FYS4411 - Project 2
The restricted Boltzmannn machine applied to the quantum many body problem

Daniel Heinesen[1], Gunnar Lange[2] & Aram Salihi[2]
[1]Department of Theoretical Astrophysics, University of Oslo, N-0316 Oslo, Norway
[2]Department of Physics, University of Oslo, N-0316 Oslo, Norway

May 10, 2018

Abstract

NICE ABSTRACT

Contents

1 Introduction

COOL INTRO

2 Theory

Our system of P electrons confined in an isotropic harmonic oscillator potential is described by the Hamiltonian:

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

Where this is in natural units, i.e. $\hbar = c = e = m_e = 1$. Here ω is the frequency of the oscillator trap and $1/r_{ij}$ is the distance between particle i and j. The first term describes the interaction of the electrons with the potential, and we refer to it as the noninteracting part. The second term describes the Coloumb interaction between the electrons, and we refer to it as the interacting part.

We now wish to find the energy, E of our system by solving the time-independent Schrödinger equation:

$$H\Psi = E\Psi \tag{2}$$

2.1 Analytic expression

We are chiefly interested in the ground-state energy, as this is the state that our system will fall into if unperturbed by e.g. thermal excitation. If we ignore the interacting part, then we have an analytic solution to the ground-state energy for two electrons, given by:

$$\Psi(\mathbf{r}_1, \mathbf{r}_2) = C \exp\left(-\omega \frac{\boldsymbol{r_1}^2 + \boldsymbol{r_2}^2}{2} \right) \tag{3}$$

Where C is a normalization constant. Furthermore, the ground-state energy in such a harmonic oscillator is, according to **REFERENCE**, given by:

$$E = P\frac{D}{2} \tag{4}$$

Where D is the number of dimensions and P is the number of particles. **FINISH - DISCUSS SPIN, TWO PARTICLES, ANALYTIC SOLUTION FOR SIMPLE INTERACTION**

2.2 The variational principle

In the more complicated case of interacting fermions, We achieve such an upper bound by invoking the variational principle, as formulated in **GRIFFITHS**, which states that:

$$E_0 \leq \frac{\langle \Psi_T | H | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle} \tag{5}$$

Where E_0 is the ground-state energy and Ψ_T is any (not necessarily normalized) wavefunction. Our approach is therefore to choose a trial wavefunction, Ψ_T, and the compute $E' = \langle \Psi_T | H | \Psi_T \rangle$ by using a variational Monte Carlo approach. The variational method then guarantees that our approximation to the energy, E' will be larger than the ground state energy E_0.

2.3 Neural-network quantum states

We follow the approach outlined by **REFERENCE** to determine our trial wavefunction. They choose to interpret the problem of solving equation 2 as a multi-dimensional minimization and feature extraction problem. We therefore employ a neural network to find the optimal wavefunction.

Thus, we represent our wavefunction by a set of parameters, which we then seek to optimize. We choose to follow **REFERENCE** and choose a restricted Boltzmann machine (RBM) appproach to this neural network.

2.4 Restricted Boltzmann Machines

Restricted Boltzmann Machines (RBM) are described in great detail in the literature, see e.g. **REFERENCE**. We therefore only outline how to apply them. The basic idea consists of setting up a vector of M visible nodes $\mathbf{X}$, which corresponds to the position states of our system (with one state per particle, P and dimension, D, i.e. $M = PD$), and a vector of N hidden nodes, $\mathbf{H}$ which correspond to some feature of our system **DANIEL CHECK THIS PLEASE**. Each of these nodes have weights associated with them. Furthermore, there are also weights coupling the hidden nodes to the visible nodes, but there are no connections between nodes in the same layer (hence *restricted* Boltzmann machine). We then model the probability distribution for our particles, given a position $\mathbf{X}$ and a set of hidden nodes $\mathbf{H}$. by a Boltzmann distribution function of the form:

$$F_{rbm}(\mathbf{X}, \mathbf{H}) = \frac{1}{Z} \exp\left(-\frac{1}{T_0} E(\mathbf{X}, \mathbf{H})\right) \tag{6}$$

Where Z is a normalization constant, T_0 is some reference temperature (set to 1 in what follows), and the energy is chosen to represent a continuous function of the position, given as:

$$E(\mathbf{X}, \mathbf{H}) = \sum_{i=1}^{M} \frac{(X_i - a_i)^2}{2\sigma^2} - \sum_{j=1}^{N} b_j H_j - \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{X_i w_{ij} H_j}{\sigma^2} \tag{7}$$

Where a_i are the weights associated with the visible nodes, b_i are the weights associated with the visible nodes and w_{ij} are the weights of the coupling between the visible and hidden nodes. The σ models the spread of the energy.

As our wavefunction is a function of the position coordinates only, we choose to interpret the marginal distribution of the visible nodes as our wavefunction. This marginal distribution is given by:

$$F_{rbm}(\mathbf{X}) = \sum_{h_i \in \mathbf{H}} F_{rbm}(\mathbf{X}, \mathbf{H}) \tag{8}$$

Which gives, after some algebra done in **REFERENCE**, our wavefunction as:

$$\Psi(\mathbf{X}) = F_{rbm}(\mathbf{X}) = \frac{1}{Z} \exp\left(-\sum_{i=1}^{M} \frac{(X_i - a_i)^2}{2\sigma^2}\right) \prod_{j=1}^{N} \left(1 + \exp\left(b_j + \sum_{i=1}^{M} \frac{X_i w_{ij}}{\sigma^2}\right)\right) \tag{9}$$

This is therefore our trial wavefunction that we use in equation 5 to find an upper bound on the ground-state energy of our system. We seek to optimize this wavefunction with respect to the weights a_i, b_i and w_{ij}. We also investigate how this energy behaves as a function of σ, as well as for different dimensions D and particles P.

3 Methods

MUCHO METHOD

4 Results

GROUNDBREAKING RESULTS

5 Discussion

EXCELLENT DISCUSSION

6 Conclusion

BRILLIANT CONCLUSION

6.1 Outlook

MAGNIFICIENT OUTLOOK

References

Appendices

A Finding the derivatives

In this section, we derive the expressions for the various derivatives stated in section **SECTION**. We begin with the local energy.

A.1 The local energy

The local energy is given by:

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

Thus, we must compute:

$$\frac{1}{\Psi}\nabla_i^2\Psi \tag{11}$$

This may be rewritten as:

$$\frac{1}{\Psi}\nabla\left(\Psi\frac{1}{\Psi}\nabla\Psi\right) = \left(\frac{1}{\Psi}\nabla\Psi\right)^2 + \nabla\left(\frac{1}{\Psi}\nabla\Psi\right) = \left[\nabla\log\Psi\right]^2 + \nabla^2\log\Psi \tag{12}$$

The logarithm of our trial wavefunction is given by:

$$\log\Psi = -\log Z - \sum_{i=1}^{M}\left(\frac{(X_i - a_i)^2}{2\sigma^2}\right) + \sum_{j=1}^{N}\log\left(1 + \exp\left(b_j + \sum_{i=1}^{M}\frac{X_i w_{ij}}{\sigma^2}\right)\right) \tag{13}$$

The derivative with respect to one coordinate is now given by:

$$\begin{aligned}\frac{\partial\log\Psi}{\partial X_k} &= \frac{(a_k - X_k)}{\sigma^2} + \sum_{j=1}^{N}\frac{w_{kj}\exp\left(b_j + \sum_{i=1}^{M}\frac{X_i w_{ij}}{\sigma^2}\right)}{\sigma^2\left(1 + \exp\left(b_j + \sum_{i=1}^{M}\frac{X_i w_{ij}}{\sigma^2}\right)\right)} \\ &= \frac{a_k - X_k}{\sigma^2} + \sum_{j=1}^{N}\frac{w_{kj}}{\sigma^2\left(1 + \exp\left(-b_j - \sum_{i=1}^{M}\frac{X_i w_{ij}}{\sigma^2}\right)\right)}\end{aligned} \tag{14}$$

Whereas the second derivative is:

$$\frac{\partial^2\log\Psi}{\partial X_k^2} = -\frac{1}{\sigma^2} + \sum_{j=1}^{N}\frac{w_{kj}^2\exp\left(-b_j - \sum_{i=1}^{M}\frac{X_i w_{ij}}{\sigma^2}\right)}{\sigma^4\left(1 + \exp\left(-b_j - \sum_{i=1}^{M}\frac{X_i w_{ij}}{\sigma^2}\right)\right)^2} \tag{15}$$

The local energy can now be found by using equation 12, and inserting equation 14 and equation 15.

A.2 The derivatives with respect to the parameters

For our optimization method, we require:

$$\frac{1}{\Psi}\frac{\partial\Psi}{\partial\alpha_k} = \frac{\partial}{\partial\alpha_k}\log\Psi \tag{16}$$

Where α_k is any of the variational parameters a, b or w. These derivatives are given by:

$$\frac{\partial\log\Psi}{\partial a_k} = \frac{X_k - a_k}{\sigma^2} \tag{17}$$

$$\frac{\partial \log \Psi}{\partial b_k} = \frac{\exp\left(b_k + \sum_{i=1}^{M} \frac{X_i w_{ik}}{\sigma^2}\right)}{1 + \exp\left(b_k + \sum_{i=1}^{M} \frac{X_i w_{ik}}{\sigma^2}\right)} = \frac{1}{1 + \exp\left(-b_k - \sum_{i=1}^{M} \frac{X_i w_{ik}}{\sigma^2}\right)} \tag{18}$$

Finally, the derivative with respect to the weights, w_{kl} is given by:

$$\frac{\partial \log \Psi}{\partial w_{kl}} = \frac{X_k \exp\left(b_l + \sum_{i=1}^{M} \frac{X_i w_{il}}{\sigma^2}\right)}{\sigma^2\left(1 + \exp\left(b_l + \sum_{i=1}^{M} \frac{X_i w_{il}}{\sigma^2}\right)\right)} = \frac{X_k}{\sigma^2\left(1 + \exp\left(-b_l - \sum_{i=1}^{M} \frac{X_i w_{il}}{\sigma^2}\right)\right)} \tag{19}$$

A.3 The derivatives with Gibbs sampling

In Gibbs sampling, we represent the wavefunction as $\Psi = \sqrt{F_{rbm}}$, instead of $\Psi = F_{rbm}$. Note, however, that we only ever differentiate the logarithm of the wavefunction. As $\log \sqrt{\Psi} = \frac{1}{2} \log \Psi$, however, this is only a marginal change. Specifically:

$$E_L = [\nabla \log \sqrt{\Psi}]^2 + \nabla^2 \log \sqrt{\Psi} = \frac{1}{4}[\nabla \log \Psi]^2 + \frac{1}{2}\nabla^2 \log \Psi \tag{20}$$

$$\frac{\partial}{\partial \alpha_k} \log \sqrt{\Psi} = \frac{1}{2} \frac{\partial}{\partial \alpha_k} \Psi \tag{21}$$

$$\frac{\partial}{\partial b_k} \log \sqrt{\Psi} = \frac{1}{2} \frac{\partial}{\partial b_k} \log \Psi \tag{22}$$

$$\frac{\partial}{\partial w_{kl}} \log \sqrt{\Psi} = \frac{1}{2} \frac{\partial}{\partial w_{kl}} \log \Psi \tag{23}$$

Where the derivatives on the right-hand side are given in the previous section.

B Deriving the spin of the ground state wavefunction

The Pauli exclusion principle