

STUDY OF THE IMPLEMENTATION OF AN ELEMENTARY VARIATIONAL MONTE-CARLO SCHEME AND DEVELOPMENTS OF THE BLOCKING METHOD.

Marius Jonsson

<http://github.com/kingoslo/postmann.pat>

June 2, 2017

ABSTRACT

This is a report submission for the second project of «Computational physics 2» at the Institute of Physics, University of Oslo, autumn 2016.

We implement a variational Monte-Carlo scheme using Hastings-Metropolis algorithm to compute observables for a quantum dot of $n = 2, 6, 12, 20$ interacting electrons at frequency $\omega = 1, 0.5, 0.1, 0.05, 0.01$. In particular, we found the trial energy of a Slater-Jastrow type trial state function for theory of a harmonic oscillator potential with Coulomb repulsion. We also found mean distance between electrons, their expected kinetic- and potential energy. We found ideal variational parameters for the trial state function. We computed that the proportion of the load which is parallelizable for our scheme is $\langle f \rangle = 0.9999988$ with variance $\text{Var}(f) = 6 \cdot 10^{-14}$. We found that expected maximal speed up was 800 000 for 10^6 Monte-Carlo cycles. We found that there was no effect of vectorization ($p = 0.7634$ df = 18). Furthermore, we develop the mathematics of the Blocking method. We prove that the error of the blocking method is reduced with every iteration and it eventually becomes constant once the observations become independent. We conjecture that it should be possible to strengthen the result further. We obtain a formula which determines the rate of convergence. We conjecture that it is possible to obtain a formula determining the ideal number of blocking iterations.

INTRODUCTION

The implementation of a Variational Monte Carlo scheme using Hastings-Metropolis algorithm is possibly the most robust way of estimating observables if a proposal of the state function ψ_T exist. It's robustness, as we shall see, is mainly due to that the properties and construction of stationary Markov chains are well understood. Their construction was first made popular by a paper Metropolis et al. 1953. In 1971, Canadian statistician Wilfred Keith Hastings provided modern rigorous proof of what became the Hastings-Metropolis theorem, contained in the appendix. His proof was constructive, in the sense that it shows that any irreducible Markov chain generates a time reversible Markov chain, $\{X_i, i \in \mathbb{N}\}$, with stationary probability distribution ψ_T . Using X_i we can compute any desired statistics, such as the distribution of the spectrum of the energy E_i . The application to quantum theory is a corollary of Chebyshev's inequality. It proves that $\langle E \rangle = \lim_{m \rightarrow \infty} (1/m) \sum_{i=1}^m E_i$ with convergence in probability (Varadhan 2001). That means the observations E_i immediately represent estimates for any observable! In this project we will investigate

the ground state of a quantum dot with $n \geq 2$ electrons. Assuming that $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ is the distance between electron i and j , our theory is

$$E_T = \sum_{i=1}^N \left[-\frac{1}{2} \frac{\nabla_i^2 \psi_T}{\psi_T} + \frac{1}{2} \omega^2 r_i^2 + \sum_{j=i+1}^N \frac{1}{r_{ij}} \right], \quad (1)$$

$$\psi_T = C \psi_0 \psi_C, \quad \psi_0(\mathbf{r}_1, \mathbf{r}_2) = \exp\left(-\frac{1}{2} \alpha \omega (r_1^2 + r_2^2)\right), \quad \psi_C(\mathbf{r}_1, \mathbf{r}_2) = \exp\left(\frac{r_{12}}{\beta r_{12} + 1}\right). \quad (2)$$

We will generalize ψ_T for any natural number of electrons. We assume the parameters α, β are given. In our case they are not known, but will be fixed such that $\langle E(\alpha, \beta) \rangle$ is a minimum with respect to α, β .

Although the variational Monte-Carlo scheme is invaluable to any physicist interested in computational physics, the highlights of this report are new contributions to the theory of the Blocking method. In the evaluation of the implementation of the variational Monte-Carlo scheme, we will use the Blocking method to compute the variance of the means of our statistics, for example $\text{Var}(\langle E \rangle)$. In the results we develop the mathematics to contain 3 immediate lemmas of the work by Flyvbjerg and Petersen, and two new propositions. We aimed to provide proof of the Blocking method. At the time of writing its not clear whether such a proof exist already. Unfortunately I was pressed on time, and finalizing the proof of was a bit more difficult than first anticipated. However, I hope to jolt readers of this report with the strength of the new advertised propositions and their corollaries. Hopefully the proof of the Blocking method come over the summer if there is interest in developing the Blocking method.

The report is structured by «introduction»-, «methods»-, «results and discussion»- and finally a «conclusion and perspectives»-sections.

METHODS

Let's first discuss the Hastings-Metropolis theorem. Its constructive proof is contained in the appendix. Often whilst proving results about Markov chains we are interested in whether the Markov chain can reach every state from any other state. For the Hastings-Metropolis method, this is important because it ensures that all the divisions which are made in the proof are non-zero. Let's make this notion precise. Suppose q_{ij} is the transition probability matrix of a Markov chain X_n and S is the state space. We say that X_n is **irreducible** if for all $i, j \in S$ there exist an $n \in \mathbb{N}$ such that $(q^n)_{ij}, (q^n)_{ji} > 0$.

Suppose that X_n is irreducible and not deterministically periodic, then we say that the probabilities $\pi_i = P(X_n = i)$ is the **stationary distribution** of X_n .

We will say that X_n is **time reversible** if the conditional probability $P(X_n = j, \text{ given that } X_{n+1} = i) = P_{ij}$ for all $i, j \in S$ and $n \in \mathbb{N}$. Interestingly, it is easily shown that a Markov chain is time reversible if and only if $\pi_i P_{ij} = \pi_j P_{ji}$ for all $i, j \in S$. If you wish to prove this yourself, the proposition follows from Markov chain property and Bayes' theorem. With these definitions, we can understand the Hastings-Metropolis theorem.

THEOREM 1 (Hastings-Metropolis theorem). *Suppose that $C\pi_i$ is a discrete probability distribution. If q_{ij} is any irreducible transition probability matrix, and X_n is a Markov chain with transition probability matrix*

$$P_{ij} = \begin{cases} \alpha_{ij} q_{ij}, & j \neq i \\ q_{ii} + \sum_{k=0}^{\infty} q_{ik} (1 - \alpha_{ik}) & j = i \end{cases}, \quad \text{where} \quad \alpha_{ij} = \min\left(\frac{\pi_j q_{ji}}{\pi_i q_{ij}}, 1\right), \quad (3)$$

then X_n is time reversible with stationary distribution π_i .

It is immediate that the entire variational Monte-Carlo scheme is specified if we define π_i and q_{ij} . Suppose $\mathbf{r}_k^{(i)}$ denote the coordinates of electron k at iteration number i . We will take π_i to equal $\psi_T(\mathbf{r}_1^{(i)}, \mathbf{r}_2^{(i)}, \dots, \mathbf{r}_n^{(i)})$ for all $i \in \mathbb{N}$. We will refer to the transition probability matrix q_{ij} as the *importance sampling*.

As we noted, theorem 3 implies that the variational Monte-Carlo scheme is uniquely defined once we specify ψ_T and the importance sampling. We will take the trial state function to be a Slater-Jastrow product. This is desirable since it exhibits the correct cusp behaviour for electrons with parallel and anti-parallel spin and it means that the variational parameters has simple physical interpretation. We will see that this produces ground state energy error smaller than 0.02%. Specifically, we define

$$\psi_T = \det(D_\uparrow) \det(D_\downarrow) \psi_C, \quad \psi_C(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n; \beta) = \exp \left(\sum_{i=1}^n \sum_{j=i+1}^n \frac{\gamma_{ij} r_{ij}}{1 + \beta r_{ij}} \right) \quad (4)$$

$$\text{such that} \quad (D_\uparrow)_{ij} = \phi_{2j}(\mathbf{r}_{2i}; \alpha), \quad (D_\downarrow)_{ij} = \phi_{2j+1}(\mathbf{r}_{2i+1}; \alpha) \quad i, j \in \{0, 1, \dots, \frac{n}{2} - 1\}$$

$$\text{and} \quad \phi_j(\mathbf{r}_i; \alpha) = H_{n_{xj}}((\alpha\omega)^{1/2} x_i) H_{n_{yj}}((\alpha\omega)^{1/2} y_i) \exp \left(-\frac{1}{2} \alpha \omega r_i^2 \right) \quad (5)$$

j	0	1	2	3	4	5	6	7	8	9	10	11	...
n_{xj}	0	0	1	1	0	0	2	2	1	1	0	0	...
n_{yj}	0	0	0	0	1	1	0	0	1	1	2	2	...
Spin	\uparrow	\downarrow	\uparrow	\downarrow	\uparrow	\downarrow	\uparrow	\downarrow	\uparrow	\downarrow	\uparrow	\downarrow	...

for all $i \in \{0, 1, 2, \dots, n-1\}$.

It remains to define the importance sampling. The generating Markov chain with probability density q_{ij} will be taken to be a stationary isotropic diffusion process. We have seen in lectures that it is immediate from the Fokker-Planck equation that if $r_k^{(i)}, r_k^{(j)}$ represent two different position of particle number k and $\Delta t > 0$ is a parameter, then such a process must satisfy

$$q_{ij} = G(\mathbf{r}_k^{(i)}, \mathbf{r}_k^{(j)}; \Delta t) = \frac{1}{(2\pi\Delta t)^{1/2}} \exp \left(-\frac{(\mathbf{r}_k^{(j)} - \mathbf{r}_k^{(i)} - \Delta t(\nabla \psi_T / \psi_T))^2}{2\Delta t} \right), \quad \text{for all } k \in \{1, 2, \dots, n\}$$

Which is nothing but the multinomial normal probability distribution with mean $\Delta t \nabla \psi_T / \psi_T$ and variance Δt (Devore and Berk 2012). That means if $\xi \sim N(0, 1)$ is some standard normal random variable we can write

$$\mathbf{r}_k^{(j)} = \mathbf{r}_k^{(i)} + \frac{1}{2} \frac{\nabla \psi_T}{\psi_T} + \xi(\Delta t)^{1/2}. \quad (\text{DeGroot and Schervish 2012, p. 338})$$

These expressions suffice to fix the implementation of the importance sampling, as we shall see in a practical example at the end of this section. As announced we want to estimate the value of the variational parameters α and β . The simplest implementation is possibly the steepest decent method.

Suppose $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a function, then Press et al. 2007, p. 515 characterises the steepest decent method as a *not-very-good method* that works by «Start at a point $\mathbf{r}^{(0)}$. As many times as needed, move from $\mathbf{r}^{(i)}$ to $\mathbf{r}^{(i+1)}$ minimizing along the line from $\mathbf{r}^{(i)}$ in the direction of the local downhill gradient $-\nabla f(\mathbf{r}^{(i)})$.». This means that given $\mathbf{r}^{(0)}$ we can iterate by

$$\mathbf{r}^{(i+1)} = \mathbf{r}^{(i)} - \zeta_i \nabla f(\mathbf{r}^{(i)}) \quad \text{for all } i \in \mathbb{N} \quad (\text{Hazewinkel 1989}) \quad (6)$$

Some choices of sequences $\{\zeta_i\}_{i=1}^\infty$ does not ensure convergence. Therefore the choice of ζ_i is interesting. Ideally we choose ζ_i such that $f(\mathbf{r}^{(i)} - \zeta_i \nabla f(\mathbf{r}^{(i)}))$ is maximised for all $i \in \mathbb{N}$ (Hazewinkel 1989). This is most conveniently done by differentiating equations (19) and (20) with respect to ζ_i . By the defining

equation (6) we therefore need to calculate $\nabla \langle E \rangle$. By using the definition of expected value and pulling the derivative into the integral, we discover that for any variational parameter θ ,

$$\frac{\partial \langle E \rangle}{\partial \theta} = 2 \left[\left\langle \frac{E}{\psi_T} \frac{\partial \psi_T}{\partial \theta} \right\rangle - \langle E \rangle \left\langle \frac{1}{\psi_T} \frac{\partial \psi_T}{\partial \theta} \right\rangle \right]$$

Using this equation we express $\nabla \langle E \rangle$ by the derivatives of the wave function. And estimate all the relevant expectation values by the mean. But we know how to differentiate, so this follows easily after one step of Jacobi's formula. Please see appendix for details of the derivation and the resulting equations. Using these estimates we will eventually be able to estimate the expected observables using the mean as the estimator. We will however be interested in estimating the variance of the mean to quantify the errors and produce an indication of the appropriateness of the state function. For these purposes, the blocking method is perfect.

Suppose \bar{X}_m is the mean of m random variables and we want to compute $\text{Var}(\bar{X}_m)$. The blocking method is a way of quickly computing $\text{Var}(\bar{X}_m)$, and is implemented as follows: Start by taking the average of every two sequential random variables from a time-series. That means that we obtain $m/2$ new random variables. If we keep on repeating this process we will obtain $m/2^k$ random variables after k repetitions. We will call each iteration the process a blocking transformation. It turns out that under certain conditions, there is some k , such that after k blocking transformations, the variance of the blocked variables are essentially the same as $\text{Var}(\bar{X}_m)$. Let's make this precise.

Suppose that X_1, \dots, X_m are $m = 2^k$ identically distributed random variables. If we let $X_i^{(0)} = X_i$ for all $i \in \mathbb{N} = \{1, 2, \dots\}$ and

$$X_i^{(k+1)} = \frac{1}{2} (X_{2i-1}^{(k)} + X_{2i}^{(k)}), \quad \text{for all } k \in \mathbb{N}.$$

If $k \in \mathbb{N}$ and $m^{(k)} = 2^{-k}m$, we will say that the set $\{X_i^{(k)} \mid i \in \{1, 2, \dots, m^{(k)}\}\}$ are **subject to k blocking transformations**. We will later show that $n^{(k)} = 2^{-k}m$ is the appropriate number of random variables subject to k blocking transformations. Also, we will define the autocovariance $\gamma_{i,j} = \text{Cov}(X_i, X_j)$. In the case that the Markov Chain or time series is stationary, the autocovariance is stationary. This means that $\gamma_{i,j} = \gamma_{i+t,j+t}$ for all $i, j, t \in \mathbb{N}$ (Shumway and Stoffer 2011, p. 24). As a corollary of theorem 3, this is the case for Markov Chains of the type we are considering, in which case it is convenient to define $\gamma_t = \gamma_{i+t,j+t}$ for all $i, j, t \in \mathbb{N}$. Moreover we will let $\gamma_t^{(k)}$ denote the autocovariance subject to k blocking transformations. Finally, we have everything we require to talk about the implementation of the variational Monte-Carlo scheme at program level.

In particular, let's talk about how to implement the variational Monte-Carlo scheme in C++. For brevity, let's show explicitly how this is implemented for two particles, then talk about additional complexity induced by multiple particles. We have an analytical expression for ψ_T from equation (2) and expression of all the observables we want to sample. For the most elementary implementation we will only sample the energy. Details of its derivation are found in the appendix. However, the energy we arrive at using equation (1) for two particles is

$$E_T = \frac{1}{r_{12}} + \frac{1}{2}\omega^2(1 - \alpha^2)(r_1^2 + r_2^2) + 2\alpha\omega + \frac{\alpha\omega r_{12}}{(1 + \beta r_{12})^2} - \frac{(1 + r_{12} - \beta^2 r_{12}^2)}{r_{12}(1 + r_{12}\beta)^4}.$$

If we let m , denote the number of Monte-Carlo cycles, r and rpp denote the new and old position matrices for the two particles then a simple program using importance sampling for two particles is quickly built using the Armadillo library by

Interestingly, for $\omega = 1$ the program returns that the energy is $\langle E \rangle = 3.00052 \pm 10^{-5}$ after only $2^{20} \approx 10^6$ Monte-Carlo cycles. For reference, the exact result is 3.

```

for (int i = 0; i < m; i++) {
    // selecting particle to move from uniform distribution
    int k = rand_particle(gen);
    int not_k = (k+1) % 2;

    // computing qij and qji (the importance sampling)
    rijpp = norm(rpp.col(0)-rpp.col(1));
    vec Fpp = -2*a*w*rpp.col(k) + 2*c*(rpp.col(k) - rpp.col(not_k))/( (1 + b*rijpp)*(1 +
        b*rijpp)*rijpp );
    r.col(k) = rpp.col(k) + D*Fpp*dt + randn<vec>(2)*sqrt(dt);
    rij = norm(r.col(0)-r.col(1));
    F = -2*a*w*r.col(k) + 2*c*(r.col(k) - r.col(not_k))/( (1 + b*rij)*(1 + b*rij)*rij );
    vec p = rpp.col(k) - r.col(k) - D*dt*F;
    vec q = r.col(k) - rpp.col(k) - D*dt*Fpp;
    double qji = exp(- dot(p,p)/(4*D*dt));
    double qij = exp(- dot(q,q)/(4*D*dt));

    // computing new state function
    wf = exp(-0.5*a*w*( dot(r.col(0),r.col(0)) + dot(r.col(1),r.col(1)) ) + r12/(1 +
        b*r12) );

    // one iteration of Hastings-Metropolis theorem
    if ( wf*wf*qji/(wfpp*wfpp*qij) > rand_double(gen) ) {
        rpp = r; wfpp = wf; rij = norm(r.col(0) - r.col(1) );
    }

    // sample energy
    double e = 1/rij + 0.5*w*w*(1-a*a)*( dot(rpp.col(0),rpp.col(0)) +
        dot(rpp.col(1),rpp.col(1)) )
        + 2*a*w + a*w*c*rij/pow(1 + rij*b,2) - c*(1+rij*c-b*b*rij*rij)/( rij*pow(1 +
            rij*b,4) );
    E += e/iterations;
}

```

Table 1: Program for Hastings-Metropolis-style implementation of a quantum dot of two electrons. The variables `r`, `rpp` contain the positions at sequential steps, `F`, `Fpp` contain the quantity $\nabla\psi_T/\psi_T$ for sequential steps and `rij`, `rijpp` contain the distance between electrons at sequential steps. The program returns the energy with uncertainty, $\text{Var}(\langle E \rangle) = 10^{-10}$ and rejection ratio $r < 10^{-4}$ for the importance sampling.

We want to generalize the program to more than two, $n > 2$, electrons. This adds the complication of adding a Slater determinant. Further consequences are that the analytical expression of the energy becomes large. This is chiefly due to the dependence of α in the functions ϕ_{ij} in the Slater determinant and their complicated interactions. Therefore $\nabla^2 \psi_T / \psi_T$ becomes a complicated expression. However, note that for the trail state function the chain rule implies

$$\frac{\nabla_i^2 \psi_T}{\psi_T} = \frac{\nabla_i^2 \det D_\uparrow}{\det D_\uparrow} + \frac{\nabla_i^2 \det D_\downarrow}{\det D_\downarrow} + \frac{\nabla_i^2 \psi_C}{\psi_C} + 2 \left(\frac{\nabla_i \det D_\uparrow}{\det D_\uparrow} + \frac{\nabla_i \det D_\downarrow}{\det D_\downarrow} \right) \cdot \frac{\nabla_i \psi_C}{\psi_C}$$

We wrote functions which evaluated each of these terms after first differentiating each term using pen and paper. The exact details of these differentiations are of little interest to post-calculus-students, and relegate these to the appendix for your reference.

However more interesting is perhaps discussing how to evaluate the Hastings-Metropolis tests for larger systems of electrons. It turns out that this adds some complications that are easily dealt with. Suppose now that $X \sim \text{unif}(0, 1)$ is a uniformly continuously distributed random variable on $(0, 1]$. As we saw from theorem 3 we have to evaluate the following inequality quickly and reliably at each Monte-Carlo cycle:

$$X < \frac{\pi_j q_{ji}}{\pi_i q_{ij}} = \frac{q_{ji}}{q_{ij}} \left(\frac{\det(D_\uparrow D_\downarrow)_j (\psi_C)_j}{\det(D_\uparrow D_\downarrow)_i (\psi_C)_i} \right)^2 \quad (7)$$

The trouble is that due to the exponentials inside the Jastrow factor, these terms will become unpractically large and overflow for large systems of electrons. The path taken in the implementation of this project was the following. Suppose $f(x)$ is some monotonous $\mathbb{R} \rightarrow \mathbb{R}$ -function. Since f is monotonous, that means by definition that $x < y$ if and only if $f(x) < f(y)$. Since the natural logarithm \log is monotonous we can take $x = X$, and y equal the right hand side of inequality (7). Hence inequality (7) is true if and only if

$$\log(X) < \log q_{ji} - \log q_{ij} + 2(\log(\det(D_\uparrow D_\downarrow)_j) - \log(\det(D_\uparrow D_\downarrow)_i) + \log(\psi_C)_j - \log(\psi_C)_i)$$

Since $X > 0$ the logarithm and exponential are reverse functions for the terms in question, and hence the evaluation of the Jastrow factors are reduced to magnitudes smaller than about 1000. Expressions to implement the Steepest decent method are also contained in the appendix.

Care was taken to check that the program ran correctly.

- > For two particles, there were exact results for the ground state energy. The program reproduced this benchmark, as becomes apparent in the results. This was verified in the case that the program was running with and without the Slater determinant. In addition, for 6 and 12 particles the expected energy was compared to M. Pedersen Lohne et al. 2011. As will become evident in the discussion, there was satisfactory agreement.
- > For two particles, there were analytical expressions available for the trail energy, the Laplacians of both the harmonic oscillator basis states and the Jastrow factor. There were also analytical expressions for the gradients of the same functions. When a Slater determinant was introduced, the program was individually checked against all four of these, and the program passed millions of iterations to check all of these without a glitch.
- > For two particles there were analytical expressions for the derivatives required for the steepest decent method. After a Slater determinant was introduced, the program was checked against both of these for millions of iterations, there were no discrepancies.
- > It was verified that when the perturbations were switched off, we recovered the unperturbed energies for $n = 2, 6, 12$. The simulated values were 2.0002ω , 10.0031ω and 28.0002ω respectively.

n	ω	$\langle E \rangle$	$\langle K \rangle$	$\langle V \rangle$	$\langle r_{ij} \rangle$
2	1.00	$3.0057 \pm 8 \cdot 10^{-4}$	$0.9217 \pm 1 \cdot 10^{-3}$	$2.0840 \pm 2 \cdot 10^{-2}$	$1.6318 \pm 3 \cdot 10^{-2}$
	0.50	$1.6595 \pm 1 \cdot 10^{-3}$	$0.4521 \pm 6 \cdot 10^{-2}$	$1.2073 \pm 1 \cdot 10^{-2}$	$2.4424 \pm 1 \cdot 10^{-1}$
	0.10	$0.4588 \pm 6 \cdot 10^{-3}$	$0.0984 \pm 1 \cdot 10^{-3}$	$0.3603 \pm 6 \cdot 10^{-3}$	$5.4414 \pm 1 \cdot 10^{-1}$
	0.05	$0.2621 \pm 6 \cdot 10^{-4}$	$0.0472 \pm 4 \cdot 10^{-4}$	$0.2148 \pm 8 \cdot 10^{-4}$	$8.6576 \pm 6 \cdot 10^{-2}$
	0.01	$0.0874 \pm 3 \cdot 10^{-3}$	$0.0103 \pm 4 \cdot 10^{-3}$	$0.0770 \pm 4 \cdot 10^{-3}$	$18.740 \pm 1 \cdot 10^{-0}$
6	1.00	$20.231 \pm 9 \cdot 10^{-3}$	$3.9071 \pm 7 \cdot 10^{-2}$	$16.320 \pm 6 \cdot 10^{-2}$	$2.1835 \pm 3 \cdot 10^{-2}$
	0.50	$11.808 \pm 1 \cdot 10^{-2}$	$1.7694 \pm 4 \cdot 10^{-2}$	$10.038 \pm 1 \cdot 10^{-1}$	$3.3861 \pm 1 \cdot 10^{-1}$
	0.10	$3.5704 \pm 2 \cdot 10^{-3}$	$0.3373 \pm 3 \cdot 10^{-3}$	$3.2330 \pm 3 \cdot 10^{-3}$	$9.2390 \pm 1 \cdot 10^{-1}$
	0.05	$2.1679 \pm 2 \cdot 10^{-3}$	$0.1716 \pm 7 \cdot 10^{-3}$	$2.0001 \pm 5 \cdot 10^{-3}$	$14.058 \pm 4 \cdot 10^{-2}$
	0.01	$0.7177 \pm 6 \cdot 10^{-3}$	$0.0049 \pm 1 \cdot 10^{-2}$	$0.7127 \pm 1 \cdot 10^{-2}$	$32.747 \pm 8 \cdot 10^{-1}$
12	1.00	$65.881 \pm 2 \cdot 10^{-2}$	$9.4946 \pm 7 \cdot 10^{-2}$	$56.387 \pm 2 \cdot 10^{-2}$	$2.6842 \pm 7 \cdot 10^{-3}$
	0.50	$39.250 \pm 3 \cdot 10^{-2}$	$4.3144 \pm 9 \cdot 10^{-2}$	$34.936 \pm 8 \cdot 10^{-2}$	$4.1283 \pm 1 \cdot 10^{-1}$
	0.10	$12.270 \pm 2 \cdot 10^{-3}$	$0.8171 \pm 1 \cdot 10^{-2}$	$11.503 \pm 1 \cdot 10^{-2}$	$11.369 \pm 3 \cdot 10^{-2}$
	0.05	$7.6022 \pm 9 \cdot 10^{-3}$	$0.4222 \pm 1 \cdot 10^{-2}$	$7.1799 \pm 1 \cdot 10^{-2}$	$17.734 \pm 4 \cdot 10^{-2}$
	0.01	$2.5671 \pm 3 \cdot 10^{-3}$	$0.0913 \pm 3 \cdot 10^{-3}$	$2.4757 \pm 4 \cdot 10^{-3}$	$44.243 \pm 2 \cdot 10^{-1}$
20	1.00	$156.25 \pm 4 \cdot 10^{-2}$	$18.124 \pm 3 \cdot 10^{-1}$	$138.13 \pm 3 \cdot 10^{-2}$	$3.1754 \pm 4 \cdot 10^{-2}$
	0.50	$93.918 \pm 6 \cdot 10^{-2}$	$8.1185 \pm 2 \cdot 10^{-1}$	$85.863 \pm 1 \cdot 10^{-1}$	$4.8527 \pm 1 \cdot 10^{-1}$
	0.10	$30.237 \pm 1 \cdot 10^{-2}$	$1.6442 \pm 1 \cdot 10^{-2}$	$28.592 \pm 8 \cdot 10^{-2}$	$13.807 \pm 1 \cdot 10^{-2}$
	0.05	$18.763 \pm 3 \cdot 10^{-2}$	$0.8454 \pm 5 \cdot 10^{-2}$	$17.917 \pm 2 \cdot 10^{-2}$	$20.942 \pm 5 \cdot 10^{-2}$
	0.01	$6.4086 \pm 5 \cdot 10^{-2}$	$0.1699 \pm 6 \cdot 10^{-2}$	$6.2386 \pm 2 \cdot 10^{-3}$	$52.237 \pm 3 \cdot 10^{-1}$

Table 2: Estimates of expected value for the state function specified by equation (4). The error bars given represent one standard deviation from the mean. $\langle E \rangle$, $\langle K \rangle$ and $\langle V \rangle$ represent the expectation of ground state-, kinetic- and potential energy respectively, whilst $\langle r_{ij} \rangle$ represent the mean distance between any two of the n electrons at the frequency ω .

- > The program passed over 100 tests of tabulated values. A reference table of expected values were provided by the course administration on Piazza, for any number of particles, and chosen frequencies ω . The program was checked against this table, and as will become evident in the results, the program produced reasonable results for all of these tests. Since there were 65 observables, this represented 65 tests of the program. There was also another table of values containing suggestions for variational parameters, this table contained 40 values. The program reproduced all these values satisfactorily.

For your reference, the source code is available from the github URL provided at the heading of this report.

RESULTS AND DISCUSSION

Our results encompassed both physics and mathematics results. We start by presenting the physics results. We obtained estimates for the expected kinetic-, potential-, and total energy for $n = 2, 6, 12, 20$ electrons, as well as the expected interparticle distance for $\omega = 1, 0.5, 0.1, 0.05, 0.01$ with confidence intervals for all estimates using the blocking method. Please see table 2. Using the method of steepest decent we were able to find parameters α, β such that $\psi(\mathbf{r}_1, \dots, \mathbf{r}_n; \alpha, \beta)$ minimize the local energy. We present these along with suitable choice of values of Δt since we found that the choice of Δt was just as important as the variational parameters. We also see that the acceptance for all simulations were larger than 99.9999%. Please see table 3 for full details. Next we completed a performance analysis in the case $n = 6$ electrons in order to

n	ω	α	β	Δt	Acceptance
2	1.00	0.988664	0.397451	0.0010	> 99.999%
	0.50	0.942741	0.358363	0.0025	> 99.999%
	0.10	0.952742	0.354292	0.0050	> 99.999%
	0.05	0.913124	0.236505	0.0750	> 99.999%
	0.01	0.911693	0.203917	0.0100	> 99.999%
6	1.00	1.040159	0.469328	0.0010	> 99.999%
	0.50	0.929001	0.395588	0.0025	> 99.999%
	0.10	0.831105	0.211448	0.0500	> 99.999%
	0.05	0.851754	0.155691	0.0750	> 99.999%
	0.01	0.839981	0.106890	0.0100	> 99.999%
12	1.00	1.059616	0.475495	0.0050	> 99.999%
	0.50	0.945597	0.414537	0.0025	> 99.999%
	0.10	0.841061	0.217988	0.0500	> 99.999%
	0.05	0.842746	0.153921	0.0750	> 99.999%
	0.01	0.835881	0.092328	0.5000	> 99.999%
20	1.00	1.079881	0.475432	0.0010	> 99.999%
	0.50	0.944121	0.424048	0.0025	> 99.999%
	0.10	0.857020	0.200731	0.1000	> 99.999%
	0.05	0.834301	0.153331	0.0750	> 99.999%
	0.01	0.835321	0.092253	0.5000	> 99.999%

Table 3: Suitable choice of paramaters. Here α and β refer to the variational parameters of the Slater-Jastrow-type state function chosen whilst Δt give the variance of the positions $r_k^{(i)}$ for all $i, k \in N$. Acceptance refers to the ratio of the accepted values of $r_k^{(i)}$ to the total number of Monte-Carlo cycles. Typically the first non-nine decimal place was the 5th or 6th digit.

study the effect of vectorization and parallelization on the scheme. Since the majority of the work is the evaluation of functions and updating the variables in memory, almost all of the work could be parallelized. In fact, suppose you regard f as the random variable which represent the fraction of time that was perfectly parallelizable relative to the whole, then we found that

$$\langle f \rangle = 0.9999988 \quad \text{and} \quad \text{Var}(f) = 6.05 \cdot 10^{-14}.$$

We know by Jensen inequality that if X is a random variable and $f : \mathbb{R} \rightarrow \mathbb{R}$ is a convex function (\dagger), then $f(\langle X \rangle) \leq \langle f(X) \rangle$ with equality if and only if X has zero variance. It turns out that $x \mapsto x^{-1}$ is a convex function, as you can check, and therefore

$$\langle \text{speedup} \rangle = \left\langle \frac{1}{(1-f) - f/p} \right\rangle \stackrel{(\dagger)}{>} \frac{1}{\langle (1-f) - f/p \rangle} = \frac{1}{(1-\langle f \rangle) - \langle f \rangle/p} \quad (8)$$

Assuming that f is a continuous random variable with continuous probability density function, then limits commute with expectation by the Dominated convergence theorem (DCT), i.e

$$\begin{aligned} \langle \text{max speedup} \rangle &= \left\langle \lim_{p \rightarrow \infty} \frac{1}{(1-f) - f/p} \right\rangle \stackrel{(\text{DCT})}{=} \lim_{p \rightarrow \infty} \left\langle \frac{1}{(1-f) - f/p} \right\rangle \stackrel{(8)}{>} \lim_{p \rightarrow \infty} \frac{1}{\langle (1-f) - f/p \rangle} \\ &= \lim_{p \rightarrow \infty} \frac{1}{(1-\langle f \rangle) - \langle f \rangle/p} = \frac{1}{1-\langle f \rangle} = 833333. \end{aligned}$$

Unfortunately, computing the variance of the maximal speed up is more complicated.

The code was vectorized to investigate the implied performance increase. Surprisingly, we found no significant mean increase in runtimes by vectorization at 10^6 Monte-Carlo cycles. Suppose the vectorized runtimes are X_1, \dots, X_n with mean μ_x and the serial runtimes are Y_1, \dots, Y_n with mean μ_y . If all the observation are independent, normally distributed with some variance σ^2 , then we are in the perfect setting to use the Analysis of Variance method to check if $\mu_x = \mu_y$ (Devore and Berk 2012). The result was that Analysis of Variance found no significant difference between μ_x and μ_y ($p = 0.7634$). In fact $\sigma = 24.18$ and if $\mu_x = \mu_y = \mu$, then $\mu = 75.1218$ seconds on 18 degrees of freedom. Please see table 4 for full details

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Vectorization	1	54.6	54.64	0.0934	0.7634
Residuals	18	10528.1	584.9	0	

Table 4: By Analysis of variance, we found no significant difference between the runtimes of the vectorized and serial code, given that the 20 runtimes were independent, normally distributed with identical variance on 18 degrees of freedom.

The remaining results develop the mathematics of the blocking method. Let's start with a soft result on which we will build more structure.

PROPOSITION 1. Suppose X_1, \dots, X_m are $m = 2^d$ identically distributed random variables with stationary non-negative covariances. Let $\bar{X}_m = (1/m) \sum_{j=1}^m X_j$ denote their average and $X_i^{(j)}$ be a random variable subject to j blocking transformations. Then

$$\varepsilon_k = \text{Var}(\bar{X}_m) - \text{Var}(X_i^{(k)}) = \frac{2}{n^{(k)}} \sum_{t=1}^{n^{(k)}-1} \left(1 - \frac{t}{n^{(k)}}\right) \gamma_t^{(k)} \quad (9)$$

and

$$\gamma_j^{(k)} = \begin{cases} \frac{1}{2}\gamma_{2j}^{(k-1)} + \frac{1}{2}\gamma_{2j+1}^{(k-1)} & \text{if } j = 0 \\ \frac{1}{4}\gamma_{2j-1}^{(k-1)} + \frac{1}{2}\gamma_{2j}^{(k-1)} + \frac{1}{4}\gamma_{2j+1}^{(k-1)} & \text{if } j > 0 \end{cases} \quad (10)$$

Proof. See Flyvbjerg and Petersen 1989. ■

Next let's make a couple of simple observations

LEMMA 1. Suppose $m = 2^d$, $d > 1$ denotes the number of observations and k denote the number of blocking transformations. Then

$$n^{(k)} = 2^{-k}m \quad \text{and} \quad k < \log_2 m \quad \text{for all} \quad k < d.$$

Proof. We proof the first part by induction. Suppose $k = 0$. By definition, $n^{(0)} = m$ and so

$$n^{(0)} = m = 2^{-0}m,$$

Suppose now that there exist some $k \in \{0, 1, 2, \dots\}$ such that $n^{(k)} = 2^{-k}m$ and $n^{(k+1)}$ exist, this is possible since $d > 1$ by hypothesis. By construction each blocking transformation halve the number of observations, so $n^{(k+1)} = n^{(k)}/2$ and therefore

$$n^{(k+1)} = \frac{n^{(k)}}{2} = \frac{2^{-k}m}{2} = 2^{-(k+1)}m.$$

To see the last observation is true, just observe that if $k < d$ and $m = 2^d$, then $k < d = d \log_2 2 = \log_2 m$. ■

LEMMA 2. Suppose $\gamma_t^{(k)}$ denote the correlation between two observations X_i and X_j such that $|i - j| = t$. Suppose also that $\gamma_t^{(k)} \geq 0$ for all t . Then

$$\gamma_t^{(k)} \leq \gamma_0^{(k)} \quad \text{for all} \quad t \geq 0.$$

Proof. We will need the formula $|\text{Cov}(X, Y)|^2 \leq \text{Cov}(X, X)\text{Cov}(Y, Y)$ (*) which is proven in the appendix. Write

$$\begin{aligned} (\gamma_t^{(k)})^2 &= |\gamma_t^{(k)}|^2 = |\gamma_{ij}^{(k)}|^2 = |\text{Cov}(X_i^{(k)}, X_j^{(k)})|^2 \\ &\stackrel{(*)}{\leq} \text{Cov}(X_i^{(k)}, X_i^{(k)})\text{Cov}(X_j^{(k)}, X_j^{(k)}) = \gamma_{ii}^{(k)}\gamma_{jj}^{(k)} = (\gamma_0^{(k)})^2 \end{aligned}$$

Since the function $f(\gamma) = \gamma^2$ is increasing on $[0, \infty)$, we know $\gamma_t^{(k)} \geq \gamma_0^{(k)}$ follows. ■

I doubt that the next proposition alone is enough to imply the blocking method, but its strength become evident once we consider its corollaries.

PROPOSITION 2. Suppose ε_k denotes the error after k blocking iterations and $\gamma_t^{(k)} \geq 0$ for all $k, t \in \mathbb{N}$. Let $\{\varepsilon_k\}_{k=1}^d$ denote the sequence of errors. Then the rate of change of the sequence is

$$\varepsilon_k - \varepsilon_{k+1} = \frac{\gamma_1^{(k)}}{n^{(k)}} \quad (11)$$

Proof. It remains to prove formula (11). We will manipulate sums and will be interested in which function $f(t)$ appear in the terms $\gamma_{f(t)}^{(k)}$ inside the summation. Define a functional $S_{f(j)} : \{2x-1, 2x, 2x+1\} \rightarrow \mathbb{R}$ by:

$$S_{f(j)} \equiv \frac{1}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2-1} \left(1 - \frac{2t}{n^{(k)}}\right) \gamma_{f(t)}^{(k)}.$$

That means, using $n^{(k+1)} = n^{(k)}/2$, we can rewrite ε_{k+1} as

$$\begin{aligned} \varepsilon_{k+1} &\stackrel{(9)}{=} \frac{2}{n^{(k+1)}} \sum_{t=1}^{n^{(k+1)}-1} \left(1 - \frac{t}{n^{(k+1)}}\right) (\gamma_{k+1})_t \stackrel{(10)}{=} \frac{1}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2-1} \left(1 - \frac{2t}{n^{(k)}}\right) (\gamma_{2t-1}^{(k)} + 2\gamma_{2t}^{(k)} + \gamma_{2t+1}^{(k)}) \\ &= S_{2j-1} + 2S_{2j} + S_{2j+1} = 2(S_{2j-1} + S_{2j}) + S_{2j+1} - S_{2j-1} \end{aligned} \quad (12)$$

where we added and subtracted S_{2j-1} in the last step. Write out the sums explicitly to obtain:

$$S_{2j-1} = \frac{1}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2-1} \left(1 - \frac{2t}{n^{(k)}}\right) \gamma_{2t-1}^{(k)} = \frac{1}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2-1} \left(1 - \frac{2t-1}{n^{(k)}}\right) \gamma_{2t-1}^{(k)} - \frac{1}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2-1} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} \quad (13)$$

$$\begin{aligned} S_{2j+1} &= \frac{1}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2-1} \left(1 - \frac{2t}{n^{(k)}}\right) \gamma_{2t+1}^{(k)} = \frac{1}{n^{(k)}} \left[\frac{2}{n^{(k)}} \gamma_3^{(k)} + \frac{4}{n^{(k)}} \gamma_5^{(k)} + \dots + \frac{n^{(k)}-2}{n^{(k)}} \gamma_{n^{(k)}-1}^{(k)} \right] \\ &= \frac{1}{n^{(k)}} \sum_{t=2}^{n^{(k)}/2} \left(1 - \frac{2(t-1)}{n^{(k)}}\right) \gamma_{2t-1}^{(k)} = \frac{1}{n^{(k)}} \sum_{t=2}^{n^{(k)}/2} \left(1 - \frac{2t-1}{n^{(k)}}\right) \gamma_{2t-1}^{(k)} + \frac{1}{n^{(k)}} \sum_{t=2}^{n^{(k)}/2} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} \end{aligned}$$

Notice that if these terms are subtracted we obtain

$$S_{2j+1} - S_{2j-1} = \frac{2}{n^{(k)}} \sum_{t=2}^{n^{(k)}/2} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} - \frac{1}{n^{(k)}} \left(1 - \frac{1}{n^{(k)}}\right) \gamma_1^{(k)} \quad (14)$$

Lets also investigate the following quantity

$$\begin{aligned} n^{(k)}(S_{2j-1} + S_{2j}) &\stackrel{(13)}{=} \sum_{t=1}^{n^{(k)}/2-1} \left(1 - \frac{2t-1}{n^{(k)}}\right) \gamma_{2t-1}^{(k)} - \sum_{t=1}^{n^{(k)}/2-1} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} + \sum_{t=1}^{n^{(k)}/2-1} \left(1 - \frac{2t}{n^{(k)}}\right) \gamma_{2t}^{(k)} \\ &= \left[\left(1 + \frac{1}{n^{(k)}}\right) \gamma_1^{(k)} + \left(1 + \frac{2}{n^{(k)}}\right) \gamma_2^{(k)} + \dots + \left(1 + \frac{n^{(k)}-2}{n^{(k)}}\right) \gamma_{n^{(k)}-2}^{(k)} \right] - \sum_{t=1}^{n^{(k)}/2-1} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} \\ &= \sum_{t=1}^{n^{(k)}-2} \left(1 - \frac{t}{n^{(k)}}\right) \gamma_t^{(k)} - \sum_{t=1}^{n^{(k)}/2-1} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} = \sum_{t=1}^{n^{(k)}-1} \left(1 - \frac{t}{n^{(k)}}\right) \gamma_t^{(k)} - \sum_{t=1}^{n^{(k)}/2} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} \end{aligned} \quad (15)$$

In the last equality we added and subtracted $(1 - t/n^{(k)})\gamma_t^{(k)}$ and in addition used that for $t = n^{(k)} - 1$, $(1 - t/n^{(k)})\gamma_t^{(k)} = 1/n^{(k)}\gamma_{n^{(k)}-1}^{(k)}$. This means that if we consider

$$2(S_{2j-1} + S_{2j}) \stackrel{(15)}{=} \frac{2}{n^{(k)}} \left[\sum_{t=1}^{n^{(k)}-1} \left(1 - \frac{t}{n^{(k)}}\right) \gamma_t^{(k)} - \sum_{t=1}^{n^{(k)}/2} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} \right] = \varepsilon_k - \frac{2}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} \quad (16)$$

Now substitute (14) and (16) into (12)

$$\begin{aligned}
\varepsilon_{k+1} &\stackrel{(12)}{=} 2(S_{2j-1} + S_{2j}) + S_{2j+1} - S_{2j-1} \\
&\stackrel{(16)}{=} \varepsilon_k - \frac{2}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} + \frac{2}{n^{(k)}} \sum_{t=2}^{n^{(k)}/2} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} - \frac{1}{n^{(k)}} \left(1 - \frac{1}{n^{(k)}}\right) \gamma_1^{(k)} \\
&= \varepsilon_k - \frac{1}{n^{(k)}} \frac{\gamma^{(k)}}{n^{(k)}} - \frac{\gamma_1^{(k)}}{n^{(k)}} + \frac{1}{n^{(k)}} \frac{\gamma^{(k)}}{n^{(k)}} = \varepsilon_k - \frac{\gamma_1^{(k)}}{n^{(k)}}
\end{aligned}$$

Subtract $\varepsilon_{k+1} - \gamma_1^{(k)}/n^{(k)}$ from each side of the equation, and the proposition follows. \blacksquare

The proposition says that the rate of convergence of the method is only dependant on the correlation between sequential observations $X_i^{(k)}$, and not the full correlation structure. But the following two corollaries show that the proposition prove the expected behaviour of the blocking method.

COROLLARY 1. Assume $\gamma_t^{(k)} \geq 0$ for all t then

- if $\gamma_1^{(k)} > 0$ for all $j \leq k \leq i$, the sequence of errors ε_k is strictly decreasing for all $j \leq k \leq i$.
- if $\gamma_1^{(k)} \geq 0$ for all $j \leq k \leq i$, the sequence of errors ε_k is decreasing for all $j \leq k \leq i$.
- if there exist some k such that $\gamma_1^{(k)} = 0$, then the sequence of errors ε_k become constant.

Proof. We prove the first part of the hypothesis. Assume the hypothesis is true and $\gamma_1^{(k)} > 0$ for all $j \leq k \leq i$. That means proposition 2 is true. Suppose $u > v$ and notice that since $n^{(k)} > 0$ and $\gamma_1^{(k)} > 0$ is by hypothesis, then the sum of such terms must be positive, and therefore

$$\begin{aligned}
0 < \sum_{k=v}^{u-1} \frac{\gamma_1^{(k)}}{n^{(k)}} &= \frac{\gamma_1^{(v)}}{n^{(v)}} + \frac{\gamma_1^{(v+1)}}{n^{(v+1)}} + \cdots + \frac{\gamma_1^{(u-1)}}{n^{(u-1)}} \stackrel{(11)}{=} (\varepsilon_v - \varepsilon_{v+1}) + (\varepsilon_{v+1} - \varepsilon_{v+2}) + \cdots + (\varepsilon_{u-1} - \varepsilon_u) \\
&= \varepsilon_v - \varepsilon_u,
\end{aligned}$$

since every term of the sum cancel except from the first and the last. Now, by adding ε_u to each side of that strict inequality, the first part is proven. To obtain the rest of the corollary, just replace the strict inequality with an inequality or equality. \blacksquare

COROLLARY 2. If the blocked variables become independent, for some number of blocking transformation, then the sequence of errors ε_k become constant.

Proof. Use the previous corollary, and notice that $\gamma_1^{(k)} = 0$ if X_i, X_{i+1} are independent for all $0 \leq i, j \leq 2^{-k}m$. \blacksquare

LEMMA 3. Suppose j and k are positive natural number and the sample size $m \geq 2^k(j-1) + 2^{k+1} - 1$, then $k < \log_2 m$ and

$$\begin{aligned}
\gamma_j^{(k)} &= 2^{-2k} \left[\gamma_{2^k(j-1)+1}^{(0)} + 2\gamma_{2^k(j-1)+2}^{(0)} + 3\gamma_{2^k(j-1)+3}^{(0)} + \cdots + 2^k \gamma_{2^k(j-1)+2^k}^{(0)} \right. \\
&\quad \left. + (2^k - 1)\gamma_{2^k(j-1)+2^k+1}^{(0)} + (2^k - 2)\gamma_{2^k(j-1)+2^k+2}^{(0)} + \cdots + \gamma_{2^k(j-1)+2^k+1-1}^{(0)} \right]. \quad (17)
\end{aligned}$$

Proof. We first show that $k < \log_2 m$. Fix j and k such that $m \geq 2^k(j-1) + 2^{k+1} - 1$, then

$$m \geq \underbrace{2^k(j-1)}_{\geq 0} + \underbrace{2^{k+1} - 1}_{\geq 2^k+1} \geq 2^k+1 \quad \text{only if} \quad \log_2 m \geq \log_2(2^k+1) > \log_2 2^k = k \log_2 2 = k$$

We prove the rest of the lemma by induction. Fix j such that $m \geq 2^1(j-1) + 2^{1+1} - 1$, in particular this ensures that if $k = 1$, then $m \geq 2j + 1$ and therefore $\gamma_{2j+1}^{(0)}$ exists. Define $M = \sup_{k \in \mathbb{N}} \{m \geq 2^k(j-1) + 2^{k+1} - 1\}$. Assume $k = 1$ and write

$$\gamma_j^{(1)} \stackrel{(10)}{=} 2^{-2} \left(\gamma_{2j-1}^{(0)} + \gamma_{2j}^{(0)} + \gamma_{2j+1}^{(0)} \right)$$

Assume now that there exist a positive natural number $k < M$ such that equation (17) is true. This implies $k + 1 \leq M$ and hence $\gamma_{2^k 2j + 2^k + 2^k - 1}^{(0)}$ exists, and we can write

$$\begin{aligned} \gamma_j^{(k+1)} &\stackrel{(10)}{=} 2^{-2} (\gamma_{k,2j-1} + 2\gamma_{k,2j} + \gamma_{k,2j+1}) \stackrel{(17)}{=} \\ &2^{-2} 2^{-2k} \left(\gamma_{0,2^k(2j-2)+1} + 2\gamma_{0,2^k(2j-2)+2} + \cdots + 2^k \gamma_{0,2^k(2j-2)+2^k} \right. \\ &\quad \left. + (2^k - 1)\gamma_{0,2^k(2j-2)+2^k+1} + \cdots + \gamma_{0,2^k(2j-2)+2^k+2^k-1} \right) + \\ &2^{-2} 2^{-2k} \left(2\gamma_{0,2^k(2j-1)+1} + 4\gamma_{0,2^k(2j-1)+2} + \cdots + 2^k \gamma_{0,2^k(2j-1)+2^k} \right. \\ &\quad \left. + 2(2^k - 1)\gamma_{0,2^k(2j-1)+2^k+1} + \cdots + 2\gamma_{0,2^k(2j-1)+2^k+2^k-1} \right) + \\ &2^{-2} 2^{-2k} \left(\gamma_{0,2^k 2j+1} + 2\gamma_{0,2^k 2j+2} + \cdots + 2^k \gamma_{0,2^k 2j+2^k} \right. \\ &\quad \left. + (2^k - 1)\gamma_{0,2^k 2j+2^k+1} + \cdots + \gamma_{0,2^k 2j+2^k+2^k-1} \right). \end{aligned}$$

By using that $2^{-2} 2^{-2k} = 2^{-(k+1)}$ and factoring $\gamma_K^{(0)}$ together for all $2^{k+1}(j-1) + 1 \leq K \leq 2^{k+1}(j-1) + 2^{k+2} - 1$, the lemma follows. \blacksquare

The next proposition is useful because combined with proposition 2, it shows how the convergence of the method is uniquely determined by the initial correlation structure.

PROPOSITION 3. Suppose $(\gamma_1^{(k)}, \gamma_2^{(k)}, \dots, \gamma_{n_k}^{(k)})$ denote the correlation structure at blocking iteration number k . Suppose $m \geq 2^{k+1} - 1$, then $k < \log_2 m$ and,

$$2^{2k} \gamma_1^{(k)} = \gamma_1^{(0)} + 2\gamma_2^{(0)} + 3\gamma_3^{(0)} + \cdots + 2^k \gamma_{2^k}^{(0)} + (2^k - 1)\gamma_{2^k+1}^{(0)} + \cdots + \gamma_{2^{k+1}-1}^{(0)}$$

Proof. Use the previous lemma with $j = 1$. \blacksquare

THEOREM 2 (Blocking method). Suppose X_1, \dots, X_m are $m = 2^d$ identically distributed random variables with stationary non-negative covariances. Let $\bar{X}_m = (1/m) \sum_{j=1}^m X_j$ denotes their average and $X_i^{(j)}$ be a random variable subject to j blocking transformations. Then there exist a $k \in \mathbb{N}$ such that

$$k < \log_2 m \quad \text{and} \quad \text{Var}(\bar{X}_m) - \text{Var}(X_i^{(k)}) < \frac{8}{m} \text{Var}(X_i^{(k)}) \quad \text{for all} \quad 1 \leq i \leq 2^{-k} m.$$

Proof to come. Did not have time to finish it before the exams. Furthermore, the proof appear to be more difficult than first anticipated. If there is interest in it, I can try to work out the details after the exams finish.

This theorem says that after less than $\log_2 m$ blocking iterations, the difference between the variance of the mean and variance of the blocked variables, is essentially zero. This is the case since m is typically many orders of magnitude larger than $\text{Var}(X_i^{(k)})$.

We proceed to discuss the results in the order they were presented. Comparing the expected energies of the ground state that we calculated, with references such as CCSD done by M. Pedersen Lohne et al. 2011

for $\omega = 1$ and $n = 2, 6$ and 12 , we see that our errors are roughly 0.1%. The CCSD computations found $\langle E \rangle = 3.000282$ and 20.189900 and 65.789460 respectively, whilst we obtained $\langle E \rangle = 3.0057 \pm 8 \cdot 10^{-4}$ and $20.231 \pm 9 \cdot 10^{-3}$ and $65.881 \pm 2 \cdot 10^{-2}$ respectively. The discrepancy could be due to multiple causes. For example, the test state function, although arguably reasonable, is not exact. The variational parameters is only specified up to some finite number of digits. Moreover, it became apparent that the parameter Δt was of utmost importance for stability and actually for the value of the final estimates. It turned out that it was possible to adjust this value to coarsely adjust the estimated values. This was also true about the estimates obtained by steepest decent. This is unfortunate, because we have no reference to which we know how to fix Δt .

Surprisingly the value of the parameters to which the steepest decent algorithm converged, appeared to depend on the value of Δt . However the dependence was much less systematic than for the expectation values. This could also be contributing factors to the accuracy of our estimates. As we saw, there are several ways to optimize the steepest decent method, we proposed to maximize $f(\mathbf{r}_n - \zeta_n \nabla f(\mathbf{r}_n))$ because intuitively it ensures that each step moves as as close as possible to the minimum. The problem with this is however, that it works intuitively, but there were problems in practise. If we instead approximate the expected energy by a quadratic form, it is straight forward to implement the conjugate gradient method, which only requires two steps for convergence (Press et al. 2007). This would only be useful if the expected energy is an approximate quadratic form to begin with.

But there were more issues with the optimization with the steepest decent method: The stability of the implementation was reasonably good, but not good enough to run without supervision. Every once in a while the method diverged without warning. However, it seems reasonable that we should not only blame this on the optimization method: In fact the state function does not remain a probability distribution for all values of α, β . For example, it is easy to choose starting variational parameters such that the expected energy is no longer an expectation value for any given starting configuration of the electrons. In this way, the function at hand also poses some complications. It could be useful to investigate whether it is possible to improve stability by also minimizing with respect to some other quantities, such as the variance of the energy for example. It would be interesting to investigate if some modern stochastic minimization method is better suited for the task. In statistics, we use maximum likelihood estimation using Newton-Raphson method.

Our investigation into performance gains due to parallelization showed that the speed ups are excellent. The proportion of the workload that was parallelizable was almost exactly 1. This number would probably increase if the number of Monte Carlo cycles increase because then a larger proportion of the time would be spent on sampling, and the overhead would stay almost exactly the same. So although 800 000 was a lower bound on the expected value of the maximal speed up, this estimate is potentially unbounded if the number of Monte-Carlo cycles increase. We saw that it was possible to estimate this with Jensen equality. As we stated, Jensen inequality becomes an equality in the case that f has zero variance. In our case, the variance was 10^{-14} , so it is possible that the expected value of the estimate is not only a lower bound, but close to the exact value. Unfortunately, estimating the variance of the estimate probably require numerics. I am unable of any analytical approach, but with bootstrapping it is straight forward to make such estimates. This pose prospects for further work. Also surprisingly we saw that there was no effect of vectorization of the serial code.

The paper by Flyvbjerg and Petersen possibly give an excellent introduction for applications (Flyvbjerg and Petersen 1989, pp. 461-466). However, mathematically interested readers may be put off by the crude (and wrong) mathematical treatment. This was motivational, along with that a brief search for proof of the Blocking method returned nothing relevant. The new proofs probably deserve some refinement when there is more time, but that said, the two most important results of the new theory were propositions 2 and 3. The first proposition says that the errors of the blocking method will monotonously decrease until the blocked variables become independent. At this point the error become constant. However on first readings it may

seem we are able to use equation (9) to fix this constant to zero. If this turns out to be correct, this proves the behaviour of the blocking method, and may suffice for any tutution of the blocking method. However, I believe there is more to be had from the theory. The proposition also says that at blocking iterations number k the convergence rate of the method only sees the correlations between subsequent observations. To new readers, this may be startling. But my intuition is that the stability of the method may be much attributed to this: If the dependence of the correlation was more complicated, it is also reasonable to expect that the behaviour of the convergence was more complicated. The second proposition 3 connect proposition 2 to the initial correlation structure. This correlation structure is available before blocking starts and, as we saw from proposition 2 contain all information about the convergence. Therefore it should be possible do more work into finding the optimal number of blocking iterations. Flyvbjerg and Petersen 1989 understand the value of this, and dedicate an entire section of their paper to try to present workaround. There seems there are prospects for further work in all of these areas. Finally, we see that proposition 3 is the convolution of the initial auto-correlation structure with the Smarandache Crescendo Pyramidal sequence:

$$\begin{array}{ccccccc}
 & & & & 1 & & \\
 & & & & 1 & 2 & 1 \\
 & & & 1 & 2 & 3 & 2 & 1 \\
 & & 1 & 2 & 3 & 4 & 3 & 2 & 1 \\
 & & & & & & & & \vdots
 \end{array}$$

It could be possible to use some of the existing research on the sequence and simplify the proposition further. This is interesting since it contains all the information about the convergence, and our ability to understand influence further developments.

CONCLUSION AND PERSPECTIVES

Our main finding are propositions 2 and 3. We conjecture that together they imply a more satisfying proof of the Blocking method than the one provided in this text. Though as they stand, the corollaries of proposition 2 seem to prove the behaviour of the Blocking method behaviour if we are able to fix the constant which arises in the first corollary. Here more research is needed. There were not enough time to work out all the details before the exam. For the same reason it was not possible to find the time to complete the proof of the proposed upper bound of the blocking method. The proof seemed more difficult than first anticipated, but seems to be within reach since propositions 2 and 3 together seem to make all the information about the convergence of the Blocking method available prior to Blocking begins.

Our physics results were exciting, we made tables which characterized the behaviour of the electrons of the quantum dot for $n = 2, 6, 12$ and 20 electrons for a variety of values of ω . Together, they tell us what the expected distance between the particles were, their, kinetic and potential energy. The values of the ground state energy had tolerances within 0.1% with CCSD benchmarks. However, in this area, there is probably prospects for more work. In particular in regards to the occasional lack of stability and implementation of parameter optimization. Surprisingly we saw that there were no effect of vectorization, but the speed up due to parallelization was almost exactly 100%

Finally, we were asked to provide constructive ideas to the course administration such that the course is kept at an interactive level. At the time of writing, the execution of project 2 went relatively smoothly, and I am much impressed with the efforts, in particular of Håkon for helping me troubleshoot the code for many hours on one occasion. The only comment that I can make is that perhaps more information could have been available about the parameter Δt ; this would certainly have helped me understand why I had singular Slater determinants whilst building the program. This concludes the report of project 2.

APPENDIX

ANALYTICAL EXPRESSIONS

Definition of trail energy

$$E_T = \sum_{i=1}^N \left[-\frac{1}{2} \frac{\nabla_i^2 \psi_T}{\psi_T} + \frac{1}{2} \omega^2 r_i^2 + \sum_{j=i+1}^N \frac{1}{r_{ij}} \right] \quad (18)$$

Chain rule on smooth m -manifolds

$$\frac{\nabla_i^2 \psi_T}{\psi_T} = \frac{\nabla_i^2 \psi_0}{\psi_0} + \frac{\nabla_i^2 \psi_C}{\psi_C} + 2 \frac{\nabla_i \psi_C}{\psi_C} \frac{\nabla_i \psi_0}{\psi_0}$$

Expressions to determine energy in the case $n = 2$:

$$\psi_T(\mathbf{r}_1, \mathbf{r}_2) = C(\psi_0 \psi_C)(\mathbf{r}_1, \mathbf{r}_2), \quad \psi_0(\mathbf{r}_1, \mathbf{r}_2) = \exp\left(-\frac{1}{2} \alpha \omega (r_1^2 + r_2^2)\right), \quad \psi_C(\mathbf{r}_1, \mathbf{r}_2) = \exp\left(\frac{\gamma r_{12}}{\beta r_{12} + 1}\right)$$

$$\begin{aligned} \frac{\nabla_i^2 \psi_0}{\psi_0} &= \alpha^2 \omega^2 r_i^2 - 2\alpha\omega, & \frac{\nabla_i^2 \psi_C}{\psi_C} &= \frac{\gamma(1 + r_{ij}\gamma - \beta^2 r_{12}^2)}{r_{12}(1 + \beta r_{12})^4}, & \sum_{i=1}^2 \frac{\nabla_i \psi_C}{\psi_C} \cdot \frac{\nabla_i \psi_0}{\psi_0} &= -\frac{\alpha\omega\gamma r_{ij}}{(1 + \beta r_{12})^2} \\ \frac{\nabla_i \psi_T}{\psi_T} &= \mathbf{e}_1 \left(-\alpha\omega x_i + \frac{\gamma(x_i - x_j)}{(1 + \beta r_{ij})^2 r_{ij}} \right) + \mathbf{e}_2 \left(-\alpha\omega y_i + \frac{\gamma(y_i - y_j)}{(1 + \beta r_{ij})^2 r_{ij}} \right) \end{aligned}$$

Auxiliary expressions for steepest decent method in the case $n = 2$:

$$\begin{aligned} \frac{\partial}{\partial \theta_j} \langle E \rangle (\theta_1, \theta_2, \dots, \theta_n) &= 2 \left\langle \frac{E}{\psi_T} \frac{\partial \psi_T}{\partial \theta_j} \right\rangle - 2 \left\langle \frac{1}{\psi_T} \frac{\partial \psi_T}{\partial \theta_j} \right\rangle \langle E \rangle \\ \frac{1}{\psi_T} \frac{\partial \psi_T}{\partial \alpha} &= -\frac{\omega}{2} (r_1^2 + r_2^2) & \frac{1}{\psi_T} \frac{\partial \psi_T}{\partial \beta} &= -\gamma \left(\frac{r_{12}}{1 + \beta r_{12}} \right)^2 \end{aligned}$$

For more than two particles the derivatives of the trail state function are

$$\begin{aligned} \frac{\nabla_i \psi_T}{\psi_T} &= \frac{\nabla_i \det D_\uparrow}{\det D_\uparrow} + \frac{\nabla_i \det D_\downarrow}{\det D_\downarrow} + \frac{\nabla_i \psi_C}{\psi_C} \\ \frac{\nabla_i^2 \psi_T}{\psi_T} &= \frac{\nabla_i^2 \det D_\uparrow}{\det D_\uparrow} + \frac{\nabla_i^2 \det D_\downarrow}{\det D_\downarrow} + \frac{\nabla_i^2 \psi_C}{\psi_C} + 2 \left(\frac{\nabla_i \det D_\uparrow}{\det D_\uparrow} + \frac{\nabla_i \det D_\downarrow}{\det D_\downarrow} \right) \cdot \frac{\nabla_i \psi_C}{\psi_C} \end{aligned}$$

Derivatives of determinants

$$\frac{\nabla_i \det D_\uparrow}{\det D_\uparrow} = \sum_{j=0}^{n/2-1} (D_\uparrow^{-1})_{ji} \nabla_i \phi_j(\mathbf{r}_i), \quad \frac{\nabla_i^2 \det D_\uparrow}{\det D_\uparrow} = \sum_{j=0}^{n/2-1} (D_\uparrow^{-1})_{ji} \nabla_i^2 \phi_j(\mathbf{r}_i); \quad \text{for } D_\downarrow \text{ replace } \uparrow \text{ by } \downarrow$$

$$\nabla \phi_{nx,ny} = \mathbf{e}_1 \left(2n_x (\alpha\omega)^{1/2} \phi_{nx-1,ny} - \alpha\omega x \phi_{nx,ny} \right) + \mathbf{e}_2 \left(2n_y (\alpha\omega)^{1/2} \phi_{nx,ny-1} - \alpha\omega y \phi_{nx,ny} \right)$$

$$\begin{aligned} \nabla^2 \phi_{nx,ny} &= 4\alpha\omega (n_x(n_x - 1) \phi_{nx-2,ny} + n_y(n_y - 1) \phi_{nx,ny-2}) - 4(\alpha\omega)^{3/2} (x n_x \phi_{nx-1,ny} + y n_y \phi_{nx,ny-1}) \\ &\quad + \alpha\omega \phi_{nx,ny} (\alpha\omega r^2 - 2) \end{aligned}$$

Derivatives of the Jastrow factors are

$$\frac{\nabla_k \psi_C}{\psi_C} = \sum_{i=1, i \neq k}^n \frac{\gamma_{ik}}{(1 + \beta r_{ik})^2 r_{ik}} (\mathbf{e}_1(x_k - x_i) + \mathbf{e}_2(y_k - y_i))$$

$$\frac{\nabla_k^2 \psi_C}{\psi_C} = \left[\sum_{j=1, j \neq k}^n \frac{\gamma_{kj}(x_k - x_j)}{(1 + \beta r_{kj})^2 r_{kj}} \right]^2 + \left[\sum_{j=1, j \neq k}^n \frac{\gamma_{kj}(y_k - y_j)}{(1 + \beta r_{kj})^2 r_{kj}} \right]^2 + \sum_{j=1, j \neq k}^n \frac{\gamma_{kj}(1 - \beta r_{kj})}{r_{kj}(1 + \beta r_{kj})^3}$$

Auxiliary expressions for the steepest decent method are

$$\begin{aligned} \frac{1}{\psi_T} \frac{\partial \psi_T}{\partial \alpha} &= \frac{1}{\det D_{\uparrow}} \frac{\partial \det D_{\uparrow}}{\partial \alpha} + \frac{1}{\det D_{\downarrow}} \frac{\partial \det D_{\downarrow}}{\partial \alpha}, & \frac{1}{\psi_T} \frac{\partial \psi_T}{\partial \beta} &= - \sum_{i=1}^n \sum_{j=i+1}^n \gamma_{ij} \left(\frac{r_{ij}}{1 + \beta r_{ij}} \right)^2 \\ \frac{1}{\det D_{\uparrow}} \frac{\partial \det D_{\uparrow}}{\partial \alpha} &= \sum_{i=0}^{n/2-1} \sum_{j=0}^{n/2-1} d_{ij}^{-1} \frac{\partial \phi_{2i}}{\partial \alpha}(\mathbf{r}_{2j}), & \frac{1}{\det D_{\downarrow}} \frac{\partial \det D_{\downarrow}}{\partial \alpha} &= \sum_{i=0}^{n/2-1} \sum_{j=0}^{n/2-1} d_{ij}^{-1} \frac{\partial \phi_{2i+1}}{\partial \alpha}(\mathbf{r}_{2j+1}) \end{aligned} \quad (19)$$

$$\frac{\partial}{\partial \alpha} \phi_i(\mathbf{r}_j) = (x_j n_{xi} \phi_{n_{xi}-1, n_{yi}}(\mathbf{r}_j) + y_j n_{yi} \phi_{n_{xi}, n_{yi}-1}(\mathbf{r}_j)) \left(\frac{\omega}{\alpha} \right)^{1/2} - \frac{1}{2} \omega r_j^2 \phi_{n_{xi}, n_{yi}}(\mathbf{r}_j) \quad (20)$$

PROOFS

THEOREM 3 (Hastings-Metropolis theorem). *Suppose that $C\pi_i$ is a discrete probability distribution. If q_{ij} is any irreducible transition probability matrix, and X_n is a Markov chain with transition probability matrix*

$$P_{ij} = \begin{cases} \alpha_{ij} q_{ij}, & j \neq i \\ q_{ii} + \sum_{k=0}^{\infty} q_{ik}(1 - \alpha_{ik}) & j = i \end{cases}, \quad \text{where} \quad \alpha_{ij} = \min \left(\frac{\pi_j q_{ji}}{\pi_i q_{ij}}, 1 \right), \quad (21)$$

then X_n is time reversible with stationary distribution π_i .

Proof. Assume that the hypothesis is true. Then in particular $q_{ij} \neq 0$ for all i, j since q is irreducible. Notice that if

$$\frac{\pi_j q_{ji}}{\pi_i q_{ij}} = 1,$$

then there is nothing to prove since then $\alpha_{ij} = 1$ and $\alpha_{ji} = 1$, and therefore

$$\pi_i P_{ij} = \pi_j P_{ji} \quad (22)$$

is automatic. Hence it suffices to prove (22) for the two cases

$$\frac{\pi_j q_{ji}}{\pi_i q_{ij}} > 1 \quad \text{and} \quad \frac{\pi_j q_{ji}}{\pi_i q_{ij}} < 1,$$

separately. Suppose first that $\pi_j q_{ji} > \pi_i q_{ij}$ (\dagger). Write:

$$\pi_i P_{ij} \stackrel{(21)}{=} \pi_i q_{ij} \alpha_{ij} \stackrel{(21)(\dagger)}{=} \pi_i q_{ij} \cdot 1 = \pi_i q_{ij} \frac{\alpha_{ji}}{\alpha_{ji}} = \alpha_{ji} \pi_i q_{ij} \frac{1}{\alpha_{ji}} \stackrel{(\dagger)}{=} \alpha_{ji} \pi_i q_{ij} \frac{\pi_j q_{ji}}{\pi_i q_{ij}} = \alpha_{ji} \pi_j q_{ji} \stackrel{(21)}{=} \pi_j P_{ji}.$$

In the case that $\pi_j q_{ji} < \pi_i q_{ij}$ (\ddagger), write

$$\pi_i P_{ij} \stackrel{(21)}{=} \pi_i q_{ij} \alpha_{ij} \stackrel{(21)(\ddagger)}{=} \pi_i q_{ij} \frac{\pi_j q_{ji}}{\pi_i q_{ij}} = \pi_j q_{ji} = \pi_j q_{ji} \cdot 1 \stackrel{(21)(\ddagger)}{=} \pi_j q_{ji} \cdot \alpha_{ji} = \pi_j P_{ji},$$

which means X_n is time reversible with stationary probability π_i . ■

LEMMA. *Suppose X and Y are random variables with finite variance, then*

$$|\text{Cov}(X, Y)|^2 \leq \text{Cov}(X, X) \text{Cov}(Y, Y)$$

Proof. Suppose Ω denotes the vector space of all real random variables with finite variance over the real numbers. Suppose $X, Y \in \Omega$ and define

$$(X, Y) = \langle XY \rangle. \quad (23)$$

To see that $(\Omega, (\cdot, \cdot))$ is an inner product space, note that

$$\text{Positivity: } (X, X) = \langle X^2 \rangle \geq 0 \quad = 0 \quad \text{if and only if} \quad X = 0 \quad \text{almost surely.}$$

$$\text{Symmetry: } (X, Y) = \langle XY \rangle = \langle YX \rangle = (Y, X)$$

$$\text{Bilinearity: } (aX + bY, Z) = \langle (aX + bY)Z \rangle = a \langle XZ \rangle + b \langle YZ \rangle = a(X, Z) + b(Y, Z)$$

This proves that (\cdot, \cdot) is an inner product on Ω (McDonald and Weiss 2012). In particular, that means we can use Cauchy-Schwarz inequality (Lay 2012). Now use the definition of covariance (*) and apply Cauchy-Schwarz inequality (**). Write

$$\begin{aligned} |\text{Cov}(X, Y)|^2 &\stackrel{(*)}{=} |\langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle|^2 \quad = 0 \stackrel{(23)}{=} |(X - \langle X \rangle, Y - \langle Y \rangle)|^2 \\ &\stackrel{(**)}{\leq} (X - \langle X \rangle, X - \langle X \rangle)(Y - \langle Y \rangle, Y - \langle Y \rangle) \stackrel{(*)}{=} \text{Cov}(X, X) \text{Cov}(Y, Y) \end{aligned}$$

Which is what we wanted. ■

LITERATURE CITED

- [1] Morris H. DeGroot and Mark J. Schervish. *Probability and Statistics*. 4th ed. Pearson Education, Inc., 2012.
- [2] Jay L. Devore and Kenneth L. Berk. *Modern Mathematical Statistics with Applications*. 2nd ed. London: Springer, 2012.
- [3] H. Flyvbjerg and H.G Petersen. “Error Estimates on Averages of Correlated Data”. In: *The Journal of Chemical Physics* 91 (July 1, 1989), pp. 461–466.
- [4] Michiel Hazewinkel. *Gradient Method*. In: *Encyclopedia of Mathematics*. 1st ed. Vol. 4. 13 vols. The Netherlands: Kluwer, 1989.
- [5] David C. Lay. *Linear Algebra and Its Applications*. 4th ed. Boston: Addison-Wesley, 2012. 481 pp.
- [6] M. Pedersen Lohne et al. “Ab Initio Computation of the Energies of Circular Quantum Dots”. In: *Physical review B* 84 (2011).
- [7] John N. McDonald and Neil A. Weiss. *A Course in Real Analysis*. 2nd ed. Oxford: Academic press, 2012.
- [8] Nicholas Metropolis et al. “Equation of State Calculations by Fast Computing Machines”. In: *The Journal of Chemical Physics* 21.6 (June 1953), pp. 1087–1092.
- [9] William H. Press et al. *Numerical Recipes*. 3rd ed. Cambridge: Cambridge university press, 2007.
- [10] Robert H. Shumway and David S. Stoffer. *Time Series Analysis and Its Applications with R Examples*. 3rd ed. New York: Springer, 2011.
- [11] Sathamangalam Ranga Iyengar Srinivasa Varadhan. *Probability Theory*. 1st ed. Providence: American Mathematical Society, 2001.