Variational Monte Carlo on Bosonic Systems

Winther-Larsen, Sebastian Gregorius[1,*] and Schøyen, Øyvind Sigmundson[1,*]

[1] *University of Oslo*

(Dated: April 6, 2018)

In this project we use Quantum Variational Monte Carlo on a toy model of a bosonic gas in an elliptic trap. The framework we have built is first tested against analytic solutions of simple non-interacting systems. From there the system is extended to include perturbed harmonic oscillator traps and particle interaction. We compute one-body densities for the interacting case and discuss why this tool is important and how our toy model has similar attributes to real-world systems.

CONTENTS

I. Introduction 1

II. Variational Monte Carlo 2
- A. Monte Carlo integration 2
- B. Local energy 3
- C. The drift force 3
- D. One-Body Densities 3

III. Systems 3
- A. Non-interacting harmonic oscillators 4
  - 1. Exact variational energy 5
- B. Interacting hard sphere bosons 5
  - 1. Scaling the system 5

IV. Algorithms 6
- A. The Metropolis-Hastings Algorithm 6
  - 1. Initial distribution of particles 6
  - 2. Importance Sampling 7
- B. Statistical Analysis 7
  - 1. Blocking 7
- C. Gradient Descent 8
  - 1. Stability of the algorithm 9

V. Results 9
- A. Non-Interacting Systems 9
  - 1. Introducing Importance Sampling 10
- B. Interacting elliptical harmonic oscillator 11
  - 1. Finding the minimum 12
  - 2. One-Body Densities 12

VI. Discussion 12
- A. Validity of Results 12
- B. Energy per particle 13
- C. Low density system 13
- D. Stability of the program 14

VII. Summary Remarks 14

References 14

A. Computing the exact energy of the non-interacting system 15

B. Computing the drift force and local energy of the full system 15

C. Variational parameter gradient of the expectation energy 16

D. Brute Force Metropolis-Hastings 18

I. INTRODUCTION

We will in this project study the Variational Monte Carlo method, and use it to evaluate the ground state energy of a trapped, hard sphere Bose gas.

By building up an increasingly intricate system we eloquently illustrate the strength of the Variational Monte Carlo methods. Any Monte Carlo method draws on the fact that the computer will not complain when being tasked to perform a very similar task up to several million times in rapid succession. Monte Carlo methods is in essence a series of guesses where the current guess is only a minor alteration of the one prior to it. The trick lies in only accepting the subsequent guesses that improves upon our configuration up to a certain likelihood.

The method we have employed is called a *Variational* Monte Carlo method, because we make an educated guess at a trial wavefunction which we allow certain degrees of freedom called *variational parameters*. We shall see that the system converges to a proper [1] minimum for only a certain set of parameters.

First, we provide the theoretical foundation for the Variational Monte Carlo method. This includes explaining how Monte Carlo integration works, and introducing the important quantities *local energy*, *drift force* and *one-body densities*.

Second, we outline in more detailed algebra the nature of the systems we will be studying. For the non-interacting system we are able to derive an exact variational energy to which we can compare the results from our simulations. We furthermore give a description of an

* Project code: https://github.com/Schoyen/FYS4411

[1] Global, as far as we can tell.

interacting system implemented as if each particle is a hard sphere.

Third, we provide a thorough description of algorithms and statistical methods employed in this study. Here we introduce the Metropolis-Hastings algorithm. This algorithm can by a deft trick be expanded with the method of *importance sampling* yielding a more intelligent choice for the next step in the Monte Carlo cycling. The main point of note with regards to the statistical analysis is that the data generated by the Monte Carlo cycling is autocorrelated and computation of variance must be done with the *blocking method* for time series data. As an automated way of finding the optimal variational parameter, we sketch out the method of *gradient descent*.

Fourth, we put the machinery described in the previous sections to work. We test against the actual values in a non-interacting system before analyzing the more interesting perturbed-trap system with interacting particles.

Lastly, we highlight interesting aspects of this study in a discussion of the results before ending on some summary remarks.

II. VARIATIONAL MONTE CARLO

The variational principle states that for a given *trial wavefunction*, $|\Psi_T\rangle$, the expectation value of the Hamiltonian, H, will be an upper bound to the ground state energy, E_0 of the system.

$$E_0 \leq E = \frac{\langle\Psi_T| H |\Psi_T\rangle}{\langle\Psi_T|\Psi_T\rangle}, \tag{1}$$

The complicated part of the variational principle is finding the trial wavefunction that minimizes the energy. The strength in this method lies in introducing variational parameters in the trial wavefunction. Using these as degrees of freedom we can minimize the energy with respect to these in order to find an infimum to the ground state energy for the given trial wavefunction and Hamiltonian. For a select few systems we can calculate this analytically, but for the interesting systems we quickly hit an intractable wall as the systems become analytically unsolvable. To get around this we try to calculate the energy for a given Hamiltonian and a trial wavefunction for different variational parameters numerically.

To determine if we have located a minimum we can use the variance of the variational energy.

$$\sigma^2 = \frac{\langle\Psi_T| H^2 |\Psi_T\rangle}{\langle\Psi_T|\Psi_T\rangle} - \left(\frac{\langle\Psi_T| H |\Psi_T\rangle}{\langle\Psi_T|\Psi_T\rangle}\right)^2. \tag{2}$$

If we have found the "true" trial wavefunction $|\Psi_T\rangle$ with the correct variational parameters, i.e., the eigenfunction to the Hamiltonian we get $\sigma^2 = 0$. We are thus interested in trying to find the variational parameters in the trial wavefunction that minimizes the variance.

A. Monte Carlo integration

Unfortunately the evaluation of the expectation value of the energy for a given trial wavefunction involves a large integral.

$$E = \frac{\langle\Psi_T| H |\Psi_T\rangle}{\langle\Psi_T|\Psi_T\rangle} = \frac{\int \mathrm{d}\mathbf{r}\Psi_T^*(\mathbf{r})H\Psi_T(\mathbf{r})}{\int \mathrm{d}\mathbf{r}|\Psi_T|^2}, \tag{3}$$

where $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_N)$ are the positions[2] of the particles contained in the system.

To get around[3] the trouble of evaluating the multidimensional-integral we employ the method of Monte Carlo integration. This consists of approximating an integral of the expectation value of a certain quantitity by a sum.

$$\langle x\rangle = \int_{\mathbb{R}} xp(x)\mathrm{d}x \approx \frac{1}{N}\sum_{i=1}^{N} x_i p(x_i), \tag{4}$$

where x is a random variable with a probability density function $p(x)$ and $\{x_i\}$ for $i \in \{1, \ldots, N\}$ is the set of drawn random variables. Here N is the number of Monte Carlo samples chosen to approximate the integral. By the law of large numbers we expect that

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} x_i p(x_i) = \langle x\rangle\,. \tag{5}$$

For us to be able to take Monte Carlo integration to the quantum realm we have to construct a probability density function from the wavefunction. Luckily the wavefunction squared is just such a quantity. Given a normalized wavefunction $|\psi\rangle$ we can compute the expectation value of a quantum operator $\mathcal{O}$ using Monte Carlo integration.

$$\langle\mathcal{O}\rangle = \langle\psi|\,\mathcal{O}\,|\psi\rangle = \int \mathrm{d}\mathbf{r}\psi^*(\mathbf{r})\mathcal{O}\psi(\mathbf{r}) \tag{6}$$

$$= \int \mathrm{d}\mathbf{r}|\psi(\mathbf{r})|^2\frac{\mathcal{O}\psi(\mathbf{r})}{\psi(\mathbf{r})}, \tag{7}$$

where we have multiplied and divided by the wavefunction. We now define the probability density function, $p(\mathbf{r})$, as

$$p(\mathbf{r}) \equiv |\psi(\mathbf{r})|^2, \tag{8}$$

and the local quantum operator

$$\mathcal{O}_L(\mathbf{r}) \equiv \frac{\mathcal{O}\psi(\mathbf{r})}{\psi(\mathbf{r})}. \tag{9}$$

---

[2] This can include both spatial and spin parts, but we will in the remainder of this project ignore spin.

[3] Not exactly *get around* per se. We still have to evaluate the integral, but we make the problem solvable at all.

We see that if the chosen wavefunction is an eigenstate of the quantum operator, the local quantum operator becomes a constant. We can thus approximate the expecatation value of the quantum operator using Monte Carlo integration.

$$\langle O \rangle = \int \mathrm{d}\mathbf{r} p(\mathbf{r}) \mathcal{O}_L(\mathbf{r}) \approx \frac{1}{N} \sum_{i=1}^{N} p(\mathbf{r}_i) \mathcal{O}_L(\mathbf{r}_i). \tag{10}$$

In the case of a non-normalized wavefunction we need to divide by a normalization factor in the probability density function.

B. Local energy

As is mostly the case in physics, the quantity we are looking for is the energy. We therefore define the *local energy* by

$$E_L(\mathbf{r}) = \frac{H\Psi_T(\mathbf{r})}{\Psi_T(\mathbf{r})}. \tag{11}$$

The energy of the system is thus found by computing the expectation value of the local energy.

$$\langle E_L \rangle = \frac{\int \mathrm{d}\mathbf{r} |\Psi_T(\mathbf{r})|^2 E_L(\mathbf{r})}{\int \mathrm{d}\mathbf{r} |\Psi_T(\mathbf{r})|^2} \approx \frac{1}{N} \sum_{i=1}^{N} p(\mathbf{r}_i) E_L(\mathbf{r}_i). \tag{12}$$

To determine if we have found a minimum we compute the variance of the expectation value of the local energy.

$$\sigma^2 = \langle E_L^2 \rangle - \langle E_L \rangle^2. \tag{13}$$

One of the most computationally intensive parts of the Variational Monte Carlo algorithm will be to compute E_L. We will therefore find an analytical expression for E_L in terms of the trial wavefunctions and the system we are exploring.

C. The drift force

A disadvantage in the use of the brute force Metropolis-Hastings algorithm is that we might be spending much computational resources in an uninteresting part of configuration space. To make smarter moves we will add importance sampling (which will be discussed in due time) to the algorithm. This alteration of the algorithm needs a new quantity called the *drift force* of the system[4].

$$\mathbf{F}(\mathbf{r}) = \sum_{k=1}^{N} \mathbf{F}_k(\mathbf{r}) = \sum_{k=1}^{N} \frac{2\nabla_k \Psi_T(\mathbf{r})}{\Psi_T(\mathbf{r})}. \tag{14}$$

[4] Also known as the *quantum force*.

Using the drift force we are able to move towards more interesting parts of configuration space as we are following the gradient of the wavefunction. We will mainly be interested in the drift force of a single particle k as the Monte Carlo sampling involves choosing a single particle at a time.

D. One-Body Densities

In order to make some conclusion towards where the particles in our systems are most prone to exist, we need a good way to visualise our wavefunctions. If we are to visualize the combined wavefunction of two three-dimensional particles we need four dimensions per particle and can subtract one dimension as we look at a single probability density yielding a grand total of seven dimensions. The fact that we will be studying systems of up to $N = 500$ particles only adds to the headache. We therefore introduce the single-particle density function, or *one-body density* $\rho(\mathbf{r})$. The one-body density is defined by the integral over all coordinates except one[1],

$$\rho(\mathbf{r}_1) \equiv \int \mathrm{d}\mathbf{r}_2 \ldots \mathrm{d}\mathbf{r}_N \left|\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)\right|^2. \tag{15}$$

The one-body density represents the distribution of all the particles in the system. Furthermore, $\rho(\mathbf{r}_1)\mathrm{d}\mathbf{r}_1$ will yield the probability of finding any particle within the volume element $\mathrm{d}\mathbf{r}_1$. By convention the one-body density should be normalized such that the following relation holds,

$$N = \int \rho(\mathbf{r}_1) \mathrm{d}\mathbf{r}_1. \tag{16}$$

In other words, the integral over the entire volume in $\mathbf{r}_1$-direction gives us the total number of particles in our system. This can be a bit odd at first sight as we are used to getting unity when we integrate over the entire probability density.[2]

It is best to think of the one-body density as the multivariate probability density for the entire system collapsed down to one coordinate in a representation than looks very familiar to every physicist with a basic knowledge of quantum mechanics.

III. SYSTEMS

To model the trapped bosonic gas particles we use the potential

$$v(\mathbf{r}) = \begin{cases} \frac{1}{2} m\omega^2 r^2 & \text{(S)}, \\ \frac{1}{2} m\left[\omega^2(x^2+y^2) + \omega_z^2 z^2\right] & \text{(E)}, \end{cases} \tag{17}$$

where we can choose between a spherical (S) or an elliptical (E) harmonic oscillator trap. The full Hamiltonian

of the system is given by

$$H = \sum_{i=1}^{N} h(\mathbf{r}_i) + \sum_{i<j}^{N} w(\mathbf{r}_i, \mathbf{r}_j), \tag{18}$$

where the single particle one-body operator, $h(\mathbf{r}_i)$, is given by

$$h(\mathbf{r}_i) = -\frac{\hbar^2}{2m}\nabla_i^2 + v(\mathbf{r}_i), \tag{19}$$

for equal masses. The two-body interaction operator, $w(\mathbf{r}_i, \mathbf{r}_j)$, is

$$w(\mathbf{r}_i, \mathbf{r}_j) = \begin{cases} \infty & |\mathbf{r}_i - \mathbf{r}_j| \leq a, \\ 0 & |\mathbf{r}_i - \mathbf{r}_j| > a, \end{cases} \tag{20}$$

where a is the hard sphere of the particle. The trial wavefunction we will be looking at is given by

$$\Psi_T(\mathbf{r}) = \Phi_T(\mathbf{r})J(\mathbf{r}), \tag{21}$$

where $\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$. The Slater permanent, $\Phi_T(\mathbf{r})$, is given by the N first single-particle functions

$$\Phi_T(\mathbf{r}) = \prod_{i=1}^{N} \phi(\mathbf{r}_i) = \prod_{i=1}^{N} \exp\big[-\alpha(x_i^2 + y_i^2 + \beta z_i^2)\big]. \tag{22}$$

Here α is the variational parameter of the trial wavefunction.[5] The Jastrow factor, $J(\mathbf{r})$ contributes with correlations between the particles and is represented as

$$J(\mathbf{r}) = \prod_{j<k}^{N} f(a, \mathbf{r}_j, \mathbf{r}_k), \tag{23}$$

where the correlation wavefunctions are given by

$$f(a, \mathbf{r}_j, \mathbf{r}_k) = \begin{cases} 0 & |\mathbf{r}_j - \mathbf{r}_k| \leq a, \\ \left(1 - \frac{a}{|\mathbf{r}_j - \mathbf{r}_k|}\right) & |\mathbf{r}_j - \mathbf{r}_k| > a. \end{cases} \tag{24}$$

We will for brevity define the notation

$$\phi_i \equiv \phi(\mathbf{r}_i), \tag{25}$$

$$r_{jk} \equiv \left|\mathbf{r}_j - \mathbf{r}_k\right|. \tag{26}$$

In practice Equation 20 is unnecessary as the situtation where $r_{jk} \leq a$ automatically yields a probability density of zero from Equation 24 thus rejecting the state completely.

[5] We can also treat β as a variational parameter, but we will in this project limit ourselves to a single variational parameter, i.e., α.

A. Non-interacting harmonic oscillators

We start by looking at a simple system of non-interacting harmonic oscillators, where $a = 0$ and $\beta = 1$. This means that there is no hard-shell interaction between the particles and that we use a spherical potential trap. We thus get the trial wavefunction

$$\Psi_T(\mathbf{r}) = \Phi_T(\mathbf{r}) = \prod_{i=1}^{N} \exp\left(-\alpha|\mathbf{r}_i|^2\right). \tag{27}$$

As $a = 0$ the interaction term, $w(\mathbf{r}_i, \mathbf{r}_j)$, vanishes and the Hamiltonian is given by (in the spherical case)

$$H = \sum_{i=1}^{N} h(\mathbf{r}_i) = \sum_{i=1}^{N}\left(-\frac{\hbar^2}{2m}\nabla_i^2 + \frac{1}{2}m\omega^2 r_i^2\right). \tag{28}$$

To find the drift force and the local energy we have to compute the gradient and the Laplacian of the trial wavefunction. The gradient is given by

$$\nabla_k \Psi_T(\mathbf{r}) = -2\alpha \mathbf{r}_k \Psi_T(\mathbf{r}), \tag{29}$$

whereas the Laplacian yields

$$\nabla_k^2 \Psi_T(\mathbf{r}) = \big(-2d\alpha + 4\alpha^2 r_k^2\big)\Psi_T(\mathbf{r}). \tag{30}$$

Here d is the dimensionality of the problem determined by $\mathbf{r}_k \in \mathbb{R}^d$. We can thus use the gradient to find an expression for the drift force for particle k.

$$\mathbf{F}_k(\mathbf{r}) = -4\alpha \mathbf{r}_k. \tag{31}$$

Using the Laplacian we can compute the kinetic term in the expression for the local energy. We get

$$E_L(\mathbf{r}) = \sum_{i=1}^{N}\left(-\frac{\hbar^2}{2m}\big[-2d\alpha + 4\alpha^2 r_i^2\big] + \frac{1}{2}m\omega^2 r_i^2\right). \tag{32}$$

In natural units, $\hbar = c = 1$, and with the mass set to unity this reduces to

$$E_L(\mathbf{r}) = \alpha d N + \left(\frac{1}{2}\omega^2 - 2\alpha^2\right)\sum_{i=1}^{N} r_i^2. \tag{33}$$

It is worth noting that for $\alpha = \omega/2$[6] we get a stable value for the local energy.

$$E_L(\mathbf{r}; \alpha = \omega/2) \equiv E_L = \frac{\omega d N}{2}, \tag{34}$$

which turns out to be the exact minimum for the ground state energy of this system as we shall see.

[6] Note that α is required to be positive.

1. Exact variational energy

As the system is non-interacting and consisting of Gaussians we can find an expression for the exact energy as a function of the variational parameter α, i.e.,

$$E(\alpha) = \frac{\langle \Psi_T | \, H \, | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle} = \left(\frac{\hbar^2 \alpha}{2m} + \frac{m\omega^2}{8\alpha} \right) dN. \tag{35}$$

By minimizing this expression with respect to α we get

$$\frac{\mathrm{d}E(\alpha)}{\mathrm{d}\alpha} = 0 \implies \alpha_0 = \frac{m\omega}{2\hbar}, \tag{36}$$

which in natural units reduces to $\alpha_0 = \omega/2$. The energy at this value of α (in natural units) is then

$$E(\alpha_0) = \frac{\omega d N}{2}, \tag{37}$$

in perfect accordance with Equation 34 which yields high hopes for the latter equation and the expected minimum.

B. Interacting hard sphere bosons

Moving to the full elliptical system for $\beta \neq 1$ and setting $a \neq 0$ we use the trial wavefunction from Equation 21. We rewrite the Jastrow factor to

$$J(\mathbf{r}) = \exp\left(\sum_{j<l}^{N} u(r_{jl}) \right), \tag{38}$$

where

$$u(r_{jk}) = \ln\big[f(a, \mathbf{r}_j, \mathbf{r}_k) \big] \equiv u_{jk} \tag{39}$$

The drift force for particle k in this system becomes

$$\mathbf{F}_k(\mathbf{r}) = 2\left(\frac{\nabla_k \phi_k}{\phi_k} + \sum_{m \neq k}^{N} \nabla_k u_{km} \right). \tag{40}$$

The Laplacian of particle k divided by the trial wavefunction is then

$$\frac{\nabla_k^2 \Psi_T(\mathbf{r})}{\Psi_T(\mathbf{r})} = \frac{\nabla_k^2 \phi_k}{\phi_k} + 2\frac{\nabla_k \phi_k}{\phi_k} \sum_{m \neq k}^{N} \frac{\Delta \mathbf{r}_{km}}{r_{km}} \frac{\partial u_{km}}{\partial r_{km}} + \sum_{m \neq k}^{N} \left(\frac{d-1}{r_{km}} \frac{\partial u_{km}}{\partial r_{km}} + \frac{\partial^2 u_{km}}{\partial r_{km}^2} \right) + \left(\sum_{m \neq k}^{N} \frac{\Delta \mathbf{r}_{km}}{r_{km}} \frac{\partial u_{km}}{\partial r_{km}} \right)^2, \tag{41}$$

where $\Delta \mathbf{r}_{km} = \mathbf{r}_k - \mathbf{r}_m$ and d is the dimensionality. Evaluating the gradient and the Laplacian of the single particle functions yields

$$\frac{\nabla_k \phi_k}{\phi_k} = -2\alpha(x_k \mathbf{e}_i + y_k \mathbf{e}_j + \beta z_k \mathbf{e}_k), \tag{42}$$

$$\frac{\nabla_k^2 \phi_k}{\phi_k} = -2\alpha\big(d - 1 + \beta\big) + 4\alpha^2\big(x_k^2 + y_k^2 + \beta^2 z_k^2\big). \tag{43}$$

Repeating this for the correlation functions yields

$$\frac{\partial u_{km}}{\partial r_{km}} = \frac{a}{r_{km}(r_{km} - a)}, \tag{44}$$

$$\frac{\partial^2 u_{km}}{\partial r_{km}^2} = \frac{a^2 - 2a r_{km}}{r_{km}^2 (r_{km} - a)^2}. \tag{45}$$

The local energy is thus

$$E_L(\mathbf{r}) = \sum_{k=1}^{N} \left\{ -\frac{\hbar^2}{2m} \frac{\nabla_k^2 \Psi_T(\mathbf{r})}{\Psi_T(\mathbf{r})} + v(\mathbf{r}_k) \right\}, \tag{46}$$

where the Laplacian is given by Equation 41. We will avoid writing out the full expression for the local energy for this system due to the size of the equation.

1. Scaling the system

We now introduce a scaled distance $\mathbf{r}' = \mathbf{r}/a_{\mathrm{ho}}$ [3], where

$$a_{\mathrm{ho}} = \sqrt{\frac{\hbar}{m\omega}}. \tag{47}$$

We can then rewrite the Hamiltonian for the elliptic potential in terms of this new distance. By doing a variable substitution for each direction in the Laplace operator we get

$$\nabla_k^2 = \frac{1}{a_{\mathrm{ho}}^2} {\nabla'_k}^2. \tag{48}$$

The one-body part of the Hamiltonian then yields

$$H = \sum_{k=1}^{N} \left\{ -\frac{\hbar^2}{2m} \nabla_k^2 + \frac{1}{2} m \left[\omega^2 (x_k^2 + y_k^2) + \omega_z^2 z_k^2 \right] \right\} \tag{49}$$

$$= \frac{\hbar\omega}{2} \sum_{k=1}^{N} \left\{ -{\nabla'_k}^2 + \left[{x'_k}^2 + {y'_k}^2 + \lambda^2 {z'_k}^2 \right] \right\}, \tag{50}$$

where we have introduced the dimensionless frequency $\lambda = \omega_z/\omega$. The wavefunction does not need to be scaled as the exponentials are already dimensionless. We only

need to use the scaled hard core radius $a' = a/a_{\text{ho}}$. We are interested in using the radius $a' = 0.0043$ as in the references [3] and [4].

IV. ALGORITHMS

In this project we rely on a Monte Carlo approach of random sampling to obtain numerical results. We simulate random walks over a volume in order to find optimal parameters in our trial wavefunctions. The most common of such methods, which we make use of herein, is the Metropolis-Hastings algorithm.

A. The Metropolis-Hastings Algorithm

The Metropolis-Hastings algorithm can in our particular situation be condensed down to the following steps:

1. The system is initialised by a certain number N of randomly generated positions, $\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$, which represent the particles[7]. This allows us to evaluate the wavefunction at these points and compute the local energy $E_L(\mathbf{r})$.

2. The initial configuration is changed by setting a new position for one of these particles. The particle is picked at random and is only allowed to move a predetermined step length.

3. A ratio, q, between the probability density of the new wavefunction and the previous probability density is computed. We accept the step if $q \geq 1$, else we accept the step with a probability p. The latter is done by drawing a random uniform number, $x \in [0, 1]$, and checking if $q > x$. This transition probability decides if the particle move is rejected or accepted.

4. If the particle movement is accepted we compute the local energy for the new system. If not, we reset the system to the previous state and reuse the last local energy.

5. Accumulate the local energy. Repeat the steps until convergence is reached or a set number of iterations has been done. When finished divide the accumulated energy by the number of cycles to compute the expectation value of the energy.

The algorithm described above can be applied in an "exhaustive" search of the parameter space in order to find the optimal parameters. Whether a proposed move is accepted or not is determined by a *transition probability*. The strength of the algorithm lies in that the transition algorithm need not be known. For example, the simplest case is to accept the new state, i.e., the new position for the random walker, if the ratio

$$q(\mathbf{r}_{i+1}, \mathbf{r}_i) = \frac{\left|\Psi_T(\mathbf{r}_{i+1})\right|^2}{\left|\Psi_T(\mathbf{r}_i)\right|^2}, \tag{51}$$

where $\mathbf{r}_{i+1}$ are all the positions at step $i + 1$, is greater than a uniform probability $p \in [0, 1]$. This transition criteria results in what we call the brute force Metropolis-Hastings algorithm. By counting the number of accepted steps and dividing by the total number of Monte Carlo cycles we get the *acceptance probability*, denoted A. The acceptance probability helps in choosing a decent length for the proposed new step. If the acceptance probability is very low we can lower the step length to get more accepted steps and vice versa.

1. Initial distribution of particles

Prior to performing the Metropolis-Hastings algorithm we have to have an initial distribution of the particles. As the algorithm is supposed to move particles in the right direction this distribution can be *somewhat* arbitrary. There are however a few caveats to this approach.

1. The distribution is unphysical. This results in the particles taking up positions which can yield bogus values for the local energy. A remedy is to allow for some *thermalization steps*, i.e., scatter the particles and run the algorithm for a set number of steps prior to starting the actual sampling.

2. The initial *spread* of the particles is too small or too large. The former can lead to an acceptance ratio of zero if no steps are accepted due to overlap. This is something that can happen to the interacting system we are studying as an overlap between the particles makes the wavefunction evaluate to zero thus rejecting all steps in the ratio test. The latter can lead to the system never reaching a true state as they are scattered too far away from the interesting regions in the configuration space.

3. The system imposes some restrictions on the distribution. This can mean that we have to distribute the particles in a certain way to make sure that we don't break the restrictions. This resembles the former case when the spread is too small for our systems.

In the case of the interacting system we are looking at we have to take some measures to make sure that the initial spread does not break the Monte Carlo cycles. We therefore distribute all the particles on a grid with a distance larger than the diameter of the hard sphere before letting the system thermalize. After the thermaliztion we start the sampling.

[7] Also known as *random walkers*.

2. Importance Sampling

A problem with the naïve Metropolis-Hastings sampling approach described above is that the sampling of position space is done with no regard for where we are likely to find a particle. This problem can be remedied through *importance sampling*. It is reasonable to assume that the particles we erratically scatter in space are prone to movement towards the peaks of the probability density as dictated by the wave function. Consider therefore the Fokker-Planck equation,

$$\frac{\partial \Psi_T}{\partial t} = D\nabla \cdot (\nabla - \mathbf{F})\,\Psi_T, \tag{52}$$

which describes the evolution in time of a probability density function. In our case this is the trial wavefunction Ψ_T[8]. Originaly an equation that models diffusion, we have a diffusion term D and the drift force Equation 14. In our case the diffusion term D is simply $1/2$ from the kinetic energy (in natural units).

We use the Langevin equation to find the new position of the particle.

$$\frac{\partial \mathbf{r}}{\partial t} = D\mathbf{F}(\mathbf{r}) + \boldsymbol{\eta}, \tag{53}$$

where $\boldsymbol{\eta}$ is a uniformly distributed stochastic variable for each dimenion. Solving Langevin's equation by Euler's method gives a recursive relation for the subsequent new positions of a particle.

$$\mathbf{r}_{i+1} = \mathbf{r}_i + D\mathbf{F}(\mathbf{r}_i)\delta t + \boldsymbol{\xi}\sqrt{\delta t}, \tag{54}$$

for a given time step δt [9] and a normally distributed stochastic variable $\boldsymbol{\xi}$.

Now we need to change the ratio computation in the Metropolis-Hastings algorithm to something that takes the new sampling method into account.

$$q(\mathbf{r}_{i+1}, \mathbf{r}_i) = \frac{G(\mathbf{r}_i, \mathbf{r}_{i+1}, \delta t)\left|\Psi_T(\mathbf{r}_{i+1})\right|^2}{G(\mathbf{r}_{i+1}, \mathbf{r}_i, \delta t)\left|\Psi_T(\mathbf{r}_i)\right|^2}, \tag{55}$$

where G is the Green's function of the Fokker-Planck equation given by

$$G(\mathbf{r}_i, \mathbf{r}_{i+1}, \delta t) = \exp\left(-\frac{\left[\mathbf{r}_i - \mathbf{r}_{i+1} - D\mathbf{F}(\mathbf{r}_{i+1})\delta t\right]^2}{4D\delta t}\right) \times \frac{1}{(4\pi D\delta t)^{dN/2}}, \tag{56}$$

where d is the dimensionality. The new ratio, q, is still subjected to the same ratio test as in the brute force Metropolis-Hastings algortihm.

[8] The trial wavefunction is not a probability density, but the Fokker-Planck equation is robust enough to handle the wavefunction as is.

[9] Bear in mind that Equation 53 is only valid as $\Delta t \to 0$, a property stemming from the use of Euler's method.

B. Statistical Analysis

If the results of the metropolis sampling were completely uncorrelated, it would be enough to compute the standard deviation in a familiar way,

$$\sigma = \sqrt{\frac{1}{N}(\langle E_L^2\rangle - \langle E_L\rangle^2}, \tag{57}$$

where N is the number of samples, or Monte-Carlo cycles, in the experiment. However, it is reasonable to assume that the data we are dealing with in this study is liable to suffer from *autocorrelation* and Equation 57 does not hold. The prevailing definition of autocorrelation in a data stream or signal is correlation between a delay of the signal and the original signal. One would be interested to find the delay, or lag, in the signal at which the "self-correlation" is highest. We refer to this spacing as d, and define the following correlation function,

$$f_d = \frac{1}{n-d}\sum_{k=1}^{n-d}(x_k - \bar{x}_n)(x_{k+d} - \bar{x}_n). \tag{58}$$

The keen reader would have noticed that the function f_d in Equation 58 would be equal to the sample variance for $d = 0$. We can now define the *autocorrelation function*

$$\kappa_d = \frac{f_d}{\mathrm{Var}(x)}, \tag{59}$$

which is equal to 1 if the data exhibits no autocorrealtions, equating to $d = 0$. From the autocorrelation function in Equation 59 we in turn define the *autocorrelation time*,

$$\tau = 1 + 2\sum_{d=1}^{n-1}\kappa_d. \tag{60}$$

Notice that the autocorrelation time is 1 for a correlation free experiment.

We are now able to make a correction to the expression for the standard deviation improving on Equation 57 by taking correlation into account,

$$\sigma = \sqrt{\frac{1 + 2\tau/\Delta t}{N}(\langle E_L^2\rangle - \langle E_L\rangle^2)}, \tag{61}$$

where Δt is the time between each sample. The main problem at this point is to find τ, which we do not know for any given system and it is generally very expensive to compute. In order to find a good estimate of τ we use a procedure called *blocking*.

1. Blocking

In the method of blocking we group the samples into blocks of increasing size. If one where to compute the

standard deviation for each block, one should see the variance increase with the block size. The standard deviation would only increase up to a certain point, from whence it would stay almost constant. What is happening is that we have reached a point where a particular sample from one block is no longer correlated with a corresponding sample from an adjacent block. The block size for this point of convergence now functions as an estimate for the autocorrelation time τ.

Instead of going by this "manual" method of looking at charts, we will construct a test statistic and let the computer do the necessary considerations in a more "automatic way.

The easiest way to do this is by use of the *automatic blocking scheme*, nicely illustrated with a flow chart in Figure 1. The input parameter for the method is an array of the local energies, but it can be any array of time series data, which is why it is herein referred to as $\mathbf{X}$. Any array of sequential data, where any particular data point depends on the previous data point inhibits all attributes of a time series. Our array of local energies is therefore most definitely a time series.

From $\mathbf{X}$ we compute the variance and sample covariance between the series itself and a one-point lag of the series. This covariance is usually referred to as the first order autocovariance.

$$\gamma_i = \frac{1}{n}\left(\sum_{k=1}^{n}\left((\mathbf{X}_i)_k - \bar{\mathbf{X}}\right)\right) \times \left(\sum_{k=0}^{n-1}\left((\mathbf{X}_i)_k - \bar{\mathbf{X}}\right)\right). \tag{62}$$

Note that i represents the iteration in the blocking process whereas k is the index into the array $\mathbf{X}_i$.

Now we "block"! The next step is to perform a transformation of the our time series $\mathbf{X}_i$, where we end up with an array that is half the length of the original time series,

$$(\mathbf{X}_i)_k = \frac{1}{2}\left((\mathbf{X}_i)_{2k-1} + (\mathbf{X}_i)_{2k}\right). \tag{63}$$

This process is were the term "blocking" stems from. There are a number of ways to perform the blocking - we could have picked the data points used in the new $(i+1)$ array randomly, or more orderly. Here we have picked them sequentially. We have obtained $n/2$ new "primed" stochastic variables. Assuming that the reader is playing close attention, it is now needless to say that it is much easier to perform these sequential blocking steps if we have an array of length $n = 2^d$, where d is integer. Otherwise, problems would arise because one would have arrays of different lengths when the original array is split in two. The ailment is easy to remedy by excluding an observation from the longest resulting array after splitting.

For each new time series $\mathbf{X}_{i+1}$ a new sample variance and sample covariance is computed. The blocking transformation is allowed to continue until we run out of data, i.e. the length of $\mathbf{X}_i < 2$. In other words, there is only a single number left. We now compute a test statistic,

$$M_j = \sum_{k=j}^{d-1} n_k \frac{\left[(n_k-1)\sigma_k^2/n_k^2 + \gamma_k\right]^2}{\sigma_k^4}, \tag{64}$$

As luck would have it, this number M is χ^2-distributed. The following step would be to look up a statistical table and find the first k, such that $M_k \leq q_{d-k}(1-\alpha)$, where q is the statistical limit and α is your favorite significance level. Finally, the correct sample variance is σ_k^2/n_k. In the rest of the project we use the labeling

$$\sigma_b^2 \equiv \frac{\sigma_k^2}{n_k}, \tag{65}$$

for the correct sample variance found from the blocking method. Again, see Figure 1 for a summary of the algorithm.

C. Gradient Descent

Scanning the entire space of variational parameters can be a somewhat despairing task. To aid us in the search we therefore implement the method of gradient descent[10]. In any optimum we generally have that the derivative should equal zero. Our goal is to find the minimum of the expected local energy with respect to the variational parameters $\boldsymbol{\alpha}$.

$$\nabla_{\boldsymbol{\alpha}} \left\langle E_L(\mathbf{R};\boldsymbol{\alpha})\right\rangle = 0. \tag{66}$$

The foundation for the method of gradient descent lies in the fact that some analytic [11] function $\mathbf{F}(\boldsymbol{\alpha})$ decreases fastest in the direction of $-\nabla_{\boldsymbol{\alpha}}\cdot\mathbf{F}(\boldsymbol{\alpha})$. This means that we could eventually get to an ensuing minimum by trailing along the pathway of the following recursive relation,

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k - \gamma\nabla_{\boldsymbol{\alpha}}\cdot\mathbf{F}(\boldsymbol{\alpha}_k). \tag{67}$$

If γ is small enough we should have $\mathbf{F}(\boldsymbol{\alpha}_k) \geq \mathbf{F}(\boldsymbol{\alpha}_{k+1})$ for all k and hopefully we will experience convergence to the desired minimum.

The variational gradient of the expectation value of the energy is thus

$$\nabla_{\boldsymbol{\alpha}}\left\langle E_L\right\rangle = 2\left(\left\langle E_L\frac{1}{\Psi_T}\nabla_{\boldsymbol{\alpha}}\Psi_T\right\rangle - \left\langle\frac{1}{\Psi_T}\nabla_{\boldsymbol{\alpha}}\Psi_T\right\rangle\left\langle E_L\right\rangle\right). \tag{68}$$

To evaluate this expression we sample the local energy, the parameter gradient of the wavefunction, and the combined expression of these at each Monte Carlo cycle.

---

[10] Also known as steepest descent, easily confused with the steepest descent method for integation.

[11] We use the term analytic quite recklessly.

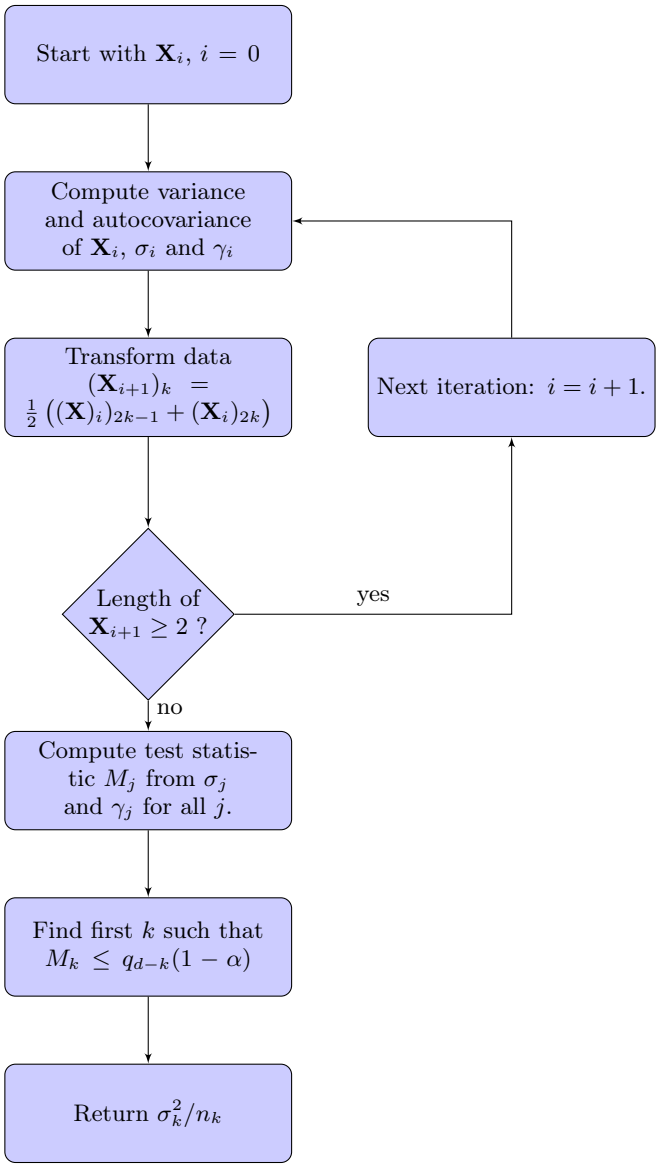

FIG. 1: Flow chart depicting the automatic blocking method for finding a better estimate for the sample variance.

1. *Stability of the algorithm*

When employing the method of gradient descent to locate a minimum value for the variational parameters we expect some issues due to stability of the Monte Carlo methods and the steepness of the expected energy curves. If our guess of the location of the minimum is far off we can expect the method to diverge as the gradient will become huge. This will yield a new value for $\boldsymbol{\alpha}$ in Equation 67 which will be further away and might result in a local minimum "trapping" the method. To prevent this from happening we have added a limit to the allowed change in $\boldsymbol{\alpha}$. If the change in $\boldsymbol{\alpha}$ is larger than a set order of magnitude we keep dividing $\boldsymbol{\alpha}$ with this number until this criteria is fulfilled.

This is a so-called "hacky" solution, but it does the trick. It is also the reason why the resultant figure in Figure 4 looks the way it does for $\alpha_0 = 0.2$ and $\alpha_0 = 0.3$.

TABLE I: One particle in one dimension for the analytic expression with 2^{21} Monte Carlo cycles and a step length of 0.5.

α	$\langle E_L \rangle$	σ	σ_b	A	t_{CPU}[s]
0.30	0.56244	0.00026	0.00129	0.892	0.43
0.34	0.53652	0.00019	0.00094	0.885	0.38
0.38	0.51863	0.00013	0.00063	0.878	0.37
0.42	0.50805	0.00009	0.00038	0.872	0.38
0.46	0.50201	0.00004	0.00017	0.866	0.37
0.50	0.50000	0.00000	0.00000	0.860	0.37
0.54	0.50183	0.00004	0.00015	0.855	0.38
0.58	0.50566	0.00007	0.00029	0.850	0.38
0.62	0.51175	0.00011	0.00041	0.845	0.37
0.66	0.51906	0.00014	0.00052	0.840	0.37
0.70	0.52883	0.00017	0.00062	0.836	0.37

V. RESULTS

A. Non-Interacting Systems

Initially we are interested in seeing how our implementation performs and to what extent it functions properly. The logical thing to do would be to look at a non-interacting system with a simple Gaussian wavefunction and harmonic oscillator potential. Such a system has a simple closed-form analytic solution[12], which is further simplified by the "god-given" analytic units (Equation 33). This expression is minimised for $\alpha = \omega/2$. We also set $\omega = 1$ and get an expression for the expected minimum local energy,

$$E_L(\mathbf{r}) = \frac{dN}{2}, \tag{69}$$

which will be used as the benchmark we are ultimately aiming for.

We will start without importance sampling, that is by employing brute force Monte Carlo sampling. Moreover, we will compute the value for the local energy analytically as well as with a numerical approach. In the numerical approach we employ the second order central difference approximation

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x} \approx \frac{f(x-h) - 2f(x) + f(x+h)}{h^2}. \tag{70}$$

[12] A fact that every undergraduate physics student should be well aware of after finishing their first course in quantum physics

TABLE II: One particle in one dimension for the numeric expression with 2^{21} Monte Carlo cycles and a step length of 0.5.

α	$\langle E_L \rangle$	σ	σ_b	A	t_{CPU}[s]
0.30	0.56675	0.00026	0.00135	0.891	0.53
0.34	0.53714	0.00019	0.00096	0.885	0.52
0.38	0.51836	0.00013	0.00062	0.878	0.52
0.42	0.50717	0.00009	0.00038	0.872	0.53
0.46	0.50184	0.00004	0.00017	0.866	0.52
0.50	0.50000	0.00001	0.00001	0.860	0.52
0.54	0.50133	0.00004	0.00015	0.855	0.52
0.58	0.50520	0.00007	0.00028	0.850	0.52
0.62	0.51120	0.00011	0.00039	0.845	0.52
0.66	0.51896	0.00014	0.00052	0.840	0.52
0.70	0.52794	0.00017	0.00061	0.835	0.51

TABLE III: Comparison of energy, standard deviation and the CPU time for $\alpha = 0.5$ for the analytic and numeric schemes in three dimensions and N particles. The step length is 0.5.

	Analytic			Numeric		
N	$\langle E_L \rangle$	σ_b	t_{CPU}[s]	$\langle E_L \rangle$	σ_b	t_{CPU}[s]
1	1.5	0.0	0.50	1.49997	0.00008	0.73
10	15.0	0.0	0.73	14.98278	0.00867	6.40
100	150.0	0.0	4.13	149.81453	0.81501	387.60
500	750.0	0.0	19.16	773.85435	20.69789	9595.02

Table I shows the expected local energy $\langle E_L \rangle$, the uncorrelated standard deviation, σ, the standard deviation found from blocking, σ_b, the accpetance probability, A, and the CPU time for the analytic computation at different values for the variational parameter α. Table II shows the same numbers but for the numeric scheme. One can clearly see, in both the analytic and numeric case, that the uncorrelated standard deviations are lower than the standard deviations found from the blocking method. This means that autocorrelated data can lead to a lower perceived uncertainty than what is the reality. The blocking method takes this fact into account and for this reason, all other standard deviations will be computed with the blocking method from here on.

The first initial observation of note is that we have found an optimum for the system in terms of a minimum energy at $\alpha = 1/2$. Moreover, notice how the standard deviation disappears completely. This should already be apparent from Equation 33, as the term involving particle positions disappears for the optimal α thus rendering the entire Monte Carlo process of moving particles void.

Systems of optimum variational parameter warrants further investigation, as summed up in Table III. The analytical results are quite predictable and monotonous; the expected local energy $\langle E_L \rangle$ is perfectly proportional by a factor 3/2 to the number of particles and the standard deviation σ_b is zero for systems of any number of particles. The numerical results tells a more tortuous tale.

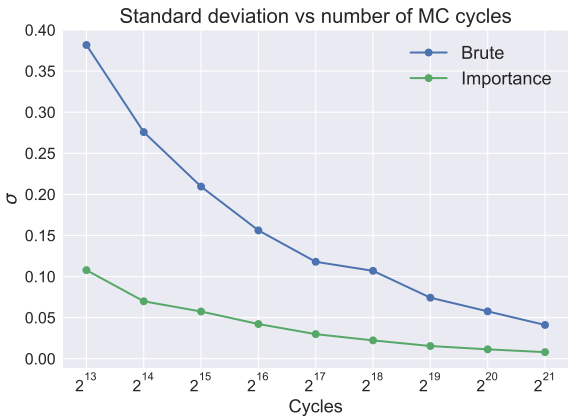

FIG. 2: As the number of Monte Carlo cycles increases, the standard deviation of the expectation value of the energy for the brute force and importance sampling schemes decreases.

As the number of particles increase, the uncertainty also increases rapidly. For the largest system, with $N = 500$ particles, the energy misses its target critically - the true energy is more than one standard deviation away. Lastly, the CPU time is notably higher for the numeric scheme.

In the appendices, Figure 7 shows how both the analytic and numeric approach matches the exact expression for the energy as a function of α for systems of increasing size. We see also here that the uncertainty increases in the numeric scheme as the system gets larger.

1. *Introducing Importance Sampling*

Next we will introduce importance sampling. The first prediction for importance sampling compared to the brute force, naïve method is that we should approach an equilibrium much faster because we are now making moves in a smarter way. This prediction is confirmed in Figure 2 and Figure 3.

In Figure 2 we can see the standard deviation of the local energies for an increasing number of Monte Carlo cycles for both the brute force sampling and the importance sampling methods. We see that the standard deviation decreases for both methods for a larger number of Monte Carlo cycles, but the standard deviation is always lower for the importance sampling method. This is in accordance with our expectations.

We have also compared the expected energy with the exact energy for a different number of Monte Carlo cycles. The result of this analysis is shown in Figure 3. We see that both methods will eventually result in expected energies that are equal to the exact energies as $\left|E - \langle E \rangle\right| \to 0$. However, for the method of importance sampling the error is consistently lower.

Table IV shows the standard deviations and acceptance ratios for different time steps, δt using the impor-

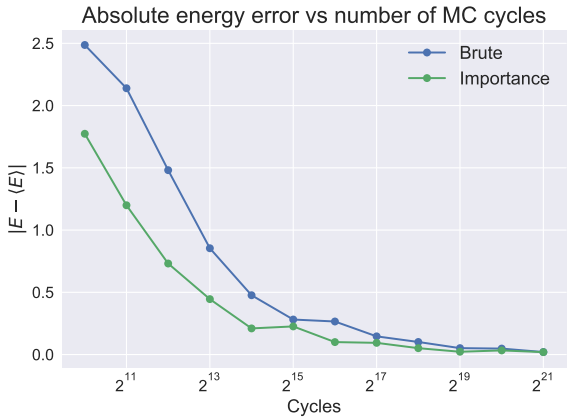

FIG. 3: The absolute difference between the expected energy and the exact energy for both the brute force method and importance sampling.

TABLE IV: Importance sampled simulations of $N = 10$ particles in $d = 3$ dimensions. Average acceptance ratios and standard deviations for eleven values of $\alpha \in [0.3, 0.7]$.

δt	$\bar{\sigma}_b$	$\bar{A}$	$\langle E \rangle \, (\alpha = 1/2)$
2^{+3}	0.09363	0.005	15.000
2^{+2}	0.01549	0.041	15.000
2^{+1}	0.00706	0.216	15.000
$2^{\ 0}$	0.00675	0.510	15.000
2^{-1}	0.00932	0.738	15.000
2^{-2}	0.01578	0.866	15.000
2^{-3}	0.02723	0.933	15.000
2^{-4}	0.05016	0.966	15.000
2^{-5}	0.08474	0.983	15.000
2^{-6}	0.11555	0.992	15.000
2^{-7}	0.17070	0.996	15.000

tance sampling method. The simulations in this table are from the same values of α as in Table I and Table II. The optimum energies for $\alpha = 1/2$ gives the same result for importance sampling as for the naïve way of sampling. Notice that by lowering δt, the uncertainty starts to decrease and then starts to increase again. Moreover, the acceptance ratios increases as $\delta t \to 0$. This happens as the change in the wavefunction grows minuscule and $q(\mathbf{r}_{i+1}, \mathbf{r}_i) \approx 1$.

B. Interacting elliptical harmonic oscillator

We now make a change to the simple system in the previous section in order to make everything a bit more interesting. The potential in the one-body Hamiltonian is perturbed in the z-direction according to Equation 50 by setting $\lambda = \sqrt{8} \approx 2.82843$. This corresponds to enclosing particles of the simulation in an elliptical trap instead of a spherical trap.

TABLE V: Simulation results for $N = 10$ bosonic interacting, elliptical, harmonic oscillators.

α	$\langle E_L \rangle$	σ_b	A	t_{CPU}[s]
0.2	35.17548	0.05399	0.990	24.5
0.3	27.62004	0.02311	0.981	13.6
0.4	24.97850	0.00827	0.972	13.5
0.5	24.39877	0.00030	0.961	13.4
0.6	24.83863	0.00604	0.950	13.3
0.7	25.82855	0.01059	0.938	13.2
0.8	27.22895	0.01436	0.924	13.1

TABLE VI: Simulation results for $N = 50$ bosonic interacting, elliptical, harmonic oscillators.

α	$\langle E_L \rangle$	σ_b	A	t_{CPU}[s]
0.2	181.69371	0.25533	0.987	199.9
0.3	142.61591	0.10782	0.978	195.8
0.4	129.88615	0.04051	0.969	190.4
0.5	127.29926	0.00595	0.957	195.1
0.6	129.97630	0.03300	0.946	193.2
0.7	135.67382	0.05785	0.933	192.0
0.8	143.23238	0.07286	0.921	192.7

We also choose a new trial wavefunction with the elliptical gaussian single particle functions and the Jastrow factor. The hard sphere radius is set to $a = 0.0043$.

Due to similarities with the spherical harmonic oscillator system we make a guess that the true minimum of this new system should be situated close to the minimum of the previous system. Choosing seven values for $\alpha \in [0.2, 0.7]$ we run importance sampling on this system for $N = \{10, 50, 100\}$ particles in $d = 3$ dimensions using 2^{21} Monte Carlo cycles[13] and an additional 10% of the Monte Carlo cycles for thermalization of the system prior to doing any sampling. We have used $\delta t = 0.1$ which yields a high acceptance ratio, but slower convergence.

In Table V the system is set to have $N = 10$ interacting particles. Notice first that the energy is higher in the interacting system than for a comparable non-interacting system (Table III). Moreover, the standard deviation gets lower as the system is close to an optimal variational parameter.

The result of simulations with a higher number of particles, $N = 50$ and $N = 100$, are shown in Table VI and Table VII respectively. We observe that the energy for systems of interacting particles increases rapidly with the size of the system.

[13] This turned out to be a little bit overkill due to the fast convergence of importance sampling.

TABLE VII: Simulation results for $N = 100$ bosonic interacting, elliptical, harmonic oscillators.

α	$\langle E_L \rangle$	σ_b	A	t_{CPU}[s]
0.2	375.04401	0.51475	0.985	745.2
0.3	296.76879	0.20214	0.975	741.0
0.4	270.60830	0.08113	0.945	723.1
0.5	266.37263	0.02020	0.953	702.8
0.6	272.51171	0.07366	0.941	689.1
0.7	285.10285	0.11381	0.929	707.5
0.8	301.40609	0.15458	0.916	707.9

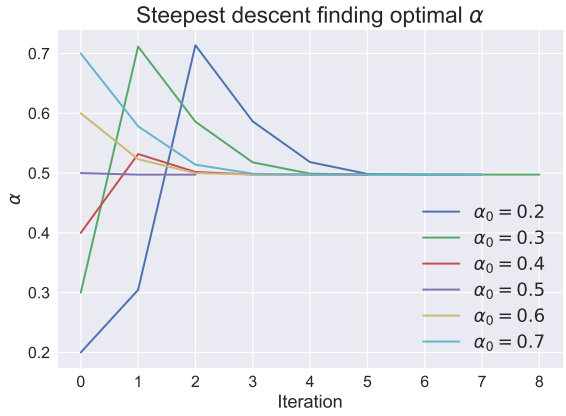

FIG. 4: In this figure we can see the convergence of different starting values for α towards the true value of α that minimizes the expected energy in the interacting case.

1. *Finding the minimum*

Using the method of gradient descent we choose a set of six starting $\alpha_0 \in [0.2, 0.8]$ and try to locate a value for α where the gradient goes to zero. We looked at $N = 10$ with $d = 3$, $\omega = 1$ and $\beta = \lambda = \sqrt{8}$. This yields the value for the minimum α to be

$$\alpha = 0.49744 \pm 0.00002. \tag{71}$$

In Figure 4 we can see how the conjugate gradient method "moves" to find the optimal value of α.

For this value of α we have computed the expected energy and standard deviation for systems of $N = \{10, 50, 100\}$ particles. These results are shown in Table VIII. We have compared with the non-interacting case, i.e., where the Jastrow factor is zero ($a = 0$).

2. *One-Body Densities*

We now look at one-body densities for the value of α found in Equation 71. The results for a system consisting of $N = 100$ particles are displayed in Figure 5. In this figure we see the densities for a system confined to a perturbed elliptic trap with and without interaction. We see that for the interacting case, the one-body density is skewed towards a larger $|\mathbf{r}|$, as is expected. The hard core prevents the bosons from stacking on top of each other spatially. By increasing the hard core radius this trend will be more noticeable. The one-body density of the "ideal" system is also included in the figure.

TABLE VIII: Comparison of an elliptical gaussian trial wavefunction with and without the Jastrow factor using the elliptical trap.

	Interacting		Non-interacting	
N	$\langle E \rangle$	σ_b	$\langle E \rangle$	σ_b
10	24.3986	0.0003	24.1418	0.0002
50	127.2852	0.0077	120.7129	0.0012
100	266.2904	0.0297	241.4248	0.0025

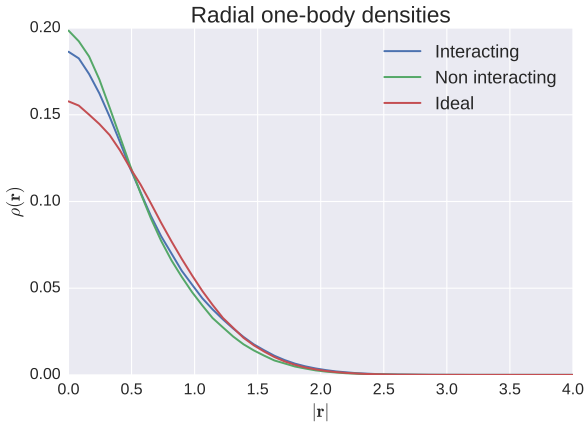

FIG. 5: In this figure the interacting and non-interacting densities correspond to the elliptical potential with and without the Jastrow factor respectively. The "ideal" system is the spherical potential with the non-interacting Gaussian trial wavefunction.

VI. DISCUSSION

A. Validity of Results

The non-interacting tests compared against the analytic "solution" showed to a very high degree the validity of the method. One can see this quite clearly from Figure 7. Because we get results from simulations that correspond to the analytic results we can be fairly certain that the method can be expanded upon to include attributes outside the realm of problems that are analytically solvable. The systems with the elliptic traps and the interacting particles have just such attributes.

That said, the way to get the best confirmation of the results presented above is in comparison to experiment.

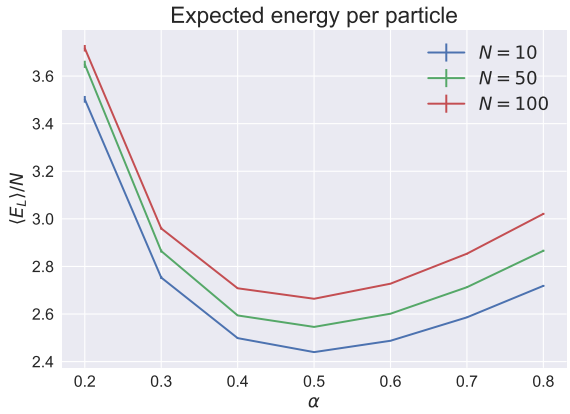

FIG. 6: Expected energy per particle as a function of variational parameter α for a varying number of particles.

This would be quite difficult as the systems presented are very theoretical in nature and most likely far from anything we would see in nature. Without being able to compare our result with anything happening in the real world this entire study would have no value except for the instrumental value, being no more than a theoretical exercise. However, the one-body density is a measure of the role of many-body correlations which in some cases can be related to experiment. For a system of charged particles it can be used to determine a charge distribution, which in turn can be linked to experiment. It would be very fruitful to simulate a system of such particles that is more natural, i.e. real, but we realise the purpose of this study is first and foremost to introduce variational Monte Carlo as a tool in computational physics. We assume, however, that this study is a good starting point for studies of alkali atoms like ^{87}Rb, ^{23}Na or ^{7}Li.

B. Energy per particle

For the non-interacting harmonic oscillator system there is no difference between a new particle and a new dimension[14]. This means that the inclusion of a new particle or dimension yields a linear scaling in the energy. Thus if we compute the energy per particle for systems of varying size we expect to get the same energy. This is most easily seen in Figure 7 and in the expression for the exact energy Equation 35.

This, however, is *not* the case in the interacting system due to the Jastrow factor. Looking at Table V and Table VII we see that multiplying the expected energy of the former with 10 yields a lower number than the latter. We can also see this behavior in Figure 6, where the energy per particle for $N = \{10, 50, 100\}$ have been plotted in the same figure.

The reason for the increase in energy per particle can be ascribed to the repulsive interaction added by the Jastrow factor. The repulsive force can be explained thusly: the Jastrow factor increases the value of the local energy for any given particle if that particle is some distance away from any other particle. The closer the particles are, the more the local energy increases. The Jastrow factor will lower the probability for the particles to get too close to one another by reducing the probability density in the vicinity of other particles. However, if there are many particles in the system they will be closer to another particle and therefore increase the energy of the system.

C. Low density system

It is safe to say that the effect of adding a Jastrow factor is significant, i.e. the increase in energy for the system as the interaction is switched on is much higher than the uncertainty of the energy. A simple look at Figure 6 will be proof enough as the error bars of the figure are not discernible. On the other hand, the energy added to the system by including interaction is not very high compared to the energy of the system as a whole. For the larger of the interacting systems we have looked at, with $N = 100$ particles, the interaction add only a mere 9% extra energy compared to the non-interacting system (see Table VIII). Studying Figure 5 will reveal the same trend in the distribution of the particles.

The reason for this is two-fold.

1. The hard sphere radius is vanishingly small. By increasing the hard sphere radius the particles will contribute more significantly to the local energy. The particles will also be more spread out thus occupying a larger configuration space.

2. The assumption which justifies that adding only two-body interaction is sufficient to model all interactions to a reasonable extent. This assumption is justified by the fact that alkali and atomic hydrogen systems[15] are dilute. For gases of alkali atoms confined in magnetic or optical traps the average distance between atoms is much larger than the range of inter-atomic interaction. The probability is high that atoms very seldom interact and if so, just with one other atom.

[14] This is not entirely true in the case of our implementation of the Metropolis-Hastings algorithm as we draw a single random particle and move all dimensions instead of iterating over all particles and all dimensions.

[15] Systems similar to our toy model.

D. Stability of the program

An interesting thing would be to model large systems of many particles and with large hard sphere radii. Unfortunately, due to the shape of the single particle functions this turns out to be quite tricky. As the single particle function consists of a sum squared positions from a given center, many particles and/or spread out systems will yield a large sum. This means that the Slater permanent grows smaller and smaller making it much harder to keep the system confined. By lowering the frequency and thus lowering the minimum α we can remedy some of this behaviour.

VII. SUMMARY REMARKS

In this study we have built a functional variational Monte Carlo solver framework for a gas following Bose-Einstein statistics confined in an elliptic trap. Monte Carlo methods is a broad class of computational algorithms, whereas the variational Monte Carlo is a subset. Other applications of such methods within the physical sciences includes, but is not limited to, quantum chromodynamics, the modeling of aerodynamic forms, particle physics detector design and galaxy evolution. Needless to say, Monte Carlo methods is a valuable tool for a computational physicist.

To prove the validity of the framework we tested it against an analytically solvable case, an ideal non-interacting system. Thereafter we have shown how our implementation can handle large systems of interacting particles and perturbed traps. Even though this study is somewhat theoretic in nature, we have discussed the possibility of expanding our machinery to a case bearing a larger resemblance to a system found in nature. To this end we have computed one-body densities for the interacting system, which can be compared to real physical properties as measured in an experiment.

[1] H. E. Kristiansen, *Time Evolution of Quantum Mechanical Many-Body Systems*, Master's thesis (2017).
[2] J. Høgberget, *Quantum monte-carlo studies of generalized many-body systems*, Master's thesis (2013).
[3] J. DuBois and H. Glyde, Physical Review A **63**, 023602 (2001).
[4] J. Nilsen, J. Mur-Petit, M. Guilleumas, M. Hjorth-Jensen, and A. Polls, Physical Review A **71**, 053610 (2005).

Appendix A: Computing the exact energy of the non-interacting system

Here we show how to arrive at Equation 35. We need to solve two integrals consisting of N, d-dimensional, Gaussians from Equation 27. We will label the position in the j-th coordinate for the i-th particle by x_{ij}. Starting with the normalization factor we get

$$\langle\Psi_T|\Psi_T\rangle = \int \mathrm{d}\mathbf{r}\big|\Psi_T(\mathbf{r})\big|^2 \tag{A1}$$

$$= \prod_{i=1}^{N}\int \mathrm{d}\mathbf{r}_i \exp\Big(-2\alpha\mathbf{r}_i^2\Big) \tag{A2}$$

$$= \prod_{i=1}^{N}\prod_{j=1}^{d}\int \mathrm{d}x_{ij} \exp\Big(-2\alpha x_{ij}^2\Big) \tag{A3}$$

$$= \prod_{i=1}^{N}\prod_{j=1}^{d}\sqrt{\frac{\pi}{2\alpha}} = \left(\sqrt{\frac{\pi}{2\alpha}}\right)^{dN}. \tag{A4}$$

From the Hamiltonian we get two integrals containing $\mathbf{r}_i^2$ for each particle i.

$$\langle\Psi_T|\,\mathbf{r}_i^2\,|\Psi_T\rangle = \int \mathrm{d}\mathbf{r}\big|\Psi_T(\mathbf{r})\big|^2\,\mathbf{r}_i^2 \tag{A5}$$

$$= \int \mathrm{d}\mathbf{r}_i \exp\Big(-2\alpha\mathbf{r}_i^2\Big)\mathbf{r}_i^2 \times \prod_{j\neq i}^{N}\int \mathrm{d}\mathbf{r}_j \exp\Big(-2\alpha\mathbf{r}_j^2\Big) \tag{A6}$$

$$= \sum_{l=1}^{d}\int\left(\prod_{k=1}^{d}\mathrm{d}x_{ik}\exp\Big(-2\alpha x_{ik}^2\Big)\right)x_{il}^2 \times\left(\sqrt{\frac{\pi}{2\alpha}}\right)^{d(N-1)} \tag{A7}$$

$$= \sum_{l=1}^{d}\frac{1}{4\alpha}\sqrt{\frac{\pi}{2\alpha}}\left(\sqrt{\frac{\pi}{2\alpha}}\right)^{d-1} \times\left(\sqrt{\frac{\pi}{2\alpha}}\right)^{d(N-1)} \tag{A8}$$

$$= \frac{d}{4\alpha}\left(\sqrt{\frac{\pi}{2\alpha}}\right)^{dN}. \tag{A9}$$

We thus have that

$$\frac{\langle\Psi_T|\,\mathbf{r}_i^2\,|\Psi_T\rangle}{\langle\Psi_T|\Psi_T\rangle} = \frac{d}{4\alpha}. \tag{A10}$$

Using Equation 30 we can find the exact energy expression we are looking for.

$$E(\alpha) = \frac{\langle\Psi_T|\,H\,|\Psi_T\rangle}{\langle\Psi_T|\Psi_T\rangle} \tag{A11}$$

$$= \sum_{i=1}^{N}\left\{-\frac{\hbar^2}{2m}\left[-2d\alpha + 4\alpha^2\frac{d}{4\alpha}\right] + \frac{1}{2}m\omega^2\frac{d}{4\alpha}\right\} \tag{A12}$$

$$= \left(\frac{\alpha\hbar^2}{2m} + \frac{m\omega^2}{8\alpha}\right)dN, \tag{A13}$$

which is what we wanted to show.

Appendix B: Computing the drift force and local energy of the full system

Computing the gradient of the wavefunction we get

$$\nabla_k\Psi_T(\mathbf{r}) = \Big[\nabla_k\Phi_T(\mathbf{r})\Big]J(\mathbf{r}) + \Phi_T(\mathbf{r})\nabla_k J(\mathbf{r}). \tag{B1}$$

The gradient of the Slater permament for particle k is given by

$$\nabla_k\Phi_T(\mathbf{r}) = \nabla_k\phi_k\prod_{i\neq k}^{N}\phi_i = \frac{\nabla_k\phi_k}{\phi_k}\Phi_T(\mathbf{r}). \tag{B2}$$

The gradient of the Jastrow factor is given by

$$\nabla_k J(\mathbf{r}) = J(\mathbf{r})\nabla_k\sum_{m<n}^{N}u_{mn} \tag{B3}$$

$$= J(\mathbf{r})\left(\sum_{m=1}^{k-1}\nabla_k u_{mk}\sum_{n=k+1}^{N}\nabla_k u_{kn}\right) \tag{B4}$$

$$= J(\mathbf{r})\sum_{m\neq k}^{N}\nabla_k u_{km}, \tag{B5}$$

where the gradient of the interaction term splits the anti-symmetric sum into two parts. As $r_{ij} = r_{ji}$ we can combine these sums into a single sum. This in total yields the gradient

$$\nabla_k\Psi_T(\mathbf{r}) = \left(\frac{\nabla_k\phi_k}{\phi_k} + \sum_{m\neq k}^{N}\nabla_k u_{km}\right)\Psi_T(\mathbf{r}). \tag{B6}$$

The Laplacian of the trial wavefunction is found by finding the divergence of Equation B6.

$$\nabla_k^2\Psi_T(\mathbf{r}) = \left(\nabla_k\left[\frac{\nabla_k\phi_k}{\phi_k}\right] + \sum_{m\neq k}^{N}\nabla_k^2 u_{km}\right)\Psi_T(\mathbf{r}) \tag{B7}$$

$$+\left(\frac{\nabla_k\phi_k}{\phi_k} + \sum_{m\neq k}^{N}\nabla_k u_{km}\right)^2\Psi_T(\mathbf{r}), \tag{B8}$$

where the squared term came from taking the gradient of the trial wavefunction. To further simplify we divide by the trial wavefunction. This yields

$$\frac{\nabla_k^2\Psi_T(\mathbf{r})}{\Psi_T(\mathbf{r})} = \frac{\nabla_k^2\phi_k}{\phi_k} + 2\frac{\nabla_k\phi_k}{\phi_k}\sum_{m\neq k}\nabla_k u_{km} + \sum_{m\neq k}^{N}\nabla_k^2 u_{km} + \left(\sum_{m\neq k}^{N}\nabla_k u_{km}\right)^2. \quad \text{(B9)}$$

From here, the next step is to find the gradient and the Laplacian of the single particle functions, ϕ_k, and the interaction functions u_{km}. For the single particle functions we use Cartesian coordinates when finding the derivatives while for the interaction functions we use spherical coordinates and do a variable substitution. Beginning with the gradient of the single particle functions we get

$$\nabla_k\phi_k = \nabla_k \exp\left[-\alpha(x_k^2 + y_k^2 + \beta z_k^2)\right] \quad \text{(B10)}$$
$$= -2\alpha(x_k\mathbf{e}_i + y_k\mathbf{e}_j + \beta z_k\mathbf{e}_k)\phi_k. \quad \text{(B11)}$$

Note that the subscripts on the unit vectors $\mathbf{e}_i$ are *not* the same as the subscripts used for its components. The Laplacian yields

$$\nabla_k^2\phi_k = \Big[-2\alpha\big(d-1+\beta\big) + 4\alpha^2\big(x_k^2 + y_k^2 + \beta^2 z_k^2\big)\Big]\phi_k, \quad \text{(B12)}$$

with d as the dimensionality of the problem. In order to derive the interaction functions we have to do a variable substitution using $r_{km} = |\mathbf{r}_k - \mathbf{r}_m|$. We can then rewrite the ∇_k-operator as

$$\nabla_k = \nabla_k \frac{\partial r_{km}}{\partial r_{km}} = \nabla_k r_{km}\frac{\partial}{\partial r_{km}} \quad \text{(B13)}$$
$$= \frac{\mathbf{r}_k - \mathbf{r}_m}{r_{km}}\frac{\partial}{\partial r_{km}}. \quad \text{(B14)}$$

Applying this version of the ∇_k-operator to u_{km} yields

$$\nabla_k u_{km} = \frac{\mathbf{r}_k - \mathbf{r}_m}{r_{km}}\frac{\partial u_{km}}{\partial r_{km}}. \quad \text{(B15)}$$

For the Laplacian we switch a little back and forth between the two ways of representing the ∇_k-operator. We thus get

$$\nabla_k^2 u_{km} = \frac{\nabla_k\mathbf{r}_k}{r_{km}}\frac{\partial u_{km}}{\partial r_{km}} + \left[\nabla_k\frac{1}{r_{km}}\right](\mathbf{r}_k - \mathbf{r}_m)\frac{\partial u_{km}}{\partial r_{km}} + \frac{\mathbf{r}_k - \mathbf{r}_m}{r_{km}}\nabla_k\frac{\partial u_{km}}{\partial r_{km}} \quad \text{(B16)}$$
$$= \frac{d}{r_{km}}\frac{\partial u_{km}}{\partial r_{km}} - \frac{(\mathbf{r}_k - \mathbf{r}_m)^2}{r_{km}^3}\frac{\partial u_{km}}{\partial r_{km}} + \frac{(\mathbf{r}_k - \mathbf{r}_m)^2}{r_{km}^2}\frac{\partial^2 u_{km}}{\partial r_{km}^2} \quad \text{(B17)}$$
$$= \frac{d-1}{r_{km}}\frac{\partial u_{km}}{\partial r_{km}} + \frac{\partial^2 u_{km}}{\partial r_{km}^2}, \quad \text{(B18)}$$

where d is again the dimensionality of the problem. In total we can state an intermediate version of the Laplacian occuring in the local energy as

$$\frac{\nabla_k^2\Psi_T(\mathbf{r})}{\Psi_T(\mathbf{r})} = \frac{\nabla_k^2\phi_k}{\phi_k} + 2\frac{\nabla_k\phi_k}{\phi_k}\sum_{m\neq k}^{N}\frac{\mathbf{r}_k - \mathbf{r}_m}{r_{km}}\frac{\partial u_{km}}{\partial r_{km}} + \sum_{m\neq k}^{N}\left(\frac{d-1}{r_{km}}\frac{\partial u_{km}}{\partial r_{km}} + \frac{\partial^2 u_{km}}{\partial r_{km}^2}\right) + \sum_{m,n\neq k}^{N}\frac{\mathbf{r}_k - \mathbf{r}_m}{r_{km}}\frac{\mathbf{r}_k - \mathbf{r}_n}{r_{kn}}\frac{\partial u_{km}}{\partial r_{km}}\frac{\partial u_{kn}}{\partial r_{kn}}. \quad \text{(B19)}$$

Moving on to the derivatives of the interaction terms, u_{km}, to get an explicit expression for the Laplacian.

$$\frac{\partial u_{km}}{\partial r_{km}} = \frac{a}{r_{km}(r_{km} - a)}, \quad \text{(B20)}$$
$$\frac{\partial^2 u_{km}}{\partial r_{km}^2} = \frac{a^2 - 2ar_{km}}{r_{km}^2(r_{km} - a)^2}. \quad \text{(B21)}$$

The local energy and the drift force can now be found by combining these expressions. We will not write out the explicit expressions as these will be called by separated functions in our programs.

Appendix C: Variational parameter gradient of the expectation energy

Here we show how to arrive at the expression shown in Equation 68. This requires us to restrict our view to real trial wavefunctions, i.e., $\Psi_T^* = \Psi_T$, which is the case for

the wavefunctions we are exploring.

$$\nabla_{\boldsymbol{\alpha}} \langle E_L \rangle = \nabla_{\boldsymbol{\alpha}} \left(\frac{\int \mathrm{d}\mathbf{r} \Psi_T^* H \Psi_T}{\int \mathrm{d}\mathbf{r} |\Psi_T|^2} \right) \tag{C1}$$

$$= -\frac{2\left(\int \mathrm{d}\mathbf{r} \Psi_T H \Psi_T\right)}{\left(\int \mathrm{d}\mathbf{r} \Psi_T\right)^2} \int \mathrm{d}\mathbf{r} \Psi_T \nabla_{\boldsymbol{\alpha}} \Psi_T \tag{C2}$$

$$+ \frac{2}{\int \mathrm{d}\mathbf{r} \Psi_T^2} \int \mathrm{d}\mathbf{r} \Psi_T H \left[\nabla_{\boldsymbol{\alpha}} \Psi_T\right] \tag{C3}$$

$$= -2 \langle E_L \rangle \frac{1}{\int \mathrm{d}\mathbf{r} \Psi_T^2} \int \mathrm{d}\mathbf{r} \Psi_T^2 \left(\frac{\nabla_{\boldsymbol{\alpha}} \Psi_T}{\Psi_T} \right) \tag{C4}$$

$$+ \frac{2}{\int \mathrm{d}\mathbf{r} \Psi_T^2} \int \mathrm{d}\mathbf{r} \Psi_T^2 E_L \left(\frac{\nabla_{\boldsymbol{\alpha}} \Psi_T}{\Psi_T} \right). \tag{C5}$$

We now use the definition of the expectation value to group terms together. We are thus left with

$$\nabla_{\boldsymbol{\alpha}} \langle E_L \rangle = 2 \left(\left\langle E_L \frac{1}{\Psi_T} \nabla_{\boldsymbol{\alpha}} \Psi_T \right\rangle - \langle E_L \rangle \left\langle \frac{1}{\Psi_T} \nabla_{\boldsymbol{\alpha}} \Psi_T \right\rangle \right), \tag{C6}$$

which is what we wanted to show.

Appendix D: Brute Force Metropolis-Hastings

In Figure 7 we see how the expectation value of the energy for the analytic and numeric expression of the local energy using Monte Carlo sampling matches the exact energy from Equation 35 for $\alpha \in [0.3, 0.7]$.

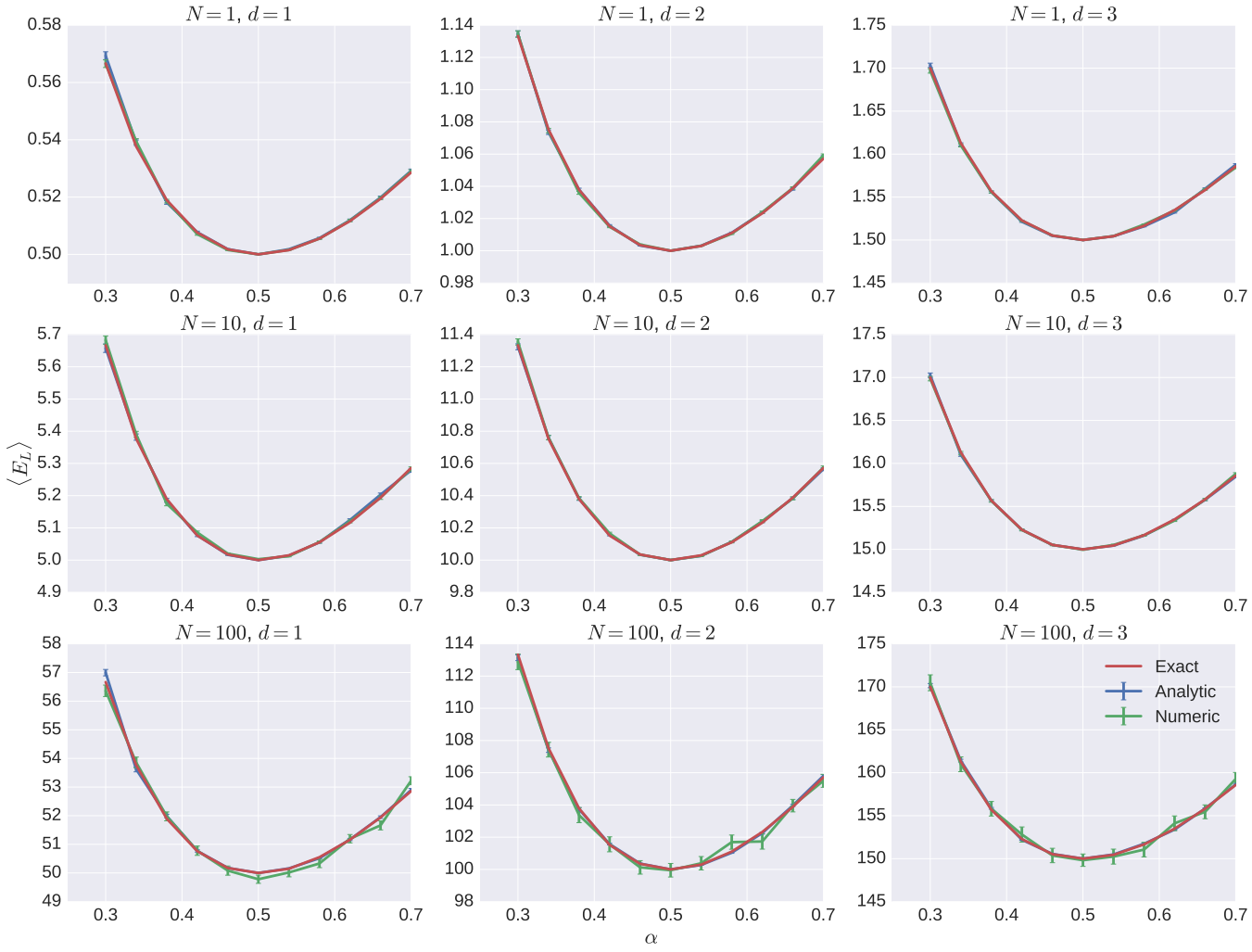

FIG. 7: Comparison of analytic and numeric results with the exact expression for the energy as a function of α. The plots show the standard deviation from the blocking method as error ticks. In the figure N is the number of particles and d the number of dimensions.