

# FYS4411 - Project 2

## The restricted Boltzmann machine applied to the quantum many body problem

Daniel Heinesen<sup>1</sup>, Gunnar Lange<sup>2</sup> & Aram Salih<sup>2</sup>

<sup>1</sup>Department of Theoretical Astrophysics, University of Oslo, N-0316 Oslo, Norway

<sup>2</sup>Department of Physics, University of Oslo, N-0316 Oslo, Norway

May 11, 2018

**Abstract**

NICE ABSTRACT

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Theory</b>	<b>3</b>
2.1	The Hamiltonian . . . . .	3
2.2	Analytic solution . . . . .	3
2.3	The variational principle . . . . .	4
2.4	Neural-network quantum states and Restricted Boltzmann Machines . . . . .	4
2.5	Variational Monte Carlo methods . . . . .	5
2.6	The local energy . . . . .	6
2.7	Sampling the position space . . . . .	6
2.8	Optimizing our wavefunction in parameter space . . . . .	7
2.9	Onebody densities . . . . .	7
<b>3</b>	<b>Methods</b>	<b>7</b>
3.1	An overview of the structure of our program . . . . .	7
3.2	Initializing our system . . . . .	8
3.3	Proposing a new position . . . . .	9
3.4	Implementing the optimization of the weights . . . . .	9
3.5	The errors in our energy . . . . .	9
3.6	Tuning the parameters . . . . .	10
3.7	Benchmarking our results . . . . .	10
3.8	The onebody density . . . . .	11
<b>4</b>	<b>Results</b>	<b>11</b>
<b>5</b>	<b>Discussion</b>	<b>11</b>
<b>6</b>	<b>Conclusion</b>	<b>11</b>
6.1	Outlook . . . . .	11
	<b>Appendices</b>	<b>12</b>
<b>A</b>	<b>Finding the derivatives</b>	<b>12</b>
A.1	The local energy . . . . .	12
A.2	The derivatives with respect to the parameters . . . . .	12
A.3	The derivatives with Gibbs sampling . . . . .	13
<b>B</b>	<b>Deriving the spin of the ground state wavefunction</b>	<b>13</b>

# 1 Introduction

There interest in applying machine learning methods to quantum mechanical system has experienced a veritable surge in recent years, and a host of different methods are currently being explored (see e.g. **REFERENCE**). A particularly interesting application is presented by **REFERENCE**. They realize that the quantum many-body wavefunction is a system of exponential complexity, where we are interested in extracting a few crucial features, such as the energy. In this context, machine learning and neural networks is a natural way to attack the problem, as these are methods designed for extracting features highly complex systems. They present an approach titled neural-network quantum states, where they use restricted Boltzmann machines to represent the wavefunction.

We apply their method to a system of fermions in an isotropic harmonic oscillator trap with Coloumb repulsion. This system has the advantage of being well-studied, as both analytic (**REFERENCE**) and numeric (**REFERENCE**) results are available for some system configurations. This enables us to benchmark our results, and compare the performance of neural-network quantum states to other algorithms.

We focus on the ground state of our system, and therefore only investigate the case of 1 and 2 electrons, in 1, 2 and 3 dimensions. We use restricted Boltzmann machines to attain the form of our wavefunction, and the use Variational Monte Carlo (VMC) methods to establish an upper bound on the ground state energy. We use a stochastic gradient approach with a fixed learning rate to minimize the weights in our Restricted Boltzmann machine. This is then compared using three different sampling methods in our VMC calculations - Metropolis sampling, importance sampling and Gibbs sampling.

We investigate how our system behaves as we vary the parameters associated with the system, and as we vary the sampling method used.s

## 2 Theory

### 2.1 The Hamiltonian

We consider a  $D$ -dimensional system of  $P$  electrons confined in an isotropic harmonic oscillator trap. Working in natural units ( $\hbar = c = e = m_e = 1$ ), the Hamiltonian of such a system is given by:

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

Here  $\omega$  is the frequency of the oscillator trap,  $r_i$  is the  $D$ -dimensional position vector of electron  $i$  and  $1/r_{ij}$  is the distance between electron  $i$  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:

$$H\Psi = E\Psi \quad (2)$$

Where  $\Psi$  is the wavefunction of the system.

### 2.2 Analytic solution

We are chiefly interested in the ground-state energy of our system, as this is the state that the system will fall into if left unperturbed by e.g. thermal excitation. If we ignore the interacting

part, then we have an analytic solution to the ground-state wavefunction  $P$  two electrons, which according to **REFERENCE** is given by:

$$\psi(r_1, \dots, r_P) = C \exp \left( -\frac{\omega}{2} \sum_{i=1}^P r_i^2 \right) \quad (3)$$

Where  $C$  is a normalization constant. Furthermore, the ground-state energy (in natural units) of such a system is, according to **REFERENCE**, given by:

$$E = P \frac{D}{2} \quad (4)$$

Note, however, that all of the above expressions are 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 particles in the ground-state, and these particles 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  $P \in \{1, 2\}$ .

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 with  $\omega = 1$ . **CHECK THIS**.

### 2.3 The variational principle

In the general case, where analytic solutions are not available, it is desirable to use a numeric method to find the ground state energy, or at least an upper bound on the ground state energy. We achieve such an upper bound by invoking the variational principle, as formulated in **GRIFFITHS**, which states that:

$$E_0 \leq \frac{\langle \Psi_T | H | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle} \quad (5)$$

Where  $E_0$  is the ground-state energy and  $\Psi_T$  is any (not necessarily normalized) trial wavefunction for our system. We therefore choose a class of trial wavefunctions,  $\Psi_T$ , and then 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  $E_0$ . We then adjust our trial wavefunction within our class of wavefunctions, to give the lowest possible energy, hoping to attain a decent estimate for the actual ground-state energy of our system.

### 2.4 Neural-network quantum states and Restricted Boltzmann Machines

To actually determine our class of trial wavefunction, we follow the approach outlined by **REFERENCE**. They choose to interpret the problem of solving equation 2 as a multi-dimensional minimization and feature extraction problem. As neural networks were designed to tackle exactly this class of problem, we choose to employ a neural network to find our trial wavefunction.

We follow the approach advocated by **REFERENCE** and choose a restricted Boltzmann machine (RBM) as the basis for our neural network. Restricted Boltzmann Machines (RBM) are described in great detail in the literature, see e.g. **REFERENCE**. We therefore provide only a rough outline on how to apply them. We begin by 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 that we wish to use to predict our wavefunction. **DANIEL CHECK THIS PLEASE**. Each of these nodes have weights associated with them. Additionally, 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:

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

Where  $Z$  is a normalization constant and  $T_0$  is some reference temperature (set to 1 in what follows). The energy,  $E(\mathbf{X}, \mathbf{H})$  relates the weights of each node to the probability distribution. As  $\mathbf{X}$  is a continuous variable, we choose the form of the energy to be a Gaussian-binary, given by:

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

Where  $a_i$  is the weight associated with the visible node  $X_i$ ,  $b_i$  is the weight associated with the hidden node  $H_i$  and  $w_{ij}$  is the weight associated with the coupling between  $X_i$  and  $H_j$ . The  $\sigma$  models the spread of the energy, and determines how the weights play into the energy **DANIEL PLEASE**.

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

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

Which gives, after some algebra done in **REFERENCE**, our class of trial wavefunctions as:

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

This trial wavefunction, given weights  $a, b, w$ , is then plugged into equation 5 to find an upper bound on the ground-state energy of our system given the weights. We seek to optimize this wavefunction with respect to these weights. **TALK ABOUT IF THIS CAN BE EXACT FOR NONINTERACTING**

## 2.5 Variational Monte Carlo methods

Having established our trial wavefunction, it remains to use the variational principle to estimate the ground state energy. We use a variational Monte Carlo approach to estimate the inner product given in equation 5. A more detailed description of this may be found in **REFERENCE**, and we therefore only sketch a procedure. We begin by considering the probability distribution associated with our trial wavefunction, given by:

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

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:

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

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

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

Thus, we can get an approximation to the ground state energy if we sample our energy  $K$  times. This is contingent upon our ability to efficiently compute the local energy, and efficiently sample our position space.

## 2.6 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 in appendix **APPENDIX**. The local energy is then given by:

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

## 2.7 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.

### 2.7.1 The Metropolis algorithm

The Metropolis algorithm is described in further detail in **REFERENCE**. We begin by noting that the wavefunction gives the probability of finding the particle at a 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:

$$\mathbf{r}_{i,\text{new}} = \mathbf{r}_i + \boldsymbol{\xi} dx \quad (14)$$

Where  $\mathbf{r}_i$  is the position of particle  $i$ ,  $dx$  is a chosen step size and  $\boldsymbol{\xi}$  is a D-dimensional vector of random numbers, whose distribution we discuss in section **SECTION**. To find out whether we should sample local energy at this new position, we compute the ratio:

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

Where  $\mathbf{X}_{\text{new}}$  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  $\mathbf{X}_{\text{new}}$  than at position  $\mathbf{X}$ . 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  $\mathbf{X}_{\text{new}}$  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.

### 2.7.2 The Metropolis-Hasting algorithm/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 positions that have 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 changes the way a new step is proposed, as discussed in section **SECTION**. This force is, on particle  $i$  given by:

$$F_i(\mathbf{X}) = \frac{2}{\Psi_T} \nabla_i \Psi_T = 2 \nabla_i \log \Psi_T \quad (16)$$

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

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

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:

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

Where  $\mathcal{D}$  is a diffusion parameter, which in our case equals 1/2. The modified transition probability is given by:

$$w_{\text{importance}} = \frac{|\Psi(\mathbf{X}_{\text{new}})|^2}{|\Psi(\mathbf{X})|^2} \frac{G(\mathbf{X}, \mathbf{X}_{\text{new}})}{G(\mathbf{X}_{\text{new}}, \mathbf{X})} \quad (19)$$

This transition probability is implemented exactly as in the previous section.

### 2.7.3 Gibbs Sampling

## 2.8 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:

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

Where  $\bar{E}_L$  is the energy computed by means of the variational Monte Carlo method (given by equation 12) and  $\gamma$  is the step size, which we vary. The derivative is with respect to all parameters. It is shown in **REFERENCE** that the derivative of this expectation value of the energy is given by:

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

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,  $\frac{\partial \bar{E}_L}{\partial \boldsymbol{\theta}}$  is small enough **DANIEL ADD STUFF**.

## 2.9 Onebody densities

The onebody density describes the probability of finding a particle at a distance  $r$  from the origin of our position space. We compute this by looking at the histogram of particle positions. This is described in further detail in section **SECTION**.

# 3 Methods

## 3.1 An overview of the structure of our program

Our program may be found on our **GITHUB**. It is an extension of the program presented in **REFERENCE**. The program consists of 3 main classes:

**main.cpp** The main class first reads in the parameter files. It then sets up the initial guess for the weights of the neural networks and runs the stochastic descent method. Finally, it runs the full simulation with the optimized weights.

**simulation.cpp** This class contains the minimization method itself, and also runs the main Monte Carlo loop.

**system.cpp** The system class contains all the functions for the wavefunction, the different sampling methods, and the acceptance check. It also contains the function for computing the local energy and the derivatives of the wavefunction as a function of the various parameters.

In addition, we implement various auxiliary classes, that read in parameters, dump variables to file and do statistical analysis.

## 3.2 Initializing our system

### 3.2.1 Initializing the position

To initialize our system, we must distribute our particles. How we distribute the particles, however, depends on the sampling method employed. When using the Metropolis algorithm or Gibbs sampling, we distribute our particles according to:

$$\mathbf{r}_i = \boldsymbol{\xi} \quad (22)$$

Where  $\boldsymbol{\xi}$  is a vector of  $D$  number, uniformly distributed in  $[0, 0.5]$  as we find that this gives a good spread.

When using importance sampling, however, we must distribute our particles slightly differently, as discussed in **REFERENCE**. In this case, the initial distribution of particles is given by:

$$\mathbf{r}_i = \boldsymbol{\xi}' \sqrt{dx} \quad (23)$$

Where  $\boldsymbol{\xi}'$  is a vector of  $D$  number distributed as  $\mathcal{N}(0, 1)$  and  $dx$  is a step size which we tune, as described further in section **SECTION**.

### 3.2.2 Initializing the parameters

We wish to minimize our wavefunction as a function of the weights  $a$ ,  $b$  and  $w$ . This is done by iterating equation 20. However, we require an initial guess for the vector of parameters,  $\boldsymbol{\theta}^0$ . We choose this initial guess to be uniformly distributed in  $[-0.01, 0.01]$ .

### 3.2.3 Initializing the simulation

To initialize the simulation, we set the parameters in the parameter file found on our **GITHUB**. This gives the option to set a number of parameters of the system. Specifically, the parameter file allows the following parameters to be set:

- The number of Monte Carlo cycles
- The number of particles,  $P$
- The number of hidden nodes,  $N$
- The number of dimensions,  $D$
- The oscillator frequency  $\omega$



- The spread in the visible nodes,  $\sigma$
- The diffusion parameter  $D$  (setting this to 0 deactivates importance sampling)
- The step size,  $dx$
- The learning rate,  $\gamma$
- Whether or not the system is interacting
- Whether the energy should be computed numerically or analytically
- Whether or not to use Gibbs sampling

### 3.3 Proposing a new position

As discussed in section **SECTION**, both the Metropolis algorithm and the Metropolis-Hasting algorithm rely on proposing a new position of the system. The general form for this updating was given in equation 14. We now discuss how this is concretely implemented in our program.

For straightforward Metropolis sampling, there is no bias in selecting this position. We therefore simply define the new position by:

$$\mathbf{r}_{i,\text{new}} = \mathbf{r}_i + \boldsymbol{\xi} dx \quad (24)$$

Where  $\boldsymbol{\xi}$  consists of  $D$  numbers, uniformly distributed in  $[-1, 1]$  and  $dx$  is a step size, which we tune for Metropolis sampling and Gibbs sampling separately, as discussed in **SECTION**.

For importance sampling, we add a bias in selecting the new position. This bias is introduced by including the quantum force, presented in equation 17. The details of how to implement this bias are presented in **REFERENCE**. This gives that the new step should be selected according to:

$$\mathbf{r}_{i,\text{new}} = \mathbf{r}_i + \boldsymbol{\xi}' \sqrt{dx} + \mathcal{D} \mathbf{F}_i dx \quad (25)$$

Where  $F_i$  is the quantum force on particle  $i$ ,  $\mathcal{D}$  is the diffusion parameter (equal to  $1/2$ ) and  $\boldsymbol{\xi}' \sim \mathcal{N}(0, 1)$ . Note that  $dx$  may in general not be the same for the two different methods, and we investigate which  $dx$  gives the best values for each case separately, as discussed in section **SECTION**.

### 3.4 Implementing the optimization of the weights

To find the best trial wavefunction, we run the optimization algorithm given in equation 20 until the absolute value of the gradient is less than  $\epsilon = 10^{-5}$ , or we have surpassed 100 iterations. We use 100 times fewer Monte Carlo cycles to compute the local energy in the optimization loop than in the actual simulation used to determine the energy.

### 3.5 The errors in our energy

We wish to provide error bounds on the energies that we find. Note that the local energies constitute a highly correlated system, as the state of the system in the current Monte Carlo step differs only by the position of a single particle from the state of the system in the previous Monte Carlo step. The variance of the energy is therefore not a good indicator of the error in our system. Instead, we use a technique known as blocking. This is described in further detail in **REFERENCE**, and we follow the same procedure as outline in **REFERENCE**.

Blocking works by averaging successive observations of the local energy. Such a pairwise operation preserves the variance of the data, but reduces the covariance of the data. We keep doing this until the covariance is small enough, as described in detail in **REFERENCE**. To ensure that we can this as frequently as possible, we let our number of Monte Carlo cycles be a power of 2.

### 3.6 Tuning the parameters

We note that our methods are contingent upon a slew of parameters, in addition to the weights that our minimization algorithm tunes. In particular, we must choose the learning rate,  $\gamma$ , the step size  $dx$ , the number of hidden nodes  $N$ , the spread in the visible nodes,  $\sigma$  and the number of Monte Carlo cycles.

We optimize these parameters separately, as they ultimately should only determine if our method converges, and how fast it converges. Note that these parameters are not directly related to the physical parameters of our system, and that only  $N$  actually appears in our trial wavefunction.

#### 3.6.1 Tuning the step size, $dx$

The step size  $dx$  determines how well we are able to find the local energy. To find the best possible step size  $dx$ , we investigate how fast the energy in our system stabilizes, by looking at the evolving variance as a function of the number of Monte Carlo cycles. We check this for the simplest possible system (1 particle in 1 dimension), and always check that the energy actually converges to the analytic value predicted by equation 4. We then choose the step size where the variance is quickest to stabilize. We do this separately for Metropolis sampling, importance sampling and Gibbs sampling. We then use this  $dx$  in both the noninteracting case and in the interacting case.

#### 3.6.2 Tuning the learning rate, $\gamma$ and the number of hidden nodes, $N$

We choose to tune the learning rate  $\gamma$  and number of hidden nodes  $N$  simultaneously, as both determine how fast our minimization method converges. We choose representative values for both parameters, and look at the number of iterations required for our minimization method to converge in the simplest case (1 particle and 1 dimension). We then choose the combination of parameters that gives the quickest convergence, always ensuring that we actually reproduce the correct energy. We again do this separately for Metropolis sampling, importance sampling and Gibbs sampling. We then use this value of  $\gamma$  and  $N$  for the interacting and noninteracting case for both 1 and 2 particles.

#### 3.6.3 Choosing the number of Monte Carlo steps

Note that the numbers of Monte Carlo steps required will depend on the method used to sample the space, as e.g. importance sampling will accept more steps and therefore require fewer cycles. As we are limited to 2 particles, however, we can afford to use a fairly large number of Monte Carlo cycles. We therefore choose to use enough Monte Carlo steps to ensure convergence in the uncertainty in the energy, and use the same number of Monte Carlo cycles independent of the sampling.

#### 3.6.4 Varying the spread in the visible nodes, $\sigma$

Finally, we also investigate how our system depends on the value of  $\sigma$ . **FINISH**

### 3.7 Benchmarking our results

Having chosen the parameters of our system, we can now benchmark our results. We look at the ground state energy for 1 and 2 particles, in 1,2 and 3 dimensions, both with and without interaction. Note that we have analytic results for the noninteracting case, given by equation 4, and that we have analytic results in the interacting case with 2 dimension, given by **REFERENCE**.

Note, that the variational principle only guarantees that we attain an upper bound on the ground state energy. As such, we can only check that our result is larger than or equal to the analytic result.

### **3.8 The onebody density**

## **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:

$$E_L = \frac{1}{\Psi} \hat{H} \Psi = \sum_{i=1}^N \left( -\frac{1}{2\Psi} \nabla_i^2 \Psi + \frac{1}{2} \omega^2 r_i^2 \right) + \sum_{i < j} \frac{1}{r_{ij}} \quad (26)$$

Thus, we must compute:

$$\frac{1}{\Psi} \nabla_i^2 \Psi \quad (27)$$

This may be rewritten as:

$$\frac{1}{\Psi} \nabla \left( \Psi \frac{1}{\Psi} \nabla \Psi \right) = \left( \frac{1}{\Psi} \nabla \Psi \right)^2 + \nabla \left( \frac{1}{\Psi} \nabla \Psi \right) = [\nabla \log \Psi]^2 + \nabla^2 \log \Psi \quad (28)$$

The logarithm of our trial wavefunction is given by:

$$\log \Psi = -\log Z - \sum_{i=1}^M \left( \frac{(X_i - a_i)^2}{2\sigma^2} \right) + \sum_{j=1}^N \log \left( 1 + \exp \left( b_j + \sum_{i=1}^M \frac{X_i w_{ij}}{\sigma^2} \right) \right) \quad (29)$$

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

$$\begin{aligned} \frac{\partial \log \Psi}{\partial X_k} &= \frac{(a_k - X_k)}{\sigma^2} + \sum_{j=1}^N \frac{w_{kj} \exp \left( b_j + \sum_{i=1}^M \frac{X_i w_{ij}}{\sigma^2} \right)}{\sigma^2 \left( 1 + \exp \left( b_j + \sum_{i=1}^M \frac{X_i w_{ij}}{\sigma^2} \right) \right)} \\ &= \frac{a_k - X_k}{\sigma^2} + \sum_{j=1}^N \frac{w_{kj}}{\sigma^2 \left( 1 + \exp \left( -b_j - \sum_{i=1}^M \frac{X_i w_{ij}}{\sigma^2} \right) \right)} \end{aligned} \quad (30)$$

Whereas the second derivative is:

$$\frac{\partial^2 \log \Psi}{\partial X_k^2} = -\frac{1}{\sigma^2} + \sum_{j=1}^N \frac{w_{kj}^2 \exp \left( -b_j - \sum_{i=1}^M \frac{X_i w_{ij}}{\sigma^2} \right)}{\sigma^4 \left( 1 + \exp \left( -b_j - \sum_{i=1}^M \frac{X_i w_{ij}}{\sigma^2} \right) \right)^2} \quad (31)$$

The local energy can now be found by using equation 28, and inserting equation 30 and equation 31.

### A.2 The derivatives with respect to the parameters

For our optimization method, we require:

$$\frac{1}{\Psi} \frac{\partial \Psi}{\partial \alpha_k} = \frac{\partial}{\partial \alpha_k} \log \Psi \quad (32)$$

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

$$\frac{\partial \log \Psi}{\partial a_k} = \frac{X_k - a_k}{\sigma^2} \quad (33)$$

$$\frac{\partial \log \Psi}{\partial b_k} = \frac{\exp\left(b_k + \sum_{i=1}^M \frac{X_i w_{ik}}{\sigma^2}\right)}{1 + \exp\left(b_k + \sum_{i=1}^M \frac{X_i w_{ik}}{\sigma^2}\right)} = \frac{1}{1 + \exp\left(-b_k - \sum_{i=1}^M \frac{X_i w_{ik}}{\sigma^2}\right)} \quad (34)$$

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

$$\frac{\partial \log \Psi}{\partial w_{kl}} = \frac{X_k \exp\left(b_l + \sum_{i=1}^M \frac{X_i w_{il}}{\sigma^2}\right)}{\sigma^2 \left(1 + \exp\left(b_l + \sum_{i=1}^M \frac{X_i w_{il}}{\sigma^2}\right)\right)} = \frac{X_k}{\sigma^2 \left(1 + \exp\left(-b_l - \sum_{i=1}^M \frac{X_i w_{il}}{\sigma^2}\right)\right)} \quad (35)$$

### 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:

$$E_L = [\nabla \log \sqrt{\Psi}]^2 + \nabla^2 \log \sqrt{\Psi} = \frac{1}{4} [\nabla \log \Psi]^2 + \frac{1}{2} \nabla^2 \log \Psi \quad (36)$$

$$\frac{\partial}{\partial \alpha_k} \log \sqrt{\Psi} = \frac{1}{2} \frac{\partial}{\partial \alpha_k} \log \Psi \quad (37)$$

$$\frac{\partial}{\partial b_k} \log \sqrt{\Psi} = \frac{1}{2} \frac{\partial}{\partial b_k} \log \Psi \quad (38)$$

$$\frac{\partial}{\partial w_{kl}} \log \sqrt{\Psi} = \frac{1}{2} \frac{\partial}{\partial w_{kl}} \log \Psi \quad (39)$$

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:

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = C \exp\left(-\omega (r_1^2 + r_2^2) / 2\right) \quad (40)$$

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:

$$|\psi\rangle = \frac{1}{\sqrt{2}} (|\uparrow\rangle - |\downarrow\rangle) \quad (41)$$

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

$$\mathbf{S}^2 |\psi\rangle = \frac{1}{\sqrt{2}} \left( \underbrace{\mathbf{S}^2 |\uparrow\rangle}_{=\frac{3}{4}\hbar^2} - \underbrace{\mathbf{S}^2 |\downarrow\rangle}_{=\frac{3}{4}\hbar^2} \right) = 0 \quad (42)$$