The Restricted Boltzmann Machine Applied to the Quantum Many-Body Problem

Erik Alexander Sandvik

June 7, 2021

Abstract

A variational Monte Carlo calculation is performed to obtain the ground state energy of a system of two interacting electrons in a harmonic oscillator potential, a system which was shown to be analytically solvable for a particular set of frequencies by Taut [1]. The trial function corresponds to a probability distribution of the visible layer in a restricted Boltzmann machine, which resembles the ground state of non-interacting harmonic oscillators. Calculations are performed for both the interacting and non-interacting case in two dimensions, for which the latter agrees with the analytical solution found in the literature. In the interacting case however, the calculated ground state energy deviates from the analytical solution by Taut by 8%, which is outside the estimated error bounds.

1 Introduction

In this project, I attempt to solve the quantum many-body problem using a variational Monte Carlo procedure. The system considered consists of two interacting electrons in a harmonic oscillator potential. The trial wavefunction is inspired by the probability distribution of the visible node configuration in a restricted Boltzmann machine - a type of neural network with a single hidden layer that is used to learn a probability distribution of the input parameters. Despite the title of this article, no actual restricted Boltzmann machine is implemented in this project. Instead, the trial function is chosen because it vaguely resembles the ground state wavefunction of the quantum harmonic oscillator.

The quantum system of two interacting electrons in a harmonic oscillator potential was shown to be analytically solvable by Taut for an infinite, but discretely spaced, set of harmonic oscillator frequencies [1]. In particular, in the two-dimensional case with a frequency of $\omega = 1$ a.u., the ground state energy was shown to be $E_0 = 3$ a.u., exactly one atomic unit above the ground state energy in the non-interacting case [2].

This project is a continuation of the article written by S. Schrader Et al. [3]. No new concepts (apart from the restricted Boltzmann machine) are provided here, but a brief recap is given in section 2.3. The system and the trial function is presented in section 2.1 and 2.2, with analytical expressions for the local energy, its gradient with respect to the variational parameters and the quantum force in section 5. Results for both the interacting and the non-interacting case are presented and discussed in section 3. Lastly, a conclusion is given and the article summarized in section 4.

2 Theory & Methods

2.1 The system

The Hamiltonian that describes a system of two interacting electrons in a harmonic oscillator potential is given by

$$H = -\frac{1}{2}\left(\nabla_1^2 + \nabla_2^2\right) + \frac{1}{2}\omega^2\left(\mathbf{r}_1^2 + \mathbf{r}_2^2\right) + \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \tag{2.1}$$

where atomic units ($\hbar = m_e = e = a_0 = 1$) will be used throughout this article. In the non-interacting case, the ground state energy is given by

$$E_0 = \frac{1}{2}\omega n d, \tag{2.2}$$

where n is the number of particles and d is the number of dimensions. The single-particle ground state wavefunction is given by [4]

$$\psi_0(\mathbf{r}) \propto e^{-\omega \mathbf{r}^2/2}. \tag{2.3}$$

2.2 The trial wavefunction

A restricted Boltzmann machine consists of an input layer $\mathbf{x}$ of M nodes and a hidden layer $\mathbf{h}$ of N nodes. The probability distribution for a configuration $(\mathbf{x},\ \mathbf{h})$ is defined as [5]

$$P(\mathbf{x},\ \mathbf{h}) = \frac{1}{Z}e^{-E(\mathbf{x},\ \mathbf{h})} \tag{2.4}$$

where Z^{-1} is a normalization constant and $E(\mathbf{x},\ \mathbf{h})$ is an "energy function", not to be confused with the energy of the quantum system. The energy function of choice corresponds to the "Gaussian-Binary" restricted Boltzmann machine,

$$E(\mathbf{x},\ \mathbf{h}) = \sum_i^M \frac{(x_i - a_i)^2}{2\sigma_i^2} - \sum_j^N b_j h_j - \sum_i^N \sum_j^N \frac{x_i W_{ij} h_j}{\sigma_i^2}, \tag{2.5}$$

where $\mathbf{a} \in \mathbb{R}^M$ and $\mathbf{b} \in \mathbb{R}^N$ are the so-called visible and hidden biases and $W \in \mathbb{R}^{M\times N}$ is a weight matrix. While the input vector is $\mathbf{x} \in \mathbb{R}^M$, the nodes of the hidden layer can only take binary values $h_j \in \{0, 1\}$. The input vector will correspond to the configuration of the system $(\mathbf{r}_1, \mathbf{r}_2)$, thus $M = nd$. The number of hidden nodes N is on the other hand a choice we make. The σ_i^2's are non-variational parameters and since every coordinate is on equal footing in the Hamiltonian I will set $\sigma_i^2 = \sigma^2\ \forall\ i$.

The marginal probability distribution of the configuration of the input layer is given by

$$P(\mathbf{x}) = \sum_{\{\mathbf{h}\}} P(\mathbf{x},\ \mathbf{h}) \tag{2.6}$$

where the sum is taken over all possible configurations of the hidden layer, i.e all binary vectors with N elements. This can in turn be written as

$$P(\mathbf{x}) \propto \exp\left\{-\sum_i^M \frac{(x_i - a_i)^2}{2\sigma^2}\right\} \sum_{\{\mathbf{h}\}} \exp\left\{\sum_j^N b_j h_j + \sum_i^N \sum_j^N \frac{x_i W_{ij} h_j}{\sigma^2}\right\} \tag{2.7}$$

$$= \exp\left\{-\sum_i^M \frac{(x_i - a_i)^2}{2\sigma^2}\right\} \sum_{\{\mathbf{h}\}} \prod_j^N \exp\left\{b_j h_j + \sum_i^N \frac{x_i W_{ij} h_j}{\sigma^2}\right\} \tag{2.8}$$

$$= \exp\left\{-\sum_i^M \frac{(x_i - a_i)^2}{2\sigma^2}\right\} \prod_j^N \sum_{h_j=0,1} \exp\left\{b_j h_j + \sum_i^N \frac{x_i W_{ij} h_j}{\sigma^2}\right\} \tag{2.9}$$

$$= \exp\left\{-\sum_i^M \frac{(x_i - a_i)^2}{2\sigma^2}\right\} \prod_j^N \left[1 + \exp\left\{b_j + \sum_i^N \frac{x_i W_{ij}}{\sigma^2}\right\}\right] \tag{2.10}$$

which is chosen as the trial function, $\Psi_T = P(\mathbf{x})$, with $\mathbf{x}$ being the configuration of the particles. Although $|\Psi_T|^2$ is the probability distribution of the particle configurations, and not Ψ_T, this is an apples and oranges sort of situation because whatever interpretation $P(\mathbf{x})$ has in a different context is not of interest here. As mentioned, $P(\mathbf{x})$ is chosen as the trial wavefunction just because it vaguely resembles the ground state of the quantum harmonic oscillator. Setting the weights and the visible biases to zero leads to

$$P(\mathbf{x}) \propto \prod_i^M \exp\left\{-\frac{x_i^2}{2\sigma^2}\right\}, \tag{2.11}$$

which shows that $P(\mathbf{x})$ becomes a product of single-particle ground states if we set $\sigma^2 = 1/\omega$.

2.3 Variational Monte Carlo

In this section I will give a brief repetition of the variational Monte Carlo procedure. For a more rigorous explanation, see [3].

The Variational Principle states that for any trial function Ψ_T, the expectation value of the Hamiltonian is bounded from below by the ground state energy E_0,

$$E_0 \leq \frac{\int d\mathbf{x} \Psi_T^*(\mathbf{x}; \boldsymbol{\alpha}) H \Psi_T(\mathbf{x}; \boldsymbol{\alpha})}{\int d\mathbf{x} |\Psi_T(\mathbf{x}; \boldsymbol{\alpha})|^2}, \tag{2.12}$$

with equality if and only if Ψ_T is the exact ground state. By tacking on some variational parameters $\boldsymbol{\alpha}$ onto the trial function, we can minimize the right hand side of Eq. (2.12) with respect to $\boldsymbol{\alpha}$ in order to approximate both the ground state and the ground state energy. The right hand side can also be written as

$$\langle E_L \rangle = \frac{\int d\mathbf{x} |\Psi_T(\mathbf{x}; \boldsymbol{\alpha})|^2 E_L(\mathbf{x}; \boldsymbol{\alpha})}{\int d\mathbf{x} |\Psi_T(\mathbf{x}; \boldsymbol{\alpha})|^2} \tag{2.13}$$

where

$$E_L(\mathbf{x}; \boldsymbol{\alpha}) \equiv \frac{1}{\Psi_T(\mathbf{x}; \boldsymbol{\alpha})} H \Psi_T(\mathbf{x}; \boldsymbol{\alpha}) \tag{2.14}$$

is the local energy. We can approximate the expectation value of the local energy with a sample mean

$$\langle E_L \rangle \approx \overline{E}_L = \frac{1}{m} \sum_i^m E_L(\mathbf{x}_i; \boldsymbol{\alpha}) \tag{2.15}$$

by drawing configurations $\mathbf{x}_i$ at random from the probability distribution $|\Psi_T(\mathbf{x}; \boldsymbol{\alpha})|^2$. This can be done using the Metropolis algorithm where, given a current configuration $\mathbf{x}_j$, a new configuration $\mathbf{x}_i$ is proposed and accepted with a probability

$$A_{i \leftarrow j} = \min \left(1, \frac{|\Psi_T(\mathbf{x}_i)|^2}{|\Psi_T(\mathbf{x}_j)|^2} \right). \tag{2.16}$$

If the new configuration is not accepted, the current configuration is kept. Specifically, the new configuration proposed by selecting a single particle in the system at random and moving it an amount $\Delta x \cdot \eta$ in all dimensions, where η is a uniformly distributed random variable in $[-1, 1]$. Since $\psi_0 \propto \exp\{-\omega \mathbf{r}^2/2\}$, the characteristic length of the system is $l = \sqrt{2/\omega}$, and so the steplength of choice is $\Delta x = l/2$.

We can also use importance sampling which involves non-symmetric transition probabilities $T_{i \leftarrow j}$ given by

$$T_{i \leftarrow j} \propto \exp \left\{ -\frac{(\mathbf{x}_i - \mathbf{x}_j - D\Delta t \mathbf{F}(\mathbf{x}_j))^2}{4D\Delta t} \right\}. \tag{2.17}$$

These guys emerge from solving the Fokker-Planck equation which describes the time evolution of a diffusion process. For wavefunctions, $\mathbf{F}(\mathbf{x})$ is the so-called "quantum force", which is given by

$$\mathbf{F}(\mathbf{x}) = \frac{2}{\Psi(\mathbf{x})} \nabla \Psi(\mathbf{x}). \tag{2.18}$$

With importance sampling, the acceptance probability is given by

$$A_{i \leftarrow j} = \min \left(1, \frac{T_{j \leftarrow i} |\Psi_T(\mathbf{x}_i)|^2}{T_{i \leftarrow j} |\Psi_T(\mathbf{x}_j)|^2} \right), \tag{2.19}$$

where the proposed configuration $\mathbf{x}_i$ is given by a discretized solution of the Langevin equation.

$$\mathbf{x}_i = \mathbf{x}_j + D\mathbf{F}(\mathbf{x}_j)\Delta t + \xi \sqrt{\Delta t} \tag{2.20}$$

where ξ is drawn from the standard normal distribution, and D is set to 0.5. Δt will be set to the same as Δx in the "brute force" Metropolis algorithm.

Then, the gradient of the expectation value of the local energy with respect to the variational parameters,

$$\nabla_{\boldsymbol{\alpha}} \langle E_L \rangle = 2(\langle E_L \nabla_{\boldsymbol{\alpha}} \ln \Psi_T \rangle - \langle E_L \rangle \langle \nabla_{\boldsymbol{\alpha}} \ln \Psi_T \rangle) \tag{2.21}$$

is calculated. This is used to update the variational parameters using a plain gradient descent iterative procedure,

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k - \eta \nabla_{\boldsymbol{\alpha}} \langle E_L \rangle, \tag{2.22}$$

where $0 < \eta \ll 1$ is the learning rate. The updating stops when $|\boldsymbol{\alpha}_{k+1} - \boldsymbol{\alpha}_k|$ is less than some small positive tolerance δ.

Finally, an estimate of the statistical error $\sqrt{\mathrm{Var}(\overline{E}_L)}$ is needed. Due to the configurations $\mathbf{x}_i$ being highly correlated, the standard way to approximate the error

$$\frac{\sigma}{\sqrt{m}}, \tag{2.23}$$

where

$$\sigma^2 = \frac{1}{m}\sum_i^m (E_L(\mathbf{x}_i) - \overline{E}_L)^2, \tag{2.24}$$

gives a too optimistic estimate. To get a proper estimate of the error, so-called blocking transformations are performed on the data set $\{E_L(\mathbf{x}_i)\}$ by replacing subsequent data points in the set by their average,

$$(E_L(\mathbf{x}_1),\ E_L(\mathbf{x}_2),\ \ldots,\ E_L(\mathbf{x}_{m-1}),\ E_L(\mathbf{x}_m)) \rightarrow \left(\frac{E_L(\mathbf{x}_1) + E_L(\mathbf{x}_2)}{2},\ \ldots,\ \frac{E_L(\mathbf{x}_{m-1}) + E_L(\mathbf{x}_m)}{2}\right), \tag{2.25}$$

where m is an integer power of 2. It can be shown that the total variance of $\overline{E}_L$, which both the sample variance σ^2/m and correlations between the data points contribute to, is conserved under blocking transformations, but the correlation contribution is reduced [6]. Therefore the sample variance of the transformed data sets will approach the total variance. Blocking transformation are performed until subsequent data points in the transformed data set are uncorrelated. Again, see [3] for more details.

3 Results & Discussion

3.1 Non-interacting electrons in two dimensions

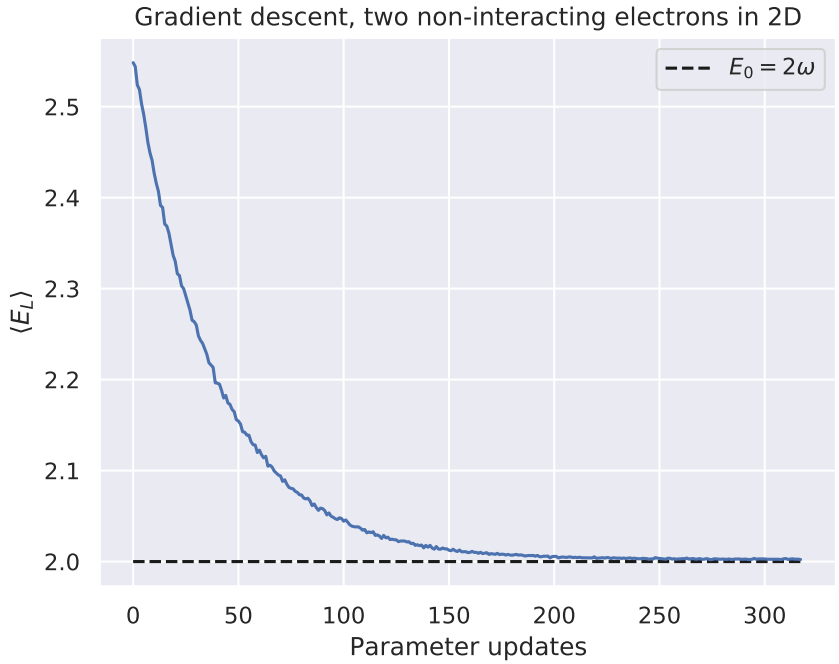


Figure 1: *The expectation value of the local energy as a function of the number of parameter updates in gradient descent. The updates were made using a learning rate of* $\eta = 0.01$. 10^6 *Monte Carlo steps were used at each update. The parameters were initialized from a Gaussian distribution with* $\sigma = 0.5$.

Figure 1 shows the expectation value of the local energy at each iteration in the gradient descent procedure. Only a single layer was used and the frequency ω was explicitly set to 1. The brute-force Metropolis method was used to compute the expectation values. The ℓ_1 norms of the vectors containing the variational parameters are shown in Figure 2. Although W is a matrix, the "ℓ_1 norm" is the sum of the magnitudes of the elements. The visible biases and the weight matrix elements were expected to approach zero with each parameter update, and so the variational parameters were initialized from a Gaussian distribution with a purposefully "large" standard deviation of $\sigma = 0.5$. This was to get an initial energy away from the analytical solution and to show that it converges towards $E_0 = 2\omega$ at each iteration. Figure 1 shows that this is indeed what happens, albeit with some small visible fluctuations. This may be due to the choice of steplength Δx, where the displacement $\Delta x \cdot \eta$ may be too large (or too small) depending on the random variable η.

Although magnitudes of the visible biases get smaller with each update, the magnitudes of the weight matrix elements do not, contrary to what I expected. This is strange first of all because while the expectation value of the local energy converges to $E_0 = 2\omega$ as expected, the trial wavefunction does not converge to a product of single-particle harmonic oscillator ground states $\psi_0(\mathbf{r}) \propto e^{-\omega \mathbf{r}^2/2}$. And secondly because setting σ to a really small number in the initialization of the parameters makes the expectation value of the local energy start very close to the analytical solution right off the bat. For small σ, the gradient descent process will stop almost instantly or not at all (at least not until after many parameter updates has passed), depending on whether the tolerance δ is "too large" or "too small" (in both cases $\delta < 10^{-4}$). In the latter case, the visible biases were observed to tend towards zero, while the weight matrix elements increased in magnitude ever so slightly, which is consistent with the behaviour shown in Figure 2. The hidden biases on the other hand stayed approximately constant.

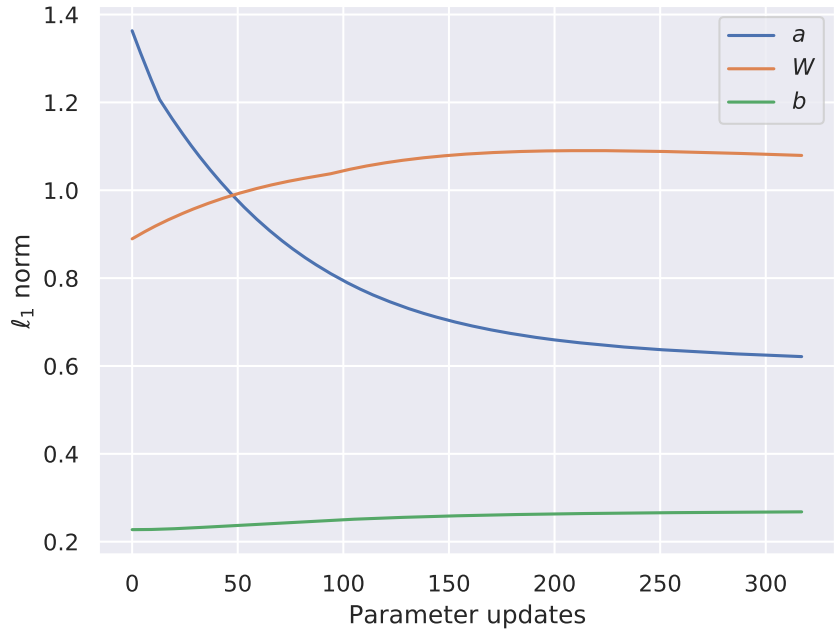


Figure 2: *The ℓ_1 norm of the vectors containing the variational parameters after each parameter update in gradient descent.*

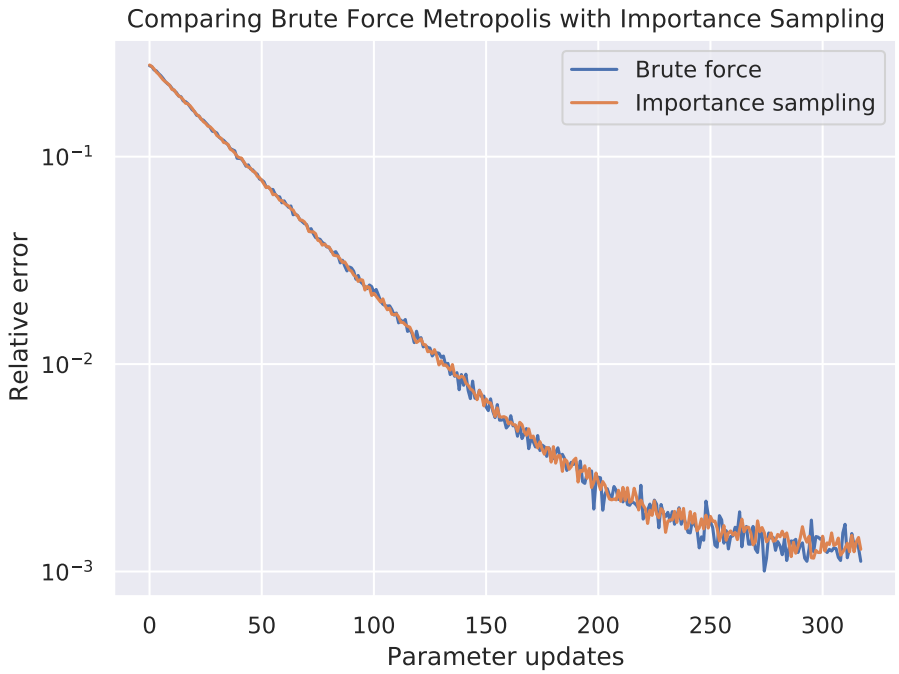
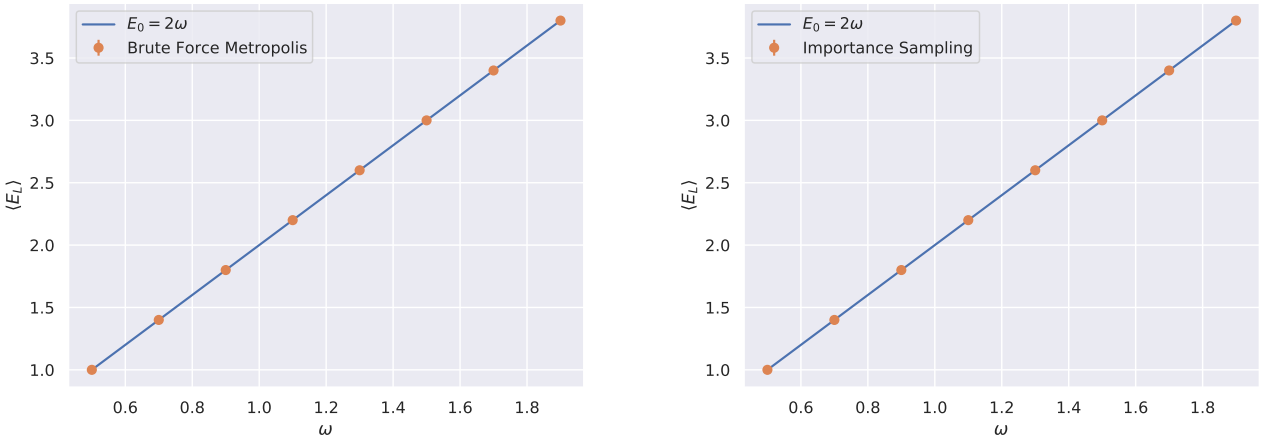


Figure 3: *The relative error $|\langle E_L\rangle - E_0|/E_0$ after each parameter update in gradient descent for both brute force Metropolis and importance sampling.*

If I were to show figures corresponding to Figures 1 and 2 for importance sampling they would look utterly identical. Figure 3 shows the relative error after each parameter update for both brute force Metropolis and importance sampling. We see that both sampling methods perform at about the same level in terms of accuracy and in terms of how quickly the expectation value of the local energy converges to the analytical solution.



(a) **(b)**

Figure 4: *Expectation value of the local energy as function of the frequency for brute force Metropolis (left) and importance sampling (right). The errors are the standard deviations of the transformed data sets using the blocking method, multiplied by* 3.29053 *which corresponds to a* 99.9% *confidence interval.*

Figure 4 shows the expectation value of the local energy as function of the harmonic oscillator frequency with error bars (believe it or not) for both brute force Metropolis and importance sampling. The errors correspond to a 99.9% confidence interval. A tolerance of $\delta = 10^{-4}$ was used to determine when to stop updating the parameters and a standard deviation of $\sigma = 10^{-3}$ was used in the parameter initialization. The orders of magnitude of the errors are $10^{-5} - 10^{-6}$ and they are thus not visible in Figure 4. That's why they are listed in Table 1. We see that brute force Metropolis and importance sampling perform about equally, but that brute force Metropolis performs slightly worse all across the board. The errors are also increasing with the frequency.

Table 1: *The errors corresponding to a* 99.9% *confidence interval for both brute force Metropolis (BF) and importance sampling (IM). All errors are in units of* 10^{-5}.

ω	0.5	0.7	0.9	1.1	1.3	1.5	1.7	1.9
BF	0.98	1.63	2.32	3.32	4.14	5.07	6.04	7.04
IM	0.79	1.28	1.83	2.40	3.10	3.68	4.59	5.27

3.2 Interacting electrons in two dimensions

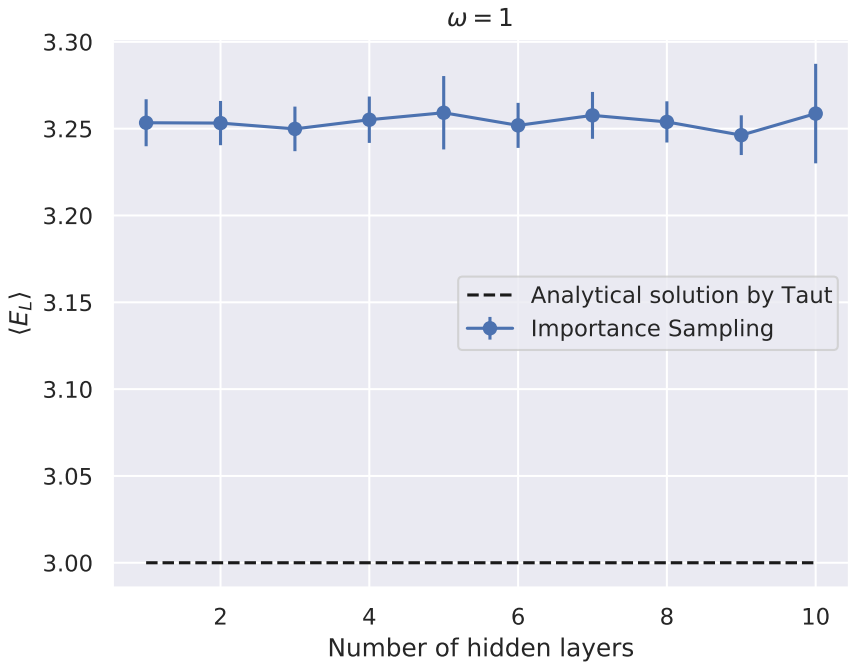


Figure 5: *The expectation value of the local energy as a function of the number of hidden layers for the frequency $\omega = 1$.*

Figure 5 shows the expectation value of the local energy as a function of the number of hidden layers for the frequency $\omega = 1$ with Taut's analytical solution for comparison. The errors have increased by 3 orders of magnitude, and it's difficult to say how many hidden layers are optimal. All errors are 0.01 except for 5 and 10 hidden layers which have an error of 0.02 and 0.03 respectively. Clearly, the analytical solution by Taut is nowhere near the calculated energy, despite the fact that the errors correspond to a 99.9% confidence interval. Therefore, there must be some major systematic errors in the calculations, one of which is easy to point out: there is no incentive for the trial function to go to zero as the electrons get closer because of the form it has. There are no variational parameters coupled to the relative distance between the electrons, and no measures taken to make sure that it goes to zero. Meanwhile, the true wavefunction has to go to zero because the Coulomb potential blows up. The trial function allows the electrons to get close more often than they have any right to, which causes it to overestimate the energy of the system by about 8%. An easy "fix" would be to not allow the electrons to get closer than some chosen amount, and then adjust it so that the energy get closer to Taut's analytical solution. However, constructing a model in order to get the result you want as opposed to making a model based on the properties of the system is not particularly satisfying. And this "fix" effectively makes the wavefunction discontinuous where it shouldn't be, and it is thus unphysical. Something more satisfying would be to include some factor that goes to zero as the electrons get close like, say, $\exp\{-\beta|\mathbf{r}_1 - \mathbf{r}_2|^{-\gamma}\}$ where β and γ are positive.

4 Conclusion

A restricted Boltzmann machine inspired trial function was applied to a variational Monte Carlo calculation of the quantum many-body problem of two electrons in a harmonic oscillator potential. In the non-interacting case, the calculation reproduced the ground state energy of non-interacting harmonic oscillators as a function of the frequency, with respectably low error estimates of about 10^{-5} for both brute force Metropolis and importance sampling, with a slightly smaller error for the latter calculations. In the interacting case however, the calculation failed to reproduce the analytical

solution provided by Taut, overestimating the ground state energy by about 8%. With no clear pattern determining how many hidden layers are optimal, and the analytical solution far away from the error bounds provided by the blocking method, it is clear that the trial function cannot possibly represent the true ground state of the system of interacting electrons, in particular because the true wavefunction must go to zero as the electrons get closer.

5 Appendix

5.1 The Local Energy

The local energy is given by

$$E_L = \frac{1}{\Psi} H\Psi \tag{5.1}$$

$$= \frac{1}{\Psi}\left(-\frac{1}{2}\nabla^2\Psi + \frac{1}{2}\omega^2\mathbf{R}^2\Psi + \sum_{i<j}\frac{1}{r_{ij}}\Psi\right) \tag{5.2}$$

$$= -\frac{1}{2\Psi}\nabla^2\Psi + \frac{1}{2}\omega^2\mathbf{R}^2 + \sum_{i<j}\frac{1}{r_{ij}}. \tag{5.3}$$

The tricky part is the Laplacian term, but we can use a trick to simplify the calculations. Since

$$\frac{\partial \ln\Psi}{\partial x} = \frac{1}{\Psi}\frac{\partial\Psi}{\partial x} \rightarrow \frac{\partial\Psi}{\partial x} = \Psi\frac{\partial \ln\Psi}{\partial x}, \tag{5.4}$$

we have that

$$\frac{\partial^2\Psi}{\partial x^2} = \frac{\partial\Psi}{\partial x}\frac{\partial \ln\Psi}{\partial x} + \Psi\frac{\partial^2 \ln\Psi}{\partial x^2} \tag{5.5}$$

$$\frac{1}{\Psi}\frac{\partial^2\Psi}{\partial x^2} = \frac{1}{\Psi}\frac{\partial\Psi}{\partial x}\frac{\partial \ln\Psi}{\partial x} + \frac{\partial^2 \ln\Psi}{\partial x^2} = \left(\frac{\partial \ln\Psi}{\partial x}\right)^2 + \frac{\partial^2 \ln\Psi}{\partial x^2}. \tag{5.6}$$

Thus we can write the local energy as

$$E_L = -\frac{1}{2}\left[(\nabla \ln\Psi)^2 + \nabla^2 \ln\Psi\right] + \frac{1}{2}\omega^2\mathbf{R}^2 + \sum_{i<j}\frac{1}{r_{ij}}. \tag{5.7}$$

The natural logarithm of the wavefunction is

$$\ln\Psi = \ln\left(\exp\left\{-\frac{1}{2\sigma^2}\sum_i^M (x_i - a_i)^2\right\}\prod_j^N\left[1 + \exp\left\{b_j + \frac{1}{\sigma^2}\sum_i^M x_i W_{ij}\right\}\right]\right) \tag{5.8}$$

$$= \ln\left(\exp\left\{-\frac{1}{2\sigma^2}\sum_i^M (x_i - a_i)^2\right\}\right) + \ln\left(\prod_j^N\left[1 + \exp\left\{b_j + \frac{1}{\sigma^2}\sum_i^M x_i W_{ij}\right\}\right]\right)) \tag{5.9}$$

$$= -\frac{1}{2\sigma^2}\sum_i^M (x_i - a_i)^2 + \sum_j^N \ln\left[1 + \exp\left\{b_j + \frac{1}{\sigma^2}\sum_i^M x_i W_{ij}\right\}\right] \tag{5.10}$$

The elements of the gradient are

$$\frac{\partial}{\partial x_k}\ln\Psi = -\frac{x_k - a_k}{\sigma^2} + \sum_j^N \frac{\frac{1}{\sigma^2}W_{kj}\exp\left\{b_j + \frac{1}{\sigma^2}\sum_i^M x_i W_{ij}\right\}}{1 + \exp\left\{b_j + \frac{1}{\sigma^2}\sum_i^M x_i W_{ij}\right\}} \tag{5.11}$$

$$= -\frac{x_k - a_k}{\sigma^2} + \frac{1}{\sigma^2}\sum_j^N \frac{W_{kj}}{1 + \exp\left\{-\left(b_j + \frac{1}{\sigma^2}\sum_i^M x_i W_{ij}\right)\right\}}. \tag{5.12}$$

If we define a vector $\mathbf{e}$ with elements

$$e_j = \left[1 + \exp\left\{-\left(b_j + \frac{1}{\sigma^2}\sum_i^M x_i W_{ij}\right)\right\}\right]^{-1} \tag{5.13}$$

Then we can write the gradient as

$$\nabla \ln\Psi = -\frac{1}{\sigma^2}(\mathbf{x} - \mathbf{a}) + \frac{1}{\sigma^2}W\mathbf{e}. \tag{5.14}$$

which also happens to be the quantum force up to a factor of 2. Meanwhile, each term in the Laplacian is given by

$$\frac{\partial^2}{\partial x_k^2}\ln\Psi = -\frac{1}{\sigma^2} + \frac{1}{\sigma^2}\sum_j^N W_{kj}\frac{\partial}{\partial x_k}\left[1 + \exp\left\{-\left(b_j + \frac{1}{\sigma^2}\sum_i^M x_i W_{ij}\right)\right\}\right]^{-1} \tag{5.15}$$

$$= -\frac{1}{\sigma^2} - \frac{1}{\sigma^2}\sum_j^N W_{kj}\left[1 + \exp\left\{-\left(b_j + \frac{1}{\sigma^2}\sum_i^M x_i W_{ij}\right)\right\}\right]^{-2}\exp\left\{-\left(b_j + \frac{1}{\sigma^2}\sum_i^M x_i W_{ij}\right)\right\}\left(-\frac{W_{kj}}{\sigma^2}\right) \tag{5.16}$$

$$= -\frac{1}{\sigma^2} + \frac{1}{\sigma^4}\sum_j^N W_{kj}^2 \frac{\exp\left\{-\left(b_j + \frac{1}{\sigma^2}\sum_i^M x_i W_{ij}\right)\right\}}{\left[1 + \exp\left\{-\left(b_j + \frac{1}{\sigma^2}\sum_i^M x_i W_{ij}\right)\right\}\right]^2}. \tag{5.17}$$

5.2 The Local Energy Gradient

For a general Hamiltonian and trial function, the gradient of the local energy can be written as

$$\nabla_{\boldsymbol{\alpha}}\langle E_L\rangle = \nabla_{\boldsymbol{\alpha}}\langle H\rangle = \nabla_{\boldsymbol{\alpha}}\left(\frac{\int d\mathbf{R}\Psi^*\mathcal{H}\Psi}{\int d\mathbf{R}|\Psi|^2}\right). \tag{5.18}$$

Using the product and quotient rules we can write this as

$$\nabla_{\boldsymbol{\alpha}}\langle E_L\rangle = \frac{\left(\int d\mathbf{R}|\Psi|^2\right)\nabla_{\boldsymbol{\alpha}}\left(\int d\mathbf{R}\Psi^*\mathcal{H}\Psi\right) - \left(\int d\mathbf{R}\Psi^*\mathcal{H}\Psi\right)\nabla_{\boldsymbol{\alpha}}\left(\int d\mathbf{R}|\Psi|^2\right)}{\left(\int d\mathbf{R}|\Psi|^2\right)^2} \tag{5.19}$$

$$= \frac{\int d\mathbf{R}\left[\nabla_{\boldsymbol{\alpha}}\left(\Psi^*\right)\mathcal{H}\Psi + \Psi^*\nabla_{\boldsymbol{\alpha}}(\mathcal{H}\Psi)\right]}{\int d\mathbf{R}|\Psi|^2} - \frac{\left(\int d\mathbf{R}\Psi^*\mathcal{H}\Psi\right)\int d\mathbf{R}\nabla_{\boldsymbol{\alpha}}|\Psi|^2}{\left(\int d\mathbf{R}|\Psi|^2\right)^2} \tag{5.20}$$

$$= \frac{\int d\mathbf{R}|\Psi|^2\frac{\nabla_{\boldsymbol{\alpha}}(\Psi^*)\mathcal{H}\Psi+\Psi^*\nabla_{\boldsymbol{\alpha}}(\mathcal{H}\Psi)}{|\Psi|^2}}{\int d\mathbf{R}|\Psi|^2} - \frac{\int d\mathbf{R}\Psi^*\mathcal{H}\Psi}{\int d\mathbf{R}|\Psi|^2}\frac{\int d\mathbf{R}|\Psi|^2\frac{\nabla_{\boldsymbol{\alpha}}|\Psi|^2}{|\Psi|^2}}{\int d\mathbf{R}|\Psi|^2} \tag{5.21}$$

$$= \left\langle \frac{\nabla_{\boldsymbol{\alpha}}\left(\Psi^*\right)\mathcal{H}\Psi + \Psi^*\nabla_{\boldsymbol{\alpha}}(\mathcal{H}\Psi)}{|\Psi|^2}\right\rangle - \langle H\rangle\left\langle\frac{\nabla_{\boldsymbol{\alpha}}|\Psi|^2}{|\Psi|^2}\right\rangle \tag{5.22}$$

$$= \left\langle \frac{\nabla_{\boldsymbol{\alpha}}\left(\Psi^*\right)\mathcal{H}\Psi + \nabla_{\boldsymbol{\alpha}}\left(\Psi^*\right)\mathcal{H}\Psi}{|\Psi|^2}\right\rangle - \langle H\rangle\left\langle\frac{2|\Psi|\nabla_{\boldsymbol{\alpha}}|\Psi|}{|\Psi|^2}\right\rangle \tag{5.23}$$

$$= \left\langle \frac{2\nabla_{\boldsymbol{\alpha}}\Psi^*}{\Psi^*}E_L\right\rangle - \langle E_L\rangle\left\langle\frac{2\nabla_{\boldsymbol{\alpha}}|\Psi|}{|\Psi|}\right\rangle = 2\left[\left\langle \frac{\nabla_{\boldsymbol{\alpha}}\Psi^*}{\Psi^*}E_L\right\rangle - \langle E_L\rangle\left\langle\frac{\nabla_{\boldsymbol{\alpha}}|\Psi|}{|\Psi|}\right\rangle\right] \tag{5.24}$$

Depending on the trial function, in particular if the trial function is an exponential or a product of exponentials, the following equivalent expression might be more practical.

$$\nabla_{\boldsymbol{\alpha}}\langle E_L\rangle = 2\left[\langle E_L\nabla_{\boldsymbol{\alpha}}\ln\Psi^*\rangle - \langle E_L\rangle\langle\nabla_{\boldsymbol{\alpha}}\ln|\Psi|\rangle\right] \tag{5.25}$$

In this case we have

$$\frac{\partial\Psi}{\partial a_k} = \frac{1}{Z}\prod_j^N\left[1+\exp\left\{b_j+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}\right]\prod_{i\neq k}^M\exp\left\{-\frac{(x_i-a_i)^2}{2\sigma^2}\right\}\frac{\partial}{\partial a_k}\exp\left\{-\frac{(x_k-a_k)^2}{2\sigma^2}\right\} \tag{5.26}$$

$$\frac{1}{\Psi}\frac{\partial\Psi}{\partial a_k} = \frac{\frac{1}{Z}\prod_j^N\left[1+\exp\left\{b_j+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}\right]\prod_{i\neq k}^M\exp\left\{-\frac{(x_i-a_i)^2}{2\sigma^2}\right\}\frac{\partial}{\partial a_k}\exp\left\{-\frac{(x_k-a_k)^2}{2\sigma^2}\right\}}{\frac{1}{Z}\prod_j^N\left[1+\exp\left\{b_j+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}\right]\prod_i^M\exp\left\{-\frac{(x_i-a_i)^2}{2\sigma^2}\right\}} \tag{5.27}$$

$$= \frac{\frac{\partial}{\partial a_k}\exp\left\{-\frac{(x_k-a_k)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(x_k-a_k)^2}{2\sigma^2}\right\}} = \frac{\exp\left\{-\frac{(x_k-a_k)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(x_k-a_k)^2}{2\sigma^2}\right\}}\cdot\left[-\frac{2(x_k-a_k)}{2\sigma^2}\cdot(-1)\right] \tag{5.28}$$

$$= \frac{x_k-a_k}{\sigma^2} \tag{5.29}$$

$$\frac{\partial\Psi}{\partial b_k} = \frac{1}{Z}\prod_i^M\exp\left\{-\frac{(x_i-a_i)^2}{2\sigma^2}\right\}\prod_{j\neq k}^N\left[1+\exp\left\{b_j+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}\right]\frac{\partial}{\partial b_k}\left[1+\exp\left\{b_k+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}\right] \tag{5.30}$$

$$\frac{1}{\Psi}\frac{\partial\Psi}{\partial b_k} = \frac{\frac{1}{Z}\prod_i^M \exp\left\{-\frac{(x_i-a_i)^2}{2\sigma^2}\right\}\prod_{j\neq k}^N\left[1+\exp\left\{b_j+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}\right]\frac{\partial}{\partial b_k}\left[1+\exp\left\{b_k+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}\right]}{\frac{1}{Z}\prod_j^N\left[1+\exp\left\{b_j+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}\right]\prod_i^M\exp\left\{-\frac{(x_i-a_i)^2}{2\sigma^2}\right\}} \tag{5.31}$$

$$= \frac{\frac{\partial}{\partial b_k}\left[1+\exp\left\{b_k+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}\right]}{1+\exp\left\{b_k+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}} = \frac{\frac{\partial}{\partial b_k}\exp\left\{b_k+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}}{1+\exp\left\{b_k+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}} = \frac{\exp\left\{b_k+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}}{1+\exp\left\{b_k+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}} \tag{5.32}$$

$$= \frac{1}{1+\exp\left\{-\left(b_k+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right)\right\}} \tag{5.33}$$

$$\frac{\partial\Psi}{\partial W_{kl}} = \frac{1}{Z}\prod_i^M\exp\left\{-\frac{(x_i-a_i)^2}{2\sigma^2}\right\}\prod_{j\neq l}^N\left[1+\exp\left\{b_j+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}\right]\frac{\partial}{\partial W_{kl}}\left[1+\exp\left\{b_l+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}\right] \tag{5.34}$$

$$\frac{1}{\Psi}\frac{\partial\Psi}{\partial W_{kl}} = \frac{\frac{1}{Z}\prod_i^M \exp\left\{-\frac{(x_i-a_i)^2}{2\sigma^2}\right\}\prod_{j\neq l}^N\left[1+\exp\left\{b_j+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}\right]\frac{\partial}{\partial W_{kl}}\left[1+\exp\left\{b_l+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}\right]}{\frac{1}{Z}\prod_j^N\left[1+\exp\left\{b_j+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}\right]\prod_i^M\exp\left\{-\frac{(x_i-a_i)^2}{2\sigma^2}\right\}} \tag{5.35}$$

$$= \frac{\frac{\partial}{\partial W_{kl}}\left[1+\exp\left\{b_l+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}\right]}{\left[1+\exp\left\{b_l+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}\right]} = \frac{\frac{\partial}{\partial W_{kl}}\exp\left\{b_l+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}}{1+\exp\left\{b_l+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}} = \frac{\exp\left\{b_l+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}}{1+\exp\left\{b_l+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right\}}\cdot\frac{x_k}{\sigma^2} \tag{5.36}$$

$$= \frac{x_k/\sigma^2}{1+\exp\left\{-\left(b_l+\sum_i^M\frac{x_iW_{ij}}{\sigma^2}\right)\right\}} \tag{5.37}$$

References

[1] M. Taut. "Two electrons in an external oscillator potential: Particular analytic solutions of a Coulomb correlation problem". In: *Phys. Rev. A* 48 (5 Nov. 1993), pp. 3561–3566. DOI: 10.1103/PhysRevA.48.3561. URL: https://link.aps.org/doi/10.1103/PhysRevA.48.3561 (cit. on p. 1).

[2] M Taut. "Two electrons in a homogeneous magnetic field: particular analytical solutions". In: *Journal of Physics A: Mathematical and General* 27.3 (Feb. 1994), pp. 1045–1055. DOI: 10.1088/0305-4470/27/3/040. URL: https://doi.org/10.1088/0305-4470/27/3/040 (cit. on p. 1).

[3] E. A. Sandvik S. Schrader. "A VMC calculation of an interacting boson gas in a harmonic oscillator trap". In: (Mar. 2021). URL: https://github.com/erikasan/comphys-2-project1/blob/main/report/report.pdf (cit. on pp. 1, 3, 5).

[4] N. Zettili. *Quantum Mechanics*. 2nd ed. Wiley, 2009 (cit. on p. 2).

[5] A Courville I. Goodfellow Y. Bengio. *Deep Learning*. 1st ed. MIT Press, 2020 (cit. on p. 2).

[6] H. Flyvbjerg and H. G. Petersen. "Error estimates on averages of correlated data". In: *The Journal of Chemical Physics* 91.1 (1989), pp. 461–466. DOI: 10.1063/1.457480. eprint: https://doi.org/10.1063/1.457480. URL: https://doi.org/10.1063/1.457480 (cit. on p. 5).