# Sample-efficient learning of quantum many-body systems*

Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, Mehdi Soleimanifar

**Background.** Recent experimental advances in quantum technologies including quantum simulators and programmable quantum devices have led to the realization of interacting quantum systems with unprecedented complexity. These experimental setups provide a direct probe into the intricate microscopic properties of quantum many-body systems and enable implementations of near-term quantum algorithms [BEJ+19, BSK+17, ZPH+17, AAB+19, HSB+18]. In order to validate the performance of these devices or study the properties of the quantum systems that they realize, a major challenge is to devise efficient methods that extract useful information from high dimensional experimental data generated by these devices. An outstanding problem in this direction is the problem of *quantum Hamiltonian learning* [BAL19, BGP+20, WGFC14, EHF19, WPS+17], where the goal is to develop methods for inferring the interactions between quantum particles using a finite number of experimentally feasible measurements.

Machine learning algorithms are natural candidates for efficiently solving the quantum Hamiltonian learning problem. For one, these algorithms have proven to be highly successful for analyzing complex data sets in various fields, and more importantly, a major set of common learning tasks such as learning Boltzmann machines can be cast as inferring a *classical Hamiltonian* (e.g., Ising model) using samples from the Gibbs distribution. There is a rich body of work in this direction, culminating in efficient algorithms for learning such classical models [Bre15, KM17, VMLC16, HKM17, RWL+10]. In spite of the analogies between learning quantum Hamiltonians and these problems, no theoretical framework exists that studies to what extent the quantum nature of interactions affects a direct application of classical learning techniques to quantum problems. Moreover, it has long remained open to devise a quantum Hamiltonian learning method with *provably efficient* performance guarantees. This lack of rigorous results is not a mere technicality and points to a fundamental distinction between quantum and classical interactions: At finite temperatures, quantum systems can violate the Markov property [LP08], the very feature that allows for the machine learning techniques to be successfully applied to inferring classical systems such as the Ising model. Does this mean that quantum interactions cannot be robustly learned with efficient resources?

In this work, we prove that despite this disparity between quantum and classical Hamiltonians, quantum systems exhibit features that can be exploited to devise rigorous learning algorithms for inferring their Hamiltonian. In particular, we show that the log of the partition function in these systems is *strongly convex*. While classically this directly follows from the Markov property, in this paper, we use a completely different route combining various recent techniques from many-body physics, machine learning theory, and quantum information to establish this property. The framework that we construct in our paper shows that polynomially many local measurements at finite temperature are necessary and sufficient for learning a quantum Hamiltonian.

**Problem statement.** We now introduce the setup more formally. Consider a $\kappa$-spatially 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 the operators $E_\ell$ are Hermitian and $\{E_\ell\}$ forms an orthogonal basis for the space of operators. For instance, in the case of qubits, $E_\ell$ are tensor product of at most $\kappa$ Pauli operators that act non-trivially only on spatially contiguous qubits. We let the vector $\mu = (\mu_1, \ldots, \mu_m)$ be the vector of *interaction coefficients*. Here, 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* $\beta$, the qudits are in the *Gibbs state* defined as

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

*ArXiv version: https://arxiv.org/abs/2004.07266 . A preliminary version of this work will appear in FOCS 2020.

In the learning problem, we are given multiple copies of $\rho_\beta(\mu)$ and can perform arbitrary *local measurements* on them. In particular, we can obtain all the $\kappa$-local *marginals* (or the local expectations) of $\rho_\beta(\mu)$ denoted by $e_\ell = \mathrm{tr}[\rho_\beta(\mu)E_\ell]$ for $\ell \in [m]$ and the goal is to learn the coefficients $\mu_\ell$ of the Hamiltonian $H$.

**Problem 1 (Hamiltonian learning problem).** *Consider a $\kappa$-spatially 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 $\max_{\ell\in[m]} |\mu_\ell| \leq 1$. 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 $\beta$. The goal is to obtain an estimate $\hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_m)$ of the coefficients $\mu_k$ such that with probability $\geq 0.99$, $\|\mu - \hat{\mu}\|_2 \leq \varepsilon$, where $\|\mu - \hat{\mu}\|_2 = \left(\sum_{\ell=1}^m |\mu_\ell - \hat{\mu}_\ell|^2\right)^{\frac{1}{2}}$ is the $\ell_2$- difference between $\mu$ and $\hat{\mu}$.*

**Further motivation.** A major challenge that accompanies the scalable development of quantum devices is to efficiently *certify* their functionality. Preparing and measuring the Gibbs state of a given Hamiltonian (*quantum Gibbs sampling*) is a widely used subroutine in various quantum algorithms [BS17, vAGGdW20, BKL+19, vG18, BKF19, Mon15, HW20, TOV+11, WKS14, MST+20]. Given that 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 given samples from a Gibbs sampler determines if the correct Hamiltonian has been implemented. More broadly, a general question in quantum many-body physics is if we can efficiently test or falsify our theories about highly interacting and complex quantum many-body systems. An example of such systems could be a newly synthesized material or a cold atom setup which realizes a novel phase of matter, e.g., the quantum spin liquid phase. A crucial task here is determining if the underlying interactions in such a system is consistent with our theoretical predictions—the Kitaev model in the quantum spin liquid case. An efficient algorithm for the quantum Hamiltonian learning problem can be used to address this question.

**Main results and proof ideas.** Our first result establishes that the Hamiltonian learning problem can be solved using a number of samples that scales *polynomially* 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. This result is stated in the following theorem:

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

$$N = \mathcal{O}\left(\frac{e^{\mathcal{O}(\beta^c)}}{\beta^{\tilde{c}}\varepsilon^2} \cdot m^3 \cdot \log(m)\right) \tag{1}$$

*copies of the Gibbs state $\rho_\beta(\mu)$, where $c, \tilde{c} \geq 1$ are constants that depend on the geometry of the Hamiltonian. In particular, for $\beta = O(1)$, the number of copies is given by $N = \tilde{\mathcal{O}}(m^3/\varepsilon^2)$.*

For spatially local Hamiltonians, the number of interaction terms $m$ scales as $O(n)$. The number of samples in (1) increases 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 $\mu$. 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 $\mu$. In either cases, more sample are required to distinguish the parameters $\mu$. To complement our upper bound, we also obtain a $\Omega(\sqrt{m})$ lower bound for the Hamiltonian learning problem 1 using a simple reduction to the state discrimination problem. Hence, our sample complexity is tight up to polynomial factors.

**Theorem 3 (A lower bound).** *The number of copies $N$ of the Gibbs state needed to solve the Hamiltonian learning problem and output a $\hat{\mu}$ satisfying $\|\hat{\mu} - \mu\|_2 \leq \varepsilon$ with probability $0.99$ is lower bounded by $N \geq \Omega\Big(\sqrt{m}/(\beta\varepsilon)\Big)$.*

**Technical contribution.** A method for learning the Hamiltonian $H$ is based on the following procedures: first measure all the $\kappa$-local marginals of the Gibbs state $e_\ell$, 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*:

$$\max_{\sigma} \quad S(\sigma) \qquad \text{s.t.} \qquad \mathrm{tr}[\sigma E_\ell] = e_\ell \quad \text{for every } \ell \in [m], \quad \sigma \succeq 0, \quad \mathrm{tr}[\sigma] = 1. \tag{2}$$

where $S(\sigma) = -\mathrm{tr}[\sigma \log \sigma]$ is the *von Neumann entropy* of the state $\sigma$. The optimal solution of this program is a quantum state with a familiar structure [Jay57, SK14, BKL+19]. Namely, it is a Gibbs state $\rho(\lambda)$ for some set of coefficients $\lambda = (\lambda_1, \ldots, \lambda_m)$. The coefficients $\lambda$ are the *Lagrange multipliers* corresponding to the dual of this program.

A major technical problem that we address in this work is analyzing the robustness of the optimization in (2) 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 $e_\ell$. As we explain next proving such a robustness is a significant issue connected to the fundamental properties of interacting quantum particles. One technique for analyzing the sample complexity is based on the notion of *strong convexity* originated in the field of stochastic optimization and machine learning. In general, a function $f : \mathbb{R}^m \mapsto \mathbb{R}$ is $\alpha$-strongly convex if its Hessian matrix satisfies $\nabla^2 f(x) \succeq \alpha \mathbb{1}$, where $\alpha$ is some positive constant. We apply the strong convexity framework to the *dual* of the convex program in (2) 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, in particular we show that for every vector $v \in \mathbb{R}^m$, we have

$$v^T \cdot \nabla^2 \log Z_\beta(\mu) \cdot v \geq e^{-\mathcal{O}(\beta^c)} \frac{\beta^{c'}}{m} \cdot \|v\|_2^2, \tag{3}$$

where $c, c'$ are $\mathcal{O}(1)$ constants which depend on the spatial dimension and the details of the Hamiltonian.

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

**Proposition 4 (Error bound from strong convexity).** *Suppose the marginals $e_\ell$ are determined up to error $\delta$, 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 (2) satisfies*

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

By substituting the expression for the strong convexity parameter $\alpha$ in Eq. (3) into the bound given in Proposition 4, we obtain our claimed sample complexity in Theorem 2. The proof of Eq. (3) 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 [LR72, NS09] and the Belief Propagation framework of [Has07] 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 [AKL16]) 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. We defer the proofs of these statements to the main text of the paper.

Our framework based on the strong convexity provides a rigorous approach and a new avenue for studying quantum many-body problems using machine learning and stochastic optimization techniques. Given the interdisciplinary nature of our work, we expect that introducing this connection to the quantum information community will have a broader impact.

## References


[AAB+19] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A Buell, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019.

[AKL16] Itai Arad, Tomotaka Kuwahara, and Zeph Landau. Connecting global and local energy distributions in quantum spin models on a lattice. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(3):033301, March 2016.

[BAL19] Eyal Bairey, Itai Arad, and Netanel H Lindner. Learning a local hamiltonian from local measurements. *Physical Review Letters*, 122(2):020504, 2019.

[BEJ+19] Tiff Brydges, Andreas Elben, Petar Jurcevic, Benoît Vermersch, Christine Maier, Ben P Lanyon, Peter Zoller, Rainer Blatt, and Christian F Roos. Probing rényi entanglement entropy via randomized measurements. *Science*, 364(6437):260–263, 2019.

[BGP+20] Eyal Bairey, Chu Guo, Dario Poletti, Netanel H Lindner, and Itai Arad. Learning the dynamics of open quantum systems from their steady states. *New Journal of Physics*, 2020.

[BKF19] Fernando GS L Brandão, Richard Kueng, and Daniel Stilck França. Faster quantum and classical SDP approximations for quadratic binary optimization. *arXiv:1909.04613*, 2019.

[BKL+19] 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)*, volume 132, pages 27:1–27:14, 2019.

[Bre15] Guy Bresler. Efficiently learning Ising models on arbitrary graphs. In *STOC'15—Proceedings of the 2015 ACM Symposium on Theory of Computing*, pages 771–782. ACM, New York, 2015.

[BS17] Fernando GSL Brandao and Krysta M Svore. Quantum speed-ups for solving semidefinite programs. In *Foundations of Computer Science (FOCS), 2017 IEEE 58th Annual Symposium on*, pages 415–426. IEEE, 2017.

[BSK+17] Hannes Bernien, Sylvain Schwartz, Alexander Keesling, Harry Levine, Ahmed Omran, Hannes Pichler, Soonwon Choi, Alexander S Zibrov, Manuel Endres, Markus Greiner, et al. Probing many-body dynamics on a 51-atom quantum simulator. *Nature*, 551(7682):579–584, 2017.

[EHF19] Tim J Evans, Robin Harper, and Steven T Flammia. Scalable bayesian Hamiltonian learning. *arXiv:1912.07636*, 2019.

[Has07] Matthew B Hastings. Quantum belief propagation: An algorithm for thermal quantum systems. *Physical Review B*, 76(20):201102, 2007.

[HKM17] Linus Hamilton, Frederic Koehler, and Ankur Moitra. Information theoretic properties of markov random fields, and their algorithmic applications. In *Advances in Neural Information Processing Systems*, pages 2463–2472, 2017.

[HSB+18] R Harris, Y Sato, AJ Berkley, M Reis, F Altomare, MH Amin, K Boothby, P Bunyk, C Deng, C Enderud, et al. Phase transitions in a programmable quantum spin glass simulator. *Science*, 361(6398):162–165, 2018.

[HW20] Aram W Harrow and Annie 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*, pages 193–212. SIAM, 2020.

[Jay57] Edwin T Jaynes. Information theory and statistical mechanics. ii. *Physical Review*, 108(2):171, 1957.

[KM17] Adam R. Klivans and Raghu Meka. Learning graphical models using multiplicative weights. In *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS*, pages 343–354. 2017.

[LP08] Matthew S Leifer and David Poulin. Quantum graphical models and belief propagation. *Annals of Physics*, 323(8):1899–1946, 2008.

[LR72] Elliott H. Lieb and Derek W. Robinson. The finite group velocity of quantum spin systems. *Communications in Mathematical Physics*, 28(3):251–257, 1972.

[Mon15] Ashley Montanaro. Quantum speedup of monte carlo methods. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2181):20150301, 2015.

[MST+20] Mario Motta, Chong Sun, Adrian TK Tan, Matthew J O'Rourke, Erika Ye, Austin J Minnich, Fernando GSL Brandão, and Garnet Kin-Lic Chan. Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution. *Nature Physics*, 16(2):205–210, 2020.

[NS09] Bruno Nachtergaele and Robert Sims. Locality estimates for quantum spin systems. In Vladas Sidoravičius, editor, *New Trends in Mathematical Physics*, pages 591–614, Dordrecht, 2009. Springer Netherlands.

[RWL+10] Pradeep Ravikumar, Martin J Wainwright, John D Lafferty, et al. High-dimensional ising model selection using l1-regularized logistic regression. *The Annals of Statistics*, 38(3):1287–1319, 2010.

[SK14] Brian Swingle and Isaac H. Kim. Reconstructing quantum states from local data. *Phys. Rev. Lett.*, 113:260501, Dec 2014.

[TOV+11] Kristan Temme, Tobias J Osborne, Karl G Vollbrecht, David Poulin, and Frank Verstraete. Quantum metropolis sampling. *Nature*, 471(7336):87, 2011.

[vAGGdW20] J. van Apeldoorn, A. Gilyén, S. Gribling, and R. de Wolf. Quantum SDP-solvers: Better upper and lower bounds. *Quantum*, 4:230, 2020.

[vG18] J. van Apeldoorn and A. Gilyén. Improvements in quantum SDP-solving with applications. 2018. arXiv:1804.05058.

[VMLC16] Marc Vuffray, Sidhant Misra, Andrey Lokhov, and Michael Chertkov. Interaction screening: Efficient and sample-optimal learning of ising models. In *Advances in Neural Information Processing Systems*, pages 2595–2603, 2016.

[WGFC14] Nathan Wiebe, Christopher Granade, Christopher Ferrie, and David Cory. Quantum hamiltonian learning using imperfect quantum resources. *Physical Review A*, 89(4):042314, 2014.

[WKS14] Nathan Wiebe, Ashish Kapoor, and Krysta M Svore. Quantum deep learning. *arXiv:1412.3489*, 2014.

[WPS+17] Jianwei Wang, Stefano Paesani, Raffaele Santagati, Sebastian Knauer, Antonio A Gentile, Nathan Wiebe, Maurangelo Petruzzella, John G O'Brien, Jeremy L. Rarity, Anthony Laing, et al. Experimental quantum hamiltonian learning. *Nature Physics*, 13(6):551–555, 2017.

[ZPH+17] Jiehang Zhang, Guido Pagano, Paul W Hess, Antonis Kyprianidis, Patrick Becker, Harvey Kaplan, Alexey V Gorshkov, Z-X Gong, and Christopher Monroe. Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator. *Nature*, 551(7682):601–604, 2017.