Neural Networks for Quantum Many Body Physics

Group L2

June 2023

1 Introduction

1.1 The Goal of this Experiment

In this AI Lab experiment, we are addressing the issue of finding a many-body quantum Hamiltonian's ground state wavefunction. In particular, we will be working with the Rydberg-atom hamiltonian. At absolute zero, where phase transitions that are only caused by quantum fluctuations are possible, such a wavefunction describes the characteristics of the quantum system. We will focus on the quantum phase change that occurs in a system of neutral atoms in an optical array that serves as a Rydberg atom quantum simulator.
Using the "Restricted Boltzmann Machine" ansatz approach, we will find the initial parameters for the many-body Hamiltonian, then reduce the energy expectation value to identify the best approximation of the ground state. To determine the critical value point for the system, we will simulate graphs for changes in order parameter values.

1.2 Introduction to the Physics Background

In our experiment, we are working with the Rydberg atom quantum simulator, where an optical lattice is used to store atoms to create a lattice geometry. In such a lattice, a collection of neutral atoms, usually alkali atoms like cesium, are trapped and cooled to extremely low temperatures using techniques such as laser cooling. This cooling process allows the atoms to be confined and manipulated for our experimentation. We have two energy levels of each atom, one of which is the highly excited Rydberg atomic state and the other is the ground state. By using laser light tuned to the energy difference between these two levels, the two energy levels are coupled. For our experiment, we are considering the Hamiltonian model as:

$$H = \frac{\Omega}{2} \sum_{n=1}^{N} \hat{\sigma_i^x} - \Delta \sum_{n=1}^{N} \hat{n_i} + \sum_{i \neq j} \frac{C}{|i-j|^6} \hat{n_i} \hat{n_j}$$

Here, N is number of atoms on the one dimensional chain, Ω , Δ, and C are parameters representing laser power, laser frequency, and inter-atom distances and $\sigma_i^x = |g\rangle_i\langle r| + |r\rangle_i\langle g|$. The g and r represent the ground and the Rydberg excited states, respectively.
We will approximate this model by assuming only nearest-neighbour site interactions, set $\Omega/2 = 1$, and perform unitary transformation $\prod_i \hat{\sigma_i}^z$ to flip the sign of the off-diagonal term, such that the Hamiltonian becomes stoquastic. In this respect, open boundary conditions indicate that the sites at the boundary couple only to interior neighbours. Now, the Hamiltonian will finally look like :

$$H = -\sum_{n=1}^{N} \hat{\sigma_i^x} - \Delta \sum_{n=1}^{N} \hat{n_i} + V \sum_{i=1}^{N-1} \hat{n_i}\hat{n}_{i+1}$$

Here, the term V represents the Vanderwaal's coefficient. In our work, we will set $V > 0$ to a sufficiently high value so as to reinforce our interest in considering only nearest neighbour interactions. When $V > 0$, we get either a disordered or an ordered phase. For $\Delta < \Delta_c$ we will observe the many-body ground state, which is the disordered one, and for $\Delta > \Delta_c$ we get an ordered phase resembling $|grgr...\rangle$ or $|rgrg...\rangle$. The ordered phase is characterised by the spontaneous breaking of lattice translational symmetry which happens when the excited states outnumber the ground states. This phase transition can be initiated by tuning the values of the parameters in our Hamiltonian. When N is odd, with open boundary conditions this phase transition can be best represented by the order parameter :

$$M = \frac{1}{N-1} \sum_{i=1}^{N-1} (-1)^i (\hat{n_i} - \hat{n}_{i+1})$$

We will be trying to estimate Δ_c which is where the phase transition takes place. We expect the order parameter M to disappear below the critical value for Δ, as we only anticipate it for the ordered phase of the lattice.

1.3 Restricted Boltzmann Machines

Restricted Boltzmann machines are a simple form of a neural network with just two layers, one visible layer $\sigma = (\sigma_1, \sigma_2, ..., \sigma_N)$ and one hidden layer $\mathbf{h} = (h_1, h_2, ..., v_{N_h})$. The visible units represent the spins of our system. Individual nodes just take values that are either +1 or -1 and the goal of the architecture is to learn a probability distribution for when the visible nodes will be activated or deactivated. The parameters of the model will be $\theta = \mathbf{W}, \mathbf{b}, \mathbf{c}$ where $\mathbf{W}$ is a weight matrix and $\mathbf{b}$ is the bias vector for the visible unit and $\mathbf{c}$ the bias vector for the hidden unit. The probability for a configuration of the visible units will be

$$P(\sigma) = \exp\left\{\sum_{i=1}^{N} b_i\sigma_i\right\} \prod_{j=1}^{N_h} 2\cosh\left\{c_j + \sum_{i=1}^{N} \sigma_i W_{ij}\right\}$$

1.4 Approximating the ground state energy

In order to approximate the ground state of the system, we will use the probability distribution of the visible units of an RBM and use it as an ansatz for our wavefunction, i.e. $\psi_\theta(\sigma) = P(\sigma)$. In the Hilbert space of N qubits we can then write the wavefunction as

$$|\psi_\theta\rangle = \sum_\sigma \psi_\theta(\sigma) \, |\sigma\rangle \, .$$

Let us denote the Energy eigenfunctions as $|\psi_\theta\rangle$ For the eigenvalues of our wavefunction we then have the inequality

$$\begin{aligned} E(\theta) &= \frac{\langle\psi_\theta| \, \hat{H} \, |\psi_\theta\rangle}{\langle\psi_\theta|\psi_\theta\rangle} \\ &= \frac{1}{\langle\psi_\theta|\psi_\theta\rangle} \sum_n \langle\psi_\theta|\psi_n\rangle \, \langle\psi_n| \, \hat{H} \, |\psi_\theta\rangle \\ &= \frac{1}{\langle\psi_\theta|\psi_\theta\rangle} \sum_n \langle\psi_\theta|\psi_n\rangle \, E_n \, \langle\psi_n|\psi_\theta\rangle \\ &>= \frac{1}{\langle\psi_\theta|\psi_\theta\rangle} \sum_n \langle\psi_\theta|\psi_n\rangle \, E_0 \, \langle\psi_n|\psi_\theta\rangle \\ &= E_0, \end{aligned}$$

with equality if $|\psi_\theta\rangle = |\psi_0\rangle$.
Therefore our goal with the ansatz is to select the parameters θ, such that the Energy is minimized, we don't have to worry about the energy being too small. This will be done with batch gradient descent. The hyperparameters of this method are the batch size and the learning rate. We will look at how to tune these parameters in subsection 2.1.

1.5 Sampling

The size of the Hilbert space of N qubits is 2^N, therefore for large enough values of N calculating the expectation values becomes infeasible. We will not calculate exact expectation values of the operators but instead sample from the distribution of the RBM with a Metropolis-Hastings algorithm. We will name the number of samples drawn before updating the parameters the batch size.

2 Results

2.1 Tuning Hyperparameters

To ensure the effectiveness and efficiency of our RBM model, we engaged in a process of tuning the hyperparameters. Our aim was to find the optimal values for the hyperparameters that would lead to the convergence of the energy in a reasonable number of batches. We optimized the following parameters:

- Alpha: The alpha hyperparameter determines the number of hidden units in the RBM, where the number of hidden units (N_h) is equal to alpha times the number of visible units (N_v).
- Batches: Batches refer to subsets of the training data used for training the RBM in mini-batch gradient descent. Each batch consists of a fixed number of samples.
- Sample size: Sample size or Batch size refers to the number of samples drawn using the metropolis hastings algorithm to calculate the energy expectation value for each batch during the RBM training process.
- Learning rate: The learning rate (η) determines the step size at which the RBM adjusts its weights during training, influencing the speed of convergence and the magnitude of weight updates.

During the training process, we encountered two primary challenges. First, we needed to ensure that the code executed within a reasonable runtime to facilitate practical usage. Second, we observed random energy spikes or noise in the Energy vs. Batch curves, which required attention to achieve a stable convergence.

We began the tuning process by selecting initial hyperparameters. We set N_v to 5 (min value of N_v to be used for the results), the maximum number of batches to 500, and the learning rate (η) to 0.01. We used an α value of 5, where N_h (the number of hidden units) is equal to $\alpha \times N_v$. We also employed a total of 100 samples generated in the RBM training.

To address the first challenge of runtime, we evaluated the convergence time for different hyperparameter choices. We ran multiple simulations (at least 5) with randomized initial samples to ensure a representative result. Based on these simulations, we adjusted the hyperparameters to achieve convergence within a reasonable timeframe. We aimed for a convergence time of approximately 1.5 to 2 minutes.

To tackle the second challenge of random energy spikes, we systematically modified one hyperparameter at a time to observe its effects on convergence and noise reduction in the Energy vs. Batch curves. By analyzing these effects, we aimed to identify the hyperparameter values that minimize energy spikes while maintaining convergence.

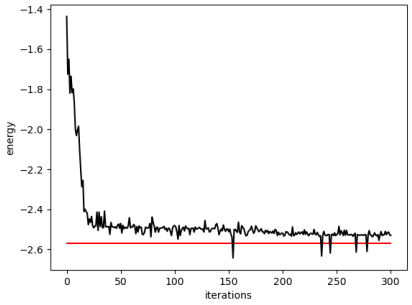

Figure 1: Convergence for initial hyper-parameters

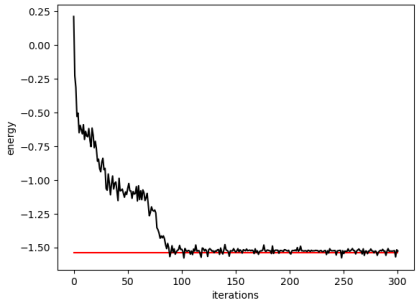

Figure 2: Convergence for $\alpha = 10$ and rest parameters same

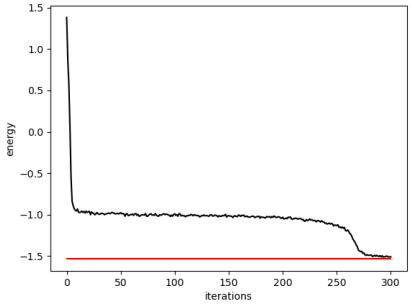

Figure 3: Convergence after sample size optimization= 1000

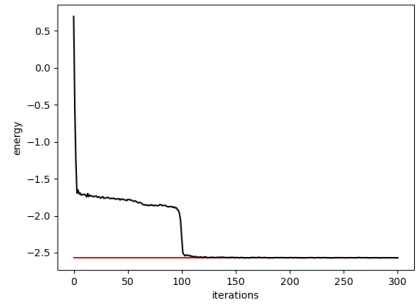

Figure 4: Convergence for the best hyperparameters

Initially, we varied the α value and observed a decrease in random energy spikes when setting it to around 10. Further increasing α resulted in an increase in noise, negating the desired effects. We also considered the convergence time during this optimization process.

After optimizing α, we focused on the number of samples per batch. By gradually increasing this parameter, we identified that using 1000 samples led to reduced noise and facilitated convergence within a reasonable runtime.

Throughout the tuning process, we periodically assessed the convergence time for the largest value of $N_v = 13$ that we planned to use in our final results. We aimed to ensure that the convergence time remained within an acceptable range.

After fine-tuning the hyperparameters, we conducted approximately 20 simulations with the finalized values. These simulations consistently achieved convergence within 150 batches. To allow for a safe margin, we set the total number of training batches to 300, ensuring ample time for convergence.

Consequently, we arrived at the following optimized hyperparameters: $\alpha = 10$,

$\eta = 0.06$, and 1000 samples. With these hyperparameters, our code had an average runtime of approximately 1.5 minutes for $N_v = 13$, and convergence was typically reached within 30 seconds. The chosen hyperparameters ensured efficient convergence and stable performance, allowing us to proceed with confidence in our subsequent analyses and results.

2.2 Order Parameter and Phase transition

Now after the training of the RBM we can use the fitted model to predict the energy as well as the order parameter as a function of Delta. To get a good comparison on how well the RBM predictions are, we computed the exact values with a numerical approach. In the following figure one can see the results of both methods.

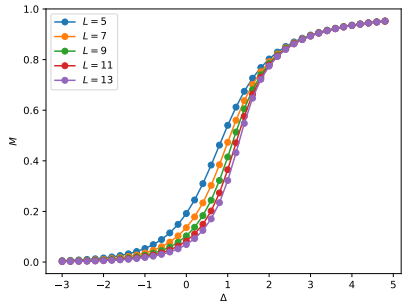

Figure 5: Numerical App. Order parameter

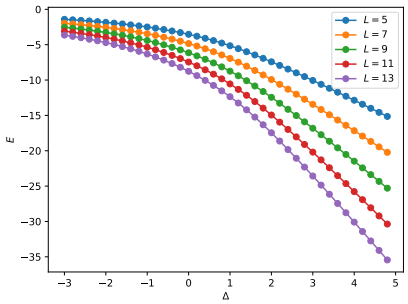

Figure 6: Numerical App. Energy

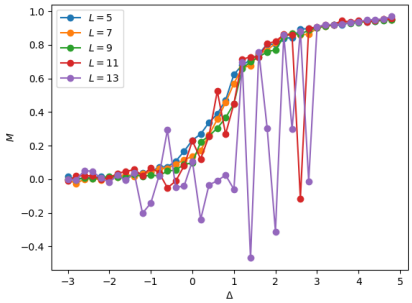

Figure 7: RBM App. Order parameter

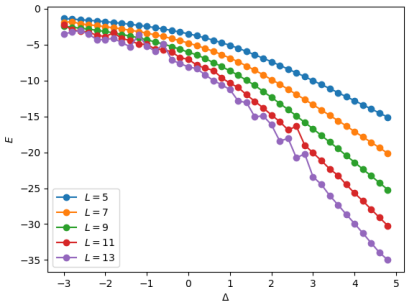

Figure 8: RBM App. Energy

One can conclude from this that the order parameter must have a critical value for Delta around zero, where below this value the order parameter vanishes. The energy does have a continuous phase transition crossing the critical value for Delta.

As we didn't check for the convergence of the order parameter, some of the values exhibit a large error compared to the exact values. We assume this is due to wavefunctions that approximate the energy pretty well but do have some difference to the exact configuration of the spins of the ground state wavefunction. An example can be found when considering the definition of the order parameter and the Hamiltonian, shifting all spins one step to the left in the chain (where the first element becomes the last), flips the sign of the order parameter but doesn't have a big effect on the energy.