

# PROOF OF THE BLOCKING METHOD

Marius Jonsson

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

May 29, 2017

## ABSTRACT

This is a report submission for the first 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 interacting electrons. We develop the mathematics of the Blocking method and we eventually prove the blocking method that Flyvbjerg and Pettersen claim in their 1989 paper. Specifically, we prove that if  $X_1, X_2, \dots, X_m$  are  $2^n$  identically distributed random variables with stationary non-negative covariances,  $\bar{X}_m = (1/m) \sum_{j=1}^m X_j$  denote their average and  $X_i^{(j)}$  be a random variable subject to  $j$  blocking transformations. Then there exist a  $k \in \mathbb{N}$  such that  $k < \log_2 m$  and

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

Next, we compute the trail energy of a Slater-Jastrow type trail wave function for theory of a harmonic oscillator with Coulomb repulsion. We also compute mean distance between electrons, their expected kinetic- and potential energy.

## INTRODUCTION

The implementation of a Variational Monte Carlo scheme using Hastings-Metropolis algorithm is a robust way of estimating observables if a proposal of the wave function  $\psi_T$  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 1953 paper (Equation of State Calculations by Fast Computing Machines 1953 ref). In the begining of 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 chains,  $\{X_i, i \in \mathbb{N}\}$ , with stationary probability distribution  $\psi_T$ . Using  $X_i$  we can compute any desired statistic, 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 (referanse). 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  $\psi_T$  for any natural number of electrons. We assume the parameters  $\alpha, \beta$  are given. In our case they are not known, but will be fixed such that  $\langle E(\alpha, \beta) \rangle$  is a minima.

Although that the variational Monte-Carlo scheme is invaluable to any physicist interested in computational physics, the highlights of this report are perhaps new contributions to the theory of the Blocking method. In the evaluation of the implementation of our variational Monte-Carlo scheme, we will use the Blocking method to compute the variance of the means of our statistics:  $\text{Var}(\overline{E})$ . In the results I develop the mathematics to contain 3 immediate lemmas of the work by Flyvbjerg and Petersen, and two propositions which combined imply the Blocking method.

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

## METHODS

This report assumes some previous knowledge of the Blocking method. These details have therefore been omitted from the methods section and been moved to the appendix. For the evaluation of this report, it is important, otherwise the methods section may appear unsatisfactory on first reading. An algebraic understanding is necessary since the report develops the mathematics of the Blocking method further. To readers unfamiliar of the blocking method, the appendix contain proof of auxiliary results which are simultaneous excellent introduction to the idea of the Blocking method itself.

Let's first discuss the Hastings-Metropolis theorem. It's constructive proof is contained in the appendix. However, the Hastings-Metropolis method is easily made precise. 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  $\pi_i$ .*

This theorem makes the entire variational Monte-Carlo scheme well defined if we define  $\pi_i$  and  $q_{ij}$ . We will take  $\pi_i$  to equal  $\psi_T$  and refer to the density  $q_{ij}$  as **importance sampling**. We will also present source

code, and show that the code produces known benchmarks and tests to help us verify that the variational Monte-Carlo-code runs correctly.

As we noted, theorem 3 implies that the variational Monte-Carlo scheme is uniquely defined once we specify  $\psi_T$  and the importance sampling. We will take the trial wave function to be of 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

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

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

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

$j$	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  $\mathbf{r}_k^{(i)}, \mathbf{r}_k^{(j)}$  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  $\Delta t$  (ref). 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 suffices to fix the implementation of importance sampling, as we shall see in practical example at the end of this section. As announced we want to estimate the value of the variational parameters  $\alpha$  and  $\beta$ . 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

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

In order to ensure convergence one has to find a sequence  $\{\gamma_i\}_{i=1}^\infty$  that ensures convergence. Ideally such that  $f(\mathbf{r}^{(i)} - \gamma_i \nabla f(\mathbf{r}^{(i)}))$  is maximised for all  $i \in \mathbb{N}$  (Hazewinkel 1989). However for this implementation, automation was not carried out. As we explain in the introduction, we define  $\alpha$  and  $\beta$  such that they minimize the energy 1. 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  $\theta$

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

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. See appendix for details of the derivation and the resulting equations.

Let's talk about how the variational Monte-Carlo scheme was implemented 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  $\psi_T$  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 the derivation are found in the appendix, for anyone following the derivation (1) once it is evaluated 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,  $\mathbf{r}$  and  $\mathbf{rpp}$  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, the program returns that the energy is  $\langle E_T \rangle = 3.00052 \pm 10^{-5}$  after only  $2^{20} \approx 10^6$  Monte-Carlo cycles. For reference, the exact result is 3.

We want to generalize the program to more than two electrons,  $n > 2$ . This adds the complication of adding a Slater determinant. First, the analytical expression of the energy becomes large. This is chiefly due to the dependence of  $\alpha$  in the functions  $\phi_{ij}$  in the Slater determinant and their complicated interactions. Therefore  $\nabla^2 \psi_T$  becomes a complicated expression. Note first that for the trial state function we specified

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

We wrote functions which evaluated each of these terms by 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. Suppose  $X \sim \text{unif}(0, 1)$  is a uniformly continuously distributed random variable on  $(0, 1]$ , another complication which arises in the implementation of many particles if evaluation of the inequality

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

The trouble is that due to the exponentials inside the Jastrow factor, these terms will become unpractically large and overflow. 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  $x < y$  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 magnitude smaller than about 1000. There were no other technical complications worth mentioning. Expressions to implement the Steepest decent method are also contained in the appendix.

Care was taken to check that the program ran correctly. - Benchmarks numerical or closed form. - Multiple electrons were backward compatibly checked: + after implementation we checked that in the case  $n = 2$  the numerical formula individually reduced to the closed form expressions. Nevn alle termene i laplacian. Nevn at dette også ble gjort for steepest decent.

```

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
    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  $\mathbf{r}$ ,  $\mathbf{rpp}$  contain the positions at sequential steps,  $\mathbf{F}$ ,  $\mathbf{Fpp}$  contain the quantity  $\nabla\psi_T/\psi_T$  for sequential steps and  $\mathbf{rij}$ ,  $\mathbf{rijpp}$  contain the distance between electrons at sequential steps. The program returns the energy with uncertainty in the variance of  $\text{Var}(\overline{E_T}) = 10^{-10}$  and rejection ratio  $r < 10^{-4}$ .

## RESULTS AND DISCUSSION

I Present your results I Give a critical discussion of your work and place it in the correct context. I Relate your work to other calculations/studies I An eventual reader should be able to reproduce your calculations if she/he wants to do so. All input variables should be properly explained. I Make sure that figures and tables should contain enough information in their captions, axis labels etc so that an eventual reader can gain a first impression of your work by studying figures and tables only.

Next we want to prove the blocking method. The paper by Flyvbjerg and Petersen gives an excellent introduction for applications (Flyvbjerg and Petersen 1989, pp. 461-466). Mathematically interested readers may be affronted by the paper because the paper does not contain proof of the method. The paper falsely claims that the blocking transformations has a fixed point and that conditions can be put on  $\gamma_t$  such that repeated blockings ensure the  $\gamma_{t0}$ s wind up at  $C\delta_{t0}$ . This is false because the blocking transformations are not a linear operator (ref), and therefore  $T$  cannot have a fixed point (ref) moreover. Since there is no fixed point, there is no basin of attraction (ref). We will however prove the blocking method, and study its properties, such as convergence rate and compute the error induced by the algorithm and present sufficient conditions such that the method is stable and converges. To my knowledge, this has never been done for the blocking method. As you will see, the method is suitable for estimation of the variance of random variables.

Suppose  $\bar{X}_m$  is the mean of  $m$  random variables and you 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 this 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

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

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  $m^{(k)} = 2^{-k}m$  is the appropriate number of random variables subject to  $k$  blocking transformations. Also, we will define the autocovariance  $\gamma_{i,j} = \text{Cov}(X_i, X_j)$ . Snakk om stasjonær tidsserie og introduser  $\gamma_t$ .

**PROPOSITION 1.** Suppose  $X_1, \dots, X_m$  are  $m = 2^l$  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

$$\varepsilon_k = \text{Var}(\bar{X}_m) - \text{Var}(X_i^{(k)}) = \frac{2}{n^{(k)}} \sum_{t=1}^{n^{(k)}-1} \left( 1 - \frac{t}{n^{(k)}} \right) (\gamma_k)_t \quad (8)$$

and

$$\gamma_{k,j} = \begin{cases} \frac{1}{2}\gamma_{k-1,2j} + \frac{1}{2}\gamma_{k-1,2j+1} & \text{if } j = 0 \\ \frac{1}{4}\gamma_{k-1,2j-1} + \frac{1}{2}\gamma_{k-1,2j} + \frac{1}{4}\gamma_{k-1,2j+1} & \text{if } j > 0 \end{cases} \quad (9)$$

*Proof.* See Flyvbjerg and Petersen 1989. A proof is also contained in the appendix. ■

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

$$k < \log_2 m, \quad n^{(k)} = 2^{-k}m.$$

*Proof.*  $n^{(k)} = 2^{-k}m$  and  $n^{(k+1)} = 2^{-(k+1)}m$ , and so

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

Rest to come ■

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

$$\gamma_t \leq \gamma_0 \quad \text{for all} \quad t \geq 0.$$

*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

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

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

**PROPOSITION 2.** Suppose  $\varepsilon_k$  denotes the error after  $k$  blocking iterations. Let  $\{\varepsilon_k\}_{k=1}^d$  denote the sequence of errors. Then this sequence is monotonously decreasing and the rate at which it decreases is

$$\varepsilon_k - \varepsilon_{k+1} = \frac{(\gamma_k)_1}{n^{(k)}} \quad (10)$$

*Proof.* It remains to prove formula (10). 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  $\varepsilon_{k+1}$  as

$$\begin{aligned} \varepsilon_{k+1} &\stackrel{(8)}{=} \frac{2}{n^{(k+1)}} \sum_{t=1}^{n^{(k+1)}-1} \left(1 - \frac{t}{n^{(k+1)}}\right) (\gamma_{k+1})_t \stackrel{(9)}{=} \frac{1}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2-1} \left(1 - \frac{2t}{n^{(k)}}\right) (\gamma_{2t-1}^{(k)} + 2\gamma_{2t}^{(k)} + \gamma_{2t+1}^{(k)}) \\ &= S_{2j-1} + 2S_{2j} + S_{2j+1} = 2(S_{2j-1} + S_{2j}) + S_{2j+1} - S_{2j-1} \end{aligned} \quad (11)$$

where we added and subtracted  $S_{2j-1}$  in the last step. Let's examine the terms in the sum individually, write:

$$S_{2j-1} = \frac{1}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2-1} \left(1 - \frac{2t}{n^{(k)}}\right) \gamma_{2t-1}^{(k)} = \frac{1}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2-1} \left(1 - \frac{2t-1}{n^{(k)}}\right) \gamma_{2t-1}^{(k)} - \frac{1}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2-1} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} \quad (12)$$

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

Notice that if we subtract these terms, then almost all terms from one of the sums cancel and we obtain

$$S_{2j+1} - S_{2j-1} = \frac{2}{n^{(k)}} \sum_{t=2}^{n^{(k)}/2} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} - \frac{1}{n^{(k)}} \left(1 - \frac{1}{n^{(k)}}\right) \gamma_1^{(k)} \quad (13)$$

Lets also investigate

$$\begin{aligned} n^{(k)}(S_{2j-1} + S_{2j}) &\stackrel{(12)}{=} \sum_{t=1}^{n^{(k)}/2-1} \left(1 - \frac{2t-1}{n^{(k)}}\right) \gamma_{2t-1}^{(k)} - \sum_{t=1}^{n^{(k)}/2-1} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} + \sum_{t=1}^{n^{(k)}/2-1} \left(1 - \frac{2t}{n^{(k)}}\right) \gamma_{2t}^{(k)} \\ &= \left[ \left(1 + \frac{1}{n^{(k)}}\right) \gamma_1^{(k)} + \left(1 + \frac{2}{n^{(k)}}\right) \gamma_2^{(k)} + \cdots + \left(1 + \frac{n^{(k)}-2}{n^{(k)}}\right) \gamma_{n^{(k)}-2}^{(k)} \right] - \sum_{t=1}^{n^{(k)}/2-1} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} \\ &= \sum_{t=1}^{n^{(k)}-2} \left(1 - \frac{t}{n^{(k)}}\right) \gamma_t^{(k)} - \sum_{t=1}^{n^{(k)}/2-1} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} = \sum_{t=1}^{n^{(k)}-1} \left(1 - \frac{t}{n^{(k)}}\right) \gamma_t^{(k)} - \sum_{t=1}^{n^{(k)}/2} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} \end{aligned} \quad (14)$$

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

$$2(S_{2j-1} + S_{2j}) \stackrel{(14)}{=} \frac{2}{n^{(k)}} \left[ \sum_{t=1}^{n^{(k)}-1} \left(1 - \frac{t}{n^{(k)}}\right) \gamma_t^{(k)} - \sum_{t=1}^{n^{(k)}/2} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} \right] = \varepsilon_k - \frac{2}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} \quad (15)$$

Now substitute (13) and (15) into (11)

$$\begin{aligned} \varepsilon_{k+1} &\stackrel{(11)}{=} 2(S_{2j-1} + S_{2j}) + S_{2j+1} - S_{2j-1} \\ &\stackrel{(15)}{=} \varepsilon_k - \frac{2}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} + \frac{2}{n^{(k)}} \sum_{t=2}^{n^{(k)}/2} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} - \frac{1}{n^{(k)}} \left(1 - \frac{1}{n^{(k)}}\right) \gamma_1^{(k)} \\ &= \varepsilon_k - \frac{1}{n^{(k)}} \frac{\gamma_1^{(k)}}{n^{(k)}} - \frac{\gamma_1^{(k)}}{n^{(k)}} + \frac{1}{n^{(k)}} \frac{\gamma_1^{(k)}}{n^{(k)}} = \varepsilon_k - \frac{\gamma_1^{(k)}}{n^{(k)}} \end{aligned}$$

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

**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

$$\begin{aligned} \gamma_{k,j} &= 2^{-2k} \left[ \gamma_{0,2^k(j-1)+1} + 2\gamma_{0,2^k(j-1)+2} + 3\gamma_{0,2^k(j-1)+3} + \cdots + 2^k \gamma_{0,2^k(j-1)+2^k} \right. \\ &\quad \left. + (2^k - 1)\gamma_{0,2^k(j-1)+2^k+1} + (2^k - 2)\gamma_{0,2^k(j-1)+2^k+2} + \cdots + \gamma_{0,2^k(j-1)+2^k+1-1} \right]. \end{aligned} \quad (16)$$

*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  $\gamma_{0,2j+1}$  exists. Define  $M = \sup_{k \in \mathbb{N}} \{m \geq 2^k(j-1) + 2^{k+1} - 1\}$ . Assume  $k = 1$  and write

$$\gamma_{1,j} \stackrel{(9)}{=} 2^{-2} (\gamma_{0,2j-1} + \gamma_{0,2j} + \gamma_{0,2j+1})$$



Assume now that there exist a positive natural number  $k < M$  such that equation (16) is true. This implies  $k + 1 \leq M$  and hence  $\gamma_{0,2^k 2j+2^k+2^k-1}$  exists, and we can write

$$\begin{aligned} \gamma_{k+1,j} &\stackrel{(9)}{=} 2^{-2} (\gamma_{k,2j-1} + 2\gamma_{k,2j} + \gamma_{k,2j+1}) \stackrel{(16)}{=} \\ &2^{-2} 2^{-2k} \left( \gamma_{0,2^k(2j-2)+1} + 2\gamma_{0,2^k(2j-2)+2} + \cdots + 2^k \gamma_{0,2^k(2j-2)+2^k} \right. \\ &\quad \left. + (2^k - 1)\gamma_{0,2^k(2j-2)+2^k+1} + \cdots + \gamma_{0,2^k(2j-2)+2^k+2^k-1} \right) + \\ &2^{-2} 2^{-2k} \left( 2\gamma_{0,2^k(2j-1)+1} + 4\gamma_{0,2^k(2j-1)+2} + \cdots + 2 \cdot 2^k \gamma_{0,2^k(2j-1)+2^k} \right. \\ &\quad \left. + 2(2^k - 1)\gamma_{0,2^k(2j-1)+2^k+1} + \cdots + 2\gamma_{0,2^k(2j-1)+2^k+2^k-1} \right) + \\ &2^{-2} 2^{-2k} \left( \gamma_{0,2^k 2j+1} + 2\gamma_{0,2^k 2j+2} + \cdots + 2^k \gamma_{0,2^k 2j+2^k} \right. \\ &\quad \left. + (2^k - 1)\gamma_{0,2^k 2j+2^k+1} + \cdots + \gamma_{0,2^k 2j+2^k+2^k-1} \right). \end{aligned}$$

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

**PROPOSITION 3.** Suppose  $\gamma_k = ((\gamma_k)_0, (\gamma_k)_1, \dots, (\gamma_k)_{n_k})$  denote the vector of correlation structure at blocking iteration number  $k$ . Suppose  $m \geq 2^{k+1} - 1$ , then  $k < \log_2 m$  and,

$$2^{2k}(\gamma_k)_1 = (\gamma_0)_1 + 2(\gamma_0)_2 + 3(\gamma_0)_3 + \cdots + 2^k(\gamma_0)_{2^k} + (2^k - 1)(\gamma_0)_{2^k+1} + \cdots + (\gamma_0)_{2^{k+1}-1}$$

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

Add def av stationary non negative decreasing covariance.

**THEOREM 2** (Blocking method). Suppose  $X_1, \dots, X_m$  are  $m = 2^l$  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.  $\blacksquare$

**COROLLARY 1.** Suppose  $X_1, \dots, X_m$  are  $m = 2^l$  are observations from a stationary time series. 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 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 do vis at positiviteten til gammaene er implisitt  $\blacksquare$

**COROLLARY 2.** Suppose  $X_1, \dots, X_m$  are  $m = 2^l$  are observations from a stationary Markov chain. 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 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 do  $\blacksquare$

Moreover  $S^2 = \text{blabla}$  is an unbiased estimator of  $\text{Var}(\bar{X}_m)$ . Bla bla.

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  $\text{Var}(X_i^{(k)})$ .

## CONCLUSION AND PERSPECTIVES

Our main finding are propositions 2 and 3. Together they imply the Blocking method. Here there are many prospects for further work. For example, combining the two propositions, we have all the information about the convergence up to a scalar factor which possibly is easy to determine using some kind of boundary- or initial condition. That would mean that one could go ahead a plot the convergence graph, and probably find a formula for the optimum stopping point of the algorithm. In order to realize automation of the algorithm Flyvbjerg and Petersen explicitly state that they wish for such formulas in their 1989 paper.

Snakk om Fysikkresultatene. Drøft om det er pass for forbedringer av programmet. I State your main findings and interpretations I Try as far as possible to present perspectives for future work I Try to discuss the pros and cons of the methods and possible improvements

## APPENDIX

### ANALYTICAL EXPRESSIONS

Definition of trail energy

$$E_T = \sum_{i=1}^N \left[ -\frac{1}{2} \frac{\nabla_i^2 \psi_T}{\psi_T} + \frac{1}{2} \omega^2 r_i^2 + \sum_{j=i+1}^N \frac{1}{r_{ij}} \right] \quad (17)$$

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  $n = 2$ :

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

$$\frac{\nabla_i^2 \psi_0}{\psi_0} = \alpha^2 \omega^2 r_i^2 - 2\alpha\omega, \quad \frac{\nabla_i^2 \psi_C}{\psi_C} = \frac{\gamma(1 + r_{ij}\gamma - \beta^2 r_{12}^2)}{r_{12}(1 + \beta r_{12})^4}, \quad \sum_{i=1}^2 \frac{\nabla_i \psi_C}{\psi_C} \cdot \frac{\nabla_i \psi_0}{\psi_0} = -\frac{\alpha\omega\gamma r_{ij}}{(1 + \beta r_{12})^2}$$

$$\frac{\nabla_i \psi_T}{\psi_T} = \mathbf{e}_1 \left( -\alpha\omega x_i + \frac{\gamma(x_i - x_j)}{(1 + \beta r_{ij})^2 r_{ij}} \right) + \mathbf{e}_2 \left( -\alpha\omega y_i + \frac{\gamma(y_i - y_j)}{(1 + \beta r_{ij})^2 r_{ij}} \right)$$

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

$$\frac{\partial}{\partial \theta_j} \langle E \rangle (\theta_1, \theta_2, \dots, \theta_n) = 2 \left\langle \frac{E}{\psi_T} \frac{\partial \psi_T}{\partial \theta_j} \right\rangle - 2 \left\langle \frac{1}{\psi_T} \frac{\partial \psi_T}{\partial \theta_j} \right\rangle \langle E \rangle$$

$$\frac{1}{\psi_T} \frac{\partial \psi_T}{\partial \alpha} = -\frac{\omega}{2} (r_1^2 + r_2^2) \quad \frac{1}{\psi_T} \frac{\partial \psi_T}{\partial \beta} = -\gamma \left( \frac{r_{12}}{1 + \beta r_{12}} \right)^2$$

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

$$\frac{\nabla_i \psi_T}{\psi_T} = \frac{\nabla_i \det D_\uparrow}{\det D_\uparrow} + \frac{\nabla_i \det D_\downarrow}{\det D_\downarrow} + \frac{\nabla_i \psi_C}{\psi_C}$$

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

Derivatives of determinants

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

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

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

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

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

## PROOFS

In this section we will prove the Hastings-Metropolis theorem. Often whilst proving results about Markov chains we are interested in whether the Markov chain can reach every state from any other state. In the following proof this is important because it ensures that all the divisions we will make 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 are ready for the main result.

**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*

$$P_{ij} = \begin{cases} \alpha_{ij} q_{ij}, & j \neq i \\ q_{ii} + \sum_{k=0}^{\infty} q_{ik}(1 - \alpha_{ik}) & j = i \end{cases}, \quad \text{where} \quad \alpha_{ij} = \min \left( \frac{\pi_j q_{ji}}{\pi_i q_{ij}}, 1 \right), \quad (18)$$

*then  $X_n$  is time reversible with stationary distribution  $\pi_i$ .*

*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

$$\pi_i P_{ij} = \pi_j P_{ji} \quad (19)$$

is automatic. Hence it suffices to prove (19) 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:

$$\pi_i P_{ij} \stackrel{(18)}{=} \pi_i q_{ij} \alpha_{ij} \stackrel{(18)(\dagger)}{=} \pi_i q_{ij} \cdot 1 = \pi_i q_{ij} \frac{\alpha_{ji}}{\alpha_{ji}} = \alpha_{ji} \pi_i q_{ij} \frac{1}{\alpha_{ji}} \stackrel{(\dagger)}{=} \alpha_{ji} \pi_i q_{ij} \frac{\pi_j q_{ji}}{\pi_i q_{ij}} = \alpha_{ji} \pi_j q_{ji} \stackrel{(18)}{=} \pi_j P_{ji}.$$

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

$$\pi_i P_{ij} \stackrel{(18)}{=} \pi_i q_{ij} \alpha_{ij} \stackrel{(18)(\ddagger)}{=} \pi_i q_{ij} \frac{\pi_j q_{ji}}{\pi_i q_{ij}} = \pi_j q_{ji} = \pi_j q_{ji} \cdot 1 \stackrel{(18)(\ddagger)}{=} \pi_j q_{ji} \cdot \alpha_{ji} = \pi_j P_{ji},$$

which means  $X_n$  is time reversible with stationary probability  $\pi_i$ . ■

**PROPOSITION.** Suppose  $X_1, \dots, X_m$  are  $m = 2^l$  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

$$\varepsilon_k = \text{Var}(\bar{X}_m) - \text{Var}(X_i^{(k)}) = 2^{k+1} \sum_{t=1}^{2^{-k}m-1} \left(1 - \frac{2^k t}{m}\right) \frac{(\gamma_k)_t}{m}$$

This is not a new result. The proof is however contained in the appendix for completeness.

*Proof.* Define the mean  $\bar{X}_n = (1/n) \sum_{k=1}^n X_n$ . Let's compute the variance of  $\bar{X}_n$ . Write

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \Sigma_{ij}, \quad (20)$$

where  $\Sigma_{ij} \equiv \text{Cov}(X_i, X_j)$  are the elements of the covariance matrix  $\Sigma$ . In particular, that means that  $\Sigma_{ii} = \text{Cov}(X_i, X_i) = \text{Var}(X_i) = \sigma^2$  (\*) and  $\Sigma_{ij} = \text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i) = \Sigma_{ji}$  ( $\dagger$ ). Then hence by first splitting the summation at the right hand side of equation (20) over the diagonal of  $\Sigma$  and triangular parts of  $\Sigma$  separately, we obtain:

$$\begin{aligned} \text{Var}(\bar{X}_n) &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \Sigma_{ij} = \frac{1}{n^2} \left( \sum_{i=1}^n \underbrace{\Sigma_{ii}}_{=\sigma^2 \text{ (*)}} + \sum_{i=1}^n \sum_{j=i+1}^n \Sigma_{ij} + \sum_{i=j+1}^n \sum_{j=1}^n \underbrace{\Sigma_{ij}}_{=\Sigma_{ji} \text{ (\dagger)}} \right) \\ &= \frac{1}{n^2} \left( \underbrace{\sum_{i=1}^n \sigma^2}_{=n\sigma^2} + \sum_{i=1}^n \sum_{j=i+1}^n \Sigma_{ij} + \underbrace{\sum_{i=1}^n \sum_{j=i+1}^n \Sigma_{ji}}_{\sum_{j=i+1}^n \sum_{i=1}^n \Sigma_{ij}} \right) = \frac{1}{n^2} \left( n\sigma^2 + 2 \sum_{i=1}^n \sum_{j=i+1}^n \Sigma_{ij} \right) \quad (21) \end{aligned}$$

Since the observations are identically distributed the covariances satisfy  $\Sigma_{ij} = \Sigma_{(i+k)(j+k)}$  for all  $i, j, k \in \mathbb{N}$  (check this). That means that along diagonals in  $\Sigma$ , the components of  $\Sigma$  are constant.

$$\text{Var}(\bar{X}_n) \stackrel{(21)}{=} \frac{1}{n} \left( \sigma^2 + 2 \sum_{i=1}^n \sum_{j=i+1}^n \frac{1}{n} \Sigma_{ij} \right) = \frac{1}{n} \left( \sigma^2 + 2 \sum_{t=1}^{n-1} \sum_{j=t+1}^n \frac{1}{n} \Sigma_{tj} \right)$$

Let's rewrite the last term. As you've seen, the sum is over the upper triangular (without the diagonal) elements of  $\Sigma$ , and arrive at the desired expression. ■

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

$$(X, Y) = \langle XY \rangle. \quad (22)$$

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

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

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

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  $\langle \cdot, \cdot \rangle$  is an inner product on  $\Omega$  (ref mcdonald and weiss). In particular, that means we can use Cauchy-Schwarz (ref). Now use the definition of covariance (\*) and apply Cauchy-Schwarz inequality (\*\*). Write

$$\begin{aligned} |\text{Cov}(X, Y)|^2 &\stackrel{(*)}{=} | \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle |^2 \quad = 0 \stackrel{(22)}{=} |(X - \langle X \rangle, Y - \langle Y \rangle)|^2 \\ &\stackrel{(**)}{\leq} (X - \langle X \rangle, X - \langle X \rangle)(Y - \langle Y \rangle, Y - \langle Y \rangle) \stackrel{(*)}{=} \text{Cov}(X, X) \text{Cov}(Y, Y) \end{aligned}$$

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] 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.
- [3] Michiel Hazewinkel. *Gradient Method*. In: *Encyclopedia of Mathematics*. 1st ed. Vol. 4. 13 vols. The Netherlands: Kluwer, 1989.
- [4] William H. Press et al. *Numerical Recipes*. 3rd ed. Cambridge: Cambridge university press, 2007.