Project 2 - An Autoregressive Model for Bosonic Quantum Dots

Erlend Lima, Kristian Wold

I. INTRODUCTION

Problems in many-particle quantum mechanics is notoriously hard, and has over many years resulted in a suit of methods and techniques for solving them efficiently and reliably. A popular family of such methods is Variational Monte-Carlo, which parameterize a trial wave function and seeks to optimize the parameters in order to minimize the expected energy, thus approximating the ground state of the system. While classic VMC methods, such as trial wave functions with Jastrow factors, have proved useful, they come with some weaknesses: The inflexibility of the trial wave function may lead to poor approximations of the real ground state, even if a global minimum in energy for the particular trial wave function was obtain. Further, observables are hard to estimate from trial wave functions, because popular methods for sampling, such as MCMC, tends to produce data with high degree of auto-correlation(sitat). This is even more true for larger systems, and leads to a high statistical uncertainty in estimated observables.

In recent years, drawing inspiration from Deep Learning, Artificial Neural Networks have been employed as trial wave functions with great success (sitat). Because of the huge number of parameters, Artificial Neural Networks tend to be much more flexible than classical Trial Wave Functions. However, this results in a greater challenge to optimize the parameters, and the usual problem with auto-correlated samples is still present.

To address the latter problem, some papers have very recently succeeded in employing auto-regressive models on spin-systems, such as Quantum Ising-model. To obtain a spin-configuration, each spin is sampled in sequence, each time conditioning on the previous sampled spins. Because of the nature of auto-regressive models, all sampled spin-configurations are uncorrelated, removing the problem with high statistical variance in estimating observables.

In this paper, we present an autoregressive model for bosonic quantum dots based on Recurrent Neural Networks, inspired by the work of (sitat).

II. THEORY

A. The System

1. The Potentials

The Hamiltonian under investigation describes N bosons in a potential trap, and is on the form

$$H = \sum_{i=1}^{N} \left(\frac{-\hbar^2}{2m} \nabla_i^2 + V_{\text{ext}}(\mathbf{r}_i) \right) + \sum_{i<j}^{N} V_{\text{int}}(\mathbf{r}_i, \mathbf{r}_j)$$

where V_{ext} is the external potential of the trap while V_{int} is the internal potential between the particles. The external potential has an elliptical form, being anisotropic in the z-direction:

$$V_{\text{ext}}(\mathbf{r}) = \frac{1}{2} m \left(\omega \left[x^2 + y^2 \right] + \omega_z z^2 \right) \tag{II.1}$$

The internal potential is a hard shell potential, being infinite for distances where two bosons overlap:

$$V_{\text{int}} = \begin{cases} \infty, & \text{for } |r_i - r_j| \leq 0 \\ 0, & \text{otherwise} \end{cases}$$

2. Non-interacting Case

For non-interacting bosons in a spherical with $\beta = 1$ and $a = 0$ the system reduces to spherical harmonic oscillators where analytical solutions are available. The trial wavefunction reduces to simply the product of one body elements

$$\Psi_T(\mathbf{r}) = \prod_i^N \exp\left[-\alpha \left(x_i^2 + y_i^2 + z_i^2 \right) \right] = \prod_i^N \exp\left(-\alpha |r_i|^2 \right)$$

while the Hamiltonian reduces to

$$H = \sum_i^N \frac{-\hbar^2}{2m}\nabla_i^2 + \frac{1}{2}m\omega^2 r_i^2$$

which in natural units is

$$H = \frac{1}{2}\sum_i^N -\frac{1}{m}\nabla_i^2 + m\omega^2 r_i^2$$

Applying the Hamiltonian gives the local energy as

$$E_L = \frac{\alpha d}{m}N + \left(\frac{1}{2}m\omega^2 - \frac{2\alpha^2}{m}\right)\sum_i^N r_i^2$$

where d is the dimension. As the factor $\sum r_i^2$ is always positive, its term should be minimized, which is accomplished by setting $\alpha = \frac{m\omega}{2}$, giving a minimal local energy of

$$E_L = \frac{\omega d N}{2} \tag{II.2}$$

3. *Interacting Case*

The local energy for the full interacting case is much more complicated. The full computation is deferred to appendix **??**. The expression evaluates to

$$\begin{aligned}
E_L = -\frac{1}{2m}\sum_i \Bigg[& 4\alpha^2\left(x_k^2\hat{\mathbf{i}} + y_k^2\hat{\mathbf{j}} + \beta^2 z_k^2\hat{\mathbf{k}}\right) \\
& - 2\alpha(d - 1 + \beta) \\
& - 4\alpha\left(x_k\hat{\mathbf{i}} + y_k\hat{\mathbf{j}} + \beta z_k\hat{\mathbf{k}}\right)\sum_{l\neq k}\frac{\mathbf{r}_k - \mathbf{r}_l}{r_{kl}}u'(r_{kl}) \\
& + \sum_{i\neq k}\sum_{j\neq k}\frac{(\mathbf{r}_k - \mathbf{r}_i)(\mathbf{r}_k - \mathbf{r}_j)}{r_{ki}r_{kj}}u'(r_{ki})\,u'(r_{kj}) \\
& + \sum_{l\neq k}\left(u''(r_{kl}) + \frac{2}{r_{kl}}u'(r_{kl})\right)\Bigg] \\
& + \sum_i V_{\text{ext}}(\mathbf{r}_i) + \sum_{i<j} V_{\text{int}}(\mathbf{r}_i, \mathbf{r}_i)
\end{aligned}$$

B. Variational Monte Carlo

In order to find a good candidate wavefunction for a given potential, one can employ the *variational principle*. One starts by guessing a trial wavefunction $|\Psi_T\rangle$ and estimating the trial energy, which is guaranteed to be equal to or higher than the true ground state energy E_0:

$$E_0 \leq E = \frac{\langle\Psi_T|H|\Psi_T\rangle}{\langle\Psi_T|\Psi_T\rangle} \tag{II.3}$$

If $|\Psi_T\rangle$ is an eigenfunction of the Hamiltonian, the variance σ^2 will be minimal

$$\sigma^2 = \frac{\langle\Psi_T|H^2|\Psi_T\rangle}{\langle\Psi_T|\Psi_T\rangle} - \left(\frac{\langle\Psi_T|H|\Psi_T\rangle}{\langle\Psi_T|\Psi_T\rangle}\right)^2 = 0$$

The variational principle expands on this idea by letting $|\Psi_T\rangle$ be a functional class of a *variational parameter* α. By varying α one can find the optimal trial wavefunction within the functional class by minimizing σ^2.

Only a small collection of potentials have analytical solution using the variational principle. For most potentials, one must numerically integrate (II.3) using Monte Carlo integration.

For a stochastic variable x with probability density function $p(x)$, the average $\langle x\rangle$ is defined as

$$\langle x\rangle = \int_{\mathbb{R}} x p(x)\mathrm{d}x$$

By sampling the stochastic variable M times, the average can be approximated by

$$\langle x\rangle = \int_{\mathbb{R}} x p(x)\mathrm{d}x \approx \frac{1}{M}\sum_{i=1}^M x_i p(x_i)$$

Applying this to an observable $\mathcal{O}$, we have

$$\begin{aligned}
\langle\mathcal{O}\rangle &= \langle\Psi|\mathcal{O}|\Psi\rangle \\
&= \int \mathrm{d}\mathbf{r}\Psi^*\mathcal{O}\Psi \\
&= \int \mathrm{d}\mathbf{r}|\Psi|^2\frac{1}{\Psi}\mathcal{O}\Psi \\
&= \frac{1}{M}\sum_{i=1}^M p(\mathbf{r})\mathcal{O}_L
\end{aligned}$$

where $|\Psi|^2$ is defined as the probability density function, and $\frac{1}{\Psi}\mathcal{O}\Psi$ the *local operator*.

The local trial energy can then be defined as

$$E_L = \frac{1}{\Psi_T}H\Psi_T$$

which can be computed using Monte Carlo integration as

$$\langle E_L\rangle \approx \frac{1}{M}\sum_{i=1}^M p(\mathbf{r}_i)E_L(\mathbf{r}_i)$$

The goal is therefore to minimize minimizing $\sigma^2 = \langle E_L^2\rangle - \langle E_L\rangle^2$ over the variational parameter α.

C. DNN and RNN

D. Autoregressive Models

E. Metropolis-Hastings

The estimate of the local energy relies on samples from the trial wave function. To get a physical value, the configuration of the particles must be as physical and probable. As the configuration achieving this is unknown, the configuration space must be explore. This is done through Monte Carlo simulation, more specifically by using the Metropolis-Hastings algorithm.

At each MC step a single particle is chosen at random, and a change to its position is proposed by moving it a fixed step length Δ and computing the ratio between new and old probability densities

$$\omega = \frac{P\left(\mathbf{r}_r, \ldots, \mathbf{r}_k^*, \ldots, \mathbf{r}_n\right)}{P\left(\mathbf{r}_r, \ldots, \mathbf{r}_k, \ldots, \mathbf{r}_n\right)} = \frac{\left|\Psi_T\left(\mathbf{r}_r, \ldots, \mathbf{r}_k^*, \ldots, \mathbf{r}_n\right)\right|^2}{\left|\Psi_T\left(\mathbf{r}_r, \ldots, \mathbf{r}_k, \ldots, \mathbf{r}_n\right)\right|^2}$$

where $\mathbf{r}^*$ denotes a modified position. If the ratio ω is greater than a uniformly distributed number $\theta \in [0, 1]$, the move is accepted. This ratio can often be reduced analytically to obviate the need for recomputing the entire probability density each step.

F. Onebody Density

A useful way to understand a many body system is to integrate over all dimensions except for one, yielding the *one body density* $\rho(\mathbf{r})$, defined as

$$\rho(\mathbf{r}) = \int \mathrm{d}\mathbf{r}_2 \ldots \mathrm{d}\mathbf{r}_N \left|\Psi_T(\mathbf{r})\right|^2$$

It is a scaled probability density function giving the number of particles within the volume $\Delta\mathbf{r}$ as $\rho(\mathbf{r})\Delta\mathbf{r}$. By convention the integral over all $\mathrm{d}\mathbf{r}$ yields the total number of particles in the system, N.

III. METHOD

A. Architecture

B. Metropolis Sampling

C. Automatic Differentiation in Tensorflow

D. Optimization

E. Statistical Treatment

F. Experimental Setup

G. One-Body Density

To extract a one-body density from a Monte-Carlo simulation, we partition space into a number of bins in an appropriate range where the wave function is large. The bin size can be chosen small to get finer details of the density, but will require more data to mitigate statistical error.

For each particle configuration produced at every Metropolis step, the number of particles coinciding with each bin is checked. The one-body density is then produced by averaging over all Metropolis steps.

IV. RESULTS AND DISCUSSION

A. One-Particle DNN Model

1. *Relu vs Tanh*

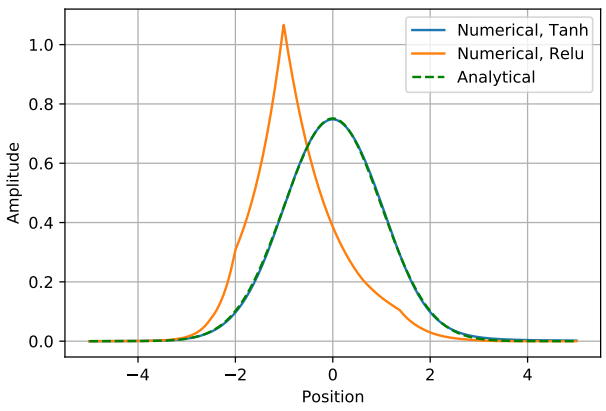

Figure IV.1

Appendix A