Sample-efficient learning of quantum many-body systems

Anurag Anshu*†, Srinivasan Arunachalam‡, Tomotaka Kuwahara§¶ and Mehdi Soleimanifar‖
*Institute for Quantum Computing and Department of Combinatorics and Optimization, University of Waterloo
†Perimeter Institute for Theoretical Physics, Waterloo, Canada. aanshu@uwaterloo.ca
‡IBM Research, Yorktown Heights, USA. Srinivasan.Arunachalam@ibm.com
§RIKEN Interdisciplinary Theoretical & Mathematical Sciences Program (iTHEMS)
¶ RIKEN Center for Advanced Intelligence Project (AIP), Tokyo, Japan. tomotaka.kuwahara@riken.jp
‖Center for Theoretical Physics, MIT, Cambridge, USA. mehdis@mit.edu

***Abstract*—We study the problem of learning the Hamiltonian of a quantum many-body system given samples from its Gibbs (thermal) state. The classical analog of this problem, known as learning graphical models or Boltzmann machines, is a well-studied question in machine learning and statistics. In this work, we give the first sample-efficient algorithm for the quantum Hamiltonian learning problem. In particular, we prove that polynomially many samples in the number of particles (qudits) are necessary and sufficient for learning the parameters of a spatially local Hamiltonian in ℓ_2-norm. Our main contribution is in establishing the strong convexity of the log-partition function of quantum many-body systems, which along with the maximum entropy estimation yields our sample-efficient algorithm. Classically, the strong convexity for partition functions follows from the Markov property of Gibbs distributions. This is, however, known to be violated in its exact form in the quantum case. We introduce several new ideas to obtain an unconditional result that avoids relying on the Markov property of quantum systems, at the cost of a slightly weaker bound. In particular, we prove a lower bound on the variance of quasi-local operators with respect to the Gibbs state, which might be of independent interest. Our work paves the way toward a more rigorous application of machine learning techniques to quantum many-body problems.**

***Keywords*-Quantum computing; many-body systems; sample complexity; quantum learning; partition function; strong convexity**

I. Introduction

The success of machine learning algorithms in analyzing high-dimensional data, has resulted in a surge of interest in applying these algorithms to study quantum many-body systems whose description requires dealing with an exponentially large state space. One important problem in this direction is the *quantum Hamiltonian learning* problem, which has been the focus of many recent theoretical and experimental works [1], [2], [3], [4], [5], [6], [7]. Here, one would like to learn the underlying Hamiltonian of a quantum system given multiple identical copies of its Gibbs (thermal) state. The classical analog of this problem is a central problem in machine learning and modern statistical inference, known as *learning graphical models* or *Boltzmann machines* (aka Ising models). Classically, understanding the learnability of Boltzmann machines was initiated by the works of Hinton and others in the 80s [8], [9]. In the past few years, there has been renewed interest in this subject and has seen significant progress resulting in *provably efficient* learning algorithms for graphical models with optimal number of samples and time complexity especially for sparse and bounded-degree graphs [10], [11], [12], [13], [14]. Thus far, a rigorous analysis of the *quantum* Hamiltonian learning problem with guaranteed sample complexity has been lacking. The main contribution of this work is to provide the first *sample-efficient* algorithm for this task.

We now introduce the quantum Hamiltonian learning problem. Consider a κ-local Hamiltonian H acting on n qudits. In general, we can parameterize H by

$$H(\mu) = \sum_{\ell=1}^{m} \mu_\ell E_\ell$$

where $\mu_\ell \in \mathbb{R}$ and $\{E_\ell\}_{\ell=1}^m$ are a set of orthogonal bases for the space of Hermitian operators. For instance in the case of qubits, E_ℓ can be taken as a tensor product of at most κ Pauli operators that act non-trivially only on spatially contiguous qubits. We let the vector $\mu = (\mu_1, \ldots, \mu_m)^\top$ be the vector of *interaction coefficients*. In our setup, without loss of generality we assume the Hamiltonian is traceless, i.e., for the identity operator $E_\ell = \mathbb{1}$, the coefficient $\mu_\ell = 0$. At a *inverse-temperature* β, the qudits are in the *Gibbs state* defined as

$$\rho_\beta(\mu) = \frac{e^{-\beta H(\mu)}}{\mathrm{tr}[e^{-\beta H(\mu)}]}.$$

In the learning problem, we are given multiple copies of $\rho_\beta(\mu)$ and can perform arbitrary *local measurements* on them. From the measurement data, we can obtain all the κ-local *marginals* of $\rho_\beta(\mu)$ denoted by

$$e_\ell = \mathrm{tr}[\rho_\beta(\mu) E_\ell] \quad \textit{for } \ell \in [m].$$

The goal is to learn the coefficients μ_ℓ of the Hamiltonian H using the result of these measurements. We call this the Hamiltonian Learning Problem. Before stating our main results, we provide further motivations to investigate this problem.

Physics perspective: Quantum many-body systems consist of many quantum particles (qudits) that *locally* interact with

2575-8454/20/$31.00 ©2020 IEEE
DOI 10.1109/FOCS46700.2020.00069

each other. The interactions between these particles are described by the Hamiltonian of the system. Even though the interactions in the Hamiltonian are local, the state of the whole system can be highly *entangled*. This leads to intriguing phenomena such as novel phases of matter. Physicists are interested in both synthesizing quantum many-body systems with a specified Hamiltonian and also characterizing the Hamiltonian of an existing material. Algorithms for Hamiltonian learning can directly address these problems and provide a rigorous procedure for inferring the interactions between quantum particles from measurement data.

Verification of quantum devices: The size of the available quantum computers is increasing and they are becoming capable of running more intricate quantum algorithms or preparing highly entangled states over larger number of qubits. Due to the noise in these devices, a major challenge that accompanies the scalable development of quantum devices is to efficiently *certify* their functionality. In recent times, one widely used subroutine in quantum algorithms is *quantum Gibbs sampling*. Preparing and measuring the Gibbs state of a given Hamiltonian is used in quantum algorithms for solving semi-definite programs [15], [16], [17], [18], quantum simulated annealing [19], [20], quantum machine learning [21], [22], or quantum simulations at finite temperature [23]. Given near term quantum devices will be noisy, an important problem when implementing these quantum subroutines is to certify the performance of the quantum Gibbs samplers and to calibrate them. More specifically, it would be ideal to have a *classical* algorithm that verifies from the output of a Gibbs sampler whether the correct Hamiltonian has been implemented.

Quantum machine learning for quantum data: A popular family of models for describing classical distributions are *graphical models* or *Markov random fields*. These models naturally encode the causal structure between random variables and have found widespread applications in various areas such as social networks, computer vision, signal processing, and statistics (see [14] for a survey). A simple and well-studied example of such a family is the classical *Ising model* (i.e. *Boltzmann machine*) defined over a graph whose vertices correspond to the random variables x_i. A natural distribution that one can associate to this model is

$$\mathrm{pr}[X = x] = \frac{1}{Z} \exp\Big(\sum_{i \sim j} J_{ij} x_i x_j + \sum_i h_i x_i \Big) \qquad (1)$$

where $J_{ij}, h_i \in \mathbb{R}$ are real coefficients and the normalization factor Z is called the *partition function*. This distribution in Eq. (1) is also known as the *Gibbs distribution*. There is a rich body of work on learnability of Ising models given samples from the Gibbs distribution. Remarkably, a sequence of works concluded in showing a classical sample-efficient algorithm with a running time quadratic in the number of vertices that outputs estimates of the coefficients J_{ij} and h_i [10], [11], [12]. Similar results have been also proved for more general graphical models.

Considering these achievements in learning theory and the broad practical application of machine learning algorithms, there has been a rising interest in connecting these techniques to problems in quantum computing and many-body physics. This along with other related problems is loosely referred to as *quantum machine learning*. Is there a natural problem that we can rigorously establish such a connection for it? Thus far, the proposals we are aware of in this direction are based on heuristic grounds. One proposal that stands out due to its similarity to the classical case is the problem of learning quantum Ising model (or quantum Boltzmann machine [22]) or more generally the Hamiltonian Learning Problem.

In this paper, we rigorously establish the sample complexity of the Hamiltonian Learning Problem, which asks how many samples are sufficient in order to estimate the Hamiltonian with a desired error. We show that one can reliably learn a Hamiltonian from a sample set whose size is *polynomial* in the number of qudits. To the best of our knowledge, this is the first such result that unconditionally obtains a non-trivial sample complexity. We believe our work opens the doors to further study of this problem using insight from machine learning and optimization theory.

The detailed proofs and further discussion on the topics presented here is given in the full version of this work [24].

II. Main results

Motivated by these applications, we now formally define the Hamiltonian learning problem.

Problem 1 (Hamiltonian learning problem)**.** *Consider a κ-local Hamiltonian $H(\mu) = \sum_{\ell=1}^{m} \mu_\ell E_\ell$ that acts on n qudits and consists of m local terms such that $\max_{\ell \in [m]} |\mu_\ell| \leq 1$. In the Hamiltonian Learning Problem, we are given N copies of the Gibbs state of this Hamiltonian*

$$\rho_\beta(\mu) = \frac{e^{-\beta H(\mu)}}{\mathrm{tr}[e^{-\beta H(\mu)}]}$$

at a fixed inverse-temperature β. Our goal is to obtain an approximate $\hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_m)$ for the coefficients μ such that with probability at least $1 - \delta$,

$$\|\mu - \hat{\mu}\|_2 \leq \varepsilon,$$

where $\|\mu - \hat{\mu}\|_2 = \big(\sum_{\ell=1}^{m} |\mu_\ell - \hat{\mu}_\ell|^2\big)^{\frac{1}{2}}$ is the ℓ_2-norm of the difference of μ and $\hat{\mu}$.

Our main result is a sample-efficient algorithm for the Hamiltonian Learning Problem.

Theorem 2 (Sample-efficient Hamiltonian learning)**.** *The Hamiltonian Learning Problem 1 can be solved using*

$$N = \mathcal{O}\left(\frac{e^{\mathcal{O}(\beta^c)}}{\beta^{\tilde{c}} \varepsilon^2} \cdot m^3 \cdot \log\left(\frac{m}{\delta}\right) \right) \qquad (2)$$

copies of the Gibbs state $\rho_\beta(\mu)$, where $c, \tilde{c} \geq 1$ are constants that depend on the geometry of the underlying graph.

Our work establishes the first unconditional and rigorous upper bounds on the sample complexity of the Hamiltonian Learning Problem. For spatially local Hamiltonians, the number of interaction terms m scales as $O(n)$. Hence, our result in Theorem 2 implies a sample complexity polynomial in the number of qudits.

We remark that the number of samples in (2) diverges to infinity as $\beta \to \infty$ or $\beta \to 0$. As the temperature increases ($\beta \to 0$), the Gibbs state approaches the maximally mixed state independent of the choice of parameters μ. At low temperatures ($\beta \to \infty$), the Gibbs state is in the vicinity of the ground space, which for instance, could be a product state $|0\rangle^{\otimes n}$ for the various choices of μ. In either cases, more sample are required to distinguish the parameters μ.

To complement our upper bound, we also obtain a $\Omega(\sqrt{m})$ lower bound for the Hamiltonian Learning Problem with ℓ_2 norm using a simple reduction to the state discrimination problem [25] . Hence, our upper bound in Theorem 2 is tight up to polynomial factors.

Theorem 3 (Lower bound on sample complexity)**.** *The number of copies N of the Gibbs state needed to solve the Hamiltonian Learning Problem and outputs a $\hat{\mu}$ satisfying $\|\hat{\mu} - \mu\|_2 \leq \varepsilon$ with probability $1 - \delta$ is lower bounded by*

$$N \geq \Omega\Big(\frac{\sqrt{m} + \log(1-\delta)}{\beta\varepsilon}\Big).$$

The proof of these results is provided in the full version of the paper [24].

Here, we give a brief summary of some of the ideas and techniques used to prove Theorem 2.

In statistical learning theory, a conventional method for obtaining the parameters of a probability distribution from data relies on the concepts of *sufficient statistics* and the *maximum entropy estimation*. Suppose $p(x;\mu)$ is a family of probability distributions parameterized by μ that we want to learn. This family could for instance be various normal distributions with different mean or variance. Let $X_1, \ldots, X_m \sim p(x;\mu)$ be m samples from a distribution in this family. A sufficient statistic is a function T of these samples $T(X_1, \ldots, X_m)$ such that conditioned on that, the original date set $X_1, \ldots, X_m$ does not depend on the parameter μ. For example, the sample mean and variance are well known sufficient statistic functions.

Given classical samples of a data set, a learning algorithm first obtains the sufficient statistic of this data set. Once we achieve this, there is a well-motivated algorithm to estimate the unknown parameter μ is the following: find the distribution, with the maximum Shannon entropy, that matches the observed statistic $T(x)$. The reason for "searching" over the largest entropy distribution is because, intuitively, this estimate will be the least biased estimation for a given set of samples from a data set [26], [27]. This technique (which is related also the maximum likelihood estimation problem), is often used for analyzing the complexity of classical statistical problems.

We first observe that the techniques can also be extended naturally to the quantum Hamiltonian Learning problem [28]. The maximum entropy principle has appeared earlier in [29] to learn the thermal state from local measurements and the work of Brandao and Svore [17] in the context of quantum algorithms for semi-definite programming. In this direction, we first formally show that the trace marginals $\mathrm{tr}[E_\ell \rho]$ for all $\ell \in [m]$ provide a sufficient statistic for the Hamiltonian Learning Problem.

Similar to the classical case discussed above, one implication of this is a method for learning the Hamiltonian H: first measure all the κ-local marginals of the Gibbs state e_ℓ, then among all the states, find the one that matches those marginals. Finding such a state can be naturally formulated in terms of an optimization problem known as the *maximum entropy problem*:

$$\begin{aligned} \max_{\sigma} \quad & S(\sigma) \\ \text{s.t.} \quad & \mathrm{tr}[\sigma E_\ell] = e_\ell, \quad \forall \ell \in [m] \\ & \sigma \succeq 0, \quad \mathrm{tr}[\sigma] = 1. \end{aligned} \tag{3}$$

where $S(\sigma) = -\mathrm{tr}[\sigma \log \sigma]$ is the *von Neumann entropy* of the state σ. The optimal solution of this program is a quantum state with a familiar structure [28]. Namely, it is a Gibbs state $\rho(\lambda)$ for some set of coefficients $\lambda = (\lambda_1, \ldots, \lambda_m)$. The coefficients λ are the *Lagrange multipliers* corresponding to the dual of this program.

The main technical problem that we address in this work is analyzing the robustness of the optimization in (3) to the statistical error in the marginals. This is necessary since practically we can only obtain estimates $\hat{e}_\ell$ for the marginals (i.e. the empirical averages) instead of the exact values e_ℓ. This is an instance of a *stochastic optimization* which is a well-studied problem in optimization.

One technique for analyzing the sample complexity in stochastic optimization is based on the notion of *strong convexity*. In general a function $f : \mathbb{R}^m \mapsto \mathbb{R}$ is α-strongly convex if its Hessian matrix satisfies $\nabla^2 f(x) \succeq \alpha \mathbb{1}$, where α is some positive constant. We apply the strong convexity framework to the *dual* of the convex program in (3) in two steps:

1) *Proving the strong convexity of the objective function*: This is equivalent to showing that the log-partition function (aka the free energy) is strongly convex. This result is the main technical contribution of our work and is stated in the following theorem:

Proposition 4 (Strong convexity of the log-partition function)**.** *Let $H = \sum_{\ell=1}^m \mu_\ell E_\ell$ be a κ-local Hamiltonian over a finite dimensional lattice with $\|\mu\|_\infty \leq 1$. For a given inverse-temperature β, there are constants $c, c' > 3$*

depending on the geometric properties of the lattice such that

$$\nabla^2 \log Z_\beta(\mu) \succeq e^{-\mathcal{O}(\beta^c)} \frac{\beta^{c'}}{m} \cdot \mathbb{1}, \tag{4}$$

i.e., for every vector $v \in \mathbb{R}^m$ *we have* $v^T \cdot \nabla^2 \log Z_\beta(\mu) \cdot v \geq e^{-\mathcal{O}(\beta^c)} \frac{\beta^{c'}}{m} \cdot \|v\|_2^2$.

2) *Bounding the error in estimating* μ *in terms of the error in estimating the marginals* e_ℓ: In this step we show that as long as the statistical error of the marginals is small, using the strong convexity property from step (1), we can prove an upper bound on the difference between the solutions of the ideal and the empirical optimizations. The result can be stated as follows:

Proposition 5 (Error bound from strong convexity)**.** *Let* $\delta, \alpha > 0$*. Suppose the marginals* e_ℓ *are determined up to error* δ*, i.e.,* $|e_\ell - \hat{e}_\ell| \leq \delta$ *for all* $\ell \in [m]$*. Additionally assume* $\nabla^2 \log Z_\beta(\lambda) \succeq \alpha \mathbb{1}$ *and* $\|\lambda\| \leq 1$*. Then the optimal solution to the dual of the optimization in* (3) *satisfies*

$$\|\mu - \hat{\mu}\|_2 \leq \frac{2\beta\sqrt{m}\delta}{\alpha}$$

By substituting the expression for the strong convexity parameter α in Proposition 4 into the bound given in Proposition 5, we obtain our claimed sample complexity in Theorem 2.

The proof of Proposition 4, involves two main ideas. First, we show that the term $v^T \cdot \nabla^2 \log Z_\beta(\mu) \cdot v$ can be lower bounded by the variance of a quasi-local operator. While this can be easily shown for classical Hamiltonians, establishing this for non-commuting Hamiltonians requires extra care. We use the Lieb-Robinson bound [30], [31] and the Belief Propagation framework of [32] to achieve this. The second issue we address is obtaining an $\Omega(1/m)$ lower-bound on this variance. This is technically challenging and is done in multiple steps. We first use the robustness of the spectrum of the Gibbs state to local unitary perturbations (following and extending the techniques of [33]) to prove that the variance of the quasi-local operator is greater than the variance of an operator acting on $O(1)$ qudits. Finally, we show that this latter variance is at least a constant by relating it to the variance of the same operator at infinite temperature.

III. Further discussions

A. Connection to previous work

A similar problem to Hamiltonian Learning Problem known as the shadow tomography has been considered before [34], [35], [17] where instead of the coefficients μ_ℓ, we want to find a state σ that approximately matches $\mathrm{tr}[E_\ell \sigma] \approx_\varepsilon \mathrm{tr}[E_\ell \rho]$ given multiple copies of an unknown state ρ. It was shown $\mathrm{poly}(\log m, \log d^n, 1/\varepsilon)$ copies of ρ are sufficient for tomography. The Hamiltonian Learning Problem differs from the shadow tomography problem. Our goal is to estimate the Hamiltonian (i.e., the coefficients μ_ℓ) within some given error bound. The shadow tomography protocol only concerns with estimating the marginals $\mathrm{tr}[E_\ell \rho]$ up to a fixed error and by itself does not imply a bound on the Hamiltonian. Moreover, since the Hamiltonians we consider are spatially local, we only need to measure *local* observables E_ℓ. This means we do not need to rely on the whole machinery of the shadow tomography which is applicable even when E_ℓ are non-local. We instead use a variant of this method introduced in [36] or other approaches such as those in [37], [38] to estimate $\mathrm{tr}[E_\ell \rho_\beta]$.

There have been a number of proposals for the Hamiltonian Learning Problem in the past. In [1], [6], [39] learning the Hamiltonian from local measurements is considered. Their approach is based on setting up a linear system of equations whose constraints (i.e., the matrix of coefficients) are determined from the measurement outcomes. The solution of these equations is the parameter μ_k of the Hamiltonian. The sample complexity in this approach depends inverse polynomially on the "spectral gap" of the matrix of coefficients which thus far has not been rigorously bounded. Another line of work considers learning the Hamiltonian using a trusted quantum simulator [4], [5], [40] which is analyzed using a combination of numerical evidence and heuristic arguments. Amin et al. [41] quantized classical Boltzmann machines and proposed a method to train and learn quantum Boltzmann machines using gradient descent.

As mentioned earlier, there has been a fruitful series of works on the classical analog of the Hamiltonian Learning Problem (see e.g., [10], [11], [12]). In our work, we assume it is a priori known that the interaction graph of the Hamiltonian is spatially local. We then estimate the parameters in ℓ_2-norm using $\mathrm{poly}(n)$ samples which is polynomially tight even for classical Hamiltonians. If we instead consider estimation in ℓ_∞-norm, the classical algorithms can achieve a stronger result. That is, given $\mathcal{O}(\log n)$ samples, they succeed in efficiently learning the structure of the underlying graph and its parameters in ℓ_∞-norm. If we apply our current analysis to this setup, we cannot improve our $\mathrm{poly}(n)$ sample complexity to $\mathcal{O}(\log n)$. This is in part because the classical results devise a more efficient convex program that learns the parameters node-wise (this relies on the commutativity of the Hamiltonian terms), and partly because their required strong convexity assumptions is based on the Markov property, none of which are known to be quantizable.

B. Open questions

In Section II we explained our approach to analyzing the Hamiltonian Learning Problem based on reducing data to its sufficient statistics and using maximum entropy estimation. An issue with this approach is the blowup in the computational complexity. It is shown in [42] that this approach basically requires approximating the partition function which is

NP-hard. Ideally, one would like to have an algorithm for the Hamiltonian Learning Problem that requires small number of samples, but also has an efficient running time. Satisfying both these constraints for all inverse-temperatures β even in the classical learning problems is quite challenging. It was only recently that more efficient algorithms are devised for learning graphical models [11], [12]. In this work, we focus on the less demanding but still non-trivial question of bounding the sample complexity and leave obtaining an efficient running time for future work. Below we mention some of the open problems in this direction.

It is an interesting open question to improve the dependency of the strong convexity parameter α ideally to a constant independent of system size, assuming physically-motivated conditions such as the decay of correlations or the decay of conditional mutual information. Another approach might be to derive such a bound at high temperatures where powerful tools such as cluster expansions are available [43]. We also expect our bounds can be improved for commuting Hamiltonians. Indeed, using structural results such as [44], [45], one should be able to follow the same strategy as in the classical spin systems to find a constant lower bound on the variance of commuting Hamiltonians.

There are recent results on efficiently computing the partition function of quantum many-body systems under various assumptions [46], [47], [43], [48], [49]. We expect by combining these results with our maximum entropy estimation algorithm, one can obtain efficient classical algorithms for the Hamiltonian Learning Problem. Another approach might be to use calibrated quantum computers (or Gibbs samplers) as in [50], [17] to solve the maximum entropy optimization using multiplicative weight update method and learn the parameters of another quantum device.

Finally, an important future direction is to devise more refined objective functions for the Hamiltonian Learning Problem that matches the performance of the learning algorithms for the classical problem as discussed in Section III-A. Given the non-commutative nature of quantum Hamiltonians, this seems to require substantially new ideas and advances in characterizing the information theoretic properties of the quantum Gibbs states.

ACKNOWLEDGEMENTS.

We thank Aram Harrow, Yichen Huang, Rolando La Placa, Sathyawageeswar Subramanian, John Wright, and Henry Yuen for helpful discussions. Part of this work was done when SA and TK were visiting Perimeter Institute. SA was supported in part by the Army Research Laboratory and the Army Research Office under grant number W1975117. AA is supported by the Canadian Institute for Advanced Research, through funding provided to the Institute for Quantum Computing by the Government of Canada and the Province of Ontario. Perimeter Institute is also supported in part by the Government of Canada and the Province of Ontario. TK was supported by the RIKEN Center for AIP and JSPS KAKENHI Grant No. 18K13475. MS was supported by NSF grant CCF-1729369 and a Samsung Advanced Institute of Technology Global Research Partnership.

REFERENCES

[1] E. Bairey, I. Arad, and N. H. Lindner, "Learning a local hamiltonian from local measurements," *Physical Review Letters*, vol. 122, no. 2, p. 020504, 2019.

[2] E. Bairey, C. Guo, D. Poletti, N. H. Lindner, and I. Arad, "Learning the dynamics of open quantum systems from their steady states," *New Journal of Physics*, 2020.

[3] J. Carrasquilla and R. G. Melko, "Machine learning phases of matter," *Nature Physics*, vol. 13, no. 5, pp. 431–434, May 2017. [Online]. Available: https://doi.org/10.1038/nphys4035

[4] N. Wiebe, C. Granade, C. Ferrie, and D. G. Cory, "Hamiltonian learning and certification using quantum resources," *Physical Review Letters*, vol. 112, no. 19, p. 190501, 2014.

[5] N. Wiebe, C. Granade, C. Ferrie, and D. Cory, "Quantum hamiltonian learning using imperfect quantum resources," *Physical Review A*, vol. 89, no. 4, p. 042314, 2014.

[6] T. J. Evans, R. Harper, and S. T. Flammia, "Scalable bayesian Hamiltonian learning," *arXiv:1912.07636*, 2019.

[7] J. Wang, S. Paesani, R. Santagati, S. Knauer, A. A. Gentile, N. Wiebe, M. Petruzzella, J. G. O'Brien, Jeremy L. Rarity, A. Laing *et al.*, "Experimental quantum hamiltonian learning," *Nature Physics*, vol. 13, no. 6, pp. 551–555, 2017.

[8] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, "A learning algorithm for Boltzmann machines," *Cognitive science*, vol. 9, no. 1, pp. 147–169, 1985.

[9] G. E. Hinton, T. J. Sejnowski *et al.*, "Learning and relearning in Boltzmann machines," *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 1, no. 282-317, p. 2, 1986.

[10] G. Bresler, "Efficiently learning Ising models on arbitrary graphs," in *STOC'15—Proceedings of the 2015 ACM Symposium on Theory of Computing*. ACM, New York, 2015, pp. 771–782.

[11] A. R. Klivans and R. Meka, "Learning graphical models using multiplicative weights," in *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS*, 2017, pp. 343–354.

[12] M. Vuffray, S. Misra, A. Lokhov, and M. Chertkov, "Interaction screening: Efficient and sample-optimal learning of ising models," in *Advances in Neural Information Processing Systems*, 2016, pp. 2595–2603.

[13] L. Hamilton, F. Koehler, and A. Moitra, "Information theoretic properties of markov random fields, and their algorithmic applications," in *Advances in Neural Information Processing Systems*, 2017, pp. 2463–2472.

[14] P. Ravikumar, M. J. Wainwright, J. D. Lafferty *et al.*, "High-dimensional ising model selection using l1-regularized logistic regression," *The Annals of Statistics*, vol. 38, no. 3, pp. 1287–1319, 2010.

[15] F. G. Brandao and K. M. Svore, "Quantum speed-ups for solving semidefinite programs," in *Foundations of Computer Science (FOCS), 2017 IEEE 58th Annual Symposium on*. IEEE, 2017, pp. 415–426.

[16] J. van Apeldoorn, A. Gilyén, S. Gribling, and R. de Wolf, "Quantum SDP-solvers: Better upper and lower bounds," *Quantum*, vol. 4, p. 230, 2020.

[17] F. G. S. L. Brandão, A. Kalev, T. Li, C. Y.-Y. Lin, K. M. Svore, and X. Wu, "Quantum SDP Solvers: Large Speed-Ups, Optimality, and Applications to Quantum Learning," in *46th International Colloquium on Automata, Languages, and Programming (ICALP)*, vol. 132, 2019, pp. 27:1–27:14.

[18] F. G. L. Brandão, R. Kueng, and D. S. França, "Faster quantum and classical SDP approximations for quadratic binary optimization," *arXiv:1909.04613*, 2019.

[19] A. Montanaro, "Quantum speedup of monte carlo methods," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 471, no. 2181, p. 20150301, 2015.

[20] A. W. Harrow and A. Y. Wei, "Adaptive quantum simulated annealing for bayesian inference and estimating partition functions," in *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2020, pp. 193–212.

[21] N. Wiebe, A. Kapoor, and K. M. Svore, "Quantum deep learning," *arXiv:1412.3489*, 2014.

[22] M. H. Amin, E. Andriyash, J. Rolfe, B. Kulchytskyy, and R. Melko, "Quantum Boltzmann machine," *Phys. Rev. X*, vol. 8, p. 021050, May 2018. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevX.8.021050

[23] M. Motta, C. Sun, A. T. Tan, M. J. ORourke, E. Ye, A. J. Minnich, F. G. Brandão, and G. K.-L. Chan, "Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution," *Nature Physics*, vol. 16, no. 2, pp. 205–210, 2020.

[24] A. Anshu, S. Arunachalam, T. Kuwahara, and M. Soleimanifar, "Sample-efficient learning of quantum many-body systems," *arXiv preprint arXiv:2004.07266*, 2020.

[25] M. Hayashi, A. Kawachi, and H. Kobayashi, "Quantum measurements for hidden subgroup problems with optimal sample complexity," *Quantum Information and Computation*, vol. 8, pp. 0345–0358, 2008.

[26] E. T. Jaynes, "Information theory and statistical mechanics," *Physical review*, vol. 106, no. 4, p. 620, 1957.

[27] ——, "On the rationale of maximum-entropy methods," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 939–952, 1982.

[28] ——, "Information theory and statistical mechanics. ii," *Physical Review*, vol. 108, no. 2, p. 171, 1957.

[29] B. Swingle and I. H. Kim, "Reconstructing quantum states from local data," *Phys. Rev. Lett.*, vol. 113, p. 260501, Dec 2014.

[30] E. H. Lieb and D. W. Robinson, "The finite group velocity of quantum spin systems," *Communications in Mathematical Physics*, vol. 28, no. 3, pp. 251–257, 1972. [Online]. Available: https://doi.org/10.1007/BF01645779

[31] B. Nachtergaele and R. Sims, "Locality estimates for quantum spin systems," in *New Trends in Mathematical Physics*, V. Sidoravičius, Ed. Dordrecht: Springer Netherlands, 2009, pp. 591–614.

[32] M. B. Hastings, "Quantum belief propagation: An algorithm for thermal quantum systems," *Physical Review B*, vol. 76, no. 20, p. 201102, 2007.

[33] I. Arad, T. Kuwahara, and Z. Landau, "Connecting global and local energy distributions in quantum spin models on a lattice," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2016, no. 3, p. 033301, March 2016. [Online]. Available: https://doi.org/10.1088

[34] S. Aaronson, "Shadow tomography of quantum states," in *STOC'18—Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. ACM, New York, 2018, pp. 325–338. [Online]. Available: https://doi.org/10.1145/3188745.3188802

[35] S. Aaronson and G. N. Rothblum, "Gentle measurement of quantum states and differential privacy," in *STOC'19—Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. ACM, New York, 2019, pp. 322–333.

[36] H.-Y. Huang, R. Kueng, and J. Preskill, "Predicting many properties of a quantum system from very few measurements," *arXiv:2002.08953*, 2020.

[37] J. Cotler and F. Wilczek, "Quantum overlapping tomography," *Physical Review Letters*, vol. 124, no. 10, p. 100401, 2020.

[38] X. Bonet-Monroig, R. Babbush, and T. E. O'Brien, "Nearly optimal measurement scheduling for partial tomography of quantum states," *arXiv:1908.05628*, 2019.

[39] X.-L. Qi and D. Ranard, "Determining a local Hamiltonian from a single eigenstate," *Quantum*, vol. 3, p. 159, 2019.

[40] G. Verdon, J. Marks, S. Nanda, S. Leichenauer, and J. Hidary, "Quantum Hamiltonian-based models and the variational quantum thermalizer algorithm," *arXiv:1910.02071*, 2019.

[41] M. H. Amin, E. Andriyash, J. Rolfe, B. Kulchytskyy, and R. Melko, "Quantum Boltzmann machine," *Physical Review X*, vol. 8, no. 2, p. 021050, 2018.

[42] A. Montanari, "Computational implications of reducing data to sufficient statistics," *Electron. J. Stat.*, vol. 9, no. 2, pp. 2370–2390, 2015.

[43] T. Kuwahara, K. Kato, and F. G. S. L. Brandão, "Clustering of conditional mutual information for quantum gibbs states above a threshold temperature," *Phys. Rev. Lett.*, vol. 124, p. 220601, Jun 2020.

[44] S. Bravyi and M. Vyalyi, "Commutative version of the k-local Hamiltonian problem and common eigenspace problem," *arXiv: quant-ph/0308021*, 2003.

[45] D. Aharonov and L. Eldar, "On the complexity of commuting local Hamiltonians, and tight conditions for topological order in such systems," in *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science—FOCS 2011*. IEEE Computer Soc., Los Alamitos, CA, 2011, pp. 334–343.

[46] S. Bravyi and D. Gosset, "Polynomial-time classical simulation of quantum ferromagnets," *Physical Review Letters*, vol. 119, no. 10, p. 100503, 2017.

[47] A. W. Harrow, S. Mehraban, and M. Soleimanifar, "Classical algorithms, correlation decay, and complex zeros of partition functions of quantum many-body systems," in *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, 2020, pp. 378–386.

[48] E. Crosson and A. W. Harrow, "Rapid mixing of path integral monte carlo for 1d stoquastic hamiltonians," *arXiv preprint arXiv:1812.02144*, 2018.

[49] E. Crosson and S. Slezak, "Classical simulation of high temperature quantum ising models," *arXiv preprint arXiv:2002.02232*, 2020.

[50] F. G. Brandão and M. J. Kastoryano, "Finite correlation length implies efficient preparation of quantum thermal states," *Communications in Mathematical Physics*, pp. 1–16, 2016.