-Project 1-
A Computational Study of the Ground-State Energy for Bosonic N-Particle Systems

Yngve Mardal Moe

Norges Miljø- og Biovitenskapelige Universitet

Fabian Maximilian Faulstich

Universitetet i Oslo

Teacher: Prof. Dr. Morten Hjorth-Jensen

Abstract

Here comes the abstract...

Contents

List of Figures

List of Tables

1 Introduction

It was more than 80 years ago that Paul Dirac made his famous statement:

The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are [...] completely known [1].

However, the development of *approximate practical methods of applying quantum mechanics*, demanded by Dirac in the same breath, is still a highly active field of research intersecting physics, chemistry, applied mathematics and computer science. In order to obtain results that can be utilized in practice, an extremely high accuracy of such practical methods is needed. On the other hand, solving the governing equation of quantum mechanics, the molecular Schrödinger equation, is an exceedingly high-dimensional and thus computationally demanding problem. Due to this very unfortunate combination of demanded high accuracy on the one side and the computational bad scaling on the other, small to medium-sized quantum chemical problems are still push the limits of commonly available computational resources.

In this project we focus on N-particle bosonic systems, i.e., particles that follow *Bose–Einstein statistics*. We present a computational study of the system's ground-state energy using statistical methods.

This work is outlined as follows: We start by introducing the considered bosonic system in a more detailed, and for our approaches suitable, way in Section 2. A for this project appropriate detailed and mathematical rigorous description of the computational schemes is presented in Section 3.

2 The Hard Sphere Bose Gas' Quantum Mechanical Model

[...] Thus, improved many-body methods like Monte Carlo calculations may be needed. The aim of this project is to use the Variational Monte Carlo (VMC) method and evaluate the ground state energy of a trapped, hard sphere Bose gas for different numbers of particles with a specific trial wave function. This trial wave function is used to study the sensitivity of condensate and non-condensate properties to the hard sphere radius and the number of particles.

2.1 The Trapped Hard Sphere Bose Gas

In this work we study the ground state energy of a trapped, hard sphere Bose gas [BOSE GAS...] The trap we use is a spherical (S) or an elliptical (E) harmonic trap in one, two and finally three dimensions, with the latter given by

$$V_{ext}(r) = \begin{cases} \frac{1}{2}m\omega_{ho}^2 r^2 & (S)\ , \\ \frac{1}{2}m[\omega_{ho}^2(x^2+y^2)+\omega_z^2 z^2] & (E)\ . \end{cases} \tag{2.1}$$

The considered two-body Hamiltonian is formally defined by

$$H\psi(r) = \sum_{i}^{N} \left(\frac{-\hbar^2}{2m} \nabla_i^2 + V_{ext}(r_i) + \sum_{i<j}^{N} V_{int}(r_i, r_j) \right) \psi(r) \ , \tag{2.2}$$

where ω_{ho}^2 defines the trap potential strength and $r = (r_1, ..., r_N)$. In the case of the elliptical trap, ω_{ho} is the trap frequency in the *perpendicular plane*, i.e., the x-y-plane, and ω_z is the frequency in the z-direction. We represent the boson's interaction by a pairwise, repulsive potential

$$V_{int}(|r_i - r_j|) = \begin{cases} \infty & |r_i - r_j| \leq a \ , \\ 0 & |r_i - r_j| > a \ , \end{cases} \tag{2.3}$$

where a is the so-called hard-core diameter of the bosons. This potential describes non-interacting bosons modeled as hard spheres of radius $a/2$ described as point mass at position r_i, i.e., the distance of two bosons positioned at r_i and r_j can not be smaller than a. Consequently, $V_{int}(|r_i - r_j|)$ is zero if the bosons are separated by a distance $|r_i - r_j|$ greater than a but infinite if they attempt to come within a distance $|r_i - r_j| \leq a$, i.e., their volumes overlap.

2.2 The Trial Function

The trial wave function is used to study the sensitivity of condensate and non-condensate properties to the hard sphere radius and the number of particles. For the ground state with N atoms it s given by

$$\psi_T(r) = \psi_T(r_1, r_2, \ldots r_N, \alpha, \beta) = \prod_i g(\alpha, \beta, r_i) \prod_{i<j} f(a, |r_i - r_j|) \ , \tag{2.4}$$

where α and β are variational parameters. The single-particle wave function g is proportional to the harmonic oscillator function for the ground state, i.e.,

$$g(\alpha, \beta, r_i) = \exp(-\alpha(x_i^2 + y_i^2 + \beta z_i^2)) \ . \tag{2.5}$$

The mean square vibrational amplitude of a single boson at $T = 0K$ in the trap (2.1) is $\langle x^2 \rangle = \hbar/2m\omega_{ho}$ so that $a_{ho} \equiv (\hbar/m\omega_{ho})^{\frac{1}{2}}$ defines the characteristic length of the trap. For spherical traps we set $\beta = 1$ and for non-interacting bosons we have $\alpha = 1/2a_{ho}^2$, setting $a = 0$. The correlation wave function is given by

$$f(a, |r_i - r_j|) = \begin{cases} 0 & \text{, for } |r_i - r_j| \leq a \ , \\ (1 - \frac{a}{|r_i - r_j|}) & \text{, for } |r_i - r_j| > a \ . \end{cases} \tag{2.6}$$

3 Variational Monte Carlo Method and Statistical Analysis

One of the main challenges in computational physics lies in the high dimensionality of the considered problems. To illustrate this we consider a single water molecule reduced to the electronic problem by the Born–Oppenheimer approximation [...]. The spacial discretization of this three dimensional ten electron problem is in $\mathcal{O}(n^{30})$, where n describes the discretization points. Hence, for $n = 10^3$ this simple toy model already exceeds the number of particles in the universe [...]. This exponential growth with respect to the dimensionality is also known as *curse of dimensionality*. Subsequently, we introduce computational methods breaking this fatal growth behavior.

3.1 Variational Monte Carlo Method

The main object of study will be the expectation value of the Hamilton operator, i.e.,

$$\langle H \rangle_\psi = \frac{\int_\Omega \psi^*(\tau) H \psi(\tau)\, d\tau}{\|\psi\|} \ .$$

We emphasize that in most cases Ω will be high dimensional, e.g., for nuclear problems $\Omega = \mathbb{R}^{3N} \times \{\pm 1/2\}^N \times \{\pm 1/2\}^N$ describing the spacial-, spin- and isospin-coordinates of the nucleons. This high dimensionality prohibits a use of conventional numerical integration schemes like the Gaussian quadrature. There are various alternative many-body methods trying to break the curse of dimensionality, however, their success depends strongly on the system's physical properties.

The method chosen for this work is the variational Monte-Carlo method (vmc). Before going into further detail, we notice that the system's energy is defined by the Rayleigh–Ritz variational principle [...], i.e.,

$$E_0 = \min_{\|\psi\|=1} \langle \psi, H\psi \rangle = \langle H \rangle_{\psi_0} \ .$$

Hence, by choosing any trial wave function ψ_T with $\|\psi_T\| = 1$ we get $E_0 \leq \langle H \rangle_{\psi_T}$. Using special properties of the Hamilton operator[1] we can expand any trial function in terms of eigenstates, i.e.,

$$\psi_T = \sum_i a_i \psi_i \ ,$$

with $H\psi_i = E_i \psi_i$. Without loss of generality we assume $\|\psi\| = 1$ which implies

$$E_0 \leq \frac{\sum_i |a_i|^2 E_i}{\sum_i |a_i|^2} \ .$$

We see that by varying the trial function's parameter a_i, the Rayleigh –Ritz variational principle yields the ground state solution. Note, that in most cases the wave function has

[1] Careful, linear self-adjoint and bounded operators do not necessarily have an eigen-basis. However, in finite dimensional Hilbert spaces as in computations this is the case.

small extension coefficients in large parts of the configuration space, i.e., a straightforward computation using homogeneously distributed points in the configuration space will most likely not be efficient. Consequently, we will use a *importance sampling* combined with the *Metropolis* algorithm.

Monte-Carlo methods are based on the *strong law of large numbers*:

Etemadi, 1981 [...]: *Let $(X_n)_{n\in\mathbb{N}} \subseteq L^1(\Omega, \mathcal{A}, \mathbb{P})$ be a sequence of pairwise independent, identically distributed random variables. Then*

$$\frac{1}{N}\sum_{n=1}^{N} X_n \xrightarrow{a.s} \mathbb{E}(X_1) \ ,$$

as $n \to \infty$.

Remark 1. *In the above theorem, we used the standard notation for probability theory. The function space $L^1(\Omega, \mathcal{A}, \mathbb{P})$ is the quotient space with respect to the kernel of $\|\cdot\|_{L^1}$ of reel-valued $\mathbb{P}$-integrable functions $f : \Omega \to \mathbb{R}$ with the underlying sigma-algebra $\mathcal{A}$. Almost sure convergence or strong convergence, denoted by $\xrightarrow{a.s}$, means that*

$$\mathbb{P}\left(\lim_{N\to\infty} X_n - \mathbb{E}(X_1)\right) = 0 \ ,$$

i.e., the set on which $\lim_{N\to\infty} X_n$ differs from $\mathbb{E}(X_1)$ is of measure zero.
The above cited version is a stronger version of the strong law of large numbers proven by Kolmogorov in 1930 [...] as it only presumes pairwise independence of the random variables.

For a chosen trial function $\psi_T(R;a) = \sum_i a_i\psi_i(R)$ we define the probability density function

$$P(R;a) = \frac{|\psi_T(R;a)|^2}{\|\psi_T(\cdot\,;a)\|}$$

and observe that

$$\begin{aligned}\langle H\rangle_{\psi_T(\cdot\,;a)} &= \int_\Omega \frac{\psi_T(R;a)^* H\psi_T(R;a)}{\|\psi_T(\cdot\,;a)\|}\,dR = \int_\Omega \frac{|\psi_T(R;a)|^2}{\|\psi_T(\cdot\,;a)\|}\frac{H\psi_T(R;a)}{\psi_T(R;a)}\,dR \\ &= \int_\Omega P(R;a)E_L(R;a)\,dR \ ,\end{aligned}$$

where $E_L(R;a) = H\psi_T(R;a)/\psi_T(R;a)$ denotes the local energy. We see that the last term corresponds to an integration with respect to the probability measure $\mathbb{P}$ induced by the density P. By **Etemadi** this yields

$$\frac{1}{N}\sum_{n=1}^{N} P(R_i;a)E_L(R_i;a) \xrightarrow{a.s} \langle H\rangle_{\psi_T(\cdot\,;a)} \ .$$

We emphasize, that in the setting of Monte–Carlo methods, the number of pairwise identically distributed, independent variables N corresponds to the number of Monte–Carlo samples.

GO INTO DETAILS, HOW OUR COMPUTATIONS WORK!!!
MENTION THE VARIANCE = 0

3.1.1 Metropolis Sampling

We start with a so-called brute-force sampling using the *metropolis sampling*, i.e., sampling based on the Metropolis–Hastings algorithm [...]. As workhorse of Markov-Chain Monte–Carlos methods, the Metropolis–Hastings algorithm (also called Metropolis algorithm) is based on the theory of Markov chains. Given the unknown ground-state solution density $p = |\psi_0|^2/\|\psi_0\|$ (also called *the target* of the Metropolis algorithm), which is defined on the sample space Ω, and computable up to a multiplying constant (due to the Rayleigh–Ritz principle $p \propto \tilde{p}$). The Metropolis–Hastings algorithm then proposes a generic way to construct a Markov chain on Ω that is *ergodic* and *stationary* with respect to p and therefore converges by the ergodic theorem in distribution to p [...]. Due to its simplicity and versatility, the Metropolis–Hastings algorithm is the workhorse of Markov-Chain Monte–Carlos methods. This makes it the first solution to consider for intractable situations after homogeneously distributed sampling algorithms (cf. Robert and Casella, 2009). The main motivation for using Markov chains is that they provide shortcuts in cases where generic sampling requires too much effort from the experimenter. Rather than aiming at the big picture, like homogeneously distributed sampling algorithm, Markov chains construct a progressive picture of the target distribution, proceeding by local exploration of the state space Ω until all the regions of interest have been uncovered.
We start this section by briefly introducing the necessary concept of Markov-chains.....

Markov chains are one of the most important classes of stochastic processes, as they appear in various time-evolution models. In the following, we restrict the introduction to discrete time-homogeneous Markov chains on a finite dimensional sample set. A sequence of Ω'-valued random variable $(X_n)_{n\in\mathbb{N}}$ is said to fulfill the Markov property if for all $k \in \mathbb{N}$ and all $\omega_1, \omega_2, ..., \omega_{k+1} \in \Omega'$ the following holds:

$$\mathbb{P}(X_{k+1} = \omega_{k+1} | X_k = \omega_k, X_{k-1} = \omega_{k-1}, ..., X_1 = \omega_1) = \mathbb{P}(X_{k+1} = \omega_{k+1} | X_k = \omega_k) \ ,$$

where $\mathbb{P}(\cdot|\cdot)$ describes the conditional probability. Hence, if a stochastic process fulfills the Markov property, the $k+1$-step is only depending on the k step, i.e., the Markov chain has no memory. (This is the major and most important difference between Markov chains and Martingales – another class of important stochastic processes). As we only consider Markov chains on a finite dimensional sample set, we use the class of stochastic

matrices $\mathcal{S}$ to describe a Markov process, i.e.,

$$\mathcal{S} = \left\{ S \in \mathbb{R}^{m\times m} | S_{i,j} \in [0,1] \, \forall i,j \in \{1,...,m\}, \sum_{i=j}^{m} S_{i,j} = 1 \, \forall i \in \{1,...,m\} \right\} .$$

We then define a Markov chain as follows:
Let $S \in \mathcal{S}$ be a stochastic matrix. A sequence of Ω'-valued random variable $(X_n)_{n\in\mathbb{N}}$ is called a time-homogeneous Markov chain with transition matrix S is for all $k \in \mathbb{N}$ and all $\omega_1, \omega_2, ..., \omega_{k+1} \in \Omega'$ with $\mathbb{P}(X_k = \omega_k, X_{k-1} = \omega_{k-1}, ..., X_1 = \omega_1) > 0$ the following holds:

$$\mathbb{P}(X_{k+1} = \omega_{k+1} | X_k = \omega_k, X_{k-1} = \omega_{k-1}, ..., X_1 = \omega_1) = S_{\omega_k, \omega_k+1} .$$

The matrix entries $S_{i,j}$ are called transition probabilities and the initial distribution μ is defined by $\mu(\omega) = \mathbb{P}(X_1 = \omega)$ for $\omega \in \Omega'$. An important feature of Markov chain a stationary distribution, i.e., a distribution ν that stay invariant under the Markov chain and therewith fulfilling the matrix equation $\nu S = \nu$. The existence and uniqueness of such stationary distributions depends strongly on the Markov chain. The for the Metropolis algorithm central theorem, is the ergodic theorem time-homogeneous Markov chains on a finite dimensional sample space. It states that for an *aperiodic* and *irreducible* Markov chain with transition matrix S, $\nu S^n \xrightarrow{\mathbb{P}} \mu$ for $n \to \infty$ and any initial distribution ν, with μ being the stationary distribution. We again used the standard notation from probability theory, where $\nu S^n \xrightarrow{\mathbb{P}} \mu$ denotes the convergence in probability, i.e., for all $\varepsilon > 0$ we have $\lim_{n\to\infty} \mathbb{P}(|\nu S^n - \mu| \geq \varepsilon) = 0$. Note that demanding the Markov chain with transition matrix S to be *aperiodic* and *irreducible* is the same as demanding the existence of $n \in \mathbb{N}$ such that $S^n_{i,j} > 0$ for all i, j, i.e., the Markov chain is ergodic.

The aim is now to construct an ergodic Markov chain that converges to the ground-state solution density p. Indeed, the Markov chain $(X_n)_{n\in\mathbb{N}}$ returned by the Metropolis-Hastings method is such that it converges to p [...]. This means that the chain can be considered as a sample, albeit a dependent sample, and approximately distributed from p. Due to the Markovian nature of the simulation, the first values are highly dependent on the initial distribution and therewith from X_1. Consequently, these are removed from the sample as warm-up, also called burn-in. While there are very few settings where the time when the chain reaches stationarity can be determined [...], there is no need to look for such an instant since the empirical average

$$\frac{1}{N} \sum_{n=1}^{N} P(X_i; a) E_L(X_i; a)$$

converges almost surely if the Markov chain is ergodic. Comparing this to (3.1) it implies that, in theory, simulating a Markov chain is intrinsically equivalent to the pairwise independent identically distributed simulation from the target. Note, that the difference

being in a loss of efficiency, i.e., it is necessary to simulate more terms to achieve a given variance for the above Monte Carlo estimator.

The Metropolis-Hastings algorithm associated with the target density p requires the choice of a conditional density q also called *proposal* or *candidate kernel*. The transition $t \to t+1$ is then given by Algorithm 1.

Algorithm 1 Metropolis–Hasting

1: **procedure** CREATE METROPOLIS-HASTING MARKOV CHAIN
2: Given $X_t = x_t$
3: Generate $Y_t \sim q(y|x_t)$
4: Take

$$X_{t+1} = \begin{cases} Y_t, \text{ with probability } \rho(x_t, Y_t), \\ x_t, \text{ with probability } 1 - \rho(x_t, Y_t), \end{cases}$$

where

$$\rho(x, y) = \min\left\{1, \frac{\tilde{p}(y)q(x|y)}{\tilde{p}(x)q(y|x)}\right\}.$$

Then, as shown in Metropolis et al. (1953), this transition preserves the stationary density p if the chain is irreducible, i.e., q has a wide enough support to eventually reach any region of the state space Ω' with positive mass under p. As mentioned before, a sufficient condition is that q is positive everywhere. The very nature of accept-reject step introduced by Metropolis et al. is therefore sufficient to turn a simulation from an almost arbitrary proposal density q into a generation that preserves p as the stationary distribution. This appears somewhat surprising but works indeed, as shown in theoretical description above. In practice, the performances of the algorithm are highly dependent on the choice of the transition q, since some choices see the chain unable to converge in a manageable time. In this work we use X_t perturbed by a normal distribution, i.e., $Y_t \sim X_t + \varepsilon_t$ with $\varepsilon_t \sim \mathcal{N}(0, 1)$.

3.1.2 Importance Sampling

Despite the Metropolis–Hasting sampling's simplicity, we replace this brute force algorithm with a walk in coordinate space biased by the trial wave function. This *importance sampling* is based on a stochastic differential equation, namely the Langevin equation

$$\frac{d^2x}{dt^2} = -\lambda\frac{dx}{dt} + \eta\,(t)\,,$$

with the position x a particle of mass one. The noise term $\eta\,(t)$ has a Gaussian probability distribution, where the correlation function depends on the physical problem. For this work, however, we are not interested in a solution of a Langevin equation for a

particular realization of the fluctuating force. Rather, we are interested in correlation functions of the slow variables after averaging over the fluctuating force. As shown in [...], such correlation functions can be described by the Fokker–Planck equation, a deterministic differential equation. Unlike Langevin equations, a Fokker–Planck equation describes a time dependent probability density of a random variable and not the random variable itself. We shall see that this sampling is not as far from the previously described Metropolis–Hasting sampling, which was derived from the theory of discrete time-homogeneous Markov chains on a finite dimensional sample set.

Subsequently, we present a quick and dirty derivation of the Fokker–Planck equation, by shortly dropping the mathematical rigor of this work. We now consider time-continuous Markov chains, i.e.,

The stochastic process $(X_t)_{t\geq 0}$ is called a time continuous Markov chain, if for all $0 \leq s_0 < s_1 < ... < s_n < s$ and all possible events $\omega_0, \omega_1, ...\omega_n, \omega_i, \omega_j \in \Omega'$ the following holds

$$\mathbb{P}(X_{t+s} = \omega_j | X_s = \omega_i, X_{s_n} = \omega_n, ..., X_{s_0} = \omega_0) = \mathbb{P}(X_t = \omega_j | X_0 = \omega_i)$$

The Markov property and the definition of the conditional probability we derive straightforwardly the Chapman-Kolmogorow equality. From the Markov property yields

$$\begin{aligned}&\mathbb{P}(X_{t_3} = \omega_3, X_{t_2} = \omega_2, X_{t_1} = \omega_1) \\ &\quad = \mathbb{P}(X_{t_1} = \omega_1)\mathbb{P}(X_{t_2} = \omega_2 | X_{t_1} = \omega_1)\mathbb{P}(X_{t_3} = \omega_3 | X_{t_2} = \omega_2)\end{aligned}$$

which is equivalent to

$$\mathbb{P}(X_{t_3} = \omega_3, X_{t_1} = \omega_1) = \mathbb{P}(X_{t_1} = \omega_1) \int_{\Omega'} \mathbb{P}(X_{t_2} = \omega_2 | X_{t_1} = \omega_1)\mathbb{P}(X_{t_3} = \omega_3 | X_{t_2} = \omega_2) d\omega_2 \ .$$

Division by $\mathbb{P}(X_{t_1} = \omega_1)$ yields the Chapman-Kolmogorow equality

$$\mathbb{P}(X_{t_3} = \omega_3 | X_{t_1} = \omega_1) = \int_{\Omega'} \mathbb{P}(X_{t_2} = \omega_2 | X_{t_1} = \omega_1)\mathbb{P}(X_{t_3} = \omega_3 | X_{t_2} = \omega_2) d\omega_2$$

and considering the times $t' \leq t' + \tau \leq t$ this reads

$$\mathbb{P}(X_t = \omega | X_{t'} = \omega') = \int_{\Omega'} \mathbb{P}(X_t = \omega | X_{t'+\tau} = \omega'')\mathbb{P}(X_{t'+\tau} = \omega'' | X_{t'} = \omega') d\omega'' \ . \quad (3.1)$$

By construction of the delta distribution, we further have

$$\mathbb{P}(X_{t'+\tau} = \omega'' | X_{t'} = \omega') = \int_{\Omega'} \mathbb{P}(X_{t'+\tau} = y | X_{t'} = \omega')\delta(y - \omega'') dy \ . \quad (3.2)$$

Inserting the Taylor expansion of the delta distribution, i.e.,

$$\delta(y - \omega'') = \sum_{n=0}^{\infty} \frac{(y - \omega')^n}{n!} \left(\frac{\partial}{\partial \omega'}\right)^n \delta(\omega' - \omega'')$$

into Eq. (3.2), we obtain

$$\mathbb{P}(X_{t'+\tau} = \omega''|X_{t'} = \omega') = \left(1 + \sum_{n=0}^{\infty} \frac{1}{n!} \underbrace{\int_{\Omega'} (y - x')^n \mathbb{P}(X_{t'+\tau} = y|X'_t = \omega')dy}_{=M_n(\omega',t',\tau)} \left(\frac{\partial}{\partial x'} \right)^n \right) \delta(\omega' - \omega'') \ .$$

Inserting this into Eq. (3.1) yields

$$\mathbb{P}(X_t = \omega|X_{t'} = \omega') = \mathbb{P}(X_t = \omega|X_{t'+\tau} = \omega') + \sum_{n=0}^{\infty} \frac{1}{n!} M_n(\omega', t', \tau) \left(\frac{\partial}{\partial \omega'} \right)^n \mathbb{P}(X_t = \omega|X_{t'+\tau} = \omega') \ .$$

As mentioned above, the Fokker-Planck equation is deterministic differential equation describing the probability density of a random variable. To meet that end, we take a closer look a differential quotient

$$\frac{\partial \mathbb{P}(X_t = \omega|X_{t'} = \omega')}{\partial t'} = -\lim_{\tau \to 0} \frac{\mathbb{P}(X_t = \omega|X_{t'} = \omega') - \mathbb{P}(X_t = \omega|X_{t'+\tau} = \omega')}{\tau} \ ,$$

together with Eq. (3.1.2) this can be written as

$$\frac{\partial \mathbb{P}(X_t = \omega|X_{t'} = \omega')}{\partial t'} = -\lim_{\tau \to 0} \frac{1}{\tau} \sum_{n=0}^{\infty} \frac{1}{n!} M_n(\omega', t', \tau) \left(\frac{\partial}{\partial \omega'} \right)^n \mathbb{P}(X_t = \omega|X_{t'+\tau} = \omega') \ .$$

Using the Taylor expansion of $M_n(\omega', t', \tau)$, i.e.,

$$\frac{M_n(\omega', t', \tau)}{n!} = 0 + D^{(n)}(\omega', t')\tau + \mathcal{O}(\tau^2), \quad \text{with } D^{(n)}(\omega', t') = \lim_{\tau' \to 0} \mathbb{E}((X_{t'+\tau'} - X_{t'})^n)\big|_{X_{t'}=\omega'}$$

yields the differential equation

$$\frac{\partial \mathbb{P}(X_t = \omega|X_{t'} = \omega')}{\partial t'} = -\mathcal{L}^{\dagger}_{KM}(\omega', t')\mathbb{P}(X_t = \omega|X_{t'} = \omega') \tag{3.3}$$

with the Kramers–Moyal operator

$$\mathcal{L}_{KM}(\omega', t') = \sum_{n=0}^{\infty} \left(-\frac{\partial}{\partial \omega'} \right)^n D^{(n)}(\omega', t') \ ,$$
$$\mathcal{L}^{\dagger}_{KM}(\omega', t') = \sum_{n=0}^{\infty} D^{(n)}(\omega', t') \left(\frac{\partial}{\partial \omega'} \right)^n \ .$$

Similar to Eq. (3.3) with find the analog deterministic differential equation

$$\frac{\partial P(\omega, t)}{\partial t} = \mathcal{L}_{KM}(\omega, t)P(\omega, t) \tag{3.4}$$

describing the time evolution of the probability density $P(\omega, t)$. The mayor question arising at this point is under which circumstances this expansion terminates. The answer is given by Pawulas theorem [...]:

For positive transition probabilities $\mathbb{P}(\omega, t|\omega', t')$ *the expansion in Eq.* (3.4) *terminates after one or two terms. If it does not terminate after the second term it never terminates.*

If the expansion in Eq. (3.4) terminates after the second term, the resulting differential equation is called a Fokker-Planck equation

$$\frac{\partial P(\omega,t)}{\partial t} = -\frac{\partial}{\partial \omega} D^{(1)}(\omega,t)P(\omega,t) + \frac{\partial^2}{\partial \omega^2} D^{(2)}(\omega,t)P(\omega,t) = \mathcal{L}_{FP}P(\omega,t) \ ,$$

with the Fokker-Planck operator

$$\mathcal{L}_{FP} = -\frac{\partial}{\partial \omega} D^{(1)}(\omega,t) + \frac{\partial^2}{\partial (\omega)^2} D^{(2)}(\omega,t) \ .$$

The term $D^{(1)}$ is called *drift coefficient* and is responsible for the time-evolution of the random variables mean value. The term $D^{(2)}$ is called *diffusion coefficient* and is responsible for the time-evolution of the random variables standard deviation. In this work we consider a constant diffusion coefficient, in atomic unties, $D^{(2)} = 1/2$. Consequently, the convergence to a stationary probability density, i.e., $\partial P(\omega, t)/\partial t = 0$, then corresponds to the solution of

$$\frac{\partial^2}{\partial (\omega)^2} P(\omega,t) = D^{(1)}(\omega,t)\frac{\partial}{\partial \omega} P(\omega,t) + P(\omega,t)\frac{\partial}{\partial \omega} D^{(1)}(\omega,t) \ .$$

We assume the drift coefficient to be of the form $D^{(1)}(\omega, t) = g(\omega, t)\partial P(\omega, t)/\partial \omega$. Together with Random–Nikodym theorem this yields

$$\frac{\partial^2 P(\omega,t)}{\partial (\omega)^2} = g(\omega,t)\left(\frac{\partial P(\omega,t)}{\partial \omega}\right)^2 + P(\omega,t)g(\omega,t)\frac{\partial^2 P(\omega,t)}{\partial \omega^2} + P(\omega,t)\frac{\partial g(\omega,t)}{\partial P(\omega,t)}\left(\frac{\partial P(\omega,t)}{\partial \omega}\right)^2 \ .$$

For $g(\omega, t) = 1/P(\omega, t)$ this equality is fulfilled. Indeed,

$$\frac{\partial^2 P(\omega,t)}{\partial (\omega)^2} = \frac{1}{P(\omega,t)}\left(\frac{\partial P(\omega,t)}{\partial \omega}\right)^2 + \frac{\partial^2 P(\omega,t)}{\partial \omega^2} + P(\omega,t)\left(-\frac{1}{P^2(\omega,t)}\right)\left(\frac{\partial P(\omega,t)}{\partial \omega}\right)^2 \ .$$

As in this work $P = |\psi_T|^2$ this yields

$$D^{(1)}(\omega,t) = \frac{1}{P}\frac{\partial P}{\partial \omega} = 2\frac{1}{\psi_T}\nabla\psi_T = F \ ,$$

which is also known as *quantum force*. As the drift coefficient describes the movement of the mean value, it is this term that forces the taken samples, also called walkers, into regions of the configuration space where the trial wave function ψ_t is large. This increases the simulation's efficiency in contrast to the here considered Metropolis–Hastings sampling where the samples are normally distributed, i.e, they have the same probability of

moving in every direction. Using the Green's functions theory, we obtain a solution of above described Fokker-Planck equation, implying the following description of transition probabilities

$$G(x, y, \Delta t) = \frac{1}{(2\pi\Delta t)} \exp\left(-\frac{(y - x - \frac{\Delta t F(x)}{2})}{2\Delta t} \right) ,$$

where we used $D^{(2)} = 1/2$.

Using the Langevin equation we can describe the random variable of the transition probability using an explicit Euler procedure with time step size Δt, i.e.,

$$Y_t = X_t + D^{(2)} F(X_t)\Delta t + \xi\sqrt{\Delta t} .$$

Using this random variable and the transition probability described by the Fokker-Planck's Greens function in the Metropolis–Hastings algorithm the Algorithm 2.

Algorithm 2 Importance Sampling

1: **procedure** Create Importance Sampling Markov Chain
2: Given $X_t = x_t$
3: Generate $Y_t \sim X_t + D^{(2)} F(X_t)\Delta t + \xi\sqrt{\Delta t}$
4: Take

$$X_{t+1} = \begin{cases} Y_t, \text{ with probability } \rho(x_t, Y_t), \\ x_t, \text{ with probability } 1 - \rho(x_t, Y_t), \end{cases}$$

where

$$\rho(x, y) = \min\left\{ 1, \frac{G(x, y, \Delta t)|\psi_T(y)|^2}{G(y, x, \Delta t)|\psi_T(x)|^2} \right\} .$$

3.2 Statistical Analysis

In this section, we describe the statistical tools that will be used for a proper evaluation of our data's statistical errors. The computational results are obtained by using Monte Carlo simulations. These simulations can be treated as computational experiments and its results can therefore be analyzed with the same statistical tools used for laboratory experiments. We hereby focus on two *resampling method*, i.e. methods that estimate the precision of sample statistics (medians, variances, percentiles), namely:

- The *bootstrap technique* also called *bootstrapping*, where the estimation is based on drawing randomly with replacement from a set of data points.
- The *blocking method* also just *blocking*, where the samples are divided in subsamples of certain size, allowing to use estimators for uncorrelated samples.

The estimators used in this work are a *mean estimator* defined by

$$\hat{\Theta}(X) = \hat{\Theta}(X_1, ..., X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

and a *variance estimator* defined by

$$\hat{\varphi}(X) = \hat{\varphi}(X_1, ..., X_n) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = S^2X \ ,$$

where $X = (X_1, ..., X_n)$ describes a set of samples.

3.2.1 Bootstrapping

In the wider sense of statistical analysis, any test that relies on random sampling with replacement is called bootstrapping. This technique allows the estimation of the sampling distribution and provides an error analysis assigning measures of accuracy defined in terms of bias, variance, confidence intervals, prediction error. The basic idea of bootstrapping is treating inference of the true probability distribution J, given the original data, as being analogous to inference of the empirical distribution of $\hat{J}$, given the resampled data. The accuracy of inferences regarding $\hat{J}$ using the resampled data is assessable because $\hat{J}$ is known. Assuming that $\hat{J}$ is a 'reasonable approximation' to the unknown distribution J, then the quality of inference on J can in turn be inferred.

In a more informal way bootstrapping can be described as follows:
Assume we are interested in the mean value of $X_1, ..., X_K$ independently and identically J-distributed with K being very large. If K becomes too large, we can have to restrict the evaluation to a subset of size n. Without loss of generality we denote this subset by $X_1, ..., X_n$ with $n \ll K$. From this sample the mean estimator yields a single estimate $\overline{X} = \hat{\Theta}(X_1, ..., X_n)$. In order to reason about $\Theta(X_1, ..., X_K)$, we need a sense of the variability of the mean $\overline{X}$. Bootstrapping now treats the sample $X_1, ..., X_n$ in the same way as $X_1, ..., X_K$. A straightforward bootstrap approach therefore involves sampling from $X_1, ..., X_n$ to form a new sample (called a *resample* or *bootstrap sample*) of the same size. Under the assumption that n is sufficiently large, there is practically zero probability that the sample will be identical to the original real sample. This process is repeated numerous times, and for each of these bootstrap samples the mean is computed, the so-called *bootstrap estimates*. This yields a histogram of bootstrap means providing an estimate of the unknown distribution's mean. How much the mean varies can then be deduced from this histogram and used to measure the quality of the mean computed from the original sample.

The above described method is also called *non-parametric Bootstrap*, and is used when the distribution J is unknown. We summarize this algorithm in the following pseudo-code Alg. (3): In the case the histogram $P(X)$ is known, we resample from this histogram and do the Bootstrap sample steps from Alg. (3). This is then known as *parametric Bootstrap*.

Algorithm 3 Non-parametric Bootstrap

1: **procedure** CREAT BOOTSTRAP SAMPLE HISTOGRAM
2: Draw histogram of $X(P^*(X))$
3: **procedure** BOOTSTRAP SAMPLE
4: Draw from $P^*(X)$, n random numbers $\{X_i\} = X^*$
5: Compute $\hat{\Theta}(X^*) = t(X^*)$
6: write t to memory
7: Use Bootstrap Samples to obtain the final histogram $P^*(\hat{\Theta})$

3.2.2 Blocking

[Something on Blocking, introduce blocks]

Consider experimental data from a Monte Carlo simulation of step size Δt. In case of uncorrelated samples the mean's standard deviation is given by

$$\sigma = \sqrt{\hat{\varphi}(X)} \ . \tag{3.5}$$

The Monte Carlo simulation however, will most likely produce a correlated sample set. In this case the standard deviation needs to be corrected as follows

$$\sigma = \sqrt{\frac{1 + 2\tau/\Delta t}{n-1}\left(\overline{X^2} - \overline{X}^2\right)} \ ,$$

where τ is the so called *correlation time*, i.e., the time between a sample and the next uncorrelated sample. In case $\Delta t \gg \tau$ the Monte Carlo samples will be uncorrelated and Eq. (3.5) can be used to compute the standard deviation. As mentioned before, it is most likely that $\Delta t < \tau$, in which case Eq. (3.5) is not valid anymore. Furthermore, we are facing the difficulty of not knowing the correlation time τ. However, with blocking we are still able to determine a valid standard deviation of the mean $\overline{X}$. Using the uncorrelated expression Eq. (3.5), we compute the mean of each block $\overline{X^{(i)}}$ with $i = 1, ..., n_b$ where n_b is the number of blocks, and calculate then the total mean and variance. The crucial part for this method is the block size, i.e., it must be so large that sample j of block i is not correlated with sample j of block $i+1$. An optimal choice would therefore be the unknown correlation time. A quick and dirty solution to this problem is a manual fix of the block size:

A graphical representation of the standard deviation as a function of the block size will show a plateau region, i.e., a connected set of block sizes for which the standard deviation is constant. As soon as the standard deviation stops increasing, we know that the blocks are uncorrelated. This number can then be manually put into the blocking algorithm.

An alternative to this brute force fix of the blocking method, is the *automatic blocking*. We here proceed as follow... INTRODUCE ALL THE NECESSARY STUFF...

Algorithm 4 Automatic Blocking

procedure ESTIMATE $\mathbb{V}(\overline{X})$
  Set $i = 1$
  Compute $\left(S^{(i)}\right)^2$, $\hat{\gamma}^i(1)$
  Transform $Q^{(i)} \to Q^{(i+1)}$
  if $|Q^{(i)}| \leq 2$ **then**
    Compute M_j using $\hat{\gamma}^j(1)$ and $\left(S^{(j)}\right)^2$
    Find the first k such that $M_k \leq q_k$
    Return $\left(S^{(k)}\right)^2/n_k$
  else
    Go back to 2

4 Implementation

In this section we present the implementation we used to derive next section's results. [SOMETHING]
We start by deriving an analytic expression for the local energy.

4.1 Analytical Solutions

We observe that the considered Hamiltonian contains the Laplace operator, which can be approximated by the method of finite differences. However, we note that the used quantities are the local energy E_l and the quantum force F, which can be described analytically, which saves computational time. We start by deriving the local energy the case with only the harmonic oscillator potential, i.e., $a = 0$ implying $f(a, |r_i - r_j|) = 1$ and $V_{\text{int}}(|r_i - r_j|) = 0$.

4.1.1 Harmonic Oscillator Potential

Following, we start with $\beta = 1$, i.e., $g(\alpha, \beta, r_i) = \exp(-\alpha r_i^2)$ and find the relevant local energies in one, two and three dimensions for one and N particles with the same mass. The trial function of a one dimensional single particle is then given by

$$\psi_T(r_1, \alpha, 1) = \exp(-\alpha r_1^2) \ .$$

We then compute

$$\begin{aligned} H\psi_T(r_1, \alpha, 1) &= \left(\frac{-\hbar^2}{2m}\Delta + \frac{1}{2}m\omega_{ho}^2 r_1^2\right)\exp(-\alpha r_1^2) \\ &= -\frac{\hbar^2}{2m}\left(4\alpha^2 r_1^2 - 2\alpha\right)\exp(-\alpha r_1^2) + V_{\text{ext}}(r_1)\exp(-\alpha r_1^2) \ , \end{aligned}$$

which yields

$$E_L(r_1, \alpha) = -\frac{\hbar^2}{2m}\left(4\alpha^2 r_1^2 - 2\alpha\right) + V_{\text{ext}}(r_1) \ .$$

Note that in atomic unites with $\alpha = 1/2$ the local energy is given by $E_L = 1/2$, which is the ground-state energy of the one dimensional harmonic oscillator. This can be straightforward generalized to the n-dimensional case, where

$$H\psi_T(r_1, \alpha, 1) = -\frac{\hbar^2}{2m}\left(4\alpha^2 r_1^2 - 2\alpha n\right)\exp(-\alpha r_i^2) + V_{\text{ext}}(r_1)\exp(-\alpha r_1^2)$$

and therewith

$$E_L(r_1, \alpha) = -\frac{\hbar^2}{2m}\left(4\alpha^2 r_1^2 - 2\alpha n\right) + V_{\text{ext}}(r_1) \ .$$

For N particles in n dimensions the trial function reads

$$\psi_T(r_1, ..., r_N, \alpha, 1) = \prod_{i=1}^{N} \exp\big(-\alpha r_i^2\big) = \exp\big(-\alpha \sum_{i=1}^{N} r_i^2\big) \ .$$

Similar to the one dimensional case we then compute

$$H\psi_T(r_1, ..., r_N, \alpha, 1) = \left(\sum_{i=1}^{N} -\frac{\hbar^2}{2m}\left(4\alpha^2 r_i^2 - 2\alpha n\right) + V_{\text{ext}}(r_i)\right)\psi_T(r_1, ..., r_N, \alpha, 1)$$

and therewith

$$E_L(r_1, ..., r_N, \alpha) = \sum_{i=1}^{N} -\frac{\hbar^2}{2m}\left(4\alpha^2 r_i^2 - 2\alpha n\right) + V_{\text{ext}}(r_i) \ .$$

Assuming $\beta \neq 1$ only influences the three dimensional case. We compute for one particle

$$H\psi_T(r_1, \alpha, \beta) = -\frac{\hbar^2}{2m}\left(4\alpha^2(x^2 + y^2 + \beta^2 z^2) - 2\alpha(2 + \beta)\right)\psi_T(r_1, \alpha, \beta) + V_{\text{ext}}(r_1)\psi_T(r_1, \alpha, \beta) \ ,$$

with the trial function

$$\psi_T(r_1, \alpha, 1) = \exp(-\alpha(x^2 + y^2 + \beta z^2)) \ .$$

Hence,

$$E_L(r_1, \alpha, \beta) = -\frac{\hbar^2}{2m}\left(4\alpha^2(x^2 + y^2 + \beta^2 z^2) - 2\alpha(2 + \beta)\right) + V_{\text{ext}}(r_1) \ .$$

We conclude for the N particle case for the trial function

$$\psi_T(r_1, ..., r_N, \alpha, \beta) = \exp(-\alpha \sum_{i=1}^{N}(x_i^2 + y_i^2 + \beta z_i^2))$$

that

$$H\psi_T(r_1,...,r_n,\alpha,\beta) = \sum_{i=1}^{N} -\frac{\hbar^2}{2m}\left(4\alpha^2(x_i^2+y_i^2+\beta^2 z_i^2) - 2\alpha(2+\beta)\right)\psi_T(r_1,...,r_n,\alpha,\beta) + V_{\text{ext}}(r_i)\psi_T(r_1,...,r_n,\alpha,\beta) \ ,$$

and therewith

$$E_L(r_1,...,r_n,\alpha,\beta) = \sum_{i=1}^{N} -\frac{\hbar^2}{2m}\left(4\alpha^2(x_i^2+y_i^2+\beta^2 z_i^2) - 2\alpha(2+\beta)\right) + V_{\text{ext}}(r_i) \ .$$

4.1.2 Quantum Force for the Harmonic Oscillator Potential

Another often used quantity is th drift coefficient of the Fokker–Planck equation, i.e.,the quantum force. As we do not want to approximate the gradient of the trial function for every sampling, we derive the expression analytically. For the harmonic oscillator potential with $\beta = 1$. Then the trial function is again given by

$$\psi_T(r_1,\alpha,1) = \exp(-\alpha r_1^2) \ ,$$

for which the quantum force is given by

$$F(r_1,\alpha,1) = \frac{1}{\psi_T(r_1,\alpha,1)}\nabla\psi_T(r_1,\alpha,1) = -2\alpha r_1 \ .$$

As the gradient is a map $\nabla : L^2(\mathbb{R}^{3N}) \to \left(L^2(\mathbb{R}^{3N})\right)^{3N}$ (using its weak formulation), we obtain the following quantum force for the N-particle case

$$F(r_1,...,r_N,\alpha,1) = \frac{1}{\psi_T(r_1,...,r_N,\alpha,1)}\nabla\psi_T(r_1,...,r_N,\alpha,1) = -2\alpha(r_i)_{i=1}^N \ ,$$

where $(r_i)_{i=1}^N = (x_1,y_1,z_1,x_2,....,z_N)^T$.

4.1.3 The Local Energy in the full Potential

Next we consider the full problem in three dimensions. The aim is to find an analytic expression for the derivative of the trial wave function

$$\frac{1}{\Psi_T(r)}\sum_i^N \nabla_i^2\Psi_T(r) \ ,$$

with the trial function given by Eq. (2.4). Rewriting this trial function yields

$$\psi_T(r_1,...,r_N,\alpha,\beta) = \prod_{i=1}^{N} g(\alpha,\beta,r_i)\exp\left(\sum_{i<j} u(r_{ij})\right) \ ,$$

where we have defined $r_{ij} = |r_i - r_j|$ and

$$f(r_{ij}) = \exp\left(\sum_{i<j} u(r_{ij})\right) ,$$

where $u : \mathbb{R} \to \mathbb{R}$ with $x \mapsto \ln\left(1 - \frac{a}{x}\right)$. Using the product rule we find

$$\begin{aligned}\nabla_k \psi_T(r_1, ..., r_N, \alpha, \beta) = (\nabla_k g(\alpha, \beta, r_k)) \prod_{i \neq k; i=1}^{N} g(\alpha, \beta, r_i) \exp\left(\sum_{i<j} u(r_{ij})\right) \\ + \prod_{i=1}^{N} g(\alpha, \beta, r_i) \exp\left(\sum_{i<j} u(r_{ij})\right) \sum_{i<j}^{N} \nabla_k u(r_{ij}) ,\end{aligned} \tag{4.1}$$

where

$$\sum_{i<j}^{N} \nabla_k u(r_{ij}) = \frac{1}{2} \sum_{j=1}^{N} \sum_{i \neq j; i=1}^{N} \nabla_k u(|r_i - r_j|) .$$

Note that we sightly abuse the notation here, namely $u(|r_i - r_j|)$ describes the map $u(|\cdot - r_j|) : \mathbb{R}^3 \to \mathbb{R}$ evaluated in r_i as well as the map $u(|r_i - \cdot|) : \mathbb{R}^3 \to \mathbb{R}$ evaluated in r_j. Consequently holds

$$\nabla_i u(|r_i - r_j|) = u'(|r_i - r_j|) \nabla_i |r_i - r_j| = \frac{a(r_i - r_j)}{(|r_i - r_j| - a)|r_i - r_j|^2} = \nabla_i u(|r_j - r_i|) , \tag{4.2}$$

and therewith

$$\sum_{i<j}^{N} \nabla_k u(r_{ij}) = \sum_{j \neq k; j=1}^{N} \nabla_k u(r_{kj}) .$$

From Eq. (4.1) we can then derive that the quantum force in the full potential is given by

$$F(r_1, ..., r_N, \alpha, \beta) = 2 \left(\left(\frac{\nabla_i g(\alpha, \beta, r_i)}{g(\alpha, \beta, r_i)} \right)_{i=1}^{N} + \left(\sum_{j \neq i} \nabla_i u(|r_i - r_j|) \right)_{i=1}^{N} \right) .$$

Deriving another time, we find

$$\begin{aligned}\Delta_k\psi_T(r_1,...,r_N,\alpha,\beta) &= (\Delta_k g(\alpha,\beta,r_k))\prod_{i\neq k;i=1}^{N} g(\alpha,\beta,r_i)\exp\left(\sum_{i<j}u(r_{ij})\right)\\ &+(\nabla_k g(\alpha,\beta,r_k))\prod_{i\neq k;i=1}^{N} g(\alpha,\beta,r_i)\exp\left(\sum_{i<j}u(r_{ij})\right)\sum_{j\neq k;j=1}^{N}\nabla_k u(r_{kj})\\ &+\prod_{i=1}^{N} g(\alpha,\beta,r_i)\exp\left(\sum_{i<j}u(r_{ij})\right)\left(\sum_{j\neq k;j=1}^{N}\nabla_k u(r_{kj})\right)^2\\ &+\prod_{i=1}^{N} g(\alpha,\beta,r_i)\exp\left(\sum_{i<j}u(r_{ij})\right)\sum_{j\neq k;j=1}^{N}\Delta_k u(r_{kj})\ .\end{aligned} \tag{4.3}$$

As for the local energy this needs to be divided by the trial function, we find the following expression for the second derivative

$$\begin{aligned}\frac{\Delta_k\psi_T(r_1,...,r_N,\alpha,\beta)}{\psi_T(r_1,...,r_N,\alpha,\beta)} &= \frac{\Delta_k g(\alpha,\beta,r_k)}{g(\alpha,\beta,r_k)}+\frac{\nabla_k g(\alpha,\beta,r_k)}{g(\alpha,\beta,r_k)}\sum_{j\neq k;j=1}^{N}\nabla_k u(r_{kj})\\ &+\sum_{i,j\neq k}\nabla_k u(r_{kj})\nabla_k u(r_{kj})+\sum_{j\neq k;j=1}^{N}\Delta_k u(r_{kj})\ ,\end{aligned} \tag{4.4}$$

with

$$\begin{aligned}\Delta_k u(r_{kj}) &= \nabla_k\cdot\frac{r_k-r_j}{|r_k-r_j|}u'(|r_k-r_j|)\\ &= u''(|r_k-r_j|)+u'(|r_k-r_j|)\left(\frac{3}{|r_k-r_j|}-(r_k-r_j)\cdot\frac{(r_k-r_j)}{|r_k-r_j|^3}\right)\\ &= u''(|r_k-r_j|)+u'(|r_k-r_j|)\frac{2}{|r_k-r_j|}\ .\end{aligned}$$

It remains to compute u' and u''. We already computed u' in Eq. (4.2), however, for the sake of completeness we list it again subsequently. We find

$$\begin{aligned}u'(r_{ij}) &= \frac{a}{(r_{i,j}-a)r_{i,j}}\ ,\\ u''(r_{ij}) &= -\frac{a}{((r_{i,j}-a)r_{i,j})^2}(2r_{ij}-a)\ .\end{aligned}$$

This together with the analytic expression for the harmonic oscillator yields an analytic expression or the local energy.

5 Numerical Results

6 Conclusion and Perspectives

References

[1] Paul AM Dirac. Quantum mechanics of many-electron systems. *Proc. R. Soc. Lond. A*, 123(792):714–733, 1929.