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

December 16, 2018

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 catched me so hardly.

Contents

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 explaination.

Source Code

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

1 Introduction

Write introduction here

- Introduce the wavefunction - Mention the uncertainity principle and also quantum entanglement to catch the readers interest

1.1 Many-Body problem

1.2 Machine learning

1.3 Goals and milestones

2 Quantum Many-Body Physics

Here I might present basic quantum mechanics briefly.

2.1 Elementary Quantum Theory

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

2.1.1 Postulates of Quantum Mechanics

We will first list the postulates of quantum mechanics, and then discuss their importance

- The state of a quantum mechanical system is completely specified by its wavefunction

- To every observable in classical mechanics there corresponds a linear, Hermitian operator in quantum mechanics.

-

2.1.2 The Schrödinger Equation

A natural starting point is with the Schrödinger equation, which gives the energy eigenstates of a system defined by a Hamiltonian $\hat{H}$ and its eigenfunctions, Ψ, which are the wavefunctions,

$$\hat{H}\Psi_n(\boldsymbol{r}) = \epsilon_n \Psi_n(\boldsymbol{r}). \tag{1}$$

Solving equation (1) with respect to the energies, we obtains some integrals,

$$\epsilon_n = \frac{\int d\boldsymbol{r}\,\Psi_n^*(\boldsymbol{r})\hat{H}\Psi_n(\boldsymbol{r})}{\int d\boldsymbol{r}\,\Psi_n^*(\boldsymbol{r})\Psi_n(\boldsymbol{r})}, \tag{2}$$

which not necessarily are trivial to solve. The integral in the nominator gives the normalization, and ensures that the wavefunction is normalized. If this already is the case, we can omit the nominator and focus on solving the integral in the denominator. The equation can be expressed more elegantly using Dirac braket notation,

$$\epsilon_n = \langle n|\hat{H}|n\rangle, \tag{3}$$

where the first part, $\langle n|$ is called a bra and the last part, $|n\rangle$ 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. [Gri05]

2.1.3   The Variational Principle

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

$$\epsilon_0 \leq \langle \Psi_T | \hat{\mathrm{H}} | \Psi_T \rangle . \tag{4}$$

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

2.2   Wavefunction

In quantum mechanical calculations, the wavefunction is often the problem. As the first postulate of quantum mechanics states, the wavefunction carries all the information about the system, which is an enormous amount of information when the system gets large. We have methods to expand the total wavefunction in Slater Determinants if various single particle functions (SPFs), known as Full Configuration Interaction (FCI), which gives the exact wavefunction when all SPFs are taken into account. The problem is that even for relatively small systems, the computational expensiveness becomes unmanageable for normal computers, and we need to look for approximative methods. An example is the Hartree-Fock method, which optimize a Slater Determinant such that the energy is minimized. More about the different methods in section 3. Before that, we will look at which properties the wavefunction needs to take.

2.2.1   Symmetry and anti-symmetry

Assume we have a permutation operator $\hat{P}$ which switches two coordinates in the wavefunction,

$$\hat{P}\Psi_n(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_i, \ldots, \boldsymbol{x}_j, \ldots, \boldsymbol{x}_M) = p\Psi_n(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_j, \ldots, \boldsymbol{x}_i, \ldots, \boldsymbol{x}_M), \tag{5}$$

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 two identical fermions, because of their antisymmetric wavefunction, 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 section. In this particular project, we are going to focus on fermions, just because they are more difficult to handle, but much of the theory and implementation 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 (SPFs) and contains all the information about the system. The way we compile the SPFs needs to be based on Pauli's exclusion as discussed above.

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

$$\Psi_T(\boldsymbol{r}_1, \boldsymbol{r}_2) = \begin{vmatrix} \phi_1(\boldsymbol{r}_1) & \phi_2(\boldsymbol{r}_1) \\ \phi_1(\boldsymbol{r}_2) & \phi_2(\boldsymbol{r}_2) \end{vmatrix} = \phi_1(\boldsymbol{r}_1)\phi_2(\boldsymbol{r}_2) - \phi_2(\boldsymbol{r}_1)\phi_1(\boldsymbol{r}_2), \tag{6}$$

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 wavefunction with the $'T'$, which indicates that it is a trial wavefunction. 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. For bosons the total wavefunction is defined similarly, but with no negative signs since the Pauli principle does not apply. This is called a permanent.

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

---

[1]This 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.

$$\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N) = \begin{vmatrix} \psi_1(\boldsymbol{r}_1) & \psi_2(\boldsymbol{r}_1) & \ldots & \psi_N(\boldsymbol{r}_1) \\ \psi_1(\boldsymbol{r}_2) & \psi_2(\boldsymbol{r}_2) & \ldots & \psi_N(\boldsymbol{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(\boldsymbol{r}_N) & \psi_2(\boldsymbol{r}_N) & \ldots & \psi_N(\boldsymbol{r}_N) \end{vmatrix} \tag{7}$$

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

$$\psi = \phi \otimes \xi \tag{8}$$

with ξ as the spin part.

2.2.3   Electron system

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

$$\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N) = \begin{vmatrix} \phi_1(\boldsymbol{r}_1)\xi_\uparrow & \phi_1(\boldsymbol{r}_1)\xi_\downarrow & \ldots & \phi_{N/2}(\boldsymbol{r}_1)\xi_\downarrow \\ \phi_1(\boldsymbol{r}_2)\xi_\uparrow & \phi_1(\boldsymbol{r}_2)\xi_\downarrow & \ldots & \phi_{N/2}(\boldsymbol{r}_2)\xi_\downarrow \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\boldsymbol{r}_N)\xi_\uparrow & \phi_1(\boldsymbol{r}_N)\xi_\downarrow & \ldots & \phi_{N/2}(\boldsymbol{r}_N)\xi_\downarrow \end{vmatrix}. \tag{9}$$

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   Basis set

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

2.2.5   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.

2.2.6 CUSP condition

Until now, we have been fiddling with the wavefunction in multiple ways. Are we really allowed to fix it in whichever way we want? The answer is yes, as long as it satisfies the CUSP condition.

The CUSP condition states that

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,

$$\hat{H} = \sum_{i=1}^{P}(-\frac{1}{2}\nabla_i^2 + \frac{1}{2}\omega^2 r_i^2) + \sum_{i<j}\frac{1}{r_{ij}} \tag{10}$$

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

$$f_n(x) = H_n(x)\exp(-x^2/2) \tag{11}$$

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.

$$\hat{H} = \sum_{i=1}^{P}(-\frac{1}{2}\nabla_i^2 + \frac{1}{2}\omega^2\frac{1}{r_i}) + \sum_{i<j}\frac{1}{r_{ij}} \tag{12}$$

In atomic units.

2.3.3 Electron gas

2.3.4 Helium gas

He^3 and He^4

3 Methods

3.1 Configurations Interaction

3.2 Hartree-Fock

3.3 Variational Monte Carlo

Based on the variational principle

3.4 Coupled Cluster

4   Machine Learning

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

4.1   Statistics

4.1.1   Bayesian statistics

4.2   Supervised Learning

4.3   Unsupervised Learning

4.3.1   Boltzmann Machines

5 Implementation
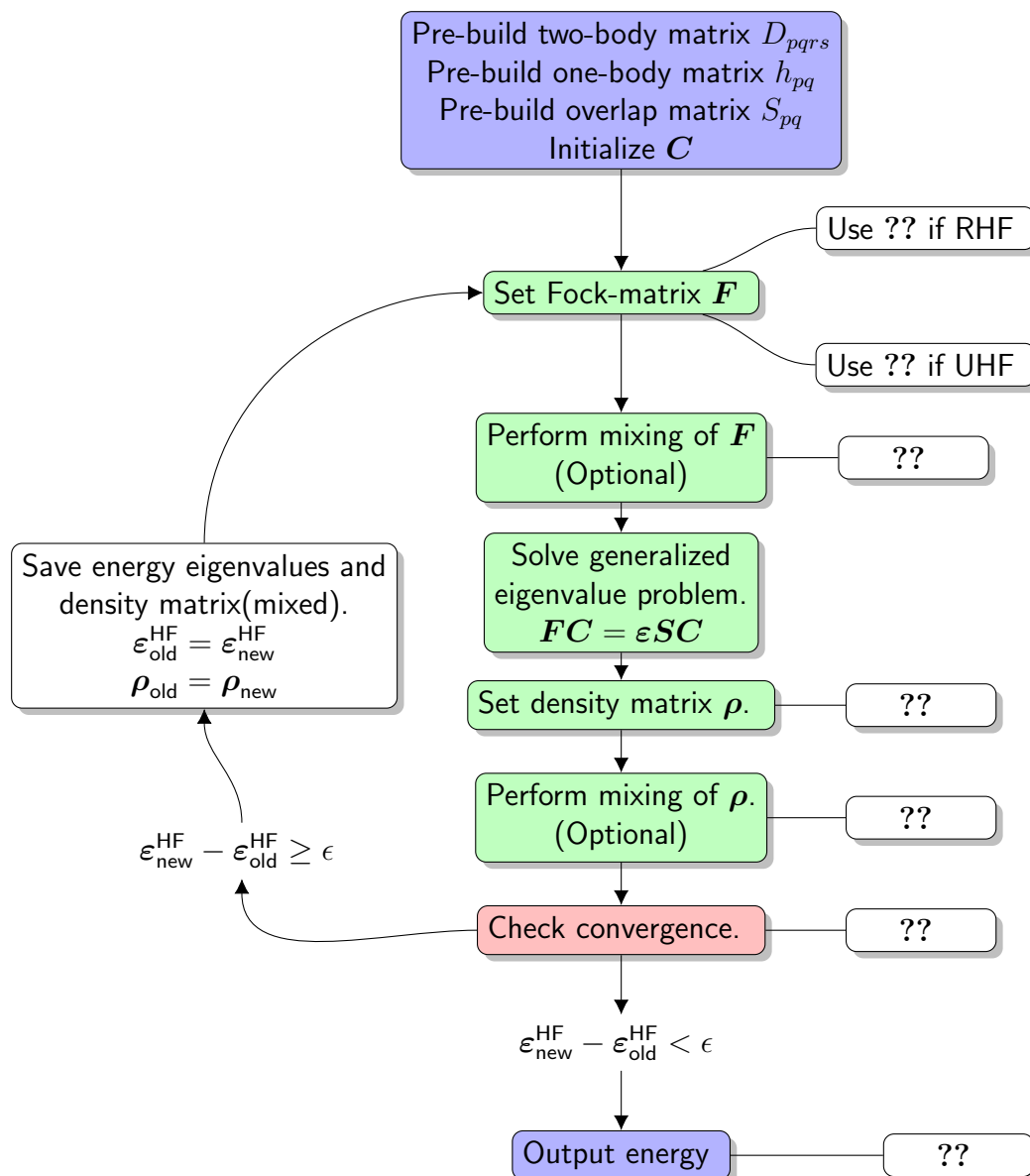
5.1 Foundation

5.2 Structure

Add structure chart

Figure 1: Hartree-Fock Algorithm.

5.3   Optimization algorithms

5.3.1   Stochastic Gradient Descent

5.3.2   ADAM

6 Results

7 Conclusion and future work

A Appendix A - RBM wavefunction

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

$$F(\boldsymbol{X}, \boldsymbol{H}) = \frac{1}{Z} e^{-\beta E(\boldsymbol{X}, \boldsymbol{H})} \tag{13}$$

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

$$E(\boldsymbol{X}, \boldsymbol{H}) = \sum_{i=1}^{M} \frac{(X_i - a_i)^2}{2\sigma_i^2} - \sum_{j=1}^{N} b_j H_j - \sum_{i,j=1}^{M,N} \frac{X_i W_{ij} H_j}{\sigma_i^2} \tag{14}$$

such that

$$F_{\mathrm{RBM}}(\boldsymbol{X}, \boldsymbol{h}) = e^{\sum_{i=1}^{M} \frac{(X_i - a_i)^2}{2\sigma^2}} e^{\sum_{j=1}^{N} \left(b_j h_j + \sum_{i=1}^{M} \frac{X_i W_{ij}}{\sigma^2} \right)}. \tag{15}$$

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 $\boldsymbol{h}$, $\{\boldsymbol{h}\}$:

$$
\begin{aligned}
F_{\mathrm{RBM}}(\boldsymbol{X}) &= \sum_{\{\boldsymbol{h}\}} e^{\sum_{i=1}^{M} \frac{(X_i - a_i)^2}{2\sigma^2}} \prod_{j=1}^{N} e^{b_j h_j + \sum_{i=1}^{M} \frac{X_i W_{ij} h_j}{\sigma^2}} \\
&= \sum_{h_1} \sum_{h_2} \ldots \sum_{h_N} e^{\sum_{i=1}^{M} \frac{(X_i - a_i)^2}{2\sigma^2}} \times \\
&\quad e^{b_1 h_1 + \sum_{i=1}^{M} \frac{X_i W_{i1} h_1}{\sigma^2}} e^{b_2 h_2 + \sum_{i=1}^{M} \frac{X_i W_{i2} h_2}{\sigma^2}} \ldots e^{b_N h_N + \sum_{i=1}^{M} \frac{X_i W_{iN} h_N}{\sigma^2}} \\
&= e^{\sum_{i=1}^{M} \frac{(X_i - a_i)^2}{2\sigma^2}} \prod_{j=1}^{N} \sum_{h_j=0}^{1} e^{b_j h_j + \sum_{i=1}^{M} \frac{X_i W_{ij} h_j}{\sigma^2}} \\
&= e^{\sum_{i=1}^{M} \frac{(X_i - a_i)^2}{2\sigma^2}} \prod_{j}^{N} (1 + e^{b_j + \frac{\boldsymbol{X}^T \boldsymbol{W}_{*j}}{\sigma^2}})
\end{aligned} \tag{16}
$$

B Appendix B - Dirac notation

References

[Gri05]   D.J. Griffiths. *Introduction to quantum mechanics.* 2nd Edition. Pearson PH, 2005. ISBN: 0131911759.