Learning Quantum Hamiltonians at Any Temperature in Polynomial Time

Ainesh Bakshi
Massachusetts Institute of Technology
Cambridge, USA
ainesh@mit.edu

Allen Liu
Massachusetts Institute of Technology
Cambridge, USA
cliu568@mit.edu

Ankur Moitra
Massachusetts Institute of Technology
Cambridge, USA
moitra@mit.edu

Ewin Tang
University of California
Berkeley, USA
ewin@berkeley.edu

ABSTRACT

We study the problem of learning a local quantum Hamiltonian H given copies of its Gibbs state $\rho = e^{-\beta H}/\mathrm{tr}(e^{-\beta H})$ at a known inverse temperature $\beta > 0$. Anshu, Arunachalam, Kuwahara, and Soleimanifar gave an algorithm to learn a Hamiltonian on n qubits to precision ε with only polynomially many copies of the Gibbs state, but which takes exponential time. Obtaining a computationally efficient algorithm has been a major open problem, with prior work only resolving this in the limited cases of high temperature or commuting terms. We fully resolve this problem, giving a polynomial time algorithm for learning H to precision ε from polynomially many copies of the Gibbs state at any constant $\beta > 0$.

Our main technical contribution is a new *flat* polynomial approximation to the exponential function, and a translation between multi-variate scalar polynomials and nested commutators. This enables us to formulate Hamiltonian learning as a polynomial system. We then show that solving a low-degree sum-of-squares relaxation of this polynomial system suffices to accurately learn the Hamiltonian.

CCS CONCEPTS

• **Theory of computation → Quantum information theory**.

KEYWORDS

Hamiltonian learning, Gibbs state, critical temperature, efficient algorithm, polynomial approximation, sum-of-squares, constraint system

ACM Reference Format:
Ainesh Bakshi, Allen Liu, Ankur Moitra, and Ewin Tang. 2024. Learning Quantum Hamiltonians at Any Temperature in Polynomial Time. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing (STOC '24), June 24–28, 2024, Vancouver, BC, Canada.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3618260.3649619

STOC '24, June 24–28, 2024, Vancouver, BC, Canada

ACM ISBN 979-8-4007-0383-6/24/06
https://doi.org/10.1145/3618260.3649619

1 INTRODUCTION

Quantum computing has sparked a major interest in increasing the scale and control of quantum systems [26]. This increased interest is accompanied with the need for better algorithms to characterize and verify these systems [20]. A central computational task in controlling and verifying quantum systems is that of *Hamiltonian learning*, where the goal is to estimate physical properties, namely the interaction strengths, of an interacting quantum many-body system from measurements [6, 8, 21, 22]. Formally, we consider n qubits (quantum particles with a local dimension of two) on a lattice.[1] The resulting system is characterized by a Hamiltonian, a $2^n \times 2^n$ complex Hermitian matrix of the form $H = \sum_{a=1}^{m} \lambda_a E_a$, where a term E_a encodes an interaction on at most $\mathfrak{K}$ of the particles, and the coefficient $\lambda_a \in [-1, 1]$ is the strength of the corresponding interaction. We assume that the system has reached thermal equilibrium at a known inverse temperature β, in which case it is in the Gibbs state with density matrix

$$\rho = \frac{e^{-\beta H}}{\mathrm{tr}\, e^{-\beta H}}.$$

The density matrix is normalized by the partition function, $\mathrm{tr}\, e^{-\beta H}$, which ensures that ρ has trace one. The goal of the Hamiltonian learning problem is to estimate the λ_a's, given the ability to prepare copies of the Gibbs state.

Problem (Hamiltonian learning)**.** Consider n qubits on a constant-dimensional lattice. Let $H = \sum_{a=1}^{m} \lambda_a E_a \in \mathbb{C}^{2^n \times 2^n}$ be a Hamiltonian whose terms E_a are known, distinct, non-identity Pauli operators supported on at most k qubits that are local with respect to the lattice. Further suppose the coefficients $\lambda_a \in \mathbb{R}$ satisfy $|\lambda_a| \leqslant 1$. Given copies of the corresponding Gibbs state ρ at a known inverse temperature $\beta > 0$, and $\varepsilon > 0$, find estimates $\tilde{\lambda}_a$ such that $|\tilde{\lambda}_a - \lambda_a| \leqslant \varepsilon$ for all $a \in [m]$.

We are interested in both the number of copies of ρ that we need, which is called the sample complexity, as well as the running time of the algorithm. Of particular interest to us is Hamiltonian learning in the *low-temperature* regime, where β is an arbitrarily large constant.

[1] Our results do not need the Hamiltonian to be *geometrically local* as described here; we merely require it to be *low-intersection* in the sense of Haah, Kothari, and Tang [30]. So, our algorithm will still work if the locality structure of the qubits is, say, an expander graph.

Motivation. As alluded to above, this problem is of fundamental importance in science and engineering. For example, in pursuit of understanding the phenomena like topological order and superconductivity that are studied in condensed matter physics, experimentalists carefully design systems which exhibit this exotic behavior. In particular, analog quantum simulators are tuned to obey poorly-understood Hamiltonians like the Fermi–Hubbard model for experimental exploration [24, 28, 36].[2] For these experiments, a natural goal is to learn the interactions which give rise to various phenomena [43, 59]. Intractability of computation is a major barrier to resolving open problems like finding the phase diagram of the 2D Fermi–Hubbard model, so having better algorithmic tools is of key importance in this domain [47]. This problem also arises when engineering quantum systems: a major challenge in building near-term quantum devices is being able to validate them—certify that they implement the desired Hamiltonian—and understand sources of error [22, 56]. Quantum devices with 100 or more qubits are challenging to simulate classically, but quantum Hamiltonian learning has emerged as an alternative strategy for benchmarking devices by combining quantum resources and classical learning techniques [20, 29].

The low-temperature setting is of particular interest because because quantum phenomena are most prominent at zero or near-zero temperature [3], precisely where high-temperature series expansions fail [47].[3] In some sense, this is the only relevant setting for analog quantum simulators, since models at high-temperature can be solved with a classical computer [51, Chapter 8], without needing to resort to a quantum simulation. More generally, low temperature is the *computationally* interesting regime, since quantum advantage is a low-temperature phenomenon: "temperature scaling" laws show that quantum annealers can only achieve large speedups over classical computers when β scales with system size [1].

Prior work. Despite its importance, the computational complexity of Hamiltonian learning from Gibbs states is not well understood. Anshu, Arunachalam, Kuwahara, and Soleimanifar gave the first polynomial sample complexity bounds for this task in 2020 [6], attaining coefficient estimates using

$$\frac{2^{\mathrm{poly}(\beta)} m^2 \log m}{\beta^c \varepsilon^2} \qquad [6]$$

copies of the Gibbs state [30, Remark 4.5]. However, their work comes with a serious drawback: it is computationally inefficient. In particular, they give a stochastic gradient descent algorithm and show that it converges to the true parameters in a small number of iterations, but actually computing an iterate involves evaluating a log-partition function, which is well-known to be computationally hard even for classical systems [50].

Prior work has obtained fast algorithms for Hamiltonian learning in limited regimes. A follow-up paper of Anshu, Arunachalam, Kuwahara and Soleimanifar [5] shows that when the terms of H commute, then a direct generalization of the classical algorithm learns the parameters efficiently. Further, [6] notes that their suggested algorithm can be performed in polynomial time for sufficiently high temperature (small β), since in this regime the log-partition function can be evaluated, using that its multivariate Taylor series expansion converges rapidly. Haah, Kothari, and Tang [30] later gave an improved algorithm that achieves the sample and time complexity

$$\frac{e^{\mathcal{O}(\beta)} \log m}{\beta^2 \varepsilon^2} \quad \text{and} \quad \frac{m e^{\mathcal{O}(\beta)} \log m}{\beta^2 \varepsilon^2}, \qquad [30]$$

respectively, which they prove is tight up to the constant factor in the exponential, even in the classical case.

However, a central open question remains [2, 4, 6, 30]:

Question 1. *Can Hamiltonian learning at low temperature be solved in time polynomial in n?*

In practice quantum many-body systems are run at low temperature, which is also when most macroscopic phenomena arise, so this is the most important regime for the problem. However, as we discuss later, no strategies had been suggested for solving Hamiltonian learning at low temperature. In fact, the situation is even more dire: all approaches to Hamiltonian learning used in prior settings fail catastrophically here, since reduction to sufficient statistics [6], efficient computation of the partition function [6], the approximate Markov property [46], and cluster expansion [30] all provably fail for sufficiently large β. This state of the literature reflects a broader scarcity of algorithmic tools known for understanding Hamiltonians outside of special settings like high temperature or one dimension. So, a negative resolution to this question seemed plausible, or even likely. Indeed, a recent survey on the complexity of learning quantum systems by Anshu and Arunachalam [4] discusses Hamiltonian learning and conclude by asking two questions:

Question 2 ([4])**.** Can we achieve Hamiltonian learning under the assumption that the Gibbs states satisfy an approximate conditional independence?[4]

Question 3 ([4])**.** Could low temperature Gibbs states be pseudorandom, which would explain the difficulty in finding a time efficient algorithm?

1.1 Our Results

Surprisingly, we provide a positive resolution to Question 1. Our main result is a computationally efficient algorithm for Hamiltonian learning that works at all temperatures. This is a fortunate development since, if learning were truly computationally hard in the low-temperature regime, then we could not understand the behavior of analog quantum simulators in precisely the regimes where they outperform classical simulators [53, Section 6.10]. As a consequence our main result, we also resolve Question 2 positively and Question 3 negatively.

Theorem 1.1 (Efficiently learning a quantum Hamiltonian (informal)). *Given $\varepsilon > 0$, $\beta \geqslant \beta_c$, for a fixed universal constant*

[2] See also [27, Section 5.4.2] for a description of this work aimed towards theoretical computer scientists.

[3] Morally, these expansions fail precisely because of the non-local quantum correlations we'd like to understand!

[4] Approximate conditional independence is a property of Gibbs states which is proven to hold in 1D and conjectured to hold in general.

$\beta_c > 0$, and $\mathfrak{n}$ copies of the Gibbs state of a low-intersection Hamiltonian, $H = \sum_{a\in[m]} \lambda_a E_a$, there exists an algorithm that runs in time $\mathfrak{n}^{\mathcal{O}(1)}$ and outputs estimates $\left\{ \hat{\lambda}_a \right\}_{a\in[m]}$ such that with probability at least 99/100, for all $a \in [m]$, $\left|\lambda_a - \hat{\lambda}_a\right| \leqslant \varepsilon$, whenever $\mathfrak{n} \geqslant \mathrm{poly}\left(m, (1/\varepsilon)^{2^{\mathcal{O}(\beta)}}\right)$.

Remark 1.2 (On temperature)**.** For our algorithm, we only need to know an upper bound on β, as we can consider a Gibbs state at temperature β to be a Gibbs state at temperature, say, 2β with Hamiltonian $H/2$. Our requirement that $\beta > \beta_c$ is for simplicity, and β_c can be any constant bounded away from zero. In particular, we can take β_c to be the temperature at which the high-temperature algorithm of [30] fails; so, when $\beta < \beta_c$, we can simply appeal to [30] to achieve a sample and time complexity of $\frac{\log m}{\beta^2\varepsilon^2}$ and $\frac{m\log m}{\beta^2\varepsilon^2}$, respectively.

As noted by prior work [6], Hamiltonian learning is a generalization of the classical and well-studied problem of learning undirected graphical models, specifically parameter learning of these models. This classical problem requires $\frac{e^{\mathcal{O}(\beta)} m\log(m)}{\beta^2\varepsilon^2}$ time (and there is an algorithm matching the lower bound), so exponential dependence on β is necessary [30].[5] Analogies with the classical setting turn out to be of limited use, since the non-commutativity and nonlocality inherent in the quantum setting rules out generalizations of classical ideas. However, with the classical setting we can identify barriers to designing a time-efficient algorithm.

A key challenge of time-efficient Hamiltonian learning is that we cannot work directly with the partition function. The only previous approach to low temperature [6] only used its copies of ρ to estimate $\mathrm{tr}(E_a\rho)$ for all $a \in [m]$. It is known in the classical literature that taking just these estimates and using them to compute the parameters λ_a is as hard as computing the partition function [50]. To avoid this barrier, we take a richer set of expectations $\mathrm{tr}(P\rho)$ that allows us to reduce learning to a tractable, but fairly involved, optimization problem instead. Along the way we develop several new tools of independent interest, and ultimately give a semi-definite programming algorithm based on the sum-of-squares hierarchy. Consequently, we show that sophisticated modern tools in optimization theory lead to a surprising resolution of the Hamiltonian learning problem.

1.2 Technical Overview

The recipe for quantum Hamiltonian learning introduced by Anshu, Arunachalam, Kuwahara, and Soleimanifar [6] is based on matching the local marginals of the Gibbs state ρ, which we can estimate with copies of ρ. Specifically, for two Hamiltonians $H = \sum \lambda_a E_a$ and $H' = \sum \lambda'_a E_a$ with respective Gibbs states

$$\rho = \frac{e^{-\beta H}}{\mathrm{tr}(e^{-\beta H})} \text{ and } \rho' = \frac{e^{-\beta H'}}{\mathrm{tr}(e^{-\beta H'})},$$

they show that $H = H'$ (and so $\lambda_a = \lambda'_a$ for all $a \in [m]$) if and only if ρ and ρ' are identical on local marginals i.e. $\mathrm{tr}(E_a\rho) = \mathrm{tr}(E_a\rho')$ for all $a \in [m]$ [6, Proposition 4]. This does not imply a bound on sample complexity, because with copies of ρ, we can only compute $\mathrm{tr}(E_a\rho)$ approximately, with noise introduced from sampling error. The key structural result of [6] is that this equivalence can be made robust, so that if H' only approximately matches marginals, then the corresponding coefficients λ'_a approximately match the true coefficients λ_a.

However, the last step of this algorithm is to invert the map $\{\lambda_a\}_{a\in[m]} \mapsto \{\mathrm{tr}(E_a\rho)\}_{a\in[m]}$, which is a computationally hard problem. Formally, for a classical Hamiltonian,[6] $\mathrm{tr}(E_a\rho)$ are *sufficient statistics* of a graphical model and it is known that estimating the parameters of a graphical model from these sufficient statistics is computationally intractable [50]. This doesn't mean that the problem is hopeless, but rather that to find a tractable algorithm, we should be looking for the opportunity to use a richer family of statistics.

Designing a new system of constraints. We interpret the previous argument as defining and then solving a constraint system in the set of unknowns, $\{\lambda'_a\}_{a\in[m]}$. The structural result in [6] shows that an approximate solution to this system will be close to the true parameters λ_a. However, this system is computationally hard to solve. Our starting point is to define a larger set of constraints which $\{\lambda_a\}_{a\in[m]}$ must satisfy, which can be verified by measuring expectations of observables slightly less local than the terms $\{E_a\}_{a\in[m]}$. Let $\mathcal{P}_{\mathrm{local}}$ be the set of Pauli matrices whose support is K-local for some large constant K. We begin with the following system of constraints:

$$\left\{\begin{array}{ll} \forall a \in [m] & -1 \leqslant \lambda'_a \leqslant 1 \\ & H' = \sum_{a\in[m]} \lambda'_a \cdot E_a \\ \forall P, Q \in \mathcal{P}_{\mathrm{local}}, & \mathrm{tr}\left(Qe^{-\beta H'}Pe^{\beta H'}\rho\right) = \mathrm{tr}(PQ\rho) \end{array}\right\}, \quad (1)$$

The constraints above are indeed satisfied for the true parameters ($\lambda' = \lambda$) since by assumption $|\lambda_a| \leqslant 1$ for all $a \in [m]$ and moreover

$$\mathrm{tr}\left(Qe^{-\beta H}Pe^{\beta H}\rho\right) = \mathrm{tr}\left(Qe^{-\beta H}Pe^{\beta H}\frac{e^{-\beta H}}{\mathrm{tr}(e^{-\beta H})}\right) = \mathrm{tr}(PQ\rho)$$

which follows from the cyclic property of the trace. Two main challenges remain: Must a solution to this system be close to the true coefficients? And how can we efficiently solve the system? Eventually we will derive a convex relaxation for it that is based on

(A) replacing the last constraint in Eq. (1) which involves the matrix exponential with low degree polynomial constraints on the indeterminates $(\{\lambda'_a\}_{a\in[m]})$ instead and
(B) showing that any choice of λ' that satisfy the constraints must also approximately match the true coefficients λ.

In general solving systems of polynomial equations is computationally hard, but because our analysis in (B) will be based on sum-of-squares proofs, there is by now standard machinery for turning it into an efficient algorithm (see the full version for detailed explanation).

[5] It is an interesting open question to improve our doubly exponential dependence to singly exponential.

[6] A classical Hamiltonian is a Hamiltonian that is diagonal, i.e. its terms are tensor products of the identity and the Pauli σ_z. For a classical Hamiltonian, the state ρ is a sample from the Gibbs distribution, and $\mathrm{tr}(E_a\rho)$ is a $\mathfrak{K}$-point correlation function.

Identifying an equivalence between nested commutators and polynomials. Working towards the goal of replacing the term

$$\mathrm{tr}\Big(Qe^{-\beta H'}Pe^{\beta H'}\rho\Big)$$

with a low-degree polynomial in the variables λ', we begin by recalling the Hadamard formula:[7]

$$e^{-\beta H'}Pe^{\beta H'} = \sum_{\ell=0}^{\infty}(-\beta)^\ell\frac{[H',P]_\ell}{\ell!}, \tag{2}$$

where $[H',P]_\ell = [H',[H',\ldots,[H',P]\ldots]]$ is the ℓ-th nested commutator. A natural first step is to truncate this series at d terms and observe that

$$\mathrm{tr}\left(Q\left(\sum_{\ell=0}^{d}(-\beta)^\ell\frac{[H',P]_\ell}{\ell!}\right)\rho\right)$$

is a low-degree polynomial in the variables λ'. For instance, observe for the order-2 nested commutator, we have

$$\begin{aligned}\mathrm{tr}\Big(\Big[\sum_{a\in[m]}\lambda'_aE_a,P\Big]_2\rho\Big) &= \mathrm{tr}\Big(\Big[\sum_{a\in[m]}\lambda'_aE_a,\sum_{a\in[m]}\lambda'_a[E_a,P]\Big]\rho\Big)\\ &= \sum_{a,b\in[m]}\lambda'_a\lambda'_b\,\mathrm{tr}([E_a,[E_b,P]]\,\rho),\end{aligned}$$

which is a degree-2 polynomial in the λ'_i indeterminates. However, the series in Eq. (2) only converges quickly when β is sufficiently small [30], so we cannot use it.[8]

Nevertheless, from this observation we can develop a general formalism for constructing polynomial approximations of evolutions of operators. We observe that in the eigenbasis of H',

$$[H',P]_\ell = P\circ\Big\{\,(\sigma_i-\sigma_j)^\ell\,\Big\}_{ij},$$

where $\{\,\sigma_i\,\}_{i\in[N]}$ are the eigenvalues of H' and $\circ$ denotes the Hadamard product. Similarly,

$$e^{-\beta H'}Pe^{\beta H'} = P\circ\Big\{\,e^{-\beta(\sigma_i-\sigma_j)}\,\Big\}_{ij},$$

and thus we can focus our attention on designing polynomials that approximate the scalar quantity $e^{-\beta(\sigma_i-\sigma_j)}$ with low-degree polynomials in $(\sigma_i-\sigma_j)$. Further, any degree-d polynomial $p(z)=\sum_{\ell=0}^{d}c_\ell z^\ell$ can be extended to commutators as follows:

$$p(H'\mid P) = P\circ\big\{\,p(\sigma_i-\sigma_j)\,\big\}_{ij} = \sum_{\ell=0}^{d}c_\ell[H',P]_\ell.$$

This allows us to translate between matrix series expansions involving nested commutators and univariate polynomials. We note that for technical reasons we need to extend our equivalence to nested commutators with two distinct operators X,Y appearing in an arbitrary order, such as $[X,[Y,[X,\ldots]\ldots],E_{a_1}]$ and bi-variate polynomials $p(x,y)$. The translation between bi-variate polynomials and nested commutators incurs additive error depending on $[X,Y]$ due to re-ordering of the X and Y operators as expected. In our full algorithm, we introduce an additional constraint to drive this additive error to zero. Focusing on the scalar polynomial approximation to the exponential function, we now formalize the notion of approximation that we require.

Constructing a new, flat approximation to the exponential. Recall that we want a polynomial such that, working in the eigenbasis of H',

$$\begin{aligned}p(H'\mid P) &= P\circ\big\{\,p(\sigma_i-\sigma_j)\,\big\}_{ij}\\ &\approx P\circ\Big\{\,e^{-\beta(\sigma_i-\sigma_j)}\,\Big\}_{ij}\\ &= e^{-\beta H'}Pe^{\beta H'},\end{aligned} \tag{3}$$

where "$\approx$" denotes an unusual notion of approximation which, for the purposes of this discussion, we can consider to mean that the matrices are close in some norm. The Taylor series approximation to the exponential would be Eq. (2), which we established is too high degree.

Our key insight is that we choose a better polynomial approximation. We begin by observing that an operator with small support is approximately band-diagonal in the basis of eigenvectors of H, which is a property of local terms proved by Arad, Kuwahara, and Landau [7]. We state a weak version of this here: let P be a Pauli operator with support size $\mathcal{O}(1)$, and let $H=\sum_i d_iv_iv_i^\dagger$ be an eigendecomposition of H. Then, considering P in the eigenbasis of H,

$$|P_{ij}| = |v_i^\dagger Pv_j| \leqslant e^{-\Omega(|d_i-d_j|)}. \tag{4}$$

A consequence of this is that entries of

$$P\circ\Big\{\,p(\sigma_i-\sigma_j)-e^{-\beta(\sigma_i-\sigma_j)}\,\Big\}_{ij},$$

which denote the error of the polynomial approximation, are weighted inverse exponentially in $\sigma_i-\sigma_j$. Therefore, our polynomial approximation need not be equally good for all $\sigma_i-\sigma_j$; rather, our approximation should be ε-good in a small range but is allowed to diverge at a sufficiently slow exponential rate outside that range. We call this a *flat* approximation. In particular, given parameters $\beta\geqslant 0$, $0<\varepsilon,\eta<1$, we construct p such that

$$\begin{cases}\Big|p(z)-e^{-\beta z}\Big|\leqslant\varepsilon & \text{if } z\in[-1,1]\\ |p(z)|\leqslant\max\Big(1,e^{-\beta z}\Big)\cdot e^{\eta\beta|z|} & \text{if } z\notin[-1,1]\end{cases} \tag{5}$$

The key difficulty in satisfying the above constraints is satisfying $|p(z)|\leqslant e^{\eta\beta|z|}$ for $z\geqslant\beta$, as standard approximations like Taylor series truncations and Chebyshev series truncations fail this condition. We explicitly construct a degree

$$\Big(2^{\mathcal{O}(1/\eta)}\cdot(\beta+\log(1/(\varepsilon\eta)))\Big)$$

polynomial that satisfies Eq. (5). This construction is inspired by the iterative "peeling" of the exponential used in proofs of Lieb-Robinson bounds [33, 48]. We can write

$$e^{-\beta z} = \underbrace{e^{-\beta_c z}\cdots e^{-\beta_c z}}_{\beta/\beta_c}$$

for a fixed small constant β_c and then truncate the Taylor series expansion of $e^{-\beta_c z}$ at different scales for all of the $\beta/\beta_c=O(\beta)$ copies in the product so that the tails of the different truncations don't "interfere".

[7] This can be derived from the Baker–Campbell–Hausdorff formula,

$$\exp(A)\exp(B)=\exp\Big(A+B+\frac{1}{2}[A,B]+\frac{1}{12}([A,[A,B]]+[B,[B,A]])+\ldots\Big).$$

[8] This expansion does converge after $\beta\|H\|$ terms, but our running time is exponential in the degree, so this would be far too large.

We show that when p is a flat approximation of the above form for some sufficiently small η, then $Qp(H|P)\rho$ is a good approximation to $Qe^{-\beta H}Pe^{\beta H}\rho$. In other words, the polynomial approximation is good when $H' = H$ and we right multiply by ρ; this is crucial for the polynomial system that we set up next to be feasible.

Formulating a polynomial system. We now have all the tools to describe a polynomial system that captures the Hamiltonian learning problem. The constraint system we describe in this section is an informal treatment of the system that appears in the full algorithm, and avoids several technical details. We show that using our flat approximation to the exponential, we can obtain a polynomial p such that

$$\operatorname{tr}\left(Qe^{-\beta H}Pe^{\beta H}\rho\right) \approx \operatorname{tr}(Qp(H \mid P)\rho).$$

Then, we can re-write Eq. (1) as the following polynomial constraint system:

$$\begin{Bmatrix} \forall a \in [m] & -1 \leqslant \lambda'_a \leqslant 1 \\ & H' = \sum_{a\in[m]} \lambda'_a E_a \\ \forall P, Q \in \mathcal{P}_{\text{local}}, & |\operatorname{tr}(Qp(H' \mid P)\rho) - \operatorname{tr}(PQ\rho)| \leqslant \varepsilon \end{Bmatrix}, \quad (6)$$

and observe that the last constraint encodes a relaxation of the last constraint in Eq. (1) and is satisfied when $H' = H$. Further, all of the constraints are indeed succinctly representable as low-degree polynomials in the indeterminates, $\{\lambda'_a\}_{a\in[m]}$, as discussed earlier. Finally, the coefficients, such as $\operatorname{tr}([E_a, [E_b, P]])$, are expectations of the Gibbs state against a slightly larger set of local observables, which are the richer class of test functions we desired. We can obtain estimates of these expectations through quantum measurements. Computing these estimates is the only quantum part of our algorithm, and the rest of the algorithm is entirely classical.

Feasibility of the polynomial system. Recall that to show that the polynomial system in (6) is feasible, we need to argue that

$$\operatorname{tr}\left(Qe^{-\beta H}Pe^{\beta H}\rho\right) \approx \operatorname{tr}(Qp(H \mid P)\rho)$$

for all P, Q. Working in the eigenbasis of H, let its eigenvalues be $\{\sigma_i\}_{i\in[2^n]}$. The key tool that we leverage is from [7] which roughly states that any local term E must be approximately diagonal in the eigenbasis of H, with off-diagonal entries decaying as $|E_{ij}| \leqslant e^{-\Omega(|\sigma_i-\sigma_j|)}$. Thus, we can decompose the matrices Q, P into two parts – parts indexed by i, j where $|\sigma_i - \sigma_j| \leqslant \beta$ and parts indexed by i, j where $|\sigma_i - \sigma_j| \geqslant \beta$. Then we use the fact that $p(x)$ is a good approximation to e^{-x} on $[-\beta, \beta]$ to prove that the error on the first part is small. We then appeal to the exponential decay of the off-diagonals to argue that the contribution from the second part in both $\operatorname{tr}\left(Qe^{-\beta H}Pe^{\beta H}\rho\right)$ and $tr(Qp(H \mid P)\rho)$ is small. Our flat approximation to the exponential is designed to ensure that it does not overwhelm the exponential decay in the off-diagonal entries in P, Q in any regime.

Efficiently optimizing polynomial systems. Now that we know that our polynomial system is feasible, we consider a convex relaxation of this system. In particular, we consider a degree-d sum-of-squares relaxation, which can be efficiently optimized by expressing it as a semi-definite program (see the full version for details), with $d = \log(1/\varepsilon) \cdot 2^{\mathcal{O}(\beta)}$. Since we have m variables and $2^{\mathcal{O}(\beta)}$ constraints, and each constraint is a degree-d polynomial, we can solve the degree-$2d$ sum-of-squares relaxation of Eq. (6) in $m^{\left(\log(1/\varepsilon)\cdot 2^{\mathcal{O}(\beta)}\right)}$ time. The main challenge in analyzing the sum-of-squares relaxation is to show that we can *round* it to estimates $\left\{\tilde{\lambda}'_a\right\}_{a\in[m]}$ such that they are close to the true parameters. Here, we adopt the so-called *proofs-to-algorithms* philosophy, where we instead work with the dual object to the sum-of-squares relaxation, namely sum-of-squares proofs (see [13, 25], and references therein). This perspective states that if the true parameters are identifiable only using the sum-of-squares proof system, then we immediately obtain an efficient algorithm and we show that we can easily and accurately *round* the solution.

We then provide a proof of identifiability, i.e. for all $a \in [m]$, the inequality $(\lambda'_a - \lambda_a) \leqslant \varepsilon$ can be derived using the system of polynomial constraints and other basic inequalities that admit sum-of-squares proofs (we refer the reader to the full version for a detailed exposition). At a high level, the proof works by arguing that when $H' - H$ is large, there are witnesses P, Q such that

$$|\operatorname{tr}(Qp(H'|P)\rho) - \operatorname{tr}(Qp(H|P)\rho)|$$

is large. Since we know that H is a feasible solution, this would imply that H' cannot be a feasible solution so any feasible solution must have $H' - H$ be small. The construction of the witnesses relies on an additional property of the polynomial p that we construct, namely that it is strongly monotone (in some appropriate quantitative sense).

For the identifiability proof, we crucially use an additional important property of local Hamiltonians. It deals with the quantity $\operatorname{tr}(A^2\rho)$, where $A = \sum_b \sigma_b P_b$ is a Hermitian linear combination of Pauli matrices with small support. Thinking of ρ as a distribution, $\operatorname{tr}(A^2\rho)$ is a second moment term with respect to ρ; we can prove this is not much smaller than $\operatorname{tr}(A^2 I/\dim) = \sum_b \sigma_b^2$, the second moment against the uniform distribution: for some constant $c > 0$,

$$\operatorname{tr}(A^2\rho) \geqslant c^{\mathcal{O}(\beta)} \max_b \sigma_b^2.$$

Intuitively, this shows that ρ is not close to zero in any local direction. This was first shown by [6] for quasi-local operators; we adjust their proof to hold for just local operators and give a tighter bound. We show that we can obtain a slightly weaker statement of this form in the sum-of-squares proof system by formulating it as a quadratic inequality. This inequality can be used to remove the dependence on ρ in expressions appearing in the proof; for example, it is used to relate $\operatorname{tr}([H, H']^2\rho)$ to the size of $[H, H']$ itself.

Finally, we observe that our identifiability proof does not use the full power of a degree-$2d$ sum-of-squares relaxation and therefore, it should suffice to solve a significantly smaller semi-definite program. We show that we can execute our proof of identifiability by only appealing to a sparse subset of monomials of degree at most $2d$ and

invoke a linearization theorem by Steurer and Teigel [57] to obtain a final running time of $\mathrm{poly}(m) \cdot (1/\varepsilon)^{2^{\mathcal{O}(\beta)}}$, as desired.

1.3 Further Related Work

Hamiltonian learning. Hamiltonian learning is a broad topic studied both in experimental and theoretical contexts. This work fits into a large body of algorithmic research about learning properties of quantum states modeling physical systems [4, 21]. Here, we point to a few lines of related work in this field.

Hamiltonian learning often focuses on the real-time evolution setting, where one can allow the system to evolve with respect to H, applying the unitary $e^{-\mathrm{i}Ht}$ [22, 38, 59]. Some algorithms consider taking time derivatives (i.e. taking $t \to 0$), which are similar to small-β algorithms in the Gibbs state setting [29, 30, 60]. There is some research on learning from (zero-temperature) ground states [54], but the algorithmic work is limited because the ground state of a Hamiltonian need not determine the Hamiltonian. We study the finite temperature case, which is both the typical temperature at which experiments are run and, in the $\beta \to \infty$ limit, a rich approximation to the much less computationally tractable ground state [2, 27].

Though our algorithm is not practical, we use constraint systems that bear some similarity to the "correlation matrix" linear constraint systems analyzed heuristically and experimentally in prior work [8, 54]. In fact, our constraint system contains these constraints for technical reasons. Our work places these works on a rigorous basis, as we prove that, though the linear constraint systems might not uniquely identify the true Hamiltonian, adding more, similar constraints eventually fully constrains the Hamiltonian.

Bounding correlations in Gibbs states. Though classical Gibbs states have extremely good locality properties, these become much weaker in the quantum setting. A series of works aims at bounding the non-locality in quantum Gibbs states with various different measures and in various different regimes [40, 45, 46], often with the goal of concluding that simulating or learning these systems can be done time-efficiently. It is an interesting open problem whether one can extract a new kind of "locality" statement from our algorithm, to understand how general our approach is for learning quantum systems. Our polynomial approximation is inspired by proofs of the Lieb–Robinson bound [33, 48], and can be viewed as a "low-degree" form of this bound. This could be of independent interest.

Parameter learning of graphical models. There is a rich body of work on the problem of learning graphical models. Our setting is that of learning Markov random fields; the literature on this topic focuses on the task of *structure learning*, which in our setting corresponds to learning the terms $\{E_a\}_{a\in[m]}$, given the guarantee that they form an (unknown) dual interaction graph with bounded degree [18, 19, 31, 42]. The problem we consider, learning the parameters with known terms, is easy in the classical setting [30, Appendix B], because classical Gibbs states satisfy the Hammersley–Clifford theorem [32], also known as the Markov property. A consequence of the Markov property is that estimating a parameter on a $\mathfrak{K}$-body term can be done by computing conditional marginals on the support of this term. It is not clear how to generalize this argument to the quantum setting, since the Markov property does not hold for low-temperature quantum Hamiltonians, even approximately [46].

The sum-of-squares meta-algorithm. The sum-of-squares hierarchy has been used to analyze several problems in quantum information, including best state separation [14, 16, 17, 23], optimizing fermionic Hamiltonians [34, 35], and a quantum analogue of max-cut [52, 58]. Additionally, the proofs-to-algorithms perspective, introduced in [13, 15], has been extensively used to design efficient algorithms for several estimation and learning tasks. In particular, this perspective has led to efficient algorithms for robust learning [9, 10, 12, 37, 41, 44, 49] and list-decodable learning [11, 39, 55].

2 FORMAL RESULTS AND PROOFS

Due to space constraints, the formal statements of our main results and full proofs are deferred to the full version.

ACKNOWLEDGMENTS

AB is supported by Ankur Moitra's ONR grant and the NSF TRIPODS program (award DMS-2022448). AL is supported in part by an NSF GRFP and a Hertz Fellowship. AM is supported in part by a Microsoft Trustworthy AI Grant, an ONR grant and a David and Lucile Packard Fellowship. ET is supported by the NSF GRFP (DGE-1762114) and the Miller Institute for Basic Research in Science, University of California Berkeley.

REFERENCES

[1] Tameem Albash, Victor Martin-Mayor, and Itay Hen. [n. d.]. Temperature scaling law for quantum annealing optimizers. 119, 11 ([n. d.]), 110502. https://doi.org/10.1103/physrevlett.119.110502 arXiv:1703.03871 [quant-ph]
[2] Álvaro M. Alhambra. 2022. Quantum many-body systems in thermal equilibrium. (2022). https://doi.org/10.48550/ARXIV.2204.08349 arXiv:2204.08349 [quant-ph]
[3] Luigi Amico, Rosario Fazio, Andreas Osterloh, and Vlatko Vedral. [n. d.]. Entanglement in many-body systems. 80, 2 ([n. d.]), 517–576. https://doi.org/10.1103/revmodphys.80.517 arXiv:quant-ph/0703044
[4] Anurag Anshu and Srinivasan Arunachalam. 2023. A survey on the complexity of learning quantum states. (2023). https://doi.org/10.48550/ARXIV.2305.20069 arXiv:2305.20069 [quant-ph]
[5] Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, and Mehdi Soleimanifar. [n. d.]. Efficient learning of commuting hamiltonians on lattices. https://anuraganshu.seas.harvard.edu/links
[6] Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, and Mehdi Soleimanifar. 2020. Sample-efficient learning of quantum many-body systems. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)* (2020-11). IEEE, 685–691. https://doi.org/10.1109/focs46700.2020.00069 arXiv:2004.07266 [quant-ph]
[7] Itai Arad, Tomotaka Kuwahara, and Zeph Landau. [n. d.]. Connecting global and local energy distributions in quantum spin models on a lattice. 2016, 3 ([n. d.]), 033301. https://doi.org/10.1088/1742-5468/2016/03/033301 arXiv:1406.3898 [quant-ph]
[8] Eyal Bairey, Itai Arad, and Netanel H. Lindner. 2019. Learning a local Hamiltonian from local measurements. *Physical Review Letters* 122 (Jan. 2019), 020504. Issue 2. https://doi.org/10.1103/PhysRevLett.122.020504 arXiv:1807.04564
[9] Ainesh Bakshi, Ilias Diakonikolas, Samuel B Hopkins, Daniel Kane, Sushrut Karmalkar, and Pravesh K Kothari. 2020. Outlier-robust clustering of gaussians and other non-spherical mixtures. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 149–159.
[10] Ainesh Bakshi, Ilias Diakonikolas, He Jia, Daniel M Kane, Pravesh K Kothari, and Santosh S Vempala. 2022. Robustly learning mixtures of k arbitrary gaussians. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. 1234–1247.
[11] Ainesh Bakshi and Pravesh K Kothari. 2021. List-decodable subspace recovery: Dimension independent error in polynomial time. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 1279–1297.
[12] Ainesh Bakshi and Adarsh Prasad. 2021. Robust linear regression: Optimal rates in polynomial time. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. 102–115.

[13] Boaz Barak. [n. d.]. Proofs, beliefs, and algorithms through the lens of sum-of-squares. ([n. d.]).

[14] Boaz Barak, Fernando GSL Brandao, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. 2012. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*. 307–326.

[15] Boaz Barak, Jonathan A Kelner, and David Steurer. 2015. Dictionary learning and tensor decomposition via the sum-of-squares method. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*. 143–151.

[16] Boaz Barak, Pravesh K Kothari, and David Steurer. 2017. Quantum entanglement, sum of squares, and the log rank conjecture. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*. 975–988.

[17] Fernando GSL Brandão, Matthias Christandl, and Jon Yard. 2011. A quasipolynomial-time algorithm for the quantum separability problem. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*. 343–352.

[18] Guy Bresler. 2015. Efficiently learning Ising models on arbitrary graphs. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing - STOC '15*. ACM Press. https://doi.org/10.1145/2746539.2746631

[19] Guy Bresler, Elchanan Mossel, and Allan Sly. 2013. Reconstruction of Markov random fields from samples: some observations and algorithms. *SIAM J. Comput.* 42, 2 (Jan. 2013), 563–578. https://doi.org/10.1137/100796029

[20] Jose Carrasco, Andreas Elben, Christian Kokail, Barbara Kraus, and Peter Zoller. [n. d.]. Theoretical and experimental perspectives of quantum verification. 2, 1 ([n. d.]), 010102. https://doi.org/10.1103/prxquantum.2.010102 arXiv:2102.05927 [quant-ph]

[21] Marcus Cramer, Martin B. Plenio, Steven T. Flammia, Rolando Somma, David Gross, Stephen D. Bartlett, Olivier Landon-Cardinal, David Poulin, and Yi-Kai Liu. [n. d.]. Efficient quantum state tomography. 1, 1 ([n. d.]). https://doi.org/10.1038/ncomms1147 arXiv:1101.4366 [quant-ph]

[22] Marcus P. da Silva, Olivier Landon-Cardinal, and David Poulin. [n. d.]. Practical characterization of quantum devices without tomography. 107, 21 ([n. d.]), 210404. https://doi.org/10.1103/physrevlett.107.210404 arXiv:1104.3835 [quant-ph]

[23] Andrew C Doherty, Pablo A Parrilo, and Federico M Spedalieri. 2002. Distinguishing separable and entangled states. *Physical Review Letters* 88, 18 (2002), 187904.

[24] M. Endres, M. Cheneau, T. Fukuhara, C. Weitenberg, P. Schauß, C. Gross, L. Mazza, M. C. Bañuls, L. Pollet, I. Bloch, and S. Kuhr. [n. d.]. Observation of correlated particle-hole pairs and string order in low-dimensional Mott insulators. 334, 6053 ([n. d.]), 200–203. https://doi.org/10.1126/science.1209284 arXiv:1108.3317 [cond-mat.quant-gas]

[25] Noah Fleming, Pravesh Kothari, Toniann Pitassi, et al. 2019. *Semialgebraic Proofs and Efficient Algorithm Design.* now the essence of knowledge.

[26] I. M. Georgescu, S. Ashhab, and Franco Nori. [n. d.]. Quantum simulation. 86, 1 ([n. d.]), 153–185. https://doi.org/10.1103/revmodphys.86.153 arXiv:1308.6253 [quant-ph]

[27] Sevag Gharibian, Yichen Huang, Zeph Landau, and Seung Woo Shin. [n. d.]. Quantum Hamiltonian complexity. 10, 3 ([n. d.]), 159–282. https://doi.org/10.1561/0400000066 arXiv:1401.3916 [quant-ph]

[28] Christian Gross and Immanuel Bloch. [n. d.]. Quantum simulations with ultracold atoms in optical lattices. 357, 6355 ([n. d.]), 995–1001. https://doi.org/10.1126/science.aal3837

[29] Andi Gu, Lukasz Cincio, and Patrick J. Coles. 2022. Practical black box hamiltonian learning. (2022). https://doi.org/10.48550/ARXIV.2206.15464 arXiv:2206.15464 [quant-ph]

[30] Jeongwan Haah, Robin Kothari, and Ewin Tang. 2022. Optimal learning of quantum Hamiltonians from high-temperature Gibbs states. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)* (2022-10). IEEE. https://doi.org/10.1109/focs54457.2022.00020 arXiv:2108.04842 [quant-ph]

[31] Linus Hamilton, Frederic Koehler, and Ankur Moitra. 2017. Information theoretic properties of markov random fields, and their algorithmic applications. *Advances in Neural Information Processing Systems* 30 (2017).

[32] J.M. Hammersley and P. Clifford. 1971. Markov field on finite graphs and lattices. (1971). Available at http://www.statslab.cam.ac.uk/~grg/books/hammfest/hamm-cliff.pdf.

[33] Matthew B Hastings. [n. d.]. Locality in quantum systems. In *Quantum Theory from Small to Large Scales*, Manfred Salmhofer, Vieri Mastropietro, Wojciech De Roeck, and Leticia F. Cugliandolo (Eds.). Lecture Notes of the les Houches Summer School Ser., Vol. 95. Oxford University Press, Incorporated, 171–212.

[34] Matthew B Hastings. 2023. Field Theory and The Sum-of-Squares for Quantum Systems. *arXiv preprint arXiv:2302.14006* (2023).

[35] Matthew B Hastings and Ryan O'Donnell. 2022. Optimizing strongly interacting fermionic Hamiltonians. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. 776–789.

[36] Sarah Hirthe, Thomas Chalopin, Dominik Bourgund, Petar Bojović, Annabelle Bohrdt, Eugene Demler, Fabian Grusdt, Immanuel Bloch, and Timon A. Hilker. [n. d.]. Magnetically mediated hole pairing in fermionic ladders of ultracold atoms. 613, 7944 ([n. d.]), 463–467. https://doi.org/10.1038/s41586-022-05437-y arXiv:2203.10027 [cond-mat.quant-gas]

[37] Samuel B Hopkins and Jerry Li. 2018. Mixture models, robustness, and sum of squares proofs. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. 1021–1034.

[38] Hsin-Yuan Huang, Yu Tong, Di Fang, and Yuan Su. [n. d.]. Learning many-body Hamiltonians with Heisenberg-limited scaling. 130, 20 ([n. d.]), 200403. https://doi.org/10.1103/physrevlett.130.200403 arXiv:2210.03030 [quant-ph]

[39] Sushrut Karmalkar, Adam Klivans, and Pravesh Kothari. 2019. List-decodable linear regression. *Advances in neural information processing systems* 32 (2019).

[40] Kohtaro Kato and Fernando G. S. L. Brandão. 2019. Quantum approximate markov chains are thermal. *Communications in Mathematical Physics* 370, 1 (June 2019), 117–149. https://doi.org/10.1007/s00220-019-03485-6

[41] Adam Klivans, Pravesh K Kothari, and Raghu Meka. 2018. Efficient algorithms for outlier-robust regression. In *Conference On Learning Theory*. PMLR, 1420–1430.

[42] Adam Klivans and Raghu Meka. 2017. Learning graphical models using multiplicative weights. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. https://doi.org/10.1109/focs.2017.39

[43] Christian Kokail, Rick van Bijnen, Andreas Elben, Benoit Vermersch, and Peter Zoller. [n. d.]. Entanglement Hamiltonian tomography in quantum simulation. 17, 8 ([n. d.]), 936–942. https://doi.org/10.1038/s41567-021-01260-w arXiv:2009.09000 [quant-ph]

[44] Pravesh K Kothari, Jacob Steinhardt, and David Steurer. 2018. Robust moment estimation and improved clustering via sum of squares. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. 1035–1046.

[45] Tomotaka Kuwahara, Álvaro M. Alhambra, and Anurag Anshu. [n. d.]. Improved thermal area law and quasilinear time algorithm for quantum Gibbs states. 11, 1 ([n. d.]), 011047. https://doi.org/10.1103/physrevx.11.011047 arXiv:2007.11174 [quant-ph]

[46] Tomotaka Kuwahara, Kohtaro Kato, and Fernando G. S. L. Brandão. 2020. Clustering of conditional mutual information for quantum Gibbs states above a threshold temperature. *Physical Review Letters* 124, 22 (June 2020). https://doi.org/10.1103/physrevlett.124.220601

[47] J. P. F. LeBlanc, Andrey E. Antipov, Federico Becca, Ireneusz W. Bulik, Garnet Kin-Lic Chan, Chia-Min Chung, Youjin Deng, Michel Ferrero, Thomas M. Henderson, Carlos A. Jiménez-Hoyos, E. Kozik, Xuan-Wen Liu, Andrew J. Millis, N. V. Prokof'ev, Mingpu Qin, Gustavo E. Scuseria, Hao Shi, B. V. Svistunov, Luca F. Tocchio, I. S. Tupitsyn, Steven R. White, Shiwei Zhang, Bo-Xiao Zheng, Zhenyue Zhu, and Emanuel Gull and. [n. d.]. Solutions of the two-dimensional hubbard model: benchmarks and results from a wide range of numerical algorithms. 5, 4 ([n. d.]), 041041. https://doi.org/10.1103/physrevx.5.041041 arXiv:1505.02290 [cond-mat.str-el]

[48] Elliott H. Lieb and Derek W. Robinson. [n. d.]. The finite group velocity of quantum spin systems. 28, 3 ([n. d.]), 251–257. https://doi.org/10.1007/bf01645779

[49] Allen Liu and Ankur Moitra. 2021. Settling the robust learnability of mixtures of gaussians. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. 518–531.

[50] Andrea Montanari. 2015. Computational implications of reducing data to sufficient statistics. (2015). arXiv:1409.3821 [stat.CO]

[51] Jaan Oitmaa, Chris Hamer, and Weihong Zheng. [n. d.]. *Series expansion methods for strongly interacting lattice models.* Cambridge University Press. https://doi.org/10.1017/cbo9780511584398

[52] Ojas Parekh and Kevin Thompson. 2021. Application of the Level-2 Quantum Lasserre Hierarchy in Quantum Approximation Algorithms. *arXiv preprint arXiv:2105.05698* (2021).

[53] John Preskill. 2018. Quantum computing in the NISQ era and beyond. *Quantum* 2 (Aug. 2018), 79. https://doi.org/10.22331/q-2018-08-06-79 arXiv:1801.00862

[54] Xiao-Liang Qi and Daniel Ranard. 2019. Determining a local Hamiltonian from a single eigenstate. *Quantum* 3 (July 2019), 159. https://doi.org/10.22331/q-2019-07-08-159

[55] Prasad Raghavendra and Morris Yau. 2020. List decodable learning via sum of squares. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 161–180.

[56] M. D. Shulman, S. P. Harvey, J. M. Nichol, S. D. Bartlett, A. C. Doherty, V. Umansky, and A. Yacoby. [n. d.]. Suppressing qubit dephasing using real-time Hamiltonian estimation. 5, 1 ([n. d.]). https://doi.org/10.1038/ncomms6156

[57] David Steurer and Stefan Tiegel. [n. d.]. SoS degree reduction with applications to clustering and robust moment estimation. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*. Society for Industrial and Applied Mathematics, 374–393. https://doi.org/10.1137/1.9781611976465.24 arXiv:2101.01509 [cs.DS]

[58] Adam Bene Watts, Anirban Chowdhury, Aidan Epperly, J William Helton, and Igor Klep. 2023. Relaxations and Exact Solutions to Quantum Max Cut via the Algebraic Structure of Swap Operators. *arXiv preprint arXiv:2307.15661* (2023).

[59] Nathan Wiebe, Cassandra Granade, Christopher Ferrie, and D. G. Cory. [n. d.]. Hamiltonian learning and certification using quantum resources. 112, 19 ([n. d.]), 190501. https://doi.org/10.1103/physrevlett.112.190501 arXiv:1309.0876 [quant-ph]

[60] Assaf Zubida, Elad Yitzhaki, Netanel H. Lindner, and Eyal Bairey. 2021. Optimal short-time measurements for Hamiltonian learning. (2021). https://doi.org/10.48550/ARXIV.2108.08824 arXiv:2108.08824

Received 06-NOV-2023; accepted 2024-02-11