

UFRRJ

**INSTITUTO DE CIÊNCIAS EXATAS
CURSO DE PÓS-GRADUAÇÃO EM MODELAGEM
MATEMÁTICA E COMPUTACIONAL**

DISSERTAÇÃO

**Desenvolvimento de uma Nova Proposta Matemática e
Computacional para a Metodologia da Matriz Densidade
Tight-Binding Ordem-N**

Julius Monteiro de Barros Filho

2017



**INSTITUTO DE CIÊNCIAS EXATAS
CURSO DE PÓS-GRADUAÇÃO EM MODELAGEM
MATEMÁTICA E COMPUTACIONAL**

**DESENVOLVIMENTO DE UMA NOVA PROPOSTA
MATEMÁTICA E COMPUTACIONAL PARA A
METODOLOGIA DA MATRIZ DENSIDADE TIGHT-BINDING
ORDEM-N**

JULIUS MONTEIRO DE BARROS FILHO

Sob a Orientação do Professor
Moisés Augusto da Silva Monteiro de Araújo

Dissertação submetida como requisito parcial para obtenção do grau de **Mestre em Ciências**, no Curso de Pós-Graduação em Modelagem Matemática e Computacional, Área de Concentração em Modelagem Matemática e Computacional

Seropédica, RJ
Março de 2017

Universidade Federal Rural do Rio de Janeiro
Biblioteca Central / Seção de Processamento Técnico

Ficha catalográfica elaborada
com os dados fornecidos pelo(a) autor(a)

M
478d Monteiro de Barros Filho, Julius , 1982-
Desenvolvimento de uma nova proposta matemática e
computacional para a metodologia da matriz densidade
tight-binding ordem-N. / Julius Monteiro de Barros
Filho. - 2017.
76 f.

Orientador: Moisés Augusto da Silva Monteiro de
Araújo. Dissertação (Mestrado). -- Universidade
Federal Rural do Rio de Janeiro, Programa de Pós
Graduação em Modelagem Matemática e Computacional,
2017.

1. DMTB. 2. Hamiltoniano Tight-Binding. 3. Matriz
Densidade. 4. Modelagem Matemática e Computacional.
5. Estrutura Eletrônica. I. Augusto da Silva Monteiro
de Araújo, Moisés , 1974-, orient. II Universidade
Federal Rural do Rio de Janeiro. Programa de Pós
Graduação em Modelagem Matemática e Computacional III.
Título.

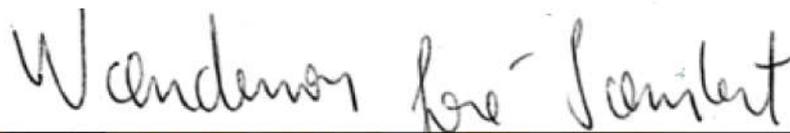
UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO DE CIÊNCIAS EXATAS
CURSO DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E
COMPUTACIONAL

JÚLIUS MONTEIRO DE BARROS FILHO

Dissertação submetida como requisito parcial para obtenção do grau de Mestre em Ciências, no Curso de Pós-Graduação em Modelagem Matemática e Computacional, área de Concentração em Modelagem Matemática e Computacional.

DISSERTAÇÃO APROVADA EM 31/03/2017


Moisés Augusto da Silva Monteiro de Araújo (Dr.) UFRRJ
(Orientador)


Wanderson José Lambert (Dr.) UFRRJ


Laércio Costa Ribeiro (Dr.) CEFET-RJ

Ao meu amado filho João Danilo e à memória de minha saudosa avó Mayte.

“Nunca se deve engatinhar quando o impulso é voar.”

Helen Keller.

Resumo

Barros Filho, Julius Monteiro. **Desenvolvimento de uma nova proposta matemática e computacional para a metodologia da matriz densidade tight-binding ordem-N**. 2017. 78p Dissertação (Mestrado em Modelagem Matemática e Computacional). Instituto de Ciências Exatas, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2017.

Neste trabalho é desenvolvido, em um contexto abrangente, uma fundamentação matemático computacional para a metodologia de cálculo de estrutura eletrônica conhecido como Método da Matriz Densidade Eletrônica Tight Binding - DMTB. A abordagem adotada deixa claro que a formulação matemática final deste método é completamente dependente das estratégias computacionais que se escolhem para a sua implementação. Desta forma, o DMTB fica estabelecido como um modelo matemático-computacional de formulação final variável. Nossa formulação usa uma estratégia computacional baseada em algoritmos de Gradientes Conjugados não linear (NLCG). O modelo final obtido é ligeiramente diferente do DMTB que foi apresentado originalmente na literatura pelo grupo de Vanderbilt, estando em acordo com a versão apresentada por Millam e Scuseria, em 1997. Nosso desenvolvimento exigiu considerar temas não standard, tanto da Matemática, quanto da Computação Numérica, tais como funções com domínios em espaços matriciais e o desenvolvimento de técnicas de cálculo diferencial nestes espaços, e modificações na estrutura do NLCG. A abordagem utilizada desenvolve os aspectos matemáticos visando a efetiva implementação computacional da metodologia. Com isso, tendo em vista que os avanços tecnológicos e os conhecimentos científicos devem estar intimamente relacionados, a presente dissertação pretende preencher uma lacuna na literatura especializada, uma vez que não existem textos com objetivos pedagógicos ou que se associem de forma natural com a modelagem matemática e computacional do DMTB.

Palavras-chave: DMTB, Matriz Densidade, Hamiltoniano Tight-Binding.

Abstract

Barros Filho, Julius Monteiro. **Development of a new mathematical and computational proposal for the density matrix tight-binding order-N methodology.** 2017. 78p Dissertation (Master in Mathematical and Computational Modeling). Instituto de Ciências Exatas, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2017.

In this work, a mathematical-computational basis for the electronic structure calculation methodology known as Electronic Density Matrix Method Tight Binding (DMTB) is developed in a comprehensive context. The adopted approach evidences that the final mathematical formulation of this method is completely dependent on the computational strategies chosen for its implementation. In this way, the DMTB is established as a mathematical-computational model of variable final formulation. Our formulation uses a computational strategy based on Nonlinear Conjugate Gradients (NLCG) algorithms. The final model obtained is slightly different from the DMTB that was originally presented in the literature by Vanderbilt group. It is in accordance with the version presented by Millam and Scuseria, in 1997. Our development considered nonstandard themes of both Mathematics and Computation such as functions with domains in matrix spaces and the development of differential calculus techniques in these spaces and modifications in the structure of NLCG. The approach which was used develops the mathematical aspects aiming at effective computational implementation of the methodology. Thus, given that technological advances and scientific knowledge must be closely related, the present dissertation intends to fulfill a gap in the specialized literature, since there are no texts with pedagogical objectives or that are associated, in a natural way, with mathematical and computational modeling aspects of the DMTB.

Key-words: DMTB, Density Matrix, Tight-Binding Hamiltonian.

Lista de Figuras

2.1	os três pares de vetores, $\{\vec{a}_1, \vec{a}_2\}$; $\{\vec{a}_1, \vec{a}'_2\}$ e $\{\vec{a}_2, \vec{a}'_1\}$ produzem todos os pontos da rede.	27
2.2	quando somamos vetores formados por arestas consecutivas de um desses hexágonos, geramos um ponto no interior do hexágono e, portanto, não pertencente a rede.	27
2.3	Rede cúbica simples - Imagem retirada de [Ashcroft and Mermin, 1976].	28
2.4	Rede cúbica de face centrada - Imagem retirada de [Ashcroft and Mermin, 1976].	29
2.5	Rede cúbica de corpo centrado - Imagem retirada de [Ashcroft and Mermin, 1976].	29
2.6	Rede triangular como rede de Bravais subjacente e base formada pelo átomo 1 e pelo átomo 2 - Imagem retirada de [Ashcroft and Mermin, 1976].	30
2.7	uma folha de grafite.	30
2.8	estrutura cristalina do diamante.	31
3.1	o polinômio purificador $f(x) = 3x^2 - 2x^3$	48
3.2	Cálculo do Parâmetro de Rede Experimental.	57

Lista de Símbolos

Alfabeto latino

\hat{H}	Operador hamiltoniano
$\hat{T}_{\vec{r}}$	Operador translação
H	Matriz hamiltoniana (Tight-Binding)
K_N	Operador energia cinética nuclear
K_e	Operador energia cinética eletrônica
V_{Ne}	Potencial coulombiano atrativo núcleo-elétron
V_{ee}	Potencial coulombiano repulsivo elétron-elétron
\vec{k}	vetor de onda de Bloch
\vec{d}_k	direção de busca do passo k , no NLCG

Alfabeto grego

- ψ Função de onda de Schrödinger
- φ Orbital atômico
- ρ Matriz densidade eletrônica
- Ω Funcional de energia eletrônica
- ϵ Erro máximo dos autovalores de uma matriz quase idempotente
- μ Multiplicador de Lagrange
- α comprimento do passo na busca linear do NLCG

Sumário

Introdução	1
1 A Equação de Schrödinger e a modelagem dos fenômenos quânticos	4
1.1 A Modelagem Matemática e Computacional	4
1.2 A equação de Schrödinger e os postulados de suporte da Teoria Quântica.	6
1.2.1 A função de onda de Schrödinger	7
1.2.2 A Equação de Schrödinger	8
1.2.3 Os Operadores Quânticos	12
1.2.4 Operadores Hermitianos	13
1.2.5 Medições de grandezas observáveis, medições simultâneas e o Princípio da Incerteza de Heisenberg	14
1.2.6 Potencial Periódico e o Teorema de Bloch	19
1.2.7 Generalização para sistemas de muitas partículas quântica e os principais métodos de solução da Equação de Schrödinger	22
2 Metodologia em sólidos cristalinos	26
2.1 Redes de Bravais e estrutura cristalina	26
2.2 O método Tight-Binding	31
2.2.1 A proposta LCAO	32
2.2.2 O Hamiltoniano Tight-Binding	33
3 O Método DMTB	37
3.1 A Matriz Densidade	38

3.2	Espaço de matrizes versus Espaços \mathbb{R}^n	40
3.3	O modelo de minimização	44
3.3.1	O vínculo do número de elétrons	45
3.3.2	O vínculo da idempotência	46
3.4	O modelo matemático final	48
3.4.1	O modelo matemático-computacional final	50
3.5	Conclusões e Perspectivas	56
3.6	Pesquisas Futuras	58

Referências Bibliográficas	60
-----------------------------------	-----------

Introdução

A Mecânica Quântica é uma das teorias mais bem sucedidas da Física. Além disso, as possibilidades de aplicação desta teoria aumentam a cada ano. Isto talvez justifique o empenho de milhares de pesquisadores ao redor do mundo sobre este tema. Apenas para citar algum exemplo, temos na *Computação Quântica* um recente interesse em interfaces silício/barreira. Recentemente ligas de GeSi têm sido utilizadas como tais barreiras. No entanto, resultados com previsões teóricas das propriedades de elétrons ligados a doadores ou a pontos quânticos próximos a interfaces Si/GeSi são escassos. Uma outra aplicação está nas chamadas *Discordâncias Cristalinas*. Estas formam a base dos processos de deformação plástica dos materiais e a teoria quântica é usada na investigação das propriedades eletrônicas das discordâncias cristalinas ([Araújo, 2006], [Oliveira, 2014]).

No entanto, sistemas quanto-mecânicos polieletrônicos possuem uma estrutura matemática bastante intrincada e as equações quanto-mecânicas só podem ser resolvidas com o uso de hipóteses simplificadoras, tais como a aproximação de Bohr-Oppenheimer e as aproximações de elétron independente. Além disso, muitas propostas de resolução de modelos matemáticos que possuem descrição quântica em suas formulações, esbarram na alta complexidade computacional dos algoritmos envolvidos. Isto limita a aplicação destes modelos a sistemas com muitos átomos, impedindo que se façam simulações para investigações mais realistas. Atualmente as investigações teóricas de propriedades eletrônicas são feitas através de simulações computacionais de modelos matemáticos baseados nas metodologias básicas da teoria quântica da matéria condensada: as metodologias *semi-empíricas* e *de primeiros princípios*, como definiremos no próximo capítulo. Várias estratégias matemático-computacionais com complexidade de ordem $O(N)$ têm sido propostas, onde N é o número de átomos envolvidos ([Stefan Goedecker, 1999], [Bollinger, 2008]). Uma delas é a metodologia semi-empírica chamada de DMTB (Método da Matriz Densidade Tight-Binding). Este método foi desenvolvido por Li, Nunes e Vanderbilt em [Li et al., 1993].

O objetivo desta dissertação é desenvolver detalhadamente uma nova proposta para a construção da metodologia DMTB em um contexto abrangente. O nível de detalhamento é aquele necessário para compreender as estruturas matemáticas envolvidas e estabelecer os resultados matemáticos e computacionais propostos. O modelo que apresentaremos se mostra ligeiramente mais geral do que o proposto originalmente por Li, Nunes e Vanderbilt, em 1993 ([Li et al., 1993]), estando em acordo com a proposta de Scuseria em [Millam and Scuseria, 1997]. Este trabalho foi desenvolvido em um grupo de pesquisa. No nosso grupo, trabalho com a integrante Fernanda Lúcia Sá Ferreira e a minha pesquisa foi a base para a implementação computacional, baseada no paradigma da Orientação a Objetos, do método DMTB para o silício, tema de trabalho da integrante Fernanda. Esta implementação foi feita em [Ferreira, 2017].

Embora a Mecânica Clássica (Newtoniana) seja um assunto bastante conhecido pela comunidade científica em geral, a Mecânica Quântica não faz parte da formação básica da grande maioria das graduações científicas do nosso país. O mesmo ocorre com a chamada Modelagem Matemática e Computacional. Por isso, dedicamos um capítulo para estes temas. No capítulo em questão, devido aos nossos objetivos finais, nossa abordagem privilegia os aspectos matemáticos, em detrimento da descrição fenomenológica. No entanto, a descrição fenomenológica é fundamental para uma aceitação racional das premissas e dos argumentos que permeiam o desenvolvimento matemático apresentado. Experimentos famosos, como o da dupla fenda, se quer são comentados. Porém, ressaltamos que o processo de interpretação destes experimentos passou pela intervenção de cientistas eminentes do século XX e, até hoje, ainda não se finalizou. Para estas questões, recomendamos o excelente [Bastos Filho and Siqueira, 1993].

Esta dissertação está estruturada em 4 capítulos, que passamos a descrever brevemente:

No primeiro capítulo descrevemos, de forma bastante breve, o Método DMTB, a sua importância e os objetivos da dissertação.

No segundo capítulo apresentamos uma breve revisão dos conceitos e resultados dos fundamentos da Mecânica Quântica não relativística, que fazemos uso neste trabalho. Ainda neste capítulo, apresentamos, de forma qualitativa, uma classificação dos métodos computacionais de solução da equação de Schrödinger que têm se tornado bastante comum nos artigos que tratam do assunto.

O terceiro capítulo trata das peculiaridades das metodologias do primeiro capítulo, quando aplicadas aos sólidos cristalinos. O método Thigt-Binding é apresentado.

Estes capítulos têm duplo objetivo. O primeiro é tornar o texto o mais auto contido possível. O se-

gundo é deixar claro como as hipóteses adotadas ao longo do desenvolvimento vão sendo usadas para a construção dos modelos. Para atingir este último, foi necessário demonstrar, de maneira rigorosa, alguns resultados já bastante conhecidos destas teorias. Isto tornou o capítulo mais extenso do que sugere uma "revisão".

No quarto capítulo, objetivo do nosso trabalho, apresentamos a construção do método DMTB. O modelo obtido é ligeiramente diferente do apresentado por Li, Nunes e Vanderbilt em [Li et al., 1993], estando em acordo com a versão presente em [Millam and Scuseria, 1997]. O método DMTB faz uso do formalismo da Matriz Densidade Eletrônica, que não abordamos neste trabalho. No entanto, enunciamos os fatos básicos necessários para a compreensão da construção apresentada e indicamos uma bibliografia específica que trata exatamente do nosso caso de interesse. Este capítulo segue o mesmo padrão de rigor teórico dos capítulos anteriores. Introduzimos um processo de vetorização de matrizes e provamos alguns resultados ligados à derivação de funções reais de variável matricial, necessários para o desenvolvimento deste capítulo. Além disso, tratamos brevemente dos detalhes matemáticos fundamentais necessários para a implementação computacional do DMTB baseada em um algoritmo de Gradiente Conjugado não Linear (NLCG). Por fim, discutimos alguns possíveis caminhos a serem seguidos a partir deste trabalho. Finalmente, gostaríamos de mencionar que adotamos um enfoque que, em grande parte, não pode ser encontrado na literatura e que permite a obtenção de resultados profundos de maneira simples e com poucos pré-requisitos, tornando assim a exposição acessível a um grande número de leitores.

Capítulo 1

A Equação de Schrödinger e a modelagem dos fenômenos quânticos

Neste texto apresentaremos uma breve revisão dos conceitos e resultados da Mecânica Quântica (não relativista). Aqui, a abordagem é feita tendo como alvo o conceito de Modelagem Matemática. Para teorias mais completas do que a que está exposta aqui, ver, por exemplo, [Pauling and Junior, 1937], [Ballentine, 2000] e [Levine, 2006]. Esta apresentação pretende ser o mais autocontida possível. Além disso, o estilo adotado pretende deixar claro como as hipóteses adotadas ao longo do desenvolvimento vão sendo usadas para a construção do modelo. Para atingir este último objetivo, foi necessário demonstrar, de maneira rigorosa, alguns resultados já bastante conhecidos desta teoria. Como referências para os fatos básicos da álgebra linear, da análise vetorial e da teoria das probabilidades que fazemos uso neste trabalho, indicamos [Lang, 1966], [Spivak, 1965] e o capítulo 4 de [Barata, 2011], e [Dantas, 2004], respectivamente. Indicamos ainda a leitura de alguns dos muitos textos introdutórios de mecânica quântica que tratam de experimentos realizados no final do século XIX, como o famoso experimento da fenda dupla, que indicaram que a mecânica clássica não era adequada para tratar da dinâmica das partículas quânticas.

1.1 A Modelagem Matemática e Computacional

Diferentemente do que ocorre em áreas como Álgebra e Mecânica Quântica, não existe um conjunto de regras sobre como criar um modelo. Atualmente o processo de modelagem é baseado em uma coleção de princípios gerais que ao longo dos anos têm se mostrado eficientes na aplicação do conhecimento matemático para analisar fenômenos de várias áreas do conhecimento. Para mais detalhes, indicamos

[Karam and Almeida, 2003]. Modelagem é o processo de construção de modelos para representar algum processo de interesse. O tipo de linguagem usada nos modelos define o tipo de modelagem. Assim, a modelagem matemática é caracterizada pela construção de equações ou estruturas matemáticas, cujos comportamentos ou propriedades correspondam ao processo alvo. Um passo seguinte, geralmente natural e importante, que é dado atualmente nos modelos matemáticos, é o processo da construção de algoritmos computacionais que implementem e simulem os comportamentos e propriedades das equações e estruturas dos modelos. Geramos assim um modelo chamado de *matemático-computacional*.

Ao construirmos um modelo, devemos fazer uso de todas as informações disponíveis sobre o processo em questão. No caso específico de propostas de modelos matemáticos para fenômenos físicos, esta base de informações usada para a fundamentação do modelo muitas vezes é oriunda da interpretação dos fatos experimentais conhecidos acerca do fenômeno que está sob análise. Formular estas interpretações em uma linguagem matemática coerente e consistente pode se tornar, a depender do fenômeno, uma tarefa altamente não trivial. Mas voltando ao caso geral, neste conjunto de informações, o ideal seria usarmos um subconjunto (preferencialmente pequeno) como base para a fundamentação do modelo e, então, poderíamos testá-lo (validá-lo) tentando explicar os outros fatos conhecidos que estão fora do subconjunto fundamental que usamos, ou seja, os fatos conhecidos acerca do processo que não foram utilizados na construção do modelo. Uma vez vencida esta etapa de, digamos, coerência, onde o modelo demonstrou ser capaz de nos fornecer entendimentos sobre os fenômenos que já eram observados do sistema modelado, chega a hora de se perguntar sobre a qualidade do modelo construído. A grandiosidade de um modelo não está na sua validação. Isto é uma obrigação para a aceitação inicial de qualquer modelagem proposta. A grande qualidade de um modelo está no seu poder de previsão. Um bom modelo é capaz de prever propriedades do processo por ele modelado que não eram conhecidas antes da sua construção. Evidentemente é preciso se aguardar um tempo para poder apreciar esta qualidade. É o caso do modelo para o Eletromagnetismo dado pelas Equações Integrais de Maxwell e as ondas de radio, por exemplo, e com a equação de Einstein do Efeito Foto Elétrico. A modelagem usada na Mecânica Quântica, que tem a Equação de Schrödinger como elemento principal, tem mostrado sua grandiosidade através do último século.

Não é nosso objetivo apresentar a construção histórica do que hoje chamamos de Mecânica Quântica. Esta opção seria muito árdua pois, assim como na construção de muitas teorias, o caminho histórico que

levou à Mecânica Quântica é tortuoso, muitas vezes contraditório e confuso (veja o livro "A Parte e o Todo", de Werner Heisenberg [Heisenberg, 1996]). Mesmo os textos especializados no assunto, destacam apenas os fatos mais relevantes. Nosso objetivo neste capítulo é ressaltar os aspectos da Modelagem Matemática dos fenômenos quânticos, apresentando o modelo de forma lógica, em acordo com a sua aceitação atual, para que sirva de guia introdutório para iniciantes na área de Modelagem Matemática e Computacional que precisem da teoria quântica em suas pesquisas.

1.2 A equação de Schrödinger e os postulados de suporte da Teoria Quântica.

O último século gerou um acúmulo de evidências experimentais que apoiam a ideia de que os fenômenos atomísticos são governados por uma Equação Diferencial Parcial (EDP) conhecida por *equação de Schrödinger*, juntamente com alguns postulados, decorrentes da interpretação de fatos experimentais¹. Antes da apresentação formal da Equação de Schrödinger, falaremos brevemente da função que é objeto desta equação, chamada de *função de onda* (de Schrödinger). Vale comentar que coube ao físico alemão Max Born a interpretação probabilística da função de onda. Hoje esta interpretação é amplamente aceita, porém alguns cientistas importantes do século passado, como Einstein e o próprio Schrödinger, jamais a aceitaram.

Chamaremos de *partículas quânticas* às partículas elementares (elétrons, prótons, nêutrons...), os átomos e as moléculas. Diferentemente do que ocorre deterministicamente com uma partícula na Mecânica Clássica, a detecção da posição de uma partícula quântica possui um forte caráter probabilístico. A natureza permite analisar a probabilidade de ocorrência de uma partícula quântica em uma certa região do espaço, mas não a sua posição precisa. Sendo assim, este postulado diz que a detecção da posição de uma partícula quântica é uma variável aleatória contínua². A respectiva função de densidade de probabilidade é modelada pela equação de Schrödinger, como descreveremos a seguir. Na verdade, todas as grandezas físicas associadas à uma partícula quântica são variáveis aleatórias. Esta afirmação pode causar estranheza quando observamos a construção da Mecânica Clássica. Porém, vale lembrar que físicos importantes do

¹Até o presente momento, não é possível se deduzir a equação de Schrödinger a partir de princípios físicos mais fundamentais. Sendo assim, a mesma se apresenta como uma "lei física", um axioma da teoria quântica moderna.

²Portanto, a probabilidade de uma partícula quântica estar em uma posição específica é nula.

século passado, como Einstein e o próprio Schrödinger, não aceitaram esta afirmação. Devemos ter em mente que os modelos matemáticos para os fenômenos físicos estão interessados em responder como a natureza funciona e não o porquê a natureza funciona de uma determinada maneira.

1.2.1 A função de onda de Schrödinger

Consideremos uma partícula quântica com massa (constante) conhecida, sujeita a uma energia potencial. A detecção posição desta partícula quântica no espaço é uma variável aleatória. A esta partícula quântica está associada uma função complexa, $\psi(t, x, y, z)$, $(x, y, z) \in \mathbb{R}^3$, $t \in \mathbb{R}_+$, chamada de *função de onda* (de Schrödinger). A função real

$$|\psi(t, x, y, z)|^2 = \psi^*(t, x, y, z)\psi(t, x, y, z),$$

onde a notação w^* representa o conjugado complexo do número complexo w , modela a densidade de probabilidade da variável posição. Ou seja, a probabilidade P da partícula quântica estar em uma certa região $\Omega \subset \mathbb{R}^3$ do espaço, no instante $t \geq 0$, é dada por

$$P = \iiint_{\Omega} |\psi(t, x, y, z)|^2 dx dy dz.$$

Consequentemente, devemos ter

$$\iiint_{\mathbb{R}^3} |\psi(t, x, y, z)|^2 dV = 1,$$

já que a probabilidade de encontrarmos a partícula quântica em algum lugar do espaço é 100%. Como discutiremos, a característica probabilística da função de onda não se revela apenas nesta densidade de probabilidade, mas em toda a teoria quântica. Vemos então que, do ponto de vista matemático, a função de onda deve ser, pelo menos, de quadrado integrável nas variáveis espaciais. Também assumiremos que as suas derivadas parciais com respeito às variáveis espaciais sejam de quadrado integrável. Mais precisamente, embora o assunto esteja fora do escopo elementar da nossa apresentação, admitiremos que as funções de onda pertençam ao espaço de Sobolev \mathbb{H}^1 (embora não façamos uso explícito aqui, para maiores detalhes sobre os espaços de Sobolev, indicamos [Medeiros and Miranda, 2011]). Em particular, as funções de onda necessariamente se anulam no infinito, contribuindo para que de fato tenhamos

$$\iiint_{\mathbb{R}^3} |\psi(t, x, y, z)|^2 dV = 1.$$

1.2.2 A Equação de Schrödinger

O coração do modelo que estamos abordando é, sem dúvida, a EDP conhecida como *Equação de Schrödinger*, proposta pelo físico austríaco Erwin Schrödinger em 1925 ([Schrödinger, 1982]). Fixada uma partícula quântica com massa (constante) conhecida m e sujeita a uma energia potencial dada por uma função real $U(t, x, y, z)$, a sua função de onda satisfaz a EDP abaixo, chamada de Equação de Schrödinger

$$i\hbar \frac{\partial \psi}{\partial t}(t, x, y, z) = \hat{H}\psi(t, x, y, z), \quad (1.1)$$

com

$$\hat{H} = -\frac{\hbar^2}{2m}\Delta + U(t, x, y, z),$$

onde i representa a unidade imaginária, Δ representa o operador Laplaciano, \hbar é uma constante conhecida como *constante de Dirac*, cujo valor é $1,054 \times 10^{-27} \text{ erg} \cdot \text{s}$. \hat{H} é o chamado operador hamiltoniano³.

É importante notar que a EDP acima não define necessariamente a função de onda. Poderíamos dizer que as funções de onda são as únicas soluções da equação de Schrödinger que possuem significado físico até o momento. Por exemplo, é fácil verificar que, para cada vetor constante (k_1, k_2, k_3) e cada constante $A \in \mathbb{R}$, a onda plana

$$Ae^{ik_1x - ik_2y - ik_3z - wt}$$

é solução da equação (1.1), no caso em que $U \equiv 0$. No entanto essas onda planas não são de quadrado integrável nas variáveis espaciais. Mais ainda, é possível mostrar que estas são as únicas soluções possíveis no caso em que $U \equiv 0$. Repare que este caso corresponderia a uma situação quanto mecânica em que a partícula quântica estaria completamente livre, sem sofrer influência de potencial algum.

Porém, mesmo que procuremos soluções⁴ em $\mathbb{H}^1(\mathbb{R}^3)$, não há obrigação que a solução tenha a integral de volume sobre \mathbb{R}^3 dando 1. Como esta equação é claramente linear, poderíamos pensar então em normalizar a solução obtida de modo que teríamos efetivamente uma densidade de probabilidade. E é exatamente isto que faremos, mas para isso precisamos de algumas observações. Primeiro, considere uma

³se a partícula quântica estiver sob efeito de um campo eletromagnético, o hamiltoniano assume uma outra forma, devido a força de Lorentz. Além disso, até o momento, são desconhecidas forças não conservativas independentes do tempo no nanomundo. Veja [Gondar and Cipolatti, 2011].

⁴A abordagem abstrata da equação (1.1) do ponto de vista da teoria das equações diferenciais parciais está além das nossas intenções. Veja [Iorio Júnior, 1987] para maiores detalhes.

solução ψ da equação (1.1). Vamos mostrar que a norma

$$N = \left(\iiint_{\mathbb{R}^3} |\psi(t, x, y, z)|^2 dx dy dz \right)^{\frac{1}{2}}$$

independe do tempo.

Proposição 1.1. Sendo $\psi(t, x, y, z)$ uma solução da Equação de Schrödinger, a integral

$$N = \left(\iiint_{\mathbb{R}^3} |\psi(t, x, y, z)|^2 dx dy dz \right)^{\frac{1}{2}}$$

não depende da variável temporal.

Para a demonstração desta proposição, faremos uso do seguinte lema preliminar:

Lema 1.2. Sendo $\psi(t, x, y, z)$ uma solução da Equação de Schrödinger, tem-se que

$$\iiint_{\mathbb{R}^3} \psi^* \Delta \psi dV = - \iiint_{\mathbb{R}^3} |\nabla \psi|^2 dV,$$

onde $V \subset \mathbb{R}^3$ é uma região compacta do espaço.

Demonstração do lema: Usando a identidade $\nabla \cdot (\phi \vec{a}) = \phi \nabla \cdot \vec{a} + \vec{a} \cdot \nabla \phi$, onde ϕ é um campo escalar e \vec{a} é um campo vetorial, e o fato de que $\nabla \cdot \nabla = \Delta$, onde ∇ representa o operador gradiente, vemos que $\nabla \cdot (\psi^* \nabla \psi) = \psi^* \Delta \psi + \nabla \psi^* \cdot \nabla \psi$. Portanto, considerando o Teorema da Divergência sobre um volume compacto $V \subset \mathbb{R}^3$, vemos que, para todo $t \geq 0$,

$$\begin{aligned} \left(\iiint_V \psi^* \Delta \psi dV \right)^{\frac{1}{2}} &= \iiint_V \nabla \cdot (\psi^* \nabla \psi) dV - \iiint_V \nabla \psi^* \cdot \nabla \psi dV \\ &= \iint_{\partial V} \langle \psi^* \nabla \psi, \vec{n} \rangle d(\partial V) - \iiint_V |\nabla \psi|^2 dV, \end{aligned}$$

onde ∂V representa a fronteira (topológica) de V .

Agora, como as funções de onda se anulam no infinito, a integral de superfície acima se anula quando tomamos o diâmetro do volume V tendendo para infinito. Então vemos que

$$\iiint_{\mathbb{R}^3} \psi^* \Delta \psi dV = - \iiint_{\mathbb{R}^3} |\nabla \psi|^2 dV.$$

■

Demonstração da proposição: Começamos multiplicando a equação de Schrödinger (1.1) pelo conjugado complexo de $i\psi$, gerando

$$\hbar\psi^*\frac{\partial\psi}{\partial t}(t,x,y,z) - i\frac{\hbar^2}{2m}\psi^*\Delta\psi(t,x,y,z) = -iU(t,x,y,z)|\psi(t,x,y,z)|^2.$$

Integrando esta última equação, sobre um volume compacto, para um $t \geq 0$ fixo, porém arbitrário, geramos

$$\begin{aligned} \hbar \iiint_{\mathbb{R}^3} \psi^* \frac{\partial\psi}{\partial t}(t,x,y,z) dV = \\ -i\frac{\hbar^2}{2m} \iiint_{\mathbb{R}^3} |\nabla\psi|^2 dV - i \iiint_{\mathbb{R}^3} U(t,x,y,z) |\psi(t,x,y,z)|^2 dV. \end{aligned}$$

Isto nos mostra que, para todo $t \geq 0$,

$$\iiint_{\mathbb{R}^3} \psi^* \frac{\partial\psi}{\partial t}(t,x,y,z) dV$$

é um número imaginário puro. É fácil ver que a parte real da integral acima coincide exatamente com

$$\frac{1}{2} \frac{d}{dt} \iiint_{\mathbb{R}^3} |\psi(t,x,y,z)|^2 dV$$

pois, considerando as partes real e imaginária de ψ , $Re(\psi)$ e $Im(\psi)$, respectivamente, e a definição do produto entre dois números complexos, vemos que

$$Re(\psi^* \frac{\partial\psi}{\partial t}) = Re(\psi) \frac{\partial Re(\psi)}{\partial t} + Im(\psi) \frac{\partial Im(\psi)}{\partial t}$$

e que

$$\frac{d|\psi|^2}{dt} = 2 \left[Re(\psi) \frac{\partial Re(\psi)}{\partial t} + Im(\psi) \frac{\partial Im(\psi)}{\partial t} \right].$$

Com isso, concluímos que

$$\frac{d}{dt} \iiint_{\mathbb{R}^3} |\psi(t,x,y,z)|^2 dV = 0.$$

Logo, a norma

$$N = \left(\iiint_{\mathbb{R}^3} |\psi(t,x,y,z)|^2 dV \right)^{\frac{1}{2}},$$

de fato, independe do tempo. ■

Então, daqui para frente, ao resolvermos a equação de Schrödinger, se a norma acima não for unitária, basta dividirmos a solução pela referida norma para obtermos a função de onda procurada. Por isso, quando tomarmos uma solução da equação de Schrödinger, já assumiremos que a mesma é uma função de onda.

Uma característica muito apreciável em modelagem matemática é a unicidade de soluções. De fato, de posse de uma propriedade de unicidade, ficamos livres para resolver o modelo com as mais variadas técnicas possíveis, tendo a certeza de que continuamos perseguindo a mesma solução. Vejamos que a unicidade de soluções da equação de Schrödinger é consequência da conservação temporal da norma.

Proposição 1.3. A Equação de Schrödinger goza da propriedade de unicidade de soluções.

Demonstração: De fato, sejam $\psi_1(t, x, y, z)$ e $\psi_2(t, x, y, z)$ duas soluções da equação (1.1) satisfazendo a mesma condição inicial $\psi_1(0, x, y, z) = \psi_2(0, x, y, z)$ para todo $(x, y, z) \in \mathbb{R}^3$. Como a equação de Schrödinger é linear, a função $\psi(t, x, y, z) = \psi_1(t, x, y, z) - \psi_2(t, x, y, z)$ também corresponde a uma solução e, portanto, devido à conservação da norma,

$$\iiint_{\mathbb{R}^3} |\psi(t, x, y, z)|^2 dV = \iiint_{\mathbb{R}^3} |\psi(0, x, y, z)|^2 dV = \iiint_{\mathbb{R}^3} 0 dV = 0,$$

o que implica, pela continuidade de ψ , em ψ ser nula. Isto revela que $\psi_1 \equiv \psi_2$. ■

Devido a esta unicidade de solução, dizemos que a função de onda define um estado (quanto-mecânico) da partícula quântica.

Feitas estas observações, podemos agora dizer precisamente que, dada a massa de uma partícula quântica e a energia potencial associada a ela, a posição desta partícula quântica no espaço, no instante $t \geq 0$, é uma variável aleatória (tridimensional)⁵ X cuja densidade de probabilidade é dada por

$$f_X(x, y, z) = |\psi(t, x, y, z)|^2,$$

onde ψ é a função de onda associada. Sendo assim, o valor esperado (esperança matemática) da posição da partícula quântica em um estado ψ é representado pelo vetor $\langle X(t) \rangle = (\langle x(t) \rangle, \langle y(t) \rangle, \langle z(t) \rangle)$, onde

⁵Deve-se observar que a aleatoriedade está presente nas variáveis espaciais, mas não na variável temporal.

cada componente é dada por

$$\langle x(t) \rangle = \iiint_{\mathbb{R}^3} x |\psi(t, x, y, z)|^2 dV = \iiint_{\mathbb{R}^3} \psi(t, x, y, z)^* x \psi(t, x, y, z) dV;$$

$$\langle y(t) \rangle = \iiint_{\mathbb{R}^3} y |\psi(t, x, y, z)|^2 dV = \iiint_{\mathbb{R}^3} \psi(t, x, y, z)^* y \psi(t, x, y, z) dV;$$

$$\langle z(t) \rangle = \iiint_{\mathbb{R}^3} z |\psi(t, x, y, z)|^2 dV = \iiint_{\mathbb{R}^3} \psi(t, x, y, z)^* z \psi(t, x, y, z) dV.$$

1.2.3 Os Operadores Quânticos

Todas as grandezas físicas (que não sejam as fundamentais) associadas à uma partícula quântica são variáveis aleatórias. Dada uma grandeza física L , diremos que um operador \hat{L} atuando sobre o espaço $\mathbb{H}^1(\mathbb{R}^3)$ é um operador quântico associado à L se

$$\langle L \rangle = \iiint_{\mathbb{R}^3} \psi^* \hat{L} \psi dV.$$

Usando uma Transformada de Fourier tridimensional e a identidade de Parseval, pode-se justificar que, para o momento linear (quantidade de movimento) p ,

$$\langle p \rangle = \iiint_{\mathbb{R}^3} \psi^* (-i\hbar) \nabla \psi dV.$$

A partir disso, mostra-se que o operador quântico proveniente da energia cinética é dado por $\frac{(-i\hbar\nabla)^2}{2m} = \frac{-\hbar^2\Delta}{2m}$, onde m é a massa da partícula quântica. Isto está feito no capítulo 7 de [Gondar and Cipolatti, 2011].

Outro operador quântico fundamental: O operador de Hamilton. Em um sistema quântico sujeito a um campo de forças $F(t, x, y, z) = -\nabla U(t, x, y, z)$, (note que não há dependência do momento linear) a *hamiltoniana* H de uma partícula quântica de massa m é

$$H = U(t, x, y, z) + \frac{|p|^2}{2m}.$$

O operador quântico associado a esta grandeza é o operador de Hamilton

$$\hat{H} = -\frac{\hbar^2}{2m} \Delta \psi(t, x, y, z) + U(t, x, y, z).$$

Se a hamiltoniana é estacionária, o operador de Hamilton representa a energia mecânica do sistema.

1.2.4 Operadores Hermitianos

A determinação de um operador quântico associado a uma grandeza física não é uma tarefa simples em geral. Mas, sendo L uma grandeza física, tanto ela, quanto o seu valor esperado são reais. A partir desta conclusão pode-se mostrar ([Gondar and Ciolatti, 2011]) que uma condição necessária para que um operador \hat{L} seja o operador quântico associado a uma grandeza física L é que \hat{L} seja hermitiano, como passamos a definir.

Definição 1.4. Um operador complexo \hat{L} é hermitiano se ele coincide com seu operador adjunto. Mais precisamente, se

$$\langle f | \hat{L} | g \rangle = \iiint_{\mathbb{R}^3} f^*(x, y, z) \hat{L} g(x, y, z) dV = \iiint_{\mathbb{R}^3} g^*(x, y, z) \hat{L} f(x, y, z) dV = \langle g | \hat{L} | f \rangle,$$

quaisquer que sejam as funções contínuas f e g para as quais as integrais acima estejam definidas.

A proposição abaixo estabelece um fato fundamental sobre os operadores hermitianos.

Proposição 1.5. Sendo \hat{L} um operador hermitiano, seus autovalores são reais e suas autofunções associadas a autovalores distintos são ortogonais (o sentido da ortogonalidade fica claro na demonstração abaixo).

Demonstração: Seja φ uma autofunção de \hat{L} associada a um autovalor $\lambda \in \mathbb{C}$. Então $\hat{L}\varphi = \lambda\varphi$. Daí,

$$\langle \varphi | \hat{L} | \varphi \rangle = \lambda \iiint_{\mathbb{R}^3} |\varphi(x, y, z)|^2 dV$$

e

$$\langle \varphi | \hat{L} | \varphi \rangle^* = \lambda^* \iiint_{\mathbb{R}^3} |\varphi|^2 dV.$$

Portanto, como \hat{L} é hermitiano e φ não é identicamente nula, vemos que $\lambda = \lambda^*$, revelando que $\lambda \in \mathbb{R}$, como queríamos. Sejam agora φ_1 e φ_2 duas autofunções de \hat{L} associadas aos autovalores distintos λ_1 e λ_2 , respectivamente. Já sabemos que estes autovalores são reais. Com isso,

$$\langle \varphi_1 | \hat{L} | \varphi_2 \rangle = \lambda_2 \iiint_{\mathbb{R}^3} \varphi_1^* \varphi_2 dV$$

e

$$\langle \varphi_2 | \hat{L} | \varphi_1 \rangle = \lambda_1 \iiint_{\mathbb{R}^3} \varphi_2^* \varphi_1 dV.$$

Porém, \hat{L} é hermitiano. Logo

$$\langle \varphi_1 | \hat{L} | \varphi_2 \rangle = \langle \varphi_2 | \hat{L} | \varphi_1 \rangle^* = \lambda_1 \iiint_{\mathbb{R}^3} \varphi_1^* \varphi_2 dV,$$

pois $\lambda_1 \in \mathbb{R}$. Daí, como estes autovalores são distintos, vemos que, obrigatoriamente,

$$\iiint_{\mathbb{R}^3} \varphi_1^* \varphi_2 dV = 0,$$

como queríamos. ■

1.2.5 Medições de grandezas observáveis, medições simultâneas e o Princípio da Incerteza de Heisenberg

Como vimos, as grandezas físicas são representadas por operadores quânticos. Esta representação tem uma importante implicação quando efetuamos medições de uma grandeza L sobre um dado estado quanto-mecânico ψ : A grandeza L se encontrará determinada no estado ψ se, e somente se, ψ for uma autofunção do respectivo operador quântico

$$\hat{L}\psi(t, x, y, z) = \lambda\psi(t, x, y, z).$$

Neste caso, o único valor que a natureza admite para a grandeza L é o autovalor λ . Existe uma interessante justificativa probabilística para este fato, dada em [Gondar and Cipolatti, 2011]: Vamos analisar um processo de medição clássico, porém hipotético. Vamos supor que em um fluxo de partículas, a cada uma das partículas está associado um valor, eventualmente diferente, de uma determinada grandeza física $L : L_1, L_2, \dots, L_k, \dots$. Imagine que tais valores possam ser medidos para cada partícula com absoluta precisão (por isso o processo é hipotético). Efetuando N medições da grandeza L , veremos que o valor L_1 se manifesta n_1 vezes, L_2 se manifesta n_2 vezes e assim sucessivamente. A probabilidade (definição frequentista de probabilidade) de que em uma medição de L , neste fluxo, se registre o valor L_j é dada por

$$P_j = \lim_{N \rightarrow \infty} \frac{n_j}{N},$$

onde

$$\sum_k n_k = N.$$

Além disso, o valor médio e a variância de L no fluxo são dados por

$$\langle L \rangle = \sum_j P_j L_j;$$

$$\langle (\Delta L)^2 \rangle = \sum_j P_j (L_j - \langle L \rangle)^2.$$

Se a variância é nula, não ocorre dispersão dos valores e então todas as partículas do fluxo estão associadas ao mesmo valor de L . Embora a natureza do fenômeno seja diferente, as considerações acima também se aplicam, do ponto vista quanto-mecânico, quando efetuamos medições da grandeza L sobre um estado $\psi(t, x, y, z)$. Considere o operador desvio médio:

$$\widehat{\Delta L} = \hat{L} - \lambda, \quad \lambda = \langle L \rangle.$$

Como $\widehat{\Delta L}$ é hermitiano, vemos que

$$\begin{aligned} \langle (\Delta L)^2 \rangle &= \iiint_{\mathbb{R}^3} \psi^* \widehat{\Delta L} \widehat{\Delta L} \psi \, dx \, dy \, dz \\ &= \iiint_{\mathbb{R}^3} \widehat{\Delta L} \psi \widehat{\Delta L} \psi^* \, dx \, dy \, dz \\ &= \iiint_{\mathbb{R}^3} |\widehat{\Delta L} \psi|^2 \, dx \, dy \, dz. \end{aligned}$$

Os estados quanto-mecânicos $\psi(t, x, y, z)$ para os quais a variância da medida da grandeza L é nula devem satisfazer

$$\widehat{\Delta L} \psi = 0,$$

$$\hat{L} \psi = \lambda \psi.$$

Existem grandezas físicas que geram operadores quânticos com espectro discreto, como é o caso da energia mecânica do (único) elétron do átomo de hidrogênio. Estes exemplos refletem o comportamento quântico da natureza.

Na mecânica clássica, a definição de momento linear permite que, uma vez conhecida a trajetória de uma partícula $S(t) = (x(t), y(t), z(t))$, conheçamos o valor do momento e, reciprocamente, uma vez medido o momento, podemos estabelecer a trajetória da partícula:

$$p = m \frac{dS}{dt}.$$

O conceito de trajetória clássica não se aplica às partículas quântica de modo que não existe nenhuma relação definida entre momento e posição. Esta última afirmação pode ser obtida a partir do chamado *Princípio da Incerteza de Heisenberg*. A desigualdade obtida originalmente por Heisenberg foi a seguinte: sendo o momento $p = (p_1, p_2, p_3)$ e as coordenadas da posição dadas por (x_1, x_2, x_3) ,

$$\langle (\Delta p_j)^2 \rangle \langle (\Delta x_j)^2 \rangle \geq \frac{\hbar^2}{4}, \quad j = 1, 2, 3.$$

Hoje, este "princípio" pode ser obtido como consequência da teoria quântica apresentada aqui e é um caso particular de um teorema geral sobre medições simultâneas de grandezas físicas que possuem operadores quânticos que não comutam (veja uma demonstração em [Gondar and Cipolatti, 2011]). O Princípio da Incerteza de Heisenberg estabelece que, quanto mais precisas forem as medidas da posição de uma partícula quântica, menos precisas serão as medidas dos momentos (evidentemente que nos estados onde estão definidos) e, reciprocamente, quanto mais precisas forem as medidas do momento, menos precisas serão as medidas da posição. Este resultado teórico desestimula qualquer tentativa de construção de um dispositivo que possa medir simultaneamente estas duas grandezas. No entanto, existem pares de grandezas físicas, L e K , que não são tão incompatíveis. São as grandezas simultaneamente medíveis. Para estas grandezas, existem estados quanto-mecânicos nos quais a medição acurada das duas grandezas é possível. Do que discutimos sobre as medições de uma grandeza física, isso significa que estes estados são simultaneamente autofunções dos operadores \hat{L} e \hat{K} . De fato, pode-se mostrar (veja [Prugovecki, 1971]) que duas grandezas L e K , tais que seus respectivos operadores quânticos, \hat{L} e \hat{K} , tenham espectro discreto e conjuntos completos de autofunções, são simultaneamente medíveis se, e só se, seus operadores quânticos comutam,

$$\hat{L}\hat{K} = \hat{K}\hat{L}.$$

Modelos Estacionários

Agora vamos nos restringir ao caso em que a energia potencial atuante sobre a partícula quântica, $U(x, y, z)$, independe do tempo e o operador de Hamilton possui espectro discreto. Com isto, o operador Hamiltoniano é estacionário:

$$\hat{H}\psi(t, x, y, z) = -\frac{\hbar^2}{2m}\Delta\psi(t, x, y, z) + U(x, y, z)\psi(t, x, y, z).$$

Vamos propor o método de separação de variáveis: $\psi(t, x, y, z) = T(t)\varphi(x, y, z)$. Daí,

$$i\hbar\frac{T'(t)}{T(t)} = \frac{\hat{H}\varphi(x, y, z)}{\varphi(x, y, z)}.$$

Como \hat{H} é estacionário, existe uma constante $\lambda \in \mathbb{C}$ tal que

$$i\hbar\frac{T'(t)}{T(t)} = \frac{\hat{H}\varphi(x, y, z)}{\varphi(x, y, z)} = \lambda.$$

Podemos ver então que o problema se reduz a determinar as autofunções do operador de Hamilton. Como \hat{H} é hermitiano, os autovalores correspondentes são reais. No jargão da Mecânica Quântica, este último problema de autovalor é chamado de *Equação de Schrödinger independente do tempo*.

Suponhamos que \hat{H} tenha um conjunto de autovalores, $\{\lambda_n; n \in \mathbb{N}\}$ e um respectivo conjunto de autoestados, $\{\varphi_n; n \in \mathbb{N}\}$, ortonormal e completo em $L^2(\mathbb{R}^3)$. Para cada autovalor λ_n vemos que existe uma constante complexa c_n , tal que

$$T_n(t) = \exp(-i\lambda_n t/\hbar)c_n.$$

Como a Equação de Schrodinger é linear, escrevemos

$$\psi(t, x, y, z) = \sum_{n=1}^{\infty} c_n \varphi_n(x, y, z) \exp(-i\lambda_n t/\hbar).$$

Dada uma condição inicial $\varphi(0, x, y, z) = \varphi_0(x, y, z)$, temos que

$$\psi_0(x, y, z) = \sum_{n=1}^{\infty} c_n \varphi_n(x, y, z),$$

com

$$c_n = \iiint_{\mathbb{R}^3} \varphi_n^*(x, y, z) \varphi_0(x, y, z) dV$$

Além disso, pela Identidade de Parseval,

$$\|\varphi_0\|^2 = \iiint_{\mathbb{R}^3} |\varphi_0(x, y, z)|^2 dV = \sum_{n=1}^{\infty} |c_n|^2.$$

Exemplo de solução analítica: o modelo da partícula quântica na caixa

Agora apresentaremos uma situação ideal onde podemos obter a solução analítica da equação de Schrödinger. Trata-se do modelo da partícula quântica na caixa. Neste sistema, as partículas quântica estão sob a ação de um campo de forças que, embora seja nulo no interior de uma região Ω , é tão intenso no exterior de Ω que as partículas quântica que estiverem inicialmente no interior de Ω , ficarão aprisionadas lá. Vamos tomar o caso unidimensional, onde $\Omega = [0, a]$. Este modelo também é chamado de "poço retangular infinito". Neste caso, o potencial do modelo é dado por

$$U_\infty = \begin{cases} 0, & \text{se } x \in [0, a] \\ +\infty & \text{se } x \notin [0, a] \end{cases}$$

O valor médio da energia mecânica em um estado ψ é dado por

$$\begin{aligned} \langle H \rangle &= \langle \psi | \hat{H} | \psi \rangle \\ &= \int_{\mathbb{R}} \psi^*(t, x) \left(-\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2}(t, x) + U_\infty(x)\psi(t, x) \right) dx \\ &= \left(-\frac{\hbar^2}{2m} \int_{\mathbb{R}} \psi^*(t, x) \frac{d^2\psi}{dx^2}(t, x) dx + \int_{\mathbb{R}} U_\infty(x) \|\psi(t, x)\|^2 dx \right) \end{aligned}$$

Note que

$$\int_{\mathbb{R}} \psi^*(t, x) \frac{d^2\psi}{dx^2}(t, x) dx = \int_{\mathbb{R}} \frac{d}{dx} \left(\psi^* \frac{d\psi}{dx} \right) dx - \int_{\mathbb{R}} \left\| \frac{d\psi}{dx}(t, x) \right\|^2 dx = - \int_{\mathbb{R}} \left\| \frac{d\psi}{dx}(t, x) \right\|^2 dx,$$

onde a última igualdade decorre do fato de ψ se anular no infinito.

Com isso, vemos que os estados ψ para os quais faz sentido definir o valor médio da energia mecânica são aqueles em que

$$\langle H \rangle = \frac{\hbar^2}{2m} \int_{\mathbb{R}} \left\| \frac{d\psi}{dx}(t, x) \right\|^2 dx + \int_{\mathbb{R}} U_\infty(x) \|\psi(t, x)\|^2 dx < +\infty.$$

Devido à intensidade de U_∞ fora do intervalo $[0, a]$, devemos ter $\psi(t, x) = 0$ sempre que $x \notin [0, a]$.

Portanto a equação de Schrödinger se torna

$$\begin{cases} -i\hbar \frac{\partial \psi}{\partial t}(t, x) = -\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2}(t, x), & t > 0, x \in (0, a) \\ \psi(t, 0) = \psi(t, a) = 0 \end{cases}$$

O problema de autovalor correspondente é então

$$\begin{cases} -i \frac{d^2 \varphi}{dx^2} = \frac{2m}{\hbar^2} E \varphi, & x \in (0, a) \\ \varphi(0) = \varphi(a) = 0 \end{cases}$$

Esta é uma EDO linear de segunda ordem. Sabemos que este tipo de equação tem solução dada em função das raízes de sua equação característica. Neste caso, a equação característica é $z^2 + \frac{2m}{\hbar^2} E = 0$, cujas raízes são $\pm \sqrt{-\frac{2m}{\hbar^2} E}$. Se $E \leq 0$ então essas raízes são reais e, devido a solução geral da EDO para esses casos e às condições de fronteira $\varphi(0) = \varphi(a) = 0$, teremos apenas a solução trivial $\varphi \equiv 0$. Portanto, devemos assumir que $E > 0$. Daí, as raízes são imaginárias e a solução geral da EDO é

$$\varphi(x) = A \cos \left(\sqrt{\frac{2m}{\hbar^2} E} x \right) + B \sin \left(\sqrt{\frac{2m}{\hbar^2} E} x \right),$$

com $A, B \in \mathbb{R}$ sendo constantes. Como $\varphi(0) = 0$, vemos que $A = 0$. Então, para termos solução não trivial, necessariamente devemos ter $B \neq 0$. Como $\varphi(a) = 0$, vemos que

$$B \sin \left(\sqrt{\frac{2m}{\hbar^2} E} a \right) = 0.$$

Como $B \neq 0$, obtemos os valores permitidos para a energia

$$E_n = \frac{n^2 \pi^2 \hbar^2}{2ma^2}, \quad n \in \mathbb{N}.$$

Esta quantização revela que a energia é uma variável aleatória discreta neste modelo.

1.2.6 Potencial Periódico e o Teorema de Bloch

O Teorema de Bloch é um dos resultados mais importantes da Física da Matéria Condensada. Com ele temos a forma geral da solução da Equação de Schrödinger para uma partícula quântica sujeita a um potencial periódico. Tanto o enunciado quanto a demonstração deste teorema estão aqui em uma forma que não encontramos na literatura. Por este motivo, não indicamos referências bibliográficas para este tema. No entanto, todos os livros que abordam a equação de Schrödinger sob um potencial periódico abordam "algum teorema de Bloch".

Consideremos a Equação de Schrödinger independente do tempo

$$\hat{H} \varphi(x, y, z) = \left[-\frac{\hbar^2}{2m} \Delta + U(x, y, z) \right] \varphi(x, y, z) = \lambda \varphi(x, y, z).$$

Admitamos que o potencial U seja periódico, digamos

$$U(x + r_1, y + r_2, z + r_3) = U((x, y, z) + \vec{r}) = U(x, y, z),$$

qualquer que seja o ponto (x, y, z) , onde $\vec{r} = (r_1, r_2, r_3)$ é um vetor fixo.

Definamos o operador translação $\hat{T}_{\vec{r}}$ por

$$\hat{T}_{\vec{r}}(f(x, y, z)) = f((x, y, z) + \vec{r}),$$

onde f é uma função qualquer. Vamos mostrar que o operador translação comuta com o operador hamiltoniano. Realmente,

$$\begin{aligned} \hat{T}_{\vec{r}}\hat{H}(\varphi(x, y, z)) &= \hat{T}_{\vec{r}}\left(\frac{-\hbar^2\nabla^2}{2m}\varphi(x, y, z) + U(x, y, z)\varphi(x, y, z)\right) = \\ &= \frac{-\hbar^2\nabla^2}{2m}\varphi((x, y, z) + \vec{r}) + U((x, y, z) + \vec{r})\varphi((x, y, z) + \vec{r}) = \\ &= \frac{-\hbar^2\nabla^2}{2m}\varphi((x, y, z) + \vec{r}) + U(x, y, z)\varphi((x, y, z) + \vec{r}) = \\ &= \hat{H}(\varphi((x, y, z) + \vec{r})) = \hat{H}\hat{T}_{\vec{r}}(\varphi((x, y, z))), \end{aligned}$$

como queríamos. Pelo o que discutimos sobre grandezas simultaneamente mensuráveis, esta comutatividade implica em os autoestados do operador Hamiltoniano também serem autofunções do operador translação. Fixemos $m \in \mathbb{Z}$ e tomemos um destes autoestados. Então

$$\varphi((x, y, z) + m\vec{r}) = \hat{T}_{\vec{r}}^m\varphi(x, y, z) = \tau^m\varphi(\vec{x}),$$

onde τ é o autovalor do operador translação associado a autofunção φ . Logo,

$$\|\varphi((x, y, z) + m\vec{r})\| = |\tau|^m\|\varphi(x, y, z)\|.$$

Como φ não é identicamente nula, existe $(x_0, y_0, z_0) \in \mathbb{R}^3$ tal que $\varphi(x_0, y_0, z_0) \neq 0$. Vamos investigar

$\lim_{m \rightarrow \pm\infty} \|\varphi((x_0, y_0, z_0) + m\vec{r})\|$. Se $|\tau| > 1$,

$$\lim_{m \rightarrow +\infty} \|\varphi((x_0, y_0, z_0) + m\vec{r})\| = \lim_{m \rightarrow +\infty} |\tau|^m\|\varphi(\vec{x})\| = +\infty.$$

Por outro lado, se $|\tau| < 1$,

$$\lim_{m \rightarrow -\infty} \|\varphi((x_0, y_0, z_0) + m\vec{r})\| = \lim_{m \rightarrow -\infty} |\frac{1}{\tau}|^{-m}\|\varphi(\vec{x})\| = +\infty.$$

Ambos os resultados são absurdos em Mecânica Quântica, uma vez que os autoestados se anulam no infinito. Daí, $|\tau| = 1$. Sendo assim, mostramos que o autovalor do operador translação é da forma $\tau = e^{i\theta}$, $\theta \in \mathbb{R}$. Portanto, $\varphi((x, y, z) + m\vec{r}) = e^{i\theta} \varphi(x, y, z)$. Esta igualdade já define as chamadas *funções de Bloch*, porém, em física do estado sólido, usamos vetores, chamados de *vetor de onda de Bloch*, \vec{k} tais que⁶ $\theta = \vec{k} \cdot \vec{r}$. Daí,

$$\varphi_k((x, y, z) + \vec{r}) = e^{i\vec{k} \cdot \vec{r}} \varphi_k(x, y, z).$$

Esta formulação para os auto estados é adequada para as nossas aplicações. No entanto, vamos obter uma versão equivalente, que é bastante popular. Defina $u_k(x, y, z) = \varphi(x, y, z) e^{-i\vec{k} \cdot \vec{x}}$. É fácil verificar que u_k é periódica de período \vec{r} . De fato,

$$u_k((x, y, z) + \vec{r}) = \varphi((x, y, z) + \vec{r}) e^{-i\vec{k} \cdot ((x, y, z) + \vec{r})} = \varphi(x, y, z) e^{-i\vec{k} \cdot (x, y, z)} = u_k(x, y, z).$$

Dessa observação decorre a seguinte versão equivalente: Os autoestados do operador Hamiltoniano com potencial periódico têm a seguinte propriedade

$$\varphi_k(x, y, z) = e^{i\vec{k} \cdot (x, y, z)} u_k(x, y, z),$$

onde u_k é uma função com a mesma periodicidade do potencial.

Resta verificarmos a recíproca desta afirmação: a aplicação

$$\varphi_k(x, y, z) = e^{i\vec{k} \cdot (x, y, z)} u_k(x, y, z),$$

onde u_k é uma função com a mesma periodicidade do potencial, é um autoestado do operador Hamiltoniano. Com efeito, esta aplicação é uma autofunção do operador translação pois,

$$\hat{T}_{\vec{r}}(\varphi_k(x, y, z)) = e^{i\vec{k} \cdot \vec{r}} e^{i\vec{k} \cdot (x, y, z)} u_k((x, y, z) + \vec{r}) = e^{i\vec{k} \cdot \vec{r}} e^{i\vec{k} \cdot (x, y, z)} u_k((x, y, z)),$$

onde para obtermos a última igualdade fizemos uso da periodicidade da função u_k . Daí, como o operador translação comuta com o operador Hamiltoniano (devido a periodicidade do potencial), a aplicação φ_k é um autoestado de \hat{H} .

Estabelecemos então o seguinte teorema:

⁶Quando estamos em uma rede cristalina, a periodicidade do potencial decorre da periodicidade da rede. Neste caso, existe um significado físico para os vetores \vec{k} relacionado ao conceito de rede recíproca e da teoria de bandas, que não abordaremos aqui. Maiores detalhes, indicamos o capítulo 8 de [Ashcroft and Mermin, 1976]

Teorema 1.6. (Teorema de Bloch) Consideremos a Equação de Schrödinger independente do tempo, onde o potencial é periódico, com período \vec{r} . Nestas condições, uma aplicação φ é um autoestado do operador hamiltoniano se, e somente se, tem a forma

$$\varphi(x, y, z) = e^{i\vec{k}\cdot(x,y,z)} u_k(x, y, z),$$

onde k é um vetor constante, chamado de *vetor de onda de Bloch*, e u_k é uma função periódica de período \vec{r} .

1.2.7 Generalização para sistemas de muitas partículas quântica e os principais métodos de solução da Equação de Schrödinger

A questão geral sobre existência de soluções da Equação de Schrödinger é tratada fazendo uso de resultados e técnicas da teoria das Equações Diferenciais Parciais em Espaços de Hilbert (para isto, recomendamos [Iorio Júnior, 1987] e [Prugovecki, 1971]). Do ponto de vista estritamente da Modelagem Matemática, a preocupação com a existência de soluções para um modelo muitas vezes não recebe tanto destaque. Se o modelo é coerente, a sua solução existe e é o objeto que está sendo modelado! Em vista da proposição 1.3, os esforços são dedicados à construção de possíveis técnicas de obtenção de solução. O escopo do nosso trabalho está incluído nos modelos estacionários. A resolução das equações da mecânica quântica é uma tarefa altamente não trivial que necessita de várias hipóteses simplificadoras específicas para o caso em que estão sendo tratadas. Mesmo com muitas destas aproximações, soluções analíticas só são conhecidas para um número extremamente pequeno de situações. Podemos citar o modelo do poço retangular infinito, o oscilador quântico isotrópico ([Gondar and Cicolatti, 2011]) e o átomo de hidrogênio. Este último exemplo, além de uma grande importância histórica, é importante para a validação do modelo de Schrödinger, pois reproduz as informações que já eram observadas sobre o (único) elétron do átomo de hidrogênio. Para este caso, se obtém um espectro discreto para o operador Hamiltoniano, revelando que a natureza apenas admite valores discretos para a energia mecânica daquele elétron. A solução analítica para o átomo de hidrogênio é bastante sutil e reconstrói todas as informações experimentais que já eram conhecidas sobre o hidrogênio. Para uma apresentação absolutamente correta e precisa, recomendamos o capítulo 19, seção 8, de [Barata, 2011]. A autofunção de um elétron é chamada de orbital espacial.

Para os outros átomos diferentes do hidrogênio, recaímos em um problema de muitos corpos, pois temos mais de um elétron, além do núcleo. No entanto, a abordagem é análoga: Para um átomo com

n elétrons, formamos a Equação de Schrödinger independente do tempo para uma função de onda, que agora depende das coordenadas de cada elétron e das coordenadas do núcleo, $\hat{H}\varphi = \epsilon\varphi$, com

$$\varphi = \varphi(g_N, g_1, g_2, \dots, g_n),$$

onde g_i representa as três coordenadas espaciais para o elétron i e g_N representa as três coordenadas espaciais para o núcleo. Temos então uma função de $3 \times (n + 1)$ variáveis. O operador Hamiltoniano por sua vez é dado pela energia mecânica do sistema:

$$\hat{H} = K_N + K_e + V_{eN} + V_{ee},$$

onde

- K_N representa o operador energia cinética do núcleo;
- K_e representa a soma dos operadores energia cinética de cada elétron;
- V_{eN} representa a soma dos potenciais coulombianos atrativos de cada par elétron-núcleo;
- V_{ee} representa a soma dos potenciais coulombianos repulsivos de cada par elétron-elétron.

Os resultados sobre o modelo Schrödinger descritos nas seções anteriores fizeram uso da Análise Matemática em três dimensões. Os mesmos resultados podem ser obtidos (inclusive o Teorema de Bloch) para este modelo multidimensional usando a Análise Matemática m -dimensional. Também temos a mesma interpretação probabilística para as funções de onda: A região do espaço $(3(n + 1))$ -dimensional é formada pelo produto cartesiano de $3 \times (n + 1)$ subregiões do \mathbb{R}^3 (topologia produto) e a respectiva integral nos revela a probabilidade de encontrarmos, simultaneamente, o núcleo na primeira subregião e o i -ésimo elétron na i -ésima subregião. No caso de sólidos ou moléculas, que são o nosso interesse no presente trabalho, a única diferença é que temos vários átomos e, portanto, vários núcleos. Com isso, o termo K_N representaria a soma do operadores energia cinética dos vários núcleos, o termo V_{eN} incluiria os pares elétron-núcleo para os vários núcleos e devemos adicionar um termo V_{NN} , representando a soma dos potenciais coulombianos repulsivos de cada par núcleo-núcleo. Surpreendentemente, não se conseguiu obter solução analítica para átomos com mais de um elétron. As soluções aproximadas são obtidas sob algumas hipóteses simplificadoras. Neste trabalho fazemos uso de duas:

1. *Aproximação de Born-Oppenheimer*

A primeira hipótese simplificadora que faremos uso é a chamada *Aproximação de Born-Oppenheimer*. Esta aproximação é baseada no fato de que a massa do núcleo atômico é muito maior (mais de mil vezes) do que a massa dos elétrons que o circundam. Daí, os núcleos se movem muito mais lentamente do que os elétrons, embora o movimento eletrônico reaja instantaneamente à mudança das posições nucleares. Portanto, tendo como objetivo a descrição quanto-mecânica eletrônica, é razoável considerar os elétrons sob um campo de forças gerado por núcleos fixos. Desta forma, pensando em apenas um elétron, a função de onda deve ser referenciada à posição do núcleo, mas tratada como função apenas da posição do elétron, como num problema de dois corpos. Isto é feito na solução do átomo de hidrogênio, por exemplo. No caso dos sólidos, é proposta uma função de onda dada pelo produto de uma função das coordenadas dos núcleos por uma função das coordenadas dos elétrons e resolve-se uma Equação de Schrödinger para o movimento nuclear e outra para o movimento eletrônico. O correspondente operador hamiltoniano eletrônico não leva em consideração a energia cinética dos núcleos e o potencial repulsivo núcleo-núcleo é constante. O nosso interesse está apenas no problema eletrônico.

2. *Aproximação de Hartree: um modelo de partículas quânticas independentes*

O potencial repulsivo V_{ee} do operador hamiltoniano contém termos que dependem das coordenadas de dois elétrons. Obviamente, estes termos não aparecem no problema do átomo de hidrogênio, por exemplo. Termos deste tipo impedem que a Equação de Schrödinger seja resolvida de forma exata. Hartree, em 1927, propôs uma estratégia para isso. Na aproximação de Hartree, os elétrons se movem independentemente uns dos outros, porém, além do potencial atrativo devido aos núcleos, cada elétron fica sujeito a um potencial repulsivo médio, devido à sua interação com os outros elétrons da molécula. Desta forma, o operador hamiltoniano é aproximado por um "operador hamiltoniano médio", formado pela soma dos vários operadores hamiltonianos de cada elétron.

Os métodos computacionais de solução aproximada para a Equação de Schrödinger são separados em dois grandes grupos: os *métodos de primeiros princípios* ou *métodos ab-initio* e os *métodos semiempíricos*.

Os métodos ab-initio são aqueles que não utilizam nenhuma informação oriunda de medidas experimentais sobre o sistema estudado. Então, a partir das posições atômicas e das interações entre os átomos,

os métodos de primeiros princípios são capazes de resolver o problema quântico e prover a descrição tanto do sistema eletrônico, quanto nuclear, do material. O primeiro método deste tipo foi o método de Hartree-Fock (HF). Outros métodos de destaque são aqueles baseados na Teoria do Funcional de Densidade (DFT). As metodologias de primeiros princípios são bastante gerais e acuradas, no entanto têm complexidade computacional muito alta (em geral, $O(N^3)$, onde N é o número de átomos do sistema estudado).

Os métodos semiempíricos são, de forma simplista, métodos baseados em aproximações feitas nas formulações ab-initio, com HF. Estes métodos negligenciam certos termos nestas formulações e compensam estas aproximações com a introdução de parâmetros experimentais que servem como nós interpoladores para ajuste dos resultados. Em geral, estes métodos têm complexidade computacional bem inferior, quando comparados aos métodos ab-initio. Com isso, para alguns sistemas formados por muitos átomos, podem ser os únicos métodos viáveis de solução. Também podem ser usados de forma encadeada com os métodos ab-initio. Um exemplo desta situação é o uso de metodologias semiempíricas para obter a melhor estrutura de partida de uma molécula de muitos átomos para o uso em cálculos de primeiros princípios. Uma grande limitação dos métodos semiempíricos é que estes só podem ser usados em sistemas para os quais já foram desenvolvidos parâmetros para todos os átomos que o compõem. Um dos objetivos da área chamada de *Química Quântica* é o desenvolvimento de conjuntos de parâmetros semiempíricos eficientes para uma ampla faixa de aplicações.

A metodologia DMTB, objetivo desta dissertação, se encontra no grupo dos métodos semiempíricos, como veremos.

Capítulo 2

Metodologia em sólidos cristalinos

Muitos sólidos usados em ciência e engenharia apresentam uma estrutura física macroscópica bastante organizada, orientada e com um alto grau de simetria. São os chamados *cristais*. Esta estrutura macroscópica organizada dos materiais pode ser explicada por um modelo matemático para o arranjo estrutural de seus átomos constituintes. Este modelo é conhecido como *estrutura cristalina*. Hoje existem técnicas experimentais (difração por raio X) para a caracterização de estruturas cristalinas, mostrando que estes arranjos estruturais internos, de fato, existem. O modelo de estruturas cristalinas é baseado no conceito matemático de *rede de Bravais*, que passamos a descrever. O objetivo deste capítulo é apenas introduzir o conceito de estrutura cristalina para ser usado no capítulo seguinte. Para uma abordagem mais completa do assunto, recomendamos os clássicos [Kittel, 1971] e [Ashcroft and Mermin, 1976].

2.1 Redes de Bravais e estrutura cristalina

No século XIX, o Físico francês August Bravais estudou as diferentes possibilidades de se posicionar pontos geométricos de forma periódica no espaço. Este estudo levou ao seguinte conceito:

Definição 2.1. Sejam n vetores de \mathbb{R}^n , $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n$. O conjunto

$$B = \left\{ \vec{x} \in \mathbb{R}^n; \vec{x} = \sum_{i=1}^n n_i \vec{a}_i, \text{ com } n_i \in \mathbb{Z} \right\}$$

é chamado de **rede de Bravais**. Os vetores \vec{a}_i são chamados de **vetores primitivos** da rede.

Vale notar que a escolha dos vetores primitivos de uma rede de Bravais não é única, como ilustra a figura de uma rede de Bravais bidimensional abaixo.

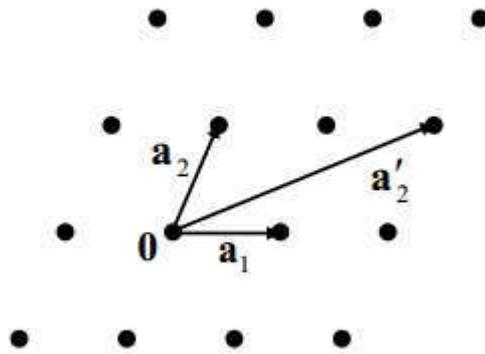


Figura 2.1: os três pares de vetores, $\{\vec{a}_1, \vec{a}_2\}$; $\{\vec{a}_1, \vec{a}'_2\}$ e $\{\vec{a}_2, \vec{a}'_2\}$ produzem todos os pontos da rede.

Decorre diretamente da definição que os pontos de uma rede de Bravais exibem um arranjo regular. Embora uma rede de Bravais não seja um subespaço vetorial, é fácil ver que ela é fechada para a operação de soma vetorial. Devido a isto, podemos verificar que o arranjo chamado de "rede Honeycomb", embora aparentemente regular, não é uma rede de Bravais, uma vez que não é fechado para a operação de soma de seus elementos.

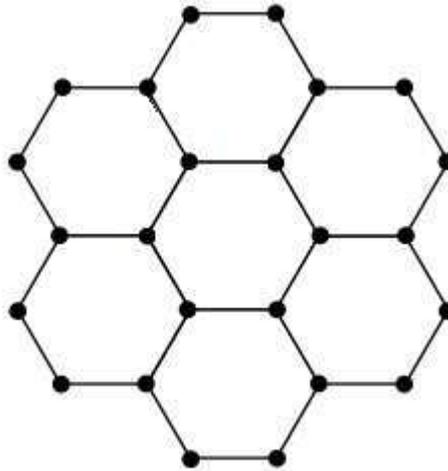


Figura 2.2: quando somamos vetores formados por arestas consecutivas de um desses hexágonos, geramos um ponto no interior do hexágono e, portanto, não pertecente a rede.

Um exemplo importante de rede de Bravais tridimensional é formado pelas chamadas *redes cúbicas*.

- **Rede Cúbica Simple**

Fixado $a \in \mathbb{R}_+^*$ a rede cúbica simples é definida pelos vetores primitivos $\vec{a}_1 = a\vec{e}_1$, $\vec{a}_2 = a\vec{e}_2$ e $\vec{a}_3 = a\vec{e}_3$, onde \vec{e}_1 , \vec{e}_2 e \vec{e}_3 são os vetores da base canônica de \mathbb{R}^3 . Uma porção desta rede de Bravais está ilustrada abaixo. Podemos notar que os pontos da rede estão dispostos nos vértices de cubos

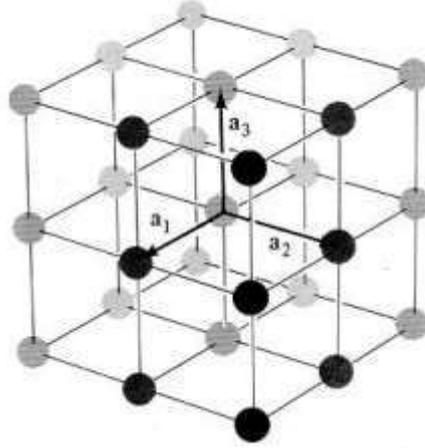


Figura 2.3: Rede cúbica simples - Imagem retirada de [Ashcroft and Mermin, 1976].

de aresta de tamanho a . Esta rede é abreviada por *sc* (simple cubic). O parâmetro a é chamado de *parâmetro de rede*.

- **Rede Cúbica de Face Centrada**

Fixado $a \in \mathbb{R}_+^*$ a rede cúbica de face centrada é definida pelos vetores primitivos $\vec{a}_1 = \frac{a}{2}(\vec{e}_2 + \vec{e}_3)$, $\vec{a}_2 = \frac{a}{2}(\vec{e}_1 + \vec{e}_3)$ e $\vec{a}_3 = \frac{a}{2}(\vec{e}_1 + \vec{e}_2)$, onde \vec{e}_1 , \vec{e}_2 e \vec{e}_3 são os vetores da base canônica de \mathbb{R}^3 . Uma porção desta rede de Bravais está ilustrada abaixo.

Podemos notar que os pontos da rede estão dispostos, além de nos vértices de cubos de aresta de tamanho a com na rede *sc*, nos centros das faces destes cubos. Esta rede é abreviada por *fcc* (face-centered cubic). O parâmetro a também é chamado de *parâmetro de rede*.

- **Rede Cúbica de Corpo Centrado**

Fixado $a \in \mathbb{R}_+^*$ a rede cúbica de corpo centrado é definida pelos vetores primitivos $\vec{a}_1 = \frac{a}{2}(-\vec{e}_1 + \vec{e}_2 + \vec{e}_3)$, $\frac{a}{2}(\vec{e}_1 - \vec{e}_2 + \vec{e}_3)$ e $\frac{a}{2}(\vec{e}_1 + \vec{e}_2 - \vec{e}_3)$, onde \vec{e}_1 , \vec{e}_2 e \vec{e}_3 são os vetores da base canônica de \mathbb{R}^3 . Uma porção desta rede de Bravais está ilustrada abaixo. Podemos notar que os pontos da rede estão

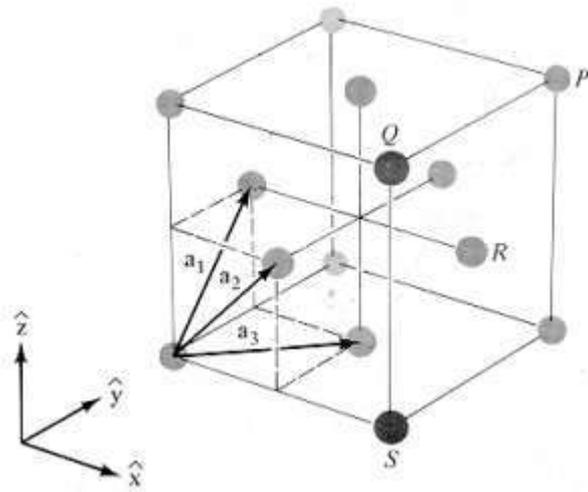


Figura 2.4: Rede cúbica de face centrada - Imagem retirada de [Ashcroft and Mermin, 1976].

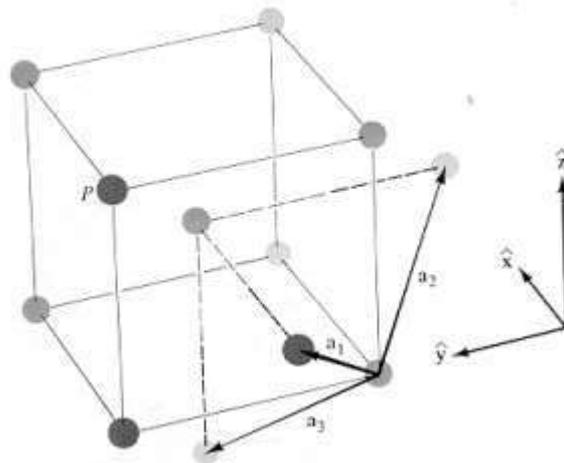


Figura 2.5: Rede cúbica de corpo centrado - Imagem retirada de [Ashcroft and Mermin, 1976].

dispostos, além de nos vértices de cubos de aresta de tamanho a com na rede sc , nos centros destes cubos. Esta rede é abreviada por bcc (body-centered cubic). O parâmetro a também é *chamado de parâmetro de rede*.

Estas três redes de Bravais formam o chamado *sistema cristalino cúbico*.

Uma *estrutura cristalina* é caracterizada por uma rede de Bravais (chamada de *rede de Bravais subjacente*) e um conjunto de posições, geradas a partir dos vetores primitivos da rede, onde são alocados um ou

mais tipos de átomos. São necessários 7 sistemas cristalinos (incluindo o sistema cúbico) para representar todas as estruturas cristalinas conhecidas na natureza (ver[Ashcroft and Mermin, 1976]). O conjunto de posições atômicas é chamado de *base da estrutura cristalina*. Por exemplo, se tomamos uma rede triangular bidimensional e posicionamos um átomo na origem e outro em $\frac{1}{2}(n_1\vec{a}_1 + n_2\vec{a}_2)$, com $n_1, n_2 \in \mathbb{Z}$, geramos a estrutura cristalina abaixo que é a rede honeycomb, que já vimos que não é uma rede de Bravais. Neste

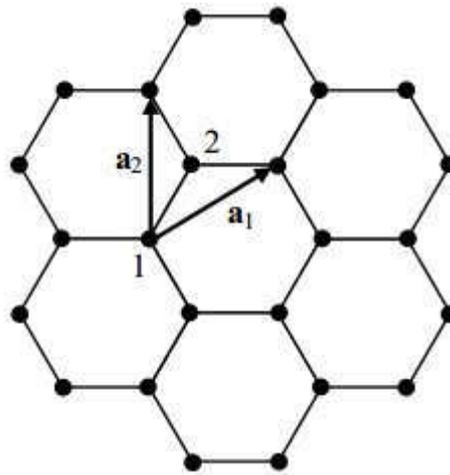


Figura 2.6: Rede triangular como rede de Bravais subjacente e base formada pelo átomo 1 e pelo átomo 2 - Imagem retirada de [Ashcroft and Mermin, 1976].

exemplo, se os dois átomos forem de carbono, geramos a estrutura cristalina do grafite (uma folha de grafite). Já, se o primeiro átomo é o boro e o segundo é o nitrogênio, formamos uma folha de nitreto de boro (NB).

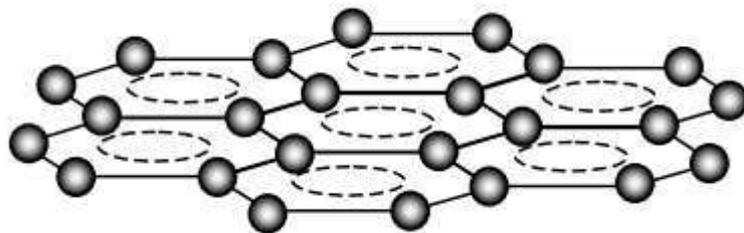


Figura 2.7: uma folha de grafite.

Alguns sólidos cristalinos são formados por um único tipo de átomo e estes átomos ocupam exatamente as posições dos pontos da rede de Bravais subjacente. Outros exemplos importantes são a *estrutura do*

diamante e a *zincblend*. A estrutura do diamante é formada por uma rede de Bravais fcc e uma base formada por dois átomos iguais. O primeiro é posicionado na origem e o segundo a um quarto da diagonal do cubo $\frac{1}{4}(\vec{e}_1 + \vec{e}_2 + \vec{e}_3)$.

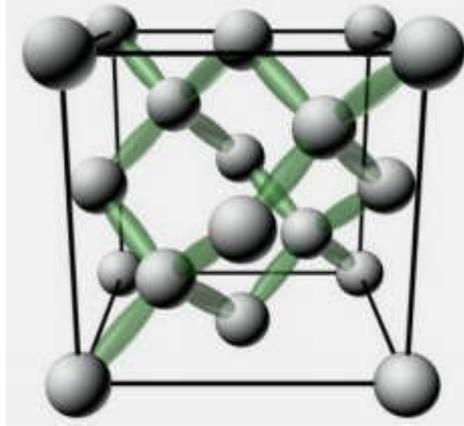


Figura 2.8: estrutura cristalina do diamante.

Sabe-se que os elementos da coluna IV da tabela periódica se cristalizam desta forma. Quando o átomo em questão é o carbono, o parâmetro de rede é aproximadamente 3,57 Å e geramos o diamante comum. Quando o átomo em questão é o silício, sabe-se que o parâmetro de rede é aproximadamente 5,43 Å. Já no caso do átomo ser o germânio, o parâmetro de rede é aproximadamente 5,66 Å.

A única diferença na chamada estrutura zincblend para a estrutura do diamante é que a base é composta por átomos diferentes.

Para fins de implementações computacionais, é interessante observar que a estrutura do diamante pode ser obtida como duas redes fcc interpenetradas.

2.2 O método Tigth-Binding

Embora a determinação do potencial U seja uma tarefa altamente não trivial (que não abordamos aqui), quando estamos trabalhando com uma estrutura cristalina, a periodicidade da rede impõe que o potencial seja periódico

$$U(\vec{x}) = U(\vec{x} + T),$$

onde T é qualquer vetor da rede de Bravais subjacente. Uma aproximação bastante geral para se obter a função de onda para esta situação é, partindo de uma base de funções conhecidas, gerar pequenas perturbações nestas funções e, de alguma forma, combiná-las entre si. Uma primeira simplificação para os tipos de perturbações e misturas que podem ser usadas consiste em impor que as funções resultantes satisfaçam o teorema de Bloch.

Como trata-se de um modelo com potencial estacionário, a discussão aqui se refere às soluções da equação de Schrödinger independente do tempo. No método Tight-Binding, as funções da base são os orbitais atômicos dos átomos que compõem o cristal. Embora esta aproximação tenha se mostrado bastante coerente para as bandas de valência de semicondutores, podem ocorrer discrepâncias na descrição de bandas de condução. O nome "tight binding" foi adotado porque era usado para o cálculo das bandas que descrevem os elétrons do cerne (internos). Hoje é possível obter bons resultados com o método para praticamente todos os tipos de sólidos cristalinos, desde que se utilizem as combinações apropriadas. Trata-se de um modelo monoelétrônico. Isto quer dizer que existe um único elétron em todo o cristal e as várias funções de onda descrevem os vários estados possíveis para este elétron. Uma vez feito o cálculo das bandas de energia, podemos preencher as bandas com o número total de elétrons do cristal.

2.2.1 A proposta LCAO

No final do primeiro capítulo desta dissertação, discutimos brevemente o modelo de Schrödinger para um sólido. Contudo, não estabelecemos qualquer relação da função de onda de um sólido com as funções de onda dos átomos que o compõem. Nesta seção apresentamos a relação proposta no método conhecido como LCAO (Linear Combination of Atomic Orbitals).

Devido ao escopo do nosso trabalho, vamos considerar o caso de um sólido cristalino monoatômico (apenas um átomo na base da rede de Bravais). Estamos limitando o espectro do operador hamiltoniano (para uma única partícula quântica) a M autovalores. Com isso, estamos usando, para cada átomo, M orbitais. Além disso, a nossa estrutura cristalina finita possui N sítios (N átomos) nas posições $T_1, T_2, \dots, T_N \in \mathbb{R}^3$.

Existem muitas formas funcionais que satisfazem o Teorema de Bloch. A forma utilizada no método LCAO é baseada no j -ésimo orbital atômico de cada átomo do cristal, como passamos a descrever.

Para cada orbital φ_j , definimos o j -ésimo orbital de Bloch como sendo a seguinte combinação linear

de orbitais atômicos:

$$\Phi_j(k, \vec{x}) = \frac{1}{\sqrt{N}} \sum_{i=1}^N e^{i\vec{k}\cdot T_i} \varphi_j(\vec{x} - T_i), \quad (j = 1, 2, \dots, M).$$

Note a dependência paramétrica dos orbitais com respeito às posições atômicas T_i (veja a seção 2.2.2). Além disso, a acurácia desta aproximação está bastante ligada com o valor do número M que controla a quantidade de orbitais atômicos que caracterizarão o orbital de Bloch.

Estas funções satisfazem o Teorema de Bloch. Realmente, seja T uma posição da rede de Bravais subjacente. Então

$$\begin{aligned} \Phi_j(k, \vec{x} + T) &= \frac{1}{\sqrt{N}} \sum_{i=1}^N e^{i\vec{k}\cdot T_i} \varphi_j(\vec{x} - T_i + T) = \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^N e^{i\vec{k}\cdot(T_i - T)} e^{i\vec{k}\cdot T} \varphi_j(\vec{x} - T_i + T) = \\ &= e^{i\vec{k}\cdot T} \left[\frac{1}{\sqrt{N}} \sum_{i=1}^N e^{i\vec{k}\cdot(T_i - T)} \varphi_j(\vec{x} - T_i + T) \right] = \\ &\quad e^{i\vec{k}\cdot T} \Phi_j(\vec{x}), \end{aligned}$$

onde, na última igualdade, usamos as condições de fronteira de Born - Von Karman que exigem que

$$\varphi(\vec{r} + T) = \varphi(\vec{r}).$$

2.2.2 O Hamiltoniano Tight-Binding

O método conhecido por *Tight-Binding* é construído com um raciocínio análogo ao método LCAO. Neste método a proposta para autofunções Ψ_j do Hamiltoniano cristalino é construída usando-se os orbitais de Bloch $\{\Phi_{j'}\}$ como base:

$$\Psi_j(\vec{k}, \vec{x}) = \sum_{j'=1}^M C_{jj'}(\vec{k}) \Phi_{j'}(\vec{k}, \vec{x}), \quad (j = 1, 2, \dots, M),$$

onde os coeficientes da combinação linear precisam ser determinados. Repare que, mais uma vez, não é informado o valor do número M que controla a quantidade de orbitais de Bloch que caracterizarão a autofunção do Hamiltoniano cristalino. A quantidade de orbitais de Bloch para se obter uma representação exata da autofunção deve tender ao infinito. Como na prática é usada uma quantidade finita de orbitais,

a acurácia destes cálculos está ligada principalmente a esta quantidade. Por fim, note que esta proposta também satisfaz o Teorema de Bloch. O j -ésimo autovalor $E_j(\vec{k})$ do operador hamiltoniano cristalino, como função de \vec{k} , é dado então por

$$E_j(\vec{k}) = \frac{\langle \Psi_j | \hat{H} | \Psi_j \rangle}{\langle \Psi_j | \Psi_j \rangle}.$$

Usando a expansão por orbitais de Bloch, obtemos

$$E_j(\vec{k}) = \frac{\sum_{j,j'=1}^M C_{jj'}^* C_{jj'} \langle \Phi_j | \hat{H} | \Phi_{j'} \rangle}{\sum_{j,j'=1}^M C_{jj'}^* C_{jj'} \langle \Phi_j | \Phi_{j'} \rangle} := \frac{\sum_{j,j'=1}^M H_{jj'} C_{jj'}^* C_{jj'}}{\sum_{j,j'=1}^M S_{jj'} C_{jj'}^* C_{jj'}},$$

onde $H_{jj'}$ e $S_{jj'}$ são as chamadas matrizes integrais de transferência e overlap, respectivamente. Fixados os M^2 valores das entradas das matrizes de transferência e overlap para um dado \vec{k} , os coeficientes $C_{jj'}^*$ da função de onda do estado fundamental do cristal (estado de menor energia) devem ser otimizados para minimizar $E_j(\vec{k})$:

$$\frac{\partial E_j(\vec{k})}{\partial C_{jj'}^*} = \frac{\sum_{j'=1}^M H_{jj'} C_{jj'} \sum_{j''=1}^M S_{jj''} C_{jj''}^* C_{jj''}}{\left(\sum_{j''=1}^M S_{jj''} C_{jj''}^* C_{jj''} \right)^2} - \frac{\sum_{j''=1}^M H_{jj''} C_{jj''}^* C_{jj''} \sum_{j'=1}^M S_{jj'} C_{jj'}^* C_{jj'}}{\left(\sum_{j'=1}^M S_{jj'} C_{jj'}^* C_{jj'} \right)^2} = 0.$$

Daí, vemos que

$$\sum_{j'=1}^n H_{jj'} C_{jj'} = E_j(\vec{k}) \sum_{j'=1}^M S_{jj'} C_{jj'},$$

para cada j . Definindo o vetor coluna $C_j = [C_{j1} \dots C_{jM}]^t$, vemos que as equações acima podem ser formuladas como

$$HC_j = E_j(\vec{k})SC_j \Leftrightarrow [H - E_j(\vec{k})S]C_j = 0$$

Portanto, para que não tenhamos todos os coeficientes nulos (ou seja, para que tenhamos efetivamente uma função de onda), a matriz $[H - E_j(\vec{k})S]$, para cada j , não pode ser inversível:

$$\det[H - ES] = 0.$$

A equação acima é chamada de *Equação Secular*. Se a base de orbitais de Bloch formar um conjunto ortonormal, a matriz de overlap será a matriz identidade e a Equação Secular se torna o polinômio característico da matriz H ,

$$\det[H - EI] = 0.$$

Resolvendo a Equação Secular, encontramos todos os valores permitidos de energia e , com isso, os coeficientes da combinação linear de orbitais de Bloch proposta podem ser calculados.

Do exposto acima, fica claro que o Tight-Binding é um método de primeiros princípios. A complexidade computacional fica por conta da equação secular que esbarra, quer em condições sobre um determinante, quer sobre a diagonalização matricial. Uma aplicação pedagogicamente interessante da metodologia Tight-Binding aplicada ao átomo de hidrogênio (para o qual conhecemos a solução analítica da equação de Schrödinger!) pode ser vista em [Custodio et al., 2002]. Já a aplicação em duas dimensões para o grafite pode ser vista no segundo capítulo de [Saito et al., 1998]. Porém, existe consequência fundamental da Equação Secular: é possível se obter parametrizações semi empíricas para a determinação dos elementos de matriz do hamiltoniano cristalino H . Portanto, ainda que sem uma definição explícita da base de orbitais de Bloch, podemos resolver esta equação secular diretamente. Ou seja, na metodologia tight-binding as bases para a representação do hamiltoniano têm influência apenas implícita, por meio destas parametrizações do hamiltoniano cristalino. Sendo assim, o Tight-Binding constitui um método semi empírico.

As representações parametrizadas foram sugeridas por Slater e Koster em [Slater and Koster, 1954]. Trata-se de uma técnica de interpolação de resultados de cálculos de estrutura eletrônica provenientes de métodos ab initio. Sem entrar em muitos detalhes técnicos, vamos dar uma ideia desta estratégia. Nos elementos da matriz de transferência

$$H_{mn} = \langle \Phi_m | \hat{H} | \Phi_n \rangle = \frac{1}{N} \sum_{R=1}^N e^{i\vec{k}R} \int \varphi_m^*(\vec{x}) \hat{H} \varphi_n(\vec{x} - T_R) dx dy dz$$

as integrais que aparecem são substituídas por parâmetros que reproduzem os valores de energia. Estes parâmetros em geral são ajustados se utilizando métodos ab initio. Além disso, se admite que a base de orbitais é ortonormal (e, portanto, a matriz de overlap é a matriz identidade). Através do procedimento de Slater e Koster, os parâmetros independem do valor de \vec{k} . Além disso, no tratamento no espaço real por esta metodologia, os elementos de matriz, além dos índices m e n , que representam os orbitais atômicos,

também são indexados fazendo referência aos átomos nos quais estão localizados os orbitais m e n :

$$H_{m_i n_j}.$$

Uma outra questão interessante do Tight - Binding parametrizado é que, como as integrais independem da estrutura cristalina, a informação relacionada com a espécie atômica em questão, contida no ajuste paramétrico, pode ser transferido de um material para outro. No entanto, como estes parâmetros dependem fundamentalmente da distância entre o átomo i e j , devemos ajustar essa distância à do material para o qual estamos transferindo os parâmetros.

Várias parametrizações de hamiltonianos cristalinos para elementos específicos têm sido produzidas. O nosso grupo de pesquisa tem trabalhado com o hamiltoniano de Kwon ([Kwon et al., 1994]) desenvolvido para o Silício.

Neste caso, se usa o conjunto de orbitais $\{s, p_x, p_y$ e $p_z\}$ e, portanto, o hamiltoniano cristalino pode ser representado computacionalmente como uma matriz quadrada de ordem N , onde cada posição é uma matriz quadrada de ordem 4:

$$H_{m,i,n,j}, \quad m, n \in \{s, p_x, p_y, p_z\} \text{ e } i, j \in \{1, 2, \dots, N\}.$$

Os elementos fora da diagonal ($i \neq j$) são chamados de *integrais de hopping*. Kwon construiu uma parametrização onde as integrais de hopping tendem a zero exponencialmente quando a distância entre os átomos tende ao infinito.

Capítulo 3

O Método DMTB

Métodos baseado na matriz densidade para a obtenção da energia do estado fundamental foram propostos simultaneamente em 1993 por Li, Nunes e Vanderbilt [Li et al., 1993] e por Daw [Daw, 1993] (baseados em argumentos e técnicas diferentes), estão diretamente ligados ao tratamento *tight-binding*, onde os estados eletrônicos cristalinos podem ser descritos explicitamente em termos de orbitais atômicos e o hamiltoniano cristalino pode ser construído a partir de parâmetros semi empíricos e, desta forma, dependo apenas implicitamente dos orbitais atômicos. Por isso, este método foi chamado de *método da matriz densidade tight-binding* - DMTB. Devido ao uso do hamiltoniano Tight-Binding, o DMTB se enquadra no grupo dos métodos semiempíricos. O método abordado neste trabalho segue as linhas do grupo do Vanderbilt. No entanto, o modelo proposto se mostra ligeiramente diferente do apresentado por Vanderbilt em [Li et al., 1993], estando de acordo com o modelo reportado por Scuseria em [Millam and Scuseria, 1997]. Esse método tem, como veremos, potencial para ser implementado com complexidade computacional linear em relação número de átomos, o que permite que sejam tratados sistemas com milhares de átomos.

O algoritmo computacional que foi implementado a partir deste trabalho ([Ferreira, 2017]) realiza cálculos baseados em uma simulação de estrutura cristalina por super-células. Esta simulação é a representação de sistemas formados por um número muito grande de átomos através de repetições no espaço de estruturas menores (as chamadas super-células), com apenas N átomos, cada qual com M autoestados eletrônicos ocupados.

Podemos então formular o problema em questão da seguinte forma: dada uma estrutura gerada pela repetição de uma super-célula com N átomos e M orbitais eletrônicos por átomo, obter a energia quântica do estado fundamental desta estrutura.

Embora o modelo que apresentaremos aqui seja absolutamente geral, por se tratar de uma metodologia semi empírica, seu uso é restrito aos elementos para os quais já foram desenvolvidos parâmetros semi empíricos acurados. Na implementação feita em [Ferreira, 2017], são usados átomos de silício apenas, via o hamiltoniano Tight-Binding de Kwon ([Kwon et al., 1994]).

3.1 A Matriz Densidade

A mecânica quântica que descrevemos no capítulo 1 foi baseada no conceito de estados quanto-mecânicos. Existe uma formulação mais geral, usando uma técnica chamada de *Operador Densidade* (ou *Matriz Densidade*). Este é um tema profundo que não poderemos abordar aqui. Recomendamos o capítulo 9 de [Amaral et al., 2011]. Segundo [Amaral et al., 2011] (cap 9 página 122)

"O operador densidade é em geral usado para indicar que nosso conhecimento é incompleto devido às imperfeições na preparação dos estados, ou devido à impossibilidade de conhecimento completo do estado quântico do sistema, o que acontece quando o estado de um sistema composto é emaranhado."

Nos limitaremos a dizer que, naquela abordagem, o conceito de estado quanto-mecânico é generalizado para elementos de um novo espaço vetorial sobre \mathbb{C} , onde os estados, tais como definimos aqui, podem ser vistos como casos particulares, e a matriz densidade ρ é um operador de projeção sobre o subespaço dos estados ocupados de H .

Não vamos entrar em detalhes, mas, formalmente, o operador densidade é definido por

$$\rho = \sum_n f_n |\psi_n\rangle \langle \psi_n|$$

onde n é o índice do espectro discreto do hamiltoniano, f_n é o número de ocupação e é usado o simbolismo introduzido por Dirac, no qual o símbolo $|\ \rangle$, chamado de *ket*, representa um vetor do espaço vetorial dos estados quanto mecânicos. O símbolo $\langle \ |$, chamado de *bra*, representa um vetor do dual desse espaço espaço vetorial. Para essa notação, recomendamos o próprio formulador da notação, [Dirac, 1947], na página 18, e [Jauch, 1968], na página 32. Escrevendo os estados como combinação linear da base finita do método Tight Binding, obtem-se a matriz do operador ρ , chamada de matriz densidade ρ_{ij} . Para uma apresentação do método Tight-Binding usando o formalismo do Operador Densidade, recomendamos [Paxton, 2009]. Na abordagem de Paxton, assim como na nossa, os resultados foram obtidos admitindo-se que a base de

orbitais forma um conjunto ortonormal. O hamiltoniano de Kwon ([Kwon et al., 1994]), com o qual o nosso grupo de pesquisa tem trabalhado, também admite esta hipótese.

Então, a partir de agora, a função de onda não será mais o conceito fundamental para a descrição do sistema, mas sim a densidade eletrônica. Através da matriz densidade ρ , podemos escrever o número de elétrons - N_e - e a energia eletrônica total do sistema E (parte atrativa da energia total do sistema, dada pela soma dos autovalores da matriz do hamiltoniano cristalino Tight-Binding) usando o operador "traço matricial"

$$N_e = tr(\rho) = \sum_i \rho_{ii}; \quad (3.1)$$

$$E = tr(\rho H) = \sum_{i,j} \rho_{ij} H_{ji} \quad (3.2)$$

Além disso, como o operador densidade é um projetor, a matriz densidade é idempotente, ou seja,

$$\rho^2 = \rho.$$

Para demonstrações das igualdades 3.1 e 3.2, a partir do formalismo do Operador Densidade, também indicamos [Paxton, 2009].

Nesse tipo de metodologia (Tigh-Binding + Matriz Densidade) a energia total de um sistema com N_e elétrons é usualmente dada por:

$$E_{tot} = E + E_{rep} + E_0 N, \quad (3.3)$$

onde E_{rep} representa o potencial repulsivo, N é número de átomos do sistema e E_0 é uma constante de energia por átomo.

Podemos agora reformular o problema que estamos atacando: dada uma estrutura gerada pela repetição de uma super-célula com N átomos e M orbitais eletrônicos por átomo, com um matriz hamiltoniana H (quadrada de ordem NM) obter a matriz densidade eletrônica ρ tal que $E = tr(\rho H)$ seja mínima. Então, estamos encarando a energia como função da matriz densidade ρ .

3.2 Espaço de matrizes versus Espaços \mathbb{R}^n

Devido à equação 3.2, estamos considerando a energia como um operador atuando no espaço das matrizes quadradas de ordem NM . A técnica de otimização que usaremos no processo de minimização da energia do sistema, pertencente a uma categoria de métodos conhecida como Gradiente Conjugado não Linear (NLGC), é uma técnica para minimização em espaços \mathbb{R}^n . Um fato básico e fundamental é que todos os espaços vetoriais de uma mesma dimensão finita são linearmente isomorfos. Portanto, podemos identificar, do ponto de vista da álgebra linear, o espaço das matrizes quadradas com o espaço dos vetores linha (ou coluna) de dimensão $(NM)^2$. Um isomorfismo canônico entre estes dois espaços é, como sabemos,

$$X \longleftrightarrow [X_{11}, \dots, X_{1n}, X_{21}, \dots, X_{2n}, \dots, X_{n1}, \dots, X_{nn}]^t.$$

Por exemplo, a matriz quadrada de ordem 3 (3 átomos e um orbital por átomo (hidrogenóides))

$$A = \begin{bmatrix} 1 & -2 & -1 \\ -5 & 1 & 0 \\ 3 & 0 & 2 \\ 1 & 0 & 1 \end{bmatrix}$$

é identificada com o vetor coluna do \mathbb{R}^9

$$\begin{bmatrix} 1 \\ -2 \\ -1 \\ -5 \\ 1 \\ 0 \\ 3 \\ 0 \\ 2 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

No entanto, além de esta operação ter complexidade computacional $(NM)^2$, no NLGC é necessário o conhecimento da derivada da função a ser minimizada. As derivadas não possuem uma identificação direta por este isomorfismo. Por isso, nesta seção, vamos desenvolver o aparato necessário para que possamos aplicar o NLGC no nosso contexto de forma nativa, no sentido de não fazer uso explícito do isomorfismo natural citado. Tendo um processo de vetorização de matrizes em mente, vamos definir a diferencial de

funções reais de variável matricial e obter algumas regras de derivação para estas funções. Como não encontramos este aparato específico na literatura, não apresentamos referências bibliográficas para este tema e só desenvolvemos o mínimo necessário para satisfazer os nossos propósitos, embora muito mais possa ser discutido.

Uma operação fundamental no nosso desenvolvimento, que também não possui uma identificação óbvia por este isomorfismo, é o traço do produto de duas matrizes. Com esta identificação, o traço do produto de duas matrizes corresponde ao produto interno canônico dos vetores produzidos pela vetorização de cada uma das matrizes, como mostra a proposição abaixo.

Proposição 3.1. Seja $n \in \mathbb{N}$. Consideremos o espaço vetorial real das matrizes quadradas de ordem n $M_n(\mathbb{R})$ e o isomorfismo linear vec entre $M_n(\mathbb{R})$ e \mathbb{R}^{n^2} dado por

$$A = (a_{uv})_{n \times n} \in M_n(\mathbb{R}) \mapsto vec(A) = [a_{11}, \dots, a_{1n}, a_{21}, \dots, a_{2n}, \dots, a_{n1}, \dots, a_{nn}]^t \in \mathbb{R}^{n^2}.$$

Então,

$$tr(AB) = vec(A) \cdot vec(B^t),$$

onde " \cdot " representa o produto interno canônico de \mathbb{R}^{n^2} , quaisquer que sejam $A, B \in M_n(\mathbb{R})$.

Demonstração: De fato, denotando por $C = AB$, pela definição de produto matricial, temos que

$$(AB)_{uv} = \sum_{k=1}^n a_{uk} b_{kv}, \quad 1 \leq u, v \leq n.$$

Daí,

$$tr(AB) = \sum_{i=1}^n (AB)_{ii} = \sum_{i=1}^n \sum_{k=1}^n A_{ik} B_{ki} = vec(A) \cdot vec(B^t). \quad \blacksquare$$

Como veremos, esta identificação será importante para a aplicação do NLCG. Vale notar que, embora seja um fato bastante conhecido que $tr(AB^t)$ define um produto interno no espaço das matrizes quadradas, a proposição anterior estabelece mais do que isso.

A vetorização de matrizes permite que consideremos o gradiente de uma função real de variável matricial, já que o isomorfismo em questão identifica esta função com um campo escalar. Sendo $f(X)$ uma

função real de variável matricial, definimos a derivada de f como sendo a seguinte matriz

$$\frac{\partial f}{\partial X} = \begin{bmatrix} \frac{\partial f}{\partial X_{11}} & \frac{\partial f}{\partial X_{12}} & \cdots & \frac{\partial f}{\partial X_{1n}} \\ \frac{\partial f}{\partial X_{21}} & \frac{\partial f}{\partial X_{22}} & \cdots & \frac{\partial f}{\partial X_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial X_{n1}} & \frac{\partial f}{\partial X_{n2}} & \cdots & \frac{\partial f}{\partial X_{nn}} \end{bmatrix}.$$

O que é muito importante para o desenvolvimento que estamos fazendo aqui é que, a derivada definida acima tem como vetorização justamente o gradiente do campo escalar f quando X é posta na forma vetorizada que nos fixamos. Por isso, sempre que estiver claro que a variável é matricial e que a vetorização escolhida é a que definimos aqui, escreveremos ∇f ao invés de $\frac{\partial f}{\partial X}$. Com isso, as propriedades básicas do operador gradiente são todas válidas para este novo operador. No nosso trabalho, é particularmente importante calcularmos o gradiente do traço da potência de matrizes. A proposição abaixo é suficiente para as nossas necessidades.

Proposição 3.2. Seja $n \in \mathbb{N}$. Consideremos o espaço vetorial real das matrizes quadradas de ordem n $M_n(\mathbb{R})$ e fixemos uma matriz $A \in M_n(\mathbb{R})$. Então valem as seguintes regras de derivação:

1. Sendo a função real de variável matricial f dada por $f(X) = \text{tr}(AX) = \text{tr}(XA)$,

$$\frac{\partial f}{\partial X} = A^t;$$

2. Sendo a função real de variável matricial f dada por $f(X) = \text{tr}(AX^2) = \text{tr}(X^2A)$,

$$\frac{\partial f}{\partial X} = (AX)^t + (XA)^t;$$

3. Sendo a função real de variável matricial f dada por $f(X) = \text{tr}(AX^3) = \text{tr}(X^3A)$,

$$\frac{\partial f}{\partial X} = (AX^2)^t + (XAX)^t + (X^2A)^t;$$

Demonstração:

1. Pela definição de traço e de produto matricial, temos que

$$f(X) = \text{tr}(AX) = \sum_{i=1}^n (AX)_{ii} = \sum_{i=1}^n \sum_{k=1}^n A_{ik} X_{ki}.$$

Dáí, fixados r e s ,

$$\frac{\partial f}{\partial X_{rs}} = A_{sr}.$$

Portanto,

$$\frac{\partial f}{\partial X} = A^t.$$

2. Neste caso,

$$f(X) = \text{tr}(AX^2) = \sum_{i=1}^n \sum_{k=1}^n A_{ik}(X^2)_{ki} = \sum_{i=1}^n \sum_{k=1}^n A_{ik} \sum_{j=1}^n X_{kj}X_{ji} = \sum_{i=1}^n \sum_{k=1}^n \sum_{j=1}^n A_{ik}X_{kj}X_{ji}.$$

Com isso, para o cálculo de $\frac{\partial f}{\partial X_{rs}}$, contribuem da expressão acima apenas as somas em que $k = r$ e $j = s$, e $j = r$ e $i = s$. Portanto, fixados r e s ,

$$\frac{\partial f}{\partial X_{rs}} = \sum_{i=1}^n A_{ir}X_{si} + \sum_{k=1}^n A_{sk}X_{kr} = (XA)_{sr} + (AX)_{sr},$$

mostrando que

$$\frac{\partial f}{\partial X} = (AX)^t + (XA)^t.$$

3. Basta escrever,

$$f(X) = \text{tr}(AX^3) \tag{3.4}$$

$$= \text{tr}((AX)X^2) \tag{3.5}$$

$$= \sum_{i=1}^n \sum_{k=1}^n (AX)_{ik}(X^2)_{ki} \tag{3.6}$$

$$= \sum_{i=1}^n \sum_{k=1}^n \left(\sum_{l=1}^n A_{il}X_{lk} \right) \left(\sum_{j=1}^n X_{kj}X_{ji} \right) \tag{3.7}$$

$$= \sum_{i=1}^n \sum_{k=1}^n \sum_{j=1}^n \sum_{l=1}^n A_{il}X_{lk}X_{kj}X_{ji}. \tag{3.8}$$

Agora notemos que, fixados r e s , para o cálculo de $\frac{\partial f}{\partial X_{rs}}$, contribuem da expressão acima apenas as

somas em que $l = r$ e $k = s$, $k = r$ e $j = s$, e $j = r$ e $i = s$. Portanto,

$$\frac{\partial f}{\partial X_{rs}} = \sum_{i=1}^n \sum_{j=1}^n A_{ir} X_{sj} X_{ji} + \sum_{i=1}^n \sum_{l=1}^n A_{il} X_{lr} X_{si} + \sum_{k=1}^n \sum_{l=1}^n A_{sl} X_{lk} X_{kr} \quad (3.9)$$

$$= \sum_{i=1}^n A_{ir} \left(\sum_{j=1}^n X_{sj} X_{ji} \right) + \sum_{i=1}^n X_{si} \left(\sum_{l=1}^n A_{il} X_{lr} \right) + \sum_{k=1}^n X_{kr} \left(\sum_{l=1}^n A_{sl} X_{lk} \right) \quad (3.10)$$

$$= \sum_{i=1}^n A_{ir} (X^2)_{si} + \sum_{i=1}^n X_{si} (AX)_{ir} + \sum_{k=1}^n X_{kr} (AX)_{sk} \quad (3.11)$$

$$= (X^2 A)_{sr} + (XAX)_{sr} + (AX^2)_{rs}, \quad (3.12)$$

revelando que

$$\frac{\partial f}{\partial X} = (AX^2)^t + (XAX)^t + (X^2 A)^t. \quad \blacksquare$$

Antes de encerrarmos esta seção, deve ficar claro que os resultados da proposição anterior nada tem a ver com qualquer processo de vetorização que se escolha. Na verdade, sendo $f(X)$ uma função de domínio matricial, a matriz $\frac{\partial f}{\partial X}$ nada tem a ver com processos de vetorização. Agora, a identificação

$$\frac{\partial f}{\partial X} \longleftrightarrow \nabla f$$

só é correta para a vetorização que definimos aqui. Como geramos fórmulas matriciais diretas para as derivações, o uso do processo de vetorização na aplicação do NLCG não é explícito. Na verdade, sempre que forem necessárias operações com gradientes de campos escalares, o processo é encapsulado através do uso das fórmulas de derivação que obtivemos. Além disso, devemos ter em mente que, embora estejamos usando matrizes ainda, elas representam vetores e, portanto, devemos ter cuidado com as interpretações dos resultados finais.

3.3 O modelo de minimização

Como o nosso objetivo é obter a energia do estado fundamental, o foco é a minimização da energia eletrônica, dada como função da matriz densidade (equação (3.2)). Porém, temos alguns vínculos, devidos à mecânica quântica que usamos até aqui, que devem ser mantidos. A minimização não pode ser feita irrestritamente no espaço de matrizes, já que o número de elétrons presentes na nossa estrutura precisa ser reproduzido pela matriz densidade como na equação (3.1). Além disso, a matriz densidade deve ser

idempotente. Do que foi discutido na seção anterior, estamos operando com a energia como sendo um campo escalar definido em $\mathbb{R}^{(NM)^2}$. Por fim, a matriz densidade eletrônica deve comutar com a matriz do hamiltoniano Tight-Binding.

3.3.1 O vínculo do número de elétrons

Para o vínculo do número de elétrons, podemos usar o Teorema dos Multiplicadores de Lagrange (para uma demonstração, veja [Lima, 2005] nas páginas 171 e 172):

Teorema 3.3. (Teorema do Multiplicador de Lagrange) Seja $f : U \rightarrow \mathbb{R}$ uma função de classe C^k ($k \geq 1$) no conjunto aberto $U \subset \mathbb{R}^{n+1}$, e $M = g^{-1}(0)$ uma hipersuperfície contida em U , imagem inversa do valor regular 0 por uma função $g : U \rightarrow \mathbb{R}$, também de classe C^k . Um ponto $p \in M$ é um ponto crítico de $f|_M$ se, e somente se, existe um número real μ (chamado de multiplicador de Lagrange) tal que

$$\nabla f(p) + \mu \nabla g(p) = 0.$$

Considerando o nosso processo de vetorização, o teorema anterior nos garante o seguinte: ρ é um ponto crítico da função $E(A) = \text{tr}(AH)$, restrita a hipersuperfície $g^{-1}(0) = \{A \in M_n(\mathbb{R}); \text{tr}(A) - N_e = 0\}$, se, e somente se, existe um número real μ que torna nulo, em ρ , o gradiente da função

$$\Omega(A) = \text{tr}(AH) + \mu(\text{tr}(A) - N_e),$$

desde que o gradiente do vínculo g não seja nulo em ρ .

Notemos que, devido a vetorização que estamos fazendo, o gradiente de g pode ser obtido facilmente usando-se a proposição (3.2):

$$\nabla(g(A)) = \nabla((\text{tr}(A) - N_e)) = I^t = I.$$

Portanto, o uso dos multiplicadores de Lagrange é permitido.

Aqui vale observar que o grupo de Vanderbilt originalmente concebe o escalar μ como um potencial químico em [Li et al., 1993]. Com isso, é necessário algum procedimento para atualizar o valor do potencial

químico durante a busca pela matriz densidade. Diferentemente disso, na nossa abordagem, a interpretação de μ como um multiplicador de Lagrange, que tem o papel de selecionar matrizes que reproduzem o número correto de elétrons, como veremos, permite usar a combinação da implementação computacional baseada no método dos gradientes conjugados não linear, que foi feita em [Ferreira, 2017], com exigência da consistência física para o número de elétrons, para determinar explicitamente o multiplicador μ como função de ρ em cada passo de iteração. Portanto, no modelo acima, μ não será uma incógnita independente. Esta mesma interpretação para μ aparece também em [Millam and Scuseria, 1997].

3.3.2 O vínculo da idempotência

A idempotência da matriz ρ também precisa ser vinculada. Mesmo que iniciemos o processo de minimização com uma matriz idempotente, os experimentos computacionais feitos em [Ferreira, 2017] mostram que a idempotência não é preservada e o algoritmo de minimização perde a convergência, gerando as "runaway solutions" comentadas por Scuseria em [Daniels and Scuseria, 1999]. Seguindo as recomendações do grupo do Vanderbilt e de Scuseria, vamos impor o vínculo da idempotência, de forma aproximada e implícita, através da chamadas "transformações de purificação", apresentadas em [McWeeny, 1960], que passamos a analisar.

Como sabemos da Álgebra Linear elementar, se λ é um autovalor de uma matriz quadrada A , então, sendo $k \in \mathbb{N}$, λ^k é um autovalor da matriz A^k . Por isso, os autovalores de uma matriz idempotente ou valem 0 ou valem 1. Em [Li et al., 1993] é enfatizado que esta última condição é que deve ser preservada.

Considere a aplicação

$$F(X) = 3X^2 - 2X^3,$$

definida no espaço das matrizes quadradas. É evidente que toda matriz idempotente é um ponto fixo desta aplicação. Além disso, a proposição abaixo revela uma relação interessante entre esta aplicação o polinômio $f(x) = 3x^2 - 2x^3$.

Proposição 3.4. Seja $n \in \mathbb{N}$. Consideremos o espaço vetorial real das matrizes quadradas de ordem n $M_n(\mathbb{R})$ e fixemos uma matriz $X \in M_n(\mathbb{R})$. Sendo λ um autovalor de X , $f(\lambda)$ é autovalor de $F(X) = 3X^2 - 2X^3$, onde f é o polinômio

$$f(x) = 3x^2 - 2x^3.$$

Demonstração: Seja λ um autovalor de X . Então, λ^2 e λ^3 são autovalores de X^2 e X^3 , respectivamente. Tomemos um autovetor \vec{v} de X , associado a λ . Então

$$F(X)\vec{v} = 3X^2\vec{v} - 2X^3\vec{v} = 3\lambda^2\vec{v} - 2\lambda^3\vec{v} = f(\lambda)\vec{v},$$

mostrando que $f(\lambda)$ é um autovalor de $F(X)$, como queríamos. ■

Seja B uma matriz "aproximadamente idempotente", no sentido de que, para um certo $\epsilon > 0$, seus autovalores são dados por h ou $1 + h$, com $h \in \mathbb{R}$ e $0 < |h| < \epsilon$. Com isso, os autovalores de $F(B)$ são $f(h)$ ou $f(1 + h)$. Pelo desenvolvimento de Taylor para f , qualquer que seja $x \in \mathbb{R}$, temos que

$$f(x + h) = f(x) + f'(x)h + \frac{1}{2}f''(\xi_x)h^2 = f(x) + (6x - 6x^2)h + \frac{1}{2}(6 - 12\xi)h^2,$$

para algum ξ_x entre x e $x + h$. Tomando $x = 0$, obtemos que

$$|f(h)| = \left| \frac{1}{2}(6 - 12\xi_0)h^2 \right| \leq |3(1 - 2\xi_0)\epsilon^2|.$$

E tomando $x = 1$, obtemos

$$|f(1 + h) - 1| = \left| \frac{1}{2}(6 - 12\xi_1)h^2 \right| \leq |3(1 - 2\xi_1)\epsilon^2|.$$

Isto nos mostra que, sendo B uma matriz aproximadamente idempotente, cujo o erro máximo absoluto da aproximação entre seus autovalores e os valores desejados, 0 e 1, seja ϵ , a matriz $F(B)$ será aproximadamente idempotente com erro máximo absoluto da ordem de $O(\epsilon^2)$. Portanto, usar $F(\rho)$ no lugar de ρ na equação (3.2) se apresenta como uma proposta de introdução implícita e aproximada do vínculo da idempotência no nosso modelo.

A aplicação F é conhecida como Transformação de Purificação de MacWeeny ([McWeeny, 1960]) e foi incorporada no modelo do grupo de Vanderbilt. Analisando o polinômio de purificação f , vemos que, estando os autovalores de uma matriz no intervalo $[-0.5; 1.5]$, os autovalores da matriz purificada permanecerão no intervalo $[0; 1]$.

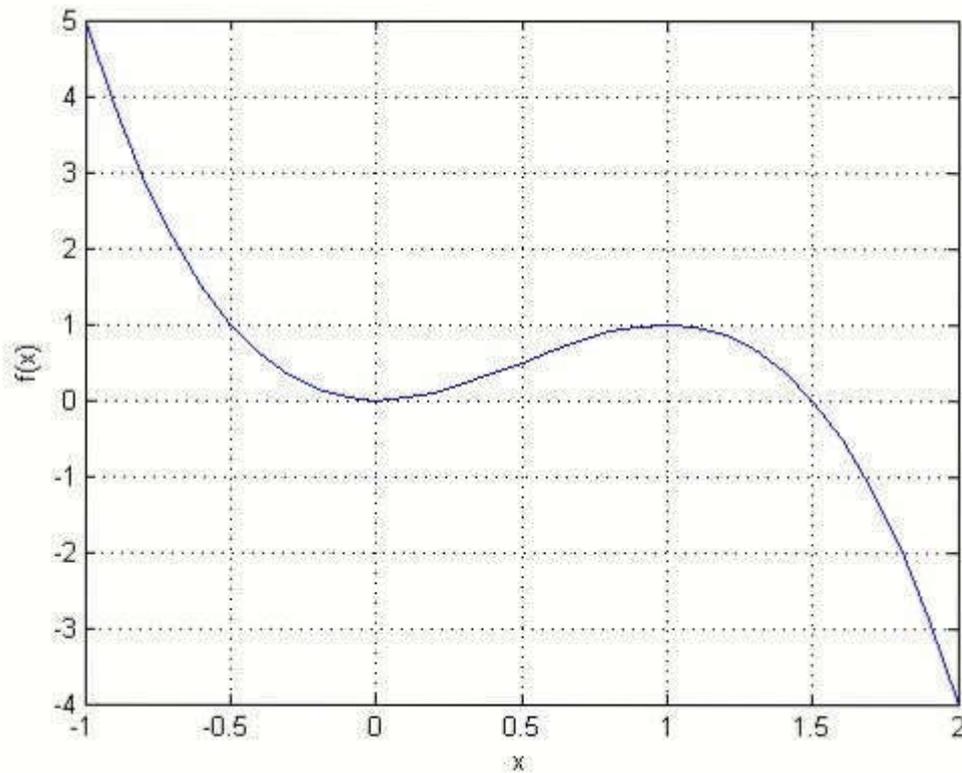


Figura 3.1: o polinômio purificador $f(x) = 3x^2 - 2x^3$

3.4 O modelo matemático final

Repitamos aqui a formulação do nosso problema:

Consideremos uma estrutura dada pela replicação de uma super-célula formada por N átomos, cada qual com M autoestados eletrônicos preenchidos. Admitamos que para este sistema quanto-mecânico exista um hamiltoniano Tigth-Binding semi-empírico H construído. Nestas condições, obter a energia quanto-mecânica do estado fundamental desta estrutura.

Como dissemos, e mostraremos na próxima subseção, o multiplicador de Lagrange não será uma incógnita independente no nosso problema. μ estará completamente determinado em função da matriz densidade, como consequência das considerações físicas do nosso sistema e do método computacional es-

colhido para a resolução do problema. Devido a isso, vamos fazer um uso um tanto incomum do método dos multiplicadores de Lagrange: a proposta de solução deste problema é dada pela minimização irrestrita do funcional abaixo

$$\Omega(A) = \text{tr}((3A^2 - 2A^3)H) + \mu(\text{tr}(A) - N_e). \quad (3.13)$$

Este funcional está definido no espaço das matrizes reais quadradas de ordem NM e, devido ao processo de vetorização de matrizes que estamos adotando, está sendo considerado como um campo escalar definido em $\mathbb{R}^{(NM)^2}$. O mínimo deste funcional obviamente tornará o seu gradiente nulo e, como o teorema dos multiplicadores de Lagrange é um teorema do tipo "se, e somente se," este mínimo é solução do nosso problema. Então, um método computacional de minimização irrestrita será utilizado, ao invés de se resolver um sistema não linear fornecido pelo método dos multiplicadores de Lagrange. Insistimos, mais uma vez, que a definição do multiplicador e o próprio funcional Ω , são, como veremos na próxima seção, completamente dependente da estratégia de implementação computacional escolhida. Portanto, o DMTB discutido aqui se trata de um modelo *matemático-computacional* com toda a força da palavra! Em princípio esta metodologia poderia ter grande risco: os mínimos locais do funcional Ω , pois estes também tornam o gradiente nulo. Em [Ferreira, 2017] são discutidas algumas estratégias computacionais para se evitar mínimos locais.

Em geral, estratégias de minimização necessitam, pelo menos, do gradiente do funcional em questão. O gradiente do funcional de energia que construímos pode ser facilmente obtido com o uso da proposição 3.2. De fato,

$$\begin{aligned} \frac{\partial \Omega}{\partial A} = \nabla \Omega &= \frac{\partial \text{tr}((3A^2 - 2A^3)H) + \mu(\text{tr}(A) - N_e)}{\partial A} \\ &= 3 \frac{\partial \text{tr}((A^2)H)}{\partial A} - 2 \frac{\partial \text{tr}((A^3)H)}{\partial A} + \mu \frac{\partial \text{tr}(A)}{\partial A} \\ &= 3((HA)^t + (AH)^t) - 2((HA^2)^t - (AHA)^t - (A^2H)^t) \\ &= 3A^t H^t + 3H^t A^t - 2(A^2)^t H^t - 2A^t H^t A^t - 2H^t (A^2)^t + \mu I, \end{aligned}$$

onde I representa a matriz identidade de ordem NM . Então temos que

$$\nabla \Omega(A) = 3(HA + AH)^t - 2(HA^2 + AHA + A^2H)^t + \mu I. \quad (3.14)$$

Restou apenas a questão da comutatividade do produto ρH , sendo ρ a verdadeira matriz densidade eletrônica, que esclareceremos nas próximas seções.

3.4.1 O modelo matemático-computacional final

O modelo matemático que construímos para o DMTB tem, além da matriz densidade, um multiplicador de Lagrange como incógnita. Se o multiplicador μ é determinado, o problema pode ser solucionado minimizando-se o funcional Ω (equação (3.13)). Na implementação computacional apresentada em [Ferreira, 2017], que tem a construção baseada na presente formulação, a minimização do funcional Ω , é feita através de um algoritmo baseado no Gradiente Conjugado não Linear (NLCG). O algoritmo do gradiente conjugado linear (CG) é um método iterativo standard para a solução de sistemas lineares (de Cramer), baseado em um processo de minimização para formas quadráticas definidas em \mathbb{R}^n . O CG é muito bem estabelecido, tanto do ponto de vista matemático, quanto computacional, e o processo de minimização é realizado em n passos, como pode ser visto nas referências [Shewchuk, 1994], [Nocedal and Wright, 2006], [Brufati, 2011] e [Ferreira, 2017]. Existem generalizações do CG para funções não quadráticas, agora já não tendo nenhuma ligação com a resolução de sistemas lineares. Estas generalizações não são equivalentes e pertencem a uma categoria de métodos de minimização irrestrita de campos escalares definidos em \mathbb{R}^n conhecida como Gradiente Conjugado não Linear (NLCG). No entanto, embora traga bons resultados com possibilidade de baixa complexidade computacional, o caso não linear ainda é alvo de investigações científicas (veja, por exemplo, [Sun and Liu, 2015], [Armand, 2007], [Yu et al., 2009] e [Zhang et al., 2012]).

Sem entrar em muitos detalhes, um método de NLCG resolve o problema de minimização irrestrita de uma função diferenciável $f: \mathbb{R}^p \rightarrow \mathbb{R}$ construindo uma sequência de vetores pela seguinte fórmula de recorrência

$$\vec{x}_{k+1} = \vec{x}_k + \alpha_k \vec{d}_k, \quad (3.15)$$

onde o escalar α_k é chamado de *comprimento do passo k* , e o vetor \vec{d}_k é chamado de *direção de busca do passo k* , definido por

$$\vec{d}_k = \begin{cases} -\nabla f(\vec{x}_k), & k = 1 \\ -\nabla f(\vec{x}_k) + \beta_k \vec{d}_{k-1}, & k \geq 2. \end{cases} \quad (3.16)$$

Existem várias propostas (não equivalentes) para o escalar β_k reportadas na literatura especializada em Otimização Numérica.

Métodos baseados no NLCG tem sido aplicados em modelagem matemática e computacional de fenômenos da física da matéria condensada desde, pelo menos, a década de 80 do século passado ([Teter et al., 1989]).

Abaixo temos o pseudocódigo da rotina fundamental de um método NLCG, do tipo Polak-Ribière, para a minimização irrestrita de um campo escalar f

Algoritmo 1: GRADIENTE CONJUGADO NÃO LINEAR (POLAK-RIBIÈRE-TYPE)

Entrada: $f, \nabla f, \vec{x}_0, tol, it_{max}$ {campo escalar, gradiente do campo, estimativa inicial e número máximo de iterações }

Saída: vetor \vec{x} que minimiza o campo escalar

```

1 início
2    $it \leftarrow 0;$ 
3    $\vec{d}_0 \leftarrow -\nabla f(\vec{x}_0);$ 
4    $\vec{x} \leftarrow \vec{x}_0;$ 
5    $\vec{d} \leftarrow \vec{d}_0;$ 
6    $G \leftarrow \vec{d}_0;$ 
7   enquanto  $\|G\| > tol$  e  $it < it_{max}$  faça
8      $G_0 \leftarrow \nabla f(\vec{x}_0);$ 
9     Determine  $\alpha;$ 
10     $\vec{x} \leftarrow \vec{x} + \alpha \vec{d};$ 
11     $G \leftarrow \nabla f(\vec{x});$ 
12     $\beta \leftarrow \frac{\langle G, (G - G_0) \rangle}{\langle G, G \rangle};$ 
13     $d \leftarrow -\nabla f(\vec{x}) + \beta d;$ 
14     $\vec{x}_0 \leftarrow \vec{x};$ 
15     $it \leftarrow it + 1;$ 
16  fim
17 fim
18 retorna  $\vec{x}$ 

```

Como podemos ver, é um método iterativo que produz o ponto \vec{x} que minimiza o campo escalar f . Na linha 10, o ponto é atualizado para um ponto da reta que contém \vec{x} , com direção \vec{d} . Esta atualização precisa ser para um ponto onde o campo escalar f assumo um valor inferior a $f(\vec{x})$. A garantia desta estimativa é dada pela escolha do passo α , determinado na linha 9 através de buscas unidirecionais, geralmente inexatas. Existem várias propostas para a escolha do passo no NLCG reportadas na literatura e este assunto ainda é alvo de discussões científicas. As buscas lineares têm influência em teoremas de convergência global para os algoritmos NLCG. O estudo convergência global para o NLCG ainda recebe muitas contribuições atuais e tem como objetivo determinar hipóteses suficientes para garantir que

$$\lim_{k \rightarrow \infty} \|\nabla f(\vec{x}_k) = 0\|.$$

É um estudo β -dependente. Além disso, na linha 13, o parâmetro β é fundamental para a atualização da direção \vec{d} . Este parâmetro, tal qual como é calculado na linha 12, é devido a Polak e Ribière

([Polak and Ribiere, 1969]). Como dissemos, existem outras propostas para β , tais como Fletcher e Reeves ([Fletcher and Reeves, 1964]). Estas escolhas, quando aplicadas ao caso CG, são equivalentes. No entanto, não há garantia geral de equivalência para o caso não linear. Diante disso, vemos que não basta informarmos que estamos usando NLCG se queremos que o leitor possa reproduzir nossos resultados.

A adaptação para o nosso funcional de energia, Ω , é feita via vetorização de matrizes que apresentamos na seção 4.2. Os produtos escalares são substituídos pelo traço do produto de matrizes, devido à proposição 3.1. Já o gradiente do campo f é substituído pela equação 3.14, necessitando, portanto, de um valor para o multiplicador de Lagrange μ em cada iteração, que obteremos abaixo fazendo uma imposição devido a considerações físicas do nosso processo. Já o vetor \vec{x} procurado é a vetorização da matriz densidade ρ .

Admitindo uma implementação para o NLCG contendo n passos, levando em consideração o nosso processo de vetorização, a versão adaptada para o DMTB faria uso de n^2 passos, uma vez que, para o cálculo da energia do sistema, em princípio seria necessário computar todos os elementos da matriz densidade. No entanto, sabe-se que

$$\lim_{R_{ij} \rightarrow \infty} \rho_{ij} = 0,$$

onde R_{ij} a distância entre os orbitais i e j ([Li et al., 1993]). Esta convergência para zero tem ordem de convergência polinomial em metais e exponencial em isolantes ([Li et al., 1993]). Diante deste fato, o grupo de Vanderbilt propôs a introdução de um raio de corte $R_c > 0$ de modo que se $R_{ij} > R_c$ então $\rho_{ij} = 0$. Sendo assim, fixado um átomo na rede, seja L o número de átomos da rede contidos na esfera de raio R_c , centrada no átomo que foi previamente fixado (devido à simetria da rede, L não depende do átomo fixado, sendo então uma constante para cada escolha do raio de corte). Com isso, para formarmos um elemento ρ_{ij} eventualmente não nulo, i pode ocupar qualquer uma das NM posições disponíveis, porém j deve variar dentre LM posições. Daí, se consideramos apenas estes elementos eventualmente não nulos da matriz densidade, o nosso processo de vetorização produz um vetor com NLM^2 coordenadas. Tendo este vetor como entrada, o algoritmo CG faria uso de NLM^2 passos, apresentando, portanto, complexidade $O(N)$. É claro que R_c precisa ser escolhido de forma a obter matrizes de densidade próximas das densidades verdadeiras. Na implementação feita em [Ferreira, 2017] usou-se $M = 4$ (um orbital s e três orbitais p).

Abaixo segue o pseudocódigo, baseado na metodologia do NLCG, para a minimização do funcional $\Omega(\rho)$, usado em [Ferreira, 2017].

Algoritmo 2: NLCG PARA O FUNCIONAL DE ENERGIA (POLAK-RIBIÈRE-TYPE)

Entrada: $\Omega, \nabla\Omega, \rho_0, tol, it_{max}$ {funcional de energia, gradiente do funcional, estimativa inicial para a matriz densidade e número máximo de iterações }

Saída: matriz densidade eletrônica ρ para o estado fundamental

```
1 início
2    $it \leftarrow 0;$ 
3    $\vec{d}_0 \leftarrow -\nabla\Omega(x_0);$ 
4    $\rho \leftarrow \rho_0;$ 
5    $\vec{d} \leftarrow \vec{d}_0;$ 
6    $G \leftarrow \vec{d}_0;$ 
7   enquanto  $\|G\| > tol$  e  $it < it_{max}$  faça
8      $G_0 \leftarrow \nabla\Omega(\rho_0);$ 
9     Determine  $\alpha$  usando uma busca unidirecional
10     $\rho \leftarrow \rho + \alpha\vec{d};$ 
11    Determine o Multiplicador de Lagrange  $\mu;$ 
12     $G \leftarrow \nabla\Omega(\rho);$ 
13     $\beta \leftarrow \frac{tr(G(G-G_0)^t)}{tr(G_0G_0^t)};$ 
14     $d \leftarrow -G + \beta\vec{d};$ 
15     $\rho_0 \leftarrow \rho;$ 
16     $it \leftarrow it + 1;$ 
17  fim
18 fim
19 retorna  $\rho$ 
```

Trata-se portanto de um processo iterativo que constrói a matriz densidade eletrônica do estado fundamental.

Mais uma vez, chamamos a atenção para o fato de que o retorno deste algoritmo é uma matriz ρ (a matriz densidade eletrônica do estado fundamental) que está representando um vetor. Portanto, o vetor final dado pelo NLCG é a vetorização (definida na proposição (3.2)) desta matriz. Como o nosso objetivo final é o valor da energia eletrônica, e não a matriz densidade, devido a fórmula (3.2), mais uma vez o processo de vetorização não se manifesta explicitamente.

Analisando o polinômio purificador (veja a figura3.1), vemos que a purificação não é garantida se a matriz possui autovalores fora do intervalo $[-0.5; 1.5]$. Com isso, devemos tomar uma matriz densidade inicial tendo autovalores neste intervalo.

Resta apenas a determinação do multiplicador de Lagrange. Considerando-se que o número de elétrons, $N_e = \text{tr}(\rho)$, deve ser mantido constante ao longo do processo, as matrizes construídas na linha 10 do código acima devem ter traço constante (igual ao traço da matriz tentativa inicial). Observando-se a atribuição realizada na linha 10, na iteração $k + 1$,

$$\rho_{k+1} \leftarrow \rho_k + \alpha \vec{d}_k;$$

vemos que, devido a linearidade do traço matricial e o fato do passo α ser não nulo, N_e será mantido se, e somente se, o traço de \vec{d} for nulo em todas as iterações do laço *while*.

Devido relação de recorrência que define \vec{d} (equação (3.16)), devemos impor que

$$\text{tr}(\nabla\Omega) = \text{tr}(\nabla\Omega^t) = 0.$$

Para isso, usamos a equação (3.14), a linearidade do traço matricial e o fato de que o traço do produto de duas matrizes não depende da ordem em que o produto é realizado:

$$\begin{aligned} & \text{tr}(3AH + 3HA - 2A^2H - 2AHA - 2HA^2 + \mu I) \\ &= 3\text{tr}(AH) + 3\text{tr}(HA) - 2\text{tr}(A^2H) - 2\text{tr}(AHA) - 2\text{tr}(HA^2) + N_e\mu \\ &= -6\text{tr}(A^2H) + 6\text{tr}(AH) + N_e\mu = 0. \end{aligned}$$

Portanto, o multiplicador de Lagrange é dado por

$$\mu = \frac{6}{N_e} [\text{tr}(A^2H) - \text{tr}(AH)]. \quad (3.17)$$

Notemos que, quando A é uma matriz idempotente, $\mu = 0$. Isto é coerente pois, neste caso, não existe o vínculo da idempotencia.

A implementação feita em [Ferreira, 2017] faz uso da definição explícita do multiplicador μ dada na equação (3.17). Esta definição é aplicada repetidamente em cada passo do processo iterativo.

Com essa justificativa física para a determinação do multiplicador μ , é fundamental que a matriz inicial ρ_0 já possua o número correto de elétrons da super célula. Desta forma, no algoritmo 2, a expressão

$$\mu(\text{tr}(A) - N_e)$$

que já era nula em ρ_0 , se mantém nula durante todo o processo. Portanto a expressão do funcional Ω dada na equação (3.14) se torna

$$\Omega(A) = \text{tr}((3A^2 - 2A^3)H). \quad (3.18)$$

Ou seja, com esta abordagem, o multiplicador de Lagrange não surte efeito algum sobre o funcional objetivo Ω . Seu efeito, no entanto, se manifesta no gradiente $\nabla\Omega$. Portanto, o problema final é a minimização irrestrita de

$$\Omega(A) = \text{tr}((3A^2 - 2A^3)H)$$

por NLCG (já tendo em mente o nosso processo de vetorização), onde o gradiente $\nabla\Omega$ é dado pela equação (3.14), com μ definido na equação (3.17).

No entanto, idempotência, em geral, não é preservada pela atualização

$$\rho_{k+1} \leftarrow \rho_k + \alpha \vec{d}_k,$$

e a perda da idempotência permite o surgimento de runaway solutions que impossibilitam a convergência do algoritmo. Por isso, em [Ferreira, 2017], aplica-se a purificação de MacWeeny depois de cada busca unidimensional. Desta forma, passa a existir uma perturbação no valor do traço das matrizes ρ ao longo do processo. Portanto, devemos reconsiderar o multiplicador de Lagrange no funcional objetivo Ω .

Por fim, fisicamente devemos ter a comutatividade do produto ρH , quando ρ for a densidade eletrônica. Esta propriedade afeta diretamente a escolha do critério de parada para o nosso algoritmo do NLCG. De fato, supondo que as matrizes construídas vão ficando aproximadamente idempotentes o suficiente para que possamos considerar $\mu = 0$, vemos que o gradiente se torna

$$\nabla\Omega = (H\rho + \rho H - 2\rho H\rho)^t.$$

É claro que, se ρ e H comutam, então $\nabla\Omega = 0$. Por outro lado, admitamos que $\nabla\Omega = 0$. Daí,

$$H\rho + \rho H = 2\rho H\rho.$$

Multiplicando-se esta equação por ρ à esquerda, vemos que

$$\rho H\rho + \rho^2 H = 2\rho^2 H\rho.$$

Considerando $\rho^2 = \rho$, obtemos que $\rho H = \rho H \rho$. De forma análoga, fazendo agora a multiplicação à direita, obtemos que $H \rho = \rho H \rho$. E, portanto $\rho H = H \rho$.

Com isso, vemos que ρ será um ponto crítico de Ω se, e somente se comutar com a matriz do hamiltoniano Tight-Binding. Isto nos leva a concluir que não devemos usar um número máximo de iterações no nosso algoritmo. O critério de parada deve estar ligado apenas à tolerância do valor $\|\nabla\Omega\|$.

Como dissemos, o NLCG não é uma técnica completamente estabelecida. Existem várias versões de implementação que são não equivalentes. Além disso, como discutimos brevemente, qualquer implementação do NLCG depende de vários parâmetros numéricos, cujas escolhas podem, em alguns casos, fazer com que o processo iterativo não convirja. Ainda assim, devido ao fato de que este procedimento computacional não exige o uso da matriz hessiana do funcional em questão e possuir baixa complexidade computacional, o NLCG tem sido usado no estudo de estrutura eletrônica de sólidos ao longo dos anos. Em [Ferreira, 2017] temos uma implementação detalhada e completa do NLCG para o modelo do DMTB construído na presente dissertação para super-células de átomos de silício, fazendo uso do hamiltoniano semi empírico de Kwon ([Kwon et al., 1994]). Além disso, em ([Kwon et al., 1994]) também é apresentada uma formulação funcional parametrizada para E_{rep} bem como o parâmetro E_0 para o caso do silício.

3.5 Conclusões e Perspectivas

O já consagrado método para o cálculo da energia quanto-mecânica do estado fundamental, conhecido como DMTB, foi construído e apresentado na presente dissertação dentro dos moldes da Modelagem Matemática e Computacional. A construção apresentada compôs a base teórica para a implementação computacional do DMTB feita em [Ferreira, 2017]. Os resultados consistentes obtidos em [Ferreira, 2017] confirmam a coerência das ideias desenvolvidas aqui. Nossa abordagem teve o objetivo de ressaltar rigorosamente os aspectos teóricos da metodologia do DMTB e para isso foi necessário desconstruirmos e reconstruirmos uma série de conceitos e técnicas tanto da matemática, quanto da teoria quântica da matéria condensada. Também foi nossa intenção tornar o texto o mais auto contido possível. Por conta disso, adotamos um enfoque que, em grande parte, não pode ser encontrado na literatura, o que permitiu a obtenção dos resultados de maneira simples e com poucos pré-requisitos, tornando assim a exposição

acessível a um grande número de leitores.

Muitas hipóteses foram feitas ao longo de todo o desenvolvimento apresentado. Nos tentamos fornecer justificativas *a priori* para muitas delas. Além disso, algumas aproximações e construções computacionais foram feitas sem invocarmos nenhuma razão física que as justificassem. Existe um argumento que é muito importante em Modelagem Matemática e Computacional que a justificativa *a posteriori*, onde os resultados justificam as hipóteses e aproximações adotadas. Ou seja, a teoria é desenvolvida e suas previsões são comparadas com os resultados experimentais conhecidos, ou com os resultados de outras teorias já pré estabelecidas, como explicamos sobre a validação de modelos. Em outras palavras, as hipóteses são consideradas aceitáveis porque o modelo final funciona! No nosso contexto, muitas medidas, sejam obtidas de forma experimental, sejam obtidas por métodos *ab initio*, estão em acordo com as previsões obtidas pelo DMTB. Isso justifica o seu largo nas últimas duas décadas. Para se ter uma ideia do tipo de validação que pode ser feita para o DMTB desenvolvido nesta dissertação, podemos tomar o parâmetro de rede experimental dos cristais de Si. Sabe-se que este parâmetro de rede mede aproximadamente 5.43 angstroms. Sendo assim, o DMTB aplicado a cristais de silício, quando alimentado com este valor de parâmetro de rede, deve apresentar energia total (equação (3.3)) com valor mais baixo do que outros valores para este parâmetro. Mais precisamente, a energia total dada pela equação (3.3), vista como função do parâmetro de rede, deve apresentar um mínimo em 5.43 Å. Como, tanto a matriz do hamiltoniano cristalino, quanto E_{rep} são dados de forma parametrizada, este tipo de validação não está ligada somente ao DMTB, mas também ao hamiltoniano escolhido. Abaixo vemos o gráfico obtido em [Ferreira, 2017], utilizando o hamiltoniano de Kwon ([Kwon et al., 1994]) Para este gráfico, foram utilizados 10 valores de energia obtidos com

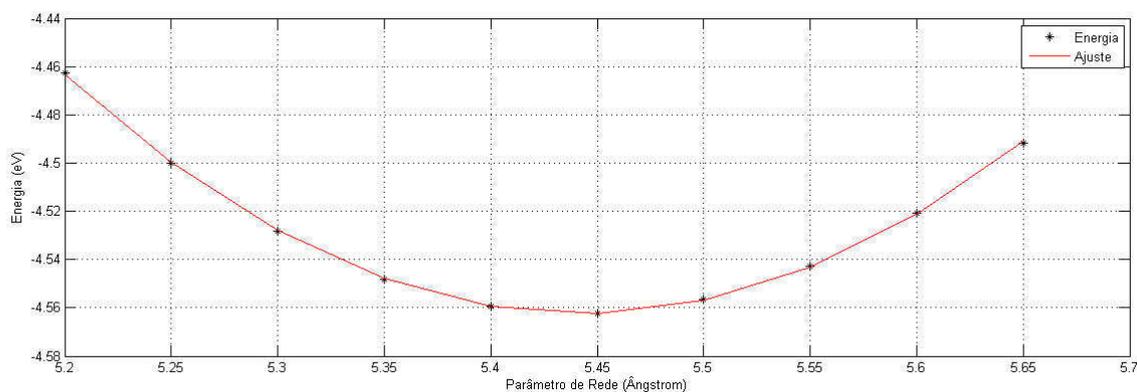


Figura 3.2: Cálculo do Parâmetro de Rede Experimental.

o DMTB, utilizando-se parâmetros de rede variando de 5.20 até 5.65, de 0.5 ângstrons em 0.5 ângstrons. Em seguida, ajustou-se, por mínimos quadrados, os pontos em uma parábola e obteve-se o seu mínimo. o valor mínimo ocorre em 5.4433 ângstrons. Com isso, vemos que o erro relativo é de aproximadamente 0.2%. Para maiores detalhes sobre este resultado, recomendamos [Ferreira, 2017].

3.6 Pesquisas Futuras

Não foi nossa intenção discutir o estado da arte do método DMTB. Desde a sua primeira aparição, em 1993 ([Li et al., 1993]), o DMTB tem sido alvo de aplicações, reformulações e investigações, tanto do ponto de vista físico-matemático ([Araújo, 2006], [Oliveira, 2014], [Kagimura, 2006], [Millam and Scuseria, 1997]), quanto do ponto de vista computacional ([Daniels and Scuseria, 1999], [Bollinger, 2008]). Em alguns momentos fomos bastante conservadores no uso de técnicas clássicas. Um exemplo foi o uso das purificações de Macwneey para a aproximação idempotente. No entanto, a construção apresentada se mostra bastante promissora para a inclusão de novas técnicas, devido à coerência lógica e forte conexão entre as seções da dissertação. Outras duas questões que não foram abordadas por nós foram o cálculo de forças a partir das expressões Hellmann-Feynman e os cálculos para simulações de dinâmica molecular. Estas consequências do DMTB estão indicadas em [Li et al., 1993]. Devido ao desenvolvimento que fizemos, a formalização e implementação destas consequências é um passo natural que devemos seguir.

Em [Li et al., 1993], o DMTB é apresentado como um método computacional de complexidade algorítmica (de pior caso) $O(N)$, onde N é o número de átomos. Para que isto realmente ocorra, uma série de estratégias computacionais precisam ser adotadas. Algumas são discutidas em [Daniels and Scuseria, 1999]. Uma próxima etapa de trabalho, partindo-se de [Ferreira, 2017], é atingir este nível de complexidade computacional. Isto inclui o uso de formatos especializados para matrizes esparsas e a paralelização computacional de certas subrotinas, como o traço do produto de matrizes esparsas. Uma vez que a implementação computacional feita em [Ferreira, 2017] faz uso do paradigma da orientação a objetos, as adaptações, reescritas e reaproveitamento de códigos podem ser realizada de modo bastante simples.

O algoritmo final apresenta uma modificação no NLCG. Desta forma, motivados pela convergência apresentada em [Ferreira, 2017] para o caso do Silício, a análise da convergência deste algoritmo "tipo"

NLCG, sob as hipóteses standard para a convergência do NLCG, também é uma etapa futura para a nossa pesquisa.

Um ponto central no nosso desenvolvimento foi o uso de bases ortonormais de orbitais. Este fato está presente nas propriedades da matriz densidade e na escolha do hamiltoniano parametrizado. Uma generalização do DMTB para bases não ortogonais de orbitais foi apresentada em [Nunes and Vanderbilt, 1994]. Desenvolver esta generalização nos moldes desta dissertação é um trabalho futuro que também pretendemos realizar.

Referências Bibliográficas

- [Amaral et al., 2011] Amaral, B., Baraviera, T., and Cunha, M. (2011). *Mecânica Quântica para Matemáticos em Formação*. 28° Colóquio Brasileiro de Matemática. IMPA.
- [Araújo, 2006] Araújo, M. M. (2006). *Teses de Doutorado: Estudos Teóricos Sobre Discordâncias Cristalinas em Silício*. PhD thesis, UFMG.
- [Armand, 2007] Armand, P. (2007). Modification of the wolfe line search rules to satisfy the descent condition in the polak-ribière-polyak conjugate gradient method. *Journal of Optimization Theory and Applications*, (2).
- [Ashcroft and Mermin, 1976] Ashcroft, N. and Mermin, N. (1976). *Solid state physics*. Science: Physics. Saunders College.
- [Ballentine, 2000] Ballentine, L. E. (2000). *Quantum Mechanics: A Modern Development*. World Scientific Publishing Co.
- [Barata, 2011] Barata, J. C. A. (2011). *Curso de Física-Matemática*. Universidade de São Paulo- Departamento de Física Matemática- Disponível no site: <http://denebola.if.usp.br/jbarata/Notas de aula>.
- [Bastos Filho and Siqueira, 1993] Bastos Filho, J. B. and Siqueira, A. F. (1993). O experimento da dupla fenda como exemplo de incognoscibilidade? *Revista Brasileira do Ensino de Física*, 15:153–162.
- [Bollinger, 2008] Bollinger, C. (2008). Term project: Linear scaling electronic structure methods. Master's thesis, ETH Zurich.

- [Brufati, 2011] Brufati, T. E. B. (2011). Trabalho de curso: Métodos de busca linear para problemas irrestritos. Master's thesis, UFP.
- [Custodio et al., 2002] Custodio, R., José Roberto dos Santos Politi, J. R. S., Segala, M., HaidukeKwon, R. L. A., and Cyrillo, M. (2002). Quatro alternativas para resolver a equação de schrödinger para o Átomo de hidrogênio. *Química Nova*, 25:150–170.
- [Daniels and Scuseria, 1999] Daniels, A. M. and Scuseria, G. E. (1999). What is the best alternative to diagonalization of the hamiltonian in large scale semiempirical calculation? *J. Chem. Phys.*, 110:5569–5577.
- [Dantas, 2004] Dantas, C. A. B. (2004). *PROBABILIDADE: Um Curso Introductório*. Edusp.
- [Daw, 1993] Daw, M. S. (1993). Model for energetics of solids based on the density matrix. *Phys. Rev. B*, 47:10895–10898.
- [Dirac, 1947] Dirac, P. (1947). *The Principles of Quantum Mechanics*. Oxford University Press.
- [Ferreira, 2017] Ferreira, F. L. S. (2017). Dissertação de mestrado: Cálculos quantum-mecânicos - implementação do método dmtb para a obtenção da energia no estado fundamental do silício. Master's thesis, UFRRJ.
- [Fletcher and Reeves, 1964] Fletcher, R. and Reeves, C. M. (1964). Function minimization by conjugate gradients. *The Computer Journal*, 7(2):149–154.
- [Gondar and Cipolatti, 2011] Gondar, J. L. and Cipolatti, R. (2011). *Iniciação à Física Matemática. Modelagem de Processos e Métodos de Solução*. IMPA.
- [Heisenberg, 1996] Heisenberg, W. (1996). *A parte e o todo (título original: PHYSICS AND BEYOND: ENCOUNTERS AND CONVERSATIONS)*. Contraponto.
- [Iorio Júnior, 1987] Iorio Júnior, R. (1987). *Tópicos na Teoria da Equação de Schrödinger*. IMPA.
- [Jauch, 1968] Jauch, J. M. (1968). *Foundation of Quantum Mechanics*. ADDISON-WESLEY PUBLISHING COMPANY.

- [Kagimura, 2006] Kagimura, R. (2006). *Tese de Doutorado: Estudo de Nanoestruturas Iônicas e Semicondutoras por Métodos de Primeiros Princípios*. PhD thesis, UFMG.
- [Karam and Almeida, 2003] Karam, F. J. and Almeida, R. C. (2003). *Introdução à Modelagem Matemática*. Notas impressas PósGraduação, LNCC.
- [Kittel, 1971] Kittel, C. (1971). *Introduction to solid state physics*. Wiley.
- [Kwon et al., 1994] Kwon, I., Biswas, R., Wang, C. Z., Ho, K. M., and Soukoulis, C. M. (1994). Transferable tight-binding models for silicon. *Phys. Rev. B*, 49:7242–7250.
- [Lang, 1966] Lang, S. (1966). *Linear Algebra*. Addison-Wesley Publishing Company.
- [Levine, 2006] Levine, I. N. (2006). *Quantum Chemistry, 5 Th. Ed.* Prentice Hall.
- [Li et al., 1993] Li, X.-P., Nunes, R. W., and Vanderbilt, D. (1993). Density-matrix electronic-structure method with linear system-size scaling. *Phys. Rev. B*, 47:10891–10894.
- [Lima, 2005] Lima, E. L. L. (2005). *Curso de análise Volume II*. IMPA.
- [McWeeny, 1960] McWeeny, R. (1960). Some recent advances in density matrix theory. *Rev. Mod. Phys.*, 32:335–369.
- [Medeiros and Miranda, 2011] Medeiros, L. A. and Miranda, M. M. (2011). *Espaços de Sobolev: Iniciação aos Problemas Elípticos não Homogêneos*. UFRJ.
- [Millam and Scuseria, 1997] Millam, J. M. and Scuseria, G. E. (1997). Liner scaling density matrix. *J. Chem. Phys.*, 106:5569–5577.
- [Nocedal and Wright, 2006] Nocedal, J. and Wright, S. J. (2006). *Numerical optimization*. Springer Series in Operations Research and Financial Engineering. Springer, Berlin. NEOS guide <http://www-fp.mcs.anl.gov/otc/Guide/>.
- [Nunes and Vanderbilt, 1994] Nunes, R. W. and Vanderbilt, D. (1994). Generalization of the density-matrix method to a nonorthogonal basis. *Phys. Rev. B*, 50:17611–17614.

- [Oliveira, 2014] Oliveira, A. C. A. (2014). Dissertação de mestrado: Modelagem computacional da interação entre discordâncias parciais a 90 graus e a superfície (111) do silício. Master's thesis, UFRRJ.
- [Pauling and Junior, 1937] Pauling, L. and Junior, E. B. W. (1937). *Introduction Quantum Mechanics with Applications to Chemistry*. McGraw-Hill Book Company.
- [Paxton, 2009] Paxton, A. T. (2009). *An Introduction to the Tigth Binding aproximation-implementation by diagonalisation*, volume 42 of *NIC series*, pages 145–176. John von Neumann Institute for Computing, Jülich. Atomistic Simulation Center-School of Mathematics and Physics, Queen's University Belfast-UK.
- [Polak and Ribiere, 1969] Polak, E. and Ribiere, G. (1969). Note sur la convergence de méthodes de directions conjuguées. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique*, 3(R1):35–43.
- [Prugovecki, 1971] Prugovecki, E. (1971). *Quantum Mechanics in Hilbert Space*. Pure and Applied Mathematics Series. Academic Press N. Y. and London.
- [Saito et al., 1998] Saito, R., Dresselhaus, G., and Dresselhaus, M. S. (1998). *Physical Properties of Carbon Nanotubes*. Imperial College Press.
- [Schrödinger, 1982] Schrödinger, E. (1982). Colected papers on wave mechanics. *Chelsea Publishing Company*.
- [Shewchuk, 1994] Shewchuk, J. (1994). *An Introduction to the Conjugate Gradient Method Without the Agonizing Pain*. Carnegie-Mellon University. Department of Computer Science.
- [Slater and Koster, 1954] Slater, J. C. and Koster, G. F. (1954). Simplified lcao method for the periodic potential problem. *Phys. Rev. B*, 94:1498–1524.
- [Spivak, 1965] Spivak, M. (1965). *Calculus on Manifolds*. Perseus Books Publishing L.L.C.
- [Stefan Goedecker, 1999] Stefan Goedecker, S. (1999). Linear scaling electronic structure methods. *Reviews of Modern Physics*, 71:7242–7250.

- [Sun and Liu, 2015] Sun, M. and Liu, J. (2015). Three modified polak-ribière-polyak conjugate gradient methods with sufficient descent property. *Journal of Inequalities and Applications*, 2015(1).
- [Teter et al., 1989] Teter, M. P., Payne, M. C., and Allan, D. C. (1989). Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. *Phys. Rev. B*, 40:1225–1263.
- [Yu et al., 2009] Yu, G., Guan, L., and Wei, Z. (2009). Globally convergent polak-ribière-polyak conjugate gradient methods under a modified wolfe line search. *Appl. Math. Comput.*, 215(8):3082–3090.
- [Zhang et al., 2012] Zhang, Y., Zheng, H., and Zhang, C. (2012). Global convergence of a modified prp conjugate gradient method. *Procedia Engineering*, 31:986 – 995.