Qubit seriation: Improving data-model alignment using spectral ordering

Atithi Acharya [1,3] **Manuel Rudolph** [2,4]

Jing Chen [3] **Jacob Miller** [2] **Alejandro Perdomo-Ortiz** [2]

[1] Department of Physics, Rutgers University, Piscataway, NJ 08854, USA
[2] Zapata Computing Canada Inc., 325 Front St W, Toronto, ON, M5V 2Y1
[3] Zapata Computing Inc., 100 Federal Street, Boston, MA 02110, USA
[4] École Polytechnique Fédérale de Lausanne, Lausanne, CH-1015, Switzerland

Abstract

With the advent of quantum and quantum-inspired machine learning, adapting the structure of learning models to match the structure of target datasets has been shown to be crucial for obtaining high performance. Probabilistic models based on tensor networks (TNs) are prime candidates to benefit from data-dependent design considerations, owing to their bias towards correlations which are local with respect to the topology of the model. In this work, we use methods from spectral graph theory to search for optimal permutations of model sites which are adapted to the structure of an input dataset. Our method uses pair-wise mutual information estimates from the target dataset to ensure that strongly correlated bits are placed closer to each other relative to the model's topology. We demonstrate the effectiveness of such prepossessing for probabilistic modeling tasks, finding substantial improvements in the performance of generative models based on matrix product states (MPS) across a variety of datasets. We also show how spectral embedding, a dimensionality reduction technique from spectral graph theory, can be used to gain further insights into the structure of datasets of interest.

Introduction

The development of increasingly powerful quantum computers has placed renewed focus on the near-term potential of quantum models and algorithms for solving problems of significant real-world value. Probabilistic modeling, where a model is trained to learn the structure of an unknown distribution from an unlabeled dataset of samples, has emerged as an application of particular promise for quantum methods [29], owing to provable advantages in expressivity [12] and generalization [1] arising from the distinct properties of quantum state spaces. Within this domain, the use of quantum-inspired tensor networks (TNs) has allowed many of the advantages of fully-quantum probabilistic models to be enjoyed in a simulated classical setting [14, 5], while also permitting the development of hybrid quantum-classical models that exploit the complementary properties of both model families for practical benefit [13, 18, 36, 32].

Although promising, the relative newness of quantum and quantum-inspired machine learning algorithms means that best practices for ensuring optimal performance remain unsettled. While a large amount of attention has been dedicated to overcoming the phenomenon of *barren plateaus* in optimization landscapes [25, 6, 17], we focus here on a less well-understood issue, namely, unveiling the dataset's geometry on the performance of the machine-learning (ML) models. The importance of optimally matching model geometry to the structure of probabilistic modeling problems are well-understood in the TN community [9, 23, 7], but prior proposals for ensuring this optimal matching

Machine Learning and the Physical Sciences workshop, NeurIPS 2022.

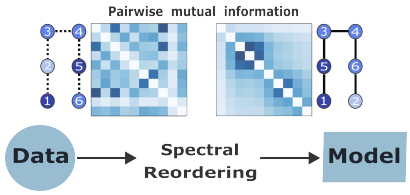

Figure 1: Schematic representation of our qubit seriation framework: Training data is frequently structured such that the most correlated variables in the data are associated with distant sites. Our spectral re-ordering method is able to ensure locality of strongly-correlated variables, as seen in the pairwise mutual information plots, where larger correlations are placed closer to the diagonal. This helps in the training of probabilistic models, notably those based on tensor networks.

have tended to rely on ad-hoc search through various model configurations [21, 16, 22], entailing a high cost due to repeated model retraining, while also failing to make use of insights present in the structure of the classical data itself. A notable exception is the problem-specific solution of Barcza et al. [4], where spectral ordering methods were shown to be capable of improving the performance of density matrix renormalization group (DMRG) calculations within quantum chemistry problems. Recent efforts to incorporate geometric considerations into quantum machine learning (QML), although of a different flavor than considered here, include the works of [26, 20, 30, 27] where the authors incorporate geometric priors arising from problem-specific symmetries into quantum models.

In this work, we introduce a simple method for ordering the variables of a classical dataset to ensure an optimal match between the correlations present in the data and the connectivity of 1D quantum or quantum-inspired models. We refer to this problem as *qubit seriation*, in recognition of its similarity with the seriation problem of linearly-ordering sequential data, which has proven important in domains as diverse as archaeology [31], DNA sequencing [11, 3], and natural language processing [34]. Our approach makes use of spectral graph theory techniques to efficiently compute an ordering directly from the pairwise mutual information between variables in a classical dataset of interest, thus guaranteeing that strongly correlated variables are mapped to nearby qubits, and weakly correlated variables to more distant qubits. We demonstrate the effectiveness of our procedure in probabilistic modeling experiments utilizing matrix product states (MPS), where reordering the variables of a classical dataset prior to optimization is shown to significantly boost the performance of the trained model. We show how spectral embedding tools can be used to extend these methods to models with more complex connectivities, and develop heuristics for understanding the impact of noise or small dataset size on the output ordering. Overall, our work emphasizes the practical importance of geometric considerations in quantum machine learning, and demonstrates the performance benefits that are possible with the use of principled approaches to solving these issues.

Method

We begin by outlining the proof of the distance-sensitive ordering and show convex relaxations of permutation vectors into eigenvectors of the Laplacian of the similarity matrix. Then we use the reordering techniques to substantially improve the performance of the TN models. Given a dataset, we construct a similarity graph Laplacian and find that its lowest eigenvectors, particularly the Fiedler vector, can be used to help in improving the learnability of TNML and QML models. We use spectral embedding, a powerful dimensionality reduction tool to enable model design in more than one dimension. We also study the limitations of our method when it is applied to the realistic case of noisy data. We build indicators for the spectral ordering solutions to analyze their applicability.

We begin by introducing spectral ordering which generates a distance sensitive ordering. It ensures the more strongly correlated sites are kept closer, whereas distant sites are less correlated. The spectral ordering solution is based on a spectral graph theory [35] approach, which uses eigenvectors of the graph Laplacian L coming from the data similarity matrix. As a measure of similarity, we use the mutual information $I(X_i; X_j) = D_{KL}(P(X_i, X_j)||P(X_i)P(X_j))$ where $X_{i,j}$ are data variables on sites i, j and P denotes their marginal probability distributions over the dataset. The graph Laplacian is then defined as $L = D - W$ with $D := (d_{ii}) = (\sum_j I(X_i; X_j))$ representing the diagonal *degree matrix* and $W := (w_{ij})_{i\neq j} = (I(X_i; X_j))_{i\neq j}$ the *weight matrix*. As such, the MI statistics of

the data can be understood as an undirected graph over the data sites with non-negative edges, as $I(X_i; X_j) \geq 0$ is always positive by definition. We are trying to ensure that after the data has been ordered, the adjacent sites have high MI, i.e., they are more strongly correlated, whereas distant sites are less correlated. This can be seen in Fig. 1, where, after ordering, next neighbor sites tend to have large MI values. For data with n sites, and an n-qubit learning problem, the ordering can be defined as the index permutation $\pi(1, 2, ..., n) = (\pi(1), \pi(2), ..., \pi(n))$. The permuted MI weight matrix is $(\pi W \pi^T)_{ij} = w_{\pi(i),\pi(j)}$. One choice of a cost function for qubit seriation i.e. distance sensitive ordering, is

$$C(\pi) = \frac{1}{2}\sum_{i,j}(i-j)^2 w_{\pi(i),\pi(j)} = \frac{1}{2}\sum_{\pi(i),\pi(j)}(i-j)^2 w_{\pi(i),\pi(j)}. \tag{1}$$

While this combinatorial cost function is likely to have no polynomial-time optimal algorithms [8], it can be tackled directly to find good orderings or model topologies. Fascinatingly, it was shown that the second eigenvector of L, commonly called the *Fiedler* vector, provides an optimal 1D solution in the noiseless case [3]. More precisely, the Fiedler vector solves a convex relaxation of C in Eq. 1 with continuous variables $x_i \in [-1, 1]$. Additionally, the shifting necessary in the discrete variables introduces a constraint $\sum_i x_i = 0$.

$$C(\pi) = \frac{1}{2}\sum_{i,j}(x_i - x_j)^2 w_{i,j} = \sum_i x_i^2 d_i - \sum_{i,j} x_i x_j w_{i,j} = \mathbf{x}^T(D - W)\mathbf{x} = \mathbf{x}^T L \mathbf{x}. \tag{2}$$

See Refs. [11, 3] for a similar analysis that was intended for applications in DNA sequencing. This relaxation requires that $\mathbf{x}^T\mathbf{x} = 1$, which can be added a lagrange multiplier λ. The stationary points of this cost function occurs when $\mathbf{x}$ is the eigenvector of the positive semidefinite operator L. Now the cost function reads as $C(\pi) = \mathbf{x}^T L\mathbf{x} - \lambda(\mathbf{x}^T\mathbf{x} - 1)$, which is identical to writing the Rayleigh quotient $R(L, x) = \frac{\mathbf{x}^T L \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$ for L and $\mathbf{x}$, and then minimizing it. We note that the lowest eigenvalue of the Laplacian is always $\lambda_0 = 0$ with the eigenvector being the trivial solution $\mathbf{x}_0 = (1, 1, ..., 1)^T$. The other $n - 1$ eigenvectors are orthogonal to $\mathbf{x}_0$ which then satisfies the $\sum_i x_i = 0$ condition naturally. This helps us in identifying that $\mathbf{x}_1$ minimizes the cost function while satisfying the constraints. By sorting the elements in $\mathbf{x}_1$ in ascending order and re-arranging the indices accordingly, we achieve the optimal ordering for the original qubit seriation problem. Other ways of relaxing the cost function (1) exist and can be exploited for some noisier datasets.

Results and discussions

To put our qubit seriation method based on spectral graph theory to the test, we train a *matrix product state* (MPS) generative model [15, 18] to fit various datasets and demonstrate advantages gained by qubit seriation. We demonstrate the success of the solution to qubit seriation with 1D, 2D and other tree structured data. After training the MPS based generative model using the negative log likelihood (NLL), we monitor the Kullback-Leibler (KL) divergence between the training data distribution $p_{\mathcal{T}}$ and MPS distribution q_θ, given by

$$\mathcal{L}(\theta) = \mathrm{KL}(p_{\mathcal{T}} || q_\theta) = \mathbb{E}_{\mathbf{x}\sim p_{\mathcal{T}}(\mathbf{x})}\left[\log \frac{p_{\mathcal{T}}(\mathbf{x})}{q_\theta(\mathbf{x})}\right] = -\log(|\mathcal{T}|) - \frac{1}{|\mathcal{T}|}\sum_{\mathbf{x}\in\mathcal{T}} \log(q_\theta(\mathbf{x})). \tag{3}$$

The numerics suggest that the qubit seriation leads to a better solution in approximately 99% of the total 1000 experiments for all three chosen datasets. The solution is accepted to be better if $\mathcal{L}_{seriation}$ is lower than $\mathcal{L}_{random}$ by at least 1% of $\mathcal{L}_{random}$. We applied seriation algorithms to samples coming from a random MPS implemented in the ITensor library [10]. It is worth noting that the improvement for random MPS is not as significant as other datasets. This can be seen as a consequence of the lack of significant gaps in the spectrum of the graph Laplacian. We also studied data samples generated from a 1D Markov chain to better understand the role of noise in the stability of spectral solutions. We infer that it is generally better to have a significant spectral gap to have a significant and robust improvement. We also tested the seriation algorithm on the bars and stripes dataset [24], ground state of the Toric code Hamiltonian [19, 33] and samples coming from a random PEPS model [28]. The KL divergence values and their variance are vastly improved using spectral ordering of the data, and for bars and stripes, we recover in all instances that a low bond dimension such as $2^3 = 8$ is sufficient to perfectly fit the dataset. See Fig. 2. We also observed similar improvements in learning the ground state of the Toric code Hamiltonian. These improvements can be qualitatively attributed

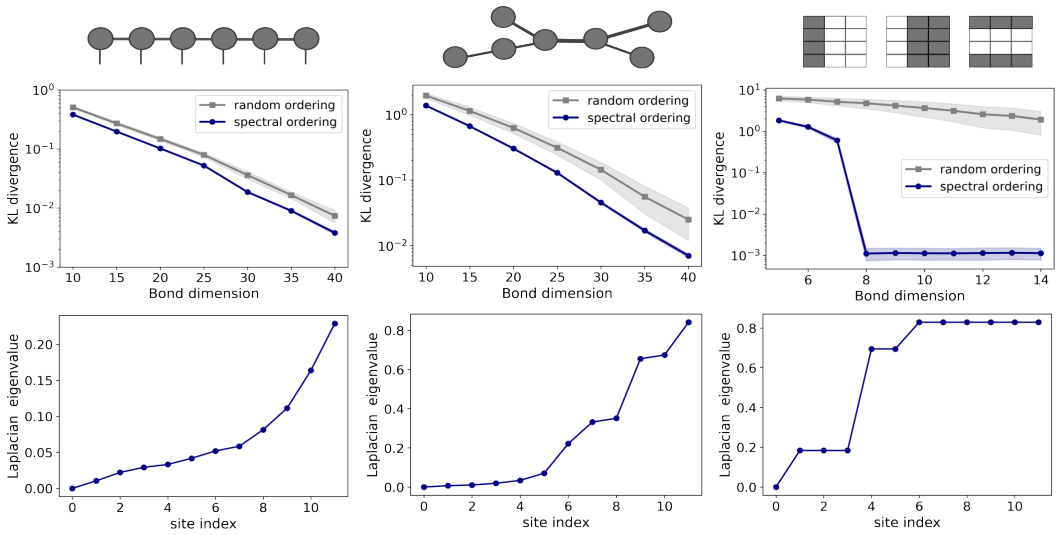

Figure 2: We plot the final KL divergence between the true data distribution and the MPS model distribution. We have data generated by a random MPS to the left, followed by samples from an Ising tree Hamiltonian in the center and finally a dataset of 4×3 bars and stripes images to the right. Results from training on data with random site ordering are marked with gray, while those from training on data which has been rearranged using spectral ordering are marked with blue. Each curve shows the median over 1000 different shufflings of the original dataset, while the fluctuations over these runs are shown as shaded regions around the curve. In all three cases, the spectral ordering solution leads to lower training error with greater than 99% confidence. In the bottom row, we plot the corresponding Laplacian eigenvalues to better understand the different types of training advantages (see text for more details).

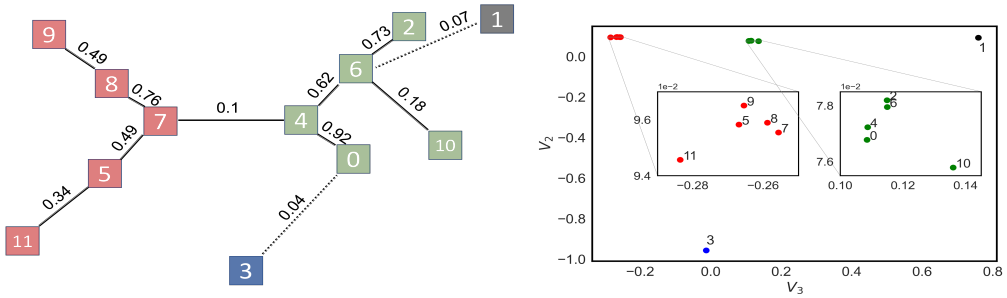

Figure 3: Demonstrating spectral embedding on an Ising tree dataset. (left): The undirected weighted graph is depicted with weights corresponding to the coupling strength J_{ij} of the Ising Hamiltonian.(right): Spectral embedding of the sites in the data using the second and the third eigenvectors of the MI graph Laplacian. We use the MI statistics of 1000 bit-strings corresponding to the ground state configurations of the Ising Hamiltonian on Sz basis. Data sites are colored for convenience. The spectral embedding correctly identifies the more strongly-correlated clusters of qubits 7, 8, 9, 5, and 11, and of qubits 2, 6, 4, 0, and 10. It is important to note that distance in this eigenvector subspace scales inversely with correlation.

to the eigenspectrum of the graph Laplacian associated with the pairwise mutual information matrix. As we see in Fig. 2, there are large gaps between the low and the intermediate eigenvalues, indicating that if we use the eigenvector corresponding to a smaller eigenvalue for spectral ordering, we are guaranteed to ensure an ordering that preserves locality. Lastly, we studied the samples coming from an Ising Tree Hamiltonian using two different methods. We used the best 1D spectral ordering to demonstrate an improvement. We also recovered the tree-like structure where the eigenvectors of the graph Laplacian ensure that the tightly coupled pairs are placed closer than the loosely coupled ones.

The stability of the spectral ordering by the Fiedler vector can be indicated by the spectral gap $\lambda_2 - \lambda_1$ of the Laplacian $L(W)$. In general the spectral gap $(\lambda_k - \lambda_{k-1})$ is used for stability analysis when the algorithm uses the subspace formed by k eigenvectors. If there is degeneracy in the system, then there might be more than one optimal solutions since the solutions form a degenerate subspace. It does not matter which direction is picked within the subspace or if a more general algorithm is employed. However, if there is no gap between the bands of eigenvalues, then it is indicative that there will be no benefit to the algorithm via seriation.

The spectral ordering solution is stable when the magnitude of unstructured noise is less than the spectral gap [11].

$$||\Delta L||_F \leq (\lambda_k - \lambda_{k-1})/\sqrt{2}$$

We also demonstrate that the sampling noise in the estimation of mutual information also leads to the lessening of the spectral gap. Thus further leading to an unstable or less useful solution to seriation (see Fig. 4). Further details can be found in [2], where the authors perform a similar analysis of stability but for closely related spectral clustering problems. We present a proof and interpretation of these results in more detail while also noting the special cases which lead to stronger guarantees on the spectral solution to seriation.

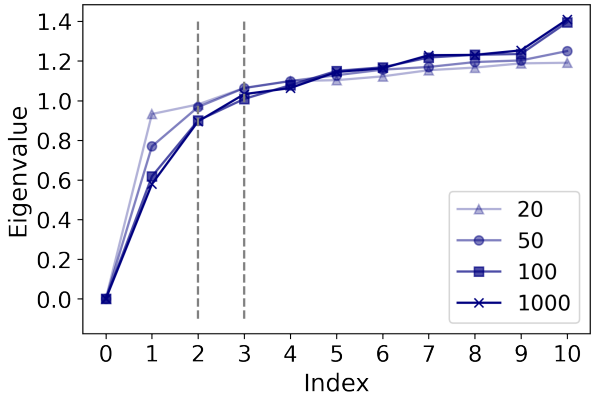

Figure 4: We see the lessening of spectral gap $(\lambda_1 - \lambda_2)$ of the normalized Laplacian if we construct noisier Laplacians by using fewer samples to estimate the pairwise mutual information matrix. This also means that the Fiedler vector is not a unique vector that will seriate the qubits, meaning that slight variations in the data will generate very different permutations as answers to the seriation problem. The data here is taken from a Markov chain, with the legend showing the number of samples used to estimate the pairwise mutual information.

.

Conclusion

In our work, we apply spectral graph theory methods for qubit seriation in tensor network based generative algorithms. Given a dataset, we constructed a similarity graph Laplacian and found that its lowest eigenvectors, and particularly the Fiedler vector, can be used to help TNML and QML in learning better. We also outlined the proof of the distance sensitive ordering and showed convex relaxations of permutation vectors into eigenvectors of the Laplacian of the similarity matrix. While we demonstrated improved training performance using MPS, other models can also benefit from this. Concretely, clustering algorithms on the spectral embedding of data can be used to design TN architectures and quantum circuit ansätze for machine learning tasks that are specific to each given dataset. Given that most negative results are derived using uninformed and generic architectures, we are optimistic that our work can lead to improve the model-data compatibility. We can also leverage the benefits of seriation in generative modeling tasks by classical machine learning models such as recurrent neural networks and its variants, using the fact that sequential learning is sensitive to data ordering [34].

Broader Impact

Overall, our results show the practical benefit of spectral graph theory methods for solving the seriation problem in probabilistic modeling, which we anticipate to be of even greater utility in problems involving larger numbers of random variables and model sizes. While our results were demonstrated with matrix product state models, we expect that these methods can additionally benefit sequential classical models, such as recurrent neural networks, as well as other graph-based probabilistic models, such as probabilistic graphical models and more general tensor networks. For these latter cases, the question of generalizing our methods for the optimal placement of random variables on a general graph (rather than on a line) remains an open question of interest.

References

[1] Amira Abbas, David Sutter, Christa Zoufal, Aurélien Lucchi, Alessio Figalli, and Stefan Woerner. The power of quantum neural networks. Nature Computational Science, 1(6):403–409, 2021.

[2] Eleonora Andreotti, Dominik Edelmann, Nicola Guglielmi, and Christian Lubich. Measuring the stability of spectral clustering, 2019.

[3] Jonathan E. Atkins, Erik G. Boman, and Bruce Hendrickson. A spectral algorithm for seriation and the consecutive ones problem. SIAM Journal on Computing, 28(1):297–310, 1998.

[4] G. Barcza, Ö. Legeza, K. H. Marti, and M. Reiher. Quantum-information analysis of electronic states of different molecular structures. Physical Review A, 83(1), jan 2011.

[5] Tai-Danae Bradley, E M Stoudenmire, and John Terilla. Modeling sequences with quantum states: a look under the hood. Machine Learning: Science and Technology, 1(3):035008, jul 2020.

[6] Marco Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J Coles. Cost function dependent barren plateaus in shallow parametrized quantum circuits. Nature communications, 12(1):1–12, 2021.

[7] Ian Convy, William Huggins, Haoran Liao, and K Birgitta Whaley. Mutual information scaling for tensor network machine learning. Machine learning: science and technology, 3(1):015017, 2022.

[8] Chris Ding and Xiaofeng He. Linearized cluster assignment via spectral ordering. In Proceedings of the twenty-first international conference on Machine learning, page 30, 2004.

[9] Glen Evenbly and Guifré Vidal. Tensor network states and geometry. Journal of Statistical Physics, 145(4):891–918, 2011.

[10] Matthew Fishman, Steven R. White, and E. Miles Stoudenmire. The itensor software library for tensor network calculations, 2020.

[11] Fajwel Fogel, Rodolphe Jenatton, Francis Bach, and Alexandre d'Aspremont. Convex relaxations for permutation problems, 2013.

[12] Xun Gao, Eric R. Anschuetz, Sheng-Tao Wang, J. Ignacio Cirac, and Mikhail D. Lukin. Enhancing generative models via quantum correlations. Phys. Rev. X, 12:021037, May 2022.

[13] Xun Gao, Zhengyu Zhang, and Luming Duan. A quantum machine learning algorithm based on generative models. Science Advances, 4(12), 2018.

[14] Ivan Glasser, Ryan Sweke, Nicola Pancotti, Jens Eisert, and Ignacio Cirac. Expressive power of tensor-network factorizations for probabilistic modeling. Advances in neural information processing systems, 32, 2019.

[15] Zhao-Yu Han, Jun Wang, Heng Fan, Lei Wang, and Pan Zhang. Unsupervised generative modeling using matrix product states. PRX, 8(3):031012, 2018.

[16] Meraj Hashemizadeh, Michelle Liu, Jacob Miller, and Guillaume Rabusseau. Adaptive learning of tensor network structures. arXiv preprint arXiv:2008.05437, 2020.

[17] Zoë Holmes, Kunal Sharma, Marco Cerezo, and Patrick J Coles. Connecting ansatz expressibility to gradient magnitudes and barren plateaus. PRX Quantum, 3(1):010313, 2022.

[18] William Huggins, Piyush Patil, Bradley Mitchell, K Birgitta Whaley, and E Miles Stoudenmire. Towards quantum machine learning with tensor networks. Quantum Science and Technology, 4(2):024001, jan 2019.

[19] A.Yu. Kitaev. Fault-tolerant quantum computation by anyons. Annals of Physics, 303(1):2–30, jan 2003.

[20] Martín Larocca, Frédéric Sauvage, Faris M. Sbahi, Guillaume Verdon, Patrick J. Coles, and M. Cerezo. Group-invariant quantum machine learning. PRX Quantum, 3(3), sep 2022.

[21] Chao Li and Zhun Sun. Evolutionary topology search for tensor network decomposition. In International Conference on Machine Learning, pages 5947–5957. PMLR, 2020.

[22] Chao Li, Junhua Zeng, Zerui Tao, and Qibin Zhao. Permutation search of tensor network structures via local sampling. In International Conference on Machine Learning, pages 13106–13124. PMLR, 2022.

[23] Sirui Lu, Márton Kanász-Nagy, Ivan Kukuljan, and J Ignacio Cirac. Tensor networks and efficient descriptions of classical data. arXiv preprint arXiv:2103.06872, 2021.

[24] David J. C. MacKay. Information Theory, Inference & Learning Algorithms. Cambridge University Press, New York, NY, USA, 2002.

[25] Jarrod Mcclean, Sergio Boixo, Vadim Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. Nature Communications, 9, 11 2018.

[26] Johannes Jakob Meyer, Marian Mularski, Elies Gil-Fuster, Antonio Anna Mele, Francesco Arzani, Alissa Wilms, and Jens Eisert. Exploiting symmetry in variational quantum machine learning. arXiv preprint arXiv:2205.06217, 2022.

[27] Quynh T Nguyen, Louis Schatzki, Paolo Braccia, Michael Ragone, Patrick J Coles, Frederic Sauvage, Martin Larocca, and M Cerezo. Theory for equivariant quantum neural networks. arXiv preprint arXiv:2210.08566, 2022.

[28] Román Orús. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. Annals of physics, 349:117–158, 2014.

[29] Alejandro Perdomo-Ortiz, Marcello Benedetti, John Realpe-Gómez, and Rupak Biswas. Opportunities and challenges for quantum-assisted machine learning in near-term quantum computers. Quantum Science and Technology, 3(3):030502, 2018.

[30] Michael Ragone, Paolo Braccia, Quynh T Nguyen, Louis Schatzki, Patrick J Coles, Frederic Sauvage, Martin Larocca, and M Cerezo. Representation theory for geometric quantum machine learning. arXiv preprint arXiv:2210.07980, 2022.

[31] W. S. Robinson. A method for chronologically ordering archaeological deposits. American Antiquity, 16(4):293–301, 1951.

[32] Manuel S. Rudolph, Jacob Miller, Jing Chen, Atithi Acharya, and Alejandro Perdomo-Ortiz. Synergy between quantum circuits and tensor networks: Short-cutting the race to practical quantum advantage. arXiv preprint arXiv:2208.13673, 2022.

[33] F. Verstraete, M. M. Wolf, D. Perez-Garcia, and J. I. Cirac. Criticality, the area law, and the computational power of projected entangled pair states. Physical Review Letters, 96(22), jun 2006.

[34] Oriol Vinyals, Samy Bengio, and Manjunath Kudlur. Order matters: Sequence to sequence for sets, 2015.

[35] Ulrike Von Luxburg. A tutorial on spectral clustering. Statistics and computing, 17(4):395–416, 2007.

[36] Michael L Wall, Matthew R Abernathy, and Gregory Quiroz. Generative machine learning with tensor networks: Benchmarks on near-term quantum computers. Physical Review Research, 3(2):023010, 2021.

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default **[TODO]** to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [Yes] See Section **??**.
- Did you include the license to the code and datasets? [No] The code and the data are proprietary.
- Did you include the license to the code and datasets? [N/A]

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [No]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [No]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [No]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [No]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...

(a) Did you include the full text of instructions given to participants and screenshots, if applicable? **[TODO]**

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? **[TODO]**

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? **[TODO]**