Solving the 1D $J_1 - J_2$ Heisenberg model using Restricted Boltzmann Machine (RBM) Neural Quantum State (NQS)

Devadevan M M

Numerical Methods (PHY 410.5)
Term Paper

Abstract

The difficulty in solving many-body quantum problems arises from the non-trivial correlations encoded in the exponential complexity of the many-body wave function. Recently it has been shown that machine learning of the wave function using neural network ansatz reduces this complexity to a tractable computational form. Ever since there was a burst in research using different network configurations to benchmark its performance with the state-of-the-art methods and was performing better. The most commonly used successful neural ansatz among them is the restricted Boltzmann machine(RBM). In this term paper I analyse the performance of RBM neural quantum state (NQS) by using it to solve the 1D spin-1/2 $J_1 - J_2$ Heisenberg spin chain in the frustration regime, which is still an unsolved problem.

1 Introduction

With the improvement in computational resources, techniques and data sets, machine learning in recent years has become an important versatile tool for solving problems which were thought impossible by a computer in near future. Deep learning, which is a branch in machine learning that makes use of artificial neural network (ANN), played an important role in this and is believed to be the key computational method of the future. Many machine learning applications have been done within the field of condensed-matter physics, in recent years. These include the classifying phases of matter, simulation of quantum computers, quantum state tomography, to classify experimental data, inference of phase transitions and many more.

Quantum many-body systems are characterized by a Hilbert space that grows exponentially with the number of particles. This fact restricts exact calculations to a few cases, while analytical treatments often require approximations that are not fully justified in the strongly interacting limit. Therefore, numerical techniques are the way. In the realm of computational quantum physics, the restricted Boltzmann machine (RBM), a type of ANN, was proposed as a variational ansatz for many-body quantum systems [1]. Since then, there has been a burst of research using different network configurations to benchmark its performance with the state-of-the-art methods and was performing better. In this term paper, I analyse the performance of RBM neural quantum state (NQS) by using it to solve the 1D spin-1/2 $J_1 - J_2$ Heisenberg spin chain in the frustration regime, which is still an unsolved problem.

2 Heisenberg $J_1 - J_2$ spin-1/2 model

We have a 1-D lattice with L lattice points and there is spin-1/2 degree of freedom at each lattice point, and we take periodic boundary condition. The Hamiltonian is defined as,

$$H = J_1 \sum_{\langle i,j \rangle} \hat{\vec{\sigma}}_i . \hat{\vec{\sigma}}_j + J_2 \sum_{\langle\langle i,j \rangle\rangle} \hat{\vec{\sigma}}_i . \hat{\vec{\sigma}}_j \tag{1}$$

were $\langle i,j \rangle$ is the nearest neighbours, $\langle\langle i,j \rangle\rangle$ is the second nearest neighbours and $\hat{\vec{\sigma}} = \left(\hat{\sigma}^x, \hat{\sigma}^y, \hat{\sigma}^z \right)$ are the Pauli matrices. We work in the parameter regime where $J_1,\ J_2 > 0$ and the system is frustrated in this regime. $\frac{J_2}{J_1}$ is defined to the frustration ratio.

Depending on the relative strengths of J_1 and J_2, the chain can exhibit different ground states. For strong J_1 and weak J_2 we get antiferromagnetic order which is the Neel phase. For strong J_2 and weak J_1 the spins form dimers and thus the dimer phase. When J_1 and J_2 are comparable the precise nature of the phase is debatable and unsettled. There's some evidence of spin liquid phase in this regime but it's not universally accepted. Thus the problem is still unsolved.

3 Casting Quantum Many Body Problem as a Machine Learning problem

The key idea is to use neural network as a variational ansatz for the variational Monte Carlo (VMC). We are given with a many body Hamiltonian H, whose ground state has to be found out.

$$H|\psi_i\rangle = E_i|\psi_i\rangle \tag{2}$$

where $i = 0, 1, 2, ...$ and $E_0 \leq E_1 \leq E_2 \leq ...$, ground state corresponds to the $|\psi_0\rangle$ with the lowest energy E_0.

3.1 Variational Monte Carlo (VMC)

We know that,

$$\langle H \rangle = \frac{\langle \psi | H | \psi \rangle}{\langle \psi | \psi \rangle} \geq E_0 \tag{3}$$

where $|\psi\rangle$ is some arbitrary state and E_0 the exact ground state energy. From this, it is clear that one can find the ground state wave function ψ_0 by the optimisation of $\langle H \rangle$ using a a variational ansatz for $\psi(x)$ which depends on some variational parameters $w_1, w_2, ..., w_p$ as $\psi(x, w)$.

$$\psi_0 = argmin_\psi \langle H \rangle \tag{4}$$

For the optimisation we do a gradient descent where the gradient is given by,

$$\partial_{w_k} \langle H \rangle = \langle \hat{E}_{loc} \hat{D}_k \rangle - \langle \hat{E}_{loc} \rangle \langle \hat{D}_k \rangle + cc \tag{5}$$

where the local operators are given by,

$$\langle x | \hat{E}_{loc} | x'' \rangle = \delta_{xx''} \sum_{x'} H_{x''x'} \frac{\psi(x')}{\psi(x'')} \tag{6}$$

$$\langle x | \hat{D}_k | x' \rangle = \delta_{xx'} \frac{\partial_{w_k} \psi(x')}{\psi(x')} \tag{7}$$

and $\{|x\rangle\}$ is the chosen set of basis, which for our case will be the tensor product of $\hat{\sigma}^z$ basis $\{|\sigma\rangle = |\sigma_1^z, \sigma_2^z, ..., \sigma_L^z\rangle\}$. The parameter updation equation is given by,

$$w_i = w_i - \eta \partial_{w_i} \langle H \rangle = w_i - \eta(\langle \hat{E}_{loc} \hat{D}_k \rangle - \langle \hat{E}_{loc} \rangle \langle \hat{D}_k \rangle + cc) \tag{8}$$

where η is the learning rate.

Thus the eigenvalue problem 2 has been made into an optimisation problem, with the use of the variational principle. But calculating exactly the expectation values in 8 is computationally heavy since we always deal with Hilbert spaces where the dimension grows exponentially with the size of the system. We can turn to Monte Carlo sampling to tackle this problem. The idea of Monte Carlo sampling is to replace the expectation value as the sample average, which converges to the actual expectation as the number of samples goes to infinity, as ensured by the weak law of large numbers. Now to perform the Monte Carlo sampling we need to find a way for sampling from our variational ansatz.

We use Metropolis Hastings algorithm for the Monte Carlo sampling, which is a Markov chain based technique, where we start from a random basis state and then propose a new state which is then accepted/rejected with some probability based on a condition and continue this process to get the sample of states which follow the desired pdf. The acceptance probability $A(x', x)$ of accepting a state $|x'\rangle$ given a state $|x\rangle$ is given by,

$$A(x', x) = min\left(1, \frac{\pi(x')}{\pi(x)}\right) = min\left(1, \frac{|\psi(x', w)|^2}{|\psi(x, w)|^2}\right) \tag{9}$$

3.2 Neural Quantum State (NQS)

The idea is to use a neural network as the variational ansatz for the VMC. We give in the basis state, which for our case is the tensor product of $\hat{\sigma}^z$ basis $|\sigma\rangle = |\sigma_1^z, \sigma_2^z, ..., \sigma_L^z\rangle$ and the neural network gives out the corresponding wave function $\psi(\sigma)$. Thus we take the variational ansatz of the VMC to be,

$$\psi(\sigma, w) = F(\sigma, w) \tag{10}$$

where F is the output of the chosen neural network and $w = (w_1, w_2, ...)$ the corresponding network parameters which acts as the variational parameters in the VMC language. For this term paper I use RBM as the NQS.

Restricted Boltzmann Machine (RBM)

RBM's are used to learn binary probability distributions. There's a theorem which states that an RBM can approximate any binary distribution to an arbitrary degree of accuracy, as measured by the Kullback-Leibler divergence (a measure of difference between two distributions). The RBM function for approximating binary distributions is defined as,

$$F_{RBM}(\sigma_1^z, \sigma_2^z, ..., \sigma_L^z) = \sum_h exp\left[\sum_{i,j} W_{ij}\sigma_i^z h_j + \sum_j h_j b_j + \sum_i \sigma_i^z a_i\right] \tag{11}$$

where W, a, b are the network parameters and h_i are the hidden nodes. For our purposes we set $a = 0$. The word restricted in RBM comes due to the fact that the connections are restricted to be between the visible and hidden nodes. The above functional form of RBM can be simplified in the following form so that it takes the form of the neural network shown in fig-1, which has a single hidden layer with $ln\,[cosh()]$ as the activation function.

$$\ln\left[F_{RBM}\right] = \sum_j \ln\left[cosh\left(\sum_i W_{ij}\sigma_i^z + b_j\right)\right] \tag{12}$$

And thus we have a map between the RBM wave function and the two-layer feed-forward neural network (FFNN) shown in fig-1,

$$\psi(\sigma, w, b) = \prod_j cosh\left[\sum_i W_{ij}\sigma_i^z + b_j\right] = exp(\text{output of FFNN}) \tag{13}$$

Note that the network parameters are complex numbers for our case.

3.3 Stochastic Reconfiguration (SR)

It is known as the natural gradient in the machine-learning community. This is a technique used for preconditioning the gradient (transforming the gradient) before being used in the parameter update equation, which improves the convergence rate to the ground state of the Hamiltonian H. SR can be thought of as an imaginary time evolution of the variational ansatz. We can Taylor expand $|\psi(w+\delta w)\rangle$ to first order as,

$$|\psi(w + \delta w)\rangle = |\psi(w)\rangle + \delta w_k \hat{D}_k |\psi(w)\rangle \tag{14}$$

where $\hat{D}_k$ is as defined in 7. The imaginary time evolution of variational ansatz give,

$$|\phi\rangle = e^{\epsilon \hat{H}}|\psi(w)\rangle \approx (I - \epsilon\hat{H})|\psi(w)\rangle \tag{15}$$

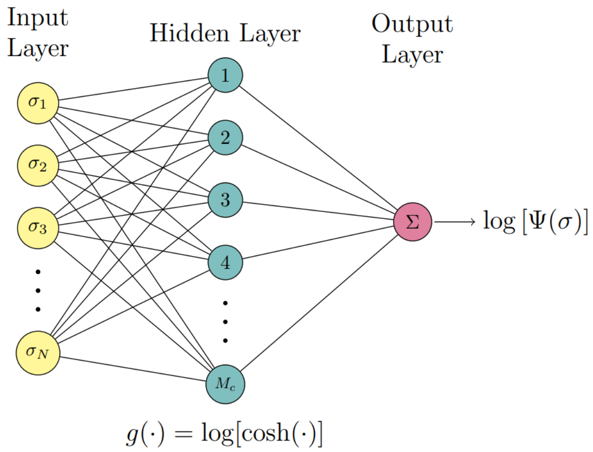


Figure 1: Schematic illustration of the feed-forward neural network representation of RBM

Now we have to minimise the norm of the state $|\phi\rangle - |\psi(w+\delta w)\rangle$ w.r.to δw_k. For calculating the distance between two states we use the Fubini-Study metric,

$$d(\psi,\phi) = arccos\sqrt{\frac{\langle\psi|\phi\rangle\langle\phi|\psi\rangle}{\langle\psi|\psi\rangle\langle\phi|\phi\rangle}} \tag{16}$$

And we are left with a linear equation to solve ,

$$S_{i,k}\delta w_k = F_i \tag{17}$$

where $F_i = \langle\hat{E}_{loc}\hat{D}_k\rangle - \langle\hat{E}_{loc}\rangle\langle\hat{D}_k\rangle$ is the gradient of $\langle H\rangle$ as shown in 5 and $S_{i,k} = \langle\hat{D}_i^\dagger\hat{D}_k\rangle - \langle\hat{D}_i^\dagger\rangle\langle\hat{D}_k\rangle$ is the quantum geometric tensor (QGT). QGT is positive definite, and can be inverted to find the solution. Since we calculate the expectation values using Monte Carlo estimates, it might happen that the QGT estimated has zero eigen values, leading to numerical stability issues while inverting. Thus to stabilize it we add a small $(10^{-5} - 10^{-2})$ shift to the diagonal of the QGT.
Thus finally the network parameters are updated using the following equation,

$$\mathbf{w} = \mathbf{w} - \eta\mathbf{S}^{-1}\mathbf{F} \tag{18}$$

4 Results and Discussions

The following are the definitions of the observables calculated.
Spin-spin correlation structure factor,

$$S_f = \frac{1}{L}\sum_{j,k}\hat{\sigma}_j^z\hat{\sigma}_k^z e^{i\pi(j-k)} \tag{19}$$

Total spin,

$$S^2 = \sum_i(\hat{\sigma}_i^x)^2 + (\hat{\sigma}_i^y)^2 + (\hat{\sigma}_i^z)^2 \tag{20}$$

Spin-spin correlation function,

$$C^{\mu\mu}(r) = \langle \frac{1}{L} \sum_j \hat{\sigma}_j^{\mu} \hat{\sigma}_{j+r}^{\mu} \rangle \tag{21}$$

where $\mu = x, y, z$. And it's Fourier transform,

$$S^{\mu\mu}(k) = \sum_r e^{-ikr} C^{\mu\mu}(r) \tag{22}$$

The NQS calculations were done with the help of the python package NetKet [3].
For all the calculations I have used the number of Monte Carlo Samples = 2000, the number of Monte Carlo steps = 1000, learning rate in gradient descend = 10^{-3}, diagonal shift for stochastic reconfiguration = 10^{-2}, and the number of hidden nodes of RBM = L = number of lattice points. The exact ground state is computed using exact diagonalisation and these exact values are used to analyse the performance of the RBM NQS results.
The following results are obtained,

1. Fig-2 shows the error in ground state energy computed from RBM NQS w.r.to exact ground state energy at different frustration values for different numbers of lattice points. It's evident from this plot [2] that RBM NQS performs good for frustration ratios less than 0.5 and above 0.5 the accuracy drops. This is in agreement with the observations made in [4].

2. The evolution of the observables with Monte Carlo steps given in plots fig-345, shows that the observables converge to the exact ground state (global minimum) for $\frac{J_2}{J_1} < 0.5$ but they converge to some excited state (local minimum) for $\frac{J_2}{J_1} > 0.5$ which is the reason for the accuracy drop that we observed. This is also in agreement with the observations made in [4].

3. The plots of correlation function fig-6 and Fourier transform of correlation function fig-7 also shows the same trend in observation that the RBM NQS works well for $\frac{J_2}{J_1} < 0.5$.

5 Conclusion

The performance of RBM NQS for 1D $J_1 - J_2$ Heisenberg model in the frustration regime was analysed. It was found that the RBM NQS performs well for frustration ratios less than 0.5, which is in agreement with [4]. It's shown in [4] that by imposing the translational symmetry of the problem to the RBM wave function one can improve the accuracy of the RBM NQS, which makes it even to perform well for frustration ratios greater than 0.5. In future, I would like to analyse the performance of such symmetrised RBM NQS.

References

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

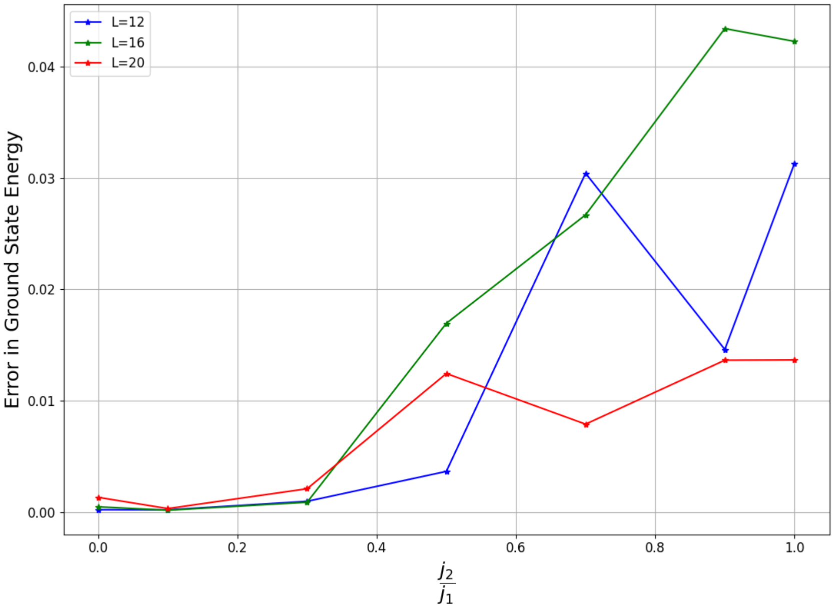
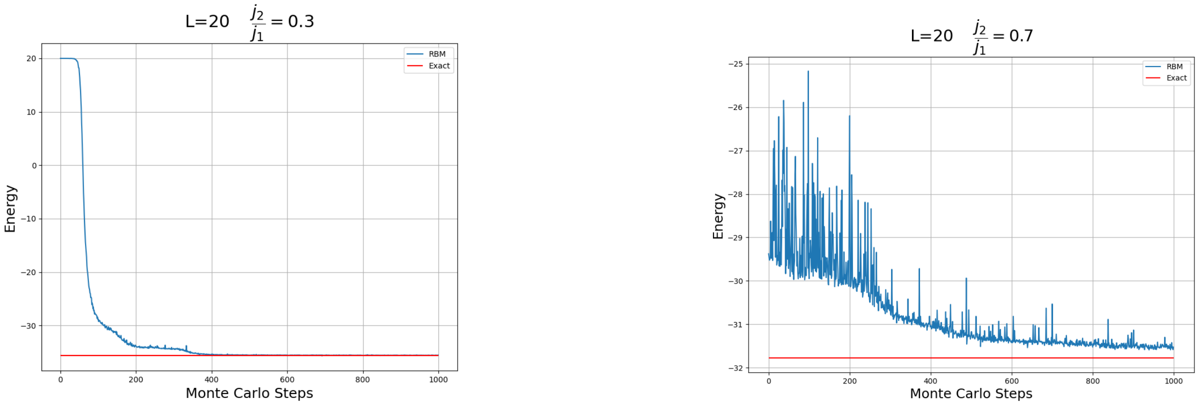


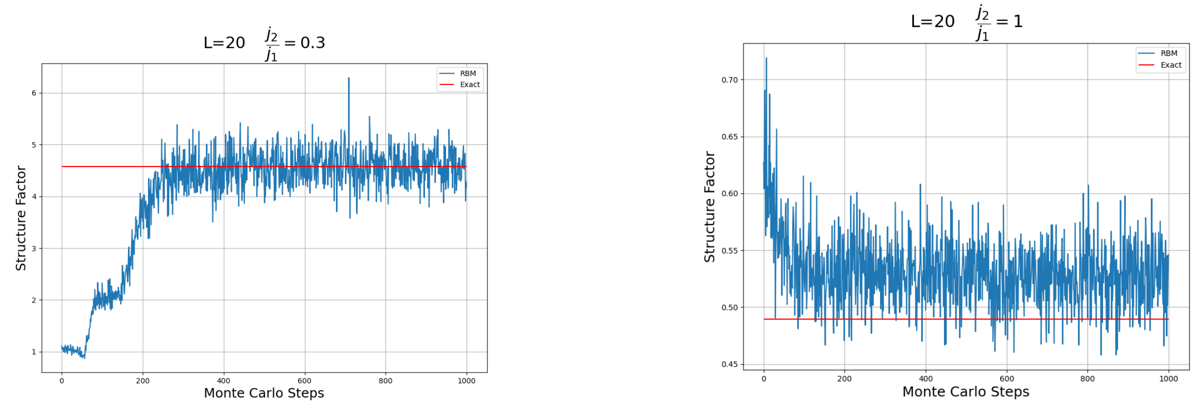
Figure 2: Error in ground state energy computed from RBM NQS w.r.to exact ground state energy at different frustration values for different numbers of lattice points



(a) L=20, $\frac{J_2}{J_1} = 0.3$

(b) L=20, $\frac{J_2}{J_1} = 0.7$

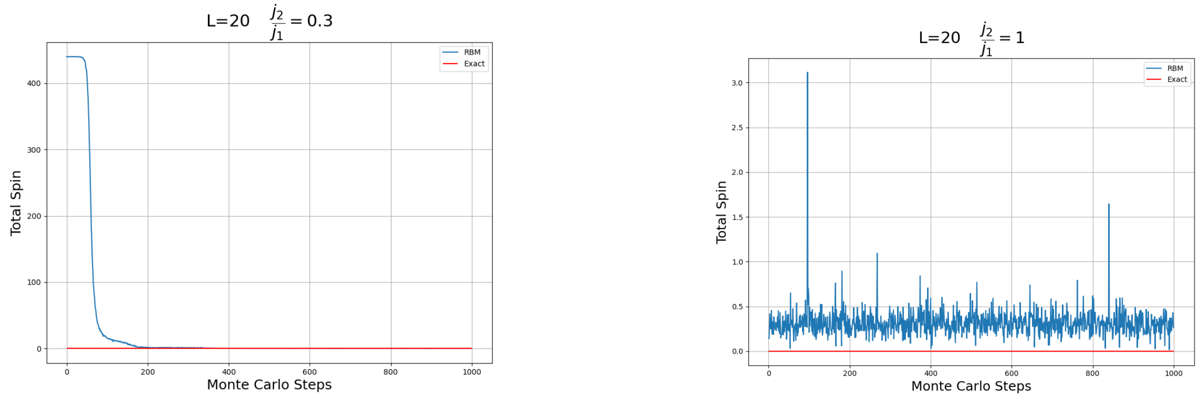
Figure 3: Evolution of ground state energy computed from RBM NQS w.r.to Monte Carlo steps



(a) L=20, $\frac{J_2}{J_1} = 0.3$

(b) L=20, $\frac{J_2}{J_1} = 1$

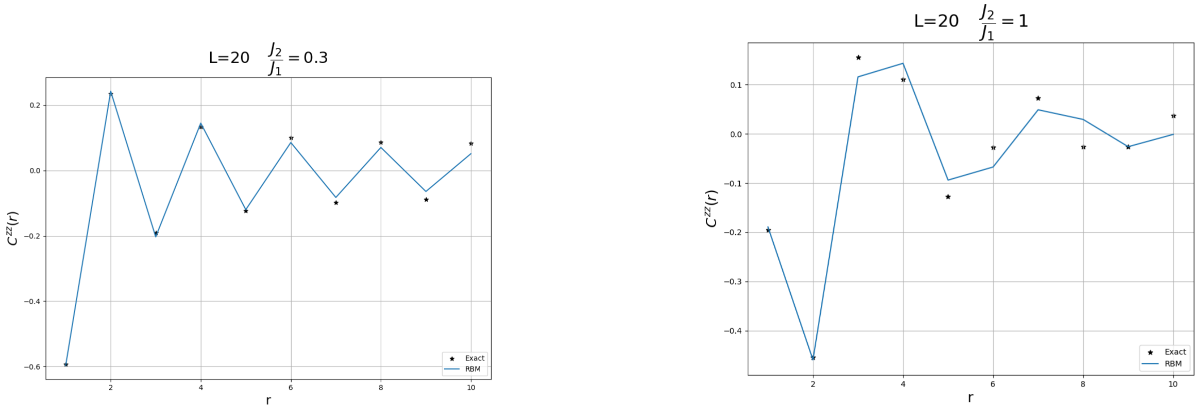
Figure 4: Evolution of spin-spin correlation structure factor [19] computed from RBM NQS w.r.to Monte Carlo steps



(a) L=20, $\frac{J_2}{J_1} = 0.3$

(b) L=20, $\frac{J_2}{J_1} = 1$

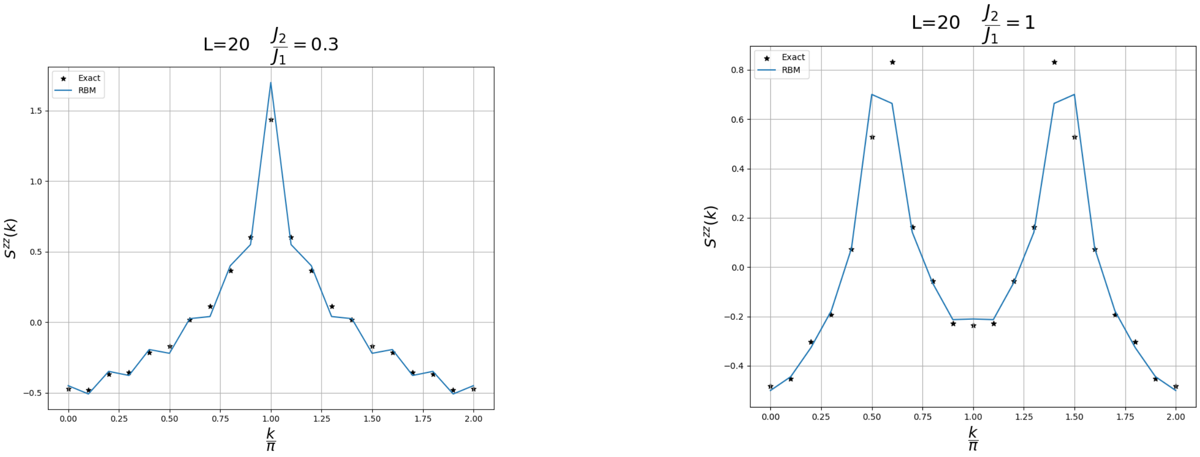
Figure 5: Evolution of total spin [20] computed from RBM NQS w.r.to Monte Carlo steps



(a) L=20, $\dfrac{J_2}{J_1} = 0.3$

(b) L=20, $\dfrac{J_2}{J_1} = 1$

Figure 6: z-z correlation function [21] computed from RBM NQS



(a) L=20, $\dfrac{J_2}{J_1} = 0.3$

(b) L=20, $\dfrac{J_2}{J_1} = 1$

Figure 7: Fourier transform of z-z correlation function [22] computed from RBM NQS

[2] K. Choo, T. Neupert, and G. Carleo. Two-dimensional frustrated ¡mml:math xmlns:mml="http://www.w3.org/1998/math/mathml"¿¡mml:mrow¿¡mml:msub¿¡mml:mi¿j¡/mml:mi¿¡mml:m
model studied with neural network quantum states. *Physical Review B*, 100(12), Sept. 2019.

[3] F. Vicentini, D. Hofmann, A. Szabó, D. Wu, C. Roth, C. Giuliani, G. Pescia, J. Nys, V. Vargas-Calderón, N. Astrakhantsev, and G. Carleo. Netket 3: Machine learning toolbox for many-body quantum systems. *SciPost Physics Codebases*, Aug. 2022.

[4] L. L. Viteritti, F. Ferrari, and F. Becca. Accuracy of restricted boltzmann machines for the one-dimensional $j_1 - j_2$ heisenberg model. *SciPost Physics*, 12(5), May 2022.