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. Note that this expression is only valid up to $P = 2$, as we are dealing with spin-1/2 fermions, to which the Pauli exclusion principle applies. Therefore, we can only have two of them in the same energy state, and they must have opposite spin. More precisely, they must be in the singlet state (total spin 0) in the noninteracting case, as discussed in appendix **APPENDIX**. In what follows, we therefore restrict ourselves to $P \leq 2$.

We also have analytic results for some special cases in the interacting case. In particular, according to **REFERENCE**, the energy is 3 (in natural units) for 2 particles in the 2-dimensional case. **CHECK THIS**.

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 and Restricted Boltzmann Machines

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.

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:

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

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. **TALK ABOUT IF THIS CAN BE EXACT FOR NONINTERACTING**

2.4 Variational Monte Carlo methods

Having determined our trial wavefunction, it remains to use the variational principle to estimate the ground state energy. Our presentation here will closely follow that of **REFERENCE**. We use a variational Monte Carlo approach to estimate the inner product given in equation 5. To do

so, we consider the integrand function of this integral, given by:

$$P(\mathbf{X}) = \frac{|\Psi_T|^2}{\int d\mathbf{X}|\Psi_T|^2} \tag{10}$$

Where the integral is to be carried out over all particles and dimension (i.e. M times). If we now define the local energy (by solving equation 2 for E) as:

$$E_L(\mathbf{X}) = \frac{1}{\Psi_T} H\Psi_T \tag{11}$$

Then the expectation value of the Hamiltonian (which gives the energy) is given by:

$$E_0 \leq E[H] = \int d\mathbf{X} P(\mathbf{X}) E_L(\mathbf{X}) \approx \frac{1}{K} \sum_{i=1}^{K} E_{L,i} \tag{12}$$

Thus, we can get an approximation to the ground state energy if we can efficiently compute the local energy, and efficiently sample our space close to the optimal energy.

2.5 The local energy

The local energy is defined in equation 11, with the Hamiltonian given in equation 1. Note that the most complicated part of this expression is the second derivative of the wavefunction. We derive an analytic expression for this is in appendix **APPENDIX**. The local energy is then given by:

$$\frac{1}{\Psi_T} H\Psi_T = \sum_{k=1}^{M} -\frac{1}{2} \left[\frac{a_k - X_k}{\sigma^2} + \sum_{j=1}^{N} \frac{w_{kj}}{\sigma^2 \left(1 + \exp\left(-b_j - \sum_{i=1}^{M} \frac{X_i w_{ij}}{\sigma^2}\right)\right)} \right]^2$$
$$+ \sum_{k=1}^{M} \frac{1}{2\sigma^2} - \frac{1}{2} \sum_{k=1}^{M} \sum_{j=1}^{N} \frac{w_{kj}^2 \exp\left(-b_j - \sum_{i=1}^{M} \frac{X_i w_{ij}}{\sigma^2}\right)}{\sigma^4 \left(1 + \exp\left(-b_j - \sum_{i=1}^{M} \frac{X_i w_{ij}}{\sigma^2}\right)\right)^2} + \frac{1}{2} \sum_{i=1}^{M} \omega^2 X_i^2 + \sum_{i=1}^{P} \sum_{j=1}^{i} \frac{1}{r_{ij}} \tag{13}$$

2.6 Sampling the position space

To make sense of equation 12, we must sample our position space, to get a good estimate for the local energy of the system. We employ three different sampling techniques.

2.6.1 The Metropolis algorithm

The Metropolis algorithm is described in some detail **REFERENCE**. Note that the wavefunction gives the probability of finding the particle at point in position space. We will therefore get the most reliable estimates for our local energy if we sample where the wavefunction (and thus the probability distribution) is large, and we avoid wasting CPU cycles. The Metropolis algorithm is one way to achieve this. This is achieved by proposing to move a single particle according to:

$$\boldsymbol{r}_{\text{i,new}} = \boldsymbol{r}_{\text{i}} + \boldsymbol{\xi} dx \tag{14}$$

Where $\boldsymbol{r}_{\text{i}}$ is the position of particle i, dx is a chosen step size and $\boldsymbol{\xi}$ is a D-dimensional vector of random numbers distributed as $\xi_i \sim 0.5 \cdot \mathcal{N}(0,1)$. To find out whether we should sample local energy at this new position, we compute the ratio:

$$w = \frac{P(\boldsymbol{X}_{\text{new}})}{P(\boldsymbol{X})} = \frac{|\Psi(\boldsymbol{X}_{\text{new}})|^2}{|\Psi(\boldsymbol{X})|^2} \tag{15}$$

Where $\boldsymbol{X}_{\mathrm{new}}$ is the position of the particles after updating the position of particle i according to 14. Note that this ratio determines if the particle is more likely to be found at the position $\boldsymbol{X}_{\mathrm{new}}$ than at position $\boldsymbol{X}$. The move is only accepted if $w \geq \theta$, where $\theta \in [0, 1]$ is a uniformly distributed random variable. Thus, if the probability of finding the particle at $\mathbf{X}_{\mathrm{new}}$ is larger than the probability of finding the particle at $\mathbf{X}$, then we accept the move. If not, then the larger the energy difference the less probable it is that the move is accepted. As such, we should eventually end up moving fairly close to the most likely position, without getting stuck in it.

2.6.2 The Metropolis-Hasting alogrithm/importance sampling

We compare the Metropolis algorithm to an alternative, known as importance sampling. This is described in further detail in **REFERENCE**. Note that the Metropolis sampling described in the previous section picks positions at random, and as such does not have a preference for proposing steps that move the system to a region of high probability density (given by $|\Psi|^2$). As such, some CPU cycles will be wasted on position that a very small probability of occurring. Importance sampling seeks to alleviate this problem by introducing a quantum force, which "pushes" the system in the direction of high probability. This force is, on particle i given by:

$$F_i(\boldsymbol{X}) = \frac{2}{\Psi_T} \nabla_i \Psi_T = 2 \nabla_i \log \Psi_T \tag{16}$$

We compute this in appendix **APPENDIX** and show that this force is given by:

$$F_i(\boldsymbol{X}) = 2 \sum_{k=1}^{D} \left(\frac{a_k - X_k}{\sigma^2} + \sum_{j=1}^{N} \frac{w_{kj}}{\sigma^2 \left(1 + \exp\left(-b_j - \sum_{i=1}^{M} \frac{X_i w_{ij}}{\sigma^2} \right) \right)} \right) \tag{17}$$

The analysis in **REFERENCE** shows that the correct way to include this in our algorithm is by multiplying the transition probability w defined in equation 15 by a Green's function given by:

$$G(\mathbf{X}, \mathbf{X}') = \frac{1}{(4D\Delta t)^{3/2}} \sum_{i=1}^{M} \exp\left(-\frac{(X_i - X_i' - D\Delta t F_i(\mathbf{X}'))^2}{4D\Delta t} \right) \tag{18}$$

And the modified transition probability is given by:

$$w_{\mathrm{importance}} = \frac{|\Psi(\boldsymbol{X}_{\mathrm{new}})|^2}{|\Psi(\boldsymbol{X})|^2} \frac{G(\boldsymbol{X}, \boldsymbol{X}_{\mathrm{new}})}{G(\boldsymbol{X}_{new}, \boldsymbol{X})} \tag{19}$$

This transition probability is implemented exactly as in the previous section. However, we use a slightly different initialization and we draw the next step, $\boldsymbol{\xi} dx$ in a slightly different way when using importance sampling. This is described in further detail in **SECTION**.

2.6.3 Gibbs Sampling

2.7 Optimizing our wavefunction in parameter space

For our neural network to be successful, we must be able to minimize the energy of our wavefunction. This is done by varying the parameters a_i, b_i and w_{ij}, and computing the local energy. The variational method then guarantees that the lowest energy we find will be larger than the ground state energy. We employ a stochastic gradient method to find the minimum in our parameter space, as described in **REFERENCE**. Defining a vector of all parameters, $\boldsymbol{\theta}$, we iterate this vector according to:

$$\boldsymbol{\theta}^{i+1} = \boldsymbol{\theta}^i + \gamma \frac{\partial \overline{E}_L}{\partial \boldsymbol{\theta}^i} (\boldsymbol{\theta}^i) \tag{20}$$

Where $\overline{E}_L$ is the energy computed by means of the variational Monte Carlo method (given by equation 12) and γ is the step size, which we vary. The derivative is with respect to all parameters.

We showed in **REFERENCE** that the derivative of this expectation value of the energy is given by:

$$\frac{\partial \overline{E}_L}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}) = 2 \left(\left\langle \frac{1}{\Psi_T} \frac{\partial \Psi_T}{\partial \boldsymbol{\theta}} \overline{E}_L(\boldsymbol{\theta}) \right\rangle - \left\langle \frac{1}{\Psi_T} \frac{\partial \Psi_T}{\partial \boldsymbol{\theta}} \right\rangle \left\langle \overline{E}_L(\boldsymbol{\theta}) \right\rangle \right) \tag{21}$$

We must therefore compute the derivative of our trial wavefunction, given in equation 9, with respect to all of our variational parameters. This is done in detail in appendix **APPENDIX**, both for the case of Metropolis sampling and Gibbs sampling. We iterate this method until the absolute value of our gradient, $\frac{\partial \overline{E}_L}{\partial \boldsymbol{\theta}}$ is small enough **DANIEL ADD STUFF**.

2.8 Onebody densities

3 Methods

3.1 An overview of the structure of our program

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:

$$E_L = \frac{1}{\Psi}\hat{H}\Psi = \sum_{i=1}^{N}\left(-\frac{1}{2\Psi}\nabla_i^2\Psi + \frac{1}{2}\omega^2 r_i^2\right) + \sum_{i<j}\frac{1}{r_{ij}} \tag{22}$$

Thus, we must compute:

$$\frac{1}{\Psi}\nabla_i^2\Psi \tag{23}$$

This may be rewritten as:

$$\frac{1}{\Psi}\nabla\left(\Psi\frac{1}{\Psi}\nabla\Psi\right) = \left(\frac{1}{\Psi}\nabla\Psi\right)^2 + \nabla\left(\frac{1}{\Psi}\nabla\Psi\right) = [\nabla\log\Psi]^2 + \nabla^2\log\Psi \tag{24}$$

The logarithm of our trial wavefunction is given by:

$$\log\Psi = -\log Z - \sum_{i=1}^{M}\left(\frac{(X_i - a_i)^2}{2\sigma^2}\right) + \sum_{j=1}^{N}\log\left(1 + \exp\left(b_j + \sum_{i=1}^{M}\frac{X_i w_{ij}}{\sigma^2}\right)\right) \tag{25}$$

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

$$\begin{aligned}\frac{\partial\log\Psi}{\partial X_k} &= \frac{(a_k - X_k)}{\sigma^2} + \sum_{j=1}^{N}\frac{w_{kj}\exp\left(b_j + \sum_{i=1}^{M}\frac{X_i w_{ij}}{\sigma^2}\right)}{\sigma^2\left(1 + \exp\left(b_j + \sum_{i=1}^{M}\frac{X_i w_{ij}}{\sigma^2}\right)\right)} \\ &= \frac{a_k - X_k}{\sigma^2} + \sum_{j=1}^{N}\frac{w_{kj}}{\sigma^2\left(1 + \exp\left(-b_j - \sum_{i=1}^{M}\frac{X_i w_{ij}}{\sigma^2}\right)\right)}\end{aligned} \tag{26}$$

Whereas the second derivative is:

$$\frac{\partial^2\log\Psi}{\partial X_k^2} = -\frac{1}{\sigma^2} + \sum_{j=1}^{N}\frac{w_{kj}^2\exp\left(-b_j - \sum_{i=1}^{M}\frac{X_i w_{ij}}{\sigma^2}\right)}{\sigma^4\left(1 + \exp\left(-b_j - \sum_{i=1}^{M}\frac{X_i w_{ij}}{\sigma^2}\right)\right)^2} \tag{27}$$

The local energy can now be found by using equation 24, and inserting equation 26 and equation 27.

A.2 The derivatives with respect to the parameters

For our optimization method, we require:

$$\frac{1}{\Psi}\frac{\partial\Psi}{\partial\alpha_k} = \frac{\partial}{\partial\alpha_k}\log\Psi \tag{28}$$

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

$$\frac{\partial\log\Psi}{\partial a_k} = \frac{X_k - a_k}{\sigma^2} \tag{29}$$

8

$$\frac{\partial \log \Psi}{\partial b_k} = \frac{\exp\left(b_k + \sum_{i=1}^{M} \frac{X_i w_{ik}}{\sigma^2}\right)}{1 + \exp\left(b_k + \sum_{i=1}^{M} \frac{X_i w_{ik}}{\sigma^2}\right)} = \frac{1}{1 + \exp\left(-b_k - \sum_{i=1}^{M} \frac{X_i w_{ik}}{\sigma^2}\right)} \tag{30}$$

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

$$\frac{\partial \log \Psi}{\partial w_{kl}} = \frac{X_k \exp\left(b_l + \sum_{i=1}^{M} \frac{X_i w_{il}}{\sigma^2}\right)}{\sigma^2 \left(1 + \exp\left(b_l + \sum_{i=1}^{M} \frac{X_i w_{il}}{\sigma^2}\right)\right)} = \frac{X_k}{\sigma^2 \left(1 + \exp\left(-b_l - \sum_{i=1}^{M} \frac{X_i w_{il}}{\sigma^2}\right)\right)} \tag{31}$$

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:

$$E_L = [\nabla \log \sqrt{\Psi}]^2 + \nabla^2 \log \sqrt{\Psi} = \frac{1}{4}[\nabla \log \Psi]^2 + \frac{1}{2} \nabla^2 \log \Psi \tag{32}$$

$$\frac{\partial}{\partial \alpha_k} \log \sqrt{\Psi} = \frac{1}{2} \frac{\partial}{\partial \alpha_k} \Psi \tag{33}$$

$$\frac{\partial}{\partial b_k} \log \sqrt{\Psi} = \frac{1}{2} \frac{\partial}{\partial b_k} \log \Psi \tag{34}$$

$$\frac{\partial}{\partial w_{kl}} \log \sqrt{\Psi} = \frac{1}{2} \frac{\partial}{\partial w_{kl}} \log \Psi \tag{35}$$

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

B   Deriving the spin of the ground state wavefunction

Consider the unperturbed wavefunction for the ground state of the two electron system, given by:

$$\psi(\boldsymbol{r}_1, \boldsymbol{r}_2) = C \exp\left(-\omega \left(r_1^2 + r_2^2\right)/2\right) \tag{36}$$

Since this unperturbed wavefunction are symmetric under permutation of coordinate. Pauli exclusion principle then states that the spin wavefunction must then be a antisymmetric function in order to achieve total spin 0 under permutation of coordinate. This wave function must then be singlet, and will look like:

$$|\psi\rangle = \frac{1}{\sqrt{2}} \left(|\uparrow\rangle - |\downarrow\rangle\right) \tag{37}$$

We will now apply the spin operator $\boldsymbol{S}^2$ on this spin state to prove that total spin is 0:

$$\boldsymbol{S}^2 |\psi\rangle = \frac{1}{\sqrt{2}} \left(\underbrace{\boldsymbol{S}^2 |\uparrow\rangle}_{=\frac{3}{4}\hbar^2} - \underbrace{\boldsymbol{S}^2 |\downarrow\rangle}_{=\frac{3}{4}\hbar^2}\right) = 0 \tag{38}$$