.1: Diagrama das etapas do processo de otimização empregadas no método variacional de Monte Carlo. Cada execução das 4 etapas é uma iteração e elas são realizadas repetidamente até a minimização da energia.

Portanto, a partir das M configurações geradas e amostradas de acordo com  $\psi^2$ , a integral da Eq. 5.14 pode ser estimada através de uma média da energia local sobre essas configurações, isto é,

$$E = \frac{1}{M} \sum_{i} E_L(R_i) . \qquad (5.35)$$

Esse procedimento é conhecido como integração ou quadratura de Monte Carlo. No contexto de machine learning, a estimativa da energia E do sistema é equivalente à função perda da tarefa a ser resolvida. Por sua vez, a tarefa consiste em obter uma representação funcional do estado fundamental de um sistema quântico de muitos corpos. Para atingir esse objetivo, a função perda, ou seja, a energia precisa ser minimizada. Note que o procedimento de minimizar a energia é equivalente ao treinamento de redes neurais e que a função auxiliar na forma do modelo tem uma correspondência com a densidade de probabilidade  $\psi^2$ . Por outro lado, as M configurações amostradas seriam equivalentes ao conjunto de dados de treinamento, com a diferença de que esse conjunto não é conhecido a priori, ao contrário, ele é gerado a partir da densidade de probabilidade do próprio modelo. Traçando um paralelo com os comentários feitos no Cap. 4, o número de configurações Mé equivalente ao hiper parâmetro batch size, ou tamanho do batch.

Na terceira etapa, os gradientes com relação aos parâmetros variacionais da Eq. 5.20 e os elementos da matriz de Fisher  $\mathcal{F}$  da Eq. 5.30 também podem ser estimados utilizando integração de Monte Carlo e usando o mesmo conjunto de M configurações empregado para computar a estimativa da energia. Na última etapa, os parâmetros da função tentativa são atualizados de acordo com a Eq. 5.32. Além disso, variações das ferramentas de regularização, taxas de aprendizagem adaptativas e outros hiper parâmetros também podem ser incluídos nessa etapa de maneira razoavelmente simples. O conjunto dessas quatro etapas configura uma iteração do método variacional de Monte Carlo. As iterações são repetidas sequencialmente até uma quantidade de iterações pré-definida ou até a precisão desejada ser alcançada.

## 6 Construção dos estados quânticos variacionais

Para explorar a correspondência entre a função de onda tentativa no método variacional e o modelo composto por redes neurais artificiais no aprendizado de máquina, diversos trabalhos propuseram maneiras de incorporar a estrutura de redes neurais na representação de estados quânticos e aplicar o método variacional. Algumas das abordagens para representar estados quânticos empregam um tipo de aprendizado de máquina conhecido como máquinas de Boltzmann restrita [30], outras utilizam redes tensoriais [112] e redes neurais convolucionais [113]. Para aplicar modelos de *machine learning* em conjunto com o método variacional de Monte Carlo, funções tentativas baseadas em rede neurais foram propostas para estudar sistemas de spin [114], bósons [42, 115, 116] e férmions [31] e apresentaram resultados competitivos com métodos estado-da-arte em diversas configurações de sistema [25].

Aproveitando a mobilização de estudos científicos nessa direção, as funções tentativas propostas neste trabalho visam expandir essas abordagens e aplicar em sistemas ainda pouco explorados, em particular, os sistemas de aglomerados de <sup>4</sup>He e os pontos quânticos apresentados nos Cap. 2 e Cap. 3, respectivamente. Nas próximas seções, é apresentado a maior parte dos desenvolvimentos e contribuições, especialmente os relativos à elaboração das funções de onda tentativa baseadas em redes neurais e especializações para abordar e representar estados fundamentais dos sistemas de aglomerados de hélio e pontos quânticos. Com os estados tentativos propostos, uma combinação entre o aprendizado de máquina e o método variacional apresentados respectivamente nos Cap. 4 e Cap. 5 pode ser utilizada para encontrar representações de estados quânticos através de redes neurais.

#### 6.1 Explorando estados quânticos como inspiração

Para iniciar a construção da função de onda tentativa genérica, considere um sistema de partículas não interagentes. Sem a inclusão da estatística e devido à ausência de interações, o problema completo de muitas partículas pode ser desacoplado e a solução se resume a um produto tensorial de estados de uma partícula, onde esses estados são soluções do hamiltoniano de um corpo. Nesse caso, a solução completa de um sistema de N partículas não interagentes confinadas, por exemplo, em um potencial harmônico tridimensional é dado por:

$$\chi_{\alpha}(\mathbf{R}) = \prod_{i} \phi_{\alpha_{i}}(\mathbf{r}_{i}) , \qquad (6.1)$$

onde  $\alpha = \{\alpha_1, \alpha_2, ..., \alpha_N\}, \phi_{\alpha_i}$  são as soluções do oscilador harmônico e  $\alpha_i$  condensa os números quânticos associados ao estado correspondente. Cada conjunto possível de  $\alpha$  resulta em uma função distinta e as diferentes funções  $\chi_{\alpha}$  resultantes formam uma base. Recordando que, na mecânica quântica, um estado genérico é descrito como uma combinação linear de estados de uma base, uma função de onda genérica é representada na forma:

$$\psi(\mathbf{R}) = \sum_{\alpha} \omega_{\alpha} \chi_{\alpha}(\mathbf{R}) , \qquad (6.2)$$

onde  $\omega_{\alpha}$  são coeficientes. Entretanto, no cenário de um sistema formado por partículas idênticas, isto é, partículas que possuem as mesmas propriedades físicas, os valores esperados de observáveis devem se manter inalterados sob uma troca de partículas, uma vez que é inviável fazer uma observação que possa distingui-las. Do ponto de vista da teoria, essa indistinguibilidade ocorre se o hamiltoniano e todos os observáveis físicos são simétricos com respeito à troca de duas partículas. Por causa dessa propriedade, os estados da Eq. 6.1 precisam ser corrigidos, já que estes podem distinguir as partículas através da sua identificação numérica *i* e do estado  $\alpha_i$  bem definido em que estão. Como, na prática, o máximo que pode ser dito é que  $N_1$  partículas estão no estado  $\alpha_1$ ,  $N_2$  partículas estão no estado  $\alpha_2$  e similarmente para todos os outros estados, com a restrição do número total de partículas ser N, então existem  $N!/\prod_i N_i!$  combinações possíveis de estados equivalentes. Essa ambiguidade é resolvida através do postulado de simetrização, que afirma:

# Os estados de um sistema contendo N partículas idênticas são necessariamente ou todos simétricos ou todos antissimétricos com respeito às permutações das N partículas [65].

Qual das circunstâncias deve ser aplicada dependerá da natureza das partículas. Dentre elas, são chamadas de bósons aquelas que possuem estados simétricos e constata-se que o spin dessas partículas é inteiro, como por exemplo os fótons. Por outro lado, as partículas que possuem estados antissimétricos são chamadas de férmions e verifica-se que seu spin é semi-inteiro, como por exemplo os prótons, nêutrons e elétrons. O caminho mais comum para corrigir a Eq. 6.1 no caso de férmions é o determinante de Slater, uma vez que o operador de anti simetrização se reduz a essa operação. Contudo, para bósons, o operador de simetrização pode ser substituído pela operação do permanente, que nem sempre é factível de efetuar computacionalmente. Uma maneira não ideal, porém simples e mais fácil de implementar ao invés de considerar apenas uma função  $\phi_{\alpha_i}$  para cada partícula na Eq. 6.1, utilizar uma soma sobre todos os valores de  $\alpha_i$  que pertencem ao conjunto  $\alpha$ , isto é,

$$\phi_{\alpha_i}(\mathbf{r}) \to \sum_{\mu} \phi_{\alpha\mu}(\mathbf{r}) \ .$$
 (6.3)

O índice  $\mu$ , que percorre os diferentes estados, foi usado para evitar confusão com o índice *i* do produtório, que percorre as partículas. Dessa forma, todos os termos no produtório terão a mesma forma funcional e, portanto, uma mudança na ordem dos termos é equivalente a uma troca de partículas. Como o produto é comutativo, a função é simétrica em relação à troca de duas partículas. Nesse caso, são necessários dois índices, um para percorrer as partículas e o outro para percorrer os estados, de modo que a forma resultante da função de onda é escrita como:

$$\chi_{\alpha}(\mathbf{R}) = \prod_{i} \sum_{\mu} \phi_{\alpha\mu}(\mathbf{r}_{i}) .$$
(6.4)

Em particular, supõe-se que nenhum estado no conjunto  $\alpha$  se repete e que o tamanho do conjunto é igual ao número de partículas N.

É natural que para sistemas não interagentes os estados dependam apenas das coordenadas de uma única partícula. Entretanto, ao incluir a interação, é razoável imaginar que a correlação entre as coordenadas de partículas diferentes passe a ser relevante para a descrição dos fenômenos. Nesse sentido, é importante incluir essas correlações de alguma forma ao mesmo tempo que preservar a simetria proveniente da estatística do tipo de partícula. Considerando que  $\mathbf{r}_{i}$  é o conjunto das coordenadas das partículas do sistema que não são a partícula *i*, um recurso para esse problema é incluir a dependência desse conjunto nas funções  $\phi_{\alpha\mu}$  restringindo que essa dependência seja invariante em relação à ordem das partículas do conjunto  $\mathbf{r}_{i}$ , resultando na forma da função:

$$\chi_{\alpha}(\mathbf{R}) = \prod_{i} \sum_{\mu} \phi_{\alpha\mu}(\mathbf{r}_{i}, \mathbf{r}_{/i}) . \qquad (6.5)$$

Dessa forma, a permutação de duas partículas continua equivalente à troca de posição de dois termos do produtório e, consequentemente, a função continua simétrica.

Especialmente quando se considera sistemas finitos, como átomos, moléculas e aglomerados, suas funções de onda devem ser necessariamente normalizáveis em todo espaço, isto é, a integral volumétrica em todo o espaço deve ser finita. Para expor essa propriedade explicitamente na expressão da função da Eq. 6.5, um decaimento exponencial dependente da distância ao centro de massa  $q_i$  do aglomerado é incluído, chegando em

$$\chi_{\alpha}(\mathbf{R}) = \prod_{i} \sum_{\mu} \phi_{\alpha\mu}(\mathbf{r}_{i}, \mathbf{r}_{/i}) \exp[-a_{\alpha\mu}q_{i}] , \qquad (6.6)$$

onde  $a_{\alpha\mu}$  são coeficientes relacionados com a coesão do *cluster*. Posteriormente, esses coeficientes serão vistos como parâmetros variacionais. A partir daqui, a função  $\phi$  será

tomada como uma rede neural cujas características de entrada podem ser expressas em termos da partícula principal i e do conjunto das outras partículas /i.

## 6.2 Incorporação de simetrias físicas de aglomerados na rede neural

As redes neurais do tipo *feedforward* apresentadas no Cap. 4 são poderosas e amplamente utilizadas em diversos modelos e aplicações [16]. Entretanto, aplicá-las isoladamente não é o suficiente para muitos problemas, sendo necessário ajustar e complementar com outras técnicas de *machine learning*. Um exemplo são os chamados *autoenconders* utilizados em modelos generativos [117], ou seja, modelos que produzem dados similares aos fornecidos, como por exemplo a geração de imagens baseadas em um estilo artístico.

Para atingir a intricada tarefa de representar um estado quântico, a rede neural construída se baseou em uma estrutura de duplo canal, algo semelhante a duas redes neurais onde a informação flui em paralelo. Também utilizou-se simetrias físicas presentes em aglomerados para construir a representação das características de entrada, isto é, a maneira como a informação é fornecida para a rede neural processar. Devido à complexidade da arquitetura, um diagrama esquemático é apresentado na Fig. 6.1. O esquema representa



Figura 6.1: Modelo esquemático da rede neural artificial que compõe a função de onda tentativa. O canal de partícula única é representado nos quadrados azuis, enquanto o canal de duas partículas está em verde. A fase final de montagem da função é mostrada em amarelo.

os átomos de hélio por esferas azul escuro, onde cada uma é indexada por um valor  $i = \{1, ..., N\}$  e está conectada ao seu vetor posição  $\mathbf{r}_i$ . As informações de todos os átomos é transmitida para computar o vetor centro de massa  $\mathbf{r}_{cm}$  e para calcular as distâncias relativas  $\mathbf{r}_{ij}$ .

A estrutura de duplo canal mencionada anteriormente que compõe a rede neural também é esquematizada na Fig. 6.1, sendo que os canais são divididos em canal de partí-
cula única e de duas partículas, onde por canal de partícula única se entende a sequência de dados representada na cor azul, enquanto o canal de duas partículas é a sequência representada em verde. Particularmente, as informações provenientes do aglomerado, isto é, as coordenadas das partículas são usadas para alimentar esses canais. Entretanto, sistemas finitos que interagem através de potenciais dependentes apenas das distâncias relativas e na ausência de ações externas possuem a característica intrínseca de que uma translação global das partículas não altera a situação física do problema. Em outras palavras, o hamiltoniano é invariante por uma translação global e, consequentemente, o sistema apresenta essa simetria. Nesse sentido, uma representação mais adequada para as configurações dos átomos seria em termos das posições em relação ao centro de massa, ou seja,  $\mathbf{q}_i = \mathbf{r}_i - \mathbf{r}_{cm}$ . Juntando isso com as posições relativas  $\mathbf{r}_{ij}$ , as representações escolhidas para as informações de entrada fornecidas aos canais de partícula única e de duas partículas são denotadas por

$$\mathbf{h}_{i}^{(0)} = (\mathbf{q}_{i}, q_{i}) \; ; \; \mathbf{h}_{ij}^{(0)} = (\mathbf{r}_{ij}, r_{ij}) \; , \qquad (6.7)$$

respectivamente. Note que cada índice *i* dispõe de um conjunto de dados diferentes. Contudo, essa representação ainda não é invariante sob a troca de posições entre partículas do conjunto { $\mathbf{r}_{i}$ }, uma vez que as posições relativas se modificariam ao trocar os índices. Para solucionar isso, uma operação intermediária é efetuada a fim de garantir uma representação apropriada. Esse procedimento consiste em realizar uma média sobre os elementos do canal de duas partículas e concatenar o resultado com  $\mathbf{h}_i^{(0)}$ . Dessa forma, as informações referentes a cada átomo são apresentadas no vetor intermediário

$$\mathbf{f}_{i}^{(0)}(\mathbf{r}_{i}, \{\mathbf{r}_{/i}\}) = \left(\mathbf{q}_{i}, q_{i}, \sum_{j} \frac{\mathbf{r}_{ij}}{N}, \sum_{j} \frac{r_{ij}}{N}\right)$$
(6.8)

Repare que os elementos desse vetor são quantidades invariantes sobre translações globais, além de também serem invariantes sob a troca de partículas no conjunto  $\mathbf{r}_{/i}$ . Essa segunda invariância se mostrará importante para satisfazer o caráter bosônico desejado para a função de onda tentativa. Isto posto, as características de entrada da rede neural estão definidas.

Com base nessas quantidades, as camadas ocultas da rede neural são montadas, onde cada camada  $\ell$  dispõe de componentes  $\mathbf{h}_{i}^{(\ell)} \in \mathbf{h}_{ij}^{(\ell)}$  do canal de partícula única e de duas partículas, respectivamente. As informações fluirão na rede neural pelos dois canais paralelamente, tal que o fluxo de dados no canal de partícula única recebe sinais do seu canal irmão através de operações intermediárias. Considerando isso, o vetor intermediário de partícula única é definido como

$$\mathbf{f}_{i}^{(\ell)} = \left(\mathbf{h}_{i}^{(\ell)} - \sum_{i} \frac{\mathbf{h}_{i}^{(\ell)}}{N}, \sum_{j} \frac{\mathbf{h}_{ij}^{(\ell)}}{N}\right) .$$
(6.9)

Na sequência, os canais de cada camada  $\ell$  são conectados com seus sucessivos na camada seguinte através de uma combinação linear seguida de uma operação não linear. Essencialmente, essa conexão ocorre de modo semelhante ao das camadas totalmente conectadas, como exposto no Cap. 4. Conjuntamente, essa operação é representada pelo símbolo de camada oculta na Fig. 6.1. De maneira adicional, nos casos em que as camadas possuem tamanhos iguais, uma conexão residual dos canais é transmitida para a próxima camada, operação também representada na Fig. 6.1 pelo símbolo de resíduo. Essas operações podem ser resumidas nas expressões a seguir

$$\mathbf{h}_{i}^{(\ell+1)} = \tanh\left(\mathbf{V}^{(\ell)}\mathbf{f}_{i}^{(\ell)} + \mathbf{b}^{(\ell)}\right) + \mathbf{h}_{i}^{(\ell)} , \qquad (6.10)$$

$$\mathbf{h}_{ij}^{(\ell+1)} = \tanh\left(\mathbf{W}^{(\ell)}\mathbf{h}_{ij}^{(\ell)} + \mathbf{c}^{(\ell)}\right) + \mathbf{h}_{ij}^{(\ell)} , \qquad (6.11)$$

onde os pesos  $\mathbf{V}^{(\ell)}$  e  $\mathbf{W}^{(\ell)}$ , bem como os vieses  $\mathbf{b}^{(\ell)}$  e  $\mathbf{c}^{(\ell)}$  são parâmetros variacionais a serem otimizados. Por fim, com o objetivo de ajustar a quantidade de dados de saída da rede neural ao formato da Eq. 6.6, realiza-se um último procedimento nas componentes do vetor intermediário  $\mathbf{f}_i^{(L)}$  da última camada L. Portanto, os vetores de saída da rede neural podem ser escritos como

$$\mathbf{h}_{i\alpha} = \sigma \left( \mathbf{K}_{\alpha} \mathbf{f}_{i}^{(L)} + \mathbf{j}_{\alpha} \right) , \qquad (6.12)$$

onde os pesos  $\mathbf{K}_{\alpha}$  e vieses  $\mathbf{j}_{\alpha}$  são parâmetros variacionais,  $\sigma$  é a função sigmoid da Eq. 4.5 e  $\alpha$  é um hiper parâmetro que assume valores inteiros de 1 até m. A arquitetura da rede neural é pré moldada tal que o tamanho desses vetores seja igual ao número de partículas N presentes no sistema. Consequentemente, pode-se mapear um elemento  $\mu$  desses vetores a uma das funções  $\phi$  presentes na Eq. 6.6,

$$\phi_{\alpha\mu}(\mathbf{r}_i, \mathbf{r}_{/i}) = h_{i\alpha\mu} , \qquad (6.13)$$

onde  $h_{i\alpha\mu} = (\mathbf{h}_{i\alpha})_{\mu}$  é um elemento do vetor  $\mathbf{h}_{i\alpha}$ . Com essa relação estabelecida, a montagem final da função de onda pode ser implementada, sendo este o processo que está representado em amarelo na Fig. 6.1. Os valores  $h_{i\alpha\mu}$  serão multiplicados por decaimentos exponenciais seguindo inspirações da Eq. 6.6. Na sequência, uma operação de soma sobre o índice  $\mu$ seguido de um produtório sobre o índice *i* são realizadas nessa ordem. Portanto, o estado da Eq. 6.2 pode ser reescrito da seguinte maneira

$$\psi(\mathbf{R}) = \sum_{\alpha} \omega_{\alpha} \left[ \prod_{i} \sum_{\mu} (\mathbf{h}_{i\alpha})_{\mu} \exp(-a_{\alpha\mu}q_{i}) \right] .$$
 (6.14)

Em princípio, baseado no teorema de Cybenko [37], essa é uma excelente função tentativa, com uma ótima capacidade de aproximar densamente funções que satisfazem algumas condições leves mencionadas no Cap. 4, em particular, estados quânticos de sistemas finitos compostos por bósons.

A fundação da arquitetura da rede neural ilustrada na Fig. 6.1 é inspirada na função tentativa FermiNet [31], em particular sua estrutura de dois canais. Contudo, existem algumas distinções importantes entre as duas. As características de entrada em  $\mathbf{f}_i^{(0)}$  e dos vetores intermediários  $\mathbf{f}_{i}^{(\ell)}$  deste trabalho são especializadas para partículas de spin  $\mathbf{s} = 0$ . No caso da FermiNet, há uma separação no tratamento entre partículas de spins com projeções opostas, uma vez que elétrons tem spin  $\mathbf{s} = 1/2$ . Adicionalmente, devido à simetria de translação global dos aglomerados, a distância relativa ao centro de massa foi empregada na construção das características de entrada de partícula única  $\mathbf{h}_{i}^{(0)}$  neste trabalho. Essa particularidade surge do tratamento do átomo de hélio como uma partícula sem graus de liberdade internos, isto é, considerando o conjunto núcleo e elétrons como uma partícula única. Isso elimina a necessidade de aplicar a aproximação de Born-Oppenheimer feita no contexto de química quântica. Essa distinção é evidente também nos vetores intermediários da Eq. 6.9, onde é realizada uma subtração ao invés de uma concatenação das médias dos vetores de partícula única  $\mathbf{h}_{i}^{(\ell)}$ . Além disso, as abordagens divergem consideravelmente a partir dos vetores de saída  $\mathbf{h}_{i}^{(L)}$  da rede neural, onde a montagem final da função é feita com o objetivo de capturar os comportamentos apresentados por sistemas compostos por bósons. Ainda, é importante notar que a combinação linear de funções  $\chi_{\alpha}$  na Eq. 6.2 pode ser vista como algo semelhante à expansão de múltiplos determinantes introduzida pela FermiNet, estabelecendo uma última conexão entre as duas funções.

Por outro lado, a quantidade de parâmetros variacionais depende essencialmente de quatro hiper parâmetros relacionados com a montagem do estado tentativo. Um deles é o número m de funções simétricas  $\chi_{\alpha}$  na expansão de múltiplos estados, enquanto o número de camadas L é outro. Apesar de não ser mandatório, estabeleceu-se usar camadas de tamanho iguais para explorar os resíduos presentes nas equações 6.10 e 6.11. Dessa forma, os dois hiper parâmetros faltantes serão as larguras dos canais de partícula única e de duas partículas, definindo as dimensões das matrizes correspondentes aos pesos e dos vetores dos vieses. Vale ressaltar que a tangente hiperbólica foi escolhida como função de ativação, com exceção apenas da Eq. 6.12, que usa a função sigmoid. Além disso, o canal de duas partículas flui de maneira independente do canal de partícula única.

## 6.3 A função tentativa para aglomerados e suas variações

Apesar da função mostrada na Eq. 6.14 ser suficiente para descrever sistemas compostos por bósons, ela não inclui explicitamente condições de cúspide para nenhum potencial de interação específico. Entretanto, tanto o potencial de Aziz quanto o potencial de Lennard-Jones possuem comportamentos consideravelmente abruptos para distâncias menores que metade do seus respectivos comprimentos característicos. Por esse motivo, a função tentativa proposta neste trabalho para estudar aglomerados de hélio é dada por

$$\psi_{\rm BHC}(R) = \exp\left[-\frac{1}{2}\sum_{i$$

onde a função tentativa BHC (bosonic neural network for helium clusters) é composta por dois termos, sendo o primeiro um termo de McMillan [58] como apresentado na Eq. 2.6, porém com os parâmetros b = 1 e m = 5 fixos. A necessidade de fixar esses parâmetros é devido a instabilidades apresentadas durante o processo de otimização ao incluí-los. O segundo termo é a combinação linear da Eq. 6.14, onde as funções  $\chi_{\alpha}$  são dadas pela expressão

$$\chi_{\alpha}(R) = \prod_{i} \sum_{\mu} \phi_{\alpha\mu}(\mathbf{r}_{i}, \{\mathbf{r}_{/i}\}) \exp[-a_{\alpha\mu}q_{i}], \qquad (6.16)$$

sendo que a parte da função tentativa baseada em redes neurais está encapsulada nas funções  $\phi_{\alpha\mu}$ , que são os elementos de saída da rede proposta. Em particular, a inicialização dos parâmetros variacionais é feita de maneira aleatória, sendo que os parâmetros  $\omega_{\alpha}$  e todos os vieses são amostrados de uma distribuição normal, enquanto os pesos são amostrados de uma distribuição normal escalada pelo largura da camada em questão. Já os decaimentos exponenciais são inicializados uniformemente iguais à  $10^{-1}$  em unidades adimensionais.

Apesar de fundamentalmente as funções  $\phi_{\alpha}$  dependerem das coordenadas da partícula principal e do conjunto de coordenadas das outras partículas, elas também podem ser escritas em termos dos atributos de entrada da rede neural, ou seja,

$$\phi_{\alpha\mu}(\mathbf{r}_i, \{\mathbf{r}_{/i}\}) = \phi_{\alpha\mu}\left(\{\mathbf{h}_i^0\}, \{\mathbf{h}_{ij}^0\}\right) . \tag{6.17}$$

Como consequência, é possível pensar em duas variações para a função tentativa BHC. A primeira variação seria ignorar os atributos de entrada de partícula única  $\mathbf{h}_i^{0\alpha}$ , de modo que a informação fluiria através da Eq. 6.11 até a última camada L, enquanto o vetor intermediário  $f_i^{(L)}$  seria dado por

$$\mathbf{f}_i^{(L)} = \sum_j \frac{\mathbf{h}_{ij}^{(L)}}{N} \ . \tag{6.18}$$

Dessa forma, a sequência da montagem da função tentativa não sofreria alterações. Essa mudança utiliza informações das coordenadas relativas  $\mathbf{r}_{ij}$  e ignora as informações das coordenadas  $\mathbf{q}_i$ . Por esse motivo, essa versão é denotada por  $\psi_{\mathbf{r}_{ij}}$ . Restringindo ainda mais as informações utilizadas, uma segunda variação para a função tentativa se limita ao uso somente das distâncias relativas  $r_{ij}$  como atributos de entrada, isto é,

$$h_i^{(0)} = \sum_j \frac{r_{ij}}{N} \ . \tag{6.19}$$

Por motivos semelhantes, essa versão será denotada por  $\psi_{r_{ij}}$ . Essas diferentes maneiras de apresentar os atributos de entrada para a rede neural são chamadas de representações. Diferentes representações tem capacidades de aprendizado distintas, mesmo que sejam baseadas em um mesmo conjunto de informações, que nesse caso são as coordenadas de todas as partículas. Em particular, esse aspecto é explorado no Cap. 7, onde os resultados são apresentados.

## 6.4 Comentários acerca da função para pontos quânticos

Já para o estudo de quantum dots, a função tentativa utilizada se baseou em uma reinterpretação do Ansatz FermiNet [31] para atender as necessidades específicas de um sistema de elétrons aprisionados em um potencial harmônico em duas dimensões. Para um sistema composto de  $n^{\uparrow}$  elétrons com *spin* para cima e  $n^{\downarrow}$  elétrons com *spin* para baixo, sua forma geral é composta por uma combinação linear de produtos de determinantes dado por

$$\psi_{\rm FN}(\mathbf{r}_1^{\uparrow},...,\mathbf{r}_{n\downarrow}^{\downarrow}) = \sum_k \omega_k \left( \det \left[ \phi_i^{k\uparrow} \left( \mathbf{r}_j^{\uparrow}; \left\{ \mathbf{r}_{/j}^{\uparrow} \right\}; \left\{ \mathbf{r}^{\downarrow} \right\} \right) \right] \det \left[ \phi_i^{k\downarrow} \left( \mathbf{r}_j^{\downarrow}; \left\{ \mathbf{r}_{/j}^{\downarrow} \right\}; \left\{ \mathbf{r}^{\uparrow} \right\} \right) \right] \right) ,$$

$$(6.20)$$

onde os coeficientes  $\omega_k$  são parâmetros variacionais, as coordenadas do elétron j com spin  $\alpha$  são denotadas por  $\mathbf{r}_j^{\alpha}$ . Além disso,  $\{\mathbf{r}^{\alpha}\}$  representa o conjunto de coordenadas de todos elétrons com  $spin \ \alpha = \{\uparrow, \downarrow\}$ , enquanto  $\{\mathbf{r}_{j}^{\alpha}\}$  representa o conjunto dos elétrons de spin  $\alpha$  que não são o elétron principal, isto é, o elétron j. De maneira semelhante à função tentativa BHC, as funções  $\phi_i^{k\alpha}$  são construídas empregando redes neurais como base e podem ser escritas como

$$\phi_i^{k\alpha} \left( \mathbf{r}_j^{\alpha}; \left\{ \mathbf{r}_{j}^{\alpha} \right\}; \left\{ \mathbf{r}^{\bar{\alpha}} \right\} \right) = \left( \mathbf{w}_i^{k\alpha} \cdot \mathbf{h}_j^{L\alpha} + g_i^{k\alpha} \right) \exp\left( -|\boldsymbol{\Sigma}_i^{k\alpha} \mathbf{r}_j^{\alpha}| \right) , \qquad (6.21)$$

onde  $\mathbf{w}_i^{k\alpha}$ ,  $g_i^{k\alpha} \in \mathbf{\Sigma}_i^{k\alpha}$ são parâmetros variacionais e  $\mathbf{h}_j^{L\alpha}$ são os elementos de saída da rede neural. Para computar esses elementos utiliza-se as seguintes relações

$$\mathbf{h}_{i}^{\ell+1\alpha} = \tanh\left(\mathbf{V}^{\ell}\mathbf{f}_{i}^{\ell\alpha} + \mathbf{b}^{\ell}\right) + \mathbf{h}_{i}^{\ell\alpha} , \qquad (6.22a)$$

$$\mathbf{h}_{ij}^{\ell+1\alpha\beta} = \tanh\left(\mathbf{W}^{\ell}\mathbf{h}_{ij}^{\ell\alpha\beta} + \mathbf{c}^{\ell}\right) + \mathbf{h}_{ij}^{\ell\alpha\beta} , \qquad (6.22b)$$

onde, semelhantemente ao caso apresentado anteriormente,  $\mathbf{V}^{\ell} \in \mathbf{W}^{\ell}$  são pesos, enquanto  $\mathbf{b}^{\ell} \in \mathbf{c}^{\ell}$  são vieses da rede neural. Entretanto, os elementos de uma camada qualquer  $\ell$  são especializados para spin  $\mathbf{s} = 1/2$ . Assim sendo, os elementos  $\mathbf{h}_{i}^{\ell\alpha}$  correspondem ao canal de um elétron, enquanto os elementos  $\mathbf{h}_{ij}^{\ell\alpha\beta}$  correspondem ao canal de dois elétrons. Para o canal de dois elétrons,  $\alpha$  indica o spin do elétron *i* e  $\beta$  o do elétron *j*. Além disso, os vetores intermediários  $\mathbf{f}_{i}^{\ell\alpha}$  também são especializados para elétrons e assumem a seguinte forma

$$\mathbf{f}_{i}^{\ell\alpha} = \left(\mathbf{h}_{i}^{\ell\alpha}, \frac{1}{n^{\uparrow}} \sum_{j=1}^{n^{\uparrow}} \mathbf{h}_{j}^{\ell\uparrow}, \frac{1}{n^{\downarrow}} \sum_{j=1}^{n^{\downarrow}} \mathbf{h}_{j}^{\ell\downarrow}, \frac{1}{n^{\uparrow}} \sum_{j=1}^{n^{\uparrow}} \mathbf{h}_{ij}^{\ell\alpha\uparrow}, \frac{1}{n^{\downarrow}} \sum_{j=1}^{n^{\downarrow}} \mathbf{h}_{ij}^{\ell\alpha\downarrow}\right) .$$
(6.23)

Por fim, os atributos de entrada de cada canal são dados pela concatenação das coordenadas das partículas e das coordenadas relativas, sendo assim, eles podem ser escritos como

$$\mathbf{h}_{i}^{0\alpha} = (\mathbf{r}_{i}^{\alpha}, |\mathbf{r}_{i}^{\alpha}|) \; ; \; \mathbf{h}_{ij}^{0\alpha\beta} = \left(\mathbf{r}_{i}^{\alpha} - \mathbf{r}_{j}^{\beta}, |\mathbf{r}_{i}^{\alpha} - \mathbf{r}_{j}^{\beta}|\right) \; . \tag{6.24}$$

Essa função possui todos elementos necessários para descrever um sistema composto de elétrons e aprisionados por um potencial harmônico, como é o caso do modelo para pontos quânticos apresentado anteriormente. Em particular, a anti-simetria da função de onda está presente nos determinantes empregados na Eq. 6.20, o decaimento exponencial com à distância até a origem devido ao potencial harmônico está incluído nas funções  $\phi_i^{k\alpha}$ . Os demais comportamentos relevantes do sistema, a rede neural tem capacidade de aprender.

# Parte III

Apresentação e discussão dos resultados

O estudo de aglomerados de hélio através de métodos de Monte Carlo na literatura é bastante extenso [3,43,49,50,63,118,119]. Por ser um sistema formado por bósons, alguns métodos de Monte Carlo quântico são capazes de alcançar resultados de energia considerados exatos dentro da sua incerteza estatística. Ainda que resultados exatos possam ser alcançados nesse caso, aglomerados de hélio continuam atraindo interesse em contextos como da unitariedade fraca [73], comportamentos universais [120] e estudo de correlações de curto alcance [4]. Assim, de posse das funções tentativas introduzidas no Cap. 6 e do método variacional descrito no Cap. 5, foi realizado um estudo dos aglomerados de hélio através dessa metodologia com o objetivo de verificar se as redes neurais são capazes de aprender o estado fundamental desses sistemas e, consequentemente, contribuir para seu melhor entendimento.

Para estudar os *clusters* de hélio, foi utilizado um hamiltoniano largamente empregado na literatura que é dado por

$$\mathcal{H} = -\frac{\hbar^2}{2m} \sum_i \nabla_i^2 + \sum_{i < j} V(r_{ij}) , \qquad (7.1)$$

onde  $m = 4.0026 m_u$  é a massa de um átomo de hélio-4 [121],  $m_u$  é a constante de massa atômica e o potencial de interação empregado será sempre o potencial de Aziz da Eq. 2.2 a menos que explicitado o contrário. Vale ressaltar que para as unidades do potencial e da energia cinética serem compatíveis é necessário dividir a energia cinética pela constante de Boltzmann  $k_B$ . Como resultado, as energias serão exibidas em unidade de Kelvin.

A principal função de onda tentativa adotada nas simulações foi a da Eq. 6.15, cuja rede neural está associada com 4 hiper parâmetros que definem sua arquitetura, além de alguns outros relacionados mais especificamente com o processo de aprendizagem. A seleção cuidadosa destes desempenha um papel importante para o processo de aprendizagem, garantindo sua eficiência e eficácia. Por isso, a Tab. 7.1 exibe valores típicos para esses hiper parâmetros. Devido a construção dessas redes terem inspirações retiradas da FermiNet, alguns dos parâmetros tipicamente usados por este Ansatz foram reutilizados. Além disso, sistemas com números de partículas diferentes exigem algum aprimoramento dos hiper parâmetros, em particular daqueles que definem a arquitetura da rede, uma vez que sistemas com mais partículas precisam de redes neurais maiores para atingir comparativamente o mesmo grau de acurácia.

Os resultados obtidos foram divididos em três partes, iniciando pela validação da metodologia através de comparações com métodos bem estabelecidos. A segunda parte da

Tamanho do <i>batch</i>	8192
Largura do canal de partícula única	128
Largura do canal de duas partículas	16
Número de funções simétricas	8
Número de camadas	4
Taxa de aprendizado inicial	0.001
Taxa de decaimento	1.0
Taxa de retardamento	10000
Iterações de treinamento	$2 \times 10^5$
Amortecimento	0.001
Momento	0.0
Decaimento da matriz de covariâncias	0.95
Constrição de norma	0.001

Tabela 7.1: Típicos valores dos hiper parâmetros empregados nas simulações de aglomerados de hélio.

ênfase no cálculo de propriedades além da energia e nas conexões com conceitos de unitariedade fraca. Por fim, a última parte explora diferentes possibilidades nas características de entrada da rede neural e algumas variações das arquiteturas. Além disso, uma versão *open source* do programa empregado na obtenção destes resultados pode ser encontrado em [122].

### 7.1 Validando a metodologia empregada

Para verificar se a função de onda tentativa proposta na Eq. 6.15 é capaz de representar suficientemente bem estados fundamentais de aglomerados de hélio, o primeiro passo foi aplicar o processo de otimização representado na Fig. 5.1 com intuito de minimizar a energia do sistema. Inicialmente, um aglomerado composto por 6 partículas foi analisado e o processo de treinamento da rede neural é mostrado na Fig. 7.1 em azul. Por propósitos de clareza, as incertezas dos valores não foram exibidas. Durante o processo, observou-se que sua evolução ocorreu de forma aquém do ideal, exibindo algumas instabilidades na fase final de otimização, particularmente após encontrar parâmetros que descrevem qualitativamente bem os valores para energia. Nesse ponto, conjecturou-se que as flutuações tinham relação com as condições de cúspide e com a presença de pares de partículas extremamente próximas em algumas configurações amostradas. Como mostrado no Ap. B, o termo de McMillan da Eq. 2.6 com os devidos parâmetros é a solução para curtas distâncias considerando o potencial de Lennard-Jones e, portanto, atende as condições de cúspide. Assim, para verificar a veracidade da hipótese, uma simulação equivalente usando o potencial de Lennard-Jones da Eq. 2.1 foi feita para o mesmo número de partículas e é mostrada no destaque da Fig. 7.1 em verde. Embora as simulações não possam ser diretamente comparadas, houve uma considerável melhora nas flutuações, confirmando a sugestão anterior.



Figura 7.1: Evolução da energia total com o número de passos de otimização para um aglomerado composto por 6 átomos interagindo pelo potencial de Aziz (círculos azuis). O efeito da inclusão do potencial de esfera dura é mostrado em laranja. O destaque ilustra uma otimização considerando o potencial de Lennard-Jones.

Por consequência, com o objetivo de transpor estas instabilidades, um potencial de esfera dura de raio  $a_R$  foi incorporado ao potencial de Aziz exclusivamente durante as otimizações. Assim, quando qualquer configuração amostrada tivesse algum par de partículas cuja distância entre elas fosse menor do que  $a_R$ , essa configuração seria automaticamente rejeitada. Com alguns testes observou-se que um raio  $a_R$  de aproximadamente 1.8Å foi capaz de controlar as flutuações o suficiente, além de não ser crítico nem prejudicar significativamente as estimativas das energias. Consequentemente, outra simulação equivalente considerando essa inclusão do potencial de esfera dura no potencial de Aziz foi realizada e é apresentada na Fig. 7.1 em laranja. Dessa forma, as flutuações foram controladas efetivamente. Em particular, como abordado no Cap. 2, as propriedades de unitariedade fraca satisfeitas por *clusters* de hélio garantem que a inclusão do potencial de esfera dura seja pouco relevante do ponto de vista da função de onda tentativa.

Com essa experiência, padronizou-se utilizar a combinação do potencial de Aziz e da esfera dura no decorrer exclusivamente do processo de treinamento. Sendo assim, o passo seguinte foi comparar os resultados com cálculos obtidos através do método de difusão de Monte Carlo (DMC). O método de Monte Carlo por difusão destaca-se como uma técnica computacional poderosa. Por meio de uma rotação de Wick, ele reformula a equação de Schrödinger dependente do tempo em uma equação de difusão no tempo imaginário. Ao evoluir iterativamente um conjunto de configurações do sistema amostradas de uma função de onda tentativa, que representam as posições das partículas no espaço, por tempo imaginário suficientemente longo, o DMC projeta a componente do estado fundamental dessas configurações. A partir das configurações, chamadas de caminhantes aleatórios, distribuídas de acordo com a função de onda do estado fundamental, o DMC fornece estimativas precisas das energias do estado fundamental, sujeitas apenas a incertezas de natureza estatística. Em particular, os resultados são considerados exatos dentro dessa incerteza para sistemas compostos puramente por bósons.

Para exemplificar a comparação com esse método, a Fig. 7.2 mostra o processo de otimização para aglomerados contendo de 2 até 10 átomos de hélio, onde a evolução da energia do sistema é mostrada em azul em função do número de passos, enquanto os resultados de DMC [43] são apresentados em laranja tracejado. Alguns aspectos chamam



Figura 7.2: Diversos processos de treinamento mostrando o progresso da energia (azul) em função das iterações de otimização. Os aglomerados considerados continham de 2 até 10 átomos. Para comparação, resultados de DMC são exibidos por um linha tracejada laranja.

atenção, por exemplo, é possível dividir o processo qualitativamente em duas fases. A primeira fase se estende até por volta de alguns milhares de passos e o foco do treinamento reside em encontrar parâmetros que descrevam qualitativamente bem a energia do sistema. Desse ponto em diante, sem qualquer diretiva explicita incluída no programa, a rede neural passa a aprender como reduzir as flutuações entre passos de otimização, levando a um resultado mais acurado como consequência. Novamente, as incertezas não são apresentadas para maior clareza das imagens.

Vale ressaltar também que, em diversos momentos do processo, é notável que os resultados parecem estar abaixo da energia considerada exata, especialmente para o caso de N = 9 partículas, o qual apresenta uma evolução não usual da energia do sistema.

Esse comportamento tem origem no caráter estocástico da simulação e na maneira como as configurações iniciais são amostradas, como comentado no Cap. 5. De modo especial, as estimativas iniciais de energia são imprecisas por efeito de escolhas desavantajosas das posições dos átomos no começo do processo, bem como o nível de aceitação das novas amostras nas primeiras etapas do método ainda está alto. Não só as estimativas de energia são imprecisas como também as variâncias associadas são consideráveis. Por esse motivo, os resultados parecem aumentar a energia em vez de diminuir, contudo, as incertezas estatísticas no início da otimização são substanciais o suficiente para tornar as energias associadas inconclusivas. Entretanto, com a evolução do treinamento, a aceitação se ajusta gradualmente até atingir níveis esperados e as configurações amostradas começam a gerar resultados relevantes, os quais possuem incertezas razoáveis. Assim, as energias passam a concordar dentro das incertezas estatísticas com os resultados de DMC.

Após a otimização, os parâmetros da função de onda tentativa são salvos e não serão mais alterados. De posse dessa boa aproximação para o estado fundamental do sistema, são realizadas novas simulações levando em conta somente o potencial de Aziz. As estimativas da energia e de outras propriedades são obtidas utilizando integração de Monte Carlo e considerando amostras maiores, isto é, simulações mais longas, com intuito de reduzir as incertezas. Esses resultados são exibidos na Tab. 7.2, onde podem ser comparados com os mesmos resultados de DMC mencionados anteriormente. Na tabela pode ser encontrado

Tabela 7.2: Em unidades de Kelvin, são apresentadas a energia cinética  $\langle \mathcal{T} \rangle$  e a energia total  $\langle E \rangle$  obtidas com a função tentativa  $\psi_{\text{BHC}}$  para aglomerados de <sup>4</sup>He contendo N átomos. A quarta coluna mostra resultados de DMC da literatura [43].

N	$\langle \mathcal{T}  angle$	$\langle E \rangle$	DMC
2	0.1246(4)	-0.002142(6)	-
3	1.695(2)	-0.13323(9)	-0.135(2)
4	4.353(3)	-0.5775(1)	-0.573(2)
5	7.745(4)	-1.3341(2)	-1.334(2)
6	11.666(4)	-2.3710(3)	-2.367(3)
7	16.145(5)	-3.6510(4)	-3.646(4)
8	20.997(6)	-5.1448(4)	-5.144(5)
9	26.212(7)	-6.8207(9)	-6.827(6)
10	31.868(8)	-8.6766(7)	-8.673(6)

estimativas variacionais da energia cinética e da energia total para *clusters* contendo de 2 até 10 átomos. A última coluna apresenta os valores obtidos através dos cálculos de DMC. Valores adicionais para resultados de energia podem ser encontrados no Ap. D.

Para analisar a qualidade de propriedades além da energia, a quantidade escolhida foi a distribuição de distâncias interatômicas para o trímero (N = 3), a qual foi calculada através do valor esperado do operador da Eq. 2.16. Dessa forma, os resultados obtidos de P(r) são apresentados na Fig. 7.3 em círculos laranjas, enquanto os dados representados pela linha azul são cálculos utilizando DMC e foram extraídos de [123]. A concordância



Figura 7.3: Representação das distribuições P(r) (Eq. 2.16) para o trímero obtidas através de redes neurais e de cálculos de DMC em laranja e azul, respectivamente.

entre os métodos é excelente e sugere uma ótima capacidade de representação de estados quânticos através das redes neurais nos cálculos de propriedades além da energia. É importante notar que, para propriedades diferentes da energia, resultados de DMC devem ser extrapolados, o que pode introduzir um viés variacional em certos casos [124].

Outro teste realizado foi examinar as energias obtidas sob a ótica do modelo de gota, onde as energias escalam de acordo com

$$E_{\rm g}(N) = C_v N + C_s N^{2/3} + C_c N^{1/3} , \qquad (7.2)$$

onde os parâmetros  $C_v$ ,  $C_s \in C_c$  correspondem aos termos de volume, superfície e curvatura, respectivamente. Foram realizadas otimizações para *clusters* contendo de 11 até 24 átomos, sendo que de 11 até 14 partículas foi realizado  $10^5$  iterações de otimização, enquanto de 15 até 24 átomos foram executados apenas  $10^4$  passos, uma vez que para aglomerados maiores as simulações são consideravelmente mais lentas. As energias obtidas são apresentadas na Fig. 7.4 em pontos pretos, enquanto a linha vermelha representa um ajuste desses dados ao modelo de gota. Além disso, a linha tracejada azul mostra um ajuste similar feito através de resultados obtidos com DMC [43]. Os coeficientes de ambos ajustes podem ser consultados no Ap. D. Até por volta de 17 partículas, os ajustes são bastante similares, região na qual os resultados tem maior precisão. Conforme o sistema aumenta e também devido ao número reduzido de passos de otimização, os ajustes terminam por divergir entre si. Contudo, os resultados anteriores indicam que, utilizando tempo e recurso computacionais maiores, seria possível inclusive atingir resultados com menores desvios estatísticos ao comparar com resultados obtidos com o método DMC, já que as funções tentativas obtidas teriam uma qualidade maior.

Como última analise desta etapa, o tempo computacional gasto (walltime) no pro-



Figura 7.4: Energias totais em função do número de partículas no aglomerado é exibido em círculos pretos. Um ajuste desses dados ao modelo de gota é representado por uma linha vermelha. Por razões de comparação, um ajuste similar feito com base em resultados de DMC é apresentado em azul tracejado.

cesso de otimização de aglomerados compostos de 2 até 14 átomos é mostrado na Fig. 7.5 em azul. Todos os tempos  $t_N$  de simulação são exibidos em relação ao tempo  $t_2$  de apro-



Figura 7.5: Dimensionamento do tempo computacional gasto  $t_N$  em termos de *walltime* para diversos tamanhos de *clusters* em relação ao tempo de treinamento do dímero  $t_2$ . A linha tracejada em laranja representa um ajuste polinomial de segundo grau.

ximadamente 7 minutos e meio empregado no dímero (N = 2) para uma simulação com  $10^4$  iterações de otimização. Um ajuste polinomial de segundo grau também é exibido em tracejado laranja. Para este caso, foram utilizadas quatro GPUs NVIDIA A100 interconectadas.

A partir destes resultados apresentados, comprovou-se que o emprego de redes neurais na construção da função de onda tentativa é um boa abordagem para representar o estado fundamental na investigação de sistemas formados por bósons, em particular, os aglomerados de átomos de hélio. As sinergias existentes no contexto de *machine learning* com o método variacional de Monte Carlo mostraram-se eficiente, especialmente utilizando o otimizador KFAC.

Ainda sobre o procedimento de otimização, alguns comentários técnicos relevantes se fazem necessários. O algoritmo de Metropolis utilizado para amostrar as configurações é baseado em aceitar ou não configurações propostas, como comentado no Cap. 5. Como as novas amostras são propostas com base nas atuais, o grau de aceitação varia com quão distante a nova configuração foi proposta. Geralmente, isso é controlado por um hiper parâmetro que define a largura do movimento proposto, sendo que quanto maior esse parâmetro, menor é a aceitação. Para a versão utilizada desse algoritmo, a porcentagem de aceitação das novas amostras deve ser em torno de 40% à 50% [124] para maior eficiência. Entretanto, o procedimento adotado busca adaptar a largura dos movimentos propostos durante a otimização de maneira que no início do treinamento a aceitação seja próxima de 99% e, gradualmente, a aceitação diminua até atingir valores em torno de 40%. Esse processo é feito principalmente durante os primeiros milhares de iterações, de forma que a maior parte da otimização é conduzida com aceitação de 50%.

Outro aspecto prático das simulações realizadas reside no tamanho do batch. Tipicamente, o valor utilizado para M é da ordem de 10<sup>4</sup> configurações. Para o hardware usado nas simulações, essa ordem de grandeza é geralmente suficiente para utilizar eficientemente as GPUs e conduzir o processo de treinamento de maneira eficaz. Entretanto, esse número de amostras é razoavelmente pequeno e, portanto, as incertezas das estimativas da energia ao empregar amostras desse tamanho são razoáveis, geralmente uma ordem de grandeza acima do desejado mesmo ao final do processo de treinamento. Por esse motivo, como observado anteriormente, é necessário conduzir simulações mais longas ao final da otimização. Além disso, devido a alta aceitação no início da simulação e em virtude das configurações iniciais das posições dos átomos serem amostradas aleatoriamente de uma distribuição gaussiana, as flutuações e os desvios padrão das energias iniciais são consideráveis. Isso pode gerar comportamentos atípicos da curva de otimização, como por exemplo as estimativas de energia iniciais terem valores abaixo da energia do estado fundamental. Entretanto, devido aos erros estatísticos serem consideravelmente altos nessa parte da simulação, isso torna esses valores inconclusivos. Além disso, esse comportamento pode ser evitado escolhendo adequadamente a dispersão da distribuição gaussiana utilizada para amostrar as posições iniciais.

## 7.2 Cálculo de propriedades e a unitariedade fraca

Nesta etapa do trabalho, as atenções foram direcionadas aos cálculos de propriedades além da energia do sistema, particularmente, muitas das quantidades estudadas tinham como objetivo sondar os aspectos universais relacionados com aglomerados de hélio. Embora universalidade não seja esperada nesses sistemas devido às suas distâncias características, um caráter universal parcial pode ser observado, como comentado no Cap.2. Ainda vale mencionar que os resultados dessa seção foram obtidos empregando as funções resultantes do processo de otimização mencionado anteriormente.

Primeiramente, com o intuito de visualizar a representação do estado fundamental gerado pela rede neural, um corte no plano z = 0 da função de onda do dímero é apresentada na Fig. 7.6. Para produzir a imagem, um dos átomos foi fixado na origem e a função é



Figura 7.6: Função de onda do dímero projetada no plano xy em função das coordenadas após a otimização da energia. Um dos átomos está fixo na origem e o eixo z mostra em uma escala arbitrária o valor da função quando o outro átomo está nas coordenadas (x, y, 0). Para enfatizar as regiões de alta probabilidade, uma mapa de cores foi utilizado.

avaliada em diferentes posições do segundo átomo no plano xy, gerando valores em uma escala arbitrária, uma vez que a função não está normalizada. Interessantemente, apesar de não ser incluído explicitamente na construção da função de onda, a rede neural foi capaz de capturar a simetria esférica do problema. Não apenas isso, como também o comportamento segue o esperado, onde se presume que o estado se anule rapidamente quando as partículas se aproximam, tenha um máximo em uma posição intermediária e decaia lentamente conforme elas se afastam. Inclusive, através da região de decaimento da função tentativa, a aproximação da Eq. 2.11 pode ser usada para estimar o comprimento de espalhamento do potencial usando um ajuste dessa região. Ao realizar esse cálculo, a estimativa do comprimento de espalhamento foi de 115Å, enquanto o valor calculado para o potencial de Aziz é de 124Å [73]. Outra quantidade calculada foi o comprimento de ligação do dímero, determinada pelo valor médio da distância do par  $\langle r_{ij} \rangle = 36.6 \pm 0.6$ Å, o qual é um pouco menor que o valor experimental de  $52 \pm 4$ Å [125]. Essa discrepância muito provavelmente está relacionada com efeitos relativísticos e de eletrodinâmica quântica [126].

Antes de adentrar as propriedades relacionadas com a unitariedade fraca, uma análise do perfil de densidade n(r) dos aglomerados foi realizada. Na Fig. 7.7 é apresentado



Figura 7.7: Perfis de densidade em função da distância ao centro de massa para aglomerados compostos de 2 até 14 átomos, onde as cores vão do amarelo para o azul com o aumento do número de partículas. São considerados apenas *clusters* com números pares de átomos.

esses perfis para aglomerados contendo N = 2, 4, 6, 8, 10, 12 e 14 átomos de hélio, onde as cores seguem do amarelo para o azul nessa ordem de tamanhos. Repare que o centro de massa do dímero é essencialmente vazio, evidenciando o largo comprimento de ligação. Além disso, conforme o número de partículas aumenta, o comportamento do perfil de densidade para pequenas distâncias parece convergir, indicando que no centro dos aglomerados a densidade está convergindo para a densidade de equilíbrio do *bulk* de 0.0218 Å<sup>-3</sup> [127]. Note que a normalização do perfil de densidade é sobre todo volume, isto é, a integral radial dessas funções pode ser reescrita como  $\int n(r)r^2 dr = N/4\pi$ . Diferentemente, a normalização da densidade de pares  $\rho_N$  não leva em conta a integral angular da maneira como foi definida.

Já para estudar aspectos da unitariedade fraca, as funções densidade de pares  $\rho_N$  obtidas através do valor esperado da Eq. 2.14 foram calculadas para vários tamanhos de aglomerados. Essas funções são mostradas na Fig. 7.8 levando em conta *clusters* contendo N = 4, 6, 8, 10, 12 e 14 átomos nas cores do amarelo para o roxo, nessa ordem. Sua estrutura geral acompanha o esperado, onde comportamento delas apresenta um pico seguido de um decaimento suave, além de se anularem rapidamente para pequenas distâncias. Entretanto, devido à universalidade fraca presente nesses sistemas, o comportamento dessas funções no regime de curtas distâncias é definido pela densidade de pares do dímero a menos de um fator multiplicativo, como visto no Cap. 2. Esse fator é o contato  $C_2^{(N)}$  e indica uma medida da correlação de dois corpos para curtas distâncias. Para estimar o valor desses fatores, foi feito um ajuste através da minimização do valor da integral  $\int (\rho_N(r) - C_2^{(N)}\rho_2(r)) dr$  no intervalo de zero até o valor de r correspondente ao valor máximo da densidade de pares  $\rho_N(r)$ . Na Fig. 7.9 são apresentados os resultados para os contatos em função do tamanho do aglomerado. Os valores numéricos obtidos são apresentados no Ap. D na Tab. D.3.

Já para observar o colapso da densidade de pares, a Fig. 7.10 mostra a razão entre



Figura 7.8: Função densidade de pares para aglomerados de  ${}^{4}\text{He}_{N}$  contendo  $N = \{4, 6, 8, 10, 12, 14\}$  átomos em função da distância e apresentadas nas cores do amarelo para o azul, nessa ordem.



Figura 7.9: Valores para os contatos  $C_2^{(N)}$  em função do número de partículas N em escala logarítmica.

as densidades de pares  $\rho_N$  e o contato  $C_2^{(N)}$  equivalente para aglomerados compostos de N = 2, 4, 6, 8, 10, 12 e 14 átomos, lembrando que por definição  $C_2^{(2)} = 1$ . Novamente, as cores do amarelo para o azul representam tamanhos de *clusters* do menor para o maior. O colapso das densidades para N > 3 sobrepondo aproximadamente a densidade de pares do dímero na região de distâncias curtas é notável para todos os tamanhos analisados. Esse resultado é um indício de que esses sistemas apresentam universalidade fraca e são consistentes com previsões existentes na literatura [128].

Um procedimento semelhante ao realizado para observar a densidade de pares também foi empregado para a analise de correlações de três corpos nesses sistemas. Nesse caso, a função utilizada foi a distribuição de três corpos de Jacobi  $P_N^{(3,\text{jacobi})}$  da Eq. 2.18. Na Fig. 7.11 é apresentado o comportamento das distribuições de Jacobi em função de  $\rho$ para aglomerados contendo de 3 até 14 átomos. Vale ressaltar que  $\rho$  deve ser visto como a distância de um átomo até o centro de massa de algum outro par. Para caracterizar



Figura 7.10: Colapso para curtas distâncias das funções densidade de pares  $\rho_N$  normalizadas pelos seus respectivos contatos  $C_2^{(N)}$ . As diferentes cores do amarelo para o azul mostram resultados para aglomerados contendo de N = 2 até N = 14.



Figura 7.11: Distribuições de Jacobi $P_N^{(3,\text{jacobi})}$  de três corpos para diferentes tamanhos de aglomerado em função da distância  $\rho$  de um átomo até o centro de massa de um par de outros dois átomos.

as correlações de curto alcance de três corpos, em particular as correlações de par-átomo, os contatos de par-átomo  $C_{2+1}^{(N)}$  foram calculados utilizando o mesmo procedimento empregado na estimativa dos contatos  $C_2^{(N)}$ , ou seja, através da minimização do valor da integral  $\int \left(P_N^{(3,\text{jacobi})} - C_{2+1}^{(N)}P_3^{(3,\text{jacobi})}\right)$ . O colapso das distribuições de Jacobi é mostrado na Fig. 7.12. A sobreposição para curtas distâncias das distribuições de aglomerados com mais de três átomos em relação à distribuição do trímero é evidente. Além disso, é interessante notar que o comportamento dessas funções no limite de  $\rho \to 0$  é mais suave do que no caso das densidades de pares, as quais se anulam rapidamente para curtas distâncias, como pode ser visto na Fig. 7.10. Por fim, alguns resultados adicionais podem ser encontrados no Ap. D.



Figura 7.12: Colapso para curtas distâncias  $\rho$  das distribuições de Jacobi  $P_N^{(3,\text{jacobi})}$  normalizadas pelos seus respectivos contatos par-átomo  $C_{2+1}^{(N)}$ .

## 7.3 Influência das características de entrada

Com o objetivo de analisar a relevância da representação dos dados para a rede neural, três diferentes formas de construir os atributos de entrada para a função  $\psi_{\rm BHC}$  foram considerados. Evidentemente, um deles considera a função tentativa completa como escrita na Eq. 6.15. Como ela utiliza informação de  $\mathbf{q}_i \in \mathbf{r}_{ij}$  na construção dos atributos  $\mathbf{h}_i^{(0)} \in \mathbf{h}_{ij}^{(0)}$ , essa versão da função tentativa é denotada como  $\psi_{\mathbf{q}_i,\mathbf{r}_{ij}}$ . Como comentado nas variações da função tentativa no Cap. 6, um segundo Ansatz que ignora os termos dependentes de  $\mathbf{q}_i$  e  $q_i$  é expressado como  $\psi_{\mathbf{r}_{ij}}$ , onde essa versão depende somente das coordenadas relativas  $\mathbf{r}_{ij}$ . Finalmente, uma função tentativa dependente apenas das distâncias relativas  $r_{ij}$  também foi examinada e denotada por  $\psi_{r_{ij}}$ .

O processo de otimização da energia para um aglomerado de 5 átomos é apresentado na Fig. 7.13 considerando esses diferentes níveis de informação presentes na função tentativa. A evolução da energia com o decorrer das iterações durante o processo de otimização é mostrado em azul, verde e laranja para as funções  $\psi_{r_{ij}}$ ,  $\psi_{\mathbf{r}_{ij}}$  e  $\psi_{\mathbf{q}_i,\mathbf{r}_{ij}}$ , respectivamente. Pode-se observar que todas as funções alcançam um grau similar de acurácia. Entretanto, um conjunto de informações mais completo facilita a otimização, tornando-a mais rápida e eficiente. É interessante notar que, embora o hamiltoniano do sistema essencialmente depender somente das coordenadas relativas, adicionar informação referente à distância ao centro de massa de cada átomo melhora substancialmente o processo de aprendizagem. Além disso, no destaque da Fig. 7.13 é apresentado resultados da energia variacional para cada versão da função tentativa após o processo de otimização.

Também foram conduzidos experimentos numéricos para o dímero de <sup>4</sup>He. Devido a simetria radial do potencial de interação, o estado fundamental do sistema depende



Figura 7.13: Processo de otimização da energia para diferentes representações dos atributos da rede neural, isto é, para diferentes funções tentativa. O pontos azuis utilizaram a função  $\psi_{\mathbf{r}_{ij}}$ , os verdes empregaram a função  $\psi_{\mathbf{r}_{ij}}$  e os laranjas a função completa  $\psi_{\mathbf{q}_i,\mathbf{r}_{ij}}$ . O destaque mostra as energias variacionais estimadas após o processo de otimização para as respectivas funções tentativa.

somente da distância entre o par de partículas. Como resultado, todas as três formas da função tentativa devem satisfazer essa propriedade após a otimização. Contudo, por causa do modo como foram construídas, somente a densidade de probabilidade  $\psi_{r_{ij}}^2$  é intrinsecamente radial dentre as três formas. Para mostrar que a rede neural é capaz de adquirir razoavelmente essa propriedade durante o processo de treinamento, a Fig. 7.14 mostra o módulo das diferenças entre a densidade de probabilidade  $\psi_{r_{ij}}^2$  e as médias angulares dos outros dois formatos da densidade de probabilidade  $\psi_{\mathbf{q}_i,\mathbf{r}_{ij}}^2$  e  $\psi_{\mathbf{r}_{ij}}^2$ . A curva em



Figura 7.14: Módulo das diferenças de probabilidade das funções não radiais com relação à função radial, onde os pontos laranja são dados por  $|\psi_{r_{ij}} - \langle \psi_{\mathbf{q}_i,\mathbf{r}_{ij}} \rangle|$  e os pontos verde por  $|\psi_{r_{ij}} - \langle \psi_{\mathbf{r}_{ij}} \rangle|$ . O destaque mostra a densidade de probabilidade  $\psi_{r_{ij}}^2$ .

verde representa  $|\psi_{r_{ij}}^2 - \psi_{\mathbf{q}_i,\mathbf{r}_{ij}}^2|$ , enquanto a curva laranja mostra os valores de  $|\psi_{r_{ij}}^2 - \psi_{\mathbf{r}_{ij}}^2|$  em

função da distância do par. Além disso, a densidade de probabilidade não normalizada  $\psi_{r_{ij}}^2$ é mostrada em azul no destaque da figura. Essas comparações mostram que as funções com dependência angular apresentam a maior diferença em torno do máximo de  $\psi_{r_{ij}}^2$ . Porém, essa diferença é pequena, sendo três ordens de magnitude menor que os próprios valores. Adicionalmente, a variância das médias angulares das densidades de probabilidade são usualmente da ordem de  $10^{-6}$  ou menor nas unidades arbitrárias utilizadas. Portanto, as redes neurais com dependência angular foram capazes de aprender a simetria do sistema sem serem explicitamente programadas para isso.

Portanto, a representação dos dados demonstrou relevância na obtenção de resultados. Isso também sugere que melhorias na representação podem conduzir a processos de treinamento mais eficientes, onde as redes neurais ainda sejam capazes de aprender as simetrias relevantes do sistema mesmo se os atributos de entrada não incluem diretamente essas propriedades. Ao contrário dos aglomerados de hélio, os pontos quânticos descritos no Cap. 3 são sistemas constituídos por elétrons, isto é, férmions. Adicionalmente, uma limitação dos métodos de Monte Carlo quântico empregados no estudo do problema de elétrons correlacionados é o problema do sinal fermiônico [129]. Em particular, como citado anteriormente, o método de difusão de Monte Carlo apresenta essa limitação. Especificamente, a maneira usual que o DMC contorna essa dificuldade é utilizando a aproximação de nós fixos [130], entretanto, o método permanece exato somente se a estrutura nodal da função guia usada for igual a estrutura nodal da função exata. Caso contrário, para estruturas nodais distintas, a energia calculada do sistema é um limite superior para a energia exata do estado fundamental [131].

Funções tentativas cujas estruturas nodais são aproximadamente iguais à do estado fundamental são amplamente desejadas, pois o problema do sinal fica atenuado ao se aplicar o método DMC. Além disso, com boas funções de onda tentativas, a aplicação do método variacional é vantajosa, porque no cálculo de propriedades diferentes da energia a avaliação é imediata, ao contrário do DMC, onde ainda é necessário extrapolações. Portanto, um dos objetivos desse estudo com pontos quânticos foi descobrir se função tentativa da Eq. 6.20, isto é, se redes neurais são capazes de produzir resultados melhores que os calculados pelo método DMC para esse sistema e encontrar funções tentativas com estruturas nodais melhores que as disponíveis na literatura. Outro objetivo foi analisar a natureza estocástica do otimizador e fazer uma avaliação quantitativa usando ferramentas estatísticas. Além disso, uma investigação de propriedades relacionadas com o escalonamento do tamanho da rede neural empregada também foi realizado.

Resultados encorajadores foram obtidos utilizando uma máquina de Boltzmann restrita para pontos quânticos circulares. Entretanto, a abordagem da Ref. [132] tem limitações, como a necessidade de incorporar *insights* analíticos sobre o sistema e as restrições da arquitetura implementada. Neste trabalho, melhores resultados utilizando técnicas de *deep learning* foram obtidos, como será discutido nas seções seguintes.

As simulações foram majoritariamente conduzidas em conjuntos de 4 placas gráficas NVIDIA A100 GPUs interconectadas. Dentre os hiper parâmetros relacionados com regularização e com o procedimento de atualização dos parâmetros, o momento, a atenuação, a constrição de norma, o decaimento da covariância móvel, a taxa de decaimento e a taxa de retardamento foram os mesmos utilizados para o caso dos aglomerados de hélio. Porém, a taxa de aprendizado foi usualmente escolhida igual à 0.05 e o tamanho do *batch* igual à 1024.

### 8.1 Analisando o caráter estocástico do otimizador

A natureza estocástica do otimizador tem sua origem na forma como as integrais necessárias são estimadas, uma vez que a integração de Monte Carlo é empregada para tal fim. Junto a isso, pela forma como a rede neural é construída e devido à quantidade de parâmetros, o funcional da energia variacional apresenta um amplo espaço de parâmetros com diversos mínimos locais equivalentes e de alta qualidade do ponto de vista da função perda [133]. Dessa forma, simulações independentes, isto é, simulações cujos números pseudo aleatórios utilizados são sequências diferentes e sem correlação entre si, chegam em resultados essencialmente independentes. Como a sequência de números aleatórios gerados é diferente, cada otimização segue um caminho diferente no espaço de parâmetros e, consequentemente, os parâmetros das funções de onda tentativa após a otimização são conjuntos distintos.

Para analisar a influência disso, 100 simulações independentes de  $10^5$  passos cada foram realizadas para um sistema de 12 elétrons. A arquitetura da função tentativa para esse caso utilizou 3 camadas, 8 determinantes, o canal de um elétron com largura de 56 e o de dois elétrons com largura de 20 neurônios. Especificamente, para amenizar as flutuações entre passos de otimização, foi utilizado médias móveis de m = 1600 passos conforme a Eq. A.9. Na Fig. 8.1 é apresentado em azul escuro a evolução da média das 100 simulações em função do passo de otimização. As áreas hachuradas em tons de azul mais claro são



Figura 8.1: Média móvel da energia total entre 100 simulações independentes em função da evolução das iterações do processo de treinamento em azul escuro para um ponto quântico contendo 12 elétrons. O desvio padrão móvel também é mostrado nas faixas em tons de azul claro, bem como a energia calculada por DMC em tracejado vermelho.

uma, duas e três vezes o desvio padrão móvel calculado através da Eq. A.10, indicando a incerteza relativa à média móvel. Ainda nessa figura, a linha vermelha tracejada representa um resultado obtido por DMC [134]. Fica claro que, mesmo com o ruído estocástico do

otimizador, as simulações consistentemente encontram energias menores que o resultado de referência.

Entretanto, não fica claro qual o tipo de distribuição os resultados das simulações seguem. Portanto, para identificar essa distribuição e sua evolução com os passos de otimização, a Fig. 8.2 mostra histogramas de frequência da média móvel da energia para três momentos das simulações. Do azul mais escuro para o mais claro, os histogramas



Figura 8.2: Histogramas de frequência da distribuição de energia para diferentes momentos do processo de otimização, sendo que os histogramas são construídos a partir das médias móveis nas iterações  $2^{12}$ ,  $2^{13}$  e  $2^{15}$  do azul escuro para o claro, respectivamente. As médias de cada distribuição são representadas por linhas tracejadas.

foram construídos com as energias nos passos de otimização  $2^{12}$ ,  $2^{13}$  e  $2^{15}$ . As linhas tracejadas indicam as médias de cada distribuição. Conforme o processo de otimização avança, a distribuição inicialmente larga evolui gradualmente para uma distribuição mais estreita. Ao mesmo tempo, a média da distribuição é deslocada para esquerda, mostrando a evolução da energia durante a otimização. A distribuição final da média móvel de energias é apresentada na Fig. 8.3, onde a linha tracejada vermelha representa o resultado de DMC apontado anteriormente e a tracejada azul representa a média da distribuição.

Como esperado, o processo estocástico junto com as medias moveis produziram uma distribuição que se assemelha à uma gaussiana. O erro proveniente dos ruídos do otimizador é dependente da quantidade de passos de otimização realizados e ele reduz gradualmente com o tempo de simulação. Além disso, também por ser um processo estocástico, esse erro tem flutuações, como se pode observar na Fig. 8.1 e na Fig. 8.2. Mesmo assim, a distribuição final demonstra a capacidade das redes neurais de produzirem resultados melhores que um método de Monte Carlo quântico como o DMC, que até o momento havia produzido as melhores energias obtidas. Note também que toda distribuição apresenta energias menores que as obtidas pelo método DMC.



Figura 8.3: Histograma da distribuição de energia ao final do processo de treinamento. A linha tracejada preta representa a média da distribuição, enquanto a tracejada vermelha representa a energia calculada através do DMC. Note a escala diferente das energias no eixo x com respeito a da Fig. 8.2.

## 8.2 Investigando os resultados e sucesso das redes neurais

#### Otimização da energia

Sabendo da capacidade de acurácia das redes neurais em representar estados quânticos, foram estudados quantum dots contendo 6, 12, 20 e 30 elétrons. Sistemas com esses números de partículas preenchem totalmente estados até um determinado nível de energia no caso não interagente, isto é, são camadas fechadas. Para determinar resultados de referência, os hiper parâmetros que definem o tamanho da rede neural considerada foram 4 camadas, 8 determinantes, 128 de largura para o canal de um elétron e de 32 para o canal de dois elétrons. Para computar a estimativa final da energia variacional, assim como feito no caso dos aglomerados de hélio, após o procedimento de otimização foi realizado um calculo de Monte Carlo variacional padrão, o qual manteve os parâmetros da função tentativa fixos e, através de uma simulação mais longa para obter resultados com uma variância menor, a estimativa foi computada junto com seu desvio de acordo com o procedimento descrito no Ap. A. O resumo dos resultados obtidos é apresentado na Tab. 8.1, onde os resultados também são comparados com outros cálculos presentes na literatura [134, 135]. As energias comparadas são provenientes de dois cálculos de difusão de Monte Carlo. A diferença entre os resultados das literaturas citadas e a energia obtida neste trabalho também são mostradas na coluna  $\Delta E$  em unidades de  $mH^*$ .

O processo de otimização da energia também pode ser visualizado na Fig. 8.4, onde é mostrado a evolução da energia variacional em função do número de iterações de otimização. Além disso, em tracejado laranja, energias calculadas pelo método DMC [135] são apresentadas para comparação. Pelo tamanho da rede ser padronizado, nota-se que

Ν	$\psi_{ m FN}$	$\Delta E \ [mH^*]$	DMC
6	7.59699(4)	3.11	7.6001(1)
12	25.62608(7)	9.52	25.6356(1)
20	61.9074(1)	14.60	61.922(2)
30	123.9469(2)	21.40	123.9683(2)

Tabela 8.1: Comparação entre valores de energia total em unidades de  $H^*$  obtidas com a função tentativa  $\psi_{\rm FN}$  e com o método DMC para pontos quânticos de camada fechada contendo N = 6, 12, 20 e 30 elétrons. A terceira coluna mostra a diferença entre os dois métodos em  $mH^*$ . Os dados do DMC foram extraídos da Ref. [134], com exceção para N = 30, que foi extraído da Ref. [135].



Figura 8.4: Evolução da energia durante o processo de otimização para pontos quânticos contendo 6 e 12 elétrons nos quadros superiores, 20 e 30 elétrons nos inferiores, ambos da direita para esquerda. A energia é mostrada em azul, enquanto os tracejados laranjas representam as energia calculadas pelo método DMC.

conforme o número de elétrons aumenta, a rede neural leva mais iterações para aprender a representar o estado fundamental. Dado que em todos os casos a rede foi capaz de aprender, essa arquitetura tem capacidade de representação de até pelo menos 30 elétrons.

#### Perfil de densidade bidimensional

De posse da representação do estado fundamental dada pela rede neural após a otimização, outras propriedades interessantes baseadas em valores esperados de operadores podem ser calculadas através da integração de Monte Carlo tradicional. Uma delas é perfil de densidade em duas dimensões. Essa quantidade é semelhante à função da Eq. 2.17, porém com resolução em direções. Nesse sentido, o operador pode ser escrito como:

$$\hat{\eta}_{2\mathrm{D}}(\mathbf{r}) = \sum_{i} \delta(\mathbf{r}_{i} - \mathbf{r}) , \qquad (8.1)$$

onde o perfil de densidade em duas dimensões é normalizado tal que  $\int \eta_{2D}(\mathbf{r}) dA = 1$ .

A Fig. 8.5 mostra um exemplo dessa propriedade para um ponto quântico contendo 20 elétrons. Qualitativamente, o perfil resultante apresenta simetria radial e um padrão de



Figura 8.5: Histograma bidimensional normalizado das posições dos elétrons em um ponto quântico contendo 20 elétrons. O perfil de densidade foi construído utilizando  $100 \times 2^{10}$  configurações amostradas da função de onda tentativa após o processo de otimização.

picos e vales envoltos por um envelope decrescente.

#### Estrutura nodal do estado fundamental

Outro aspecto interessante é a estrutura nodal da função de onda obtida. Em especial, esse aspecto pode trazer esclarecimentos do sucesso das redes neurais em descrever estados quânticos. Como mencionado anteriormente, métodos de Monte Carlo quântico como o DMC usualmente utilizam aproximação de nós fixos para transpor o problema do sinal fermiônico. Nesse caso, quanto melhor é a estrutura nodal da função guia, isto é, quanto mais as regiões onde a função guia se anula forem próximas da região em que a solução da equação se anula, melhores são os resultados obtidos por esses métodos. Portanto, para analisar a estrutura nodal, a Fig. 8.6 apresenta um mapa de cores do módulo da função no domínio do logaritmo para um ponto quântico contendo 20 elétrons. A figura reflete uma configuração especifica que foi amostrada da função tentativa após o procedimento de otimização. Para montar a imagem, a função de onda foi avaliada em diferentes coordenadas espaciais de um dos elétrons, enquanto todos os outros foram mantidos fixos. As setas pretas indicam a posição e a projeção do *spin* dos elétrons fixos e a seta em vermelho representa a coordenada original do elétron da configuração amostrada que não foi fixado. Vale ressaltar que a função tentativa não está normalizada e que o modelo assume um hamiltoniano bidimensional. Note que as linhas brancas revelam a estrutura nodal do estado para esta projeção da hiper superfície que define os zeros da função. Repare também que elétrons com a mesma projeção de *spin* que a da seta vermelha estão sempre sobre essa linhas brancas, como esperado de uma função anti-simétrica.



Figura 8.6: Mapa de cores do logaritmo da função de onda não normalizada em função das coordenadas e em escala arbitrária. Para compor o mapa, uma configuração foi amostrada da densidade de probabilidade gerada pela função de onda e a função foi computada ao varrer um dos elétrons por todo espaço bidimensional, com todos os outros elétrons fixos.

#### 8.3 Estabelecendo critérios de acurácia

Na seção anterior, foi mostrado que redes neurais são capazes de representar estados fundamentais de pontos quânticos. Além disso, os resultados obtidos foram consistentemente melhores que os alcançados pelo método DMC. Parte desse sucesso foi atribuído à capacidade das redes de melhorar as estruturas nodais das funções empregadas no DMC. Entretanto, as diversas possibilidades de tamanho da rede neural adicionam uma questão importante quando se trata da obtenção de resultados competitivos com os de referência. Em que ponto a rede é grande o suficiente para alcançar resultados desejados, porém não maior do que o necessário para que as simulações não percam em eficiência. Em outras palavras, qual seria o número mínimo de parâmetros necessário para atingir uma certa acurácia. Assim, o objetivo desta parte do trabalho foi entender melhor as características de escalonamento da rede neural, sem necessariamente obter os melhores resultados possíveis, mas sim como uma forma de entender o comportamento dos resultados com o tamanho da rede neural.

Para fazer um estudo nesse sentido é preciso definir um critério de qualidade. O critério adotado foi baseado na distância relativa da energia variacional com relação à energia do método DMC [134], isto é,

$$\Delta E = E_{\rm DMC} - E_{\rm VMC} \ . \tag{8.2}$$

Se o valor de  $\Delta$  for menor que 2  $mH^*$ , o critério de qualidade é alcançado. Além disso, devido a natureza estocástica do otimizador demonstrada anteriormente, o procedimento de cálculo de  $E_{\rm VMC}$  para realizar a comparação considera uma média entre 10 simulações independentes da média dos últimos 1600 passos de otimização de cada uma delas. O desvio padrão entre as 10 simulações também é calculado.

Com esse procedimento em mente, diversas simulações com tamanhos diferentes de rede neural foram realizadas. Por completeza, vale relembrar que os hiper parâmetros que definem a quantidade de parâmetros variacionais são o número de camadas L, o número de termos considerados na expansão de determinantes da Eq. 6.20, a largura do canal de um elétron e a largura do canal de dois elétrons. Dessa forma, as simulações realizadas consideraram redes com duas camadas e apenas um termo de determinante. Mantendo esses dois hiper parâmetros fixos, as larguras dos canais foram variadas independentemente a fim de estudar o comportamento do escalonamento da rede em função desses hiper parâmetros.

Para um ponto quântico contendo N = 6 elétrons, as energias variacionais  $E_{\rm VMC}$ em função da largura do canal de um elétron são apresentadas na Fig. 8.7, onde as di-



Figura 8.7: Comportamento da energia de um ponto quântico contendo 6 elétrons com o aumento da largura do canal de um elétron. As diferentes cores representam larguras diferentes para o canal de dois elétrons, sendo que o azul, laranja, verde e vermelho correspondem à 8, 12, 16 e 24 de largura.

ferentes cores dos pontos indicam diferentes larguras para o canal de dois elétrons, sendo que as cores azul, laranja, verde e vermelho representam larguras de 8, 12, 16 e 24, respectivamente. Junto a isso, a linha tracejada preta indica a energia de referência e a linha tracejada cinza indica quando o critério de qualidade é atingido. Equivalentemente, na Fig. 8.8 são apresentados resultados semelhantes para um ponto quântico contendo N = 12 elétrons. Em ambas figuras é possível identificar uma tendência da rede neural de atingir a acurácia cada vez com larguras menores para o canal de um elétron conforme o canal de dois elétrons é aumentado. Particularmente na Fig. 8.8 é notável as flutuações geradas pela estocasticidade do otimizador, uma vez que redes neurais de tamanhos semelhantes apresentam resultados estatisticamente equivalentes, mesmo que aumentando o número de parâmetros. Esse comportamento pode ser observado principalmente nos pontos azuis e laranjas nas regiões de 135 e 110 de largura, respectivamente.



Figura 8.8: Comportamento da energia de um ponto quântico contendo 12 elétrons com o aumento da largura do canal de um elétron. As diferentes cores representam larguras diferentes para o canal de dois elétrons, sendo que o azul, laranja, verde e vermelho correspondem à 8, 12, 16 e 24 de largura.

Por último, com o intuito de definir critérios qualitativos de quando uma rede neural é grande o suficiente para produzir resultados competitivos com outros métodos de Monte Carlo quântico, a Fig. 8.9 apresenta um mapa de cores da energia relativa  $\Delta E$ entre a energia calculada através do DMC e as energias obtidas empregando funções  $\psi_{\rm FN}$ com diferentes tamanhos de rede neural para um ponto quântico contendo 6 elétrons. O



Figura 8.9: Mapa de calor da variação de energia  $\Delta E$  entre os resultados obtidos usando o estado tentativo  $\psi_{\rm FN}$  e cálculos de DMC em função das larguras do canal de um e dois elétrons. As retas tracejadas separam qualitativamente as regiões onde o tamanho da rede neural é insuficiente, indeterminado e suficiente para desempenhar tão bem quanto o método DMC.

eixo horizontal mostra a largura do canal de um elétron, enquanto o eixo vertical mostra a largura do canal de dois elétrons. Nesse caso, para todas as energias calculadas foi considerado apenas um determinante na expansão multi orbital e redes neurais compostas de duas camadas. O mapa de cores é uma interpolação de uma grade 10 por 10, onde as larguras consideradas para o canal de um e dois elétrons foram números pares de 14 até 32 e de 6 até 24, respectivamente. Duas retas traçadas qualitativamente separam a região do espaço de hiper parâmetros considerada em três partes. A região abaixo da linha tracejada preta são redes neurais pequenas demais para serem competitivas, alcançando apenas energias maiores que a energia de DMC, mesmo considerando o erro estatístico desses resultados da literatura. A região entre as duas retas engloba redes neurais com tamanhos intermediários e que apresentam resultados com desempenho equivalente ou muito próximo ao DMC. Essa região pode ser vista como uma transição na qual é indeterminado dizer quais resultados são melhores. Por fim, a região acima da linha tracejada vermelha define as redes neurais cujos tamanhos são grandes o suficiente para atingir resultados melhores que de difusão de Monte Carlo.

Nesse sentido, para obter resultados com energias estatisticamente melhores que aquelas calculadas usando o DMC, deve-se escolher hiper parâmetros na região azul escuro o mais próximo possível da reta vermelha, porém ainda acima dela. Note que essa escolha é qualitativa e que não necessariamente será a menor energia alcançável pela rede neural, entretanto, é a escolha mais eficiente computacionalmente para obter energias menores que as obtidas pelo DMC, mesmo que minimamente. Além disso, empregando esse critério e extrapolando o comportamento da energia otimizada em função do tamanho de cada canal para outros números de partículas, uma boa proporção seria de sempre utilizar o tamanho do canal de um elétron aproximadamente 1.4 vezes maior que o canal de duas partículas. Na busca para entender o sucesso das ferramentas de inteligência artificial aplicadas à pesquisa científica, especialmente no estudo de sistemas finitos, investigou-se a utilização de redes neurais artificias para representar funções de onda tentativa com o intuito de descrever estados quânticos de aglomerados de hélio e pontos quânticos. Ao explorar essa representação através do método variacional, foi possível verificar que as funções tentativas propostas foram capazes de representar o estado fundamental de ambos sistemas precisamente.

O processo de otimização da energia demonstrou ser sensível ao comportamento do potencial de interação entre partículas, já que, para aglomerados de hélio, o treinamento apresentou instabilidades. Contornar esse problema foi desafiador, porém, abusando da universalidade das correlações de curto alcance esperadas para esse sistema, adicionar o potencial de esfera dura durante a etapa de treinamento resolveu esse impasse. Outro fator importante para obter resultados competitivos foi a inclusão de uma etapa após o procedimento de otimização na qual aumenta-se a estatística mantendo os parâmetros variacionais fixos para obter estimativas mais precisas das propriedades de interesse. Além disso, a construção dos atributos de entrada que respeitassem simetria de translação global foram cruciais para tornar o treinamento factível, mostrando a importância de construir redes neurais que contenham informações sobre a tarefa que se deseja desempenhar.

Mesmo sendo treinadas apenas através da energia, as redes neurais também foram capazes de obter resultados relevantes e de ótima qualidade para outras propriedades, incluindo aquelas cujos os operadores não comutam com o hamiltoniano. Isso demonstra a aptidão de generalização dessa classe de funções, aspecto fundamental que diferencia um simples ajuste dos algoritmos de aprendizagem, como também é uma vantagem com relação a outros métodos de Monte Carlo quântico. Particularmente, para os aglomerados de hélio, foram calculados o perfil de densidades, a distribuição e densidade de pares, a distribuição de Jacobi e os contatos. Com essas quantidades, especificamente o colapso das densidades de pares e da distribuição de Jacobi, analisou-se o comportamento universal de curto alcance dos aglomerados. Devido essa situação intermediária chamada de universalidade fraca, na qual a distância média entre as partículas nos aglomerados são da mesma ordem de grandeza que o comprimento de van der Waals, esses sistemas exibem uma riqueza de fenômenos interessante, evidenciada pelos resultados obtidos. Boa parte dos resultados referentes esse estudo foram publicados [115] e os adicionais estão em processo de escrita.

Por outro lado, na investigação do sucesso das redes neurais descrevendo aglomerados, também foi analisado o comportamento do processo de treinamento ao considerar redes com diferentes representações para os atributos de entrada, ou seja, como a informação é disponibilizada para ser processada pela rede. Observou-se principalmente que a qualidade da representação é essencial para garantir um treinamento eficiente e que as redes são capazes de aprender características presentes no sistema mesmo não sendo construídas sujeitas as restrições dessas características, como no caso das versões da rede neural com dependência angular aplicada ao dímero de hélio. Essa análise, o estudo do modelo de gota e do escalonamento de tempo de simulação também foram publicados [116].

Já no estudo dos pontos quânticos, o aspecto estocástico do otimizador foi explorado para entender como o fator aleatório intrínseco ao método poderia influenciar nos resultados. Através de simulações independentes e da análise por médias móveis, percebeu-se que as simulações tendem à chegar em resultados equivalentes se estendidas por iterações suficientes. Equivalentemente, a distribuição respeitada pelas energias otimizadas das simulações independentes tende a ser uma distribuição gaussiana, cuja média e variância móveis ficam cada vez menores com o aumento do número de iterações.

Ainda analisando pontos quânticos, porém sob outro enfoque, a influência do tamanho da rede neural no treinamento também foi investigada. O mapa de calor resultante desse estudo foi dividido em três regiões de hiper parâmetros de maneira qualitativa. Uma das regiões condiz com as redes neurais com tamanho grande o suficiente para obter valores de energia tão bom quanto ou melhores que cálculos de Monte Carlo quântico. Uma segunda região define as redes pequenas, que não são capazes de alcançar a qualidade de outros métodos de Monte Carlo. Por fim, a terceira região corresponde uma fase de transição, na qual não é claro a qual das outras duas regiões pertence o resultado. Essa região de transição tem origem na natureza estocástica do otimizador. Com esse estudo, foi possível estabelecer critérios qualitativos para a escolha do tamanho da rede neural nesses sistemas.

Outra importante análise realizada foi a investigação da qualidade da estrutura nodal das funções tentativa baseadas em redes neurais para sistemas compostos de férmions, particularmente pontos quânticos. Esse tipo de estudo é pouco explorado na literatura [130], entretanto, comparar as estruturas nodais das funções utilizadas em métodos de Monte Carlo quânticos e funções baseadas em redes neurais é essencial para entender o sucesso dessa ferramenta, já que a qualidade dos resultados de métodos como o DMC são intimamente relacionada com a estrutura nodal empregada. Os resultados referentes aos pontos quânticos também estão sendo preparados para publicação.

Por último, há algumas direções subsequentes interessantes de se explorar. Dentro do contexto de universalidades e aglomerados de hélio, seria vantajoso fazer ajustes na forma da função  $\psi_{\rm BHC}$  da Eq. 6.15 com intuito de tornar o treinamento mais eficiente e habilitar a aplicação das redes neurais em sistemas compostos de mais partículas. Usualmente, o estudo de super-sólidos e superfluidos empregam simulações contento em torno de 64 até 128 átomos de hélio, entretanto, o primeiro objetivo seria alcançar uma simulação eficiente e com resultados acurados para até 32 átomos. Nesse mesmo sentido, adaptar o programa [122] para sistemas com condições periódicas de contorno e dimensões reduzidas ampliaria as possíveis aplicações e estudos.

Outra via a ser pesquisada compete aos interesses recentes dentro do estudo da unitariedade e comportamentos universais [4,73]. Uma alternativa de perspectiva seria avançar nessa direção, explorando invariância de escala ao olhar para estados excitados em sistemas compostos por bósons interagindo perto da ressonância, por exemplo. Também razoavelmente relacionado essa área, a pesquisa em gases de Fermi aprisionados e na presença de impurezas [136] também parece promissora para a aplicação de redes neurais como funções tentativa. Já dentro da área de computação quântica, uma extensão do estudo dos pontos quânticos poderia ser feita através da investigação de íons aprisionados [137]. Com relação aos métodos e as redes neurais artificiais, duas ótimas ferramentas que podem ajudar na pesquisa desses temas são os pacotes DeepMD-kit [138] e NetKet [139]. A inclusão desses pacotes em conjunto com as ideias apresentadas para construção de funções tentativa no Cap. 6 formam uma caixa de ferramentas poderosa para auxiliar no desenvolvimento da pesquisa científica.

Em resumo, esse trabalho trouxe contribuições relevantes para o entendimento do sucesso das redes neurais aplicadas na representação de estados quânticos variacionais. Servindo-se da sinergia presente entre algoritmos de *deep learning* e o método variacional, sistemas finitos de muitos corpos formados por bósons ou por férmions foram analisados, proporcionando uma descrição acurada do estado fundamental e indicando novos caminhos para se explorar na física de sistemas complexos, em particular, sistemas finitos. Esses resultados tem o potencial de pavimentar o desenvolvimento de funções variacionais ainda mais precisas e eficientes, capazes de descrever uma ampla gama de sistemas além de aglomerados de hélio e pontos quânticos. Finalmente, a trajetória traçada até aqui foi muitas vezes desafiadora e algumas vezes gratificante, além de ser cheia de picos e vales. Ao seguir para as próximas etapas, a esperança que fica é de que as ideias apresentadas sirvam de inspiração e inovação, ajudando a desenvolver os esforços coletivos na pesquisa científica.

## BIBLIOGRAFIA

- Schmitt, A. Introduction to superfluidity field-theoretical approach and applications (2014). URL https://doi.org/10.48550/arXiv.1404.1284. 1404.1284.
- [2] Szalewicz, K. Interplay between theory and experiment in investigations of molecules embedded in superfluid helium nanodroplets. *International Reviews in Physical Chemistry* 27, 273–316 (2008). URL https://doi.org/10.1080/01442350801933485.
- Barranco, M. et al. Helium nanodroplets: An overview. Journal of Low Temperature Physics 142, 1–81 (2006). URL https://doi.org/10.1007/s10909-005-9267-0.
- Bazak, B., Valiente, M. & Barnea, N. Universal short-range correlations in bosonic helium clusters. *Physical Review A* 101, 010501 (2020-01). URL https://doi.org/ 10.1103/physreva.101.010501.
- Reimann, S. M. & Manninen, M. Electronic structure of quantum dots. *Rev. Mod. Phys.* 74, 1283-1342 (2002). URL https://link.aps.org/doi/10.1103/ RevModPhys.74.1283.
- [6] Harju, A. Variational monte carlo for interacting electrons in quantum dots. Journal of Low Temperature Physics 140, 181–210 (2005). URL https://doi.org/10.1007/s10909-005-6308-7.
- TURING, A. M. I.—COMPUTING MACHINERY AND INTELLIGENCE. Mind LIX, 433-460 (1950). URL https://doi.org/10.1093/mind/LIX.236.433.
- [8] Uchendu, A., Ma, Z., Le, T., Zhang, R. & Lee, D. TURINGBENCH: A benchmark environment for turing test in the age of neural text generation. *CoRR* abs/2109.13296 (2021). URL https://arxiv.org/abs/2109.13296.
- [9] Dargan, S., Bansal, S., Kumar, M., Mittal, A. & Kumar, K. Augmented reality: A comprehensive review. ARCHIVES OF COMPUTATIONAL METHODS IN ENGINEERING 30, 1057–1080 (2023). URL https://doi.org/10.1007/s11831-022-09831-7.
- [10] Wang, Z. et al. A review of vehicle detection techniques for intelligent vehicles. IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS 34, 3811–3831 (2023). URL https://doi.org/10.1109/TNNLS.2021.3128968.
- [11] Raucci, U. et al. Voice-controlled quantum chemistry. NATURE COM-PUTATIONAL SCIENCE 1, 42–45 (2021). URL https://doi.org/10.1038/ s43588-020-00012-9.
- [12] Li, L., Lei, B. & Mao, C. Digital twin in smart manufacturing. JOURNAL OF INDUSTRIAL INFORMATION INTEGRATION 26 (2022). URL https://doi. org/10.1016/j.jii.2021.100289.
- [13] Goldstine, H. H. The computer from Pascal to von Neumann (Princeton University Press, 1993).
- [14] Lovelace, A. A. Sketch of the analytical engine invented by charles babbage, by lf menabrea, officer of the military engineers, with notes upon the memoir by the translator. *Taylor's Scientific Memoirs* 3, 666-731 (1842). URL https://www. fourmilab.ch/babbage/sketch.html.
- [15] Moore, G. E. Cramming more components onto integrated circuits. Proceedings of the IEEE 86, 82-85 (1998). URL http://www.computer-architecture.org/ textual/Moore-Cramming-More-Components-1965.pdf.
- [16] Goodfellow, I., Bengio, Y. & Courville, A. Deep learning (MIT press, 2016).
- [17] McCulloch, W. S. & Pitts, W. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics 5, 115–133 (1943). URL https://doi.org/10.1007/BF02478259.
- [18] Rosenblatt, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review* 65, 386 (1958). URL https:// psycnet.apa.org/doi/10.1037/h0042519.
- [19] Fukushima, K. Cognitron: A self-organizing multilayered neural network. Biological Cybernetics 20, 121–136 (1975). URL https://doi.org/10.1007/BF00342633.
- [20] Rumelhart, D. E., McClelland, J. L. & Group, P. R. Parallel Distributed Processing, Volume 1: Explorations in the Microstructure of Cognition: Foundations (The MIT Press, 1986). URL https://doi.org/10.7551/mitpress/5236.001.0001.
- [21] Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986). URL https://doi.org/10. 1038/323533a0.
- [22] Hinton, G. E. To recognize shapes, first learn to generate images. Progress in brain research 165, 535-547 (2007). URL https://api.semanticscholar.org/ CorpusID:10327263.
- [23] LeCun, Y., Bengio, Y. et al. Scaling learning algorithms towards ai. Large-scale kernel machines 5, 127–168 (2007). URL https://api.semanticscholar.org/CorpusID: 15559637.

- [24] Carrasquilla, J. & Melko, R. G. Machine learning phases of matter. Nature Physics 13, 431–434 (2017). URL https://doi.org/10.1038/nphys4035.
- [25] Carleo, G. et al. Machine learning and the physical sciences. Rev. Mod. Phys. 91, 045002 (2019). URL https://link.aps.org/doi/10.1103/RevModPhys.91. 045002.
- [26] Györgyi, G. First-order transition to perfect generalization in a neural network with binary synapses. *Phys. Rev. A* 41, 7097–7100 (1990). URL https://link.aps.org/ doi/10.1103/PhysRevA.41.7097.
- [27] Albert, J. & Swendsen, R. H. The inverse ising problem. *Physics Procedia* 57, 99-103 (2014). URL https://www.sciencedirect.com/science/article/pii/S1875389214002855. Proceedings of the 27th Workshop on Computer Simulation Studies in Condensed Matter Physics (CSP2014).
- [28] Ntampaka, M. et al. A machine learning approach for dynamical mass measurements of galaxy clusters. The Astrophysical Journal 803, 50 (2015). URL https://dx. doi.org/10.1088/0004-637X/803/2/50.
- [29] Guest, D., Cranmer, K. & Whiteson, D. Deep learning and its application to lhc physics. Annual Review of Nuclear and Particle Science 68, 161– 181 (2018). URL https://www.annualreviews.org/content/journals/10.1146/ annurev-nucl-101917-021019.
- [30] Deng, D.-L., Li, X. & Das Sarma, S. Quantum entanglement in neural network states. Phys. Rev. X 7, 021021 (2017). URL https://link.aps.org/doi/10.1103/ PhysRevX.7.021021.
- [31] Pfau, D., Spencer, J. S., Matthews, A. G. D. G. & Foulkes, W. M. C. Ab initio solution of the many-electron schrödinger equation with deep neural networks. *Phys. Rev. Res.* 2, 033429 (2020). URL https://link.aps.org/doi/10.1103/ PhysRevResearch.2.033429.
- [32] Wu, D., Wang, L. & Zhang, P. Solving statistical mechanics using variational autoregressive networks. *Phys. Rev. Lett.* **122**, 080602 (2019). URL https: //link.aps.org/doi/10.1103/PhysRevLett.122.080602.
- [33] Wen, T., Zhang, L., Wang, H., E, W. & Srolovitz, D. J. Deep potentials for materials science. *Materials Futures* 1, 022601 (2022). URL https://dx.doi.org/10.1088/2752-5724/ac681d.

- [34] Bardeen, J., Cooper, L. N. & Schrieffer, J. R. Theory of superconductivity. *Phys. Rev.* 108, 1175–1204 (1957). URL https://link.aps.org/doi/10.1103/PhysRev. 108.1175.
- [35] Anderson, P. Resonating valence bonds: A new kind of insulator? Materials Research Bulletin 8, 153-160 (1973). URL https://www.sciencedirect.com/science/ article/pii/0025540873901670.
- [36] Laughlin, R. B. Anomalous quantum hall effect: An incompressible quantum fluid with fractionally charged excitations. *Phys. Rev. Lett.* 50, 1395–1398 (1983). URL https://link.aps.org/doi/10.1103/PhysRevLett.50.1395.
- [37] Cybenko, G. Approximation by superpositions of a sigmoidal function. Mathematics of Control, Signals and Systems 2, 303-314 (1989). URL https://doi.org/10. 1007/BF02551274.
- [38] Nagy, A. & Savona, V. Variational quantum monte carlo method with a neuralnetwork ansatz for open quantum systems. *Physical Review Letters* 122, 250501 (2019). URL https://link.aps.org/doi/10.1103/PhysRevLett.122.250501.
- [39] Roth, C., Szabó, A. & MacDonald, A. H. High-accuracy variational monte carlo for frustrated magnets with deep neural networks. *Physical Review B* 108, 054410 (2023). URL https://link.aps.org/doi/10.1103/PhysRevB.108.054410.
- [40] Yang, L. et al. Deep learning-enhanced variational monte carlo method for quantum many-body physics. Physical Review Research 2, 012039 (2020). URL https:// link.aps.org/doi/10.1103/PhysRevResearch.2.012039.
- [41] Qian, Y., Fu, W., Ren, W. & Chen, J. Interatomic force from neural network based variational quantum monte carlo. The Journal of Chemical Physics 157, 164104 (2022). URL https://pubs.aip.org/jcp/article/157/16/164104/ 2841958/Interatomic-force-from-neural-network-based.
- [42] Pescia, G., Han, J., Lovato, A., Lu, J. & Carleo, G. Neural-network quantum states for periodic systems in continuous space. *Phys. Rev. Res.* 4, 023138 (2022). URL https://link.aps.org/doi/10.1103/PhysRevResearch.4.023138.
- [43] Guardiola, R., Kornilov, O., Navarro, J. & Peter Toennies, J. Magic numbers, excitation levels, and other properties of small neutral he4 clusters (n≤50). The Journal of Chemical Physics 124, 084307 (2006). URL https://doi.org/10.1063/1.2140723.
- [44] Pescia, G., Han, J., Lovato, A., Lu, J. & Carleo, G. Neural-network quantum states for periodic systems in continuous space. *Phys. Rev. Research* 4, 023138 (2022). URL https://link.aps.org/doi/10.1103/PhysRevResearch.4.023138.

- [45] Whitlock, P. A., Chester, G. V. & Kalos, M. H. Monte carlo study of <sup>4</sup>He in two dimensions. *Phys. Rev. B* 38, 2418–2425 (1988). URL https://link.aps.org/doi/ 10.1103/PhysRevB.38.2418.
- [46] Krishnamachari, B. & Chester, G. V. Monte carlo studies of two-dimensional phases of helium using a shadow wave function. *Phys. Rev. B* 61, 9677–9685 (2000). URL https://link.aps.org/doi/10.1103/PhysRevB.61.9677.
- [47] Ceperley, D. M. Path integrals in the theory of condensed helium. *Rev. Mod. Phys.* 67, 279-355 (1995). URL https://link.aps.org/doi/10.1103/RevModPhys.67. 279.
- [48] Pedroso, V. Z., Pedroso, V. Z. & Vitiello, S. Metastable solid <sup>4</sup>he and the possible role of point defects. *Journal of Physics: Condensed Matter* (2020). URL https: //doi.org/10.1088/1361-648x/abc5d6.
- [49] Guardiola, R. & Navarro, J. Minimal mass size of a stable <sup>3</sup>He cluster. *Phys. Rev.* A 71, 035201 (2005). URL https://link.aps.org/doi/10.1103/PhysRevA.71. 035201.
- [50] Pandharipande, V. R., Pieper, S. C. & Wiringa, R. B. Variational monte carlo calculations of ground states of liquid <sup>4</sup>He and <sup>3</sup>He drops. *Phys. Rev. B* 34, 4571– 4582 (1986). URL https://link.aps.org/doi/10.1103/PhysRevB.34.4571.
- [51] Mani, D. et al. A helium nanodroplet setup for mid and far-infrared spectroscopy using pulsed-free-electron lasers: vibrational spectra of propargyl alcohol. RSC Adv. 7, 54318–54325 (2017). URL http://dx.doi.org/10.1039/C7RA08102B.
- [52] Andronikashvili, E. E11 a direct observation of two kinds of motion in helium ii - j. phys. (u.s.s.r.) 10, 201 (1946). In Galasiewicz, Z. M. (ed.) *Helium 4*, 154– 165 (Pergamon, 1971). URL https://www.sciencedirect.com/science/article/ pii/B9780080158167500158.
- [53] Ceperley, D. M. & Manousakis, E. Path integral monte carlo applications to quantum fluids in confined geometries. *The Journal of Chemical Physics* 115, 10111–10118 (2001). URL https://aip.scitation.org/doi/abs/10.1063/1.1418443.
- [54] Grebenev, S., Toennies, J. P. & Vilesov, A. F. Superfluidity within a small helium-4 cluster: The microscopic andronikashvili experiment. *Science* 279, 2083–2086 (1998). URL https://www.science.org/doi/abs/10.1126/science.279.5359.2083.
- [55] Zeng, T., Li, H. & Roy, P.-N. Potential generation and path-integral monte carlo in study of microscopic superfluidity. *International Journal of Quantum Chemis*-

*try* **115**, 535-540 (2014). URL https://onlinelibrary.wiley.com/doi/abs/10. 1002/qua.24815.

- [56] Allen, M. P. & Tildesley, D. J. Computer Simulation of Liquids (Oxford University Press, 2017). URL https://doi.org/10.1093/oso/9780198803195.001.0001.
- [57] Lennard-Jones, J. E. Cohesion. Proceedings of the Physical Society 43, 461 (1931).
   URL https://dx.doi.org/10.1088/0959-5309/43/5/301.
- [58] McMillan, W. L. Ground state of liquid he<sup>4</sup>. Phys. Rev. **138**, A442-A451 (1965). URL https://link.aps.org/doi/10.1103/PhysRev.138.A442.
- [59] Scoles, G. Two-body, spherical, atom-atom, and atom-molecule interaction energies. Annual Review of Physical Chemistry 31, 81-96 (1980). URL https://doi.org/ 10.1146/annurev.pc.31.100180.000501.
- [60] Ronald A. Aziz, F. R. M. & Wong, C. C. A new determination of the ground state interatomic potential for he2. *Molecular Physics* 61, 1487–1511 (1987). URL https://doi.org/10.1080/00268978700101941.
- [61] Chester, G. & Reatto, L. The ground state of liquid helium four. *Physics Letters* 22, 276-277 (1966). URL https://www.sciencedirect.com/science/article/pii/003191636690610X.
- [62] O'Reilly, D. E. & Tsang, T. New theory of superfluid <sup>4</sup>He. *Phys. Rev. B* 19, 5634–5640 (1979). URL https://link.aps.org/doi/10.1103/PhysRevB.19.5634.
- [63] Krishna, M. V. R. & Whaley, K. B. Wave functions of helium clusters. The Journal of Chemical Physics 93, 6738-6751 (1990). URL http://dx.doi.org/10.1063/1. 458943.
- [64] Naidon, P. & Endo, S. Efimov physics: a review. Reports on Progress in Physics 80, 056001 (2017). URL https://dx.doi.org/10.1088/1361-6633/aa50e8.
- [65] Messiah, A. Quantum Mechanics. No. v. 2 in Dover books on physics (Dover Publications, 1999).
- [66] Braaten, E. & Hammer, H.-W. Universality in few-body systems with large scattering length. *Phys. Rep.* 428, 259–390 (2006). URL http://dx.doi.org/10.1016/j. physrep.2006.03.001.
- [67] Shea, P., van Zyl, B. P. & Bhaduri, R. K. The two-body problem of ultra-cold atoms in a harmonic trap. American Journal of Physics 77, 511–515 (2009). URL https://doi.org/10.1119/1.3013812.

- [68] Bhaduri, R. K., Chatterjee, A. & van Zyl, B. P. An elementary exposition of the Efimov effect. American Journal of Physics 79, 274–281 (2011). URL https://doi. org/10.1119/1.3533428.
- [69] Efimov, V. Energy levels arising from resonant two-body forces in a three-body system. *Physics Letters B* 33, 563-564 (1970). URL https://www.sciencedirect. com/science/article/pii/0370269370903497.
- [70] Tan, S. Energetics of a strongly correlated Fermi gas. Ann. Phys. 323, 2952
   2970 (2008). URL http://www.sciencedirect.com/science/article/pii/ S0003491608000456.
- [71] Tan, S. Large momentum part of a strongly correlated Fermi gas. Ann. Phys. 323, 2971-2986 (2008). URL http://www.sciencedirect.com/science/article/pii/ S0003491608000432.
- [72] Tan, S. Generalized virial theorem and pressure relation for a strongly correlated Fermi gas. Ann. Phys. 323, 2987 – 2990 (2008). URL http://www.sciencedirect. com/science/article/pii/S0003491608000420.
- [73] Yates, A. J. & Blume, D. Structural properties of <sup>4</sup>He<sub>N</sub> (N=2-10) clusters for different potential models at the physical point and at unitarity. *Physical Review A* 105, 022824 (2022-02). URL https://doi.org/10.1103/physreva.105.022824.
- [74] Brack, M. The physics of simple metal clusters: self-consistent jellium model and semiclassical approaches. *Rev. Mod. Phys.* 65, 677–732 (1993). URL https://link. aps.org/doi/10.1103/RevModPhys.65.677.
- [75] Leggett, A. J. Bose-einstein condensation in the alkali gases: Some fundamental concepts. *Rev. Mod. Phys.* 73, 307–356 (2001). URL https://link.aps.org/doi/ 10.1103/RevModPhys.73.307.
- [76] Agarwal, K., Rai, H. & Mondal, S. Quantum dots: an overview of synthesis, properties, and applications. *Materials Research Express* 10, 062001 (2023). URL https://dx.doi.org/10.1088/2053-1591/acda17.
- [77] Lin, X. et al. Electrically-driven single-photon sources based on colloidal quantum dots with near-optimal antibunching at room temperature. Nature Communications 8, 1132 (2017). URL https://doi.org/10.1038/s41467-017-01379-6.
- [78] Vandersypen, L. M. K. et al. Interfacing spin qubits in quantum dots and donors hot, dense, and coherent. npj Quantum Information 3, 34 (2017). URL https: //doi.org/10.1038/s41534-017-0038-y.

- [79] Jian, H.-J. et al. Super-cationic carbon quantum dots synthesized from spermidine as an eye drop formulation for topical treatment of bacterial keratitis. ACS Nano 11, 6703–6716 (2017). URL https://doi.org/10.1021/acsnano.7b01023. PMID: 28677399.
- [80] Meinardi, F. et al. Highly efficient luminescent solar concentrators based on earthabundant indirect-bandgap silicon quantum dots. Nature Photonics 11, 177–185 (2017). URL https://doi.org/10.1038/nphoton.2017.5.
- [81] Lee, K.-H. et al. Highly efficient, color-reproducible full-color electroluminescent devices based on red/green/blue quantum dot-mixed multilayer. ACS Nano 9, 10941-10949 (2015). URL https://doi.org/10.1021/acsnano.5b05513. PMID: 26435403.
- [82] Tarucha, S., Austing, D. G., Honda, T., Van Der Hage, R. & Kouwenhoven, L. P. Atomic-like properties of semiconductor quantum dots. *Japanese journal of applied physics* 36, 3917 (1997).
- [83] Ashcroft, N. & Mermin, N. Solid State Physics. HRW international editions (Holt, Rinehart and Winston, 1976).
- [84] Ashoori, R. C. Electrons in artificial atoms. Nature 379, 413-419 (1996). URL https://doi.org/10.1038/379413a0.
- [85] Ciftja, O. Understanding electronic systems in semiconductor quantum dots. *Physica Scripta* 88, 058302 (2013). URL https://dx.doi.org/10.1088/0031-8949/88/05/058302.
- [86] Wigner, E. On the interaction of electrons in metals. *Phys. Rev.* 46, 1002–1011 (1934). URL https://link.aps.org/doi/10.1103/PhysRev.46.1002.
- [87] Pederiva, F., Umrigar, C. J. & Lipparini, E. Diffusion monte carlo study of circular quantum dots. *Phys. Rev. B* 62, 8120-8125 (2000). URL https://link.aps.org/ doi/10.1103/PhysRevB.62.8120.
- [88] Cohen-Tannoudji, C., Diu, B., Laloe, F. & Dui, B. Quantum Mechanics (2 vol. set) (Wiley-Interscience, 2006).
- [89] SILVA, W. F. e. Propriedades de pontos quânticos circulares. Master's thesis, Universidade Estadual de Campinas, Instituto de Física Gleb Wataghin (2020). URL https://doi.org/10.47749/T/UNICAMP.2020.1157335.
- [90] Mehta, P. et al. A high-bias, low-variance introduction to machine learning for physicists. Physics Reports 810, 1-124 (2019). URL https://www.sciencedirect.com/

science/article/pii/S0370157319300766. A high-bias, low-variance introduction to Machine Learning for physicists.

- [91] Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In Pereira, F., Burges, C., Bottou, L. & Weinberger, K. (eds.) Advances in Neural Information Processing Systems, vol. 25 (Curran Associates, Inc., 2012). URL https://proceedings.neurips.cc/paper\_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.
- [92] Dahl, G. E., Yu, D., Deng, L. & Acero, A. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on Au*dio, Speech, and Language Processing 20, 30–42 (2012). URL https://doi.org/ 10.1109/TASL.2011.2134090.
- [93] Goodfellow, I. J., Mirza, M., Xiao, D., Courville, A. & Bengio, Y. An empirical investigation of catastrophic forgetting in gradient-based neural networks. arXiv.org (2015). URL https://arxiv.org/abs/1312.6211.
- [94] Mitchell, T. Machine learning. Publisher: McGraw Hill (1997). URL http://www. cs.cmu.edu/~tom/mlbook.html.
- [95] Taigman, Y., Yang, M., Ranzato, M. & Wolf, L. Deepface: Closing the gap to humanlevel performance in face verification. In 2014 IEEE Conference on Computer Vision and Pattern Recognition, 1701–1708 (2014).
- [96] Cavalcante, R. C., Brasileiro, R. C., Souza, V. L. P., Nobrega, J. P. & Oliveira, A. L. I. Computational intelligence and financial markets: A survey and future directions. *EXPERT SYSTEMS WITH APPLICATIONS* 55, 194–211 (2016). URL https://doi.org/10.1016/j.eswa.2016.02.006.
- [97] Bottou, L. On-line Learning and Stochastic Approximations, 9-42. Publications of the Newton Institute (Cambridge University Press, 1999). URL https://leon. bottou.org/publications/pdf/online-1998.pdf.
- [98] Polyak, B. T. & Juditsky, A. B. Acceleration of stochastic approximation by averaging. SIAM Journal on Control and Optimization 30, 838-855 (1992). URL https://doi.org/10.1137/0330046.
- [99] Martens, J. & Grosse, R. B. Optimizing neural networks with kronecker-factored approximate curvature. In ICML'15: Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37 (2015). URL https://dl.acm.org/doi/10.5555/3045118.3045374.

- [100] Amari, S.-i. Natural Gradient Works Efficiently in Learning. Neural Computation 10, 251–276 (1998). URL https://doi.org/10.1162/089976698300017746.
- [101] Botev, A. & Martens, J. KFAC-JAX (2022). URL https://github.com/ google-deepmind/kfac-jax.
- [102] James S. Spencer, D. P. & Contributors, F. FermiNet (2020). URL http://github. com/deepmind/ferminet.
- [103] Bolotin, A. Computational solution to quantum foundational problems. *Physical Science International Journal* 4, 1145-1157 (2014). URL https://journalpsij.com/index.php/PSIJ/article/view/63.
- [104] Du, D.-Z., Pardalos, P. M. & Wu, W. History of optimization (Springer US, Boston, MA, 2009). URL https://doi.org/10.1007/978-0-387-74759-0\_268.
- [105] Snyman, J. A. & Wilke, D. N. Practical Mathematical Optimization: Basic Optimization Theory and Gradient-Based Algorithms (Springer International Publishing, Cham, 2018). URL https://doi.org/10.1007/978-3-319-77586-9\_1.
- [106] Cauchy, A. et al. Méthode générale pour la résolution des systemes d'équations simultanées. Comp. Rend. Sci. Paris 25, 536–538 (1847).
- [107] Sorella, S. Generalized lanczos algorithm for variational quantum Monte Carlo. Phys. Rev. B 64, 024512 (2001). URL http://dx.doi.org/10.1103/PhysRevB. 64.024512.
- [108] Casula, M. & Sorella, S. Geminal wave functions with jastrow correlation: A first application to atoms. *The Journal of Chemical Physics* **119**, 6500–6511 (2003). URL https://doi.org/10.1063/1.1794632.
- [109] Goldberg, D. What every computer scientist should know about floating-point arithmetic. ACM Comput. Surv. 23, 5–48 (1991). URL https://doi.org/10.1145/ 103162.103163.
- [110] Kalos, M. H. & Whitlock, P. A. Monte Carlo Methods (Wiley-VCH, 2008), second edn.
- [111] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics* 21, 1087–1092 (1953). URL https://doi.org/10.1063/1.1699114.
- [112] Chen, J., Cheng, S., Xie, H., Wang, L. & Xiang, T. Equivalence of restricted boltzmann machines and tensor network states. *Phys. Rev. B* 97, 085104 (2018). URL https://link.aps.org/doi/10.1103/PhysRevB.97.085104.

- [113] Saito, H. & Kato, M. Machine learning technique to find quantum many-body ground states of bosons on a lattice. *Journal of the Physical Society of Japan* 87, 014001 (2018). URL https://doi.org/10.7566/JPSJ.87.014001.
- [114] Choo, K., Carleo, G., Regnault, N. & Neupert, T. Symmetries and many-body excitations with neural-network quantum states. *Phys. Rev. Lett.* 121, 167204 (2018). URL https://link.aps.org/doi/10.1103/PhysRevLett.121.167204.
- [115] Freitas, W. & Vitiello, S. A. Synergy between deep neural networks and the variational Monte Carlo method for small <sup>4</sup>He<sub>N</sub> clusters. Quantum 7, 1209 (2023). URL https://doi.org/10.22331/q-2023-12-18-1209.
- [116] Freitas, W., Abreu, B. & Vitiello, S. A. Modeling <sup>4</sup>He<sub>N</sub> clusters with wave functions based on neural networks. *Journal of Low Temperature Physics* (2024). URL https: //doi.org/10.1007/s10909-024-03061-w.
- [117] Razavi, A., van den Oord, A. & Vinyals, O. Generating diverse high-fidelity images with vq-vae-2 (2019). URL https://doi.org/10.48550/arXiv.1906.00446.
- [118] Barnett, R. N. & Whaley, K. B. Variational and diffusion Monte Carlo techniques for quantum clusters. *Phys. Rev. A* 47, 4082–4098 (1993). URL https://doi.org/ 10.1103/physreva.47.4082.
- [119] Lewerenz, M. Structure and energetics of small helium clusters: Quantum simulations using a recent perturbational pair potential. The Journal of Chemical Physics 106, 4596–4603 (1997). URL https://doi.org/10.1063/1.473501.
- [120] Kievsky, A., Viviani, M., Alvarez-Rodríguez, R., Gattobigio, M. & Deltuva, A. Universal behavior of few-boson systems using potential models. *Few-Body Systems* 58 (2017-01). URL https://doi.org/10.1007/s00601-017-1228-z.
- [121] de Laeter, J. R. et al. Atomic weights of the elements. review 2000 (iupac technical report). Pure and Applied Chemistry 75, 683-800 (2003). URL https://doi.org/10.1351/pac200375060683.
- [122] Freitas, W. Bosenet helium clusters (2023). URL https://github.com/ freitas-esw/bosenet-helium-clusters.
- [123] Stipanović, P., Markić, L. V. & Boronat, J. Elusive structure of helium trimers. J. Phys. B: At., Mol. Opt. Phys. 49, 185101 (2016). URL https://doi.org/10.1088% 2F0953-4075%2F49%2F18%2F185101.
- [124] Hammond, B. L., Lester Jr, W. A. & Reynolds, P. J. Monte Carlo methods in ab initio quantum chemistry (World Scientific, 1994).

- [125] Grisenti, R. E. et al. Determination of the bond length and binding energy of the helium dimer by diffraction from a transmission grating. *Phys. Rev. Lett.* 85, 2284– 2287 (2000). URL https://link.aps.org/doi/10.1103/PhysRevLett.85.2284.
- [126] Przybytek, M. et al. Relativistic and quantum electrodynamics effects in the helium pair potential. Phys. Rev. Lett. 104, 183003 (2010). URL https://link.aps.org/ doi/10.1103/PhysRevLett.104.183003.
- [127] Boninsegni, M. & Ceperley, D. M. Path integral monte carlo simulation of isotopic liquid helium mixtures. *Phys. Rev. Lett.* 74, 2288–2291 (1995). URL https://link. aps.org/doi/10.1103/PhysRevLett.74.2288.
- [128] Werner, F. & Castin, Y. General relations for quantum gases in two and three dimensions. ii. bosons and mixtures. *Phys. Rev. A* 86, 053633 (2012). URL https: //link.aps.org/doi/10.1103/PhysRevA.86.053633.
- [129] Loh, E. Y. et al. Sign problem in the numerical simulation of many-electron systems. *Phys. Rev. B* 41, 9301–9307 (1990). URL https://link.aps.org/doi/10.1103/ PhysRevB.41.9301.
- [130] Ceperley, D. M. Fermion nodes. Journal of Statistical Physics 63, 1237–1267 (1991).
   URL https://doi.org/10.1007/BF01030009.
- [131] Ceperley, D. M. Path-integral calculations of normal liquid <sup>3</sup>He. Phys. Rev. Lett. 69, 331–334 (1992). URL https://link.aps.org/doi/10.1103/PhysRevLett.69.331.
- [132] Nordhagen, E. M., Kim, J. M., Fore, B., Lovato, A. & Hjorth-Jensen, M. Efficient solutions of fermionic systems using artificial neural networks. *Frontiers in Physics* 11 (2023). URL https://www.frontiersin.org/articles/10.3389/fphy.2023. 1061580.
- [133] Choromanska, A., Henaff, M., Mathieu, M., Arous, G. B. & LeCun, Y. The loss surfaces of multilayer networks (2015). URL https://arxiv.org/abs/1412.0233. 1412.0233.
- [134] Pedersen Lohne, M., Hagen, G., Hjorth-Jensen, M., Kvaal, S. & Pederiva, F. Ab initio computation of the energies of circular quantum dots. *Phys. Rev. B* 84, 115302 (2011). URL https://link.aps.org/doi/10.1103/PhysRevB.84.115302.
- [135] Høgberget, J. Quantum Monte-Carlo studies of generalized many-body systems. Master's thesis, Faculty of Mathematics and Natural Sciences, Department of Physics, University of Oslo (2013). URL https://www.duo.uio.no/handle/10852/37167.

- [136] Dean, D. S., Le Doussal, P., Majumdar, S. N. & Schehr, G. Impurities in systems of noninteracting trapped fermions. SCIPOST PHYSICS 10 (2021). URL https: //doi.org/10.21468/SciPostPhys.10.4.082.
- [137] Häffner, H., Roos, C. & Blatt, R. Quantum computing with trapped ions. *Physics Reports* 469, 155-203 (2008). URL https://www.sciencedirect.com/science/article/pii/S0370157308003463.
- [138] Wang, H., Zhang, L., Han, J. & E, W. Deepmd-kit: A deep learning package for many-body potential energy representation and molecular dynamics. *Computer Phy*sics Communications 228, 178-184 (2018). URL https://www.sciencedirect. com/science/article/pii/S0010465518300882.
- [139] Vicentini, F. et al. NetKet 3: Machine Learning Toolbox for Many-Body Quantum Systems. SciPost Phys. Codebases 7 (2022). URL https://scipost.org/10.21468/ SciPostPhysCodeb.7.
- [140] Reis, M. A. d. & Vitiello, S. Método variacional com monte carlo aplicado ao oscilador harmônico quântico. Revista Brasileira de Ensino de Física 28, 45–50 (2006). URL https://doi.org/10.1590/S1806-11172006000100007.

## APÊNDICE A - INTEGRAÇÃO DE MONTE CARLO E FERRAMENTAS ESTATÍSTICAS

Para computar a estimativa de uma integral através da quadratura de Monte Carlo, considere a seguinte forma para integral que se quer calcular

$$G = \int_{\Omega_0} g(x)f(x)dx , \qquad (A.1)$$

onde  $f(x) \ge 0$  é uma densidade de probabilidade normalizada no espaço  $\Omega_0$  que se quer integrar, isto é,

$$\int_{\Omega_0} f(x)dx = 1 . \tag{A.2}$$

Seja a sequência de variáveis aleatórias  $\{x_1, ..., x_M\}$  amostradas de acordo com a densidade f(x). Assim sendo, a integral G pode ser estimada através de uma média aritmética [110] dada por

$$G \simeq \frac{1}{M} \sum_{i} g(x_i) . \tag{A.3}$$

Como as variáveis aleatórias são distribuídas de acordo com a densidade de probabilidade f(x), a média aritmética da função g sobre essa sequência é equivalente a uma média da função g ponderada pela probabilidade f. Um procedimento equivalente pode ser empregado para computar integrais de funções de multi variáveis.

#### A.1 Algoritmo de Metropolis

Uma maneira de gerar a sequência de variáveis aleatórias é através do algoritmo de Metropolis [111]. Esse método é um algoritmo iterativo de aceitação e rejeição, isto é, uma variável  $x_t$  é proposta a partir de uma variável aleatória inicial  $x_i$  e é aceitada ou rejeitada dependendo da função f nesses pontos. explora

Para descrever o procedimento, seja  $T(x_t|x_i)$  uma distribuição de probabilidade que da a probabilidade de transição de  $x_i$  para  $x_t$ , ou seja,  $T(x_t|x_i)$  fornece a probabilidade do sistema sair de  $x_i$  e chegar em  $x_t$ . As principais distribuições T utilizadas são a distribuição uniforme e a gaussiana. Considere também a razão  $q(x_i|x_t)$  definida como

$$q(x_i|x_t) = \frac{T(x_t|x_i)f(x_i)}{T(x_i|x_t)f(x_t)} .$$
(A.4)

Note que  $q(x_i|x_t) \ge 0$ . O algoritmo propõe que a variável aleatória  $x_t$  seja aceita ou rejeitada se baseando na probabilidade de aceitação dada por  $A(x_i|x_t) = \min[1, q(x_i|x_t)]$ .

Assim, o algoritmo de Metropolis segue a seguinte sequência de passos

- 1)  $x_t$  é amostrado da distribuição  $T(x_t|x_i)$ ;
- 2) a probabilidade de aceitação  $A(x_i|x_t)$  é computada;
- 3) se  $A(x_i|x_t)$  for maior que um número pseudo aleatório  $\xi \in (0, 1)$ , então  $x_t$  é aceito e a próxima variável aleatória da sequência será  $x_{i+1} = x_t$ . Caso contrário,  $x_{i+1} = x_i$ .

Note que a densidade de probabilidade f não precisa ser normalizada para empregar esse algoritmo, uma vez que apenas razões entre valores de f são utilizados no procedimento. Usualmente, a porcentagem de aceitação de novas amostras é em torno de 40% ou 50% para essa versão do algoritmo. Outros detalhes podem ser consultados na Ref. [110].

#### A.2 Estatística por blocos e médias móveis

Devido a natureza estocástica do procedimento, as estimativas devem ser tratadas estatisticamente de maneira adequada. Nesse caso, a análise por blocos [140] é um procedimento apropriado para isso.

Considere que  $g_i = g(x_i)$  e que a sequência de variáveis aleatórias tenha um total de M amostras. Portanto, a estimativa da integral G é dada por

$$\bar{G} = \frac{1}{M} \sum_{i=1}^{M} g_i$$
 (A.5)

Dividindo essas amostras em m blocos, a média de um bloco qualquer  $\bar{G}_i$  é escrita como

$$\bar{G}_i = \frac{1}{n} \sum_{j=(i-1)n+1}^{in} g_j , \qquad (A.6)$$

onde n = M/m é o número de amostras em cada bloco. Por simplicidade, supõe-se que o número de amostras M é divisível pelo número de blocos m. Outra forma de computar a estimativa da integral também é a média dos blocos, ou seja,

$$\bar{G} = \frac{1}{m} \sum_{i=1}^{m} \bar{G}_i$$
 (A.7)

Por outro lado, para calcular a variância dessa estimativa da integral é necessário empregar a seguinte forma

var(G) = 
$$\frac{1}{m(m-1)} \sum_{i=1}^{m} (\bar{G}_i - \bar{G})^2$$
. (A.8)

Esse calculo assume que as médias por blocos  $\bar{G}_i$  envolvem um número suficientemente grande de amostras. Dessa forma, o erro padrão da estimativa  $\bar{G}$  é dado pela raiz quadrada  $\sqrt{\operatorname{var}(G)}$ .

Outra métrica interessante para analisar dados com flutuações é a média móvel. O cálculo dessa quantidade envolve computar médias de seções do conjunto total de dados. Assim, seja uma série temporal de dados  $E_i$ , onde  $i = \{1, ..., M\}$ . Então, a média móvel para um determinado ponto k da série temporal é dada por

$$MM_{k} = \frac{1}{m} \sum_{i=k+1-m}^{k} E_{i} , \qquad (A.9)$$

sendo que, nesse caso, m é o tamanho da seção sobre a qual a média móvel é calculada. Também é possível definir um desvio padrão móvel dessa métrica, o qual é escrito como

$$DPM_{k} = \sqrt{\sum_{i=k+1-m}^{k} \frac{(E_{i} - MM_{k})^{2}}{m}}.$$
 (A.10)

## APÊNDICE B - SOLUÇÃO LIMITE PARA O POTEN-CIAL DE LENNARD-JONES

Considere a equação de Schrodinger independente do tempo para uma partícula de massa m sob o efeito do potencial da Eq. 2.1. Como o potencial é radial, o primeiro passo é transformar as variáveis cartesianas para esféricas e utilizar o método de separação de variáveis para resolver parcialmente a equação diferencial parcial. Fazendo isso, o problema se reduz a resolver a seguinte equação:

$$\left[-\frac{\hbar^2}{2mr}\frac{d^2}{dr^2}r + V(r) + \frac{\ell(\ell+1)\hbar^2}{2mr^2}\right]R_{k,\ell}(r) = E_{k,\ell}\ R_{k,\ell}(r)\ , \tag{B.1}$$

onde  $R_{k,\ell}(r)$  são as soluções da equação radial, sendo que a função de onda  $\psi(\mathbf{r})$  solução da equação de Schrodinger é dada por

$$\psi(\mathbf{r}) = \sum_{\ell} a_{\ell} R_{k,\ell}(r) P_{\ell}(\cos \theta) , \qquad (B.2)$$

onde  $P_{\ell}$  são polinômios de Legendre e  $a_{\ell}$  são coeficientes.

É interessante escrever a Eq. B.1 em termos de quantidades adimensionais, por isso, a distância é reescrita como  $r = \hat{r}\sigma$ , onde  $\sigma$  é o comprimento característico do potencial de Lennard-Jones, e a energia é reescrita como  $E = \hat{E}\epsilon$ , onde  $\epsilon$  é a escala de energia também do potencial  $V_{\rm LJ}(r)$ . Definindo a constante adimensional  $\lambda$  como

$$\lambda \epsilon = \frac{\hbar^2}{2m\sigma^2 k_B} , \qquad (B.3)$$

é possível reescrever a equação radial na seguinte forma

$$\left\{-\frac{\lambda}{\hat{r}}\left[\hat{r}\frac{d^2}{d\hat{r}^2} + 2\frac{d}{d\hat{r}}\right] + 4\left[\left(\frac{1}{\hat{r}}\right)^{12} - \left(\frac{1}{\hat{r}}\right)^6\right] + \frac{\lambda\ell(\ell+1)}{\hat{r}^2}\right\}\epsilon R_{k,\ell}(\hat{r}) = \hat{E}\epsilon R_{k,\ell}(\hat{r}) . \quad (B.4)$$

Para encontrar o comportamento a curtas distâncias da função radial, suponha que  $R(\hat{r}) \propto \exp(\alpha \hat{r}^{\beta})$ . Devido a condição R(0) = 0, o valor de  $\beta$  será negativo. Substituindo  $R(\hat{r})$  na Eq. B.4, chega-se na igualdade a seguir

$$\hat{E} = -\lambda\alpha\beta \left[\alpha\beta\hat{r}^{2\beta-2} + (\beta-1)\hat{r}^{\beta-2} + 2\hat{r}^{\beta-1}\right] + 4\left[\left(\frac{1}{\hat{r}}\right)^{12} - \left(\frac{1}{\hat{r}}\right)^{6}\right] + \frac{\lambda\ell(\ell+1)}{\hat{r}^{2}} . \quad (B.5)$$

Olhando para o regime de curtas distâncias, os dois termos dominantes dessa equação correspondem à  $\hat{r}^{2\beta-2}$  e  $\hat{r}^{-12}$ . Exigindo que esses termos se anulem, a condição necessária é dada por

$$\lambda \left(\alpha\beta\right)^2 \hat{r}^{2\beta-2} = 4\left(\frac{1}{\hat{r}}\right)^{12} . \tag{B.6}$$

Igualando as potencias da distância adimensional e os coeficientes que acompanham  $\hat{r}$ , os valores obtidos para  $\alpha$  e  $\beta$  são dados por

$$\beta = -5, \ \alpha = -\frac{2}{5}\sqrt{\frac{1}{\lambda}} \ . \tag{B.7}$$

# APÊNDICE C - ALGORITMO DE RETRO-PROPAGAÇÃO PARA REDES NEURAIS

As saídas da rede neural são dadas por:

$$a_i^{(L)} = \sigma_L \left( z_i^{(L)} \right) , \qquad (C.1)$$

onde os elementos da pré ativação são escritos como

$$z_i^{(L)} = \sum_j W_{ij}^{(L)} a_j^{(L-1)} + b_i^{(L)} .$$
 (C.2)

Primeiramente, calcula-se as derivadas com relação aos parâmetros da última camada. Para os pesos a expressão fica

$$\frac{\partial a_i^{(L)}}{\partial W_{mn}^{(L)}} = \sigma_L'\left(z_i^{(L)}\right) \sum_j a_j^{(L-1)} \delta_{im} \delta_{jn} = \sigma_L'\left(z_i^{(L)}\right) a_n^{(L-1)} \delta_{im} , \qquad (C.3)$$

enquanto para os vieses é dada por

$$\frac{\partial a_i^{(L)}}{\partial b_m^{(L)}} = \sigma_L' \left( z_i^{(L)} \right) \delta_{im} .$$
 (C.4)

Uma conta importante nesse processo é a derivada dos resultados da rede neural com relação às pré ativações de uma camada  $\ell$  qualquer, que pode ser computada por

$$\frac{\partial a_i^{(L)}}{\partial z_m^{(\ell)}} = \sum_j \frac{\partial a_i^{(L)}}{\partial z_j^{(\ell+1)}} \frac{\partial z_j^{(\ell+1)}}{\partial z_m^{(\ell)}} \tag{C.5a}$$

$$=\sum_{j}\sum_{k}\frac{\partial a_{i}^{(L)}}{\partial z_{j}^{(\ell+1)}}\frac{\partial z_{j}^{(\ell+1)}}{\partial a_{k}^{(\ell)}}\frac{\partial a_{k}^{(\ell)}}{\partial z_{m}^{(\ell)}} \tag{C.5b}$$

$$=\sum_{j}\sum_{k}\frac{\partial a_{i}^{(L)}}{\partial z_{j}^{(\ell+1)}}\frac{\partial}{\partial a_{k}^{(\ell)}}\left[\sum_{n}W_{jn}^{(\ell+1)}a_{n}^{(\ell)}+b_{j}^{(\ell+1)}\right]\frac{\partial\sigma_{\ell}\left(z_{k}^{(\ell)}\right)}{\partial z_{m}^{(\ell)}}\tag{C.5c}$$

$$=\sum_{j}\sum_{k}\sum_{n}\frac{\partial a_{i}^{(L)}}{\partial z_{j}^{(\ell+1)}}W_{jn}^{(\ell+1)}\sigma_{\ell}'\left(z_{k}^{(\ell)}\right)\delta_{nk}\delta_{km}$$
(C.5d)

$$=\sum_{j} \frac{\partial a_i^{(L)}}{\partial z_j^{(\ell+1)}} W_{jm}^{(\ell+1)} \sigma_\ell' \left( z_m^{(\ell)} \right) . \tag{C.5e}$$

Note que a regra da cadeia foi feita utilizando as pré ativações da camada seguinte da

camada  $\ell < L$  para que uma relação de recorrência surja. Assim, para o cálculo das derivadas com relação aos vieses de uma determinada camada  $\ell$ , segue que

$$\frac{\partial a_i^{(L)}}{\partial b_m^{(\ell)}} = \sum_j \sum_k \frac{\partial a_i^{(L)}}{\partial z_j^{(\ell+1)}} \frac{\partial z_j^{(\ell+1)}}{\partial a_k^{(\ell)}} \frac{\partial a_k^{(\ell)}}{\partial b_m^{(\ell)}} \frac{\partial a_k^{(\ell)}}{\partial b_m^{(\ell)}} \tag{C.6a}$$

$$=\sum_{j}\sum_{k}\sum_{n}\frac{\partial a_{i}^{(L)}}{\partial z_{j}^{(\ell+1)}}W_{jn}^{(\ell+1)}\sigma_{\ell}'\left(z_{k}^{(\ell)}\right)\frac{\partial b_{k}^{(\ell)}}{\partial b_{m}^{(\ell)}}\,\delta_{nk}\tag{C.6b}$$

$$=\sum_{j} \frac{\partial a_i^{(L)}}{\partial z_j^{(\ell+1)}} W_{jm}^{(\ell+1)} \sigma_\ell' \left( z_m^{(\ell)} \right) = \frac{\partial a_i^{(L)}}{\partial z_m^{(\ell)}} .$$
(C.6c)

Por outro lado, para os pesos, as derivadas assumem a forma

$$\frac{\partial a_i^{(L)}}{\partial W_{mn}^{(\ell)}} = \sum_j \sum_k \frac{\partial a_i^{(L)}}{\partial z_j^{(\ell+1)}} \frac{\partial z_j^{(\ell+1)}}{\partial a_k^{(\ell)}} \frac{\partial a_k^{(\ell)}}{\partial W_{mn}^{(\ell)}} \tag{C.7a}$$

$$=\sum_{j}\sum_{k}\sum_{p}\frac{\partial a_{i}^{(L)}}{\partial z_{j}^{(\ell+1)}}W_{jp}^{(\ell+1)}\frac{\partial a_{k}^{(\ell)}}{\partial W_{mn}^{(\ell)}}\delta_{pk}$$
(C.7b)

$$=\sum_{j}\sum_{k}\frac{\partial a_{i}^{(L)}}{\partial z_{j}^{(\ell+1)}}W_{jk}^{(\ell+1)}\sigma_{\ell}^{\prime}\left(z_{k}^{(\ell)}\right)\frac{\partial}{\partial W_{mn}^{(\ell)}}\left[\sum_{p}W_{kp}^{(\ell)}a_{p}^{(\ell-1)}+b_{k}^{(\ell)}\right]$$
(C.7c)

$$=\sum_{j}\sum_{k}\sum_{p}\frac{\partial a_{i}^{(L)}}{\partial z_{j}^{(\ell+1)}}W_{jk}^{(\ell+1)}\sigma_{\ell}'\left(z_{k}^{(\ell)}\right)a_{p}^{(\ell-1)}\,\delta_{mk}\delta_{np}\tag{C.7d}$$

$$= \sum_{j} \frac{\partial a_{i}^{(L)}}{\partial z_{j}^{(\ell+1)}} W_{jm}^{(\ell+1)} \sigma_{\ell}' \left( z_{m}^{(\ell)} \right) a_{n}^{(\ell-1)} = \frac{\partial a_{i}^{(L)}}{\partial z_{m}^{(\ell)}} a_{n}^{(\ell-1)}$$
(C.7e)

Empregando estes cálculos para computar o gradiente com relação aos pesos e vieses da Eq. 4.9, as derivadas com relação aos parâmetros da rede neural de uma camada  $\ell < L$  podem ser escritos resumidamente como

$$\frac{\partial \mathbf{a}^{(L)}}{\partial \mathbf{W}^{(\ell)}} = \frac{\partial \mathbf{a}^{(L)}}{\partial \mathbf{z}^{(\ell)}} \otimes \mathbf{a}^{(\ell-1)} , \qquad (C.8)$$

$$\frac{\partial \mathbf{a}^{(L)}}{\partial \mathbf{b}^{(\ell)}} = \frac{\partial \mathbf{a}^{(L)}}{\partial \mathbf{z}^{(\ell)}} , \qquad (C.9)$$

onde a matriz Jacobiana das ativações  $\mathbf{a}^{(L)}$  com relação às pré ativações  $\mathbf{z}^{(\ell)}$  é dada por

$$\frac{\partial \mathbf{a}^{(L)}}{\partial \mathbf{z}^{(\ell)}} = \left[\frac{\partial \mathbf{a}^{(L)}}{\partial \mathbf{z}^{(\ell+1)}} \mathbf{W}^{(\ell+1)}\right] \odot \sigma'_{\ell} \left(\mathbf{z}^{(\ell)}\right) . \tag{C.10}$$

A operação  $\odot$  denota multiplicação elemento a elemento ou produto de Hadamard, en-

quanto a operação  $\otimes$  indica produto tensorial. Para o caso em que  $\ell = L$ , a expressão da matriz se resume a

$$\frac{\partial \mathbf{a}^{(L)}}{\partial \mathbf{z}^{(L)}} = \mathbb{1} \odot \sigma'_L \left( \mathbf{z}^{(L)} \right) . \tag{C.11}$$

Com essa quantidade e utilizando as relações das equações C.8, C.9 e C.10, é possível computar todas as derivadas com relação aos parâmetros do modelo. Note que a informação flui do fim para o começo da rede neural nesse caso e, por isso, esse procedimento é chamado de algoritmo de retro propagação [21].

### C.1 A matriz de covariâncias no contexto de *machine learning*

Uma importante conexão entre abordagens de aprendizado de máquina e simulações de Monte Carlo quântico é a chamada matriz de informação de Fisher. Essencialmente, a matriz de Fisher é equivalente à matriz de covariâncias da Eq. 5.30, conectando o método do gradiente natural descendente com o método de reconfiguração estocástica [31].

Para enxergar essa relação, considere a densidade de probabilidade não normalizada  $\tilde{p}(R) = |\psi_{\theta}(R)|^2$ . Então, sua normalização é definida como

$$Z(\boldsymbol{\theta}) = \int dR \ |\psi_{\boldsymbol{\theta}}(R)|^2 , \qquad (C.12)$$

onde  $\theta$  são os parâmetros com relação aos quais se deseja computar a matriz de Fisher. Dessa maneira, a densidade de probabilidade normalizada pode ser escrita como

$$p(R) = \frac{|\psi_{\theta}(R)|^2}{Z(\theta)} = \frac{\tilde{p}(R)}{Z(\theta)} .$$
(C.13)

Por construção, a derivada logarít<br/>mica de  $\psi_{\theta}$  e  $\tilde{p}$  com relação à um parâmetro  $\theta_i$  são con<br/>ectadas através da igualdade

$$O^{i}(R) = \frac{1}{\psi(R)} \frac{\partial \psi_{\theta}(R)}{\partial \theta_{i}} = \frac{\partial \ln |\psi_{\theta}(R)|}{\partial \theta_{i}} = \frac{1}{2} \frac{\partial \ln \tilde{p}(R)}{\partial \theta_{i}} .$$
(C.14)

Para encontrar a conexão entre as matrizes de covariâncias e de Fisher, considere o valor esperado das derivadas logarítmicas de  $\psi_{\theta}$  dado por

$$\mathbb{E}_{p(R)}[O^{i}(R)] = \frac{1}{2} \mathbb{E}_{p(R)} \left[ \frac{\partial \ln \tilde{p}(R)}{\partial \theta_{i}} \right]$$
(C.15a)

$$=\frac{1}{2}\int dR \ p(R)\frac{\partial \ln \tilde{p}(R)}{\partial \theta_i}$$
(C.15b)

$$= \frac{1}{2} \int dR \; \frac{\tilde{p}(R)}{Z(\boldsymbol{\theta})} \frac{\partial \ln \tilde{p}(R)}{\partial \theta_i} \tag{C.15c}$$

$$= \frac{1}{2} \int dR \; \frac{1}{\tilde{p}(R)} \frac{\tilde{p}(R)}{Z(\boldsymbol{\theta})} \frac{\partial \tilde{p}(R)}{\partial \theta_i} \tag{C.15d}$$

$$=\frac{1}{2Z(\boldsymbol{\theta})}\int dR \;\frac{\partial \tilde{p}(R)}{\partial \theta_i} \tag{C.15e}$$

$$= \frac{1}{2Z(\boldsymbol{\theta})} \frac{\partial}{\partial \theta_i} \int dR \; \tilde{p}(R) \tag{C.15f}$$

$$=\frac{1}{2Z(\boldsymbol{\theta})}\frac{\partial Z(\boldsymbol{\theta})}{\partial \theta_i} \tag{C.15g}$$

$$= \frac{1}{2} \frac{\partial \ln Z(\boldsymbol{\theta})}{\partial \theta_i} . \tag{C.15h}$$

Portanto, a derivada logarítmica do fator de normalização  $Z(\boldsymbol{\theta})$  é igual ao valor esperado de  $O^i(R)$ . Além disso, essa quantidade está conectada com o cálculo da derivada logarítmica da densidade de probabilidade p(R), que pode ser escrita como

$$\frac{\partial \ln p(R)}{\partial \theta_i} = \frac{Z(\boldsymbol{\theta})}{\tilde{p}(R)} \left( \frac{1}{Z(\boldsymbol{\theta})} \frac{\partial \tilde{p}(R)}{\partial \theta_i} - \frac{\tilde{p}(R)}{[Z(\boldsymbol{\theta})]^2} \frac{\partial Z(\boldsymbol{\theta})}{\partial \theta_i} \right) = \frac{\partial \ln \tilde{p}(R)}{\partial \theta_i} - \frac{\partial \ln Z(\boldsymbol{\theta})}{\partial \theta_i} \ .$$
(C.16)

De posse dessas relações, os elementos da matriz de covariâncias podem ser escritos em termos dos valores esperados de  $O^i$  como

$$\mathcal{M}_{ij} = \mathbb{E}_{p(R)} \left[ \left( O^i - \mathbb{E}_{p(R)}[O^i] \right) \left( O^j - \mathbb{E}_{p(R)}[O^j] \right) \right]$$
(C.17a)

$$= \frac{1}{4} \mathbb{E}_{p(R)} \left[ \left( \frac{\partial \ln \tilde{p}(R)}{\partial \theta_i} - \frac{\partial \ln Z(\boldsymbol{\theta})}{\partial \theta_i} \right) \left( \frac{\partial \ln \tilde{p}(R)}{\partial \theta_j} - \frac{\partial \ln Z(\boldsymbol{\theta})}{\partial \theta_j} \right) \right]$$
(C.17b)

$$= \frac{1}{4} \mathbb{E}_{p(R)} \left[ \frac{\partial \ln p(R)}{\partial \theta_i} \frac{\partial \ln p(R)}{\partial \theta_j} \right] = \frac{\mathcal{F}_{ij}}{4} .$$
(C.17c)

Assim, pode-se identificar a matriz de covariâncias com a matriz de Fisher  $\mathcal{F}$  da Eq. 4.19.

Tabela D.1: Em unidades de Kelvin, são apresentadas a energia cinética  $\langle \mathcal{T} \rangle$  e a energia total  $\langle E \rangle$  obtidas com a função tentativa  $\psi_{\text{BHC}}$  para aglomerados de <sup>4</sup>He contendo N átomos.

N	$\langle \mathcal{T}  angle$	$\langle E \rangle$
11	37.72(1)	-10.665(2)
12	43.50(1)	-12.801(2)
13	48.79(1)	-15.069(2)
14	55.92(1)	-17.444(2)
15	61.96(1)	-19.806(1)
16	67.76(1)	-22.401(2)
17	74.82(1)	-25.039(2)
18	82.60(1)	-27.755(2)
19	90.20(1)	-30.544(3)
20	99.17(1)	-33.381(2)
21	103.16(1)	-36.426(2)
22	108.26(1)	-38.949(3)
23	118.91(2)	-41.639(3)
24	118.16(2)	-44.615(3)

Tabela D.2: Coeficientes ajustados ao modelo de gota considerando estimativas de energia através do método DMC e através das estimativas variacionais obtidas pela representação da função de onda tentativa  $\psi_{\rm BHC}$ .

	$C_v$	$C_s$	$C_c$
DMC	-7.842	23.126	- 17.423
Bosenet	-5.887	13.767	-6.242

Tabela D.3: Estimativas dos valores de contato por partícula  $C_2^{(N)}/N$  para vários tamanhos de aglomerados de <sup>4</sup>He<sub>N</sub>.

n	3	4	5	6	8	10	11	12	13	14
-	4.96	10.19	15.23	19.67	27.88	34.83	37.86	39.94	42.13	44.73



Figura D.1: Tempo de execução  $t_N$  e seu inverso em função do número de GPUs empregadas no cálculo, ambos normalizados pelo tempo de execução ao usar apenas uma GPU. As simulações foram realizadas em aglomerados de hélio com N = 5 partículas.



Figura D.2: Tempo de execução (azul) e porcentagem de utilização das GPUs (vermelho) em função do número de GPUs empregadas no cálculo. As simulações foram realizadas em aglomerados de hélio com N = 5 partículas.



Figura D.3: Otimização da energia para o caso N = 3 do gás de elétrons não interagente aprisionado em um potencial harmônico tridimensional. Nesse caso, a energia do estado fundamental é 5.5 em unidades adimensionais. Das cores claras paras as escuras aumenta-se o tamanho da rede neural considerada. A linha tracejada é o valor da energia do estado fundamental.