Learning ground-state properties with graph neural networks

Oriel Kiss [1, 2, *]

[1] European Organization for Nuclear Research (CERN), Geneva 1211, Switzerland
[2] Department of Nuclear and Particle Physics, University of Geneva, Geneva 1211, Switzerland

Ground state problems are an important part of physics, chemistry and material sciences. Recent theoretical works showed that learning based approach enjoy better performance over standard approaches, since the problem is QMA-hard in general. In this work, we use graph neural networks to learn ground state properties of the XXZ model on a two-dimensional lattice. We empirically study the dependence on the size of training size and level of noise at the data level and show that the model can learn from few or noisy data. This framework can be applied to generalize DMRG calculations, or in conjunction with classical shadows obtained from a quantum computer.

I. INTRODUCTION

Predicting ground-state properties of quantum many-body systems is a key challenge in many areas of physics, such as condensed-matter, high-energy physics, chemistry and material science. The exponential scaling of the Hilbert space makes it a challenging task in general, as representing the full wavefunction requires larger memory capabilities that we could ever build. In certain cases, e.g., if the entanglement present in the wavefunction is bounded, it is still possible to obtain an efficient representation using tensor networks. However, those calculations can be expensive, specially when the system exhibits a large amount of connectivity; it is therefore valuable to find ways to reduce the workload. One promising venue of research is given by machine learning, which can be used to learn an explicit mapping between the Hamiltonian and the target properties. While this has been initially proposed to learn in conjunction with classical shadows [1–3], which is an optimal way to extract information form a quantum computer, we will only focus ground states obtained with the density matrix renormalization group (DMRG) [4, 5] techniques. Later, we will mimic the effect of classical shadows by introducing noise in the training labels.

II. DATASET

We consider a family of Hamiltonian defined on a two-dimensional $L \times L$ lattice

$$H = \sum_{\langle i,j \rangle} J_{ij}^X (X_i X_j + Y_i Y_j) + J_{ij}^Z Z_i Z_j + \sum_i h_i Z_i \quad (1)$$

$$J_{ij}^X, J_{ij}^Z \sim \mathcal{U}(0,1) \quad h_i \sim \mathcal{U}(-2,2), \quad (2)$$

which are refereed as random XXZ Hamiltonian with an external field. The ground states are obtained with DMRG and a bond dimension $\chi = 64$. As target, we consider symmetric correlators of the form

$$C_{ij} = \frac{1}{3}\langle X_i X_j + Y_i Y_j + Z_i Z_j \rangle, \quad (3)$$

which are useful quantities to describe the quantum information and thermalization dynamics, as well as to detect quantum phases transitions [6]. We compute 600 data points, chosen uniformly at random, which are represented as graphs, with the coupling J_{ij}^X and J_{ij}^Z are associated to the corresponding edge, and the external field h_i are one-hot encoded onto the i-th node. The major difference between the earlier papers [2, 3, 7], is that the symmetry along the z-axis is broken and the magnetic field is non-trivial, giving rise to a more richer phase diagram.

III. METHODS

A. Model

Following Tran *et al.* [7], we use graph neural networks [8, 9] to learn a hidden representation of the Hamiltonian, before learning the correlators with a multi-layer perceptrons (MLP) as a regression task. The motivation resides in using the graph structure of the Hamiltonian as an inductive-biased in hope to enjoy better generalisation properties and learning from fewer data points, which are usually expensive to collect. To be more precise, for a system of $N = L^2$ spins, the data consist of $N \times N$ diagonal matrix for the external field, and a $N \times N \times 2$ adjacency matrix with weights $e_{ij} \in [0,1]^2$. The data point x is mapped through the graph convolution [10] to

$$x_i' = \mathrm{MLP}(x_i + \sum_{j \in \mathcal{N}(i)} ((\sigma(x_j + e_{ji}) + \epsilon), \quad (4)$$

where $\sigma(\cdot)$ is the activation function (ReLU by default), $\mathcal{N}(i)$ the set of nodes adjacent to i, and ϵ a small constant added for numerical stabilities. We choose the number of hidden layers as $[N, 64, 128, 128, 64, 64, 64, 32, N]$. The data are then processed by a MLP with number of hidden layers $[N, 32, 64, 128, 64, 32, N]$, producing an image Z of size $N \times N$. Since the correlators are symmetric and

* oriel.kiss@cern.ch

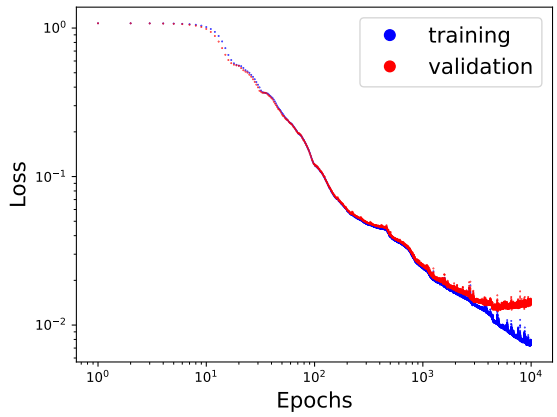


FIG. 1: **Loss:** training (blue) and validation losses as a function of the number of optimization steps in a log-log scale.

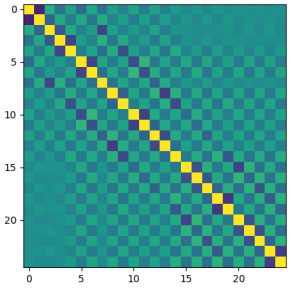
normalized, we do the same by normalizing the output and considering the final output to be ZZ^T. We note that the final activation function is the tanh , since it has the same range as the target. The architecture has been optimized through manual search.

B. Training

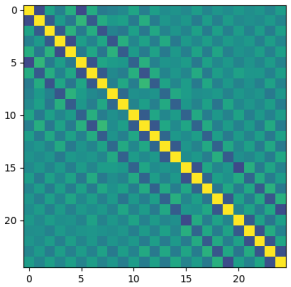
The models are trained using the ADAM optimizer on the L_2 loss, with an initial learning rate of $\text{lr} = 10^{-3}$, which is exponentially reduced every 50 epochs until it reaches 10^{-5}. Moreover, the gradients are clipped to one in order to avoid instabilities. To avoid overfitting, we add a L_1 regularization term with weight 10^{-5}, dropout layers with $p = 0.1$ and an early stopping strategy.

IV. RESULTS

We consider a system of size 5×5 with 600 data points sampled uniformly. We train the model for 10000 epochs, or until the validation loss starts to increase. Fig. 1 shows the evolution of training and validation loss using a $80\% - 20\%$ split, which is selected as the best out of five independent runs. An example from the validation set versus the corresponding prediction is shown in Fig. 2. We observe that the high level features are well reconstructed but discrepancies are still visible.



(a) Data example



(b) Prediction

FIG. 2: **Example** from the validation set and its corresponding prediction by the GNN.

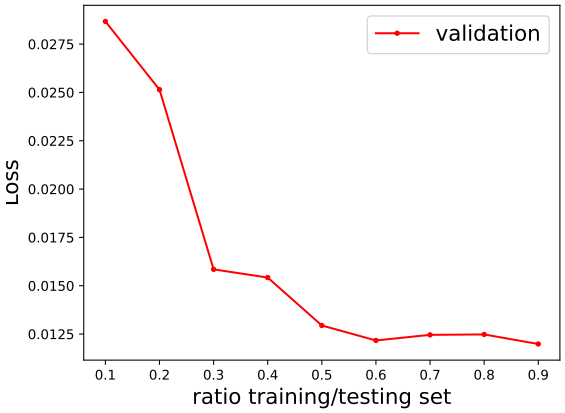


FIG. 3: **Minimum Validation Loss** as a function of the amount of training data.

A. Study on the size of the training set

We study the effect of the amount of training samples. Fig. 3 shows the minimal validation loss (best out of five runs) as a function of the amount of training data. While the loss decreases when given access to more data, as expected, we notice that the improvement quickly saturates. This means that the model only need a reduced amount of data, since the improvement is minimal after using 30%, i.e., 180 data points.

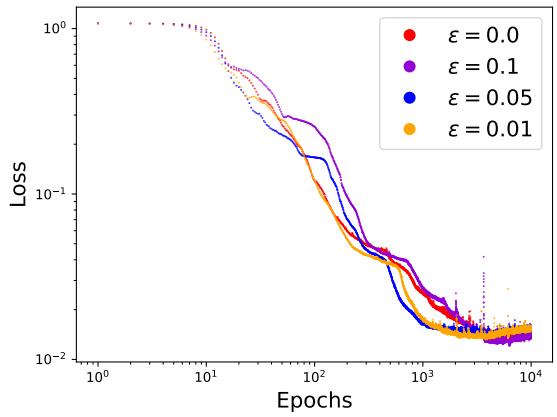


FIG. 4: **Validation Loss:** as a function of the number of optimization steps in a log-log scale. The colors correspond to different level of noise.

B. Study on the level of noise

Target observable are expected to deviate from the ground truth. In the case of DMRG, using a small bond dimension imposes a cutoff on the level of entanglement of the wavefunction, resulting in approximation errors. On the other hand, if the data are collected with classical shadows from a quantum computers, they are subject to statistical noise due to finite amount of measurements. To mimic this effect, we consider the observable as a binomial random variable with parameter $p = \frac{1+C_{ij}}{2}$ and approximate it by drawing $1/\epsilon^2$ samples, where ϵ describes the amount of noise. The model is then trained on noisy data and validated on the true target. The results, shown in Fig. 4, hints that the model is robust to noise, as the same validation loss is obtained at the different noise level.

V. CONCLUSIONS

In this work, we implement a graph neural network to learn ground state properties on XXZ model a 5×5 spin systems. We observe that the model is robust to noise and can learn from few examples. However, it is difficult to obtain lower validation loss, which is motivated by the amount of experiments run by varying the architecture. Most importantly, it seems that the model is only able to learn the symmetric correlators from Eq. 3. Hence, the model is unable to generalise when trained only on $\langle Z_i Z_j \rangle$. This can be problematic as the value of information on the physics of the sytem obtained form the symmetric correlators is limited. We believe that this observable is easier to learn since it has the same form as the Hamiltonian, and is smoother due to symmetry.

While graph neural networks are promising models to learn properties of Hamiltonian with a graph structure, more work needs to be done to better understand which properties can be learned. In addition, it could be beneficial to learn a representation of the state, instead of only the target observable.

[1] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nature Physics **16**, 1050 (2020).

[2] L. Lewis, H.-Y. Huang, V. T. Tran, S. Lehner, R. Kueng, and J. Preskill, Improved machine learning algorithm for predicting ground state properties, Nature Commun. **15**, 1740 (2024).

[3] M. Wanner, L. Lewis, C. Bhattacharyya, D. Dubhashi, and A. Gheorghiu, Predicting Ground State Properties: Constant Sample Complexity and Deep Learning Algorithms, arXiv eprint 2405.18489 (2024).

[4] S. R. White, Density matrix formulation for quantum renormalization groups, Phys. Rev. Lett. **69**, 2863 (1992).

[5] K. G. Wilson, The renormalization group: Critical phenomena and the kondo problem, Rev. Mod. Phys. **47**, 773 (1975).

[6] M. Brenes, S. Pappalardi, M. T. Mitchison, J. Goold, and A. Silva, Out-of-time-order correlations and the fine structure of eigenstate thermalization, Phys. Rev. E **104**, 034120 (2021).

[7] V. T. Tran, L. Lewis, H.-Y. Huang, J. Kofler, R. Kueng, S. Hochreiter, and S. Lehner, Using shadows to learn ground state properties of quantum hamiltonians, Machine Learning and the Physical Sciences workshop, (Neurips22) (2022).

[8] T. N. Kipf and M. Welling, Semi-supervised classification with graph convolutional networks, arXiv preprint arXiv:1609.02907 (2016).

[9] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, Neural message passing for quantum chemistry, in *International conference on machine learning* (PMLR, 2017) pp. 1263–1272.

[10] G. Li, C. Xiong, A. Thabet, and B. Ghanem, Deepergcn: All you need to train deeper gcns, arXiv preprint arXiv:2006.07739 (2020).