

Maximum-likelihood-estimate Hamiltonian learning via efficient and robust quantum likelihood gradient

Tian-Lun Zhao, Shi-Xin Hu, and Yi Zhang^{*}

International Center for Quantum Materials, School of Physics, Peking University, Beijing 100871, China



(Received 18 January 2023; accepted 15 May 2023; published 30 May 2023)

Given the recent developments in quantum techniques, modeling the physical Hamiltonian of a target quantum many-body system is becoming an increasingly practical and vital research direction. Here, we propose an efficient strategy combining maximum-likelihood estimation, gradient descent, and quantum many-body algorithms. Given the measurement outcomes, we optimize the target model Hamiltonian and density operator via a series of descents along the quantum likelihood gradient, which we prove is negative semidefinite with respect to the negative-log-likelihood function. In addition to such optimization efficiency, our maximum-likelihood-estimate Hamiltonian learning respects the locality of a given quantum system, therefore, extends readily to larger systems with available quantum many-body algorithms. Compared with previous approaches, it also exhibits better accuracy and overall stability toward noises, fluctuations, and temperature ranges, which we demonstrate with various examples.

DOI: [10.1103/PhysRevResearch.5.023136](https://doi.org/10.1103/PhysRevResearch.5.023136)

I. INTRODUCTION

Understanding the quantum states and the corresponding properties of a given quantum Hamiltonian is a crucial problem in quantum physics. Many powerful numerical and theoretical tools have been developed for such purposes and made compelling progress [1–5]. However, with the rapid experimental developments of quantum technology, e.g., near-term quantum computation [6,7] and simulation [8–14], it is also vital to explore the inverse problem, e.g., Hamiltonian learning—optimize a model Hamiltonian characterizing a quantum system with respect to the measurement results. Given the knowledge and assumption of a target system, researchers have achieved many resounding successes modeling quantum Hamiltonians with physical pictures and phenomenological approaches [15,16]. However, such subjective perspectives may risk biases and are commonly insufficient on detailed quantum devices. Therefore, the explorations for objective Hamiltonian learning strategies have attracted much recent attention [17–25].

There are mainly two categories of Hamiltonian-learning strategies, based upon either quantum measurements on a large number of (identical copies of) quantum states, e.g., Gibbs states or eigenstates [17–22], or initial states' time evolution dynamics [23–26], corresponding to the target quantum system. For example, given the measurements of the correlations of a set of local operators, the kernel of the

resulting correlation matrix offers a candidate model Hamiltonian [17–19]. However, while established theoretically, most approaches suffer from elevated costs and are limited to small systems in experiments or numerical simulations [19,20,27,28]. Besides, there remains much room for improvements in stability toward noises and temperature ranges.

Maximum-likelihood estimation (MLE) is a powerful tool that parameterizes and then optimizes the probability distribution of a statistical model so that the given observed data is most probable. MLE's intuitive and flexible logic makes it a prevailing method for statistical inference. Adding to its wide range of applications, MLE has been applied successfully to quantum state tomography [29–33], providing the most probable quantum states given the measurement outputs.

Inspired by MLE's successes in quantum problems, we propose a general MLE Hamiltonian learning protocol: given finite-temperature measurements of the target quantum system in thermal equilibrium, we optimize the model Hamiltonian toward the MLE step-by-step via a “quantum likelihood gradient.” We show that such quantum likelihood gradient, acting collectively on all presenting operators, is negative semidefinite with respect to the negative-log-likelihood function and thus provides efficient optimization. In addition, our strategy may take advantage of the locality of the quantum system, therefore allowing us to extend studies to larger quantum systems with tailored quantum many-body ansatzes such as Lanczos, quantum Monte Carlo (QMC), density matrix renormalization group (DMRG), and finite-temperature tensor network (FTTN) [34,35] algorithms in suitable scenarios. We also demonstrate that MLE Hamiltonian learning is more accurate, less restrictive, and more robust against noises and broader temperature ranges. Further, we generalize our protocol to measurements on pure states, such as the target quantum systems' ground states or quantum chaotic eigenstates. Therefore, MLE Hamiltonian learning

^{*}frankzhangyi@gmail.com

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

enriches our arsenal for cutting-edge research and applications of quantum devices and experiments, such as quantum computation, quantum simulation, and quantum Boltzmann machines [36].

We organize the rest of the paper as follows: In Sec. II, we review the MLE context and introduce the MLE Hamiltonian learning protocol; especially, we show explicitly that the corresponding quantum likelihood gradient leads to a negative semidefinite change to the negative-log-likelihood function. Via various examples in Sec. III, we demonstrate our protocol's capability, especially its robustness against noises and temperature ranges. We generalize the protocol to quantum measurements of pure states in Sec. IV and Appendix D, with consistent results for exotic quantum systems such as quantum critical and topological models. We summarize our studies in Sec. V with a conclusion on our protocol's advantages (and limitations), potential applications, and future outlooks.

II. MAXIMUM-LIKELIHOOD-ESTIMATE HAMILTONIAN LEARNING

To start, we consider an unknown target quantum system $\hat{H}_s = \sum_j \mu_j \hat{O}_j$ in thermal equilibrium, and measurements of a set of observables $\{\hat{O}_i\}$ on its Gibbs state $\hat{\rho}_s = \exp(-\beta \hat{H}_s) / \text{tr}[\exp(-\beta \hat{H}_s)]$, where β is the inverse temperature. Given a sufficient number N_i of measurements of the operator \hat{O}_i , the occurrence time f_{λ_i} of the λ_i^{th} eigenvalue o_{λ_i} approaches:

$$f_{\lambda_i} = p_{\lambda_i} N_i \approx \text{tr}[\hat{\rho}_s \hat{P}_{\lambda_i}] N_i, \quad (1)$$

where $p_{\lambda_i} = f_{\lambda_i} / N_i$ denotes the statistics of the outcome o_{λ_i} , and \hat{P}_{λ_i} is the corresponding projection operator to the o_{λ_i} sector. Our goal is to locate the model Hamiltonian \hat{H}_s for the quantum system, which commonly requires the presence of all \hat{H}_s 's terms in the measurement set $\{\hat{O}_i\}$.

Following previous MLE analysis [29–33], the statistical weight of any given state $\hat{\rho}$ is

$$\mathcal{L}(\hat{\rho}) \propto \prod_{i, \lambda_i} \left\{ \text{tr}[\hat{\rho} \hat{P}_{\lambda_i}] \frac{f_{\lambda_i}}{N_{\text{tot}}} \right\}^{N_{\text{tot}}}, \quad (2)$$

up to a trivial factor, where $N_{\text{tot}} = \sum_i N_i$ is the total number of measurements. For Hamiltonian learning, we search for (the set of parameters μ_j of) the MLE Hamiltonian \hat{H} , whose Gibbs state $\hat{\rho}$ maximizes the likelihood function in Eq. (2). The maximum condition for Eq. (2) can be re-expressed as

$$\begin{aligned} \hat{R}(\hat{\rho})\hat{\rho} &= \hat{\rho}, \\ \hat{R}(\hat{\rho}) &= \sum_{i, \lambda_i} \frac{f_{\lambda_i}}{N_{\text{tot}}} \frac{\hat{P}_{\lambda_i}}{\text{tr}[\hat{\rho} \hat{P}_{\lambda_i}]}, \end{aligned} \quad (3)$$

see Appendix A for a detailed review. Solving Eq. (3) is a nonlinear and nontrivial problem, for which many algorithms have been proposed [30–33]. For example, we can employ iterative updates $\hat{\rho}_{k+1} \propto \hat{R}(\hat{\rho}_k)\hat{\rho}_k\hat{R}(\hat{\rho}_k)$ until Eq. (3) is fulfilled [31]. These algorithms mostly center around the parametrization and optimization of a quantum state $\hat{\rho}$, whose cost is exponential in the system size. Besides, such iterative updates do not guarantee that the quantum state $\hat{\rho}$ remains a Gibbs form, especially when the measurements are insufficient to

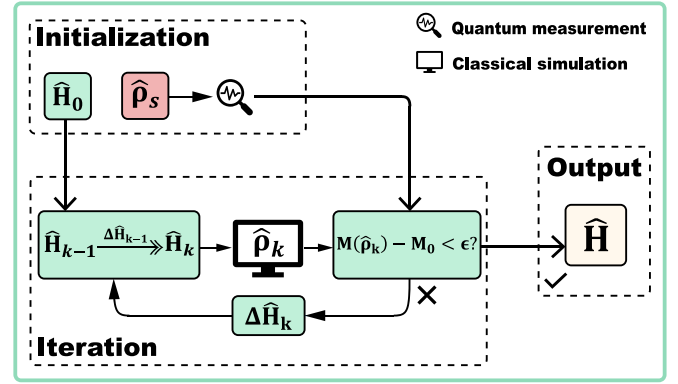


FIG. 1. An illustration of the MLE Hamiltonian learning algorithm: given the quantum measurements on the Gibbs state $\hat{\rho}_s$ of the target quantum system, we update the candidate Hamiltonian iteratively until the negative-log-likelihood function (or relative entropy) converges below a given threshold ϵ , after which the output yields the MLE Hamiltonian. Within each iterative step, we evaluate the operator expectation values with respect to the Gibbs state $\hat{\rho}_k = \exp(-\beta \hat{H}_k) / \text{tr}[\exp(-\beta \hat{H}_k)]$, which directs the model update $\Delta \hat{H}_k = -\gamma \hat{R}_k$ for the next iterative step.

uniquely determine the state [e.g., large noises or small numbers of measurements and there are many quantum states satisfying Eq. (3)]. Consequently, extracting $\hat{H} \propto -\frac{1}{\beta} \ln \hat{\rho}$ from $\hat{\rho}$ further adds up to the inconvenience.

Considering that the operator $\hat{R}(\hat{\rho})$ has the same operator structure as the Hamiltonian, we take an alternative stance for the Hamiltonian learning task and update the candidate Hamiltonian \hat{H}_k , i.e., the model parameters, collectively and iteratively. In particular, we integrate the corrections to the Hamiltonian coefficients to the operator $\hat{R}(\hat{\rho})$, which offers such a quantum likelihood gradient (Fig. 1):

$$\begin{aligned} \hat{H}_{k+1} &= \hat{H}_k - \gamma \hat{R}_k, \\ \hat{\rho}_{k+1} &= \frac{e^{-\beta \hat{H}_{k+1}}}{\text{tr}[e^{-\beta \hat{H}_{k+1}}]} = \frac{e^{-\beta(\hat{H}_k - \gamma \hat{R}_k)}}{\text{tr}[e^{-\beta(\hat{H}_k - \gamma \hat{R}_k)}]}, \end{aligned} \quad (4)$$

where $\gamma > 0$ is the learning rate—a small parameter controlling the step size. We denote $\hat{R}_k \equiv \hat{R}(\hat{\rho}_k)$ for short here afterwards. Compared with previous Hamiltonian extractions from MLE quantum state tomography, the update in Eq. (4) possesses several advantages in Hamiltonian learning. First, we can utilize the Hamiltonian structure (e.g., locality) to choose suitable numerical tools (e.g., QMC and FTTN) and even calculate within the subregions—we circumvent the costly parametrization of the quantum state $\hat{\rho}$. Also, the update guarantees a state in its Gibbs form. Last but not least, we will show that for $\gamma \ll 1$, such a quantum likelihood gradient in Eq. (4) yields a negative semidefinite contribution to the negative-log-likelihood function, guaranteeing the MLE Hamiltonian (up to a trivial constant) at its convergence and an efficient optimization toward it.

Theorem. For $\gamma \ll 1$, $\gamma > 0$, the quantum likelihood gradient in Eq. (4) yields a negative semidefinite contribution to the negative-log-likelihood function $M(\hat{\rho}_{k+1}) = -\frac{1}{N_{\text{tot}}} \log \mathcal{L}(\hat{\rho}_{k+1})$.

Proof. We note that up to linear order in $\gamma \ll 1$,

$$\begin{aligned} e^{-\beta \hat{H}_{k+1}} &= e^{-\beta \hat{H}_k} \prod_{n=0}^{\infty} \exp \left[\frac{(-1)^n}{(n+1)!} \text{ad}_{-\beta \hat{H}_k}^n (\beta \gamma \hat{R}_k) + o(\gamma^2) \right] \\ &\approx e^{-\beta \hat{H}_k} \left[1 + \sum_{n=0}^{\infty} \frac{(-1)^n}{(n+1)!} \text{ad}_{-\beta \hat{H}_k}^n (\beta \gamma \hat{R}_k) \right] \\ &= e^{-\beta \hat{H}_k} \left(1 + \beta \gamma \int_0^1 e^{\beta s \hat{H}_k} \hat{R}_k e^{-\beta s \hat{H}_k} ds \right), \end{aligned} \quad (5)$$

where $\text{ad}_A^j \hat{B} = [\hat{A}, \text{ad}_A^{j-1} \hat{B}]$ and $\text{ad}_A^0 \hat{B} = \hat{B}$ are the adjoint action of the Lie algebra. The first and third lines are based on the Zassenhaus formula [37] and the Baker-Hausdorff formula [37], respectively, while the second line neglects terms above the linear order of γ .

Following this, we can re-express the quantum state in Eq. (4) as

$$\hat{\rho}_{k+1} = \hat{\rho}_k \frac{1 + \beta \gamma \int_0^1 e^{\beta s \hat{H}_k} \hat{R}_k e^{-\beta s \hat{H}_k} ds}{1 + \beta \gamma}, \quad (6)$$

where we have used $\text{tr}[\hat{\rho}_k \hat{R}_k] = 1$ as a direct consequence of \hat{R}_k 's definition in Eq. (3).

Subsequently, after introducing the quantum likelihood gradient, the negative-log-likelihood function becomes

$$\begin{aligned} M(\hat{\rho}_{k+1}) &= -\frac{1}{N_{\text{tot}}} \log \mathcal{L}(\hat{\rho}_{k+1}) \\ &= -\sum_{i, \lambda_i} \frac{f_{\lambda_i}}{N_{\text{tot}}} \log \text{tr}[\hat{\rho}_{k+1} \hat{P}_{\lambda_i}] \\ &\approx M(\hat{\rho}_k) + \beta \gamma (1 - \Delta_k), \end{aligned} \quad (7)$$

where we keep terms up to linear order of γ in the log expansion.

We can establish the following inequality:

$$\begin{aligned} \Delta_k &= \text{tr} \left[\hat{\rho}_k \int_0^1 e^{\beta s \hat{H}_k} \hat{R}_k e^{-\beta s \hat{H}_k} \hat{R}_k ds \right] \\ &= \int_0^1 \text{tr}[\hat{\rho}_k e^{\beta s \hat{H}_k} \hat{R}_k e^{-\beta s \hat{H}_k} \hat{R}_k] \text{tr}[\hat{\rho}_k] ds \\ &= \int_0^1 \|e^{-\beta s \hat{H}_k/2} \hat{R}_k e^{-\beta(1-s) \hat{H}_k/2}\|_F^2 \|e^{-\beta \hat{H}_k/2}\|_F^2 \frac{ds}{Z_k} \\ &\geq \int_0^1 \text{tr}[\hat{\rho}_k \hat{R}_k]^2 ds = 1, \end{aligned} \quad (8)$$

where $Z_k = \text{tr}[e^{-\beta \hat{H}_k}]$ is the partition function, $\|A\|_F = \sqrt{\text{tr}[A^\dagger A]}$ is the Frobenius norm of matrix A , and the nonnegative definiteness of $\hat{\rho}_k$ allows $\hat{\rho}_k = (\hat{\rho}_k^{1/2})^2 = (e^{-\beta \hat{H}_k/2})^2/Z_k$. The inequality in the fourth line follows the Cauchy-Schwarz inequality.

We note that the equality—the convergence criteria of our MLE Hamiltonian learning protocol—is established if and only if

$$e^{-\beta s \hat{H}_k/2} \hat{R}_k e^{-\beta(1-s) \hat{H}_k/2} = e^{-\beta \hat{H}_k/2}, \quad (9)$$

which implies the conventional MLE optimization target $\hat{R} \hat{\rho} = \hat{\rho}$ in Eq. (3). We can also establish such consistency from our iterative convergence [38] following Eq. (4):

$$\hat{\rho}_{k+1} = \frac{e^{-\beta(\hat{H}_k - \gamma \hat{R}_k)}}{\text{tr}[e^{-\beta(\hat{H}_k - \gamma \hat{R}_k)}]} = \frac{e^{\beta \gamma \hat{R}_k} \hat{\rho}_k}{\text{tr}[e^{\beta \gamma \hat{R}_k} \hat{\rho}_k]} = \hat{\rho}_k, \quad (10)$$

where we have used the commutation relation $[\hat{R}_k, \hat{H}_k] = 0$ between the Hermitian operators \hat{R}_k and \hat{H}_k following $[\hat{R}_k, \hat{\rho}_k] = [\hat{R}_k, e^{-\beta \hat{H}_k}] = 0$.

Finally, combining Eqs. (7) and (8), we have shown that $M(\hat{\rho}_{k+1}) - M(\hat{\rho}_k) \leq 0$ is a negative semidefinite quantity, which proves the theorem. ■

We conclude that the quantum likelihood gradient in Eq. (4) offers an efficient and collective optimization toward the MLE Hamiltonian, modifying all model parameters simultaneously. For each step of quantum likelihood gradient, the most costly calculation is on $\hat{\rho}_{k+1}$, or more precisely, the expectation value $\text{tr}[\hat{\rho}_{k+1} \hat{P}_{\lambda_i}]$ from \hat{H}_{k+1} . Fortunately, this is a routine calculation in quantum many-body physics and condensed matter physics with various tailored candidate algorithms under different scenarios. For example, we may resort to the FTTN, or the QMC approaches, which readily apply to much larger systems than brute-force exact diagonalization. Thus, we emphasize that MLE Hamiltonian learning works with evaluations of the expectation values of quantum states instead of the more expensive quantum states themselves in their entirety.

Interestingly, MLE Hamiltonian learning also allows a more local stance. For a given Hamiltonian, the necessary expectation value of its Gibbs state $\text{tr}[\hat{\rho} \hat{P}_{\lambda_i}]$ takes the form

$$\begin{aligned} \text{tr}[\hat{\rho} \hat{P}_{\lambda_i}] &= \text{tr}[\hat{\rho}_{\text{eff}}^A \hat{P}_{\lambda_i}], \\ \hat{\rho}_{\text{eff}}^A &= \text{tr}_A[\hat{\rho}] = \frac{e^{-\beta \hat{H}_{\text{eff}}^A}}{\text{tr}[e^{-\beta \hat{H}_{\text{eff}}^A}]}, \end{aligned} \quad (11)$$

where $\hat{\rho}_{\text{eff}}^A$ is the reduced density operator defined upon a relatively local subsystem A still containing \hat{P}_{λ_i} . The effective Hamiltonian $\hat{H}_{\text{eff}}^A = \hat{H}_A + \hat{V}_{\text{eff}}^A$ of the subregion A contains the existing terms \hat{H}_A within the subsystem and the effective interacting terms \hat{V}_{eff}^A from the trace operation [39]. According to the conclusions of the quantum belief propagation theory [39,40], the locality of the interaction in the latter term $\|\hat{V}_{\text{eff}}^A(l_A)\|_F \propto (\beta/\beta_c)^{l_A/r}$, where β (β_c) denotes the current (model-dependent critical) inverse temperature, l_A is the distance between a specific site in the bulk of A and the boundary of the subregion A , and r is the maximum acting distance(diameter) of a single operator in the original Hamiltonian(similar to the k -local in the next section). Thus, when $\beta < \beta_c$ (especially when $\beta \ll \beta_c$), \hat{V}_{eff}^A is exponentially localized around the boundary of A , and the effective Hamiltonian in the bulk of A remains the same as that of the original \hat{H}_A of the entire system. Therefore, we may further boost the efficiency of MLE Hamiltonian learning by redirecting the expectation-value evaluations of the global quantum system to that of a series of local patches, as we will show in the next section.

In summary, given the quantum measurements of a thermal (Gibbs) state: $\{\hat{O}_i\}$, N_i , and f_{λ_i} , we can perform MLE

Hamiltonian learning to obtain the MLE Hamiltonian via the following steps (Fig. 1):

(i) Initialization/Update:

For initialization, start with a random model Hamiltonian \hat{H}_0 :

$$\hat{H}_0 = \sum_i \mu_i \hat{O}_i, \quad (12)$$

or an identity Hamiltonian.

For update, carry out the quantum likelihood gradient:

$$\hat{H}_{k+1} = \hat{H}_k - \gamma \hat{R}_k, \quad (13)$$

where \hat{R}_k is defined in Eq. (3) or Eq. (16).

(ii) Evaluate the properties $\text{tr}[\hat{\rho}_k \hat{P}_{\lambda_i}]$ of the quantum state:

$$\hat{\rho}_k = \frac{e^{-\beta \hat{H}_k}}{\text{tr}[e^{-\beta \hat{H}_k}]}, \quad (14)$$

with suitable numerical methods.

(iii) Check for convergence: loop back to step (1) to update, $k \rightarrow k+1$, if the relative entropy $M(\hat{\rho}_k) - M_0 \geq \epsilon$ is above a given threshold ϵ ; otherwise, terminate the process, and the final \hat{H}_k is the result for the MLE Hamiltonian. Here, M_0 is the theoretical minimum of the negative-log-likelihood function:

$$M_0 = - \sum_{i, \lambda_i} \frac{N_i}{N_{\text{tot}}} p_{\lambda_i} \log p_{\lambda_i}. \quad (15)$$

In practice, \hat{R}_k in Eq. (4) is singular for small values of $\text{tr}[\hat{\rho}_k \hat{P}_{\lambda_i}]$ and may become numerically unstable, which requires a minimal or dynamical learning rate γ to maintain the range of quantum likelihood gradient properly. Instead, we may employ a rescaled version of \hat{R}_k :

$$\tilde{R}_k = \sum_{i, \lambda_i} \frac{N_i}{N_{\text{tot}}} f_g(p_{\lambda_i} / \text{tr}[\hat{\rho}_k \hat{P}_{\lambda_i}]) \hat{P}_{\lambda_i}, \quad (16)$$

where f_g is a monotonic tuning-function:

$$f_g(x) = \frac{gx}{x + g - 1}, \quad g > 1, g \in \mathbb{N}, \quad (17)$$

which maps its argument in $(0, \infty)$ to a finite range $(0, g)$. Such a rescaled \tilde{R}_k regularizes the quantum likelihood gradient and allows a simple yet relatively larger learning rate γ for more efficient MLE Hamiltonian learning. We also have $f_g(1) = 1$, therefore $\tilde{R}_k \rightarrow \hat{R}_k$ as we approach convergence. We will mainly employ \tilde{R}_k for our examples in the following sections.

In addition to the negative-log-likelihood function $M(\hat{\rho}_k)$, we also consider the Hamiltonian distance as another criterion on the quality of Hamiltonian learning:

$$\Delta \vec{\mu}_k = \frac{\|\vec{\mu}_s - \vec{\mu}_k\|_2}{\|\vec{\mu}_s\|_2}, \quad (18)$$

where $\vec{\mu}_s$ and $\vec{\mu}_k$ are the (vectors of) coefficients [41] of the target Hamiltonian and the learned Hamiltonian after k iterations, respectively. However, we do not recommend $\Delta \vec{\mu}_k$ ($\Delta \mu$

for short) as convergence criteria as $\vec{\mu}_s$ is generally unknown aside from benchmark scenarios [28].

III. EXAMPLE MODELS AND RESULTS

In this section, we demonstrate the performance of the MLE Hamiltonian learning protocol. For better numerical simulations, we consider the k -local Hamiltonians, with operators acting nontrivially on no more than k contiguous sites in each direction. For example, for a one-dimensional spin- $\frac{1}{2}$ system, a k -local operator for $k = 2$ takes the form $\hat{S}_i^\alpha \hat{S}_{i+1}^\beta$ or \hat{S}_i^α , $\alpha, \beta \in \{x, y, z\}$, where \hat{S}_i^α denotes the spin operator. In particular, we focus on general 1D quantum spin chains with $k = 2$, taking the following form:

$$\hat{H}_s = \sum_{i, \alpha, \beta}^{L-1} J_i^{\alpha\beta} \hat{S}_i^\alpha \hat{S}_{i+1}^\beta + \sum_{i, \alpha}^L h_i^\alpha \hat{S}_i^\alpha, \quad (19)$$

where \hat{S}_i^α denotes the spin operator on site i , $\alpha, \beta \in \{x, y, z\}$. There are $12L - 9$ 2-local operators under the open boundary condition, where L is the system size. We generate the model parameters $\vec{\mu}_s = \{J_i^{\alpha\beta}, h_i^\alpha\}$ randomly following a uniform distribution in $[-1, 1]$. This Hamiltonian \hat{H}_s , specifically the model parameters $\vec{\mu}_s$, will be our target for MLE Hamiltonian learning. As the protocol's inputs, we simulate quantum measurements of all 2-local operators $\{\hat{O}_i\}$ on the Gibbs states of \hat{H}_s numerically via exact diagonalization on small systems and FTTN for large systems. For the latter, we use a tensor network ansatz called the ‘‘ancilla’’ method [34], where we purify a Gibbs state with some auxiliary qubits $\hat{\rho}_s = \text{tr}_{\text{aux}} |\psi_s\rangle \langle \psi_s|$, and obtain $|\psi_s\rangle = e^{-\beta \hat{H}_s/2} |\psi_0\rangle$ from a maximally entangled state $|\psi_0\rangle$ via imaginary time evolution. In addition, given a large number n of Trotter steps, the imaginary time evolution operator $e^{-\beta \hat{H}_s/2}$ is decomposed into Trotter gates' product as $(\prod_i e^{-\mu_i \hat{O}_i \beta/2n})^n + O(\beta^2/n)$. Here, we set the Trotter step $\delta t = \beta/n \in [0.01, 0.1]$, for which the Trotter errors of order $O(\beta^2/n)$ show little impact on our protocol's accuracy. Without loss of generality, we employ the integrated FTTN algorithm in the ITensor numerical toolkit [35], and set the number of measures $N_i = N$ for all operators in our examples for simplicity.

As we demonstrate in Fig. 2, MLE Hamiltonian learning obtains the target Hamiltonians with high accuracy and efficiency under various settings of system sizes and inverse temperatures β . Besides, instead of the original quantum likelihood gradient in Eq. (3), we may obtain a faster convergence with the rescaled \tilde{R}_k in Eq. (16) and a larger learning rate, as we discuss in Appendix B. In the following numerical examples, we use the rescaled quantum likelihood gradient \tilde{R}_k and set $g = 2$ for the tuning function in Eq. (17). Within the given iterations, not only have we achieved results [Hamiltonian distance $\Delta \mu \sim O(10^{-12})$ and relative entropy $M(\hat{\rho}_k) - M_0 \sim O(10^{-16})$] comparable to, if not exceeding, previous methods [19] for $L = 10$ systems and $\beta = 1$ straightforwardly, but we have also achieved satisfactory consistency [$\Delta \mu \sim O(10^{-2})$ and $M(\hat{\rho}_k) - M_0 \sim O(10^{-9})$] for large systems $L = 100$ and low temperatures $\beta = 3$ that were previously inaccessible.

MLE Hamiltonian learning is also relatively robust against temperature and noises, two key factors impacting accuracy

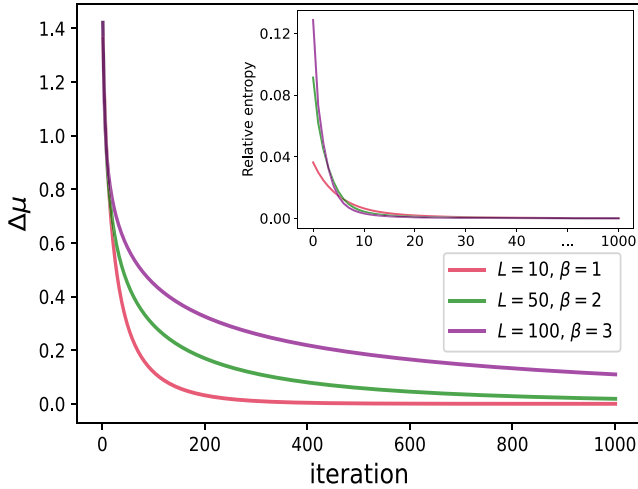


FIG. 2. Both the Hamiltonian distance $\Delta\mu$ defined in Eq. (18) and the negative-log-likelihood function $M(\hat{\rho}_{k+1})$ [or relative entropy $M(\hat{\rho}_{k+1}) - M_0$] show successful convergence of the iterations in MLE Hamiltonian learning, albeit a variety of system sizes and temperature ranges. We simulate the target Hamiltonian and the iteration process by FTTN with Trotter step $\delta t = 0.1$. Each curve is averaged on 10 trials of random \hat{H}_0 initializations. We set the learning rate $\gamma = 0.1$. The maximum number of iterations here is 1000.

in Hamiltonian learning. For illustration, we include random errors $\delta\langle\hat{O}_i\rangle$ following Gaussian distribution with zero mean and standard deviation δ to all quantum measurements: $\langle\hat{O}_i\rangle \rightarrow \langle\hat{O}_i\rangle + \delta\langle\hat{O}_i\rangle$. We note that such δ may also depict the quantum fluctuations [19,21] from a finite number of measurements $\delta \propto N_i^{-1/2}$. We also focus on smaller systems with $L = 10$ and employ exact diagonalization to avoid confusion from potential Trotter error of the FTTN ansatz [34]. We summarize the results in Fig. 3.

Most previous algorithms on Hamiltonian learning have a rather specific applicable temperature range. For example, the high-temperature expansion of $e^{-\beta\hat{H}}$ only works in the $\beta \ll 1$ limit [42,43]. Besides, gradient descent on the log partition function, despite a convex optimization, performs well in a narrow temperature range [20]. The gradient of this algorithm is proportional to the inverse temperature, so the algorithm's convergence slows at high temperatures. Also, the gradient descent algorithm cannot extend to the $\beta \rightarrow \infty$ limit—the ground state, while our protocol is directly applicable to the ground states of quantum systems, as we will generalize and justify later.

MLE Hamiltonian learning is also more robust to noises, with an accuracy of Hamiltonian distance $\Delta\mu \sim O(10^{-11})$ across a broad temperature range at noise strength $\delta \sim O(10^{-12})$. Such noise level is hard to realize in practice; nevertheless, it is necessary to safeguard the correlation matrix method [17,19,44]. Even so, due to the uncontrollable spectral gap, the correlation matrix method is susceptible to high temperature, and its accuracy drastically decreases to $\Delta\mu \sim O(10^{-3})$ at $\beta = 0.01$. In comparison, MLE Hamiltonian learning is more versatile, with an approximately linear dependence between its accuracy $\Delta\mu$ and the noise strength δ across a broad range of temperatures and noise strengths,

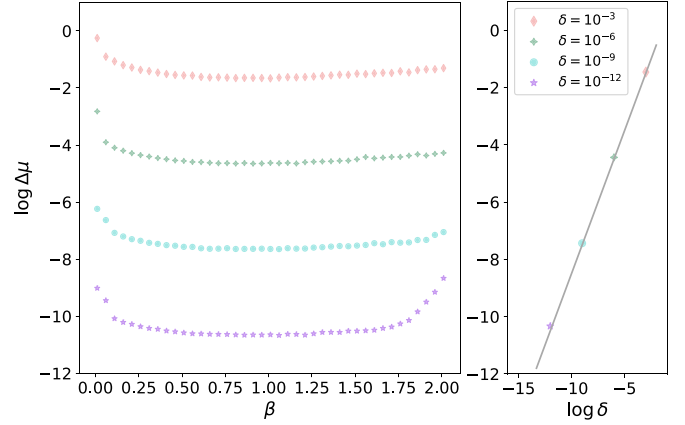


FIG. 3. The performance of MLE Hamiltonian learning maintains relatively well against noises and, especially, broader temperature ranges. Left: the Hamiltonian distance versus the inverse temperature β shows a broader applicable temperature range. Each data point contains 10 trials. Right: the performance (left figure's data averaged over temperature) versus the noise strength δ shows the impact of noises and the protocol's relative robustness against them. The slope of the straight line is ~ 1 , indicating a linear relationship between $\Delta\mu$ and δ . Note the log scale $\log(\Delta\mu)$ for the vertical axis. We set $L = 10$ for the system size, and learning rate $\gamma = 1$.

saturating the previous bound [20]; see the right panel of Fig. 3. We also provide more detailed comparisons between the algorithms in Appendix C.

Despite efficient quantum likelihood gradient and applicable quantum many-body ansatz, the computational cost of MLE Hamiltonian learning still increases rapidly with the system size L . Fortunately, as stated above in Eq. (11), we may resort to calculations on local patches, especially for low dimensions and high temperatures due to their quasi-Markov property. In particular, when $\beta < \beta_c$ ($T > T_c$), the difference between the cutoff Hamiltonian \hat{H}_A and the effective Hamiltonian \hat{H}_{eff}^A in a local subregion A , \hat{V}_{eff}^A , should be weak, short-ranged, and localized at A 's boundary [39,40]; therefore, for those operators \hat{P}_{λ_i} adequately deep inside A , we can use $\hat{\rho}^A$, the Gibbs state defined by \hat{H}_A , to estimate the corresponding $\text{tr}[\hat{\rho}\hat{P}_{\lambda_i}]$; see illustration in Fig. 4 upper panel.

For example, we apply MLE Hamiltonian learning on $L = 100$ systems, where we iteratively calculate the necessary expectation values on different local patches of size $L_A = 10, 16$. We also choose different cutoffs Λ , and evaluate $\text{tr}[\hat{\rho}^A\hat{P}_{\lambda_i}]$ for those operators at least Λ away from the boundaries and sufficiently deep inside the subregion A , so that the effective potential \hat{V}_{eff}^A may become negligible. We also employ a sufficient number of local patches to guarantee full coverage of necessary observables—operators outside A or in Λ_A are obtainable from another local patch B , as shown in the upper panel of Fig. 4, and so on so forth. Both the $L = 100$ target system and the local patches for MLE Hamiltonian learning are simulated via FTTN. We have no problem achieving convergence, and the resulting Hamiltonians' accuracy, the Hamiltonian distance $\Delta\mu$ versus the inverse temperature β , is summarized in the lower panel of Fig. 4. Indeed, the local-patch approximation is more reliable at higher temperatures, as well as with larger subsystems and cutoffs, albeit

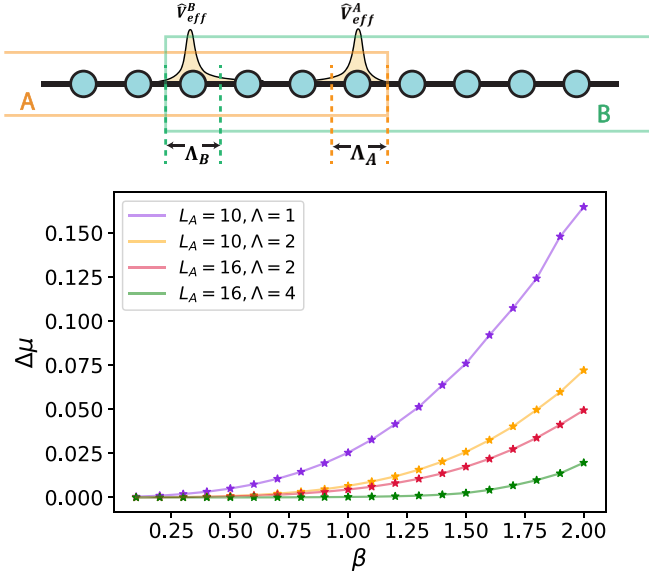


FIG. 4. Upper: MLE Hamiltonian learning's need for evaluations of local observables can be satisfied among local patches for thermal states at sufficiently high temperatures, where the effective potential \hat{V}_{eff}^A (\hat{V}_{eff}^B) is weak and localized on the boundaries of subregion A (B) within a cutoff range Λ_A (Λ_B). Consequently, for \hat{P}_{λ_i} defined sufficiently deep inside A (B), we can estimate its expectation value via $\text{tr}[\hat{\rho}^A \hat{P}_{\lambda_i}]$ ($\text{tr}[\hat{\rho}^B \hat{P}_{\lambda_i}]$). Lower: the Hamiltonian distance $\Delta\mu$ of the results after MLE Hamiltonian learning indicates better validity of the local-patch approximation at higher temperatures. The total system size is $L = 100$, while the local patches are of sizes $L_A = 10, 16$ with cutoffs $\Lambda = 1, 2, 4$, respectively. Each data point contains 10 trials. We set $\delta t = 0.1$ for the Trotter step in FTTN ansatz and learning rate $\gamma = 1$.

with rising costs. We also note that we can achieve much larger systems with the local patches than $L = 100$ we have demonstrated.

IV. MLE HAMILTONIAN LEARNING FOR PURE EIGENSTATES

In addition to the Gibbs states, MLE Hamiltonian learning also applies to measurements of certain eigenstates of target quantum systems:

(i) The ground states are essentially the $\beta \rightarrow \infty$ limit of the Gibbs states. However, due to the order-of-limit issue, the $\gamma \rightarrow 0$ requirement of the theorem on Gibbs states forbids a direct extension to the ground states. In the Appendix D, we offer rigorous proof of the effectiveness of quantum likelihood gradient based on ground-state measurements, along with several nontrivial MLE Hamiltonian learning examples on quantum critical and topological ground states. We note that Ref. [45] offers preliminary studies on pure-state quantum state tomography, inspiring this work.

(ii) A highly excited eigenstate of a (nonintegrable) quantum chaotic system \hat{H}_s is believed to obey the eigenstate thermalization hypothesis (ETH), that its density operator $\hat{\rho}_s = |\psi_s\rangle\langle\psi_s|$ behaves locally indistinguishable from a Gibbs

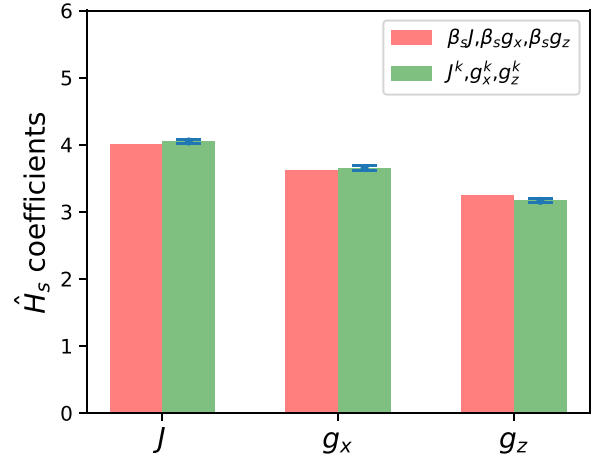


FIG. 5. The coefficients obtained via MLE Hamiltonian learning (green columns) compare well with those of the target Hamiltonian \hat{H}_s even though the quantum measurements are based upon a low-lying (first) excited state. The red columns denote the coefficients of \hat{H}_s in Eq. (21) multiply the effective (inverse) temperature $\beta_s = 4$. The error bars demonstrate the variances over the lattices and trials. We set the system size $L = 80$, learning rate $\gamma = 0.1$, and the Trotter step $\delta t = 0.1$.

state $\hat{\rho}_{s,A}$ in thermal equilibrium [46]:

$$\hat{\rho}_{s,A} = \text{tr}_A[\hat{\rho}_s] \approx \frac{e^{-\beta_s \hat{H}_A}}{\text{tr}[e^{-\beta_s \hat{H}_A}]}, \quad (20)$$

where β_s is an effective temperature determined by the energy expectation value $\langle\psi_s|\hat{H}_s|\psi_s\rangle = \frac{\text{tr}[e^{-\beta_s \hat{H}_s} \hat{H}_s]}{\text{tr}[e^{-\beta_s \hat{H}_s}]}$. As MLE Hamiltonian learning only engages local operators, its applicability directly generalizes to such eigenstates $|\psi_s\rangle$ following ETH.

(iii) In general, ETH applies to eigenstates in the center of the spectrum of quantum chaotic systems, while low-lying eigenstates are too close to the ground state to exhibit ETH [47]. However, in the rest of the section, we demonstrate numerically that MLE Hamiltonian learning still works well for low-lying eigenstates.

We consider the 1D longitudinal-transverse-field Ising model [46,47] as our target quantum system:

$$\hat{H}_s = J \sum_j \hat{\sigma}_j^z \hat{\sigma}_{j+1}^z + g_z \sum_j \hat{\sigma}_j^z + g_x \sum_j \hat{\sigma}_j^x, \quad (21)$$

where the system size is $L = 80$. We set $J = 1$, $g_x = 0.9045$, and $g_z = 0.8090$. The quantum system is strongly nonintegrable under such settings. Previous studies mainly focused on eigenstates in the middle of the energy spectrum. In contrast, we pick the first excited state—a typical low-lying eigenstate considered asymptotically integrable and ETH-violating [47]—for quantum measurements (via DMRG) and then MLE Hamiltonian learning for its candidate Hamiltonian (via FTTN).

We summarize the results in Fig. 5. Further, the model Hamiltonian we established is approximately equivalent to the target quantum Hamiltonian at an (inverse) temperature $\beta_s \approx 4$ [46], which we have absorbed into the unit of our \hat{H}_k . Therefore, we have accurately established the model

Hamiltonian and derived the effective temperature consistent with previous results [46] for a low-lying excited eigenstate not necessarily following ETH. The physical reason for quantum likelihood gradient applicability in such states is an interesting problem that deserves further studies.

V. DISCUSSIONS

We have proposed a novel MLE Hamiltonian learning protocol to achieve the model Hamiltonian of the target quantum system based on quantum measurements of its Gibbs states. The protocol updates the model Hamiltonian iteratively with respect to the negative-log-likelihood function from the measurement data. We have theoretically proved the efficiency and convergence of the corresponding quantum likelihood gradient and demonstrated it numerically on multiple nontrivial examples, which show more accuracy, better robustness against noises, and less temperature dependence. Indeed, the accuracy is almost linear to the imposed noise amplitude, thus inverse proportional to the square root of the number of samples, the asymptotic upper bound [20]. Further, MLE Hamiltonian learning directly rests on the Hamiltonians and their physical properties instead of direct and costly access to the quantum many-body states. Consequently, we can resort to various quantum many-body ansatzes in our systematic quantum toolbox and even local-patch approximation when the situation allows. These advantages allow applications to larger systems and lower temperatures with better accuracy than previous approaches. However, while our protocol is generally applicable for learning any Hamiltonian, its advantages are most apparent for local Hamiltonians, where various quantum many-body ansatzes and local-patch approximation shine. Despite such limitations, we note that the physical systems are characterized by local Hamiltonians in a significant proportion of scenarios.

In addition to the Gibbs states, we have generalized the applicability of MLE Hamiltonian learning to eigenstates of the target quantum states, including ground states, ETH states, and even selected cases of low-lying excited states. We have also provided theoretical proof of quantum likelihood gradient rigor and convergence in the Appendix D, along with several other numerical examples.

Our strategy may apply to the entanglement Hamiltonians and the tomography of the quantum states under the maximum-likelihood-maximum-entropy assumption [48]. Besides, our algorithm may also provide insights into the quantum Boltzmann machine [36]—a quantum version of the classical Boltzmann machine with degrees of freedom that obey the distribution of a target quantum Gibbs state. Instead of brute-force calculations of the loss function derivatives with respect to the model parameters or approximations with the gradients' upper bounds, our protocol provides an efficient optimization that updates the model parameters collectively.

ACKNOWLEDGMENTS

We thank Jia-Bao Wang for insightful discussions. We acknowledge support from the National Key R&D Program of China (Grant No. 2021YFA1401900) and the National Natural Science Foundation of China (Grants No. 12174008 and

No. 92270102). The calculations of this work are supported by HPC facilities at Peking University.

APPENDIX A: MAXIMUM CONDITION FOR MLE

In this Appendix, we review the derivation of the maximum condition [30] in Eq. (3) in the main text.

A general quantum state takes the form of a density operator:

$$\hat{\rho} = \sum_j p_j |\psi_j\rangle\langle\psi_j|, \quad (\text{A1})$$

where $p_j \geq 0$, $\sum_j p_j = 1$, and $|\psi_j\rangle$ is a set of orthonormal basis. The search for the quantum state that maximizes the likelihood function:

$$\mathcal{L}(\hat{\rho}) = \prod_{i, \lambda_i} \{ \text{tr}[\hat{\rho} \hat{P}_{\lambda_i}]^{f_{\lambda_i}} \}^{N_{\text{tot}}}, \quad (\text{A2})$$

can be converted to the optimization problem:

$$\begin{aligned} \min_{\hat{\rho} \in \mathcal{D}} \quad & M(\hat{\rho}) = -\frac{1}{N_{\text{tot}}} \log[\mathcal{L}(\hat{\rho})] \\ \text{subject to} \quad & \hat{\rho} \geq 0, \text{tr}[\hat{\rho}] = 1. \end{aligned} \quad (\text{A3})$$

It is hard to solve this semidefinite programming problem directly and numerically. Instead, forgoing the nonnegative definiteness, we adopt the Lagrangian multiplier method:

$$\frac{\partial}{\partial \langle \psi_j |} \{ M(\hat{\rho}) + \lambda \text{tr}[\hat{\rho}] \} = 0, \quad (\text{A4})$$

where λ is a Lagrangian multiplier. Given Eq. (A1), we obtain the following solution:

$$\begin{aligned} \hat{R}|\psi_j\rangle &= |\psi_j\rangle, \\ \hat{R} &= \sum_{i, \lambda_i} \frac{f_{\lambda_i}}{N_{\text{tot}}} \frac{\hat{P}_{\lambda_i}}{\text{tr}[\hat{\rho} \hat{P}_{\lambda_i}]}, \end{aligned} \quad (\text{A5})$$

and $\lambda = 1$. Combining Eqs. (A1) and (A5), we obtain the maximum condition:

$$\hat{R}\hat{\rho} = \hat{\rho}. \quad (\text{A6})$$

We note that Eq. (A5) does not guarantee the positive semidefiniteness of the density operator. Instead, one may search within the density-operator space (the space of positive semidefinite matrix with unit trace) to locate the MLE quantum state fulfilling Eq. (A5) or Eq. (A6). For the Hamiltonian learning task in this work, the search space is naturally the space of Gibbs states (under selected quantum many-body ansatz).

APPENDIX B: MLE HAMILTONIAN LEARNING WITH RESCALING FUNCTION

In this Appendix, we compare the MLE Hamiltonian learning with the quantum likelihood gradient \hat{R}_k and the rescaled counterpart $\hat{\tilde{R}}_k$. As we state in the main text, $\hat{\tilde{R}}_k$ regularizes the gradient, allowing us to employ a larger learning rate $\gamma = 1$, which leads to a faster convergence (Fig. 6) and a higher accuracy (Table I) given identical number of iterations.

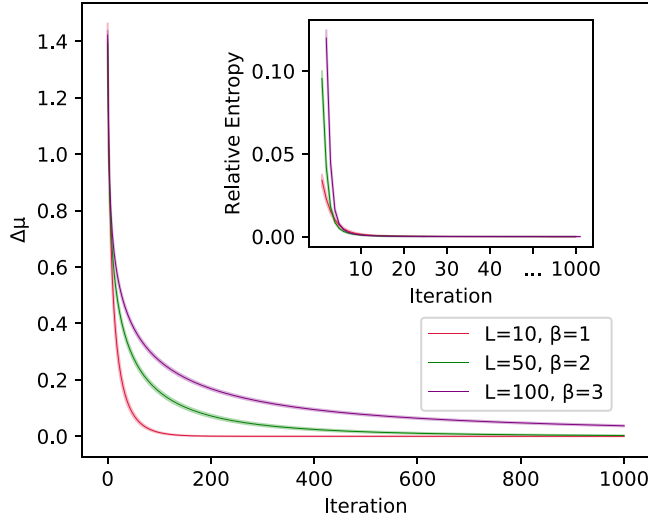


FIG. 6. Both the Hamiltonian distance $\Delta\mu$ defined in Eq. (18) and the negative-log-likelihood function $M(\hat{\rho}_{k+1})$ [or relative entropy $M(\hat{\rho}_{k+1}) - M_0$] show successful convergence of the iterations in MLE Hamiltonian learning, albeit a variety of system sizes and temperature ranges. We simulate the target Hamiltonian and the iteration process by FTTN with Trotter step $\delta t = 0.1$. Each curve is averaged on 10 trials of random \hat{H}_0 initializations. The maximum number of iterations here is 1000. In comparison with Fig. 2 in the main text, we use the rescaled \tilde{R}_k in Eq. (16) in the main text with $g = 2$, which allows us to employ a larger learning rate $\gamma = 1$ and obtain a faster convergence.

APPENDIX C: COMPARISONS BETWEEN HAMILTONIAN LEARNING ALGORITHMS

In this Appendix, we compare different Hamiltonian learning algorithms, including the correlation matrix (CM) method [17,19], the gradient descent (GD) method [20], and the MLE Hamiltonian learning (MLEHL) algorithm, by looking into some of their numerical results and performances. We consider general 2-local Hamiltonians in Eq. (10) in the main text for demonstration and measurements $\{\hat{O}_i\}$ over all the 2-local operators (instead of all 4-local operators as in Ref. [19]).

We summarize the results in Fig. 7: the accuracy of CM is unstable and highly sensitive to temperature; while GD performs similarly to the proposed MLEHL algorithm at low temperatures, its descending gradient becomes too small at high temperatures to allow a satisfactory convergence within the given maximum iterations.

We also compare the convergence rates of the MLEHL and GD algorithms with the same learning rate. As in Fig. 8, the

TABLE I. The algorithm's accuracy (Hamiltonian distance $\Delta\mu$) further improves with a rescaled quantum likelihood gradient \tilde{R}_k under various system sizes L and (inverse) temperatures β .

	$L = 10, \beta = 1$	$L = 50, \beta = 2$	$L = 100, \beta = 3$
\hat{R}_k	$O(10^{-6})$	$O(10^{-2})$	$O(10^{-1})$
\tilde{R}_k	$O(10^{-12})$	$O(10^{-3})$	$O(10^{-2})$

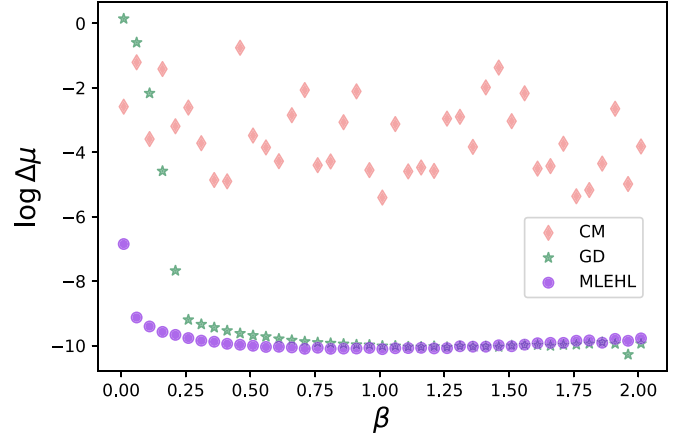


FIG. 7. The performances (logarithm of Hamiltonian distances) of different algorithms versus the inverse temperature β show the advantages of the MLEHL algorithm. Each data point contains 10 trials of random Hamiltonians. We also include noises following a normal distribution with zero means and $O(10^{-12})$ standard deviation. For both GD and MLEHL algorithms, we employ a learning rate $\gamma = 1$ and a maximum number of iterations of 7000. The system size is $L = 7$.

MLEHL algorithm exhibits a faster convergence and a smaller computational cost, which is similar under both algorithms for each iteration.

APPENDIX D: HAMILTONIAN LEARNING FROM GROUND STATE

In this Appendix, we prove the effectiveness of the quantum likelihood gradient based on measurements of the target quantum system's ground state and provide several nontrivial numerical examples, including 1D quantum critical states and 2D topological states.

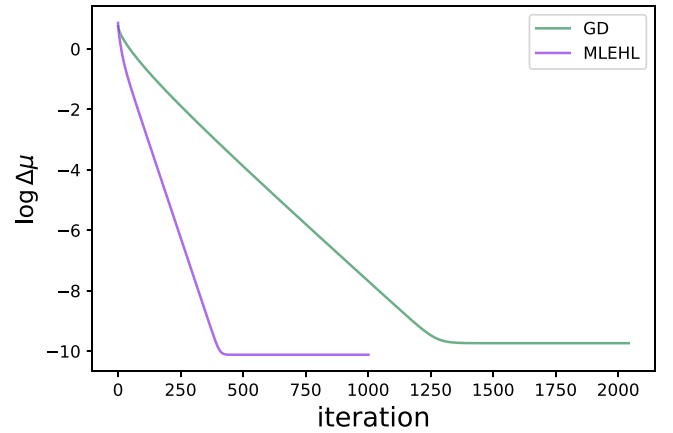


FIG. 8. The logarithm of the Hamiltonian distances versus the number of iterations shows a faster convergence under the proposed MLEHL algorithm. We also include noises following a normal distribution with zero means and $O(10^{-12})$ standard deviation. We employ a learning rate of $\gamma = 1$ for both GD and MLEHL. We set the system size $L = 7$ and the (inverse) temperature $\beta = 1$.

1. Proof for ground-state-based quantum likelihood gradient

Given a sufficient number N_i measurements of the operator \hat{O}_i on the nondegenerate ground state $|\psi_s\rangle$ of a target system \hat{H}_s , we obtain a number of outcomes as the λ_i^{th} eigenvalue of \hat{O}_i as

$$f_{\lambda_i} = p_{\lambda_i} N_i \approx \langle \psi_s | \hat{P}_{\lambda_i} | \psi_s \rangle N_i, \quad (\text{D1})$$

where $p_{\lambda_i} = f_{\lambda_i}/N_i$, and \hat{P}_{λ_i} is the projection operator of the eigenvalue o_{λ_i} .

Our MLE Hamiltonian learning follows the iterations

$$\begin{aligned} \hat{H}_{k+1} &= \hat{H}_k - \gamma \hat{R}_k, \\ \hat{R}_k &= \sum_{i, \lambda_i} \frac{f_{\lambda_i}}{N_{\text{tot}}} \frac{\hat{P}_{\lambda_i}}{\langle \psi_k^{\text{gs}} | \hat{P}_{\lambda_i} | \psi_k^{\text{gs}} \rangle}, \end{aligned} \quad (\text{D2})$$

where $|\psi_k^{\text{gs}}\rangle$ is the nondegenerate ground state of \hat{H}_k .

Theorem. For $\gamma \ll 1, \gamma > 0$, the quantum likelihood gradient in Eq. (D2) yields a negative semidefinite contribution to the negative-log-likelihood function $M(|\psi_{k+1}^{\text{gs}}\rangle) = -\frac{1}{N_{\text{tot}}} \log \mathcal{L}(|\psi_{k+1}^{\text{gs}}\rangle)$ following Eq. (2) in the main text.

Proof. At the linear order in γ , we may treat the addition of $-\gamma \hat{R}_k$ to \hat{H}_k at the k th iteration as a perturbation:

$$|\psi_{k+1}^{\text{gs}}\rangle = |\psi_k^{\text{gs}}\rangle - \gamma \hat{G}_k \hat{R}_k |\psi_k^{\text{gs}}\rangle + O(\gamma^2), \quad (\text{D3})$$

where \hat{G}_k is the Green's function in the k th iteration:

$$\hat{G}_k = \hat{Q}_k \frac{1}{E_k^{\text{gs}} - \hat{H}_k} \hat{Q}_k, \quad (\text{D4})$$

where $\hat{Q}_k = I - |\psi_k^{\text{gs}}\rangle \langle \psi_k^{\text{gs}}|$ is the projection operator orthogonal to the ground space $|\psi_k^{\text{gs}}\rangle \langle \psi_k^{\text{gs}}|$, and E_k^{gs} is the ground-state energy. Keeping terms up to the linear order of γ in the log expansion of the negative-log-likelihood function, we have

$$\begin{aligned} M(|\psi_{k+1}^{\text{gs}}\rangle) &= -\frac{1}{N_{\text{tot}}} \log \mathcal{L}(|\psi_{k+1}^{\text{gs}}\rangle) \\ &= -\sum_{i, \lambda_i} \frac{f_{\lambda_i}}{N_{\text{tot}}} \log \langle \psi_{k+1}^{\text{gs}} | \hat{P}_{\lambda_i} | \psi_{k+1}^{\text{gs}} \rangle, \\ &\approx M(|\psi_k^{\text{gs}}\rangle) + 2\gamma \Delta_k. \end{aligned} \quad (\text{D5})$$

where difference takes the form:

$$\begin{aligned} \Delta_k &= \langle \psi_k^{\text{gs}} | \hat{R}_k \hat{G}_k \hat{R}_k | \psi_k^{\text{gs}} \rangle \\ &= \sum_{l \neq \text{gs}} \frac{|\langle \psi_k^{\text{gs}} | \hat{R}_k | \psi_l^{\text{gs}} \rangle|^2}{E_k^{\text{gs}} - E_l^{\text{gs}}} \leq 0. \end{aligned} \quad (\text{D6})$$

Here, $E_l^{\text{gs}} > E_k^{\text{gs}}$ because E_l^{gs} denotes the energy for eigenstates other than the ground state. Our iteration converges when the equality in Eq. (D6) is established. This happens when $|\psi_k^{\text{gs}}\rangle$ is an eigenstate of \hat{R}_k , consistent with the MLE condition $\hat{R}|\psi\rangle = |\psi\rangle$ (or $\hat{R}\hat{\rho} = \hat{\rho}$).

Finally, combining Eqs. (D5) and (D6), we have shown that $M(|\psi_{k+1}^{\text{gs}}\rangle) - M(|\psi_k^{\text{gs}}\rangle)$ is a negative semidefinite quantity, which proves the theorem. ■

One potential complication to the proof is that Eq. (D3) needs to assume there is no ground-state-level crossing or degeneracy after adding the quantum likelihood gradient. A

potential remedy is to keep some low-lying excited states together with the ground state and compare them for maximum likelihood, especially for steps with singular behaviors. Otherwise, we can only hope such transitions are sparse, especially near convergence, and they establish a new line of iterations heading toward the same convergence. A more detailed discussion is available in Ref. [45].

2. Example: $c = \frac{3}{2}$ CFT ground state of Majorana fermion chain

Here, we consider the spinless 1D Majorana fermion chain model of length $2L$ as an example [49]:

$$\hat{H}_s = \sum_j it \hat{\gamma}_j \hat{\gamma}_{j+1} + g \hat{\gamma}_j \hat{\gamma}_{j+1} \hat{\gamma}_{j+2} \hat{\gamma}_{j+3}, \quad (\text{D7})$$

where $\hat{\gamma}_j$ is the Majorana fermion operator obeying

$$\hat{\gamma}_j^\dagger = \hat{\gamma}_j, \{\hat{\gamma}_i, \hat{\gamma}_j\} = \delta_{ij}, \quad (\text{D8})$$

and t and $g = -1$ are model parameters. This model presents a wealth of nontrivial quantum phases under different t/g . We focus on the model parameters in $t/g \in (-2.86, -0.28)$, where the ground state of Eq. (D7) is a $c = \frac{3}{2}$ CFT composed of a critical Ising theory ($c = \frac{1}{2}$) and a Luttinger liquid ($c = 1$).

Through the definition of the complex fermions followed by the Jordan-Wigner transformation:

$$\begin{aligned} \hat{c}_j &= \frac{\hat{\gamma}_{2j} + i\hat{\gamma}_{2j+1}}{2}, \\ \hat{\sigma}_j^z &= 2\hat{n}_j - 1, \\ \hat{\sigma}_j^\pm &= e^{-i\pi \sum_{i<j} \hat{n}_i} \hat{c}_j^\dagger, \end{aligned} \quad (\text{D9})$$

where $\hat{n}_j = \hat{c}_j^\dagger \hat{c}_j$ is the complex fermion number operator, we map Eq. (D7) to a 3-local spin chain of length L :

$$\begin{aligned} \hat{H}_s &= t \sum_j \hat{\sigma}_j^z - t \sum_j \hat{\sigma}_j^x \hat{\sigma}_{j+1}^x \\ &\quad - g \sum_j \hat{\sigma}_j^z \hat{\sigma}_{j+1}^z - g \sum_j \hat{\sigma}_j^x \hat{\sigma}_{j+2}^x. \end{aligned} \quad (\text{D10})$$

We employ quantum measurements on the ground state $|\psi_s\rangle$ of this Hamiltonian, based on which we carry out our MLE Hamiltonian learning protocol. Here, we evaluate the ground-state properties via exact diagonalization. The numerical results for two cases of $t = 0.5, 1.5$ are in Fig. 9. We achieve successful convergence and satisfactory accuracy on the target Hamiltonian. The relative entropy's instabilities are mainly due to the ground-state-level crossing and degeneracy.

3. Example: Alternative Hamiltonian for ground state

We have seen that MLE Hamiltonian learning can retrieve the unknown target Hamiltonians via quantum measurements of its Gibbs states, even its ground states. For pure states, however, one interesting byproduct is that the relation between Hamiltonian and eigenstates is essentially many-to-one. Therefore, it is possible to obtain various candidate Hamiltonians \hat{H}_k sharing the same ground state as the original target \hat{H}_s , especially by controlling the operator/observable set. Here, we show such numerical examples.

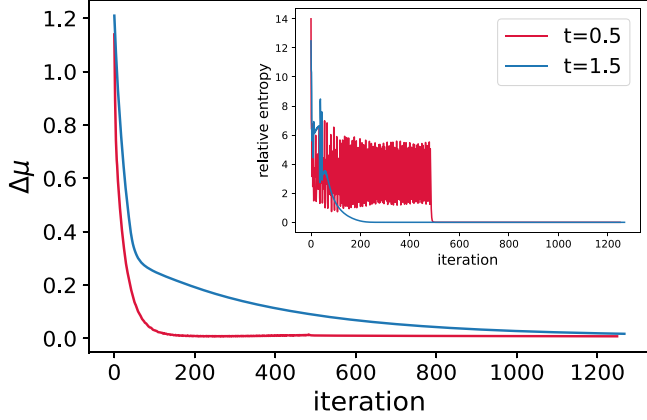


FIG. 9. Both the Hamiltonian distance $\Delta\mu$ and the relative entropy $M(|\psi_{k+1}^{\text{gs}}\rangle) - M_0$ (inset) as defined in the main text indicate successful convergence of the iterations during MLE Hamiltonian learning. We set $g = -1$, $t = 0.5$ (red curve) or $t = 1.5$ (blue curve), and system size $L = 12$ for the target quantum system, and learning rate $\gamma = 0.1$ ($\gamma = 0.05$) before (after) the 490th iteration.

As our target quantum system, we consider the transverse field Ising model (TFIM) of length $L = 15$

$$\hat{H}_s = J \sum_j \hat{S}_j^z \hat{S}_{j+1}^z + g \sum_j \hat{S}_j^x, \quad (\text{D11})$$

at its critical point $J = g = 1$. Its ground state is $|\psi_s\rangle$. However, instead of the operators presenting in \hat{H}_s , we employ a different operator set for $|\psi_s\rangle$'s quantum measurements:

$$\{\hat{O}_i\} = \{\hat{S}_i^z \hat{S}_{i+1}^z, \hat{S}_i^x \hat{S}_{i+1}^x\}. \quad (\text{D12})$$

We evaluate the ground-state properties via DMRG.

The subsequent MLE Hamiltonian learning results are in Fig. 10. Since we obtain a candidate Hamiltonian with the operators in Eq. (D12) and destined to differ from \hat{H}_s , the Hamiltonian distance is no longer a viable measure of its accuracy. Instead, we introduce the ground-state fidelity

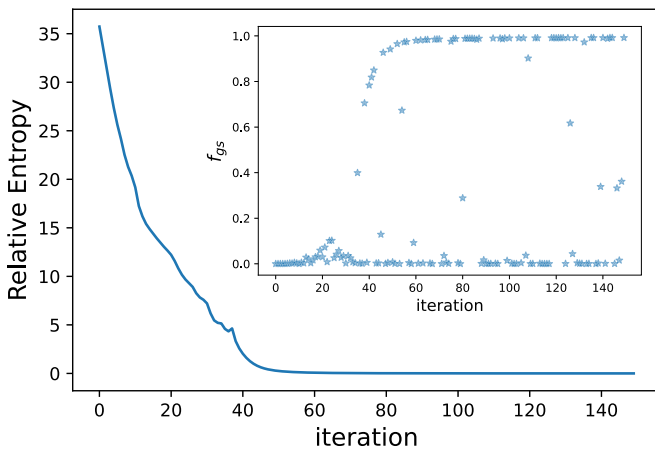


FIG. 10. Both the relative entropy and the fidelity $f_{\text{gs}} = \langle \psi_s | \psi_k^{\text{gs}} \rangle$ (inset) indicate successful convergence of the iterations during MLE Hamiltonian learning, yielding a consistent yet different Hamiltonian from the original quantum system. We set the system size $L = 15$ and learning rate $\gamma = 0.005$.

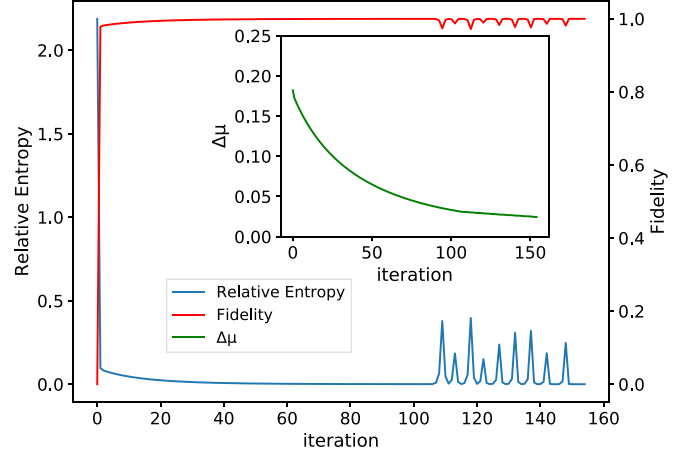


FIG. 11. The relative entropy, the fidelity $f_{\text{gs}} = \langle \psi_s | \psi_k^{\text{gs}} \rangle$, and the Hamiltonian distance $\Delta\mu$ (inset) show distinct convergence behaviors in the iterations of MLE Hamiltonian learning for a 2D topological CSL system. Our system size is 4×4 , and we set the learning rate $\gamma = 0.1$.

$f_{\text{gs}} = \langle \psi_s | \psi_k^{\text{gs}} \rangle$, where $|\psi_s\rangle$ ($|\psi_k^{\text{gs}}\rangle$) is the ground state of \hat{H}_s (\hat{H}_k). Interestingly, while the relative entropy shows full convergence, the fidelity f_{gs} jumps between $\sim 99.5\%$ and $\sim 10^{-3}\%$. This is understandable, as the quantum system is gapless, and the ground and low-lying excited states have similar properties under quantum measurements.

4. Example: Two-dimensional topological states

Here, we consider MLE Hamiltonian learning on two-dimensional topological quantum systems. In particular, we consider the chiral spin liquid (CSL) on a triangular lattice:

$$\hat{H}_s = J_1 \sum_{\langle ij \rangle} \vec{S}_i \cdot \vec{S}_j + J_2 \sum_{\langle\langle ij \rangle\rangle} \vec{S}_i \cdot \vec{S}_j + K \sum_{i,j,k \in \nabla / \Delta} \vec{S}_i \cdot (\vec{S}_j \times \vec{S}_k), \quad (\text{D13})$$

where the first and second terms are Heisenberg interactions, and the last term is a three-spin chiral interaction. Previous DMRG studies have established \hat{H}_s 's ground state as a CSL under the model parameters $J_1 = 1.0$, $J_2 = 0.1$, and $K = 0.2$ [50], which we set as the parameters of the target Hamiltonian. Here, we employ exact diagonalization on a 4×4 system. Based upon entanglement studies of the lowest-energy eigenstates, we verify that both the modular SU matrix corresponding to C_6 rotations and the entanglement entropy fit well with a CSL topological phase [51]. Subsequently, we perform MLE Hamiltonian learning based on quantum measurements of the ground state, focusing on the operators presenting in \hat{H}_s . We summarize the results in Fig. 11. The Hamiltonian distance indicates a stable converging accuracy, yet the relative entropy and the fidelity $f_{\text{gs}} = \langle \psi_s | \psi_k^{\text{gs}} \rangle$ witness certain instabilities. Indeed, being a topological phase means ground-state degeneracy—competing low-energy eigenstates with global distinctions yet similar local properties.

- [1] C. Lanczos, An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, *J. Res. Natl. Bureau Stand.* **45**, 255 (1950).
- [2] S. R. White, Density Matrix Formulation for Quantum Renormalization Groups, *Phys. Rev. Lett.* **69**, 2863 (1992).
- [3] U. Schollwoeck, The density-matrix renormalization group in the age of matrix product states, *Ann. Physics (NY)* **326**, 96 (2011).
- [4] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Quantum monte carlo simulations of solids, *Rev. Mod. Phys.* **73**, 33 (2001).
- [5] X. W. Zhang and Y. L. Liu, Electronic transport and spatial current patterns of 2D electronic system: A recursive Green's function method study, *AIP Adv.* **9**, 115209 (2019).
- [6] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information, 10th Anniversary Edition* (Cambridge University Press, Cambridge, 2010).
- [7] C. Nayak, S. H. Simon, A. Stern, M. Freedman, and S. Das Sarma, Non-Abelian anyons and topological quantum computation, *Rev. Mod. Phys.* **80**, 1083 (2008).
- [8] I. Buluta and F. Nori, Quantum simulators, *Science* **326**, 108 (2009).
- [9] I. M. Georgescu, S. Ashhab, and F. Nori, Quantum simulation, *Rev. Mod. Phys.* **86**, 153 (2014).
- [10] P. Barthélemy and L. M. K. Vandersypen, Quantum dot systems: A versatile platform for quantum simulations, *Ann. Phys.* **525**, 808 (2013).
- [11] A. Browaeys and T. Lahaye, Many-body physics with individually controlled Rydberg atoms, *Nat. Phys.* **16**, 132 (2020).
- [12] P. Scholl, M. Schuler, H. J. Williams, A. A. Eberharter, D. Barredo, K.-N. Schymik, V. Lienhard, L.-P. Henry, T. C. Lang, T. Lahaye, A. M. Läuchli, and A. Browaeys, Quantum simulation of 2D antiferromagnets with hundreds of Rydberg atoms, *Nature (London)* **595**, 233 (2021).
- [13] S. Ebadi, A. Keesling, M. Cain, T. T. Wang, H. Levine, D. Bluvstein, G. Semeghini, A. Omran, J.-G. Liu, R. Samajdar, X.-Z. Luo, B. Nash, X. Gao, B. Barak, E. Farhi, S. Sachdev, N. Gemelke, L. Zhou, S. Choi, H. Pichler *et al.*, Quantum optimization of maximum independent set using Rydberg atom arrays, *Science* **376**, 1209 (2022).
- [14] D. Bluvstein, A. Omran, H. Levine, A. Keesling, G. Semeghini, S. Ebadi, T. T. Wang, A. A. Michailidis, N. Maskara, W. W. Ho, S. Choi, M. Serbyn, M. Greiner, V. Vuletić, and M. D. Lukin, Controlling quantum many-body dynamics in driven Rydberg atom arrays, *Science* **371**, 1355 (2021).
- [15] R. Bistritzer and A. H. MacDonald, Moire bands in twisted double-layer graphene, *Proc. Natl. Acad. Sci. USA* **108**, 12233 (2011).
- [16] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Theory of superconductivity, *Phys. Rev.* **108**, 1175 (1957).
- [17] X.-L. Qi and D. Ranard, Determining a local Hamiltonian from a single eigenstate, *Quantum* **3**, 159 (2019).
- [18] M. Dupont, N. Macé, and N. Laflorencie, From eigenstate to Hamiltonian: Prospects for ergodicity and localization, *Phys. Rev. B* **100**, 134201 (2019).
- [19] E. Bairey, I. Arad, and N. H. Lindner, Learning a Local Hamiltonian from Local Measurements, *Phys. Rev. Lett.* **122**, 020504 (2019).
- [20] A. Anshu, S. Arunachalam, T. Kuwahara, and M. Soleimanifar, Sample-efficient learning of interacting quantum systems, *Nat. Phys.* **17**, 931 (2021).
- [21] J. Zhou and D. L. Zhou, Recovery of a generic local Hamiltonian from a steady state, *Phys. Rev. A* **105**, 012615 (2022).
- [22] X. Turkeshi, T. Mendes-Santos, G. Giudici, and M. Dalmonte, Entanglement-Guided Search for Parent Hamiltonians, *Phys. Rev. Lett.* **122**, 150606 (2019).
- [23] A. Valenti, G. Jin, J. Léonard, S. D. Huber, and E. Greplova, Scalable Hamiltonian learning for large-scale out-of-equilibrium quantum dynamics, *Phys. Rev. A* **105**, 023302 (2022).
- [24] W. Yu, J. Sun, Z. Han, and X. Yuan, Practical and efficient Hamiltonian learning, *arXiv:2201.00190*.
- [25] H.-Y. Huang, Y. Tong, D. Fang, and Y. Su, Learning Many-Body Hamiltonians with Heisenberg-Limited Scaling, *Phys. Rev. Lett.* **130**, 200403 (2023).
- [26] F. Wilde, A. Kshetrimayum, I. Roth, D. Hangleiter, R. Sweke, and J. Eisert, Scalably learning quantum many-body Hamiltonians from dynamical data, *arXiv:2209.14328*.
- [27] J. Wang, S. Paesani, R. Santagati, S. Knauer, A. A. Gentile, N. Wiebe, M. Petruzzella, J. L. O'Brien, J. G. Rarity, A. Laing, and M. G. Thompson, Experimental quantum Hamiltonian learning, *Nat. Phys.* **13**, 551 (2017).
- [28] J. Carrasco, A. Elben, C. Kokail, B. Kraus, and P. Zoller, Theoretical and experimental perspectives of quantum verification, *PRX Quant.* **2**, 010102 (2021).
- [29] Z. Hradil, Quantum-state estimation, *Phys. Rev. A* **55**, R1561 (1997).
- [30] J. Řeháček, Z. Hradil, E. Knill, and A. I. Lvovsky, Diluted maximum-likelihood algorithm for quantum tomography, *Phys. Rev. A* **75**, 042108 (2007).
- [31] A. I. Lvovsky, Iterative maximum-likelihood reconstruction in quantum homodyne tomography, *J. Opt. B: Quant. Semiclass. Opt.* **6**, S556 (2004).
- [32] Y. S. Teo, B. Stoklasa, B.-G. Englert, J. Řeháček, and Z. Hradil, Incomplete quantum state estimation: A comprehensive study, *Phys. Rev. A* **85**, 042317 (2012).
- [33] J. Fiurásek and Z. Hradil, Maximum-likelihood estimation of quantum processes, *Phys. Rev. A* **63**, 020101(R) (2001).
- [34] A. E. Feiguin and S. R. White, Finite-temperature density matrix renormalization using an enlarged Hilbert space, *Phys. Rev. B* **72**, 220401(R) (2005).
- [35] M. Fishman, S. R. White, and E. M. Stoudenmire, The ITensor software library for tensor network calculations, *SciPost Phys. Codebases* **4** (2022).
- [36] M. H. Amin, E. Andriyash, J. Rolfe, B. Kulchytskyy, and R. Melko, Quantum Boltzmann Machine, *Phys. Rev. X* **8**, 021050 (2018).
- [37] T. Kimura, Explicit description of the Zassenhaus formula, *Prog. Theor. Exp. Phys.* **2017**, 041A03 (2017).
- [38] In practice, given sufficient measurements, we have $\hat{R}_k \sim \hat{I}$ dictating the quantum likelihood gradient at the iteration's convergence.
- [39] T. Kuwahara, K. Kato, and F. G. S. L. Brandão, Clustering of Conditional Mutual Information for Quantum Gibbs States Above a Threshold Temperature, *Phys. Rev. Lett.* **124**, 220601 (2020).

- [40] E. Bilgin and D. Poulin, Coarse-grained belief propagation for simulation of interacting quantum systems at all temperatures, *Phys. Rev. B* **81**, 054106 (2010).
- [41] We typically perform MLE Hamiltonian learning and update the model Hamiltonian on the projection-operator basis; therefore, we transform the Hamiltonian back to the original, ordinary operator basis before $\Delta\tilde{\mu}_k$ evaluations.
- [42] J. Haah, R. Kothari, and E. Tang, Optimal learning of quantum Hamiltonians from high-temperature Gibbs states, in *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, Piscataway, NJ, 2022), pp. 135–146.
- [43] K. Rudinger and R. Joynt, Compressed sensing for Hamiltonian reconstruction, *Phys. Rev. A* **92**, 052322 (2015).
- [44] Tim J. Evans, R. Harper, and S. T. Flammia, Scalable Bayesian Hamiltonian learning, [arXiv:1912.07636](https://arxiv.org/abs/1912.07636).
- [45] J.-B. Wang and Y. Zhang, Single-shot quantum measurements sketch quantum many-body states, *Phys. Rev. B* **107**, L161101 (2023).
- [46] J. R. Garrison and T. Grover, Does a Single Eigenstate Encode the Full Hamiltonian? *Phys. Rev. X* **8**, 021026 (2018).
- [47] H. Kim, T. N. Ikeda, and D. A. Huse, Testing whether all eigenstates obey the eigenstate thermalization hypothesis, *Phys. Rev. E* **90**, 052105 (2014).
- [48] Y. S. Teo, H. Zhu, B.-G. Englert, J. Řeháček, and Z. Hradil, Quantum-State Reconstruction by Maximizing Likelihood and Entropy, *Phys. Rev. Lett.* **107**, 020404 (2011).
- [49] A. Rahmani, X. Zhu, M. Franz, and I. Affleck, Phase diagram of the interacting Majorana chain model, *Phys. Rev. B* **92**, 235123 (2015).
- [50] S.-S. Gong, W. Zhu, J.-X. Zhu, D. N. Sheng, and K. Yang, Global phase diagram and quantum spin liquids in a spin- $\frac{1}{2}$ triangular antiferromagnet, *Phys. Rev. B* **96**, 075116 (2017).
- [51] Y. Zhang, T. Grover, A. Turner, M. Oshikawa, and A. Vishwanath, Quasiparticle statistics and braiding from ground-state entanglement, *Phys. Rev. B* **85**, 235151 (2012).