Restricted Boltzmann Machines and their Information Geometry in Quantum Many-Body Problems

Author: Alex Esteban Folch
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisor: Maria Moreno Cardoner
(Dated: 14th June of 2024)

Quantum many-body problems face a scalability challenge, since the wave function scales exponentially with the number of bodies in the system. Fortunately, several machine learning approaches have been recently proposed to overcome this challenge. Specially promising are the Restricted Boltzmann Machines (RBMs), to turn the problem into a manageable computational form. Moreover, information geometry has been studied in RBMs, and the Fisher Matrix has shown to reveal relevant information about the system. Here, we create from scratch a RBM representation of the ground state for the transverse Ising model (short and long-range), and analyze the corresponding Fisher Matrix across different quantum phases, and its potential to signal the phase transition.

I. INTRODUCTION

Recently, machine learning techniques have been applied to the field of quantum many-body physics with many promising results. One method that has gathered a lot of attention is Complex Restricted Boltzmann Machine (RBM) state Ansatz using stochastic reconfiguration optimization. These variational ansätze are able not only to reproduce the ground-state energy of many-body quantum systems, but they outperform state-of-the-art tensor networks methods previously used [1].

Furthermore, work has been done in trying to extract valuable data of the inner-patterns of RBMs [5]. In fact, RBMs learn using stochastic reconfiguration, and such a method implies an energy gradient descent weighted by a Quantum Fisher Matrix (QFM), which is the quantum equivalent of the Fisher Information Matrix. The Fisher information matrix is known to be the unique Riemannian metric associated to a probability space, invariant under sufficient statistics (known as Chentsov's theorem [2]). Therefore, valuable information is contained in the metric associated with RBMs.

The purpose of this work is to implement a RBM to reproduce the ground state of a one-dimensional Ising model accross its emerging quantum phase transition, and analyse the spectrum and distribution of the associated QFM. We explore both the short-range and long-range versions of the model.

II. CONCEPTS AND METHODS

A. Restricted Boltzmann Machines

A Restricted Boltzmann Machine (RBMs) is a generative model that represents a probability distribution. It contains two types of units that are connected: visible and hidden neurons, which can be thought of as two different layers. The first layer contains information about a given real system (*e.g.,* a spin-state), while the second layer comprises the hidden units, which model the dependencies between different visible units, and act as non-linear feature detectors. In fact, RBMs can be interpreted as deterministic feed-forward neural networks with a non-linear activation function (sigmoid) (see [3] for a detailed discussion).

Boltzman Machines (BM) are a special type of Markov Random Field (MRF) with a bipartite structure. A Markov Random Field is a graphical model that represents the join probability distribution of a set of random variables $\{\vec{x} = (x_1, \cdots, x_N)\}$ with a given graph structure. Nodes represent the random variables and edges their conditional dependencies. It is well known [3] that the joint probability of a certain configuration given by the MRF can be expressed with a general factorization form (Hammersley-Clifford Theorem):

$$p(\vec{x}) = \frac{1}{Z} e^{-E(\vec{x})}, \tag{1}$$

where Z is the associated partition function and $E(\vec{x})$ is the energy functional given the configuration $\vec{x}$. Interestingly, this form follows closely the equilibrium distribution typically used in statistical physics.

In this work, we focus on Complex Restricted Boltzmann Machines which are RBMs with complex values. They have proven to efficiently describe the quantum state of a many-body system [1]. In this case, the visible and hidden units are conditionally independent of themselves, that is, no edge exists between any two visible units or hidden units (see Fig. 1 for a sketch of the RBM). The n visible and m hidden units, which are denoted respectively by $\vec{v} = \{v_1, \cdots, v_n\}$ and $\vec{h} = \{h_1, \cdots, h_m\}$, are binary variables, *i.e.*, $(\vec{v}, \vec{h}) \in \{0, 1\}^{n+m}$.

Following Eq.(1), the amplitude probability associated with certain configuration for the neurons in the visible layer (*e.g.,* corresponding to the wave function amplitude of a real spin configuration) is given by:

$$\Psi_\theta(\vec{v}) = \frac{1}{\sqrt{Z}} \sum_{\vec{h}} e^{-E_\theta(\vec{v}, \vec{h})}, \tag{2}$$

with $E_\theta(\vec{v},\vec{h}) = -\vec{h}^T \cdot \overleftrightarrow{W} \cdot \vec{v} - \vec{a}^T \cdot \vec{v} - \vec{b}^T \cdot \vec{h}$ the energy function, with variational parameters $\theta = \left\{\overleftrightarrow{W}, \vec{a}, \vec{b}\right\}$. Here the matrix $\overleftrightarrow{W} = \{w_{ij},\ (i = 1, \cdots, n),\ (j = 1, \cdots, m)\}$ connects the units from the visible and the hidden layer, and $\vec{a} = \{a_1, \cdots, a_n\}$ and $\vec{b} = \{b_1, \cdots, b_m\}$ represent local biases acting as a scaling factor. The prefactor $1/\sqrt{Z}$ guarantees the normalization of the wave function over all possible configurations of $\vec{v}$.

We emphasize that in the case of *Complex* RBM, the amplitude wave function is a complex number, implying that $\vec{a}$, $\vec{b}$ and $\overleftrightarrow{W}$ are, in general, complex numbers. These parameters are fitting parameters and need to be learnt using different numerical methods, such as the Stochastic Reconfiguration method and Monte-Carlo sampling that will be discussed later. They allow us to learn the interesting underlying distribution, which will be the wave function of the quantum many-body problem.

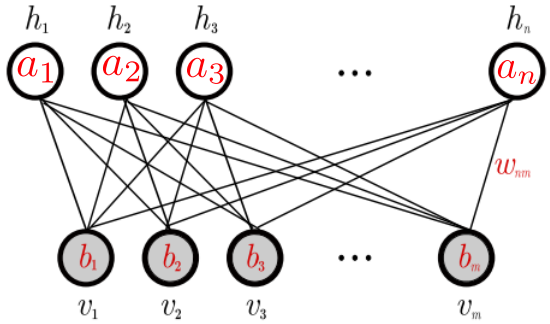

Figure 1: Structure of a RBM: a visible and a hidden layer of binary neurons denoted by $\vec{v} = \{v_1, \cdots, v_n\}$ and $\vec{h} = \{h_1, \cdots, h_m\}$, are connected via the weight matrix $\overleftrightarrow{W} = \{w_{ij};\ i = 1, \cdots, n;\ j = 1, \cdots, m\}$. No visible layer node is connected to another visible layer node, and no hidden layer node is connected to another hidden layer node. The local biases are indicated by $\vec{a} = \{a_1, \cdots, a_n\}$ and $\vec{b} = \{b_1, \cdots, b_m\}$. (Figure extracted from [3].)

B. Stochastic Reconfiguration Method and Monte-Carlo Sampling

Using the variational amplitude probability Eq.(2), the goal is to optimize the parameters θ that minimize the energy of the system. The Stochastic Reconfiguration (SR) method efficiently fits the trial wave function while moving it towards the ground state, by defining the path using the projection of $\mathbb{I} - \delta H$. Here H is the Hamiltonian of the system, and δ is an adjustable parameter that ensures the expression is non-negative.

In the following we sketch the main idea behind the method, but a more detailed analysis can be found in [5]. In general, we can write:

$$(\mathbb{I} - \delta H)|\Psi_\theta\rangle = a_0|\Psi_\theta\rangle + \sum_i a_i \frac{\partial}{\partial\theta_i}|\Psi_\theta\rangle + |\Psi_\theta^\perp\rangle\,, \quad (3)$$

where a_k are coefficients to be determined, and $|\Psi_\theta^\perp\rangle$ is some state in the orthogonal subspace. Multiplying Eq.(6) by $\langle\Psi_\theta|$ and $\partial\langle\Psi_\theta|/\partial\theta_i$, we obtain solving for a_0:

$$\sum_j \mathbf{F}_{ij} a_j = -\delta \mathbf{R}_i, \quad (4)$$

where we define $\mathbf{F}_{ij} = \langle O_i^\dagger O_j\rangle - \langle O_i^\dagger\rangle\langle O_j\rangle$ and $\mathbf{R}_i = \langle O_i^\dagger H\rangle - \langle O_i^\dagger\rangle\langle H\rangle$. Here the operator O_i is acting on state $|x\rangle$ as $O_i|x\rangle = (\partial \log\langle x|\phi_\theta\rangle/\partial\theta_i)|x\rangle$.

We can now identify a_k as the coefficients of the update parameters and a_0 as the learning rate η. Then we have:

$$\theta_i^{k+1} = \theta_i^k - \eta \sum_j \mathbf{F}_{ij}^{-1}\mathbf{R}_j, \quad (5)$$

where k represents the step in our learning procedure. Moreover, we introduce a regularization to ensure invertibility of the form :

$$\theta_i^{k+1} = \theta_i^k - \eta \sum_j (\mathbf{F} + \epsilon(k)\mathbb{I})_{ij}^{-1}\mathbf{R}_j, \quad (6)$$

where $\mathbb{I}$ is the identity matrix, $\eta = 0.01$, and we have introduced $\epsilon(k)$ as the regularization parameter function. Following [1], we use the regularization function $\epsilon(k) = \max(\epsilon_0 b^k, \epsilon_{\min})$, with $\epsilon_0 = 100$, $b = 0.9$ and $\epsilon_{\min} = 10^{-4}$.

C. Monte-Carlo Sampling

In general, it is not practical to evaluate the expectation values in Eq.(6) in an exact way, due to the exponentially increasing Hilbert space dimension with system size. Following [1], we use a Monte-Carlo Metropolis Hastings algorithm to estimate these expectation values. Starting with a random configuration, at each step of the sampling a spin flip is performed, and the new state is accepted with probability $\mathcal{P}\left(\mathcal{S}^{(k)} \to \mathcal{S}^{(k+1)}\right) = \min\left(1, \left|\Psi_\theta\left(\mathcal{S}^{(k+1)}\right)/\Psi_\theta\left(\mathcal{S}^{(k)}\right)\right|^2\right)$, where $\mathcal{S}^{(k)}$ is the spin configuration. To estimate the expectation values in Eq.(6), we will sample the operators over a large set of spin configurations generated this way.

D. Quantum Fisher Matrix and its Spectral Analysis

It is well known that the SR method can be interpreted with a purely geometrical approach. In fact, for positive wavefunctions SR may be interpreted as a Natural Gradient, which uses the information geometry provided by the energy function to maximize convergence, and proposes to "flatten" the space locally before making the parameter update.

The natural Riemannian metric in probability space is chosen to be the Fisher information matrix $\mathbb{F}$, as it is the only one invariant under sufficient statistics. It

turns out that for positive wavefunctions, this is directly related to the previously defined matrix $\mathbf{F} = \mathbb{F}/4$, and the squared distance between two wavefunctions in the parameter space can be written as $ds^2 = \sum_{i,j} \mathbf{F}_{ij} d\theta_i^* d\theta_j$. Thus, the matrix $\mathbf{F}$, also called Quantum Fisher Matrix, plays the role of a Riemannian metric.

Interestingly, it has been shown in [5] that the spectral properties of the QFM can reveal insights of the many-body state. Since the QFM is positive semi-definite, we can estimate a set of eigenvectors and non-negative real eigenvalues. The eigenvalue can be thought as the steepness in parameter space in the associated eigenvector direction. Moreover, the eigenvectors of the QFM contain information about the real spin correlations of the quantum many-body state. Let us first remind that the QFM is a $N + M + NM$ square matrix that we can divide into two first blocks associated with the biases and a third block associated with the weights between hidden and visible layers. As previously mentioned, the correlation between two nodes from the visible layer are only induced via the hidden layer, which means that all spin correlation information is already encoded into the weights subblock. We can then diagonalize the QFM where the biases part has been truncated, and quantify the entanglement between the visible and hidden layers of the resulting renormalized eigenvectors $|\Psi_i^w\rangle$. The entanglement entropy of the reduced density matrix describing the visible layer, $\hat{\rho}_i^{\vec{v}} = \mathrm{Tr}_{\vec{h}} \left[|\Psi_i^w\rangle \langle \Psi_i^w| \right]$, is given by $S[\Psi_i] \equiv -\mathrm{Tr} \left[\hat{\rho}_i^{\vec{v}} \log \hat{\rho}_i^{\vec{v}} \right]$.

III. RESULTS: APPLICATION TO THE FERROMAGNETIC ISING MODEL

The previous defined concepts will be now applied to the paradigmatic one-dimensional tranverse Ising Model, whose Hamiltonian can be expressed as:

$$H_{\mathrm{Ising}} = -\sum_{i=1}^{N} \sigma_i^z \sigma_{i+1}^z - h \sum_{i=1}^{N} \sigma_i^x. \tag{7}$$

Where N is the total number of spins, σ_i^k is the Pauli operator in the k-direction acting on site i, and h is the (dimensionless) transverse field. We will consider periodic boundary conditions, by setting $N+1 \to 1$. Despite the model can be analytically solved, we will use here exact diagonalization to benchmark our results.

Moreover, we will generalize the model and consider also a tunable range for the ZZ interactions:

$$H_\gamma = -\sum_{i=1}^{N} \sum_{j=1}^{N} \sigma_i^z \sigma_j^z \cdot \frac{1}{|r(i,j)|^\gamma} - h \sum_{i=1}^{N} \sigma_i^x. \tag{8}$$

In order to preserve periodic boundary conditions in this model, we define the distance $r_{i,j}$ as the minimal distance $r(i,j) = \min(|i-j|, |N+i-j|)$. We will refer to Eq.(7) by $\gamma = \infty$.

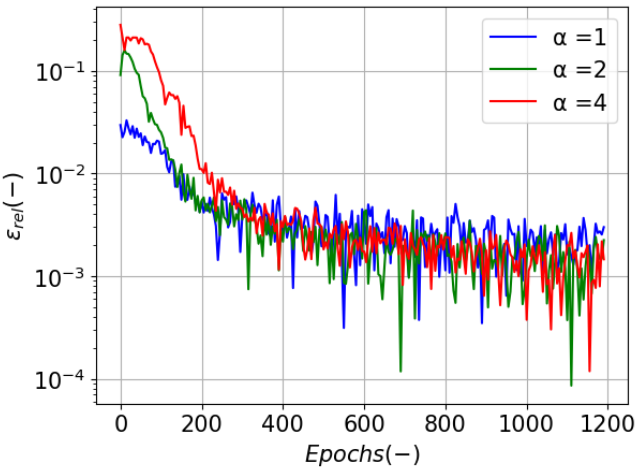

Figure 2: Relative error of ground state energy ϵ_{rel} versus number of epochs, for the case $h = 1$ and $N = 20$. Each line is for a different value of the hidden layer density $\alpha = M/N$.

A. Ground State Energy

First, we test the ability of the RBMs to find the exact ground state energy, for different densities of the hidden layer $\alpha = M/N$, where where N is the number of spins in the visible layer and M is the number of hidden nodes. We evaluate the relative error in energy as $\epsilon_{\mathrm{rel}} = |E_{\mathrm{RBM}} - E_{\mathrm{exact}}|/|E_{\mathrm{exact}}|$ as a function of the epochs or iterations, where each epoch corresponds to a Monte Carlo sampling with N iterations .

The result for $h = 1$ and $N = 20$ is shown in Fig.2. For bigger value of α we expect to get a better representation of the ground-state energy, but the we found that gains are marginal for $\alpha > 2$. Even more, the convergence is much slower for bigger α. From our testing, we decide to use the results with $\alpha = 2$, and keep N constant, since the relative error for all values of h is practically the same. We suspect, that bigger α might lead to better representation of higher order moments.

We now evaluate ϵ_{rel} per epoch for different values of h, as shown in Fig. 3(a). We observe that convergence is slower close to the critical point ($h = 1$). In the same figure, we perform the same analysis for the long-range model, with $\gamma = 1$, and find that the neural network is still able to make a good representation of the wave function. It seems in this case convergence is faster, but it decreases as h is increased. This could be explained as for this model the critical point is shifted towards larger values of h.

B. Spectral Analysis of the QFM

In this section we explore the spectral features of the QFM accross the quantum phase transition emerging in the Ising model. According to [5], we expect that the

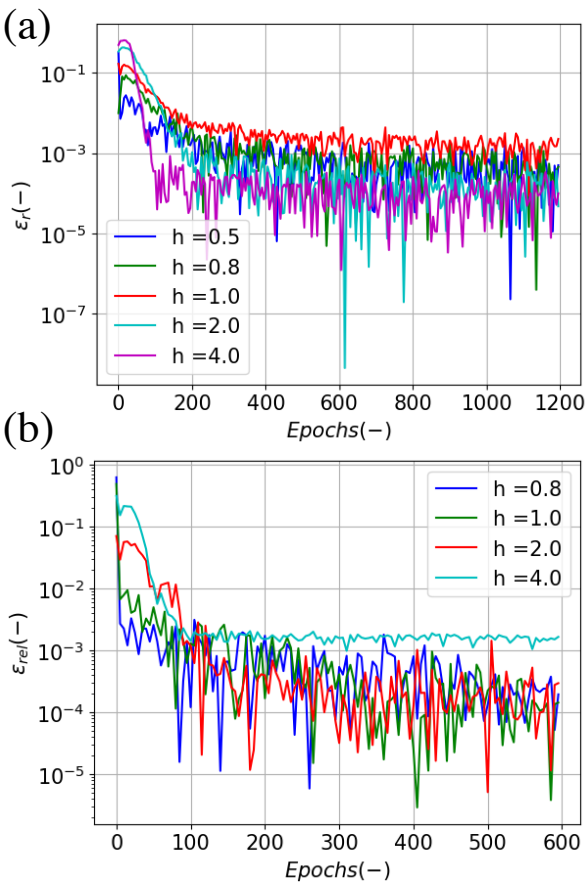

Figure 3: Relative error of ground state energy ϵ_{rel} versus number of epochs for different values of h. (a) Short-range Ising model, and (b) long-range Ising model ($\gamma = 1$). ($N = 20$, $\alpha = 2$).

spectral features that the spectral distribution of any system should hold relatively constant and only depends on the phase it is, and how far it is from a critical point. However, it is important to keep in mind that our system is finite-sized, so the transition will be smooth. For the first-neighbour interacting Ising model ($\gamma = \infty$), we compute the eigenvalues and order them to obtain the spectra as shown in Fig.4(a). The corresponding entanglement entropies are also plotted in Fig.4(b).

Interestingly, we observe that in general the spectra are very similar when the system is in the same phase. They deviate smoothly close to the phase transition, as we do not have an infinite number of spins.

We observe that in the ferromagnetic phase, the eigenvectors with the biggest eigenvalues tend to have the smallest entropy. Since the entanglement entropy between visible and hidden layer is low, this means that the eigenvectors do not encode correlations between real spins, and the system is close to a product state. In the extreme case of $h = 0$, where all spins are aligned up or down, it can be proved that the QFM has rank one, with a single eigenvalue and exactly zero entropy [5]. In these eigenvectors, most of the information is contained in the biases part of the QFM, and they will mostly contribute to first moments or expectation values of the observables (e.g., the mean spin value), but not to the spin correlations.

The entanglement entropy behavior in the paramagnetic phase ($h > 1$) looks very different. In this case we expect a smoother, exponentially decaying spectra, and a QFM that is almost a random matrix for the weights [5]. Here we observe that the eigenvalues have much larger entanglement entropies.

In the following, we will study how this behavior changes when modifying the range of the interactions given by the parameter γ. For a fixed value of h, we compare the spectra and entropy for the cases of $\gamma = \infty$ (first neighbour interaction), $\gamma = 3$ and $\gamma = 1$ (long-range). In general, we find that the cases of $\gamma = 3$ and $\gamma = \infty$ are qualitatively similar. In contrast, the case $\gamma = 1$ shows a different behavior as soon as $h > 1$. As an example, we plot the result for $h = 4$ in Fig. 5. We observe that for this value, while the $\gamma = \infty$ and $\gamma = 3$ are already in the paramagnetic phase, the case $\gamma = 1$ shows a spectra more similar to the ferromagnetic phase, but with larger value of the entropy. This could be a signature

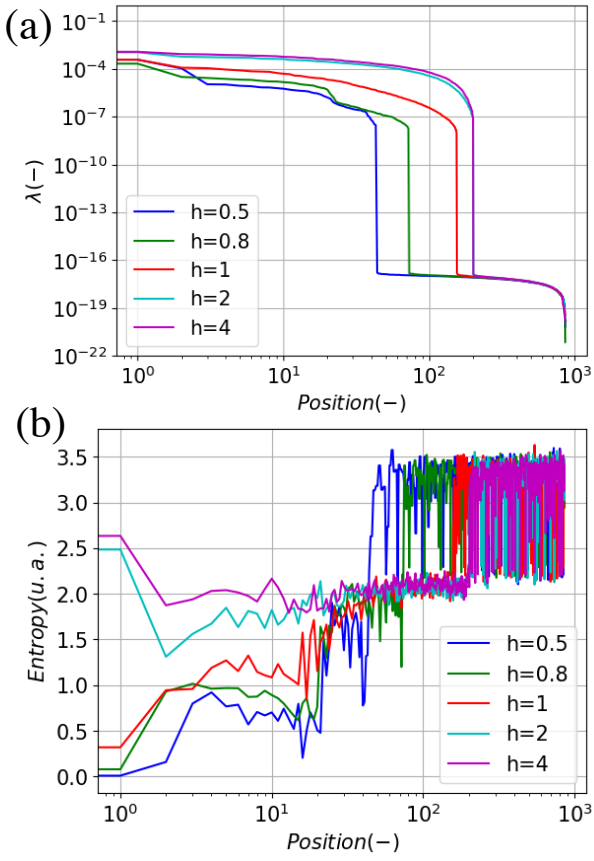

Figure 4: (a) Eigenvalues and (b) entanglement entropy of the associated wave-vectors of the reduced QFM for the short-range Ising model and different values of h. ($N = 20$, $\alpha = 2$).

that the quantum phase transition occurs at larger value of h, and that larger spin correlations are built in the system due to the long range nature of the interactions. This is in agreement with the expectation that for $\gamma = 1$, and in the thermodynamic limit ($N \to \infty$), $h_c \to \infty$ [4]. In this case the system becomes superextensive, and no true transition should be happening.

Finally, we study the long-range case $\gamma = 1$ further varying the field h. The spectra is presented in Fig. 6, where we observe a smooth transition around $4 \leq h \leq 6$, possibly due to the system finite size ($N = 20$).

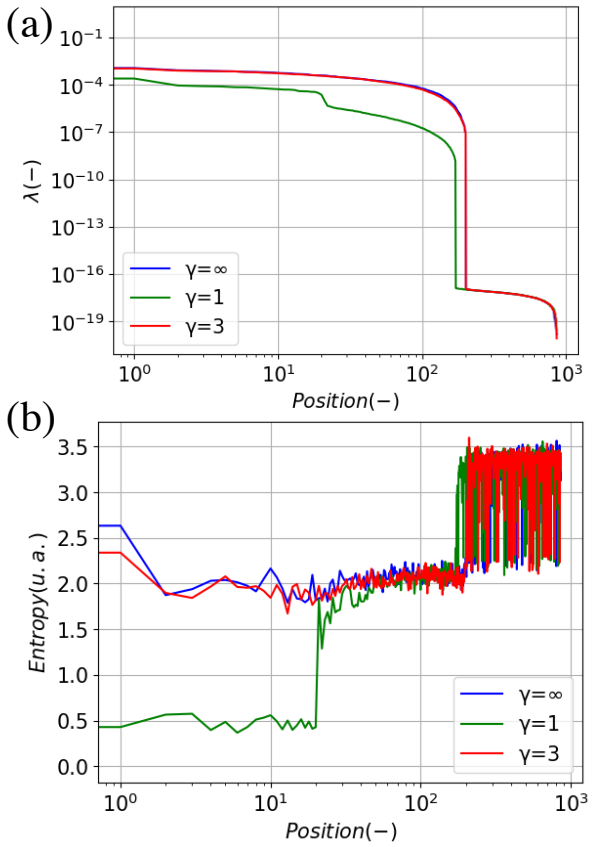

Figure 5: (a) Eigenvalues and (b) entanglement entropy of the associated wave-vectors of the reduced QFM for the long-range Ising model ($\gamma = 1$) and different values of h. ($N = 20$, $\alpha = 2$).

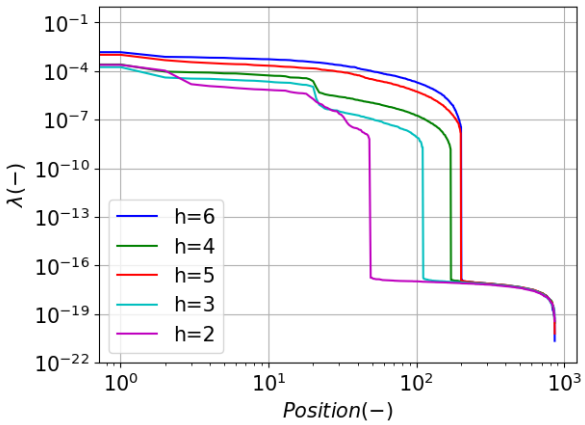

Figure 6: Eigenvalues of the QFM for the long-range Ising model ($\gamma = 1$) and different values of h. ($N = 20$, $\alpha = 2$).

IV. CONCLUSIONS

Using RBM we have simulated the quantum many-body state of the Ising model with tunable interaction range, and have analysed the corresponding QFM for the different phases of the model. We find that the spectra and entanglement entropy of the eigenvectors associated with the QFM contain valuable information about the many-body phases of the system and correlations, and can be used to signal a quantum phase transition. It would be worth to extrapolate these results to larger system sizes, to see if the expected thermodynamic limit is recovered, and sharper signatures of the quantum phase transition arise.

It would be interesting to perform a similar study of the QFM with other models beyond the Ising model, to see if this can capture more exotic types of ordering, as well as apply it to open quantum systems or to the time-evolution of a quantum state after a quench.

Acknowledgments

I'd like to thank my advisor Maria Moreno Cardoner for her help in developing such an interesting topic.

[1] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, February 2017.

[2] N. N. Cencov. Statistical decision rules and optimal inference. *Translations of Mathematical Monographs*, 53, 2000.

[3] Asja Fischer and Christian Igel. An introduction to restricted boltzmann machines. pages 14–36, 01 2012.

[4] Jan Alexander Koziol, Anja Langheld, Sebastian C. Kapfer, and Kai Phillip Schmidt. Quantum-critical properties of the long-range transverse-field ising model from quantum monte carlo simulations. *Phys. Rev. B*, 103:245135, Jun 2021.

[5] Chae-Yeun Park and Michael J. Kastoryano. Geometry of learning neural quantum states. *Physical Review Research*, 2(2), May 2020.

V. APPENDIX

A. Spectra and entropies of $h = 1$

Here we include the results in the critical point of the short-range, since the it differentiates the different behaviour of the phases.

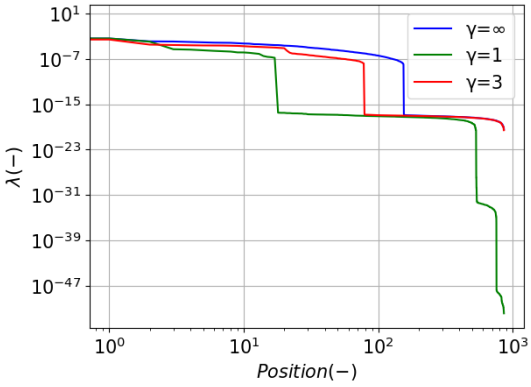

Figure 7: Magnitude of the eigenvalue for all the ordered eigenvectors in function with its position in the ordered list for different values of γ of the coupling interaction with $h = 1$.

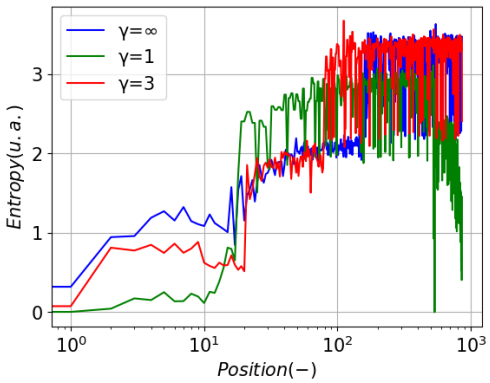

Figure 8: Entropy of the ordered eigenvectors with respect to its position in the list for different values of γ with $h = 1$.

B. Time evolution

Using the same procedures as in [1], we also were able to reproduce a quantum quench. We present the equation that evolution of the parameters satisfies:

$$\frac{d\theta(t)}{dt} = -i\mathbf{F}^{-1} \cdot R \tag{9}$$

We decided to do a quantum quench from $h = 0.5$ to $h = 1.0$ in the short-range. And decided to use as an observable $\langle \sigma_x \rangle$:

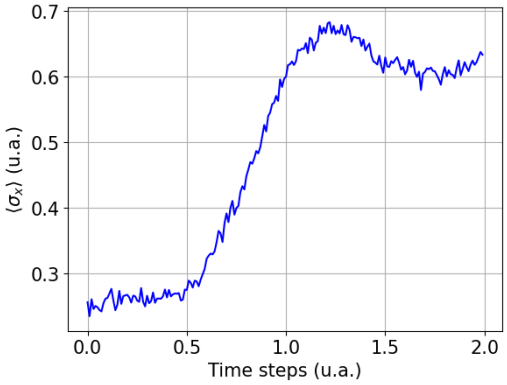

Figure 9: $\langle \sigma_x \rangle$ with respect to the arbitrary time.

Qualitatively this quantum quench is similar to the one found in [1], which presents the exact solution.