



Sample-efficient learning of interacting quantum systems

Anurag Anshu¹✉, Srinivasan Arunachalam²✉, Tomotaka Kuwahara³✉ and Mehdi Soleimanifar⁴✉

Learning the Hamiltonian that describes interactions in a quantum system is an important task in both condensed-matter physics and the verification of quantum technologies. Its classical analogue arises as a central problem in machine learning known as learning Boltzmann machines. Previously, the best known methods for quantum Hamiltonian learning with provable performance guarantees required a number of measurements that scaled exponentially with the number of particles. Here we prove that only a polynomial number of local measurements on the thermal state of a quantum system are necessary and sufficient for accurately learning its Hamiltonian. We achieve this by establishing that the absolute value of the finite-temperature free energy of quantum many-body systems is strongly convex with respect to the interaction coefficients. The framework introduced in our work provides a theoretical foundation for applying machine learning techniques to quantum Hamiltonian learning, achieving a long-sought goal in quantum statistical learning.

The fundamental interactions in materials are long known to be determined by electromagnetic forces between electrons and nuclei, but they are too complicated to be grasped in their entirety. This has shifted the attention to understanding the effective interactions between particles, described by local Hamiltonians, that, if accurately chalked out, can be used to describe a variety of properties of the system and novel phases of matter. Recent advances in experimental techniques allow for the synthesis and study of increasingly complex interacting quantum systems^{1–4}. This prompts an important question: how can we infer the interactions between particles in such complicated systems or certify that their Hamiltonians indeed match the theoretically predicted models? A similar question arises in the context of near-term quantum computers. A major application of quantum computers is to simulate the dynamics or particular states of a quantum system governed by a given Hamiltonian. In particular, the task of preparing the thermal state of a target Hamiltonian, known as quantum Gibbs sampling, is the backbone of many quantum algorithms such as semi-definite programming solvers^{5–7}, quantum simulated annealing^{8,9}, quantum machine learning¹⁰ and quantum simulations at finite temperature¹¹. Because near-term quantum computers will be noisy, an important problem in the scalable development of these devices is to verify their performance and calibrate them. For various quantum algorithms this means devising classical algorithms that, using measurement data, determine if the correct Hamiltonian has been implemented on the quantum device. We refer to this task as the ‘quantum Hamiltonian learning problem’.

Learning the Hamiltonian of a quantum system has a natural classical analogue, known as learning Boltzmann machines or, more generally, graphical models, which is a central problem in machine learning and modern statistical inference. When expressed using physics terminology, Boltzmann machines correspond to Ising models, a prototypical classical spin Hamiltonian, and the learning problem is equivalent to inferring the corresponding Hamiltonian using samples from its Gibbs (thermal) distribution. Owing to the

widespread application of these models¹², a large body of work has been devoted to studying them, resulting in highly efficient methods for learning Boltzmann machines that operate even in the regime where simulating the equivalent Ising model is computationally NP-hard (NP, non-deterministic polynomial-time hard)^{13–17}.

Given the practical importance of quantum Hamiltonian learning and the remarkable achievements of machine learning methods for the classical version of this problem, various theoretical studies^{18–20}, experimental implementations^{21,22} and a great number of heuristic algorithms^{18,23–27} for this task have appeared. Despite these efforts, the primary challenge in the quantum Hamiltonian learning problem has long remained open, that is, to devise learning methods that are provably efficient in terms of the resources required for inferring the interactions with high accuracy. Most importantly, a prerequisite for any useful Hamiltonian learning method is that the number of performed measurements or, equivalently, the number of identically prepared samples of the system, should ideally scale efficiently with the number of particles. The previous results, however, use a number of samples that in general can grow exponentially, lack a rigorous performance guarantee, assume additional control of the system or only apply to special systems with few particles.

A fundamental obstacle to achieving efficient methods for quantum Hamiltonian learning is a striking feature of interacting quantum systems, in contrast their classical counterparts. At thermal equilibrium, classical spin systems always satisfy what is known as the Markov property. That is, the correlations between two distant spins are mediated by the intermediate spins located in between them such that by conditioning on the state of these intermediate spins, the two distant spins become independent of each other (Supplementary Section 4.1 provides a discussion on this). A crucial implication of the Markov property for classical many-body systems is that their Hamiltonian can be robustly learned using local measurements. Surprisingly, quantum spin systems are known to violate the Markov property in its exact form²⁸. This makes it unclear whether, in principle, recovering quantum interactions from the

¹Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, USA. ²IBM T. J. Watson Research Center, Yorktown Heights, NY, USA. ³Mathematical Science Team, RIKEN Center for Advanced Intelligence Project (AIP), Tokyo, Japan. ⁴Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA, USA. ✉e-mail: anuraganshu@berkeley.edu; Srinivasan.Arunachalam@ibm.com; tomotaka.kuwahara@riken.jp; mehdis@mit.edu

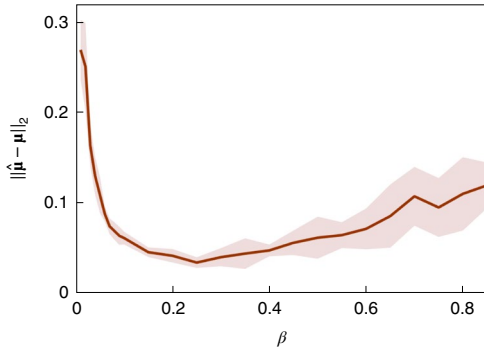


Fig. 1 | Dependence of Hamiltonian learning on the inverse temperature β . The ℓ_2 error in the inferred interaction coefficient $\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|_2$ versus inverse-temperature β is presented for a generic two-body Hamiltonian with site-dependent interactions $H = \sum_{i \sim j} (\mu_{xx}^{ij} X_i X_j + \mu_{yy}^{ij} Y_i Y_j + \mu_{zz}^{ij} Z_i Z_j) + \sum_i (\mu_x^i X_i + \mu_y^i Y_i + \mu_z^i Z_i)$ over a chain of $n=5$ qubits. The data are obtained by implementing the empirical convex optimization (4) using the gradient descent algorithm. The measured local expectations \hat{e}_ℓ are simulated by adding a Gaussian error with standard deviation 10^{-3} to the exact values e_ℓ . At each β , this procedure is repeated ten times and the mean and standard deviation are presented. These data and our theoretical bounds (6) show that for a fixed accuracy δ in the estimated thermal expectations, the error ε in the inferred interaction coefficients increases both for small and large values of the inverse temperature β . This dependency on temperature is essentially optimal^{15,34} and also holds for classical spin systems. This is expected, because the Gibbs state approaches the maximally mixed state for any choice of interactions as $\beta \rightarrow 0$. Moreover, as $\beta \rightarrow \infty$, the system approaches its ground space, which again could be the same state for various choices of interaction coefficients, such as in ferromagnetic systems.

results of local measurements is possible and complicates efforts to extend the machine learning techniques from classical to quantum Hamiltonians.

In this Article, we overcome this difficulty by establishing a structural property of quantum many-body systems at finite temperatures. We show that the log of the partition function (or the free energy) of these systems is strongly convex with respect to the interaction parameters. This results in a framework that, without relying on the Markov property, establishes the robustness against the measurement errors required for the task of Hamiltonian learning. Our framework allows classical learning techniques to be naturally carried over to the quantum case, justifying the applicability of related proposals in quantum machine learning. In particular, we show that an extension of the maximum entropy estimation method, a common tool in machine learning and statistics, can be used to accurately learn the local Hamiltonian from the results of simple local measurements on the thermal state of the system. The number of measurements in our algorithm scales only polynomially in the number of particles. We also prove that a polynomial number of such measurements is necessary for learning the Hamiltonian, rendering our results polynomially tight in the sample complexity. In addition to unconditionally resolving the sample complexity of the quantum Hamiltonian learning, our results imply that the computational resources needed for analysing the measurement outcomes is efficient in various situations. This includes when the temperature of the target system can be adjusted to be higher than a constant threshold value that can be determined in advance.

Efficient sample complexity. Our results in this Article apply to a collection of n quantum particles (qudits) of dimension d that interact with each other via a geometrically local Hamiltonian of the form

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

where E_ℓ form a standard operator basis for κ local interactions, which could be, for instance, the Pauli operators acting on κ neighbouring qudits. The coefficients μ_ℓ , which we collectively refer to as $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)$, are real numbers normalized for convenience to satisfy $-1 \leq \mu_\ell \leq 1$. These coefficients are not necessarily equal, making our set-up also applicable to systems that are not translationally invariant. For geometrically local Hamiltonians as in equation (1), m , the number of interaction coefficients μ_ℓ , scales linearly with the number of particles n , that is, $m = \mathcal{O}(n)$.

Our goal in the quantum Hamiltonian learning problem is to learn the interaction coefficients μ_ℓ for a suitably chosen basis in equation (1). The obtained estimates for these coefficients, which we denote by $\hat{\boldsymbol{\mu}}$, are required to be close to the original values $\boldsymbol{\mu}$ under some notion of distance. We will choose this distance to be the conventional Euclidean or ℓ_2 distance defined as

$$\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\|_2 = \sqrt{\sum_{\ell=1}^m (\mu_\ell - \hat{\mu}_\ell)^2} \leq \varepsilon.$$

Because there are no direct observables for measuring the interaction coefficients $\boldsymbol{\mu}$, inferring them relies on measuring other properties of the system. We assume these measurements are performed at some fixed temperature T while the system is in the Gibbs or thermal state. This state is given by

$$\rho_\beta(\boldsymbol{\mu}) = \frac{e^{-\beta H(\boldsymbol{\mu})}}{Z_\beta(\boldsymbol{\mu})}$$

where $\beta = 1/(k_B T)$ is the inverse temperature, k_B is the Boltzmann constant and $Z_\beta(\boldsymbol{\mu}) = \text{Tr}[e^{-\beta H(\boldsymbol{\mu})}]$ is the partition function of the system. Additionally, we do not presume an additional control of the quantum system such as accessing the dynamics or non-equilibrium states, which have appeared in previous proposals^{25–27}.

We assume an access to N identical samples from the Gibbs state, which can be used to accurately estimate the properties of the system. This is commonly referred to as the ‘sample complexity’ in learning problems. Our main result shows that it suffices to choose N to scale polynomially in the number of interactions m or equivalently the number of particles n . More precisely, we show that

$$N = \beta^{-\tilde{c}} \times \exp(\beta^c) \times e^{-2} \times n^3 \log(n) \quad (2)$$

where $c, \tilde{c} > 4$ depend on the locality κ and the lattice dimension. A natural question is whether the number of measurements given by this bound can be substantially improved, for instance to a polylogarithmic scaling with n . In Supplementary Section 2, we use information theoretic arguments and a reduction to the problem of quantum state discrimination to show this is not the case: we prove that at least $\sqrt{n}/(\varepsilon\beta)$ samples are necessary to learn the coefficients $\boldsymbol{\mu}$ up to error ε . This makes our sample complexity bound polynomially tight. For the rest of the Article, we focus on proving the non-trivial upper bound on the sample complexity.

Learning algorithm. Every algorithm that solves the Hamiltonian learning problem needs to address the following regarding the measurements performed on the system: (1) what observables to measure and (2) how accurately these observables need to be estimated.

To address the first question, we observe that the relevant information or sufficient statistics for inferring the interaction parameters $\boldsymbol{\mu}$ are encoded in the thermal expectations $e_\ell = \text{Tr}[\rho_\beta(\boldsymbol{\mu}) E_\ell]$ for $\ell \in \{1, 2, \dots, m\}$ (refer to refs.^{19,29} or proposition 13 in Supplementary Section 3.1 for the proof of this claim). In other words, the knowledge of the precise value of the local expectations e_ℓ is sufficient for uniquely determining the whole state $\rho_\beta(\boldsymbol{\mu})$ and reconstructing the

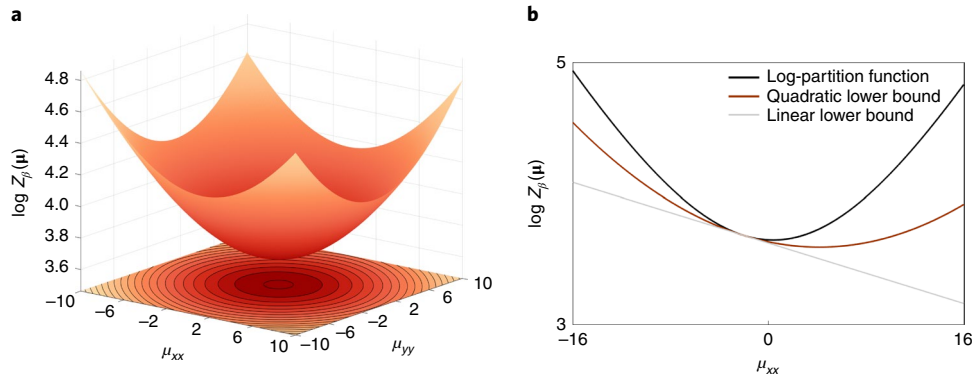


Fig. 2 | Strong convexity of the log-partition function. **a**, The log-partition function as function of the interaction coefficients μ_{xx} and μ_{yy} for the Hamiltonian $H = \sum_{i,j} (\mu_{xx} X_i X_j + \mu_{yy} Y_i Y_j + \mu_{zz} Z_i Z_j) + \sum_i (\mu_x X_i + \mu_y Y_i + \mu_z Z_i)$ over a chain of $n=5$ qubits with $\beta=0.05$, $\mu_{zz}=2$, $\mu_x=\mu_y=\mu_z=3$, and Pauli operators X_i , Y_i and Z_i . It can be seen that this function is convex with respect to the interaction coefficients. **b**, The log-partition function in **a** restricted to the $\mu_{yy}=-2$ plane. Because this function is convex, the tangential line at a point always lower-bounds the function. Strong convexity improves this to a quadratic lower bound (red curve), allowing us to analyse the effect of statistical errors on the optimization (3).

Hamiltonian $H(\mu)$. This is, of course, a special property of the Gibbs state and does not hold for an arbitrary quantum state.

After obtaining the sufficient statistics in terms of local expectations e_ℓ , there is a natural method for obtaining the Gibbs state $\rho_\beta(\mu)$ and estimating the parameter μ : among all the quantum states σ that match the observed expectations e_ℓ (that is, $\text{Tr}[\sigma E_\ell] = e_\ell$), find the one that maximizes the von Neumann entropy $S(\sigma) = -\text{Tr}[\sigma \log \sigma]$. Intuitively, this provides us with the least biased estimate given the current samples and is often referred to as the ‘maximum entropy principle’^{29,30}. This task corresponds to a convex optimization, which can be algorithmically solved. It is more convenient to consider the convex dual of this optimization, which directly yields the exact value of the interaction coefficients μ :

$$\mu = \arg \min_{\substack{\lambda=(\lambda_1, \dots, \lambda_m) \\ \forall \ell: -1 \leq \lambda_\ell \leq 1}} \log Z_\beta(\lambda) + \beta \sum_{\ell=1}^m \lambda_\ell e_\ell \quad (3)$$

where, as before, $Z_\beta(\lambda) = \text{Tr}[e^{-\beta H(\lambda)}]$ is the partition function.

In practice, we are only able to obtain approximate estimates for the expectations e_ℓ and, instead of the exact optimization in equation (3), we can only solve its empirical version. In summary, our learning algorithm consists of the following two steps:

- Step 1 (measurements): repeatedly measure the observables E_ℓ for $\ell \in \{1, \dots, m\}$ to obtain estimates \hat{e}_ℓ for their thermal expectations $e_\ell := \text{Tr}[\rho_\beta(\lambda) E_\ell]$ such that $|e_\ell - \hat{e}_\ell| \leq \delta$ for a fixed accuracy δ .
- Step 2 (post-processing): solve the empirical version of the convex optimization in equation (3) given below using standard methods such as the gradient descent to obtain the estimates $\hat{\mu}$ for the interaction coefficients:

$$\hat{\mu} = \arg \min_{\substack{\lambda=(\lambda_1, \dots, \lambda_m) \\ \forall \ell: -1 \leq \lambda_\ell \leq 1}} \log Z_\beta(\lambda) + \beta \sum_{\ell=1}^m \lambda_\ell \hat{e}_\ell \quad (4)$$

Step 1 can be carried out in several ways. One method for obtaining estimates for the expectations e_ℓ , considered in refs. ^{31,32}, is to group the operators E_ℓ into sets of mutually commuting observables and simultaneously measure them at once. Alternatively, one could use the recent procedure in ref. ³³, based on a variant of shadow tomography. In either case, the number of copies of the state needed

to find all the marginals with accuracy δ and with constant success probability, which could for instance be set as 0.99, is

$$N = 2^{\mathcal{O}(\kappa)} \times \delta^{-2} \times \log m. \quad (5)$$

Optimization (4) in step 2 can be carried out in polynomial time whenever the log-partition function and the entries of its gradient in step 2 can be efficiently computed up to an error of $1/\text{poly}(n)$. In general, achieving such an approximation is known to be NP-hard³⁵, but there are also efficient algorithms if we assume additional conditions on the system. For example, polynomial time algorithms are known if the temperature of the system can be adjusted to be higher than some threshold value that can be determined for any Hamiltonian^{36,37} or when $H(\mu)$ is ‘stochastic’ (or sign-free) and quantum Monte Carlo methods can be applied.

An important remaining issue is to determine the required accuracy δ when estimating the expectations e_ℓ such that an overall error $\|\hat{\mu} - \mu\|_2 \leq \epsilon$ is obtained for the interaction coefficients. As we explain next, this is an important problem connected to the fundamental properties of interacting quantum particles. Our analysis shows that for some constants c, \tilde{c} it holds that

$$\|\hat{\mu} - \mu\|_2 \leq e^{\mathcal{O}(\beta^c)} \times \beta^{-\tilde{c}/2} \times n^{3/2} \times \delta. \quad (6)$$

By using a sufficiently small value of δ , we can make the right-hand side of equation (6) smaller than any desired ϵ . This, along with the number of samples N to achieve an error δ given in the inequality (5), yields our claimed sample complexity in equation (2).

In Fig. 1, we carry out this learning method to infer the Hamiltonian of a spin chain with a generic two-body Hamiltonian. These numerical simulations are in agreement with the predicted performance of our learning method. We now move to theoretically establishing equation (6).

Robustness to statistical error. In this section, we explain how to bound the difference between equation (3) and equation (4) and in turn prove equation (6). We first show why the following simple approach does not work.

As mentioned earlier, the knowledge of the local expectations e_ℓ uniquely determines the Gibbs state $\rho_\beta(\mu)$. In fact, as shown in remark 14 in Supplementary Section 3.1, an extension of this argument implies that if we instead have access to the empirical values \hat{e}_ℓ that are $|\hat{e}_\ell - e_\ell| \leq \delta$, we can approximately determine the Gibbs state up to an error $\mathcal{O}(\sqrt{n\delta})$ in the trace distance.

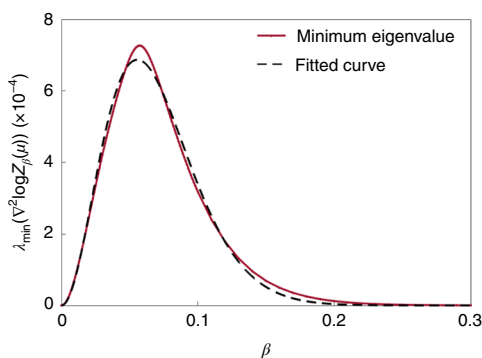


Fig. 3 | Dependence of the strong convexity parameter on the inverse temperature β . The minimum eigenvalue of the Hessian of $\log Z_\beta(\mu)$ is shown for $\beta \in [0, 0.3]$. The Hamiltonian is the translationally invariant Heisenberg model for a chain of five qubits: $H = 10 \times \sum_{i,j} (X_i X_j + Y_i Y_j + Z_i Z_j) + 2 \times \sum_i (X_i + Y_i + Z_i)$. The general behaviour predicted by bound equation (8) is observed here. The fitted curve is of the form $a\beta^{1/2} \exp(-b\beta^2)$ with $a = 0.1062$ and $b = 231.4$, which is similar to the dependence given in equation (8).

However, even if we can obtain a good approximation in terms of the Gibbs state, it does not necessarily imply the Hamiltonian is also inferred with a small error. Because the dependence on the Hamiltonian $H(\mu)$ appears in the exponent $\rho_\beta(\mu) \propto e^{-\beta H(\mu)}$, an initial analysis might suggest that a small error in the estimated Gibbs state can result in an exponentially large error in $\log \rho_\beta(\mu)$ or in the interaction coefficients. This in turn means that the number of measurements N should be exponentially large with n .

To improve this upper bound and establish equation (6), we use the notion of strong convexity, which originates in the fields of convex optimization and machine learning. We begin by giving an intuitive explanation of the strong convexity for a simple univariate function.

Suppose we wish to optimize a quadratic function $f(x) = ax^2 + bx + c$, where the parameter b is inferred from sample data that potentially include statistical errors. The convexity of f is solely quantified by the parameter a . The function is optimized for $x_0 = -b/(2a)$. On the other hand, if b includes a statistical error of δb , the solution changes to $x' = -(b + \delta b)/2a$. To satisfy $|x' - x_0| \ll 1$, we need to ensure $\delta b \ll a$. Therefore, as the convexity parameter becomes weaker, the statistical error needs to be made smaller. This simple intuition can be mathematically generalized to arbitrary optimization problems with multivariate convex functions. In these cases too, the degree of the convexity quantitatively determines how much the minimum point is shifted by the deformation of the function.

We now state things more precisely. By definition, a convex function can be lower-bounded at any point by a tangential line (or hyperplane) $f(y) \geq f(x) + \nabla f(x)^T (y - x)$ where $\nabla f(x)$ is the gradient of the function $f(x)$. Strong convexity improves this to a quadratic lower bound:

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{1}{2} \alpha \|y - x\|_2^2.$$

The coefficient α is determined by the smallest eigenvalue of the Hessian matrix $\nabla^2 f(x)$, which includes the second-order derivatives of $f(x)$, denoted by $\sigma_{\min}(\nabla^2 f(x))$.

In our settings, we think of $f(\cdot)$ as the objective function in optimization (3) or its empirical version equation (4). The strong convexity of the target function in equations (3) and (4) is determined only by the log-partition function $\log Z_\beta(\mu)$. Figure 2 provides

an illustration of the log-partition function of a spin chain with Heisenberg interactions.

Note that modulo a factor of $-1/\beta$, the log-partition function is equal to the free energy of the system. Given this, our goal is to characterize the estimation error $\|\mu - \hat{\mu}\|_2$ in terms of the statistical errors $|\hat{e}_\ell - e_\ell|$ and the strong convexity of $\log Z_\beta(\lambda)$. In Theorem 17 in Supplementary Section 3.2, we prove a clear relationship between these: if the log-partition function $\log Z_\beta(\lambda)$ satisfies the strong convexity with a parameter α , the error $\|\mu - \hat{\mu}\|_2$ is proportional to the inverse of α :

$$\|\mu - \hat{\mu}\|_2 \leq \frac{2\beta\sqrt{m}\delta}{\alpha}, \quad (7)$$

where we assume $|\hat{e}_\ell - e_\ell| \leq \delta$ holds for all ℓ . Therefore, to show that Hamiltonian learning can be achieved with a polynomial number of samples N , we need to prove that $\alpha = \sigma_{\min}(\nabla^2 \log Z_\beta(\mu))$ can only be polynomially small in the number of particles n for any choice of interaction coefficients μ .

Our main technical contribution is in establishing

$$\sigma_{\min}(\nabla^2 \log Z_\beta(\mu)) \geq e^{-\mathcal{O}(\beta^c)} \times \beta^{\tilde{c}/2+1} \times n^{-1}. \quad (8)$$

By applying the above lower bound to equation (7) with $m = \mathcal{O}(n)$, we can immediately obtain the main inequality equation (8). In Fig. 3, we present the result of numerical simulations that illustrates $\sigma_{\min}(\nabla^2 \log Z_\beta(\mu))$ for various inverse-temperatures β in a quantum spin chain with generic two-body interactions. Our predictions in bound equation (8) have a similar functional dependence on β as obtained in the numerical simulations.

Strong convexity for local Hamiltonians. We now give a very high-level sketch establishing the strong convexity in equation (8). We first observe that, in the case of classical spin systems, establishing strong convexity is closely related to a familiar problem in statistical physics, namely to determine the variance of a macroscopic observable in thermal equilibrium. More precisely, it can be shown by using the Markov property of classical Gibbs distributions that such a thermal variance cannot be smaller than a certain bounded value (Supplementary Section 4.1 provides a thorough discussion). This in turn implies a bound on $\sigma_{\min}(\nabla^2 \log Z_\beta(\mu))$ as required by the strong convexity.

Any attempt to extend these ingredients from classical to quantum settings faces some immediate obstacles. Owing to the non-commutativity of the interaction terms E_ℓ , it is not clear if the strong convexity relates to the thermal variance of observables. More importantly, in remarkable contrast to classical Gibbs distributions, quantum Gibbs states can violate the Markov property defined in terms of the quantum conditional mutual information²⁸. As stated earlier, we are able to unconditionally, and without relying on the Markov property, prove the strong convexity for quantum systems as in equation (8), albeit at the cost of a slightly weaker bound than the classical case. We achieve this by introducing several ideas and techniques that are overviewed in Supplementary Section 4.2 and explained in detail in Supplementary Sections 4 and 5.

Outlook. Although we use the maximum entropy estimation to learn the Hamiltonian, our approach based on the strong convexity can also be applied to a wide range of heuristic algorithms. This is especially useful as it is known that carrying out the maximum entropy estimation can be computationally intractable in the worst case³⁵. Such limitations also exist in classical machine learning and apply to learning Boltzmann machines, but, in practice, various other algorithms based on the mean field approximation, belief propagation or pseudo-likelihood estimators are used instead¹². Similar approximate methods have also been proposed in

quantum settings^{38–41}. Our strong convexity result and the error analysis in equation (6) extend to such methods, providing a criterion for evaluating their performance (remark 18 in Supplementary Section 3.2).

We expect our theoretical framework to motivate further development of quantum machine learning algorithms and provide a new approach to study quantum algorithms whose efficiency has so far relied on the Markov property^{38,42}.

Our method only requires simple local Pauli measurements to be performed on the system, making it a suitable algorithm for learning the interactions in various state-of-the-art quantum simulators and devices^{1–3,43,44}. A combination of our techniques and the recent tomography protocols³³ provides feasible methods for certifying various near-term applications of quantum computers, a timely challenge in the field of quantum technologies.

Finally, an important future direction is devising Hamiltonian learning algorithms with both provably efficient sample and time complexities. This is possible in classical spin systems if, instead of the local observables E_i , we allow more sophisticated observables to be measured^{15–17}. Given the non-commutativity of quantum Hamiltonians, extending these results will require substantial new advances in our understanding of the information theoretic properties of the quantum Gibbs states, which is an exciting future direction to pursue.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41567-021-01232-0>.

Received: 4 September 2020; Accepted: 25 March 2021;

Published online: 24 May 2021

References

- Bernien, H. et al. Probing many-body dynamics on a 51-atom quantum simulator. *Nature* **551**, 579–584 (2017).
- Zhang, J. et al. Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator. *Nature* **551**, 601–604 (2017).
- Arute, F. et al. Quantum supremacy using a programmable superconducting processor. *Nature* **574**, 505–510 (2019).
- Simon, J. et al. Quantum simulation of antiferromagnetic spin chains in an optical lattice. *Nature* **472**, 307–312 (2011).
- Brandao, F. G. S. L. & Svore, K. M. Quantum speed-ups for solving semidefinite programs. In *Proc. 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)* 415–426 (IEEE, 2017).
- van Apeldoorn, J., Gilyén, A., Gribling, S. & de Wolf, R. Quantum SDP-solvers: better upper and lower bounds. *Quantum* **4**, 230 (2020).
- Brandão, F. G. S. L., Kueng, R. & França, D. S. Faster quantum and classical SDP approximations for quadratic binary optimization. Preprint at <https://arxiv.org/abs/1909.04613> (2019).
- Montanaro, A. Quantum speedup of Monte Carlo methods. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **471**, 20150301 (2015).
- Harrow, A. W. & Wei, A. Y. Adaptive quantum simulated annealing for Bayesian inference and estimating partition functions. In *Proc. Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (ed. Chawla, S.) 193–212 (SIAM, 2020).
- Wiebe, N., Kapoor, A. & Svore, K. M. Quantum deep learning. *Quantum Inf. Comput.* **16**, 541–587 (2016).
- Motta, M. et al. Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution. *Nat. Phys.* **16**, 205–210 (2020).
- Wainwright, M. J. & Jordan, M. I. Graphical models, exponential families and variational inference. *Found. Trends Mach. Learn.* **1**, 1–305 (2008).
- Chow, C. & Liu, C. Approximating discrete probability distributions with dependence trees. *IEEE Trans. Inf. Theory* **14**, 462–467 (1968).
- Hinton, G. E. et al. Learning and relearning in Boltzmann machines. *Parallel Distrib. Process. Explorations Microstruct. Cognition* **1**, 282–317 (1986).
- Bresler, G. Efficiently learning Ising models on arbitrary graphs. In *Proc. 2015 ACM Symposium on Theory of Computing* 771–782 (ACM, 2015).
- Klivans, A. R. & Meka, R. Learning graphical models using multiplicative weights. In *Proc. 58th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 343–354 (IEEE, 2017).
- Vuffray, M., Misra, S., Lokhov, A. & Chertkov, M. Interaction screening: efficient and sample-optimal learning of Ising models. In *Advances in Neural Information Processing Systems* (eds Lee, D. et al.) 2595–2603 (NIPS, 2016).
- Bairey, E., Arad, I. & Lindner, N. H. Learning a local Hamiltonian from local measurements. *Phys. Rev. Lett.* **122**, 020504 (2019).
- Swingle, B. & Kim, I. H. Reconstructing quantum states from local data. *Phys. Rev. Lett.* **113**, 260501 (2014).
- Qi, X.-L. & Ranard, D. Determining a local Hamiltonian from a single eigenstate. *Quantum* **3**, 159 (2019).
- Wang, J. et al. Experimental quantum Hamiltonian learning. *Nat. Phys.* **13**, 551–555 (2017).
- Senko, C. et al. Coherent imaging spectroscopy of a quantum many-body spin system. *Science* **345**, 430–433 (2014).
- Evans, T. J., Harper, R. & Flammia, S. T. Scalable bayesian Hamiltonian learning. Preprint at <https://arxiv.org/abs/1912.07636> (2019).
- Bairey, E., Guo, C., Poletti, D., Lindner, N. H. & Arad, I. Learning the dynamics of open quantum systems from their steady states. *New J. Phys.* **22**, 032001 (2020).
- Shabani, A., Mohseni, M., Lloyd, S., Kosut, R. L. & Rabitz, H. Estimation of many-body quantum Hamiltonians via compressive sensing. *Phys. Rev. A* **84**, 012107 (2011).
- Wiebe, N., Granade, C., Ferrie, C. & Cory, D. G. Hamiltonian learning and certification using quantum resources. *Phys. Rev. Lett.* **112**, 190501 (2014).
- Wiebe, N., Granade, C., Ferrie, C. & Cory, D. Quantum Hamiltonian learning using imperfect quantum resources. *Phys. Rev. A* **89**, 042314 (2014).
- Leifer, M. S. & Poulin, D. Quantum graphical models and belief propagation. *Ann. Phys.* **323**, 1899–1946 (2008).
- Jaynes, E. T. On the rationale of maximum-entropy methods. *Proc. IEEE* **70**, 939–952 (1982).
- Jaynes, E. T. Information theory and statistical mechanics. *Phys. Rev.* **106**, 620–630 (1957).
- Cotler, J. & Wilczek, F. Quantum overlapping tomography. *Phys. Rev. Lett.* **124**, 100401 (2020).
- Bonet-Monroig, X., Babbush, R. & O’Brien, T. E. Nearly optimal measurement scheduling for partial tomography of quantum states. *Phys. Rev. X* **10**, 031064 (2020).
- Huang, H.-Y., Kueng, R. & Preskill, J. Predicting many properties of a quantum system from very few measurements. *Nat. Phys.* **16**, 1050–1057 (2020).
- Santhanam, N. P. & Wainwright, M. J. Information-theoretic limits of selecting binary graphical models in high dimensions. *IEEE Trans. Inf. Theory* **58**, 4117–4134 (2012).
- Montanari, A. Computational implications of reducing data to sufficient statistics. *Electron. J. Stat.* **9**, 2370–2390 (2015).
- Kuwahara, T., Kato, K. & Brandão, F. G. S. Clustering of conditional mutual information for quantum Gibbs states above a threshold temperature. *Phys. Rev. Lett.* **124**, 220601 (2020).
- Harrow, A. W., Mehraban, S. & Soleimanifar, M. Classical algorithms, correlation decay and complex zeros of partition functions of quantum many-body systems. In *Proc. 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC)* 378–386 (ACM, 2020).
- Poulin, D. & Hastings, M. B. Markov entropy decomposition: a variational dual for quantum belief propagation. *Phys. Rev. Lett.* **106**, 080403 (2011).
- Ferris, A. J. & Poulin, D. Algorithms for the Markov entropy decomposition. *Phys. Rev. B* **87**, 205126 (2013).
- Verdon, G., Marks, J., Nanda, S., Leichenauer, S. & Hidary, J. Quantum Hamiltonian-based models and the variational quantum thermalizer algorithm. Preprint at <https://arxiv.org/abs/1910.02071> (2019).
- Wiebe, N. & Wossnig, L. Generative training of quantum Boltzmann machines with hidden units. Preprint at <https://arxiv.org/abs/1905.09902> (2019).
- G.S.L. Brandão, F. & Kastoryano, M. J. Finite correlation length implies efficient preparation of quantum thermal states. *Commun. Math. Phys.* **365**, 1–16 (2019).
- Brydges, T. Probing Rényi entanglement entropy via randomized measurements. *Science* **364**, 260–263 (2019).
- Harris, R. et al. Phase transitions in a programmable quantum spin glass simulator. *Science* **361**, 162–165 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2021

Data availability

The data presented in the figures are available at <https://github.com/gitmehdis/hamiltonian-learning>.

Code availability

The codes used to generate the figures are available at <https://github.com/gitmehdis/hamiltonian-learning>.

Acknowledgements

We thank A. Harrow, Y. Huang, R. La Placa, S. Subramanian, J. Wright and H. Yuen for helpful discussions. Part of this work was done when S.A. and T.K. were visiting Perimeter Institute. S.A. was supported in part by the Army Research Laboratory and the Army Research Office under grant no. W911NF-20-1-0014. The work was done when A.A. was affiliated to the Institute for Quantum Computing and Department of Combinatorics and Optimization, University of Waterloo and the Perimeter Institute for Theoretical Physics. A.A. was 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. T.K. was supported by the RIKEN Center for AIP and JSPS KAKENHI grant no. 18K13475. M.S. was supported by NSF grant no. CCF-1729369, a Samsung Advanced Institute

of Technology Global Research Cluster and grant no. FXQi-RFP-1811A from the Foundational Questions Institute and Fetzer Franklin Fund, a donor advised fund of Silicon Valley Community Foundation.

Author contributions

A.A., S.A., T.K. and M.S. contributed equally in developing the main ideas, technical aspects and the writing of this work. The authors are arranged alphabetically based on the last name.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41567-021-01232-0>.

Correspondence and requests for materials should be addressed to A.A., S.A., T.K. or M.S.

Peer review information *Nature Physics* thanks Vedran Dunjko and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.