

Solving Many-body Quantum Problem using Machine Learning

by

Even Marius Nordhagen

THESIS

for the degree of

MASTER OF SCIENCE



Faculty of Mathematics and Natural Sciences
University of Oslo

January 18, 2019

Abstract

Abstract should be written here

Acknowledgements

First of all I would like to thank Christine for cooking delicious food and feeding me when I need it the most. She and her food have maintained my motivation in the best way, and they are the main reason why I've survived those two years of masters.

Secondly, my annoying cohabitants have forced me to spend more time at university. I'm so glad you behave the way you do, if you were more cozy I would certainly spend more time at home and this thesis would not be the same.

Acknowledgements should be written here

-One of my mottos used to be that everything has a reason, which I used to point out whenever people were talking about things that apparently could not be described immediately. It was first when I learned about quantum mechanics that I understood I was wrong, and that is one of the reasons why quantum mechanics caught me so hardly.

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 9 |
| 1.1 | Many-Body problem | 9 |
| 1.2 | Machine learning | 9 |
| 1.3 | Goals and milestones | 9 |
| 2 | Quantum Many-Body Physics | 11 |
| 2.1 | Electronic structure calculations | 11 |
| 2.1.1 | The Schrödinger Equation | 11 |
| 2.1.2 | The Variational Principle | 12 |
| 2.1.3 | Assumptions raised | 12 |
| 2.2 | Wave function properties | 12 |
| 2.2.1 | Bosons and fermions | 12 |
| 2.2.2 | Slater determinant | 13 |
| 2.2.3 | Electron systems | 14 |
| 2.2.4 | Cusp condition | 14 |
| 2.2.5 | Basis set | 14 |
| 2.2.6 | Jastrow factor | 14 |
| 2.3 | Systems and basis sets | 14 |
| 2.3.1 | Quantum dots | 14 |
| 2.3.2 | Atomic and molecular systems | 15 |
| 3 | Methods | 17 |
| 3.1 | Quantum Monte Carlo | 17 |
| 3.1.1 | Variational Monte Carlo | 18 |
| 3.2 | Hartree-Fock | 19 |
| 3.3 | Configuration Interaction | 19 |
| 3.4 | Coupled Cluster | 19 |
| 4 | Machine Learning | 21 |
| 4.1 | Supervised Learning | 21 |
| 4.2 | Unsupervised Learning | 21 |
| 4.2.1 | Statistical foundation | 21 |
| 4.2.2 | Boltzmann Machines | 21 |
| 4.2.3 | Restricted Boltzmann Machines | 21 |
| 5 | Implementation | 23 |
| 5.1 | Foundation | 23 |
| 5.2 | Structure | 23 |
| 5.3 | Optimization algorithms | 24 |
| 5.3.1 | Stochastic Gradient Descent | 24 |

| | | |
|----------|---|-----------|
| 5.3.2 | ADAM | 24 |
| 6 | Results | 25 |
| 7 | Conclusion and future work | 27 |
| A | Appendix A - RBM wavefunction | 29 |
| B | Appendix B - Dirac notation | 31 |
| C | Appendix C - Scaling | 33 |
| | C.0.1 Harmonic Oscillator - Natural units | 33 |
| | C.0.2 Atomic systems - Atomic units | 33 |
| D | Appendix D - Local Energy Calculations | 35 |
| | D.0.1 Slater determinant | 35 |
| | D.0.2 NQS | 35 |
| | D.0.3 Jastrow factor | 35 |

List of abbreviations

| Letters | Meaning |
|---------|--------------------------------|
| RBM | - Restricted Boltzmann Machine |
| MC | - Monte Carlo |
| VMC | - Variational Monte Carlo |
| DMC | - Diffusion Monte Carlo |
| ML | - Machine Learning |
| WF | - Wave Function |
| SPF | - Single Particle Function |

Table 1: List of symbols used with explanation.

Source Code

The source code is given in <https://github.com/evenmn>

Chapter 1

Introduction

Write introduction here

- Low frequency (weakly interacting electrons) field of interest - Multi scale calculations
- Cartesian - Introduce the wavefunction - Mention the uncertainty principle and also quantum entanglement to catch the readers interest

1.1 Many-Body problem

Not possible to solve analytically

1.2 Machine learning

Branch of artificial intelligence

1.3 Goals and milestones

- Investigate a new method to solve the Many-Body problem

Chapter 2

Quantum Many-Body Physics

Here I might present basic quantum mechanics briefly.

2.1 Electronic structure calculations

Even though it is called elementary, quantum mechanics is not basic at all.

2.1.1 The Schrödinger Equation

The Schrödinger equation is a natural starting point, which gives the energy eigenstates of a system defined by a Hamiltonian \hat{H} and its eigenfunctions, Ψ , which are the wave functions. The time-independent Schrödinger equation reads

$$\hat{H}\Psi_n(\mathbf{r}) = \epsilon_n \Psi_n(\mathbf{r}) \quad (2.1)$$

where the electronic Hamiltonian takes the form

$$\hat{H} = - \sum_i^N \frac{\hbar^2}{2m} \nabla_i^2 + \sum_i^N u_i + \sum_i^N \sum_{j>i}^M k \frac{e^2}{r_{ij}} \quad (2.2)$$

with u_i as an arbitrary external potential and the last term as the Coulomb potential between electrons. r_{ij} is the relative distance between electron i and j , defined by $r_{ij} \equiv |\mathbf{r}_i - \mathbf{r}_j|$.

Setting up equation (2.1) with respect to the energies, we obtain some integrals,

$$\epsilon_n = \frac{\int d\mathbf{r} \Psi_n^*(\mathbf{r}) \hat{H} \Psi_n(\mathbf{r})}{\int d\mathbf{r} \Psi_n^*(\mathbf{r}) \Psi_n(\mathbf{r})}, \quad (2.3)$$

which not necessarily are trivial to solve. If we take the wave function squared we get the probability of finding a particle at a certain position,

$$P(\mathbf{r}) = \Psi_n^*(\mathbf{r}) \Psi_n(\mathbf{r}) = |\Psi_n(\mathbf{r})|^2 \quad (2.4)$$

so the nominator is simply the integral over all probabilities. If the wave function is normalized correctly, this should always give 1. Assuming that is the case, the expectation value can be expressed more elegantly by using Dirac notation,

$$E[\Psi] = \langle \Psi | \hat{H} | \Psi \rangle, \quad (2.5)$$

where the first part, $\langle\Psi|$ is called a bra and the last part, $|\Psi\rangle$ is called a ket. At first this might look artificial and less informative, but it simplifies the notation significantly. More information about the notation is found in Appendix B.

We often do not know the exact wave function, and need to guess a trial wave function. Henceforth, we will use Ψ as the exact total wave function, ψ as the exact single particle function (SPF), Φ as the total trial wave function and ϕ as the trial SPF. [Gri05]

2.1.2 The Variational Principle

In the equations above, the presented wave functions are assumed to be the exact eigen functions of the Hamiltonian. But often we do not know the exact wave functions, and we need to guess what the wave functions might be. In those cases we make use of the variational principle, which states that only the groundstate wave function is able to give the groundstate energy. All other wave functions with the required properties (see section 2.2) give higher energies, and mathematically we can express the statement with

$$\epsilon_0 \leq \langle\Phi|\hat{H}|\Phi\rangle. \quad (2.6)$$

We have here introduced a trial wave function Ψ_T which not necessary is the true ground state wave function. This means that we can adjust our trial wave function with respect to the energy, and the lower energy we get the better is our wave function.

2.1.3 Assumptions raised

- Non-relativistic
- Point-like particles
- Born-Oppenheimer

2.2 Wave function properties

By the first postulate of quantum mechanics, the wave function contains all the information specifying the state of the system. This means that all observable in classical mechanics can also be measured from the wave function, so finding the wave function is the goal. We will focus on Variational Monte Carlo (VMC), which is a method that calculates the energy and optimizes the wave function iteratively based on an initial wave function guess. There are some requirements that the wave function needs to fulfill: It needs to be symmetric or anti-symmetric under exchange of two coordinates, the CUSP condition should be satisfied and it needs to approach zero at the edges. As long as these conditions are fulfilled, the variational principle is fulfilled.

2.2.1 Bosons and fermions

An introductory sentence is needed to increase the FLOW. Assume we have a permutation operator \hat{P} which exchanges two coordinates in the wave function,

$$\hat{P}(i \rightarrow j)\Psi_n(\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_j, \dots, \mathbf{x}_M) = p\Psi_n(\mathbf{x}_1, \dots, \mathbf{x}_j, \dots, \mathbf{x}_i, \dots, \mathbf{x}_M), \quad (2.7)$$

where p is just a factor which comes from the transformation. If we again apply the \hat{P} operator, we should switch the same coordinates back, and we expect to end up with the initial wave function. For that reason, $p = \pm 1$.¹

The particles that have an antisymmetric (AS) wavefunction under exchange of two coordinates are called fermions, named after Enrico Fermi, and have half integer spin. On the other hand, the particles that have a symmetric (S) wavefunction under exchange of two coordinates are called bosons, named after Satyendra Nath Bose, and have integer spin.

It turns out that because of their antisymmetric wavefunction, two identical fermions cannot be found at the same position at the same time, known as the Pauli principle. This causes some difficulties when dealing with multiple fermions, because we always need to ensure that the total wavefunction becomes zero if two identical particles happen to be at the same position. To do this, we introduce a Slater determinant as described in the next chapter. In this particular project, we are going to focus on electrons and therefore fermions. Anyway, much of the theory applies for bosons as well.

Read <https://manybodyphysics.github.io/FYS4480/doc/pub/secondquant/html/secondquant-bs.html>

2.2.2 Slater determinant

For a system of more particles we can define a total wavefunction, which is a composition of all the single particle wavefunctions (SPF) and contains all the information about the system. For fermions we need to compile the SPFs such that the Pauli principle is fulfilled at all times. One way to do this is by setting up the SPFs in a determinant, known as a Slater determinant.

Consider a system of two identical fermions with SPFs ϕ_1 and ϕ_2 at positions \mathbf{r}_1 and \mathbf{r}_2 respectively. The way we define the wavefunction of the system is then

$$\Psi_T(\mathbf{r}_1, \mathbf{r}_2) = \begin{vmatrix} \phi_1(\mathbf{r}_1) & \phi_2(\mathbf{r}_1) \\ \phi_1(\mathbf{r}_2) & \phi_2(\mathbf{r}_2) \end{vmatrix} = \phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2) - \phi_2(\mathbf{r}_1)\phi_1(\mathbf{r}_2), \quad (2.8)$$

which is set to zero if the particles are at the same position. This is called a Slater determinant, and yields the same no matter how big the system is.

Notice that we denote the wave function with the ' T ', which indicates that it is a trial wave function. We do this because the spin part is avoided with ψ as the radial parts only, thus this wavefunction is not the true wavefunction. We will look closer at how we can factorize out the spin part later. The spin part is assumed to not affect the energies.

A general Slater determinant for a system of N particles takes the form

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \begin{vmatrix} \psi_1(\mathbf{r}_1) & \psi_2(\mathbf{r}_1) & \dots & \psi_N(\mathbf{r}_1) \\ \psi_1(\mathbf{r}_2) & \psi_2(\mathbf{r}_2) & \dots & \psi_N(\mathbf{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(\mathbf{r}_N) & \psi_2(\mathbf{r}_N) & \dots & \psi_N(\mathbf{r}_N) \end{vmatrix} \quad (2.9)$$

where the ψ 's are the true single particle wave functions, which are the tensor products

$$\psi = \phi \otimes \xi \quad (2.10)$$

with ξ as the spin part.

¹This was true until 1976, when J.M. Leinaas and J. Myrheim discovered the anyon, <https://www.uio.no/studier/emner/matnat/fys/FYS4130/v14/documents/kompendium.pdf>.

There is a similar "Slater permanent" which applies for bosons. It is similar to a determinant, but all negative signs are replaced by positive signs.

2.2.3 Electron systems

For our purpose we will study fermions with spin $\sigma = \pm 1/2$ only, which can be seen as electrons. In this particular case, the SPFs can be arranged in spin-up and spin-down parts, such that the Slater determinant can be simplified to

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \begin{vmatrix} \phi_1(\mathbf{r}_1)\xi_{\uparrow} & \phi_1(\mathbf{r}_1)\xi_{\downarrow} & \dots & \phi_{N/2}(\mathbf{r}_1)\xi_{\downarrow} \\ \phi_1(\mathbf{r}_2)\xi_{\uparrow} & \phi_1(\mathbf{r}_2)\xi_{\downarrow} & \dots & \phi_{N/2}(\mathbf{r}_2)\xi_{\downarrow} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\mathbf{r}_N)\xi_{\uparrow} & \phi_1(\mathbf{r}_N)\xi_{\downarrow} & \dots & \phi_{N/2}(\mathbf{r}_N)\xi_{\downarrow} \end{vmatrix}. \quad (2.11)$$

This is called the wavefunction ansatz, because assumptions are raised, like two particles with opposite spins are found to be at the same position all the time, i.e, an equal number of fermions have spin up as spin down, are applied. Further the...

SHOULD END UP WITH SPLITTED DETERMINANTS HERE

For a detailed walkthrough, see appendix I in REF(The Stochastic Gradient Approximation: an application to Li nanoclusters, Daniel Nissenbaum).

2.2.4 Cusp condition

The Coulomb potential gives a cusp when two charged particles approach each other. This means that the wave function should also have a cusp, known as the Coulomb cusp condition, Kato theorem or just the cusp condition.

2.2.5 Basis set

To go further, we need to define a basis set, $\phi_n(\mathbf{r})$ which should be chosen carefully based on the system.

2.2.6 Jastrow factor

Often we add a Jastrow factor to improve the flexibility of the wave function. The Slater determinant itself is usually not able to capture interaction properties, so by multiplying with another function, we might get a better estimate.

2.3 Systems and basis sets

Specific systems need specific basis sets

2.3.1 Quantum dots

Quantum dots are very small particles, and contain fermions or bosons hold together by an external potential. Since these particles have discrete electronic states like an atom, they are often called artificial atoms.

In this thesis we will study electrons trapped in harmonic oscillators,

$$\hat{H} = \sum_{i=1}^P \left(-\frac{1}{2} \nabla_i^2 + \frac{1}{2} \omega^2 r_i^2 \right) + \sum_{i<j} \frac{1}{r_{ij}} \quad (2.12)$$

and we will use the Hermite functions as our basis set. The one dimensional Hermite functions read

$$f_n(x) = H_n(x) \exp(-x^2/2) \quad (2.13)$$

and are known to be the exact SPFs in a harmonic oscillator.

2.3.2 Atomic and molecular systems

We will also investigate real atoms, which have Hamiltonians defined by the Born-Oppenheimer approximation.

$$\hat{H} = \sum_{i=1}^P \left(-\frac{1}{2} \nabla_i^2 + \frac{1}{2} \omega^2 \frac{1}{r_i} \right) + \sum_{i<j} \frac{1}{r_{ij}} \quad (2.14)$$

In atomic units.

2.3.2.1 Electron gas

2.3.2.2 Helium gas

He³ and He⁴

Chapter 3

Methods

Some great methods are developed over the past decade. We will focus on the Hartree-Fock and variational Monte Carlo methods and give a detailed explanation of those methods. Additionally, the well-known methods configuration interaction and coupled cluster will be described briefly for some kind of completeness.

3.1 Quantum Monte Carlo

Monte Carlo methods in quantum mechanics are a bunch of methods that are built on diffusion processes, and includes Variational Monte Carlo (VMC), Diffusion Monte Carlo (DMC) and others. The common denominator is that we move particles in order to find the optimal configuration, usually where the energy is minimized. The particles can be moved isotropic, i.e., uniformly in all directions, or they can be affected by a drift force which makes the process anisotropic.

3.1.0.1 Isotropic processes

In isotropic processes, we have random walks where the particles move randomly in space which falls under the category Markov chains. If we assume constant timestep, the step index can be considered the time, thus we have a time-dependent probability density $P(\mathbf{x}, t)$. This probability density needs to satisfy the isotropic diffusion equation,

$$\frac{\partial P(\mathbf{x}, t)}{\partial t} = D \frac{\partial^2}{\partial \mathbf{x}^2} P(\mathbf{x}, t). \quad (3.1)$$

3.1.0.2 Anisotropic processes

For anisotropic processes, we have a drift and the moves are no longer considered random but falls still under the category Markov chains. Because of the drift, we need to rewrite the diffusion equation and we end up with the Fokker-Planck equation,

$$\frac{\partial P(\mathbf{x}, t)}{\partial t} = D \frac{\partial}{\partial \mathbf{x}} \left(\frac{\partial}{\partial \mathbf{x}} - F \right) P(\mathbf{x}, t) \quad (3.2)$$

which needs to be satisfied. The new positions in coordinate space are given as solution of the Langevin equation

$$\frac{\partial \mathbf{x}(t)}{\partial t} = D \mathbf{F}(\mathbf{x}(t)) + \eta \quad (3.3)$$

3.1.1 Variational Monte Carlo

The Variational Monte Carlo (hereafter, VMC) method is today widely used when it comes to the study of ground state properties of quantum mechanical systems. It is a Monte Carlo method which makes use of Metropolis sampling, and has been used in studies of fermionic systems since the 1970's. [Deb14] If we go back to the variational principle in equation (2.6), we see that by choosing a wave function which satisfies the criteria, we will get an energy larger or equal to the ground state energy.

There are two main problems we need to solve

1. We seldomly know the correct wave function
2. The integral we need to find the energy is hard or impossible to solve

Let us take the last problem first. The solution is to approximate the integral with a sum,

$$\begin{aligned}
 E &\leq \frac{\int \Phi(\mathbf{r})^* \hat{H} \Phi(\mathbf{r}) d\mathbf{r}}{\int \Phi(\mathbf{r})^* \Phi(\mathbf{r}) d\mathbf{r}} \\
 &= \frac{\int \Phi(\mathbf{r})^* \Phi(\mathbf{r}) (1/\Phi(\mathbf{r})) \hat{H} \Phi(\mathbf{r}) d\mathbf{r}}{\int \Phi(\mathbf{r})^* \Phi(\mathbf{r}) d\mathbf{r}} \\
 &= \frac{\int P(\mathbf{r}) E_L(\mathbf{r}) d\mathbf{r}}{\int \Phi(\mathbf{r})^* \Phi(\mathbf{r}) d\mathbf{r}} \\
 &\approx \frac{1}{N} \sum_{i=1}^N E_L(\mathbf{r}_i)
 \end{aligned} \tag{3.4}$$

where the local energy is defined as

$$E_L(\mathbf{r}) \equiv \frac{1}{\Phi(\mathbf{r})} \hat{H} \Phi(\mathbf{r}) \tag{3.5}$$

and the \mathbf{r}_i is withdrawn from the probability distribution $P(\mathbf{r})$. The energy found from equation (3.4) is an expectation value, and we therefore know that the true energy lies within the standard error. For more statistical details, see [Deb2014].

Given a wave function we are now able to estimate the corresponding energy, but how do we find a good wave function? In VMC, we define a wave function with variational parameters, which are adjusted in order to minimize the energy for every iteration. For every iteration, we run N Monte-Carlo cycles where we withdraw a new position \mathbf{r}_i . Whether or not the proposed move should be accepted is determined by the Metropolis algorithm.

3.1.1.1 The Metropolis Algorithm

The genius of the metropolis algorithm, is that the acceptance of a move is not based on the probabilities themselves, but the ratio between the new and the old probability. In that way, we avoid calculating the sum over all probabilities, which often is expensive or even impossible to calculate.

In its simplest form, the move is proposed randomly, and it is accepted if the ratio is larger than a random number between 0 and 1. However, with this approach a lot of moves will be rejected, which wastes CPU time. A better method is **importance sampling**, which makes a educated guess of the best way to move based on diffusion processes, and move the particle in that direction.

A time-dependent probability density needs to satisfy the Fokker-Planck equation

3.2 Hartree-Fock

Describe this detailed

3.3 Configuration Interaction

The completeness relation

3.4 Coupled Cluster

The coupled cluster method is the *de facto* standard wave function-based method for electronic structure calculations. [Pal05] The method approximates the wave function with an exponential expansion,

$$|\Psi_{\text{CC}}\rangle = e^{\hat{T}} |\Phi_0\rangle \quad (3.6)$$

where \hat{T} is the cluster operator, entirely given by $\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \dots$ with

$$\hat{T}_n = \left(\frac{1}{n!}\right)^2 \sum_{abc\dots} \sum_{ijk\dots} t_{ijk\dots}^{abc\dots} a_a^\dagger a_b^\dagger a_c^\dagger \dots a_k a_j a_i. \quad (3.7)$$

We again want to solve the Schrödinger equation,

$$\hat{H} |\Psi\rangle = \hat{H} e^{\hat{T}} |\Phi_0\rangle = \epsilon e^{\hat{T}} |\Phi_0\rangle, \quad (3.8)$$

which can be simplified by multiplying with $e^{-\hat{T}}$ from the left. This introduces us to the **similarity transformed Hamiltonian**

$$\bar{H} = e^{-\hat{T}} \hat{H} e^{\hat{T}}. \quad (3.9)$$

If we on one hand now multiply with the reference bra on the left hand side, we easily observe that

$$\langle \Phi_0 | \bar{H} | \Phi_0 \rangle = \epsilon \quad (3.10)$$

which is the coupled cluster energy equation. On the other hand, we can multiply with an excited bra on left hand side, and find that

$$\langle \Phi_{ijk\dots}^{abc\dots} | \bar{H} | \Phi_0 \rangle = 0 \quad (3.11)$$

which are the coupled cluster amplitude equations. The similarity transformed Hamiltonian can be rewritten using the Baker-Campbell-Hausdorff expansion

$$\begin{aligned} \bar{H} &= \hat{H} + [\hat{H}, \hat{T}] \\ &\quad + \frac{1}{2} [[\hat{H}, \hat{T}], \hat{T}] \\ &\quad + \frac{1}{6} [[[[\hat{H}, \hat{T}], \hat{T}], \hat{T}], \hat{T}] \\ &\quad + \frac{1}{24} [[[[[[\hat{H}, \hat{T}], \hat{T}], \hat{T}], \hat{T}], \hat{T}], \hat{T}] \\ &\quad + \dots \end{aligned} \quad (3.12)$$

and we are in principle set to solve the amplitude equations with respect to the amplitudes $t_{ijk\dots}^{abc\dots}$ and then find the energy. The expansion is able to reproduce the true wave function exactly using a satisfying number of terms and an infinite basis. This is, of course, not possible, but even by limiting us to the first few coupled cluster operators, the results are often good compared to other methods. [DCFSI07]

Chapter 4

Machine Learning

Parameter estimation is a large part of science, because we are often not able to measure things directly. For instance, when Millikan in ... performed his famous electron charge experiment or when we estimate Sun's mass.

Sometimes when people talk about machine learning, it sounds like its a totally new field, but the truth is that many techniques that used to have other names, are now classified as machine learning. As the UiO professor Anne Solberg said during one of her lectures

"In the early 1990's I was working a lot with machine learning, but at that time we called it pattern recognition and regression"

Machine learning includes all methods where we want to fit a model to a data set, and pattern recognition and regression are clearly some of those methods. In both cases we have concrete targets for the training, and the training is therefore called supervised training. In other cases we do not have targets, but want to train our model based on a probability distribution or so, which is called unsupervised learning.

4.1 Supervised Learning

Regression, pattern recognition, Feed-forward neural networks,

4.2 Unsupervised Learning

How can we train using unsupervised learning?

4.2.1 Statistical foundation

Bayesian statistics

4.2.2 Boltzmann Machines

Based on a Hopfield network How the things are derived, including Boltzmann distribution and stuff from statistical mechanics

4.2.3 Restricted Boltzmann Machines

Chapter 5

Implementation

5.1 Foundation

5.2 Structure

Add structure chart

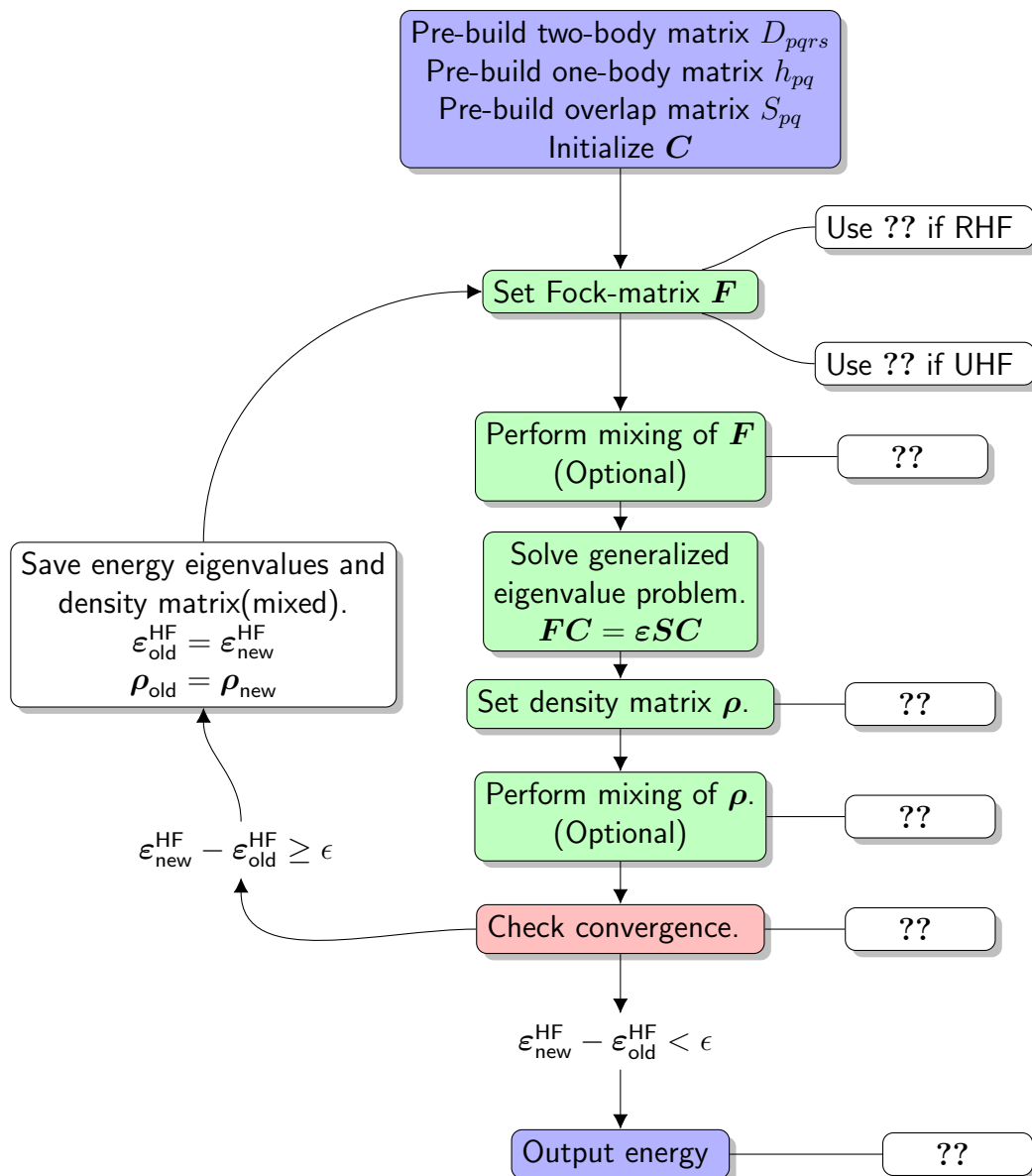


Figure 5.1: Hartree-Fock Algorithm.

5.3 Optimization algorithms

5.3.1 Stochastic Gradient Descent

5.3.2 ADAM

Chapter 6

Results

Chapter 7

Conclusion and future work

Appendix A

Appendix A - RBM wavefunction

We have seen that the probability of having a set of positions \mathbf{X} with a set of hidden nodes \mathbf{h} is given by

$$F(\mathbf{X}, \mathbf{H}) = \frac{1}{Z} e^{-\beta E(\mathbf{X}, \mathbf{H})} \quad (\text{A.1})$$

where we set $\beta = 1/kT = 1$, Z is the partition function and $E(\mathbf{X}, \mathbf{h})$ is the system energy

$$E(\mathbf{X}, \mathbf{H}) = \sum_{i=1}^M \frac{(X_i - a_i)^2}{2\sigma_i^2} - \sum_{j=1}^N b_j H_j - \sum_{i,j=1}^{M,N} \frac{X_i W_{ij} H_j}{\sigma_i^2} \quad (\text{A.2})$$

such that

$$F_{\text{RBM}}(\mathbf{X}, \mathbf{h}) = e^{\sum_{i=1}^M \frac{(X_i - a_i)^2}{2\sigma_i^2}} e^{\sum_{j=1}^N \left(b_j h_j + \sum_{i=1}^M \frac{X_i W_{ij} h_j}{\sigma_i^2} \right)}. \quad (\text{A.3})$$

We skip the partition function because it will not affect the results (it is just a normalization constant). The probability of a set of positions only is therefore the sum over all sets of \mathbf{h} , $\{\mathbf{h}\}$:

$$\begin{aligned} F_{\text{RBM}}(\mathbf{X}) &= \sum_{\{\mathbf{h}\}} e^{\sum_{i=1}^M \frac{(X_i - a_i)^2}{2\sigma_i^2}} \prod_{j=1}^N e^{b_j h_j + \sum_{i=1}^M \frac{X_i W_{ij} h_j}{\sigma_i^2}} \\ &= \sum_{h_1} \sum_{h_2} \dots \sum_{h_N} e^{\sum_{i=1}^M \frac{(X_i - a_i)^2}{2\sigma_i^2}} \times \\ &\quad e^{b_1 h_1 + \sum_{i=1}^M \frac{X_i W_{i1} h_1}{\sigma_i^2}} e^{b_2 h_2 + \sum_{i=1}^M \frac{X_i W_{i2} h_2}{\sigma_i^2}} \dots e^{b_N h_N + \sum_{i=1}^M \frac{X_i W_{iN} h_N}{\sigma_i^2}} \\ &= e^{\sum_{i=1}^M \frac{(X_i - a_i)^2}{2\sigma_i^2}} \prod_{j=1}^N \sum_{h_j=0}^1 e^{b_j h_j + \sum_{i=1}^M \frac{X_i W_{ij} h_j}{\sigma_i^2}} \\ &= e^{\sum_{i=1}^M \frac{(X_i - a_i)^2}{2\sigma_i^2}} \prod_{j=1}^N \left(1 + e^{b_j + \frac{\mathbf{x}^T \mathbf{w}_{*j}}{\sigma^2}} \right) \end{aligned} \quad (\text{A.4})$$

Appendix B

Appendix B - Dirac notation

The Dirac notation, also called bracket notation, was suggested by Paul Dirac in a 1939 paper with the purpose of improving the reading ease. [Dir39]

Appendix C

Appendix C - Scaling

C.0.1 Harmonic Oscillator - Natural units

Hamiltonian is in one dimension given by

$$H = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + \frac{1}{2} m \omega^2 x^2 \quad (\text{C.1})$$

which has corresponding wavefunctions

$$\psi_n(x) = \frac{1}{\sqrt{2^n n!}} \cdot \left(\frac{m\omega}{\pi\hbar} \right)^{1/4} \exp\left(-\frac{m\omega}{2\hbar} x^2\right) H_n\left(\sqrt{\frac{m\omega}{\hbar}} x\right). \quad (\text{C.2})$$

We want to get rid of \hbar and m in equation (C.1), and we start with scaling $H' \equiv H/\hbar$, such that the Hamiltonian reduces to

$$H' = -\frac{\hbar}{2m} \frac{\partial^2}{\partial x^2} + \frac{1}{2} \frac{m\omega^2}{\hbar} x^2 \quad (\text{C.3})$$

One can now observe that the fraction \hbar/m comes in both terms, which we can take out by scaling the length $x' \equiv x \cdot \sqrt{m/\hbar}$. The final Hamiltonian is

$$H = \frac{1}{2} \frac{\partial^2}{\partial x^2} + \frac{1}{2} \omega^2 x^2 \quad (\text{C.4})$$

which corresponds to setting $\hbar = m = 1$. In natural units, one often sets $\omega = 1$ as well by scaling $H' = H/\hbar\omega$, but since we want to keep the ω -dependency, we do it slightly different. This means that the exact wavefunctions for the one-particle-one-dimension case is

$$\psi_n(x) = \exp\left(-\frac{\omega}{2} x^2\right) H_n(\sqrt{\omega} x) \quad (\text{C.5})$$

where we take advantage of the Metropolis algorithm and ignore the normalization constant.

C.0.2 Atomic systems - Atomic units

Appendix D

Appendix D - Local Energy Calculations

The local energy is found by

$$\frac{1}{\Psi_T} \hat{H} \Psi_T = \sum_{k=1}^M \left[\frac{1}{\Psi_T} \nabla_k \Psi_T + U_k + V_k \right]. \quad (\text{D.1})$$

The kinetic energy term is what we will evaluate here. An useful relation is

$$\frac{1}{\Psi_T} \nabla_k \Psi_T = \nabla_k^2 \ln \Psi_T + (\nabla_k \ln \Psi_T)^2 \quad (\text{D.2})$$

which is convenient since the logarithm of the wavefunction is simple.

D.0.1 Slater determinant

D.0.2 NQS

D.0.3 Jastrow factor

Bibliography

- [Dir39] P. A. M. Dirac. “A new notation for quantum mechanics”. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 35.3 (1939), 416–418. DOI: 10.1017/S0305004100021162.
- [Gri05] D.J. Griffiths. *Introduction to quantum mechanics*. 2nd Edition. Pearson PH, 2005. ISBN: 0131911759.
- [Pal05] Josef Paldus. *The beginnings of coupled-cluster theory: An eyewitness account*. Amsterdam: Elsevier, 2005, pp. 115 –147. ISBN: 9780444517197. DOI: 10.1016/B978-044451719-7/50050-0.
- [DCFSI07] T Daniel Crawford and Henry F. Schaefer III. “An Introduction to Coupled Cluster Theory for Computational Chemists”. In: vol. 14. Jan. 2007, pp. 33 –136. ISBN: 9780470125915. DOI: 10.1002/9780470125915.ch2.
- [Deb14] Sukanta Deb. “Variational Monte Carlo technique”. In: *Resonance* 19.8 (2014), pp. 713–739. ISSN: 0973-712X. DOI: 10.1007/s12045-014-0079-x. URL: <https://doi.org/10.1007/s12045-014-0079-x>.