Efficient learning of ground & thermal states within phases of matter

Emilio Onorati, Cambyse Rouzé, Daniel Stilck França, James D. Watson

Efficient learning of ground & thermal states within phases of matter

Emilio Onorati,[1,*] Cambyse Rouzé ,[1,†] Daniel Stilck França,[2] and James D. Watson[3]

[1] *Zentrum Mathematik, Technische Universität München, 85748 Garching, Germany*
[2] *Univ Lyon, ENS Lyon, UCBL, CNRS, Inria, LIP, F-69342, Lyon Cedex 07, France*[‡]
[3] *University of Maryland, College Park, QuICS 3353 Atlantic Building, MD 20742-2420, USA* [§]

We consider two related tasks: (a) estimating a parameterisation of an unknown Gibbs state and expectation values of Lipschitz observables on this state; and (b) learning the expectation values of local observables within a thermal or quantum phase of matter. In both cases, we wish to minimise the number of samples we use to learn these properties to a given precision.

For the first task, we develop new techniques to learn parameterisations of classes of systems, including quantum Gibbs states of non-commuting Hamiltonians under the condition of exponential decay of correlations and the approximate Markov property, thus improving on work by [RF21]. We show that it is possible to infer the expectation values of all extensive properties of the state from a number of copies that not only scales polylogarithmically with the system size, but polynomially in the observable's locality — an exponential improvement — hence partially answering conjectures stated in [RF21] and [AAKS21] in the positive. This class of properties includes expected values of quasi-local observables and entropic quantities of the state.

For the second task, we turn our tomography tools into efficient algorithms for learning observables in a phase of matter of a quantum system. By exploiting the locality of the Hamiltonian, we show that M local observables can be learned with probability $1-\delta$ and up to precision ε with access to only $N = \mathcal{O}\big(\log\big(\frac{M}{\delta}\big)e^{\mathrm{polylog}(\varepsilon^{-1})}\big)$ samples — an exponential improvement in the precision over the best previously known bounds [HKT+22]. Our results apply to both families of ground states of Hamiltonians displaying local topological quantum order, and thermal phases of matter displaying exponential decay of correlations. In addition, our sample complexity applies to the worse case setting whereas previous results only applied to the average case setting.

To prove our results, we develop new tools of independent interest, such as robust shadow tomography algorithms for ground and Gibbs states, Gibbs approximations of locally indistinguishable ground states, and generalisations of transportation cost inequalities for Gibbs states of non-commuting Hamiltonians.

* Emilio Onorati emilio.onorati@tum.de
† Cambyse Rouzé cambyse.rouze@tum.de
‡ Daniel Stilck França daniel.stilck_franca@ens-lyon.fr
§ James D. Watson jdwatson@umd.edu

I. INTRODUCTION

Tomography of quantum states is among the most important tasks in quantum information science. In quantum tomography, we have access to one or more copies of a quantum state and wish to understand the structure of the state. However, for a general quantum state, all tomographic methods inevitably require resources that scale exponentially in the size of the system [HHJ+17, OW16]. This is due to the curse of dimensionality: the number of parameters needed to fully describe a quantum system scales exponentially with the number of its constituents. Obtaining these parameters often necessitates the preparation and destructive measurement of exponentially many copies of the quantum system, as well as their storage in a classical memory. In particular, as the size of quantum devices continues to increase beyond what can be easily simulated classically, the community faces new challenges to characterise their output states in a robust and efficient manner.

Thankfully, only a few physically relevant observables are often needed to describe the physics of a system, e.g. its entanglement or energy. Recently, new methods of tomography have been proposed which precisely leverage this important simplification to develop efficient state learning algorithms. One highly relevant development in this direction is that of classical shadows [HKP20]. This new set of protocols allows for estimating physical observables of quantum spin systems that only depend on local properties from a number of measurements that scales logarithmically with the total number of qubits. However, the number of required measurements still faces an exponential growth with respect to the size of the observables that we want to estimate. Thus, using such protocols to learn the expectation values of physical observables that depend on more than a few qubits quickly becomes unfeasible.

Gibbs State Tomography. Some simplification can be achieved from the fact that physically relevant quantum states, such as ground and Gibbs states of a locally interacting spin system, are themselves often described by a number of parameters which scales only polynomially with the number of qubits. From this observation, another direction in the characterisation of large quantum systems that has received considerable attention is that of Hamiltonian learning and many-body tomography, where it was recently shown that it is possible to robustly characterise the interactions of a Gibbs state with a few samples [Ans, HKT21]. However, even for many-body states, recovery in terms of the trace distance requires a number of samples that scales polynomially in the number of qubits, in contrast to shadows for which the scaling is logarithmic.

These considerations naturally lead to the question of identifying settings where it is possible to combine the strengths of shadows and many-body tomography. In [RF21], some of the authors proposed a first solution by combining these with new insights from the emerging field of quantum optimal transport. They obtained a tomography algorithm that only requires a number of samples that scales logarithmically in the system's size and learns all quasi-local properties of a state. These properties are characterised by so-called "Lipschitz observables". However, that first step was confined to topologically trivial states such as high-temperature Gibbs states of commuting Hamiltonians or outputs of shallow circuits. Here, we significantly extend these results to all states exhibiting exponential decay of correlations and the approximate Markov property.

Learning Phases of Matter. Tomographical techniques by themselves are somewhat limited in that they tell us nothing about nearby related states – often states belong to a phase of matter in which the properties of the states vary smoothly and are in some sense "well behaved", and we wish to learn properties of this entire phase of matter. A recent line of research in this direction

that has gained significant attention from the quantum community is that of combining machine learning methods with the ability to sample complex quantum states from a phase of matter to efficiently characterise the entire phase [BWP+17, CM17]. A landmark result in this direction is [HKT+22]. There the authors showed how to use machine learning methods combined with classical shadows to learn local linear and nonlinear functions of states belonging to a gapped phase of matter with a number of samples that only grows logarithmically with the system's size. That is, given states from that phase drawn from a distribution and the corresponding parameters of the Hamiltonian, one can train a classical algorithm that would predict local properties of other points of the phase. However, there are some caveats to this scheme: (i) the scaling of the number of samples in terms of the precision is exponential, (ii) it does not immediately apply to phases of matter beyond gapped ground states, (iii) the results only come with guarantees on the errors in the prediction in expectation. That is, given another state sampled from the same distribution as the one used to train, only *on average* is the error made by the ML algorithm proven to be small.

In this work, we address all of these shortcomings. First, our result extends to thermal phases of matter which exhibit exponential decay of correlations, which includes all thermal systems away from criticality/poles in the partition function [HMS20, Section 5]. Our result also extends to gapped phases that satisfy local topological quantum order [MZ13, BHM10, NSY22]. Furthermore, the sample complexity of our algorithm is quasi-polynomial in the desired precision, which is an exponential improvement over previous work [HKT+22]. And, importantly, it comes with pointwise guarantees on the quality of the recovery, as opposed to average guarantees.

Interestingly, our results are easier to grasp through the lens of the concentration of measure phenomenon rather than machine learning: we show that local expectation values of quantum states are quite smooth under perturbations in the same class of states. And, as is showcased by the concentration of measure phenomenon, smooth functions on high-dimensional spaces do not show a lot of variability. Thus, it suffices to collect a few examples to be able to predict what happens in the whole space, while the price we pay for these stronger recovery guarantees is that our algorithm does not work for any distribution over states, but needs some form of anti-concentration which holds e.g. for the uniform distribution (see Appendix D for a technical discussion). In other words, our algorithm necessitates to "see" enough of the space to work and struggles if there are large, low-probability corners.

II. SUMMARY OF MAIN RESULTS

In this paper, we consider a quantum system defined over a D-dimensional finite regular lattice $\Lambda = [-L, L]^D$, where $n = (2L+1)^D$ denotes the total number of qubits constituting the system. We assume for simplicity that each site of the lattice hosts a qubit, so that the total system's Hilbert space is $\mathcal{H}_\Lambda := \bigotimes_{j\in\Lambda} \mathbb{C}^2$. All of the results presented here easily extend to qudits, but we will focus on qubits for simplicity.

Our focus in this work are nontrivial statements about what can be learned about many-body states of n qubits in the setting where we are only given $\Theta(\text{polylog}(n))$ copies. The common theme is that we will assume exponential decay of correlations for our class of states, but will show results in two different regimes. In Section II A we summarise our results on how to estimate *all quasi-local properties of a given state given identical copies of it.* This is the traditional setting of quantum tomography. In contrast, in Section II B we summarise our findings on how to learn *local properties of a class of states given samples from different states from that class.* This is the

setting of [HKT+22] where ground states of gapped quantum phases of matter were studied. Here we consider (a) thermal phases of matter with exponentially decaying correlations and (b) gapped ground states with local topological quantum order.

A. Optimal Tomography of Many-Body Quantum States

We first consider the task of obtaining a good approximation of expected values of extensive properties of a fixed unknown n-qubit state over Λ. The state is assumed to be a Gibbs state of an unknown local Hamiltonian $H(x) := \sum_{j\in\Lambda} h_j(x_j)$, $x = \{x_j\} \in [-1,1]^m$, defined through interactions $h_j(x_j)$, each depending on parameters $x_j \in [-1,1]^\ell$ for some fixed integer ℓ and supported on a ball A_j around site $j \in \Lambda$ of radius r_0. We also assume that the matrix-valued functions $x_j \mapsto h_j(x_j)$ as well as their derivatives are uniformly bounded: $\|h_j\|_\infty, \|\nabla h_j\|_\infty \leq h$. The corresponding Gibbs state at inverse temperature $\beta > 0$, and the ground state as $\beta \to \infty$ take the form

$$\sigma(\beta, x) := \frac{e^{-\beta H(x)}}{\operatorname{tr}\left[e^{-\beta H(x)}\right]} \qquad \text{and} \qquad \psi_g(x) := \lim_{\beta\to\infty} \sigma(\beta, x)\,. \tag{II.1}$$

In the case when $[h_j(x_j), h_{j'}(x_{j'})] = 0$ for all $j, j' \in \Lambda$, the Hamiltonian $H(x)$ and its associated Gibbs states $\sigma(\beta, x)$ are said to be commuting.

1. *Preliminaries on Lipschitz observables*

Extensive properties of a state are well-captured by the recently introduced class of Lipschitz observables [RD19, DPMTL21].

Definition II.1 (Lipschitz Observable [DPMTL21])**.** *An observable L on $\mathcal{H}_\Lambda$ is said to be* Lipschitz *if $\|L\|_{\text{Lip}} := \max_{i\in\Lambda} \min_{L_{i^c}} 2\|L - L_{i^c} \otimes I_i\|_\infty = \mathcal{O}(1)$, where i^c is the complement of the site i in Λ and the scaling is in terms of the number of qubits in the system.*

In words, $\|L\|_{\text{Lip}}$ quantifies the amount by which the expectation value of L changes for states that are equal when tracing out one site. By a simple triangle inequality together with [DPMTL21, Proposition 15], one can easily see that $\|L\|_\infty \leq n\|L\|_{\text{Lip}}$. Given the definition of the Lipschitz constant, we can also define the quantum Wasserstein distance of order 1 by duality [DPMTL21].

Definition II.2 (Wasserstein Distance [DPMTL21])**.** *The* Wasserstein distance *between two n qubit quantum states ρ_1, ρ_2 is defined as $W_1(\rho_0, \rho_1) := \sup_{\|L\|_{\text{Lip}}\leq 1} \operatorname{tr}\left[L(\rho_0 - \rho_1)\right] \leq n\|\rho - \sigma\|_1$.*

Having $W_1(\rho, \sigma) = \mathcal{O}(\varepsilon n)$ is sufficient to guarantee that the expectation value of ρ and σ on *extensive, quasi-local* observables is the same up to a *multiplicative* error εn. This justifies why we focus on learning states up to an error $\mathcal{O}(\varepsilon n)$ in Wasserstein distance instead of the usual trace distance bound of order $\mathcal{O}(\varepsilon)$: although a trace distance guarantee of order $\mathcal{O}(\varepsilon)$ gives the same error estimate, it requires exponentially more samples even for product states, as shown in [RF21, Appendix G]. In Appendix B, we argue that Lipschitz observables and the induced Wasserstein distance capture linear and nonlinear extensive properties of many-body quantum states.

2. Gibbs state tomography

In this section, we turn our attention to the problem of obtaining approximations of linear functionals of the form $f_L(\beta, x) := \mathrm{tr}[L\sigma(\beta, x)]$ for all Lipschitz observables L from the measurement and classical post-processing of as few copies of the associated unknown Gibbs state $\sigma(\beta, x)$ as possible. We will further require that the state satisfies the property of exponential decay of correlations: for any two observables X_A, resp. X_B, supported on region A, resp. B,

$$\mathrm{Cov}_{\sigma(\beta,x)}(X_A, X_B) \leq C \, \min\{|A|, |B|\} \, \|X_A\|_\infty \, \|X_B\|_\infty \, e^{-\nu \, \mathrm{dist}(A,B)} \,, \tag{II.2}$$

for some constants $C, \nu > 0$, where $\mathrm{dist}(A, B)$ denotes the distance between regions A and B, and where the covariance is defined by

$$\mathrm{Cov}_\sigma(X, Y) := \frac{1}{2} \, \mathrm{tr}\left[\sigma \left\{X - \mathrm{tr}[\sigma X], Y - \mathrm{tr}[\sigma Y]\right\}\right] . \tag{II.3}$$

Our first main result is a method to learn Gibbs states with few copies of the unknown state:

Theorem II.3 (Tomography algorithm for decaying Gibbs states (informal))**.** *For any unknown commuting Gibbs state $\sigma(\beta, x)$ satisfying Equation* (II.2)*, there exists an algorithm that provides the description of parameters x' such that the state $\sigma(\beta, x')$ approximates $\sigma(\beta, x)$ to precision $n\varepsilon$ in Wasserstein distance with probability $1-\delta$ with access to $N = \mathcal{O}\big(\log(\delta^{-1}) \, \mathrm{polylog}(n) \, \varepsilon^{-2}\big)$ samples of the state (see Appendix C 3 a). The result extends to non-commuting Hamiltonians whenever one of the following two assumptions is satisfied:*

(i) the high-temperature regime, $\beta < \beta_c$ (see Appendix C 3 b).

(ii) uniform clustering/Markov conditions (see Corollary C.12).

In case (ii), we find good approximation guarantees under the following slightly worst scaling in the precision ε: $N = \mathcal{O}(\varepsilon^{-4} \, \mathrm{polylog}(n\delta^{-1}))$.

The results for commuting Hamiltonians and in the high-temperature regime proceed directly from the following continuity bound on the Wasserstein distance between two Gibbs states, whose proof requires the notion of quantum belief propagation in the non-commuting case (see Corollary C.4): for any $x, y \in [-1, 1]^m$,

$$W_1(\sigma(\beta, x), \sigma(\beta, y)) = \, \|x - y\|_{\ell_1} \, \mathcal{O}(\mathrm{polylog}(n)) \, . \tag{II.4}$$

Furthermore, this inequality is tight up to a $\mathrm{polylog}(n)$ factor for $\beta = \Theta(1)$. Equation (II.4) reduces the problem of recovery in Wasserstein distance to that of recovering the parameters x up to an error $\varepsilon n / \mathrm{polylog}(n)$ in ℓ_1 distance. This is a variation of the Hamiltonian learning problem for Gibbs states [AAKS21, HKT21] which relies on lower bounding the ℓ_2 strong convexity constant for the log-partition function.

In [Ans], the authors give an algorithm estimating x with $e^{\mathcal{O}(\beta k^D)}\mathcal{O}(\log(\delta^{-1}n)\varepsilon^{-2})$ copies of $\sigma(\beta, x)$ up to ε in ℓ_∞ distance when $\sigma(\beta, x)$ belongs to a family of commuting, k-local Hamiltonians on a D-dimensional lattice. If we assume $m = \mathcal{O}(n)$, this translates to an algorithm with sample complexity $e^{\mathcal{O}(\beta k^D)}\mathcal{O}(\varepsilon^{-2}\mathrm{polylog}(\delta^{-1}n))$ to learn x up to εn in ℓ_1 distance. It should also be noted that the time complexity of the algorithm in [Ans] is $\mathcal{O}(ne^{\mathcal{O}(\beta k^D)}\varepsilon^{-2}\mathrm{polylog}(\delta^{-1}n))$. Thus,

any commuting model at constant temperature satisfying exponential decay of correlations can be efficiently learned with polylog(n) samples. We refer the reader to Appendix C 3 for more information and classes of commuting states that satisfy exponential decay of correlations. In the high-temperature regime, we rely on a result of [HKT21] where the authors give a computationally efficient algorithm to learn x up to error ε in ℓ_∞ norm from $\mathcal{O}(\varepsilon^{-2}\text{polylog}(\delta^{-1}n))$ samples. This again translates to a $\mathcal{O}(\varepsilon n)$ error in ℓ_1 norm thanks to (II.4).

Furthermore, in Appendix C 3 c we more directly extend the strategy of [AAKS21] by introducing the notion of a W_1 strong convexity constant for the log-partition function and showing that it scales linearly with the system size under (a) uniform clustering of correlations and (b) uniform Markov condition. This result also generalises the strategy of [RF21] which relied on the existence of a so-called transportation cost inequality previously shown to be satisfied for commuting models at high-temperature. For the larger class of states satisfying conditions (a) and (b), we are able to find x' s.t. $W_1(\sigma(\beta,x),\sigma(\beta,x')) = \mathcal{O}(\varepsilon n)$ with $\mathcal{O}(\varepsilon^{-4}\text{polylog}(\delta^{-1}n))$ samples. Note that the uniform Markov condition is expected to hold for a large class of models that goes beyond high-temperature Gibbs states [KB19, KKBa20].

3. *Beyond linear functionals*

So far, we considered properties of the quantum system which could be related to local linear functionals of the unknown state. In [HKP20, HKT+22], the authors propose a simple trick in order to learn non-linear functionals of many-body quantum systems, e.g. their entropy over a small subregion. However, such methods require a number of samples scaling exponentially with the size of the subregion, and thus very quickly become inefficient as the size of the region increases. Here instead, we make use of the continuity of the entropy functional with respect to the Wasserstein distance, mentioned in Equation (B.6), together with the following Wasserstein continuity bound in order to estimate the entropic quantities of Gibbs states over regions of arbitrary size (see Corollary C.6): assuming Equation (II.2), for any region S of the lattice and any two $x, y \in [-1,1]^m$

$$W_1(\text{tr}_{S^c}(\sigma(\beta,x)), \text{tr}_{S^c}(\sigma(\beta,y))) \leq \|x|_{\mathcal{S}(r_S)} - y|_{\mathcal{S}(r_S)}\|_{\ell_1} \text{ polylog}(|S(r_S)|)\,, \tag{II.5}$$

where $r_S = \max\left\{r_0,\, 2\xi \log\left(2|S|C_1\|x|_{\mathcal{S}(r_0)} - y|_{\mathcal{S}(r_0)}\|_{\ell_1}^{-1}\right)\right\}$ with r_0 being the smallest integer such that $x|_{\mathcal{S}(r_0)} \neq y|_{\mathcal{S}(r_0)}$, $\mathcal{S}(r_S) := \{x_j |\ \text{supp}(h_j(x_j)) \cap S(r_S) \neq \emptyset\}$, $S(r_S) := \{i \in \Lambda : \text{dist}(i,S) \leq r_S\}$, and $C_1, \xi > 0$ are constants introduced in Lemma C.5.

Let us recall a few definitions: denoting by $\rho_R := \text{tr}_{R^c}(\rho)$ the marginal of a state $\rho \in \mathcal{D}(\mathcal{H}_\Lambda)$ on a region $R \subset \Lambda$, and given separated regions $A, B, C \subset \Lambda$ of the lattice: $S(A)_\rho := -\text{tr}[\rho_A \log \rho_A]$ is the von Neumann entropy of ρ on A, $S(A|B)_\rho := S(AB)_\rho - S(B)_\rho$ is the conditional entropy on region A conditioned on region B, $I(A:B)_\rho := S(A)_\rho + S(B)_\rho - S(AB)_\rho$ is the mutual information between regions A and B, and $I(A:B|C)_\rho := S(AC)_\rho + S(BC)_\rho - S(C)_\rho - S(ABC)_\rho$ is the conditional mutual information between regions A and B conditioned on region C. The following corollary is a direct consequence of Equation (B.6) together with Equation (II.5):

Corollary II.4. *Assume the decay of correlations holds uniformly, as specified in Equation* (II.2)*, for all* $\{\sigma(\beta,x)\}_{x\in[-1,1]^m}$*,* $m = \mathcal{O}(n)$*. Then, in the notations of the above paragraph, for any two Gibbs states* $\sigma(\beta,x)$ *and* $\sigma(\beta,y)$*,* $x, y \in [-1,1]^m$*, and any region* $A \subset \Lambda$*:*

$$|S(A)_{\sigma(\beta,x)} - S(A)_{\sigma(\beta,y)}| = \|x|_{\mathcal{S}(r_S)} - y|_{\mathcal{S}(r_S)}\|_{\ell_1} \mathcal{O}(\text{polylog}(|S(r_S)|))\,,$$

for $S \equiv A$. The same conclusion holds for $|S(A|B)_{\sigma(\beta,x)} - S(A|B)_{\sigma(\beta,y)}|$ ($S \equiv AB$), $|I(A:B)_{\sigma(\beta,x)} - I(A:B)_{\sigma(\beta,y)}|$ ($S \equiv AB$), and $|I(A:B|C)_{\sigma(\beta,x)} - I(A:B|C)_{\sigma(\beta,y)}|$ ($S \equiv ABC$).

Thus, given an an estimate y of x satisfying $\|x-y\|_{\ell_\infty} = \mathcal{O}(\varepsilon/\mathrm{polylog}(n))$, we can also approximate entropic quantities of the Gibbs state to a multiplicative error. More generally, entropic continuity bounds can be directly used together with Theorem II.3(ii) in order to estimate entropic properties of Gibbs states satisfying both uniform clustering of correlations and the approximate Markov condition (see Appendix C 3 c for details).

B. Learning Expectation Values of Parametrised Families of Many-Body Quantum Systems

Next, we turn our attention to the task of learning Gibbs or ground states of a parameterised Hamiltonian $H(x)$ known to the learner and sampled according to the uniform distribution U over $x \in [-1,1]^m$. More general distributions can also be dealt with under a condition of anti-concentration, see Appendix D. Here we restrict our results to local observables of the form $O = \sum_{i=1}^M O_i$ where $S_i := \mathrm{supp}(O_i)$ is contained in a ball of diameter independent of the system size. The setup in this section is similar to [HKT+22]. The idea is that we have access to some samples of a state chosen from different values of the parameterised Hamiltonian, and we want to use these to learn observables everywhere in the parameter space with high precision. We then want to know: what is the minimum number of samples drawn from this distribution which allows us to accurately predict expectation values of local observables for all choices of parameters?

1. *Learning Expectation Values in Thermal Phases of Matter*

The learner is given samples $\{(x_i, \sigma(\beta, x_i))\}_{i=1}^N$, where the parameters $x_i \sim U$, and their task is to learn $f_O(x) := \mathrm{tr}[\sigma(\beta,x)O]$ for an arbitrary value of $x \in [-1,1]^m$ and an arbitrary local observable O. We assume that everywhere in the parameter space $x \in [-1,1]^m$ the Gibbs states are in the same phase of exponentially decaying correlations. Then we have:

Theorem II.5 (Learning algorithm for quantum Gibbs states)**.** *With the conditions of the previous paragraph, given a set of N samples $\{x_i, \tilde{\sigma}(\beta,x_i)\}_{i=1}^N$, where $\tilde{\sigma}(\beta,x_i)$ can be stored efficiently classically, and $N = O\big(\log\big(\frac{M}{\delta}\big)\log\big(\frac{n}{\delta}\big)e^{\mathrm{polylog}(\varepsilon^{-1})}\big)$, there exists an algorithm that, on input $x \in [-1,1]^m$ and a local observable $O = \sum_{i=1}^M O_i$, produces an estimator $\hat{f}_O$ such that, with probability $(1-\delta)$,*

$$\sup_{x\in[-1,1]^m} |f_O(x) - \hat{f}_O(x)| \leq \varepsilon \sum_{i=1}^M \|O_i\|_\infty .$$

Moreover, the samples $\tilde{\sigma}(\beta,x_i)$ are efficiently generated from measurements of the Gibbs states $\{\sigma(\beta,x_i)\}_{i=1}^N$ followed by classical post-processing.

Our estimator $\hat{f}_O$ is constructed as follows: during a training stage, we pick N points $Y_1,\ldots,Y_N \sim U$ and estimate the reduced Gibbs states over large enough enlargements $S_i\partial$ of the supports $\mathcal{S}_i := \{x_j |\ \mathrm{supp}(h_j(x_j)) \cap S_i\partial \neq \emptyset\} \cap [x-\varepsilon, x+\varepsilon]^m$ of the observables O_i. Due to the anti-concentration property of the uniform distribution, the probability that a small region $\mathcal{S}_i\partial$ in parameter space contains t variables $Y_{i_1},\ldots,Y_{i_t}$ becomes large for $N \approx \log(M)$. We then

run the classical shadow tomography protocol on those states in order to construct efficiently describable and computable product matrices $\widetilde{\sigma}(\beta, Y_1), \ldots, \widetilde{\sigma}(\beta, Y_N)$. Then for any region S_i, we select the shadows $\widetilde{\sigma}(\beta, Y_{i_1}), \ldots \widetilde{\sigma}(\beta, Y_{i_t})$ whose local parameters are close to that of the target state and construct the empirical average $\widetilde{\sigma}_{S_i}(x) := \frac{1}{t}\sum_{j=1}^{t} \operatorname{tr}_{S_i^c}\left[\widetilde{\sigma}(\beta, Y_{i_j})\right]$. Using belief propagation methods (see Proposition D.2), it is possible to show that exponential decay of correlations ensures that the estimator is a good approximation to local observables. Thus such operators can be well approximated using the reduced state $\operatorname{tr}_{S_i^c} \sigma(\beta, x)$ for $t \approx \log(n)$. The estimator $\hat{f}_O$ is then naturally chosen as $\hat{f}_O(x) := \sum_{i=1}^{M} \operatorname{tr}[O_i\, \widetilde{\sigma}_{S_i}(x)]$. A key part of the proof is demonstrating that exponential decay of correlations implies that $f_O(x)$ does not change too much as x varies.

2. *Learning ground states under local indistinguishability*

We now move our attention to the problem of learning ground states. Again, the learner is given samples $\{x_i, \psi_g(x_i)\}_{i=1}^{N}$, $x_i \sim U$, and their task is to learn $f_{O,g}(x) := \operatorname{tr}[\psi_g(x) O]$. In fact, the previous argument for Gibbs states can be extended to the present setting as long as the condition of exponentially decaying correlations in the Gibbs state is replaced by the following condition of *local topological quantum order* (LTQO) [MZ13, BHM10, NSY22]: A quantum system satisfies LTQO if for any two regions $A \subset B \subset \Lambda$ and all $x \in [-1,1]^m$,

$$\left\| \operatorname{tr}_{A^c}(\psi_g(x, B) - \psi_g(x, \Lambda)) \right\|_1 \leq C_T\, |A|\, e^{-\frac{\operatorname{dist}(A, B^c)}{\xi_0}} \tag{II.6}$$

for some constants $C_T, \xi_0 > 0$, and where, given a region $R \subset \Lambda$ we denote by $\psi_g(x, R)$ the ground state corresponding to the Hamiltonian $H_R(x) = \sum_{j \in R} h_j(x_j)$. In words, LTQO states that observables localised away from the boundary of the volume B cannot distinguish between different ground states. Many systems of practical interest are known to satisfy Equation (II.6), including frustration-free spin chains with a unique translation-invariant matrix product ground state [AKLT88] and quantum double models, which include Kitaev's toric code [Kit06, Kit03, CDH+20]. For more details on LTQO, we refer to [NSY22] and the references therein.

Theorem II.6 (Learning algorithm for quantum ground states)**.** *With the conditions of the previous paragraph, given a set of N samples $\{x_i, \tilde{\psi}(x_i)\}_{i=1}^{N}$, where $\tilde{\psi}(x_i)$ can be stored efficiently classically, and $N = O\big(\log\left(\frac{M}{\delta}\right) \log\left(\frac{n}{\delta}\right) e^{\operatorname{polylog}(\varepsilon^{-1})}\big)$, there exists an algorithm that, on input $x \in [-1,1]^m$ and a local observable $O = \sum_{i=1}^{M} O_i$, produces an estimator $\hat{f}_O$ such that, with probability $(1-\delta)$,*

$$\sup_{x \in [-1,1]^m} |f_O(x) - \hat{f}_O(x)| \leq \varepsilon \sum_{i=1}^{M} \|O_i\|_\infty .$$

Moreover, the samples $\tilde{\psi}(x_i)$ are efficiently generated from measurements of the ground states $\{\psi_g(x_i)\}_{i=1}^{N}$ followed by classical post-processing.

To prove this statement, we reduce it to the problem of learning Gibbs states of the previous section. The LTQO condition permits approximating the expectation of the local observable O_i in the state $\psi_g(x)$ by the one in the state $\psi_g(x, S_i\partial)$. The latter is approximated by the local Gibbs state $\sigma(\beta, x, S_i\partial) \propto e^{-\beta H_{S_i\partial}(x)}$ for large but constant β (see Lemma E.5). By a continuity argument, these states are approached by $\sigma(\beta, Y_{i_t}, S_i\partial)$, which in turn are close to $\psi_g(Y_{i_t})$. This

chain of approximation steps together with a robust version of the shadow tomography protocol for ground states, stated in Proposition E.7, allows us to conclude. We expect that the assumption of LTQO is not the only assumption that can be made to achieve similar scaling. Indeed, we expect that a lower bound on the spectral gap in the parameterized region would achieve similar results.

III. COMPARISON TO PREVIOUS WORK

A. Classical literature

The problem of Hamiltonian learning for classical models has attracted a lot of attention in the last years by the computer science community [Bre15, PSBR20, LVMC18, ZKKW20] which traditionally refers to it as Ising model — or Markov field — learning. The question of what can be inferred from very few samples was also asked classically [DDDK20]. Our work sheds further light on this question and is of interest even when restricting to classicaapproximatingbservables. Indeed, to the best of our knowledge, the statements of Corollary C.4 and Corollary C.6 are new even for classical Gibbs distributions. Previous work by the authors of [RF21] already established similar learning results for measures satisfying a so-called transportation cost inequality (TC) [BG99, Tal96], although the present condition of exponential decay of correlations is more standard.

It should be noted that if a Gibbs measure satisfies TC, then any Lipschitz function of a random variable distributed according to it satisfies a Gaussian concentration bound [Led01]. This can easily be seen to imply that we can estimate the expectation value of M Lipschitz functions up to an error ε with probability of success δ from $\mathcal{O}(\varepsilon^{-2}\log(M\delta^{-1}))$ samples by taking the empirical average. At first sight this might look comparable with the sample complexity we obtain with our learning algorithm. However, this only holds for *one* basis, whereas our result holds for any basis. Furthermore, if the number of Lipschitz observables satisfies $M = e^{c\,\Omega(n)}$, then the number of samples required to obtain a good estimate through the empirical average becomes polynomial. On the other hand, given that $W_1(\sigma(\beta,x),\sigma(\beta,x')) \leq \varepsilon n$, we can evaluate as many Lipschitz observables as we wish from $\sigma(\beta,x')$ without requiring any further samples. Thus, even for observables in a fixed basis our result has advantages.

B. Previous work on many-body quantum state tomography

As mentioned before, one striking advantage of our Gibbs tomography algorithm when estimating expectation values of local observables compared to state-agnostic methods like classical shadows is the exponential speedup in the size of the support of the observable. In fact, our method gives good guarantees on the larger class of Lipschitz observables, which includes non-local observables. This advantage is even more visible when it comes to estimating entropic quantities: whereas the polynomial approximation proposed in [HKP20] works universally for any n-qubit state, it only gives good approximation guarantees for reduced states on very few qubits. Here instead, we avoid this issue by leveraging the Wasserstein continuity bounds offered in [DPMTL21].

Our framework also differs from the one of Hamiltonian learning algorithms tackled in [Ans, AAKS21, HKT21]: in these papers, the authors were interested in estimating the parameter x of a given Hamiltonian $H(x)$ given access to copies of the state $\sigma(\beta,x)$, in ℓ_2 or ℓ_∞. Here instead, we argue that a good recovery in W_1 distance is implied by the weaker condition of recovery in ℓ_1. Clearly, one can leverage these previous results to further control our ℓ_1 bound,

as we argue in Section II A 2. It should be noted however that our bound only requires that the Gibbs state $\sigma(\beta, x)$ satisfies an exponential decay of correlations, whereas these learning algorithms provide very efficient ℓ_∞ or ℓ_2 recovery either for (i) commuting Hamiltonians or (ii) in the high-temperature regime. It remains an important question whether the condition of exponential decay of correlations is enough to get good ℓ_1 recovery. Furthermore, in Appendix C 3 c we show that under the additional assumptions of uniform Markovianity and clustering of correlations, it is possible to learn in W_1 through the maximum entropy method, without resorting directly to learning the parameters x.

C. Previous work on learning observables in phases of matter

In [HKT+22], the authors found a machine learning algorithm which, for any smoothly parameterised family of local Hamiltonians $\{H(x)\}_{x\in[-1,1]^m}$ in a finite spatial dimension with a constant spectral gap, can be trained to predict expected values of sums of local observables in the associated ground state $\psi_g(x)$. More precisely, given a local observable $O = \sum_{i=1}^M O_i$ with $\mathrm{supp}(O_i) = \mathcal{O}(1)$, they construct an estimator $\hat{f}_O(x)$ of the expectation value of the observable such that

$$\mathbb{E}_{x\sim U([-1,1]^m)}\left[\left|\operatorname{tr}[O\psi_g(x)] - \hat{f}_O(x)\right|^2\right] \leq \varepsilon^2\Big(\sum_{i=1}^M \|O_i\|_\infty\Big)^2, \tag{III.1}$$

as long as the training size (i.e. the number of sampled points within the phase) is $N = \Big(\sum_{i=1}^M \|O_i\|_\infty\Big)^2 m^{\mathcal{O}(1/\varepsilon^2)}$.

In Theorem II.6, we improve this result for ground states in three ways, up to further imposing the LTQO condition: first, we can assume that the parameters x are distributed according to a much larger class of distributions than the uniform distribution. This extension does not carry so easily in the proof of [HKT+22] which uses Fourier analysis techniques involving integration over the Lebesgue measure to derive Equation (III.1). Second, theirs is a result in expectation, that is in $\|.\|_{L^2}$, whereas our bound in Theorem II.6 works in the worst-case setting associated to the stronger $\|.\|_\infty$-norm topology. Third and most importantly, the dependence of the number of training data points scales exponentially in the precision parameter ε in Equation (III.1), whereas ours scales only quasi-polynomially.

Finally, we extend the learning result beyond ground states to finite temperature phases of matter with exponential decay of correlations. This not only includes all high-temperature phases of matter (regardless of the Hamiltonian), but also low-temperature phases with the relevant correlation functions [DCGR19]. This is a particularly relevant result since zero temperature is never achieved in practice, so in reality we are always working with low-temperature thermal states.

We also recognise independent, concurrent work by [LTL+23]. Here the authors consider the same setup of gapped ground states as [HKT+22] and also improved over Equation (III.1) to achieve the same sample complexity as Theorem II.6. However, their result is not directly comparable to ours. We emphasise [LTL+23] consider gapped, ground state phases, whereas our work focuses on thermal phases and ground states with LTQO. We also note they remove all conditions on the prior distribution over the samples x, whereas we still need to assume a type of mild anti-concentration over the local marginals. However, their result is still stated as an $\|.\|_{L^2}$-bound due to the use of machine learning machinery, whereas our more straightforward Gibbs approximation tools allow us to get stronger bounds in $\|.\|_\infty$. Conceptually speaking, our methods for approximating local

expectation values requires no knowledge of machine learning techniques. Our work also shows that it is possible to go beyond gapped quantum phases and learn thermal phases with exponentially decaying correlations, as well as ground states with LTQO.

IV. DISCUSSION AND CONCLUSIONS

We have contributed to the tasks of tomography and learnability of quantum many-body states by combining previous techniques with approaches not considered so far in this field to obtain novel and powerful features.

Tomography. First, we extended the results of [RF21] on the efficient tomography of high-temperature commuting Gibbs states to Gibbs states with exponentially decaying correlations. This result permits to significantly enlarge the class of states for which we know how to learn all quasi-local properties with a number of samples that scales polylogarithmically with the system's size. In particular, our results now also hold for classes of Gibbs states of non-commuting Hamiltonians. As we require exponentially fewer samples to learn in the Wasserstein metric when compared with the usual trace distance and still recover essentially all physically relevant quantities associated to the states, we hope that our results motivate the community to consider various tomography problems in the Wasserstein instead of trace distance.

As we achieved this result by reducing the problem of learning the states to learning the parameters of the Hamiltonian in ℓ_1, we hope our work further motivates the study of the Hamiltonian learning problem in ℓ_1-norm with polylog samples. 1D Gibbs states are a natural place to start, but obtaining Hamiltonian learning algorithms just departing from exponential decay of correlations would provide us with a complete picture. In Appendix C 3 c we also partially decoupled the Hamiltonian learning problem from the W_1 learning one by resorting to the uniform Markov condition. Thus, it would be important to establish the latter for a larger number of systems.

It would be interesting to investigate the sharpness of our bounds, and to understand if exponential decay of correlations is really necessary. One way of settling this question would be to prove polynomial lower bounds for learning in Wasserstein distance for states at critical temperatures.

Learning Phases of Matter. Second, we improved the results of [HKT$^+$22] for learning a class of states in several directions, including the scaling in precision, the classes of states it applies to and the form of the recovery guarantee. In particular, the results now apply to Gibbs states, which are the states of matter commonly encountered experimentally. Interestingly, we did not need to resort to machine learning techniques to achieve an exponentially better scaling in precision by making arguably mild assumptions on the distributions the states are drawn from.

Although the results proved here push the state-of-the-art of learning quantum states, we believe that our methods, for instance the novel continuity bounds for various local properties of quantum many-body states, will find applications in other areas of quantum information.

Beyond the thermal phases and LTQO ground states studied here, it would be interesting to find other families of states which can be efficiently learned, and indeed if more restrictive assumptions on the parameterization of Hamiltonians can result in more efficient learning. One interesting open problem that goes beyond the present paper's scope is finding families of states satisfying LTQO without belonging to a common gapped phase of matter. If such a family existed, it would clarify the differences between our framework and that of [HKT$^+$22]. Finally, we realise that although the results proved here are for lattice systems, they almost certainly generalise to non-lattice configurations of particles.

V. ACKNOWLEDGMENTS

The authors gratefully recognise useful discussions with Hsin-Yuan (Robert) Huang and Haonan Zhang. We thank Laura Lewis, Viet T. Tran, Sebastian Lehner, Richard Kueng, Hsin-Yuan (Robert) Huang, and John Preskill for sharing a preliminary of the manuscript [LTL+23], which was discussed in Section III C.

EO is supported by the Munich Quantum Valley and the Bavarian state government, with funds from the Hightech Agenda Bayern Plus. CR acknowledges financial support from a Junior Researcher START Fellowship from the DFG cluster of excellence 2111 (Munich Center for Quantum Science and Technology), from the ANR project QTraj (ANR-20-CE40-0024-01) of the French National Research Agency (ANR), as well as from the Humboldt Foundation. DSF is supported by France 2030 under the French National Research Agency award number "ANR-22-PNCQ-0002". JDW acknowledges support from the United States Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Accelerated Research in Quantum Computing program, and also NSF QLCI grant OMA-2120757.

[AAKS21] Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, and Mehdi Soleimanifar. Sample-efficient learning of interacting quantum systems. *Nature Physics*, 17(8):931–935, 2021.

[ABF23] Guillaume Aubrun, Emily Beatty, and Daniel Stilck França. *In preparation*, 2023.

[AKLT88] Ian Affleck, Tom Kennedy, Elliott H Lieb, and Hal Tasaki. Valence bond ground states in isotropic quantum antiferromagnets. In *Condensed matter physics and exactly soluble models*, pages 253–304. Springer, 1988.

[Ans] Anurag Anshu. Efficient learning of commuting Hamiltonians on lattices. *Electronic notes*.

[Ara69] Huzihiro Araki. Gibbs states of a one dimensional quantum lattice. *Communications in Mathematical Physics*, 14(2):120–157, 1969.

[BCPH22] Andreas Bluhm, Ángela Capel, and Antonio Pérez-Hernández. Exponential decay of mutual information for Gibbs states of local Hamiltonians. *Quantum*, 6:650, February 2022.

[BFT17] Mario Berta, Omar Fawzi, and Marco Tomamichel. On variational expressions for quantum relative entropies. *Letters in Mathematical Physics*, 107(12):2239–2265, 2017.

[BG99] S.G Bobkov and F Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *Journal of Functional Analysis*, 163(1):1–28, April 1999.

[BH11] Sergey Bravyi and Matthew B Hastings. A short proof of stability of topological order under local perturbations. *Communications in mathematical physics*, 307(3):609–627, 2011.

[BHH16] Fernando G. S. L. Brandão, Aram W. Harrow, and Michał Horodecki. Local random quantum circuits are approximate polynomial-designs. *Communications in Mathematical Physics*, 346(2):397–434, August 2016.

[BHM10] Sergey Bravyi, Matthew B Hastings, and Spyridon Michalakis. Topological quantum order: stability under local perturbations. *Journal of mathematical physics*, 51(9):093512, 2010.

[BHV06] S. Bravyi, M. B. Hastings, and F. Verstraete. Lieb-Robinson bounds and the generation of correlations and topological quantum order. *Phys. Rev. Lett.*, 97:050401, Jul 2006.

[BK18] Fernando G. Brandão and Michael J. Kastoryano. Finite correlation length implies efficient preparation of quantum thermal states. *Communications in Mathematical Physics*, 365(1):1–16, 2018.

[BKL+17] Fernando GSL Brandão, Amir Kalev, Tongyang Li, Cedric Yen-Yu Lin, Krysta M Svore, and Xiaodi Wu. Exponential quantum speed-ups for semidefinite programming with applications to quantum learning. *arXiv preprint arXiv:1710.02581*, 2017.

[Bre15] Guy Bresler. Efficiently learning Ising models on arbitrary graphs. In *Proceedings of the forty-seventh annual ACM symposium on Theory of Computing*. ACM, June 2015.

[BWP+17] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, September 2017.

[CDH+20] Shawn X Cui, Dawei Ding, Xizhi Han, Geoffrey Penington, Daniel Ranard, Brandon C Rayhaun, and Zhou Shangnan. Kitaev's quantum double model as an error correcting code. *Quantum*, 4:331, 2020.

[CGLW13] Xie Chen, Zheng-Cheng Gu, Zheng-Xin Liu, and Xiao-Gang Wen. Symmetry protected topological orders and the group cohomology of their symmetry group. *Physical Review B*, 87(15):155114, 2013.

[CLMPG15] Toby S Cubitt, Angelo Lucia, Spyridon Michalakis, and David Perez-Garcia. Stability of local quantum dissipative systems. *Communications in Mathematical Physics*, 337(3):1275–1315, 2015.

[CM17] Juan Carrasquilla and Roger G. Melko. Machine learning phases of matter. *Nature Physics*, 13(5):431–434, February 2017.

[CPGW15] Toby S. Cubitt, David Perez-Garcia, and Michael M. Wolf. Undecidability of the spectral gap. *Nature*, 528(7581):207–211, 2015.

[DCGR19] Hugo Duminil-Copin, Subhajit Goswami, and Aran Raoufi. Exponential decay of truncated correlations for the Ising model in any dimension for all but the critical temperature. *Communications in Mathematical Physics*, 374(2):891–921, November 2019.

[DDDK20] Yuval Dagan, Constantinos Daskalakis, Nishanth Dikkala, and Anthimos Vardis Kandiros. Learning Ising models from one or multiple samples, 2020.

[Don86] Matthew J Donald. On the relative entropy. *Communications in Mathematical Physics*, 105(1):13–34, 1986.

[DPMTL21] Giacomo De Palma, Milad Marvian, Dario Trevisan, and Seth Lloyd. The quantum Wasserstein distance of order 1. *IEEE Transactions on Information Theory*, 67(10):6627–6643, 2021.

[DS87] Roland L Dobrushin and Senya B Shlosman. Completely analytical interactions: constructive description. *Journal of Statistical Physics*, 46(5):983–1014, 1987.

[GCC22] Andi Gu, Lukasz Cincio, and Patrick J. Coles. Practical black box hamiltonian learning, 2022.

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

[Has10] Matthew B Hastings. Locality in quantum systems. *Quantum Theory from Small to Large Scales*, 95:171–212, 2010.

[HHJ+17] Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory*, pages 1–1, 2017.

[HK06] Matthew B. Hastings and Tohru Koma. Spectral gap and exponential decay of correlations. *Communications in Mathematical Physics*, 265(3):781–804, April 2006.

[HKP20] H.-Y. Huang, R. Kueng, and J. Preskill. Predicting many properties of a quantum system from very few measurements. *Nat. Phys.*, 16(10):1050–1057, 2020.

[HKT21] Jeongwan Haah, Robin Kothari, and Ewin Tang. Optimal learning of quantum Hamiltonians from high-temperature gibbs states. *arXiv preprint arXiv:2108.04842*, 2021.

[HKT+22] Hsin-Yuan Huang, Richard Kueng, Giacomo Torlai, Victor V. Albert, and John Preskill. Provably efficient machine learning for quantum many-body problems. *Science*, 377(6613), 2022.

[HMS20] Aram W Harrow, Saeed Mehraban, and Mehdi Soleimanifar. Classical algorithms, correlation decay, and complex zeros of partition functions of quantum many-body systems. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 378–386, 2020.

[HP91] Fumio Hiai and Dénes Petz. The proper formula for relative entropy and its asymptotics in quantum probability. *Communications in mathematical physics*, 143(1):99–114, 1991.

[Jay57a] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.

[Jay57b] Edwin T Jaynes. Information theory and statistical mechanics II. *Physical review*, 108(2):171, 1957.

[Jay82] Edwin T Jaynes. On the rationale of maximum-entropy methods. *Proceedings of the IEEE*, 70(9):939–952, 1982.

[JRS+18] Marius Junge, Renato Renner, David Sutter, Mark M Wilde, and Andreas Winter. Universal recovery maps and approximate sufficiency of quantum relative entropy. In *Annales Henri Poincaré*,

volume 19, pages 2955–2978. Springer, 2018.

[KB19] Kohtaro Kato and Fernando GSL Brandao. Quantum approximate Markov chains are thermal. *Communications in Mathematical Physics*, 370(1):117–149, 2019.

[KGE14] Martin Kliesch, Christian Gogolin, and Jens Eisert. Lieb-Robinson bounds and the simulation of time-evolution of local observables in lattice systems. In *Many-Electron Approaches in Physics, Chemistry and Mathematics*, pages 301–318. Springer, 2014.

[KGK+14] Martin Kliesch, Christian Gogolin, MJ Kastoryano, A Riera, and J Eisert. Locality of temperature. *Physical review x*, 4(3):031019, 2014.

[Kim17] Isaac H Kim. Markovian matrix product density operators: Efficient computation of global entropy. *arXiv preprint arXiv:1709.07828*, 2017.

[Kit03] A Yu Kitaev. Fault-tolerant quantum computation by anyons. *Annals of Physics*, 303(1):2–30, 2003.

[Kit06] Alexei Kitaev. Anyons in an exactly solved model and beyond. *Annals of Physics*, 321(1):2–111, 2006.

[KKBa20] Tomotaka Kuwahara, Kohtaro Kato, and Fernando G. S. L. Brandão. Clustering of conditional mutual information for quantum gibbs states above a threshold temperature. *Phys. Rev. Lett.*, 124:220601, Jun 2020.

[Led01] Michel Ledoux. *The concentration of measure phenomenon*. Number 89. American Mathematical Soc., 2001.

[LR72] Elliott H Lieb and Derek W Robinson. The finite group velocity of quantum spin systems. In *Statistical mechanics*, pages 425–431. Springer, 1972.

[LSS19] Jingcheng Liu, Alistair Sinclair, and Piyush Srivastava. Fisher zeros and correlation decay in the Ising model. *Journal of Mathematical Physics*, 60(10):103304, 2019.

[LTL+23] Laura Lewis, Viet T. Tran, Sebastian Lehner, Richard Kueng, Hsin-Yuan Huang, and John Preskill. Improved machine learning algorithm for predicting ground state properties. *arXiv*, 2023.

[LVMC18] Andrey Y. Lokhov, Marc Vuffray, Sidhant Misra, and Michael Chertkov. Optimal structure and parameter learning of Ising models. *Science Advances*, 4(3), March 2018.

[MZ13] Spyridon Michalakis and Justyna P Zwolak. Stability of frustration-free Hamiltonians. *Communications in Mathematical Physics*, 322(2):277–302, 2013.

[NSY22] Bruno Nachtergaele, Robert Sims, and Amanda Young. Quasi-locality bounds for quantum lattice systems. part ii. perturbations of frustration-free spin models with gapped ground states. In *Annales Henri Poincaré*, volume 23, pages 393–511. Springer, 2022.

[OT19] Yoshiko Ogata and Hal Tasaki. Lieb–Schultz–Mattis type theorems for quantum spin chains without continuous symmetry. *Communications in Mathematical Physics*, 372(3):951–962, 2019.

[OW16] Ryan O'Donnell and John Wright. Efficient quantum tomography. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '16, page 899–912, New York, NY, USA, 2016. Association for Computing Machinery.

[Pet86] Dénes Petz. Sufficient subalgebras and the relative entropy of states of a von neumann algebra. *Communications in mathematical physics*, 105(1):123–131, 1986.

[Pou10] David Poulin. Lieb-Robinson bound and locality for general Markovian quantum dynamics. *Physical review letters*, 104(19):190401, 2010.

[PR22] Giacomo De Palma and Cambyse Rouzé. Quantum concentration inequalities. *Annales Henri Poincaré*, 23(9):3391–3429, April 2022.

[PSBR20] Adarsh Prasad, Vishwak Srinivasan, Sivaraman Balakrishnan, and Pradeep Ravikumar. On learning Ising models under huber's contamination model. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS'20, Red Hook, NY, USA, 2020. Curran Associates Inc.

[QW20] Yang Qiu and Zhenghan Wang. Ground subspaces of topological phases of matter as error correcting codes. *Annals of Physics*, 422:168318, 2020.

[RD19] Cambyse Rouzé and Nilanjana Datta. Concentration of quantum states from quantum functional and transportation cost inequalities. *Journal of Mathematical Physics*, 60(1):012202, January 2019.

[RF21] Cambyse Rouzé and Daniel Stilck França. Learning quantum many-body systems from a few copies.

arXiv preprint arXiv:2107.03333, 2021.

[SBT16] David Sutter, Mario Berta, and Marco Tomamichel. Multivariate trace inequalities. *Communications in Mathematical Physics*, 352(1):37–58, October 2016.

[Sid09] Vladas Sidoravicius. *New Trends in Mathematical Physics: Selected Contributions of the XVth International Congress on Mathematical Physics*. Springer Science & Business Media, 2009.

[SK22] Pavel Svetlichnyy and T. A. B. Kennedy. Decay of quantum conditional mutual information for purely generated finitely correlated states. *Journal of Mathematical Physics*, 63(7):072201, 2022.

[SM16] Brian Swingle and John McGreevy. Mixed s-sourcery: Building many-body states using bubbles of nothing. *Phys. Rev. B*, 94:155125, Oct 2016.

[T$^+$15] Joel A Tropp et al. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015.

[Tal96] M. Talagrand. Transportation cost for gaussian and other product measures. *Geometric and Functional Analysis*, 6(3):587–600, May 1996.

[Tas18] Hal Tasaki. Topological phase transition and Z_2 index for $s = 1$ quantum spin chains. *Physical Review Letters*, 121(14):140604, 2018.

[ZKKW20] Huanyu Zhang, Gautam Kamath, Janardhan Kulkarni, and Steven Wu. Privately learning Markov random fields. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 11129–11140. PMLR, 13–18 Jul 2020.

SUPPLEMENTAL MATERIAL

Appendix A: Preliminaries

Given a finite dimensional Hilbert space $\mathcal{H}$, we denote by $\mathcal{B}(\mathcal{H})$ the algebra of bounded operators on $\mathcal{H}$, whereas $\mathcal{B}_{\text{sa}}(\mathcal{H})$ denotes the subspace of self-adjoint operators. We denote by $\mathcal{D}(\mathcal{H})$ the set of positive operators on $\mathcal{H}$ of unit trace, and by $\mathcal{D}_+(\mathcal{H})$ the subset of positive, full-rank operators on $\mathcal{H}$. Schatten norms are denoted by $\|.\|_p$ for $p \geq 1$. The identity matrix in $\mathcal{B}(\mathcal{H})$ is denoted by I. Given a bipartite system AB, the normalised partial trace over a subsystem A is written τ_A, i.e. $\tau_A := 2^{-|A|} \operatorname{tr}_A$.

In this work, we consider a family of local qubit interactions $\{h_j(x_j)\}_{x_j \in [-1,1]^\ell}$, $j = 1, \ldots, n$ over the D-dimensional lattice $\Lambda = [-L, L]^D$, for some fixed integer ℓ, where $n = (2L+1)^D$ denotes the total number of qubits constituting the system. For each j and all $x_j \in [-1, 1]^\ell$, $h_j(x_j)$ is supported on a ball A_j around site $j \in \Lambda$ of radius r_0. We also assume that the matrix-valued functions $x_j \mapsto h_j(x_j)$ as well as their derivatives are uniformly bounded: $\|h_j\|_\infty, \|\nabla_x h_j(x)\|_\infty \leq h$. For sake of simplicity, we assume that the interactions are linear functions of their parameters, that is $h_j(x_j) = x_j V_j$ for some fixed operator V_j. However this assumption is not necessary in any of our proofs, as commented in Appendix F. Concatenating the vectors x_j into $x = (x_1, \ldots, x_n) = (x'_1, \ldots, x'_m)$, $m = n\ell$, the local interactions induce the following family of Hamiltonians $\{H(x)\}_{x \in [-1,1]^m}$, with:

$$H(x) = \sum_{j=1}^{m} h_j(x_j)\,. \tag{A.1}$$

More generally, given a region $B \subset \Lambda$ of the lattice, we denote by $H_B(x) := \sum_{j | A_j \subset B} h_j(x)$ the Hamiltonian restricted to B. We denote by $x|_{S(r)}$ the concatenation of vectors x_j corresponding to interactions h_j supported on regions intersecting $S(r) := \{l \in \Lambda |\ \operatorname{dist}(l, S) \leq r\}$.

For much of the following, we will be concerned with Gibbs states, defined as

$$\sigma(\beta, x) := \frac{e^{-\beta H(x)}}{\operatorname{tr}[e^{-\beta H(x)}]}.$$

In particular, we will be interested in systems satisfying the following type of correlation decay:

Condition A.1 (Exponential Decay of Correlations)**.** *For a state σ and any operator X_A, resp. X_B, supported on region A, resp. B, we say the state satisfies* exponential decay of correlations *if*

$$\operatorname{Cov}_\sigma(X_A, X_B) \leq C\, \min\{|A|, |B|\}\, \|X_A\|_\infty\, \|X_B\|_\infty\, e^{-\nu \operatorname{dist}(A,B)}\,, \tag{A.2}$$

for any choice of X_A,X_B, and for some parameters $C, \nu > 0$ which we assume independent of x and of the lattice size n, and where

$$\operatorname{Cov}_\sigma(A, B) := \frac{1}{2} \operatorname{tr}\Big[\sigma \left\{A - \operatorname{tr}[\sigma A], B - \operatorname{tr}[\sigma B]\right\}\Big]\,.$$

Condition A.1 is satisfied by many classes of Gibbs states, including high-temperature Gibbs states [HMS20, KKBa20] and 1D Gibbs states at any constant temperature [HMS20, BCPH22]. It is also known to hold for ground states of gapped Hamiltonians [HK06]. In fact, the class of Gibbs

states for which Condition A.1 holds is larger than that for which polylog algorithms to learn the parameters of the Hamiltonian are known. In Appendix C 3 we will discuss several examples for which it is known how to learn the parameters efficiently. In Appendix C 3 c we will also consider the case when we have the additional assumption of uniform Markovianity to show that then it is possible to bypass having to learn the parameters.

Appendix B: Lipschitz observables

In this appendix, we argue that Lipschitz observables and the induced Wasserstein distance capture most observables of physical interest, such as local and quasi-local observables, as well as quasi-local polynomials of the state and entropic quantities of subsystems. They can even capture global properties, including some of physical interest like global entropies. These classes of examples justify the claim that Lipschitz observables and the Wasserstein distance capture well both linear and nonlinear extensive properties of quantum states.

Let us illustrate our previous claims. An important class of Lipschitz observables are those of the form

$$\sum_{i=1}^{M} O_i, \quad M = \mathcal{O}(n), \quad \|O_i\| = \mathcal{O}(1), \quad \max_{1\leq j\leq n} |\{i : \mathrm{supp}(O_i) \cap \{j\} \neq \emptyset\}| = \mathcal{O}(1). \tag{B.1}$$

Observables like those defined in Equation (B.1) include local observables w.r.t. to a regular lattice. However, it is also not difficult to see that the expectation values of such observables are characterised by the marginals of the states on a few qubits. But Lipschitz observables capture more than strictly local properties. Indeed, as shown in [RF21], the time evolution of local observables like those in Equation (B.1) by a shallow quantum circuit or a short continuous-time evolution satisfying a Lieb-Robinson bound are Lipschitz. These include evolutions by Hamiltonians with algebraically decaying interactions, which will map strictly local Hamiltonians to quasi-local observables. In fact, recent results [ABF23] show that Lipschitz observables can distinguish two random quantum states almost optimally. As such states are locally indistinguishable [BHH16, Corollary 15], this fact shows that Lipschitz observables capture much more than just quasi-local properties of quantum states.

Although so far we only discussed how to use the Wasserstein distance to control linear functionals of the state, the fact that the Wasserstein distance behaves well under tensor products means that it is also easy to control the error for non-linear functions. Indeed, in [DPMTL21, Propostion 4], the authors show that the Wasserstein distance is additive under tensor products. i.e. for all states ρ, σ and integer k we have

$$W_1(\rho^{\otimes k}, \sigma^{\otimes k}) = kW_1(\rho, \sigma). \tag{B.2}$$

We can then combine this additivity with the standard trick that a polynomial of degree k on a quantum state can be expressed as the expectation value of a certain observable O on $\rho^{\otimes k}$. In particular, if this polynomial is an average over polynomials in reduced density matrices of constant size, it is not difficult to see that the corresponding observable on $\rho^{\otimes k}$ will be Lipschitz as well.

Let us exemplify this in the case of the average purity of a state. For a subset $A \subset [n]$ of the qubits of size l, let $\mathbb{F}_A \in \left(\mathbb{C}^2\right)^{\otimes 2l}$ be the flip operator acting on two copies of those qubits:

$$\mathbb{F}_A(|\psi\rangle \otimes |\varphi\rangle) = |\varphi\rangle \otimes |\psi\rangle\,. \tag{B.3}$$

It can be shown in a few lines that $\mathrm{tr}\left[\mathbb{F}_A \rho^{\otimes 2}\right] = \mathrm{tr}\left[\rho_A^2\right]$. Furthermore, observables of the form

$$O = \sum_{i=1}^{M} \mathbb{F}_{A_i}, \quad M = \mathcal{O}(n), \quad \max_{1\leq j\leq n} |\{i : A_i \cap \{j\} \neq \emptyset\}| = \mathcal{O}(1). \tag{B.4}$$

satisfy $\|O\|_{\mathrm{Lip}} = \mathcal{O}(1)$. Then

$$\sum_{i=1}^{M} \mathrm{tr}\left[\rho_{A_i}^2 - \sigma_{A_i}^2\right] = \mathrm{tr}\left[O(\rho^{\otimes 2} - \sigma^{\otimes 2})\right] \leq \|O\|_{\mathrm{Lip}} W_1(\rho^{\otimes 2}, \sigma^{\otimes 2}) = 2\|O\|_{\mathrm{Lip}} W_1(\rho, \sigma). \tag{B.5}$$

By a direct generalisation of the above, we see that $W_1(\rho, \sigma) = \mathcal{O}(\varepsilon n/k)$ is sufficient to ensure that degree k polynomials of the states are approximated to a multiplicative error. As we will see later in Section II A 3, this polynomial trick can be used to ensure that averages of various subsystem entropies, mutual informations and conditional mutual informations are well-approximated given a Wasserstein bound.

Once again it should be emphasised that a Wasserstein bound can be used to control global properties, even non-linear ones. A good example of that is the entropy of a quantum state. In [DPMTL21, Theorem 1], the authors show the continuity bound:

$$|S(\rho) - S(\sigma)| \leq g(W_1(\rho, \sigma)) + W_1(\rho, \sigma) \log(4n), \tag{B.6}$$

where $g(t) = (t+1)\log(t+1) - t\log(t)$. In this case, it turns out that a Wasserstein distance of $W_1(\rho, \sigma) = \mathcal{O}(\varepsilon n/\log(n))$ suffices to obtain a multiplicative error for the entropy. Finally, it is also worth mentioning observables that are not Lipschitz. Simple examples include linear combinations of high-weight Paulis.

Appendix C: Gibbs states tomography

In this section, our main goal is to devise an efficient tomography algorithm for Gibbs states $\sigma(\beta, x)$. In particular, we wish to learn the parameters x to high precision. We prove the following lemma:

Theorem C.1 (Tomography algorithm for decaying Gibbs states). *Let $H(x) = \sum_i h_i(x_i)$ be a Hamiltonian such that each $h_i(x_i)$, $x_i \in [-1, 1]^\ell$, is not more than k-local, for $k = O(1)$, and all terms commute. For some unknown x, let $\sigma(\beta, x)$ be its associated Gibbs state satisfying exponential decay of correlations as per Condition A.1. Then there exists an algorithm that provides the description of parameters x' such that the state $\sigma(\beta, x')$ satisfies:*

$$W_1(\sigma(\beta, x), \sigma(\beta, x')) \leq \varepsilon n \tag{C.1}$$

with probability greater than $1 - \delta$, such that the algorithm requires access to no more than $N = \mathcal{O}\big(\log(\delta^{-1})\,\mathrm{polylog}(n)\,\varepsilon^{-2}\big)$ samples of the state (see Appendix C 3 a).

The result extends to the case where $\{h_i(x_i)\}_i$ do not commute whenever one of the following two assumptions is satisfied:

(i) the high-temperature regime, $\beta < \beta_c$ (see Appendix C 3 b).

(ii) uniform clustering/Markov conditions (see Corollary C.12).

In case (ii), we find good approximation guarantees under the following slightly worst scaling in the precision ε: $N = \mathcal{O}(\varepsilon^{-4}\,\mathrm{polylog}(n\delta^{-1}))$.

Proof Outline. The full proof is laid out in sections C 1, C 2 and C 3.

The fundamental part of the result uses the continuity estimate of the Wasserstein distance between two Gibbs states that is of interest on its own. In Corollary C.4 we will show that under exponential decay of correlations we have:

$$W_1(\sigma(\beta,x),\sigma(\beta,y)) \leq \|x-y\|_{\ell_1}\ \mathrm{polylog}(n)\,. \tag{C.2}$$

The significance of the bound in Equation (C.2) is that it reduces the problem of obtaining a good estimate of $\sigma(\beta,x)$ in W_1 to estimating the parameters x in ℓ_1 distance. This is a variation of the Hamiltonian learning problem [AAKS21, GCC22, HKT21], and we can then directly import results from the literature for our tomography algorithm. □

As we argued before in Section II A 1, the recovery guarantee in Equation (C.1) suffices to ensure that $\sigma(\beta,x')$ mirrors all the quasi-local properties of $\sigma(\beta,x)$. Furthermore, the polylog complexity in system size is exponentially better than what is required to obtain a recovery guarantee in trace distance [RF21, Appendix G], even for product states.

1. Quantum belief propagation

We start by recalling a well-known tool in the analysis of quantum Gibbs states known as quantum belief propagation [Has07, Kim17, KB19]. We assume a parameterisation of the Hamiltonian as $H(x) = \sum_{j=1}^m x_j V_j$ for appropriate operators V_j (we will generalise this to other parameterisation later) and for some observable L we define the function $f_L(\beta,x) = \mathrm{tr}\,[\sigma(\beta,x)L]$. The belief propagation method then states that we have that for any $k\in[m]$,

$$\partial_{x'_k} f_L(\beta,x) = -\frac{\beta}{2}\ \mathrm{tr}\left[L\{\Phi_{H(x)}(\partial_{x'_k}H(x)),\sigma(\beta,x)\}\right] + \beta\,\mathrm{tr}(\partial_{x'_k}H(x)\sigma(\beta,x))\ \mathrm{tr}(L\sigma(\beta,x))\,.$$

where the quantum belief propagation operator $\Phi_{H(x)}$ is defined as

$$\Phi_{H(x)}(V) := \int_{-\infty}^{\infty} dt\,\kappa_\beta(t)\,e^{-iH(x)t}Ve^{iH(x)t}\,,$$

for some smooth, fast-decaying probability density function $\kappa_\beta(t) := \frac{1}{2\pi}\int \widetilde{\kappa}_\beta(\omega)e^{i\omega t}d\omega$ of Fourier transform

$$\widetilde{\kappa}_\beta(\omega) := \frac{\tanh(\beta\omega/2)}{\beta\omega/2}\,.$$

The function κ_β was in fact computed in [AAKS21, Appendix B]: for $t\in\mathbb{R}\backslash\{0\}$:

$$\kappa_\beta(t) := \frac{2}{\pi\beta}\,\log\frac{e^{\pi|t|/\beta}+1}{e^{\pi|t|/\beta}-1} \leq \frac{4}{\pi\beta}\,\frac{1}{e^{\pi|t|/\beta}-1} \tag{C.3}$$

Rewriting the above derivative, and using the notations $\langle O\rangle_{\beta,x} \equiv \mathrm{tr}(\sigma(\beta,x)O)$ for the expected value of an observable O in the Gibbs state $\sigma(\beta,x)$, we have that

$$\partial_{x'_k} f_L(\beta,x) = -\frac{\beta}{2}\,\langle\{L,\,\widetilde{H}_k(x) - \langle\widetilde{H}_k(x)\rangle_{\beta,x}\}\rangle_{\beta,x} \tag{C.4}$$

where $\widetilde{H}_k(x) := \Phi_{H(x)}(\partial_{x'_k}H(x))$. We define the covariance between two observables A and B in the state σ as

$$\mathrm{Cov}_\sigma(A,B) := \frac{1}{2}\,\mathrm{tr}\left[\sigma\left\{A-\mathrm{tr}[\sigma A], B-\mathrm{tr}[\sigma B]\right\}\right].$$

Therefore

$$\partial_{x'_k} f_L(\beta,x) = -\beta\,\mathrm{Cov}_{\sigma(\beta,x)}\left(L,\widetilde{H}_k(x)\right). \tag{C.5}$$

In what follows, we will need to approximate $\widetilde{H}_k(x)$ by observables supported on bounded regions. For this, we make use of Lieb-Robinson bounds for Hamiltonians of finite-range interactions [LR72, Pou10, KGE14, BHV06, Has10, Sid09, CLMPG15]. Here we choose a version proven in [CLMPG15, Lemma 5.5]: for any observable O_A supported on a region A of the lattice, and any $B \supset A$, we denote by α_t, resp. by α_t^B, the unitary evolution generated by $H(x)$, resp. by $H_B(x)$, up to time t, i.e.

$$\alpha_t(O) := e^{-iH(x)t}Oe^{iH(x)t}\,, \qquad \alpha_t^B(O) := e^{-iH_B(x)t}Oe^{iH_B(x)t}\,.$$

which then satisfy

$$\|\alpha_t(O_A) - \alpha_t^B(O_A)\|_\infty \le c\,|A|\,\|O_A\|_\infty\, e^{vt-\mu\,\mathrm{dist}(A,B^c)}\,, \tag{C.6}$$

for some parameters $c, v, \mu > 0$ which depend on the interactions h_j but can be chosen independent of n and x.

Lemma C.2. *For any region $A \subset B \subset \Lambda$ and operator O_A supported in A and all x,*

$$\|\Phi_{H(x)}(O_A) - \Phi_{H_B(x)}(O_A)\|_\infty \le c'\,|A|\,\|O_A\|_\infty\, e^{-\mu'\,\mathrm{dist}(A,B^c)}$$

for some parameters c' and μ' depending on $H(x)$ and β but independent of n.

Proof. We make use of the exponential decay of κ_β provided in Equation (C.3) together with the Lieb-Robinson bound Equation (C.6):

$$\begin{aligned}
\|\Phi_{H(x)}(O_A) - \Phi_{H_B(x)}(O_A)\|_\infty &\le \int_{-\infty}^{\infty} |\kappa_\beta(t)|\,\|\alpha_t(O_A) - \alpha_t^B(O_A)\|_\infty\,dt \\
&\le c\,|A|\,\|O_A\|_\infty e^{-\mu\,\mathrm{dist}(A,B^c)} \int_{-\delta}^{\delta} |\kappa_\beta(t)|\; e^{vt}\,dt \\
&\quad + 2\,\|O_A\|_\infty \int_{[-\delta,\delta]^c} |\kappa_\beta(t)|\,dt\,.
\end{aligned}$$

For the first integral above, we use that $|\kappa_\beta(t)| \propto \log(1/t)$ for t small. More precisely,

$$\int_{-\delta}^{\delta} |\kappa_\beta(t)|\,e^{vt}\,dt \le \frac{4e^{v\delta}}{\pi\beta}\int_0^\delta \log\left(\frac{e^{\pi t/\beta}+1}{t\pi/\beta}\right)dt \le \frac{4e^{(v+\pi/\beta)\delta}}{\pi^2}$$

For the other integral, we use the exponential decay of κ_β:

$$\int_{[-\delta,\delta]^c} |\kappa_\beta(t)|\, dt \leq \frac{8}{\pi\beta}\int_\delta^\infty \frac{1}{e^{\pi t/\beta}-1}\, dt \leq \frac{8}{\pi\beta}\int_\delta^\infty e^{-\frac{\pi t}{2\beta}} dt = \frac{16}{\pi^2}\, e^{-\frac{\pi\delta}{2\beta}}\,,$$

where the second inequality holds for $\delta \geq \frac{2\beta}{\pi}\,\mathrm{sh}^{-1}\left(\frac{1}{2}\right) \equiv \delta_1$. Choosing $\delta := \delta_1 + \mu\,\mathrm{dist}(A, B^c)/(2\big(v + \pi/\beta\big))$, we get

$$\|\Phi_{H(x)}(O_A) - \Phi_{H_B(x)}(O_A)\|_\infty \leq c'\,|A|\|O_A\|_\infty\, e^{-\mu' d(A,B^c)}\,,$$

for some constant $c' \equiv c'(\beta, v)$, where $\mu' = \mu \min\left\{\frac{1}{2},\, \frac{\pi}{4\beta(v+\pi/\beta)}\right\}$. □

2. Continuity estimate for W_1 distance on Gibbs states

In this subsection, we will prove Equation (C.2). First, we use the bound derived in Lemma C.2 together with the assumption that $\sigma(\beta, x)$ has exponential decay of correlations in order to control the derivatives $\partial_{x'_k} f_L$:

Lemma C.3. *Assume that $\sigma(\beta, x)$ satisfies the condition of decay of correlations, Condition A.1. Then for any $k \in [m]$,*

$$|\partial_{x'_k} f_L(\beta, x)| \leq \|L\|_{\mathrm{Lip}}\ \mathrm{polylog}(n)\,, \tag{C.7}$$

for some polynomial of $\log(n)$ of degree D with coefficients depending on $\beta, r_0, D, h, c', \nu, \mu'$ and C.

Proof. Denoting by j_k the index of the interaction h_{j_k} which depends on variable x'_k, we have that, given $\Phi_{H(x)}(\partial_{x'_k} h_j) = \delta_{j,j_k}\Phi_{H(x)}(\partial_{x'_k} h_{j_k})$, and denoting $\widetilde{h}_k = \Phi_{H(x)}(\partial_{x'_k} h_{j_k})$, from Equation (C.5) we have:

$$|\partial_{x'_k} f_L(\beta, x)| = \beta\ \mathrm{Cov}_{\sigma(\beta,x)}(L, \widetilde{H}_k(x)) = \beta\ \mathrm{Cov}_{\sigma(\beta,x)}(L,\, \widetilde{h}_k)\,.$$

Next, given a region $B_k \supset A_{j_k}$, define the observable

$$O_{B_k} := \Phi_{H_{B_k}(x)}(\partial_{x'_k} h_{j_k}) - \langle \Phi_{H_{B_k}(x)}(\partial_{x'_k} h_{j_k})\rangle_{\beta,x}\,. \tag{C.8}$$

Then by Lemma C.2 we have that

$$\begin{aligned}
\mathrm{Cov}_{\sigma(\beta,x)}(L,\, \widetilde{h}_k(x)) &= \mathrm{Cov}_{\sigma(\beta,x)}(L,\, \widetilde{h}_k(x) - O_{B_k}) + \mathrm{Cov}_{\sigma(\beta,x)}(L,\, O_{B_k}) \\
&\leq 2\|L\|_\infty\, \|\Phi_{H(x)}(\partial_{x'_k} h_{j_k}) - \Phi_{H_{B_k}(x)}(\partial_{x'_k} h_{j_k})\|_\infty + \mathrm{Cov}_{\sigma(\beta,x)}(L,\, O_{B_k}) \\
&\leq 2nc'(2r_0)^D\, h\, \|L\|_{\mathrm{Lip}}\, e^{-\mu'\,\mathrm{dist}(A_{j_k}, B_k^c)} + \mathrm{Cov}_{\sigma(\beta,x)}(L, O_{B_k})\,.
\end{aligned}$$

Next, we estimate the last covariance above. Denoting $B_k(r) := \{i \in \Lambda :\, \mathrm{dist}(i, B_k) \leq r\}$, we get

$$\begin{aligned}
\mathrm{Cov}_{\sigma(\beta,x)}(L, O_{B_k}) &= \mathrm{Cov}_{\sigma(\beta,x)}(L - \tau_{B_k(r)}(L), O_{B_k}) + \mathrm{Cov}_{\sigma(\beta,x)}(\tau_{B_k(r)}(L), O_{B_k}) \\
&\leq 2h\|L - \tau_{B_k(r)}(L)\|_\infty + 2C|B_k|\, h\|L\|_\infty\, e^{-\nu r} \\
&\leq 2h|B_k(r)|\, \|L\|_{\mathrm{Lip}} + 2C|B_k|\, h\, n\|L\|_{\mathrm{Lip}}\, e^{-\nu r}\,,
\end{aligned}$$

where the second line above follows from the condition of decay of correlations Condition A.1. Choosing $B_k = A_{j_k}(\lfloor \log(n)/\mu' \rfloor)$, so that $\mathrm{dist}(A_{j_k}, B_k^c) = \lfloor \log(n)/\mu' \rfloor$, and $r = \lfloor \log(n)/\nu \rfloor$, we have shown that, given $1/\nu' := 1/\mu' + 1/\nu$,

$$|\partial_{x'_k} f_L(\beta, x)| \leq 2\beta\, h\, \|L\|_{\mathrm{Lip}} \Big(c'(2r_0)^D\, h \, + (2(r_0 + \log(n)/\nu'))^D (1 + C) \Big)$$

The result follows. □

With the bound of Lemma C.3, we show that for Gibbs states belonging to a phase with exponentially decaying correlations, the difference of expected values of Lipschitz observables in two such states is controlled by the ℓ_1-norm of their associated parameters.

Corollary C.4. *With the conditions of Lemma C.3, for any $x, y \in [-1, 1]^m$,*

$$W_1(\sigma(\beta, x), \sigma(\beta, y)) \leq \, \|x - y\|_{\ell_1} \, \mathrm{polylog}(n)\,. \tag{C.9}$$

Furthermore, this inequality is tight up to a $\mathrm{polylog}(n)$ *factor for $\beta = \Theta(1)$.*

Proof. To get the upper bound Equation (C.9), it suffices to interpolate between the two states as follows: for any Lipschitz observable L, and a path $x(s) = (1 - s)x + sy$,

$$|\,\mathrm{tr}\,[L(\sigma(\beta, x) - \sigma(\beta, y))]\,| \leq \sum_{k=1}^{m} |x'_k - y'_k| \int_0^1 |\partial_k f_L(\beta, x)|\, ds\,.$$

The result follows from using Equation (C.7) above, and using the resulting inequality in the definition of Wasserstein distance, definition II.2.

To see that the inequality is tight up to the $\mathrm{polylog}(n)$ factor, consider the family of Hamiltonians $H(x) = \sum_i x_i Z_i$, which gives rise to diagonal, product Gibbs states that clearly satisfy exponential decay of correlations. We then have:

$$W_1(\sigma(\beta, x), \sigma(\beta, y)) \geq \frac{1}{2}\,\mathrm{tr}\left[\sum_i Z_i(\sigma(\beta, x) - \sigma(\beta, y))\right], \tag{C.10}$$

as $\sum_i Z_i$ has Lipschitz constant 2. A simple computation shows that:

$$\frac{1}{2}\,\mathrm{tr}\left[\sum_i Z_i(\sigma(\beta, x) - \sigma(\beta, y))\right] = \frac{1}{2}\sum_i \left(\frac{e^{-\beta x_i}}{e^{-\beta x_i} + e^{+\beta x_i}} - \frac{e^{-\beta y_i}}{e^{-\beta y_i} + e^{+\beta y_i}}\right). \tag{C.11}$$

We will assume without loss of generality that $x_i < y_i$ (as otherwise we can consider the observable with $-Z_i$ instead). Under this condition, the summands are all positive and thus:

$$\frac{1}{2}\,\mathrm{tr}\left[\sum_i Z_i\sigma(\beta, x) - \sigma(\beta, y))\right] = \frac{1}{2}\sum_i \left|\frac{e^{-\beta x_i}}{e^{-\beta x_i} + e^{+\beta x_i}} - \frac{e^{-\beta y_i}}{e^{-\beta y_i} + e^{+\beta y_i}}\right|. \tag{C.12}$$

Yet another simple computation shows that the derivative of the function $y \mapsto \frac{e^{-\beta y}}{e^{-\beta y} + e^{+\beta y}}$ is given by

$$-\frac{\beta}{2}\,\mathrm{sech}(\beta y). \tag{C.13}$$

Let c_β denote the minimum of the function in Equation (C.13) for a fixed $\beta = \Theta(1)$ over $y \in [-1, 1]$. Then, by the mean value theorem:

$$\frac{1}{2}\sum_i \left| \frac{e^{-\beta x_i}}{e^{-\beta x_i} + e^{+\beta x_i}} - \frac{e^{-\beta y_i}}{e^{-\beta y_i} + e^{+\beta y_i}} \right| \geq \frac{c_\beta}{2} \sum_i |x_i - y_i| \,, \tag{C.14}$$

from which we conclude that:

$$W_1(\sigma(\beta, x), \sigma(\beta, y)) \geq \frac{c_\beta}{2} \|x - y\|_{\ell_1}. \tag{C.15}$$

□

We next prove that when given a local observable O supported on a ball $S \subset \Lambda$ of diameter at most k_0 around site i of the lattice, to study its behaviour as $H(x)$ varies for Gibbs states, it is sufficient to only consider the components of x which parameterise local terms which are geometrically close to the observable O (up to some small error).

Before we prove this, we remember that we denote by $x|_{\mathcal{S}(r)}$ the concatenation of vectors x_j corresponding to interactions h_j supported on regions intersecting $S(r) := \{i \in \Lambda |\ \mathrm{dist}(i, S) \leq r\}$.

Lemma C.5 (Gibbs local indistinguishability)**.** *Assuming the exponential decay of correlations in Condition A.1, then for any observable O supported on region S, any $r \in \mathbb{N}$, denoting $f_O(x) := \mathrm{tr}[O\,\sigma(\beta, x)]$ and identify $x|_{\mathcal{S}(r)}$ with the vector $(x|_{\mathcal{S}(r)}, 0_{\mathcal{S}(r)^c}) \in [-1, 1]^m$, then the following bound holds:*

$$\sup_{x \in [-1,1]^m} |f_O(x) - f_O(x|_{\mathcal{S}(r)})| \leq C_1\, e^{-\frac{r}{2\xi}}\, \|O\|_\infty \,,$$

for $\mathcal{O}(1)$ constants $C_1, \xi > 0$ independent of n. In other words:

$$\sup_{x \in [-1,1]^m} \| \mathrm{tr}_{S^c}(\sigma(\beta, x) - \sigma(\beta, x|_{\mathcal{S}(r)}))\|_1 \leq C_1\, e^{-\frac{r}{2\xi}} \,. \tag{C.16}$$

Proof. We identify $x|_{\mathcal{S}(r)}$ with the vector $(x|_{\mathcal{S}(r)}, 0_{\mathcal{S}(r)^c}) \in [-1, 1]^m$. Given the path $x(s) = (1 - s)x + sx|_{\mathcal{S}(r)}$ with components $\{x_l'(s)\}_{l=1}^m$, we get

$$\begin{aligned} |f_O(x) - f_O(x|_{\mathcal{S}(r)})| &\leq \sum_{l \in \mathcal{S}(r)^c} |x_l'(0)| \int_0^1 \left| \partial_l \,\mathrm{tr}\left[O\sigma(\beta, x(s))\right] \right| ds \\ &= \beta \sum_{l \in \mathcal{S}(r)^c} |x_l'(0)| \int_0^1 \left| \,\mathrm{Cov}_{\sigma(\beta, x(s))}\left(O, \widetilde{H}_l(x(s))\right) \right| ds \,, \end{aligned} \tag{C.17}$$

for $\widetilde{H}_l(x) := \Phi_{H(x)}(\partial_l H(x))$, where the second line comes from Equation (C.5). Next, we call $j_l \in \Lambda$ the unique site such that x_l' is a coordinate of x_{j_l}, and denote A_{j_l} be the support of h_{j_l}. Now, the above covariance is small if r is large enough, since $\widetilde{H}_j(x(s))$ can be well approximated by an observable on S^c. Indeed,

$$\partial_l H(x) = \partial_l h_{j_\ell} \,,$$

where j_l denotes the index of interaction h_{j_l} which depends on variable x'_l. Therefore, whenever $A_{j_l} \cap S = \emptyset$, we proceed similarly to Lemma C.3: given a region $B_l \supset A_{j_l}$ such that $B_l \cap S = \emptyset$, denoting the observable

$$O_{B_l} := \Phi_{H_{B_l}(x)}(\partial_{x'_l} h_{j_l}) - \langle \Phi_{H_{B_l}(x)}(\partial_{x'_l} h_{j_l}) \rangle_{\beta,x} \,,$$

we have by Lemma C.2 as well as the assumption that the state $\sigma(\beta, x)$ has exponential decay of correlations we have the following (refer to Figure 1 for a diagram of the regions):

$$\begin{aligned}
\mathrm{Cov}_{\sigma(\beta,x)}(O,\, \widetilde{H}_l(x)) &= \mathrm{Cov}_{\sigma(\beta,x)}(O,\, \widetilde{H}_l(x) - O_{B_l}) + \mathrm{Cov}_{\sigma(\beta,x)}(O,\, O_{B_l}) \\
&\leq 2\|O\|_\infty \, \|\Phi_{H(x)}(\partial_{x'_l} h_{j_l}) - \Phi_{H_{B_l}(x)}(\partial_{x'_l} h_{j_l})\|_\infty + \mathrm{Cov}_{\sigma(\beta,x)}(O,\, O_{B_l}) \\
&\leq 2\|O\|_\infty c' |A_{j_l}| \, h \, e^{-\mu' \operatorname{dist}(A_{j_l}, B_l^c)} + 2C|S| \, \|O\|_\infty \, h \, e^{-\nu \operatorname{dist}(S, B_l)} \\
&\leq 2(C + c') \, \|O\|_\infty \, (2r_0 + k_0)^D \, h \left(e^{-\mu' \operatorname{dist}(A_{j_l}, B_l^c)} + e^{-\nu \operatorname{dist}(S, B_l)} \right)
\end{aligned}$$

By construction, for $r > 2r_0$, the condition that $A_{j_l} \cap S = \emptyset$ is met, and therefore the bound holds. We recall that $i \in \Lambda$ is defined as the center of S. Since $\operatorname{dist}(i, j_l) = k_0/2 + \operatorname{dist}(S, B_l) + \operatorname{dist}(A_{j_l}, B_l) + r_0$, we can choose B_l so that $\operatorname{dist}(S, B_l), \operatorname{dist}(A_{j_l}, B_l) \geq \operatorname{dist}(i, j_l)/2 - k_0/4 - r_0/2 - 1$. Therefore,

$$\mathrm{Cov}_{\sigma(\beta,x)}(O,\, \widetilde{H}_l(x)) \leq 4(C + c')C'' \|O\|_\infty (2r_0 + k_0)^D \, h e^{-\operatorname{dist}(i, j_l)/\xi}$$

where $1/\xi = \min\{\mu', \nu\}$ and $C'' := e^{\max\{\mu', \nu\}(k_0/4 + r_0/2 + 1)}$. Therefore

$$|f_O(x) - f_O(x|_{\mathcal{S}_i(r)})| \leq 4\beta(C + c') \, h \, (2r_0 + k_0)^D \, \|O\|_\infty \sum_{l \in \mathcal{S}(r)^c} e^{-\operatorname{dist}(i, j_l)/\xi} \,.$$

Upon shifting the center of the lattice at site i, we get

$$\begin{aligned}
|f_O(x) - f_O(x|_{\mathcal{S}(r)})| &\leq 4\beta(C + c')C'' \, h \, (2r_0 + k_0)^D \, \|O\|_\infty \sum_{|l| \geq r + k_0/2} e^{-|l|/\xi} \\
&= 4\beta(C + c')C'' \, h \, (2r_0 + k_0)^D \, \|O\|_\infty \sum_{a > r + k_0/2} \binom{a + D - 1}{D - 1} e^{-a/\xi} \\
&\leq 4\beta(C + c')C'' \, h \, (2r_0 + k_0)^D \, D^{D-1} \|O\|_\infty \sum_{a > r + k_0/2} a^{D-1} \, e^{-a/\xi} \\
&\leq 4\beta(C + c')C'' \, h \, (2r_0 + k_0)^D (D - 1)! (2\xi)^{D-1} \, D^{D-1} \|O\|_\infty \sum_{a > r + k_0/2} e^{-\frac{a}{2\xi}} \\
&\leq 4\beta(C + c')C'' \, h \, (2r_0 + k_0)^D (D - 1)! (2\xi)^{D-1} \, D^{D-1} \|O\|_\infty \, \frac{e^{-\frac{r + k_0/2 + 1}{2\xi}}}{1 - e^{-\frac{1}{2\xi}}} \\
&\equiv C_1 \, e^{-\frac{r}{2\xi}} \, \|O\|_\infty \,,
\end{aligned}$$

where C_1 depends upon all the parameters of the problem.

□

In the case when we are interested in distinguishing two Gibbs states with Lipschitz observables, over extended subregions of the lattice, the following extension of Corollary C.4 can be easily shown to hold:

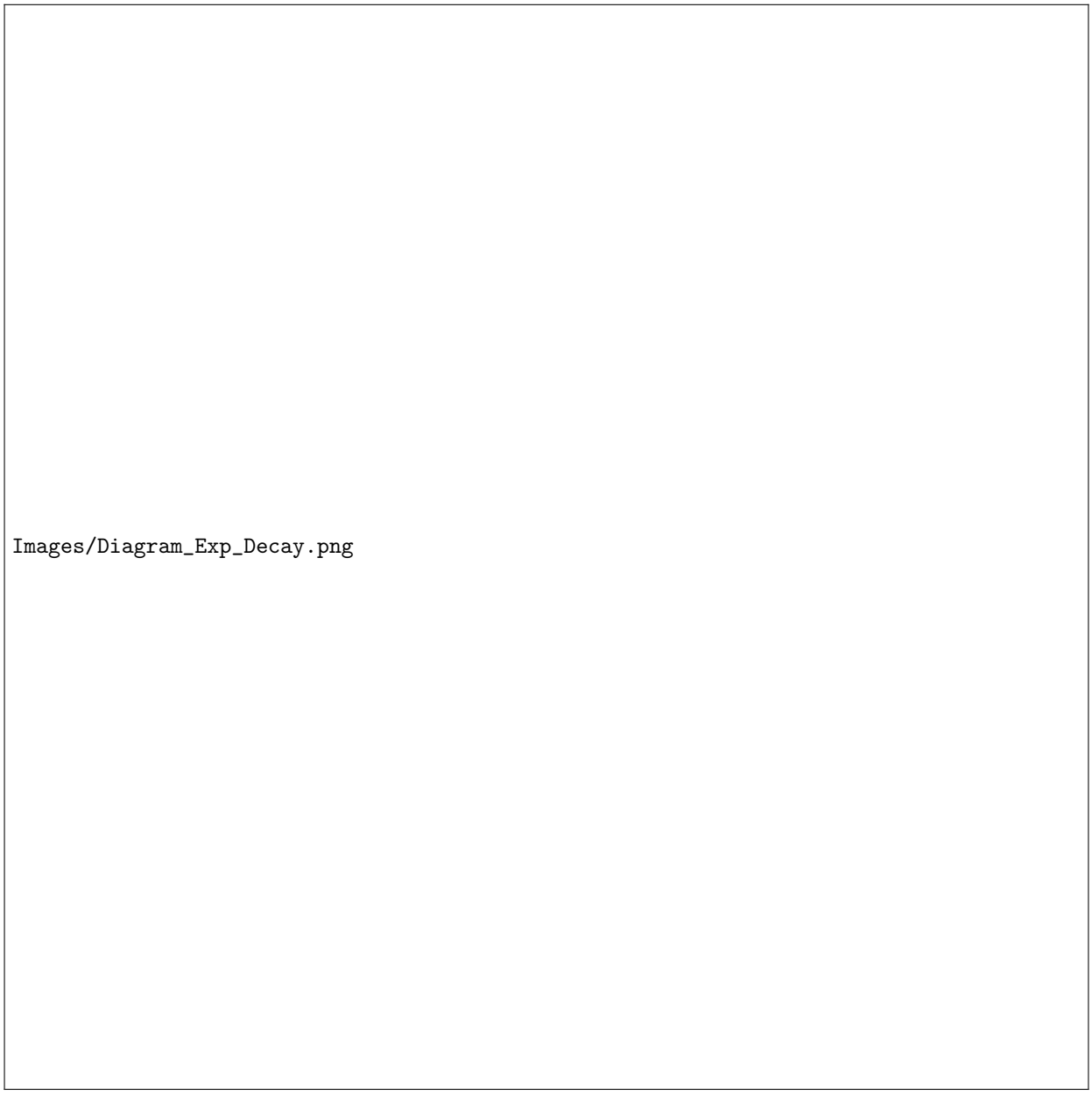

Figure 1. Diagram showing the regions involved in the proof Lemma C.5.

Corollary C.6. *Assume that the states $\sigma(\beta, x)$ satisfy the condition of decay of correlations Condition A.1. Then for any region S of the lattice and any two $x, y \in [-1, 1]^m$*

$$W_1(\mathrm{tr}_{S^c}(\sigma(\beta, x)), \mathrm{tr}_{S^c}(\sigma(\beta, y))) \leq \|x|_{\mathcal{S}(r)} - y|_{\mathcal{S}(r)}\|_{\ell_1} \ \mathrm{polylog}(|S(r)|)\,,$$

where $r = \max\left\{r_0, 2\xi \log\left(\frac{2|S|C_1}{\|x|_{\mathcal{S}(r_0)} - y|_{\mathcal{S}(r_0)}\|_{\ell_1}}\right)\right\}$ with r_0 being the smallest integer such that $x|_{\mathcal{S}(r_0)} \neq y|_{\mathcal{S}(r_0)}$, and C_1, ξ are the same constants as in Lemma C.5.

Proof. Given L_S a Lipchitz observable supported on region S of the lattice, we have for any $r \in \mathbb{N}$:

$$\begin{aligned}\left|f_{L_S}(x) - f_{L_S}(y)\right| &\leq \|L_S\|_\infty \left(\left\| \operatorname{tr}_{S^c} \left(\sigma(\beta, x) - \sigma(\beta, x|_{\mathcal{S}(r)})\right)\right\|_1 + \left\| \operatorname{tr}_{S^c} \left(\sigma(\beta, y) - \sigma(\beta, y|_{\mathcal{S}(r)})\right)\right\|_1 \right) \\ &\quad + W_1\big(\sigma(\beta, x|_{\mathcal{S}(r)}), \sigma(\beta, y|_{\mathcal{S}(r)})\big) \\ &\leq 2\,|S|\,\|L_S\|_{\mathrm{Lip}}\, C_1\, e^{-\frac{r}{2\xi}} + W_1\big(\sigma(\beta, x|_{\mathcal{S}(r)}), \sigma(\beta, y|_{\mathcal{S}(r)})\big)\,,\end{aligned}$$

where the second line follows from Equation (C.16). By Corollary C.4, we conclude that

$$\begin{aligned}W_1\big(\operatorname{tr}_{S^c}(\sigma(\beta, x)), \operatorname{tr}_{S^c}(\sigma(\beta, y))\big) &\leq 2\,|S|\,C_1\, e^{-\frac{r}{2\xi}} + W_1\big(\sigma(\beta, x|_{\mathcal{S}(r)}), \sigma(\beta, y|_{\mathcal{S}(r)})\big) \\ &\leq 2\,|S|\,C_1\, e^{-\frac{r}{2\xi}} + \|x|_{\mathcal{S}(r)} - y|_{\mathcal{S}(r)}\|_{\ell_1}\ \mathrm{polylog}(|S(r)|)\,.\end{aligned}$$

Next, we choose $r = 2\xi \log\left(\frac{2|S|C_1}{\|x|_{\mathcal{S}(r_0)} - y|_{\mathcal{S}(r_0)}\|_{\ell_1}}\right)$, where r_0 is the smallest integer such that $x|_{\mathcal{S}(r_0)} \neq y|_{\mathcal{S}(r_0)}$.

□

3. Hamiltonian estimation and optimal Gibbs state tomography

From Corollary C.4 it is immediate that we reduced the problem of obtaining a good estimate in W_1 to the problem of estimating the parameters of the Gibbs state $\sigma(\beta, x)$. Indeed, it is clear that if we can obtain an estimate x' of x satisfying

$$\|x - x'\|_{\ell_1} = \mathcal{O}(\varepsilon n/\mathrm{polylog}(n)), \tag{C.18}$$

then it suffices to ensure that $W_1(\sigma(\beta, x), \sigma(\beta, x')) = \varepsilon n$. Let us discuss some examples where we can obtain this efficiently with $\mathcal{O}(\varepsilon^{-2}\mathrm{polylog}(n))$ samples.

a. *Commuting Hamiltonians*

In [Ans], the authors give an algorithm which with

$$e^{\mathcal{O}(\beta k^D)}\mathcal{O}(\log(\delta^{-1}n)\varepsilon^{-2}) \tag{C.19}$$

copies of $\sigma(\beta, x)$ learns x up to ε in ℓ_∞ distance when $\sigma(\beta, x)$ belongs to a family of commuting, k-local Hamiltonians on a D-dimensional lattice. As we assumed that the number of parameters $m = \mathcal{O}(n)$, this translates to an algorithm with sample complexity $e^{\mathcal{O}(\beta k^D)}\mathcal{O}(\varepsilon^{-2}\mathrm{polylog}(\delta^{-1}n))$ to learn x up to εn in ℓ_1 distance. It should be noted that the time complexity of their algorithm is $\mathcal{O}(ne^{\mathcal{O}(\beta k^D)}\varepsilon^{-2}\mathrm{polylog}(\delta^{-1}n))$.

Thus, any commuting model at constant temperature satisfying exponential decay of correlations can be efficiently learned with $\mathrm{polylog}(n)$ samples. Examples of classes of commuting states that satisfy exponential decay of correlations include:

1. 1D translation-invariant Hamiltonians at any positive temperature [Ara69].
2. Commuting Gibbs states of Hamiltonians on regular lattices below a threshold temperature [KGK+14, HMS20].
3. Classical spin models away from criticality [DS87, LSS19, HMS20].
4. Ground states of uniformly gapped systems [BHM10, BH11, MZ13, NSY22].

b. High-temperature Gibbs states

Another class of states for which the conditions of our results hold are local Gibbs states on a lattice above a threshold temperature that depends on the locality of the Hamiltonian and the dimension of the lattice. These systems are known to have exponential decay of correlations [KGK+14, HMS20]. Furthermore, in [HKT21] the authors give an algorithm to learn x up to error ε in ℓ_∞ norm from $\mathcal{O}(\varepsilon^{-2}\text{polylog}(\delta^{-1}n))$ samples. This again translates to a $\mathcal{O}(\varepsilon n)$ error in ℓ_1 norm. Note that their algorithm also is computationally efficient.

We note that in [AAKS21] the authors give an algorithm to learn the Hamiltonian of any Gibbs state of positive temperature through the maximum entropy method. However, their results require a polynomial number of samples to recover the parameters in ℓ_1 distance. Thus, their results do not work for the polylog regime investigated in this work.

c. Gibbs state of exponentially decaying correlations and conditional mutual information

In the previous section, we extracted two regimes for which there exist efficient Gibbs tomography algorithms from previous works, namely the commuting and the high-temperature regimes. As said before, depending on the Hamiltonian, exponential decay of correlations can also occur in the low-temperature regime, and it is an interesting open question whether our strategy can be adapted to that setting for non-commuting interactions.

Here, we show that the Gibbs state $\sigma(\beta, x)$ of a possibly non-commuting Hamiltonian $H(x)$ can also be estimated in Wasserstein distance up to multiplicative error εn given polylog(n) copies of it as long as the latter has exponentially decaying correlations and is close to a quantum Markov chain, hence partially answering an open problem previously raised in [AAKS21].

To be more precise, in this section we will require a stronger notion of decay of correlations:

Definition C.7 (Uniform clustering)**.** *The Gibbs state $\sigma(\beta, x)$ is said to be uniformly $\zeta(\ell)$-clustering if for any $X \subset \Lambda$ and any $A \subset X$ and $B \subset X$ such that* $\text{dist}(A, B) \geq \ell$*,*

$$\text{Cov}_{\sigma(\beta,x,X)}(X_A, X_B) \leq \|X_A\|_\infty \|X_B\|_\infty \zeta(\ell)$$

for any X_A supported on A and X_B supported on B.

As pointed out in [BK18], this property is called uniform clustering to contrast with regular clustering property that usually only refers to properties of the state $\sigma(\beta, x)$.

Definition C.8 (Uniform Markov condition)**.** *The Gibbs state $\sigma_\beta(x)$ is said to satisfy the uniform $\delta(\ell)$-Markov condition if for any $ABC = X \subset \Lambda$ with B shielding A away from C and such that* $\text{dist}(i, j) \geq \ell$ *for any $i \in A$ and $j \in C$, we have*

$$I(A : C|B)_{\sigma(\beta,x,X)} \leq \delta(\ell)\,.$$

This property always holds for commuting Gibbs states for a function $\delta(\ell) = 0$ as soon as ℓ is larger than twice the interaction range. Although not proven yet, it is believed that the approximate Markov property holds with some generality for non-commuting Gibbs states. The 1D and high-temperature settings were investigated in [KB19] and [KKBa20], respectively. The decay of the conditional mutual information was also shown for finite temperature Gibbs states

of free fermions, free bosons, conformal field theories, and holographic models [SM16], as well as more recently for purely generated finitely correlated states in [SK22].

We will now show how to learn states that satisfy both the uniform Markov condition and the uniform clustering of correlations. Our strategy consists in using the maximum entropy estimation [Jay57b, Jay57a, Jay82, BKL+17], already appearing in [AAKS21], to construct an estimator $\hat{x}$ of the parameter $x \in [-1,1]^m$. The condition of exponential decay of correlations and that of approximate Markov chain will ensure that $W_1(\sigma(\beta,\hat{x}),\sigma(\beta,x)) = o(n)$. Thus, we once again emphasise that our goal is to obtain a good recovery of the state, not of the parameter x.

For sake of clarity and simplicity of presentation, we only consider the 1D setting, although our method easily extends to arbitrary dimension. We assume that each interaction $h_j(x_j)$ is of the form

$$h_j(x_j) := \sum_{l=1}^{\ell} x_{j,l}\, h_{j,l}$$

for some self-adjoint operators $h_{j,l}$ supported in $A_j := \{k \in \Lambda | \operatorname{dist}(k,j) \leq r_0\}$ with $\|h_{j,l}\| \leq h$, where we denoted by $x_{j,l}$ the entries of x_k. We also recall that given a region R of the lattice, we denote $H_R(x) := \sum_{k\in R} h_k(x_k)$. In what follows, with a slight abuse of notations, we denote by the same symbol a vector $y = \{y_{k,l}\}_{k\in N_j}$ and its embedding $(y,0)$ onto $[-1,1]^m$. Then, given an inverse temperature $\beta > 0$, we define the partition function as

$$Z_\beta(x) = \operatorname{tr}\left[e^{-\beta H(x)}\right].$$

The maximum entropy problem consists in the following strongly convex optimisation problem.

Theorem C.9 ([AAKS21])**.** *Given an unknown Hamiltonian $H(x) = \sum h_j(x_j)$, define $e_{k,l} = \operatorname{tr}[h_{k,\ell}\sigma(\beta,x)]$. Solving the following optimisation problem:*

$$\hat{x} := \operatorname*{arg\,min}_{y\in[-1,1]^m} L(y)\,, \qquad \textit{where} \qquad L(y) := \log Z_\beta(y) + \beta \sum_{k\in\Lambda}\sum_{l=1}^{\ell} y_{k,l}\, e_{k,l} \tag{C.20}$$

gives $\hat{x}$ such that $\sigma(\beta,\hat{x}) = \sigma(\beta,x)$.

In an experimental setting, we will not have access to the exact $\{e_{k,l}\}_{k,l}$, but instead may be able to approximate them using by having access to the state. However, we want to be sure that having a reasonably good approximation to $e_{k,l}$ is sufficient to approximate x. To do so one can make use of the fact that

$$\log Z_\beta(\hat{x}) \leq \log Z_\beta(x) + \beta \sum_{k\in\Lambda}\sum_{l=1}^{\ell} (x_{k,l} - \hat{x}_{k,l})\, \widetilde{e}_{k,l}\,. \tag{C.21}$$

Further assuming α_2 is a lower bound on the strong convexity constant associated to the function $x \mapsto \log Z_\beta(x)$, that is $\nabla^2 Z_\beta \geq \alpha_2\, I$, we have by Taylor expansion and since $\partial_{x_{k,l}} \log Z_\beta(x) = -\beta e_{k,l}(x)$:

$$\log Z_\beta(\hat{x}) \geq \log Z_\beta(x) - \beta \sum_{k\in\Lambda}\sum_{l=1}^{\ell} (\hat{x}_{k,l} - x_{k,l})\, e_{k,l}(x) + \frac{\alpha_2}{2}\, \|x-\hat{x}\|_{\ell_2}^2\,. \tag{C.22}$$

Combining the two bounds above, we find that

$$\|x-\hat{x}\|_{\ell_2}^2 \leq \frac{2\beta}{\alpha_2}\sum_{k,l}(x_{k,l}-\hat{x}_{k,l})(\widetilde{e}_{k,l}-e_{k,l}(x)) \leq \frac{2\beta}{\alpha_2}\|x-\hat{x}\|_{\ell_2}\|e-\widetilde{e}\|_{\ell_2}\,,$$

and hence $\|x-\hat{x}\|_{\ell_2} \leq \frac{2\beta\sqrt{\ell|\Lambda|}\,\eta}{\alpha_2}$, thus giving the following theorem:

Theorem C.10 ([AAKS21])**.** *Suppose $\widetilde{e}_{k,l}$ is an approximation of $e_{k,l}(x) := \operatorname{tr}\left[h_{k,l}\,\sigma(\beta,x)\right]$ with $\|\widetilde{e}-e(x)\|_{\ell_\infty} \leq \eta$. Assume that the following inequality is satisfied for some α_2: $\nabla^2 Z_\beta \geq \alpha I$. Solving the following optimisation problem:*

$$\hat{x} := \operatorname*{arg\,min}_{y\in[-1,1]^m} L(y)\,, \qquad \textit{where} \qquad L(y) := \log Z_\beta(y) + \beta\sum_{k\in\Lambda}\sum_{l=1}^{\ell} y_{k,l}\,\widetilde{e}_{k,l} \tag{C.23}$$

gives an output $\hat{x}$ satisfying:

$$\|\hat{x}-x\|_{\ell_2} \leq \frac{2\beta\eta\sqrt{\ell\Lambda}}{\alpha_2}.$$

Using the bound on $\hat{x}$ from Theorem C.10 the equivalence between ℓ_1 and ℓ_2-norms, we have that

$$\|x-\hat{x}\|_{\ell_1} \leq \frac{2\beta\ell n\eta}{\alpha_2}\,,$$

which provides us with the right scaling for our ℓ_1 approximation problem as long as $\eta = o(1)$ and $\alpha_2 = \Omega(1)$. Unfortunately, the constant α_2 could only be proved to scale inverse polynomially with n in [AAKS21]. A first idea from there is to try and find a constant $\alpha_1 = \Omega(n^{-1})$ such that the following strong convexity bound with respect to the ℓ_1-norm holds. As per eq. (C.22), this would imply:

$$\log Z_\beta(\hat{x}) \geq \log Z_\beta(x) - \beta\sum_{k,l}(\hat{x}_{k,l}-x_{k,l})\,e_{k,l}(x) + \frac{\alpha_1}{2}\|x-\hat{x}\|_{\ell_1}^2\,. \tag{C.24}$$

If such a bound held, we would conclude similarly to the previous setting that

$$\|x-\hat{x}\|_{\ell_1} \leq \frac{2\beta\eta}{\alpha_1} = o(\eta n)\,.$$

Which together with the continuity bound Equation (C.9) would allow us to get the desired recovery estimate in Wasserstein distance. Now, it can be seen that Equation (C.24) is equivalent to

$$\|x-\hat{x}\|_{\ell_1}^2 \leq \frac{2}{\alpha_1}D(\sigma(\beta,x)\|\sigma(\beta,\hat{x}))\,. \tag{C.25}$$

Here we recall that the relative entropy between two quantum states ρ and σ with $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$ is $D(\rho\|\sigma) := \operatorname{tr}\rho\log\rho - \operatorname{tr}\rho\log\sigma$. This together with Equation (C.9) would lead to the following local version of the *transportation cost* inequality

$$W_1(\sigma(\beta,x),\sigma(\beta,\hat{x}))^2 \leq \mathcal{O}(n\,\mathrm{polylog}(n))\,D(\sigma(\beta,x)\|\sigma(\beta,\hat{x}))\,. \tag{C.26}$$

In [PR22], such inequality was shown to hold in the high-temperature regime only for commuting H, albeit when $\sigma(\beta, x)$ can be replaced by an arbitrary state ρ on the lattice. The latter is referred to as a transportation-cost inequality for the state $\sigma(\beta, \hat{x})$. Since Equation (C.24) consists in a strengthening of Equation (C.26), proving it directly appears difficult. Here instead, we want to show the following weakening of (C.26):

$$W_1(\sigma(\beta, x), \sigma(\beta, \hat{x}))^2 \leq \mathcal{O}(n \operatorname{polylog}(n))\, D(\sigma(\beta, x) \| \sigma(\beta, \hat{x})) + o(\varepsilon n)\,,$$

for some constant δ which depends on the approximate Markov as well as the correlation decay properties of the Gibbs state $\sigma(\beta, \hat{x})$. More precisely, we show the following extension of [PR22, Theorem 4] to Gibbs states of non-commuting Hamiltonians.

Proposition C.11 (Generalised transportation-cost inequality)**.** *With the notations of the above paragraph, for all states ρ:*

$$W_1(\rho, \sigma(\beta, x)) \leq \inf_{\ell \in \mathbb{N}} \mathcal{O}(\ell\sqrt{n}) \sqrt{D(\rho \| \sigma(\beta, x))} + n^2 \big(\delta(\mathcal{O}(\ell)) + \zeta(\mathcal{O}(\ell)) + e^{-\mathcal{O}(\ell)}\big)\,.$$

In particular, if both $\zeta(l), \delta(l) = \mathcal{O}(e^{-\xi l})$, then for $l = \mathcal{O}(\xi^{-1} \log(n\varepsilon^{-1}))$ we have

$$W_1(\rho, \sigma(\beta, x)) \leq \mathcal{O}(\log(n\varepsilon^{-1})\sqrt{n}) \sqrt{D(\rho \| \sigma(\beta, x))} + o(\varepsilon n). \tag{C.27}$$

Proof. The proof is adapted from that of [PR22, Theorem 4]. We first consider a bipartite quantum subsystem $AB \subset \Lambda$ and a joint quantum state ω_{AB} of AB. We then define the so-called quantum recovery map [SBT16, JRS+18] by its action on a quantum state ρ_A on region A:

$$\Phi_{A \to AB}(\rho_A) = \int_{\mathbb{R}} \omega_{AB}^{\frac{1-it}{2}}\, \omega_A^{\frac{it-1}{2}}\, \rho_A\, \omega_A^{-\frac{1+it}{2}}\, \omega_{AB}^{\frac{1+it}{2}}\, d\mu_0(t)\,, \tag{C.28}$$

where μ_0 is the probability distribution on $\mathbb{R}$ with density

$$d\mu_0(t) = \frac{\pi\, dt}{2\,(\cosh(\pi t) + 1)}\,. \tag{C.29}$$

If A is in the state ω_A, the recovery map $\Phi_{A \to AB}$ recovers the joint state ω_{AB}, *i.e.*, $\Phi_{A \to AB}(\omega_A) = \omega_{AB}$. The relevance of the recovery map comes from the recoverability theorem [SBT16], which states that $\Phi_{A \to AB}$ can recover a generic joint state ρ_{AB} from its marginal ρ_A if removing the subsystem B does not significantly decrease the relative entropy between ρ and ω. More precisely, for any quantum state σ_{AB} of AB we have

$$D(\sigma_{AB} \| \omega_{AB}) - D(\sigma_A \| \omega_A) \geq D_{\mathbb{M}}(\sigma_{AB} \| \Phi_{A \to AB}(\sigma_A))\,, \tag{C.30}$$

where $D_{\mathbb{M}}$ denotes the measured relative entropy [Don86, Pet86, HP91, BFT17]

$$D_{\mathbb{M}}(\sigma \| \omega) := \sup_{(\mathcal{X}, M)} D(P_{\sigma, M} \| P_{\omega, M})\,, \tag{C.31}$$

where the supremum above is over all positive operator valued measures M that map the input quantum state to a probability distribution on a finite set $\mathcal{X}$ with probability mass function given by $P_{\rho, M}(x) = \operatorname{tr} \rho M(x)$.

Next, we split region A into regions A_1 and A_2 such that A_1 shields A_2 away from B, and take $\sigma_{AB} := \mathrm{tr}_{(AB)^c}(\sigma(\beta, x))$ and $\omega_{AB} = \sigma_{A_1B} \otimes \sigma_{A_2}$ In that case, (C.30) becomes

$$I(B : A_2|A_1)_\sigma \geq D_{\mathbb{M}}\big(\sigma \| \Phi_{A_1 \to A_1 B}(\sigma_A)\big)\,, \tag{C.32}$$

where we also used that the state ω is a tensor product in the cut $A_1B - A_2$, so that $\Phi_{A\to B} = \Phi_{A_1\to B}$.

Next, we pave the chain Λ into unions of intervals $A = \cup_{i=1}^M A_i$ and $B = \cup_{i=1}^M B_i$ such that $A_i \cap B_i \neq \emptyset$ and $B_i \cap A_{i+1} \neq \emptyset$. As in [BK18], we then define the channel $\mathcal{F} := \mathcal{F}_A \circ \mathcal{F}_B$ where $\mathcal{F}_B := \bigotimes_i \sigma(\beta, x, B_i) \otimes \mathrm{tr}_{B_i}$ and $\mathcal{F}_A := \prod_j \Phi_{A_i \setminus B \to A_i} \circ \mathrm{tr}_{A_i}$. In words, the channel $\mathcal{F}_B$ first prepares the Gibbs state in the region B, whereas $\mathcal{F}_A$ prepares the remaining of the Gibbs state onto region $A\setminus B$. Then, we have, for any state ρ

$$\begin{aligned}
W_1(\rho, \sigma(\beta, x)) &\leq W_1(\rho, \mathcal{F}_B(\rho)) + W_1(\mathcal{F}_B(\rho), \mathcal{F}_A \circ \mathcal{F}_B(\rho)) + W_1(\mathcal{F}(\rho), \sigma(\beta, x)) \\
&\leq \sum_i W_1(\sigma(A, i), \sigma(A, i+1)) + \sum_i W_1(\sigma(B, i), \sigma(B, i+1)) + n \left\| \mathcal{F}(\rho) - \sigma(\beta, x) \right\|_1 \\
&\overset{(1)}{\leq} R \sum_i \left\| \sigma(A, i) - \sigma(A, i+1) \right\|_1 + \left\| \sigma(B, i) - \sigma(B, i+1) \right\|_1 + n \| \mathcal{F}(\rho) - \sigma(\beta, x) \|_1 \,,
\end{aligned}$$

where $\sigma(A, i) := \bigotimes_{j<i} \sigma(\beta, x, B_j) \otimes \mathrm{tr}_{B_j}(\rho)$ and $\sigma(B, i) := \bigotimes_{j<i} \Phi_{A_j \setminus B, \to A_j} \circ \mathrm{tr}_{A_j}(\mathcal{F}_B(\rho))$, and where $R = \max\{|B_i|, |A_i|\}$ so that (1) follows from [DPMTL21, Propositiom 5]. Next, we use Pinsker's inequality, so

$$\begin{aligned}
&W_1(\rho, \sigma(\beta, x)) \\
&\quad \leq R\sqrt{2} \sum_i \sqrt{D_{\mathbb{M}}(\sigma(A, i) \| \sigma(A, i+1))} + \sqrt{D_{\mathbb{M}}(\sigma(B, i) \| \sigma(B, i+1))} + n \left\| \mathcal{F}(\rho) - \sigma(\beta, x) \right\|_1 \\
&\quad \leq 2R\sqrt{M} \left(\sum_i D_{\mathbb{M}}(\sigma(A, i) \| \sigma(A, i+1)) + D_{\mathbb{M}}(\sigma(B, i) \| \sigma(B, i+1)) \right)^{\frac{1}{2}} + n \| \mathcal{F}(\rho) - \sigma(\beta, x) \|_1 \\
&\quad \overset{(1)}{\leq} 2R\sqrt{M} \sqrt{D(\rho \| \sigma(\beta, x))} + n \| \mathcal{F}(\rho) - \sigma(\beta, x) \|_1 \\
&\quad \overset{(2)}{\leq} 2R\sqrt{M} \sqrt{D(\rho \| \sigma(\beta, x))} + n^2 \big(\delta(\mathcal{O}(R)) + \zeta(\mathcal{O}(R)) + C_1 e^{-c_2 R} \big) \,,
\end{aligned}$$

where (1) follows from multiple uses of (C.30) as well as the sub-additivity of the relative entropy under tensor products in its second argument, whereas (2) comes from [BK18, Theorem 6]. The result follows. □

We can then easily turn the previous statement into one about learning the state $\sigma(\beta, x)$:

Corollary C.12 (W_1 learning from the uniform Markov condition)**.** *Under the same conditions of Proposition C.11 assume further that $\zeta(l), \delta(l) = \mathcal{O}(e^{-\xi l})$. Then $\mathcal{O}(\varepsilon^{-4} \operatorname{polylog}(n\delta^{-1}))$ samples of $\sigma(\beta, x)$ suffice to learn a state $\sigma(\beta, x')$ s.t. with probability at least $1 - \delta$*

$$W_1(\sigma(\beta, x), \sigma(\beta, x')) = \mathcal{O}(\varepsilon n). \tag{C.33}$$

Proof. We simply need to adapt the proof of Equation (C.25). First, we recall that

$$D(\sigma(\beta,x)\|\sigma(\beta,\hat{x})) = \beta \sum_{k,l} (\hat{x}_{k,l} - x_{k,l})\, e_{k,l}(x) + \log Z_\beta(\hat{x}) - \log Z_\beta(x)\,.$$

Together with Equation (C.27), we have the following approximate strong convexity bound for the log partition function in the Wasserstein topology:

$$W_1(\sigma(\beta,x),\sigma(\beta,\hat{x}))^2 \leq \mathcal{O}(\log(n\varepsilon^{-1})^2 n)\, D(\rho\|\sigma(\beta,x)) + o(\varepsilon^2 n^2)$$

Combining with Equation (C.21) and assuming that $\|\widetilde{e} - e(x)\|_{\ell_\infty} \leq \eta$, we get

$$\begin{aligned} W_1(\sigma(\beta,x),\sigma(\beta,\hat{x}))^2 &\leq \mathcal{O}(\log(n\varepsilon^{-1})^2 n)\, \beta \sum_{k,l} (\hat{x}_{k,l} - x_{k,l})(e_{k,l}(x) - \widetilde{e}_{k,l}) + o(\varepsilon^2 n^2) \\ &\leq \mathcal{O}(\log(n\varepsilon^{-1})^2 n^2)\, \eta \, + o(\varepsilon^2 n^2)\,. \end{aligned}$$

Above, we have managed to reduce the problem to that estimating the coefficients $e_{k,l}$ to precision $\eta = \mathcal{O}(\varepsilon^2/\operatorname{polylog}(n))$. This can be done with probability $1-\delta$ given $\mathcal{O}(\varepsilon^{-4}\operatorname{polylog}(n\delta^{-1}))$ copies of the state $\sigma(\beta,x)$ through various methods (see e.g. [RF21, Appendix A] or [AAKS21, Corrolary 27] for more details).

□

Remark C.13. The main difference between the approach outlined in the proposition above and that of Corollary C.4 is that, besides requiring the additional assumption of uniform Markovianity, it has an ε^{-4} scaling with the precision instead of ε^{-2}. However, it completely bypasses the need for a good algorithm to learn the parameters x, which is required to apply Corollary C.4. Conversely, if combined with an ℓ_2 strong convexity guarantee on the partition function, the ε^{-2} scaling can be recovered (see e.g. [RF21, Theorem A.1] for more details). But, since we are currently only able to get this guarantee for commuting Hamiltonians or in the high-temperature regimes, Corollary C.4 provides a more direct path than the strategy previously exhibited in [RF21].

Appendix D: Algorithm for learning observables on Gibbs states with the same phase

Next, we assume we are given different Gibbs states $\sigma(\beta,x)$ where x is sampled according to the uniform distribution U over $x \in [-1,1]^m$, and wish to learn the expectation value of an unknown observable for all values of $x \in [-1,1]^m$. We assume that the Gibbs states in the interval $[-1,1]^m$ has exponentially decaying correlations everywhere, which can be thought of as defining a continuous phase of matter. For commuting Hamiltonians, this relationship can be made more precise [HMS20].

During a training stage, we pick N points $Y_1,\ldots,Y_N \sim U$ independently distributed uniformly at random in $[-1,1]^m$ and are given access to the Gibbs states $\sigma(\beta,Y_j)$. Next, fix $r \in \mathbb{N}$. Given an observable $O = \sum_{i=1}^M O_i$, we define $S_i = \operatorname{supp}(O_i)$ and for each S_i there is a ball of diameter at most k_0 containing S_i, and $S_i(r) := \{j \in \Lambda | \operatorname{dist}(j,S_i) \leq r\}$. We construct for any $x \in [-1,1]^m$ the estimator

$$\hat{f}_O(x) = \sum_{i=1}^M \operatorname{tr}\left[O_i\, \sigma(\beta,\hat{Y}_i(x))\right], \quad \text{with} \quad \hat{Y}_i(x) = \operatorname{argmin}_{Y_k} \|x|_{\mathcal{S}_i(r)} - Y_k|_{\mathcal{S}_i(r)}\|_{\ell^\infty}\,, \tag{D.1}$$

where we recall that we denote by $x_{\mathcal{S}_i(r)}$ the concatenation of vectors x_j corresponding to interactions h_j supported on regions intersecting $S_i(r)$. In words, we approximate the expectation value of O_i by that of the Gibbs state whose parameters in a region around S_i are the closest to the state of interest. We also denote $\mathcal{S}_i \equiv \mathcal{S}_i(0)$.

Lemma C.5 demonstrates that to measure a particular observable, it is sufficient to consider the parameters only spatially local to it. Next, we show that if we are given a number of samples uniformly sampled from $[-1,1]^m$, we can construct an estimator for the observable to high precision.

Proposition D.1. *Assuming the exponential decay of correlations in Condition A.1, the estimator* $\hat{f}_O(x) := \sum_i \operatorname{tr}[O_i\, \sigma(\beta, \hat{Y}_i(x))]$ *satisfies the bound*

$$\sup_{x\in[-1,1]^m} |f_O(x) - \hat{f}_O(x)| \leq \varepsilon \sum_{i=1}^{M} \|O_i\|_\infty \,,$$

with probability at least $1-\delta$*, whenever*

$$N = \left(\frac{\gamma}{2}\right)^{-[2(r+r_0+k_0)]^D \ell} \log\left(\frac{M}{\delta}\right) + [2(r+r_0+k_0)]^D \ell \log\left(\frac{2}{\gamma}\right)\left(\frac{\gamma}{2}\right)^{-[2(r+r_0+k_0)]^D \ell}$$

with

$$r = \left\lceil 2\xi \log\left(\frac{16\beta(C+c')\, h\, (2r_0+k_0)^D (D-1)! (2\xi)^{D-1}\, D^{D-1}}{\varepsilon\, e^{\frac{k_0+1}{2\xi}} (1-e^{-\frac{1}{2\xi}})}\right)\right\rceil \,,$$
$$\gamma = \frac{\varepsilon\, e^{-[2(r+k_0)]^D (3\log 2 + 5\beta h)}}{2[2(r+k_0)]^D h \ell} \,.$$

Before we prove this result, let us simplify its statement. First, $r = \Theta(\log(\varepsilon^{-1}))$, $\gamma = \Theta\Big(\varepsilon \frac{e^{-\log(\varepsilon^{-1})^D}}{\log(\varepsilon^{-1})^D}\Big)$, so that the number of samples needed is asymptotically

$$N = \Theta\left(\log\left(\frac{M}{\delta}\right) e^{\operatorname{polylog}(\varepsilon^{-1})}\right) \,.$$

Proof. We fix O_i and $r > 0$, and restrict ourselves to the subset of parameters $x|_{\mathcal{S}_i(r)}$. The number of parameters in that subset is bounded by the volume $V(r+r_0+k_0)$ of the ball $S_i(r+r_0)$ times the number ℓ of parameters per interaction. We denote it by $m_r := V(r+r_0+k_0)\ell$. Next, we partition the parameter space $[-1,1]^{m_r}$ onto cubes of side-size $\gamma \in (0,1)$. By the coupon collector's problem, we have that the probability that none of the sub-vectors $Y_j|_{\mathcal{S}_i(r)}$ is within one of those cubes is upper bounded by $e^{-N(\gamma/2)^{m_r} + m_r \log(2/\gamma)}$. By union bound, the probability that for any $i \in [M]$, any cube is visited by at least one sub-vector $Y_j|_{\mathcal{S}_i(r)}$ is lower bounded by $1-\delta$, $\delta := M e^{-N(\gamma/2)^{m_r} + m_r \log(2/\gamma)}$. In other words, with probability $1-\delta$ there is a $\hat{Y}_i(x)|_{\mathcal{S}_i(r)}$ in the N samples satisfying

$$\|x|_{\mathcal{S}_i(r)} - \hat{Y}_i(x)|_{\mathcal{S}_i(r)}\|_{\ell_\infty} \leq \gamma \tag{D.2}$$

for all $i \in [M]$.

Denoting $\hat{f}_{O_i}(x) := \operatorname{tr}\left[O_i\,\sigma(\beta,\hat{Y}_i(x)|_{\mathcal{S}_i(r)})\right]$, we next control $|f_{O_i}(x|_{\mathcal{S}_i(r)}) - \hat{f}_{O_i}(x)|$. This is easily done in terms of the ℓ_∞ norm distance between $\hat{Y}_i(x)|_{\mathcal{S}_i(r)}$ and $x|_{\mathcal{S}_i(r)}$ by Lipschitz estimate: writing $H(x|_{\mathcal{S}_i(r)}) \equiv H_1$ and $H(\hat{Y}_i(x)|_{\mathcal{S}_i(r)}) \equiv H_2$,

$$\begin{aligned}
\left|f_{O_i}(x|_{\mathcal{S}_i(r)}) - \hat{f}_{O_i}(x)\right| &\leq \|O_i\|_\infty \,\|\sigma(\beta, x|_{\mathcal{S}_i(r)}) - \sigma(\beta,\hat{Y}_i(x)|_{\mathcal{S}_i(r)})\|_1 \\
&\leq \|O_i\|_\infty \left\| \frac{e^{-\beta H_1}}{\operatorname{tr}[e^{-\beta H_1}]} - \frac{e^{-\beta H_2}}{\operatorname{tr}[e^{-\beta H_2}]} \right\|_1 \\
&\leq \|O_i\|_\infty \left(\frac{\|e^{-\beta H_1} - e^{-\beta H_2}\|_1}{\operatorname{tr}[e^{-\beta H_1}]} + \|e^{-\beta H_2}\|_1 \frac{|\operatorname{tr}[e^{-\beta H_1}] - \operatorname{tr}[e^{-\beta H_2}]|}{\operatorname{tr}[e^{-\beta H_1}]\operatorname{tr}[e^{-\beta H_2}]} \right) \\
&\leq \|O_i\|_\infty \,\|e^{-\beta H_1} - e^{-\beta H_2}\|_1 \,\frac{\operatorname{tr}\left[e^{-\beta H_1} + e^{-\beta H_2}\right]}{\operatorname{tr}\left[e^{-\beta H_1}\right]\operatorname{tr}\left[e^{-\beta H_2}\right]} \\
&\leq 2^{2V(r+k_0)+1} \|O_i\|_\infty \,\|e^{-\beta H_1} - e^{-\beta H_2}\|_\infty \, e^{3\beta V(r+k_0)h}\,.
\end{aligned}$$

Next, we use the integral perturbation bound for the exponential in order to bound

$$\begin{aligned}
\|e^{-\beta H_1} - e^{-\beta H_2}\|_\infty &\leq \int_0^1 \|e^{-(1-s)\beta H_1}(H_1 - H_2)e^{-\beta s H_2}\|_\infty \, ds \\
&\leq e^{2\beta V(r+k_0)h}\,\|H_1 - H_2\|_\infty \\
&\leq e^{2\beta V(r+k_0)h} \sum_{j\in S_i(r)} \|h_j(x_j - \hat{Y}_i(x)_j)\|_\infty \\
&\leq e^{2\beta V(r+k_0)h}\, V(r+k_0)h\ell\gamma\,,
\end{aligned}$$

where the last line comes from Equation (D.2). Combining this with Lemma C.5, from which we have that

$$|\operatorname{tr}\left[O_i\,(\sigma(\beta,x) - \sigma(\beta,x|_{\mathcal{S}_i(r)}))\right]|,\ |\operatorname{tr}\left[O_i\,(\sigma(\beta,\hat{Y}_i(x)) - \sigma(\beta,\hat{Y}_i(x)|_{\mathcal{S}_i(r)}))\right]| \leq C_1\, e^{-\frac{r}{2\xi}}\,\|O_i\|_\infty,$$

we have proven that for all $x \in [-1,1]^m$,

$$|f_{O_i}(x) - \hat{f}_{O_i}(x)| \leq \left(2C_1 e^{-\frac{r}{2\xi}} + C_2(r)\gamma\right)\|O_i\|_\infty\,, \tag{D.3}$$

where $C_2(r) := 2^{2V(r+k_0)+1} e^{5\beta V(r+k_0)h} V(r+k_0)h\ell$. Now, the volume $V(s)$ of a ball of radius s in Λ is equal to

$$V(s) = \sum_{a\leq s} \binom{a+D-1}{D-1} \leq (2s)^D\,.$$

We then fix r so that $2C_1 e^{-r/2\xi} \leq \varepsilon/2$, γ so that $C_2(r)\gamma \leq 2^{2^{D+1}(r+k_0)^D+1} e^{5\beta 2^D (r+k_0)^D h} 2^D (r+k_0)^D h\ell \leq \varepsilon/2$, and therefore a lower bound on N arises from the constraint

$$\delta := M e^{-N(\gamma/2)^{m_r} + m_r \log(2/\gamma)}\,.$$

Namely:

$$\begin{aligned}
r &= \left\lceil 2\xi \log\left(\frac{16\beta(C+c')\,h\,(2r_0+k_0)^D (D-1)!(2\xi)^{D-1}\,D^{D-1}}{\varepsilon\, e^{\frac{k_0+1}{2\xi}}(1 - e^{-\frac{1}{2\xi}})} \right) \right\rceil\,, \\
\gamma &= \frac{\varepsilon\, e^{-[2(r+k_0)]^D(3\log 2 + 5\beta h)}}{2[2(r+k_0)]^D h\ell}\,.
\end{aligned}$$

Therefore,

$$N = \left(\frac{\gamma}{2}\right)^{-[2(r+r_0+k_0)]^D \ell} \log\left(\frac{M}{\delta}\right) + [2(r+r_0+k_0)]^D \ell \log\left(\frac{2}{\gamma}\right)\left(\frac{\gamma}{2}\right)^{-[2(r+r_0+k_0)]^D \ell}$$

copies suffice for the approximation claimed to hold with probability $1-\delta$.

□

At this stage, we use the shadow tomography protocol to get classical shadows $\widetilde{\sigma}(\beta, \hat{Y}_i(x))$ for each of the states $\sigma(\beta, \hat{Y}_i(x))$. Since only one copy of each $\widetilde{\sigma}(\beta, \hat{Y}_i(x))$ is available, its reconstruction is likely going to be too noisy. Instead, we will use several non-identical copies $\sigma(\beta, \hat{Y}_i(x))$ which almost coincide on large enough regions surrounding the supports of observables O_i in order to improve the precision of the estimation of $\hat{f}_O$. We first develop the following Gibbs shadow tomography protocol which we believe to be of independent interest.

Consider a Gibbs state $\sigma(\beta, x)$ and a family $\sigma(\beta, x_1), \ldots, \sigma(\beta, x_N)$ of Gibbs states with the promise that for any $i \in [M]$ there exist t vectors $x_{i_1}, \ldots, x_{i_t}$ such that $\max_{j\in[t]} \|x|_{\mathcal{S}_i(r)} - x_{i_j}|_{\mathcal{S}_i(r)}\|_\infty \leq \gamma$. We run the shadow protocol and construct product operators $\widetilde{\sigma}(\beta, x_1), \ldots, \widetilde{\sigma}(\beta, x_N)$. Then for any ball B of radius k_0, we select the shadows $\widetilde{\sigma}(\beta, x_{i_1}), \ldots \widetilde{\sigma}(\beta, x_{i_t})$ and construct the empirical average

$$\widetilde{\sigma}_B(x) := \frac{1}{t}\sum_{j=1}^{t} \operatorname{tr}_{B^c}\left[\widetilde{\sigma}(\beta, x_{i_j})\right].$$

Proposition D.2 (Robust shadow tomography for Gibbs states)**.** *Fix $\varepsilon, \delta \in (0,1)$. In the notations of Proposition D.1, with probability $1-\delta'$, for any ball B of radius k_0, the shadow $\widetilde{\sigma}_B$ satisfies $\|\widetilde{\sigma}_B - \operatorname{tr}_{B^c}[\sigma(\beta, x)]\|_1 \leq 2C_1\, e^{-\frac{r}{2\xi}} + C_2(r)\gamma + \varepsilon$ as long as*

$$t \geq \frac{8.12^{k_0}}{3.\varepsilon^2} \log\left(\frac{n^{k_0} 2^{k_0+1}}{\delta'}\right). \tag{D.4}$$

Proof. In view of Proposition G.1, it is enough to show that the reduced states $\operatorname{tr}_{B^c}[\sigma(\beta, x_{i_j})]$ are close to $\operatorname{tr}_{B^c}[\sigma(\beta, x)]$. This is done by simply adapting some of the estimates in the proof of Proposition D.1. In particular, we have shown that

$$\|\operatorname{tr}_{B^c}\left[\sigma(\beta, x) - \sigma(\beta, x_{i_j})\right]\|_1 \leq 2C_1\, e^{-\frac{r}{2\xi}} + C_2(r)\gamma\,.$$

The result follows directly from Proposition G.1. □

We are now ready to state and proof the main result of this section. We denote $\widetilde{f}_{O_i}(x) = \operatorname{tr}\left[O_i\, \widetilde{\sigma}_{S_i}(x)\right]$ the function constructed from the Gibbs shadow tomography protocol of Proposition D.2, and write $\widetilde{f}_O := \sum_{i=1}^M \widetilde{f}_{O_i}$.

Theorem D.3 (Learning algorithm for quantum Gibbs states)**.** *In the notation of the previous paragraph, consider a set of N shadows $\{\widetilde{\sigma}(\beta, x_i)\}_{i=1}^N$. Given an arbitrary local observable O, we*

fix

$$r := \left\lceil 2\xi \log\left(\frac{24\beta(C+c')\,h\,(2r_0+k_0)^D(D-1)!(2\xi)^{D-1}\,D^{D-1}}{\varepsilon\, e^{\frac{k_0+1}{2\xi}}(1-e^{-\frac{1}{2\xi}})}\right)\right\rceil,$$
$$\gamma = \frac{\varepsilon\, e^{-[2(r+k_0)]^D(3\log 2+5\beta h)}}{3[2(r+k_0)]^D h\ell},$$
$$t := \left\lceil \frac{24.12^{k_0}}{\varepsilon^2}\log\left(\frac{n^{k_0}2^{k_0+1}}{\delta'}\right)\right\rceil.$$

Then, we have that with probability $(1-\delta).(1-\delta')$,

$$|f_O(x) - \widetilde{f}_O(x)| \leq \varepsilon \sum_i \|O_i\|_\infty\,,$$

as long as

$$N = t\Big(\frac{\gamma}{2}\Big)^{-[2(r+r_0+k_0)]^D\ell}\log\Big(\frac{M}{\delta}\Big) + t\log\Big[t\Big(\frac{2}{\gamma}\Big)^{[2(r+r_0+k_0)]^D\ell}\Big]\Big(\frac{\gamma}{2}\Big)^{-[2(r+r_0+k_0)]^D\ell}.$$

Once again, upon careful checking of the bounds, we have found

$$N = \Theta\left(\log\Big(\frac{M}{\delta}\Big)\,\log\Big(\frac{n}{\delta'}\Big)e^{\mathrm{polylog}(\varepsilon^{-1})}\right).$$

Proof. Adapting the proof of Proposition D.1, it is clear that with probability

$$1-\delta := 1 - Me^{-N\frac{1}{t}(\gamma/2)^{m_r}+m_r\log(2/\gamma)+\log t}$$

each cube is visited at least t times. Conditioned on that event, and choosing t such that Equation (D.4) holds, we have that with probability $1-\delta'$

$$|f_{O_i}(x) - \widetilde{f}_{O_i}(x)| \leq \Big(2C_1e^{-\frac{r}{2\xi}} + C_2(r)\gamma + \varepsilon\Big)\|O_i\|_\infty\,.$$

□

Remark D.4. We emphasise that the classical data $\{\widetilde{\sigma}(\beta,x_i)\}_{i=1}^N$ are fixed, and then any local observable O can be chosen *after the data has been taken* which will satisfy the bounds in theorem D.3.

Remark D.5. Our proof readily extends to distributions μ that satisfy the following anti-concentration property: for any $x_0 \in [-1,1]^{m_r}$ and for all π permutations of the coordinates of $[-1,1]^n$ we have:

$$\mu(A(x_0,\varepsilon,\pi)) > 0 \implies \mu(A(x_0,\varepsilon,\pi)) = \Omega((2\varepsilon)^{m_r}/\mathrm{polylog}(n))\,, \tag{D.5}$$

where $A(x_0,\varepsilon,\pi) := \pi((x_0+[-\varepsilon,\varepsilon]^{m_r})\times[-1,1]^{n-m_r})$. To see this, notice that the condition in Equation (D.5) implies that we can discretise the number of cubes with size ε and positive weight into at most $\mathcal{O}((2\varepsilon)^{-m_r}\mathrm{polylog}(n))$ cubes for any choice of m_r coordinates. By e.g. rejection sampling we can then generate a sample from the uniform distribution on those cubes by taking

at most $\mathcal{O}(\text{polylog}(n))$ samples from the distribution μ. Once given uniform samples over those cubes we can argue as in the proof above.

One distribution that satisfies the condition in Equation (D.5) but is far from uniform over the whole space is e.g. a Dirac measure on a single state. It is also satisfied for various natural distributions, such as i.i.d. distributions on each coordinate.

Remark D.6. It is clear that our stronger L_∞ recovery guarantee cannot hold for arbitrary distributions and requires some sort of anti-concentration. To see this, consider a distribution over parameters that outputs a state ρ_0 with probability $1-p$ and a different, locally distinguishable state ρ_1 with probability p. Before we have drawn $\Omega(p^{-1})$ samples it is unlikely that we gained access to even a single sample of ρ_1. But algorithms like ours with L_∞ guarantees also need to perform well on such rare outputs. Thus, we see that the sample complexity for L_∞ guarantees will have to depend on the parameter p and will blow-up as $p \to 0$. In contrast, if we wish to obtain good recovery in L_2 for this simple example as $p \to 0$, we can always output the expectation value w.r.t. ρ_0.

1. Learning the High-Temperature Phase

Theorem D.7 (Learnability of the High-Temperature Phase)**.** *Let $H(x)$ be a geometrically local Hamiltonian. Then there exists a temperature range $\beta \in [\beta_c, 0]$, such for all $x \in [-1,1]^m$ then the parameters of Theorem D.3 are sufficient to learn*

$$|f_O(x) - \widetilde{f}_O(x)| \leq \varepsilon \sum_i \|O_i\|_\infty ,$$

with probabilities and parameters as given in Theorem D.3.

Proof. From [KGK+14], we see that for sufficiently high-temperatures (low β), then the Gibbs states must satisfy exponentially decaying correlations. Thus we can utilise Theorem D.3 directly to get the parameters required to learn the high-temperature phase. □

Remark D.8. For 1D, translationally invariant Hamiltonians, the Gibbs state has exponential decay of correlations for all temperatures [BCPH22] and hence the phase can be learned efficiently everywhere.

Remark D.9. For commuting and 1D Hamiltonians we can relate the learnability of the phase to the analyticity of the free energy, and thus to a more rigorous notion of phase, defined as regions of parameter space where the free energy is analytic (assuming the Hamiltonian is parameterised in an analytic fashion). The free energy is defined as $F(\beta, x) = -\log(\text{tr}[e^{-\beta H(x)}])$. This is done using [HMS20, Theorem 32] which demonstrates that for commuting and 1D geometrically local Hamiltonians, exponential decay of correlations holds in the sense of Equation (II.2). Thus we can utilise Theorem D.3 directly to get the parameters.

One might attempt to relate this to the free energy of non-commuting Hamiltonians, however, as per [HMS20, Theorem 31], exponential decay of correlations in regions with analytic free energy is only known for observables O_1, O_2 whose supports are distance $\Omega(\log(n))$ away from each other. Although one would be able to prove learnability with more samples scaling with n, [HMS20, Theorem 31] is not strong enough to give the scaling we desire.

Remark D.10. Although we do not prove it here, it is likely we can flip the above remark on its head. If we consider a region of parameter space in which the free energy is analytic, we expect all local observables to be analytic in x in this region. As such, we should be able to approximate the local observable using polynomial interpolation (or some other technique) and learning the polynomial of this observable everywhere in the phase with small error.

It is worth noting that high-temperature Gibbs states of commuting Hamiltonians as well as those of 1D Hamiltonians can be efficiently prepared and hence desired observables could be measured directly [BK18]. Hence Theorem D.7 becomes most useful in the setting where parameters are a priori unknown to us and we wish to extract useful information.

Appendix E: Learning ground states

The previous results can be adapted to ground states with exponentially decaying correlations through the following tricks. Here by ground state we mean the limit

$$\psi_g(x) := \lim_{\beta\to\infty} \sigma(\beta, x)\,.$$

1. Commuting models

We first consider the same family of Hamiltonians $\{H(x)\}_{x\in[-1,1]^m}$ as the one defined in Equation (A.1), and assume the interactions $h_j(x_j)$ to be commuting for all x. The ground state $H(x)$ is denoted by $\psi_g(x)$. It was shown in [Ans] that the reduced Gibbs states $\mathrm{tr}_{R^c}\,\sigma(\beta, x)$ of commuting Hamiltonians can be written as

$$\mathrm{tr}_{R^c}\,\sigma(\beta, x) = \frac{e^{-\beta(H_R(x)+\Phi_R(x))}}{\mathrm{tr}\left(e^{-\beta(H_R(x)+\Phi_R(x))}\right)} \equiv \sigma'(x, \beta, R), \tag{E.1}$$

where we recall that $H_R(x)$ corresponds to the Hamiltonian restricted to the regions R, and where $\Phi_R(x)$ is an effective interaction term supported on the inner boundary $\partial_- R$ of R, which commutes with H_R and satisfies $\|\Phi_R(x)\|_\infty \leq 2|\partial R|$, where ∂R denotes the boundary of R.

Lemma E.1. *For any region R of the lattice, we denote by $\lambda_R(x)$ the spectral gap of $H_R(x) + \Phi_R(x)$. Then for any observable X_A, resp. X_B, supported on region A, resp. B, and any $\beta \geq \frac{2}{\lambda_{A\cup B}(x)} \log\left(\frac{8.2^{|A\cup B|}}{\varepsilon}\right)$,*

$$\left|\,\mathrm{Cov}_{\sigma(\beta,x)}(X_A, X_B) - \mathrm{Cov}_{\psi_g(x)}(X_A, X_B)\right| \leq \varepsilon\,\|X_A\|_\infty\,\|X_B\|_\infty\,.$$

Proof. The proof follows by rudimentary estimates: given an arbitrary region $R \subseteq \Lambda$, we denote σ_R, resp. ψ_R, the reduced state $\mathrm{tr}_{R^c}(\sigma(\beta, x))$, resp. $\mathrm{tr}_{R^c}(\psi_g(x))$. By the triangle inequality and Hölder's inequality, we have

$$\begin{aligned}
&\left|\,\mathrm{Cov}_{\sigma(\beta,x)}(X_A, X_B) - \mathrm{Cov}_{\psi_g(x)}(X_A, X_B)\right| \\
&\qquad \leq \Big[4\,\|\sigma_{AB} - \psi_{AB}\|_1 \;+\; 2\big(\|\sigma_A - \psi_A\|_1 + \|\sigma_B - \psi_B\|_1\big)\Big]\,\|X_A\|_\infty\,\|X_B\|_\infty \\
&\qquad \leq 8\,\|\sigma_{AB} - \psi_{AB}\|_1\,\|X_A\|_\infty\,\|X_B\|_\infty\,,
\end{aligned}$$

where the third line follows from the monotonicity of the trace distance under partial tracing. Next, for any region R, we denote $\psi'_g(x,R)$ the limit $\lim_{\beta\to\infty}\sigma'(x,\beta,R)$. This is a projection onto the ground eigenspace of $H_R(x)+\Phi_R(x)$. By Pinsker inequality, we have, denoting $E_0(x) < E_1(x) < \dots$ the ordered energies of $H_R(x)+\Phi_R(x)$ with corresponding multiplicities $m_j(x)$, and $\lambda_R(x)$ its gap:

$$\begin{aligned}
\|\sigma_{AB}-\psi_{AB}\|_1^2 &= \|\sigma'(x,\beta,AB)-\psi'(x,AB)\|_1^2 \\
&\leq 2\,D(\psi'_g(x,AB)\|\sigma'(x,\beta,AB)) \\
&\leq -2\log(m_0(x)) - 2\,\mathrm{tr}\left[\psi'_g(x,AB)\,\log(\sigma'(x,\beta,AB))\right] \\
&\leq -2\log(m_0(x)) - 2\,\log\left(\frac{e^{-\beta E_0(x)}}{\mathrm{tr}\left[e^{-\beta(H_R(x)+\Phi_R(x))}\right]}\right) \\
&= -2\log(m_0(x)) + 2\left(\beta E_0(x) + \log\sum_{j\geq 0} e^{-\beta E_j(x)} m_j(x)\right) \\
&= -2\log(m_0(x)) + 2\log\left(\sum_{j\geq 0} e^{-\beta(E_j(x)-E_0(x))} m_j(x)\right) \\
&= 2\log\left(1+\sum_{j\geq 1} e^{-\beta(E_j(x)-E_0(x))}\frac{m_j(x)}{m_0(x)}\right) \\
&\leq 2\,e^{-\beta\lambda_{AB}(x)}\sum_{j\geq 1}\frac{m_j(x)}{m_0(x)} \\
&\leq 2^{2|A\cup B|}\,e^{-\beta\lambda_{A\cup B}(x)}\,. \qquad \text{(E.2)}
\end{aligned}$$

The result follows after imposing the above bound to be smaller than $\varepsilon^2/8$. □

Remark E.2. Under suitable assumptions on the density of states, it is possible to improve the dependency of the temperature to $\beta \geq \Omega\left(\frac{1}{\lambda_{A\cup B}(x)}\log\left(|A\cup B|\varepsilon^{-1}\right)\right)$.

2. Non-commuting models and LTQO

To go beyond the commuting case, we make use of the notion of *local topological quantum order* (LTQO) [MZ13, BHM10, NSY22]: in words, the latter states that observables localised away from the boundary of the volume cannot distinguish between different ground states. From now on, given a region $R\subset\Lambda$ we denote by $\sigma(\beta,x,R)$, resp. $\psi_g(x,R)$, the Gibbs state corresponding to the Hamiltonian $H_R(x)$ at inverse temperature β, resp. the ground state

$$\psi_g(x,R) := \lim_{\beta\to\infty}\sigma(\beta,x,R)\,.$$

Definition E.3 (Local Topological Quantum Order)**.** *A quantum system satisfies* LTQO *if for any region $A\subset B\subset\Lambda$, all $x\in[-1,1]^m$,*

$$\left\|\,\mathrm{tr}_{A^c}(\psi_g(x,B)-\psi_g(x,\Lambda))\right\|_1 \leq C_T\,|A|\,e^{-\,\mathrm{dist}(A,B^c)/\xi_0} \qquad \text{(LTQO)}$$

for some constants $C_T,\xi_0>0$.

Remark E.4. The LTQO property is generally defined with respect to a fast-enough decaying function of $\mathrm{dist}(A, B^c)$. Here, we only considered exponential decay for simplicity, but it is not difficult to extend our proof to the general settings. It is also worth mentioning that the notion of LTQO was also extended to models with multiple distinguishable ground states, such as the Ising model, by replacing the above trace distance by an optimisation over observables X_A that satisfy a symmetry condition in [NSY22].

With this notion at hand, we can extend Lemma E.1 to the non-commutative setting:

Lemma E.5. *Assume that the quantum system is uniformly gapped,* $\inf_R \inf_x H_R(x) \geq \lambda_0 > 0$, *and satisfies LTQO. Then for any balls* A, B *of radius* r_0, *and all* $\beta \geq \frac{2}{\lambda_0} \log\left(\frac{2^{5(2(r+r_0))^D}}{\varepsilon}\right)$,

$$\left|\operatorname{Cov}_{\psi_g(x)}(X_A, X_B) - \operatorname{Cov}_{\sigma(\beta,x,A\cup B(r))}(X_A, X_B)\right| \leq \varepsilon \|X_A\|_\infty \|X_B\|_\infty ,$$

where $r := \xi_0 \log\left(\frac{16|A\cup B|C_T}{\varepsilon}\right)$. *Similarly, for any observable* X_A *supported on ball* A *of radius* r_0,

$$\left|\operatorname{tr}(X_A \psi_g(x)) - \operatorname{tr}(X_A \sigma(\beta, x, A(r)))\right| \leq \varepsilon \|X_A\|_\infty . \tag{E.3}$$

for all $\beta \geq \frac{2}{\lambda_0} \log\left(\frac{2^{2(2(r+r_0))^D}}{\varepsilon}\right)$, *where* $r = \xi_0 \log\left(\frac{2C_T|A|}{\varepsilon}\right)$.

Proof. We take a region R that includes $A \cup B$. By LTQO, we have

$$\begin{aligned}\left|\operatorname{Cov}_{\psi_g(x)}(X_A, X_B) - \operatorname{Cov}_{\psi_g(x,R)}(X_A, X_B)\right| &\leq 8 \|\operatorname{tr}_{(AB)^c}(\psi_g(x) - \psi_g(x,R))\|_1 \|X_A\|_\infty \|X_B\|_\infty \\ &\leq 8 C_T |A \cup B| e^{-\operatorname{dist}(A\cup B, R^c)/\xi_0} \|X_A\|_\infty \|X_B\|_\infty .\end{aligned} \tag{E.4}$$

Next, by the same computation to the one that leads to Equation (E.2), we have

$$\begin{aligned}\left|\operatorname{Cov}_{\psi_g(x,R)}(X_A, X_B) - \operatorname{Cov}_{\sigma(\beta,x,R)}(X_A, X_B)\right| &\leq 8 \|\psi_g(x,R) - \sigma(\beta,x,R)\|_1 \|X_A\|_\infty \|X_B\|_\infty \\ &\leq 2^{4|R|} e^{-\frac{\beta}{2}\lambda_R(x)} \|X_A\|_\infty \|X_B\|_\infty .\end{aligned} \tag{E.5}$$

We now choose R and β so that the right-hand sides of Equation (E.4) and Equation (E.5) are each bounded by $\frac{\varepsilon}{2}\|X_A\|_\infty \|X_B\|_\infty$. The proof of Equation (E.3) follows the same lines. □

Both Lemma E.1 and Lemma E.5 can be used to learn ground states of local Hamiltonians. We illustrate this in the next theorem. We first need a replacement for Proposition D.1.

Proposition E.6. *Denote* $f_{O_i,g}(x) := \operatorname{tr}(\psi_g(x) O_i)$ *and* $f_{O,g}(x) := \sum_i f_{O_i,g}(x)$. *With the assumptions of Lemma E.5, the estimator* $\hat{f}_{O,g}(x) := \sum_i \operatorname{tr}[O_i \, \psi_g(\hat{Y}_i(x))]$ *satisfies the bound*

$$\sup_{x\in[-1,1]^m} |f_{O,g}(x) - \hat{f}_{O,g}(x)| \leq \varepsilon \sum_{i=1}^M \|O_i\|_\infty ,$$

with probability at least $1 - \delta$, *whenever*

$$N = \left(\frac{\gamma}{2}\right)^{-[2(r+r_0+k_0)]^D \ell} \log\left(\frac{M}{\delta}\right) + [2(r+r_0+k_0)]^D \ell \log\left(\frac{2}{\gamma}\right)\left(\frac{\gamma}{2}\right)^{-[2(r+r_0+k_0)]^D \ell}$$

with

$$r = \xi_0 \log\left(\frac{3|S_i|C_t}{\varepsilon}\right),$$
$$\beta = \frac{2}{\lambda_0}\log\left(\frac{3.2^{(2r+k_0)^D}}{\varepsilon}\right),$$
$$\gamma = \frac{\varepsilon\, e^{-[2(r+k_0)]^D(3\log 2+5\beta h)}}{3[2(r+k_0)]^D h\ell}.$$

Proof. We simply need to adapt the proof of Proposition D.1. Using the same notations as there, with probability

$$1-\delta := 1 - Me^{-N(\gamma/2)^{m_r}+m_r\log(2/\gamma)}$$

each cube is visited at least one time. By Lemma E.5, we have that

$$\left|\operatorname{tr}\left[O_i(\psi_g(x)-\sigma(\beta,x|_{\mathcal{S}_i(r)}))\right]\right| \leq \left(C_T e^{-\frac{r}{\xi_0}}|S_i| + 2^{|S_i(r)|}e^{-\frac{\beta\lambda_0}{2}}\right)\|O_i\|_\infty \equiv \widetilde{C}_1(r,\beta)\,\|O_i\|_\infty .$$

Similarly

$$\left|\operatorname{tr}\left[O_i(\psi_g(\hat{Y}_i(x))-\sigma(\beta,\hat{Y}_i(x)|_{\mathcal{S}_i(r)}))\right]\right| \leq \widetilde{C}_1(r,\beta)\,\|O_i\|_\infty .$$

Next, we can reuse the bound on $|f_i(x|_{\mathcal{S}_i(r)})-\hat{f}_{O_i}(x)| = \left|\operatorname{tr}[O_i\sigma(\beta,x|_{\mathcal{S}_i(r)})]-\operatorname{tr}\left[O_i\sigma(\beta,\hat{Y}_i(x)|_{\mathcal{S}_i(r)})\right]\right|$ found in the proof of Proposition D.1. This leads us to the following adaptation of the bound in Equation (D.3):

$$\left|f_{O_i,g}(x)-\hat{f}_{O_i,g}(x)\right| \leq \left(C_T e^{-\frac{r}{\xi_0}}|S_i| + 2^{|S_i(r)|}e^{\frac{-\beta\lambda_0}{2}} + C_2(r)\gamma\right)\|O_i\|_\infty ,$$

where $\hat{f}_{O_i,g}(x) := \operatorname{tr}\left[O_i\psi_g(\hat{Y}_i(x)|_{\mathcal{S}_i(r)})\right]$. We conclude by choosing r so that $C_T e^{-\frac{r}{\xi_0}}|S_i| \leq \frac{\varepsilon}{3}$, β such that $2^{|S_i(r)|}e^{-\frac{\beta\lambda_0}{2}} \leq \frac{\varepsilon}{3}$ and γ such that $C_2(r)\gamma \leq \frac{\varepsilon}{3}$. □

Next, in complete analogy with Proposition D.2, we develop a robust classical shadow tomography algorithm for ground states of quantum systems with LTQO: consider a ground state $\psi_g(x)$ and a family $\psi_g(x_1),\ldots,\psi_g(x_N)$ of groups states with the promise that for any $i\in[M]$ there exist t vectors $x_{i_1},\ldots,x_{i_t}$ such that $\max_{j\in[t]}\|x|_{\mathcal{S}_i(r)}-x_{i_j}|_{\mathcal{S}_i(r)}\|_\infty \leq \gamma$. We run the shadow protocol and construct product operators $\widetilde{\psi}_g(x_1),\ldots,\widetilde{\psi}_g(x_N)$. Then for any ball B of radius k_0, we select the shadows $\widetilde{\psi}_g(x_{i_1}),\ldots\widetilde{\psi}_g(x_{i_t})$ and construct the empirical average

$$\widetilde{\psi}_B(x) := \frac{1}{t}\sum_{j=1}^{t}\operatorname{tr}_{B^c}\left[\widetilde{\psi}_g(x_{i_j})\right].$$

Proposition E.7 (Robust shadow tomography for ground states)**.** *Fix $\varepsilon,\delta\in(0,1)$. In the notations of Proposition E.6, with probability $1-\delta'$, for any ball B of radius k_0, the shadow $\widetilde{\psi}_B$ satisfies $\|\widetilde{\psi}_B - \operatorname{tr}_{B^c}[\psi_g(x)]\|_1 \leq C_T e^{-\frac{r}{\xi_0}}|B| + 2^{|B(r)|}e^{\frac{-\beta\lambda_0}{2}} + C_2(r)\gamma+\varepsilon$ as long as*

$$t \geq \frac{8.12^{k_0}}{3.\varepsilon^2}\log\left(\frac{n^{k_0}2^{k_0+1}}{\delta'}\right). \tag{E.6}$$

Proof. In view of Proposition G.1, it is enough to show that the reduced states $\mathrm{tr}_{B^c}[\psi_g(x_{i_j})]$ are close to $\mathrm{tr}_{B^c}[\psi_g(x)]$. This is done by simply adapting some of the estimates in the proof of Proposition E.7. In particular, we have shown that

$$\| \mathrm{tr}_{B^c} \left[\psi_g(x) - \psi(x_{i_j})\right] \|_1 \leq C_T e^{-\frac{r}{\xi_0}} |B| + 2^{|B(r)|} e^{\frac{-\beta\lambda_0}{2}} + C_2(r)\gamma \,.$$

The result follows directly from Proposition G.1. □

We are now ready to state and prove the main result of this section. We denote $\widetilde{f}_{O_i}(x) = \mathrm{tr}\left[O_i \widetilde{\psi}_{S_i}(x)\right]$ the function constructed from the Gibbs shadow tomography protocol of Proposition D.2, and write $\widetilde{f}_O := \sum_{i=1}^M \widetilde{f}_{O_i}$.

Theorem E.8 (Learning algorithm for ground states)**.** *With the assumptions of Proposition E.6, we fix*

$$r = \xi_0 \log\left(\frac{4|S_i|C_t}{\varepsilon}\right), \quad \beta = \frac{2}{\lambda_0}\log\left(\frac{4.2^{(2r+k_0)^D}}{\varepsilon}\right), \quad \gamma = \frac{\varepsilon\, e^{-[2(r+k_0)]^D(4\log 2 + 5\beta h)}}{3[2(r+k_0)]^D h\ell}.$$

Then, we have that with probability $(1-\delta).(1-\delta')$*,*

$$|f_O(x) - \widetilde{f}_O(x)| \leq \varepsilon \sum_i \|O_i\|_\infty \,,$$

as long as

$$N = t\Big(\frac{\gamma}{2}\Big)^{-[2(r+r_0+k_0)]^D\ell} \log\Big(\frac{M}{\delta}\Big) + t\log\Big[t\Big(\frac{2}{\gamma}\Big)^{[2(r+r_0+k_0)]^D\ell}\Big]\Big(\frac{\gamma}{2}\Big)^{-[2(r+r_0+k_0)]^D\ell}.$$

In other words, once again,

$$N = \Theta\left(\log\Big(\frac{M}{\delta}\Big)\, \log\Big(\frac{n}{\delta'}\Big) e^{\mathrm{polylog}(\varepsilon^{-1})}\right).$$

Proof. Adapting the proof of Proposition D.1, it is clear that with probability

$$1 - \delta := 1 - M e^{-N\frac{1}{t}(\gamma/2)^{m_r} + m_r \log(2/\gamma) + \log t}$$

each cube is visited at least t times. Conditioned on that event, and choosing t such that Equation (E.6) holds, we have that with probability $1-\delta'$

$$|f_{O_i}(x) - \widetilde{f}_{O_i}(x)| \leq \Big(C_T e^{-\frac{r}{\xi_0}} |B| + 2^{|B(r)|} e^{\frac{-\beta\lambda_0}{2}} + C_2(r)\gamma + \varepsilon\Big)\|O_i\|_\infty \,.$$

□

3. Examples

In this section, we gather examples of lattice quantum systems satisfying LTQO. For more details, we refer the interested reader to [NSY22] and references therein.

i. Frustration-free spin chains with a unique translation-invariant matrix product ground state, such as the famous AKLT chain [AKLT88], provide a class of families of ground states satisfying LTQO [CGLW13, OT19, Tas18].

ii. Quantum double models, among which the well-known Toric code introduced by Kitaev [Kit06, Kit03], models of commuting Hamiltonians and were recently shown to satisfy LTQO in [CDH+20].

iii. Levin-Wen models are another class of two-dimensional commuting Hamiltonians studied for their good properties as quantum error correcting codes. Their LTQO property was shown in [QW20].

iv. The stability of LTQO was proved in [NSY22, Theorem 7.2] for a large class of local perturbations under the condition that an unperturbed family of Hamiltonians is uniformly gapped and frustration-free. If the frustration-free condition is removed, stability no longer holds in general, and there exist families of Hamiltonians with uncomputable spectral gaps [CPGW15].

Appendix F: Non-Linear Parameterisations of the Hamiltonian

Here we show that taking a parameterisation of the Hamiltonian that is not a sum of Paulis does not change the results of Appendix D.

Lemma F.1. *Consider a Hamiltonian parameterised in terms of Pauli strings:*

$$H(x) = \sum_j x_j P_j$$

where P_j is a Pauli string. Consider an alternative parameterisation of the same Hamiltonian in terms of the local terms:

$$H(y) = \sum_j h_j(y^{(j)})$$

where $y^{(j)} \in [-1,1]^b$, $b = O(1)$, and h_j only depends on the coordinates in $y^{(j)}$. We will also assume that each h_j is k-local and $\|\partial_u h_j(y)\| \leq 1$.

Then, assuming all the non-zero elements of Jacobian are bounded as $1/C' \leq |\partial y_m/\partial x_k| \leq C$ for $C, C' = O(1)$, the following holds:

$$\left\| \frac{\partial h_j(y^{(j)})}{\partial x_m} \right\| \leq b 4^k C,$$

and

$$|\partial_{y_m} f_L(y)| \leq C'' \max_m |\partial_{x_i} f_L(\beta, x)|,$$

where $C' = O(1)$.

Proof. We see that:

$$\frac{\partial h_j(y^{(j)})}{\partial x_m} = \sum_i \frac{\partial y_i}{\partial x_m} \frac{\partial h_j(y^{(j)})}{\partial y_i}$$

We note that $\frac{\partial h_j(y^{(j)})}{\partial y_i}$ is only non-zero for a b terms. Furthermore, since h_j is k-local, then it can be written as a sum of $\leq 4^k$ Pauli strings. Hence:

$$\begin{aligned}\left\| \frac{\partial h_j(y^{(j)})}{\partial x_m} \right\| &\leq \sum_i \left| \frac{\partial y_i}{\partial x_m} \right| \left\| \frac{\partial h_j(y^{(j)})}{\partial y_i} \right\| \\ &\leq b4^k C \max \left\| \frac{\partial h_j(y^{(j)})}{\partial y_i} \right\| \\ &\leq b4^k C = O(1).\end{aligned}$$

We now consider the functions $f_L(y) = \mathrm{tr}[L\rho(y)]$, $f_L(\beta, x) = \mathrm{tr}[L\rho(x)]$.

$$\partial_{y_m} f_L(y) = \sum_i \frac{\partial x_i}{\partial y_m} \partial_{x_i} f_L(\beta, x)$$

Using that a given y_m can depend on at most 4^k x_m coordinates, we see that for a given y_m, at most $\mathrm{poly}(4^k) = O(1)$ many $\frac{\partial x_i}{\partial y_m}$ can be non-zero. Thus

$$\begin{aligned}|\partial_{y_m} f_L(y)| &\leq \mathrm{poly}(4^k) \left| \frac{\partial x_i}{\partial y_m} \right| |\partial_{x_i} f_L(\beta, x)| \\ &\leq \mathrm{poly}(4^k) C' |\partial_{x_i} f_L(\beta, x)|.\end{aligned}$$

The lemma statement then follows for $C'' = \mathrm{poly}(4^k)C'$. □

This lemma allows us to prove up bounds on the derivative of $f_L(y)$, and thus an equivalent to Lemma C.2 holds for local observables. The rest of the results in appendix D follow similarly.

Appendix G: Shadow tomography for non-identical copies

In this appendix, we extend the shadow tomography protocol to the case of non-identical copies. Consider a state σ and a family $\sigma_1, \ldots, \sigma_N$ of states over n qubits with the promise that for any subset A of qubits of size $|A| \leq r$ there exists a subfamily of states $\sigma_{i_1} \ldots \sigma_{i_t}$, flagged in advance, with the promise that $\max_{j \in [t]} \| \mathrm{tr}_{A^c}(\sigma_{i_j} - \sigma) \|_1 \leq \eta$.

We run the shadow protocol and construct product operators $\widetilde{\sigma}_1, \ldots, \widetilde{\sigma}_N$. Then for any region A, we select the shadows $\widetilde{\sigma}_{i_1}, \ldots \widetilde{\sigma}_{i_t}$ and construct the empirical average

$$\widetilde{\sigma}_A := \frac{1}{t} \sum_{j=1}^{t} \mathrm{tr}_{A^c}(\widetilde{\sigma}_{i_j}) \,.$$

Proposition G.1 (Shadows for non-identical copies)**.** *Fix $\varepsilon, \delta \in (0,1)$. With probability $1-\delta$, the shadow $\widetilde{\sigma}_A$ satisfies $\|\widetilde{\sigma}_A - \sigma_A\|_1 \leq \varepsilon + \eta$ as long as*

$$t \geq \frac{8.12^r}{3.\varepsilon^2} \log\left(\frac{n^r 2^{r+1}}{\delta} \right) .$$

In order to prove the above proposition, we need an extension of the matrix Bernstein inequality used in proving the convergence guarantee of the standard shadow protocol to the case of independent, non-identically distributed random matrices:

Lemma G.2 (Matrix Bernstein for non-i.i.d. random matrices [T+15])**.** *Let $S_1, \dots, S_t$ be independent, centered random matrices with common dimension $d_1 \times d_2$, and assume that each one is uniformly bounded:*

$$\mathbb{E}[S_j] = 0 \qquad \text{and} \qquad \|S_j\|_\infty \leq L \qquad \text{for all } j = 1, \dots, t.$$

Denote the sum $Z = \sum_{j=1}^t S_j$ and let $\nu(Z) := \max\left\{\|\mathbb{E}[ZZ^*]\|_\infty, \|\mathbb{E}[Z^*Z]\|_\infty\right\}$. Then,

$$\mathbb{P}\big(\|Z\|_\infty \geq s\big) \leq (d_1 + d_2) \exp\left(\frac{-s^2/2}{\nu(Z) + Ls/3}\right) \qquad \text{for all } s \geq 0\,.$$

Proof of Proposition G.1. In the notations of the previous paragraph and of Lemma G.2, we take $S_j := \operatorname{tr}_{A^c}(\widetilde{\sigma}_{i_j} - \sigma_{i_j})$, $j = 1 \dots t$, so that $Z/t = \widetilde{\sigma}_A - \mathbb{E}[\widetilde{\sigma}_A]$. Adapting the proof for the standard shadow tomography protocol (see e.g. [HKT+22]), we have

$$\|\operatorname{tr}_{A^c}(\widetilde{\sigma}_{i_j} - \sigma_{i_j})\|_\infty \leq 2^r + 1 \qquad \text{and} \qquad \frac{\nu(Z)}{t} = \frac{1}{t}\Big\|\sum_{j=1}^t \mathbb{E}[S_j^2]\Big\|_\infty \leq 3^r\,.$$

Since $\|X\|_\infty \leq \|X\|_1 \leq 2^r \|X\|_\infty$, we have

$$\mathbb{P}\Big(\|\widetilde{\sigma}_A - \mathbb{E}[\widetilde{\sigma}_A]\|_1 \geq s\Big) \leq 2^{r+1} \exp\left(\frac{-ts^2/2^{2r+1}}{3^r + (2^r+1)s/(3.2^r)}\right) \leq 2^{r+1} \exp\Big(\frac{-3ts^2}{8.12^r}\Big)\,.$$

Next, we observe that under the assumption $\max_{j\in[t]} \|\operatorname{tr}_{A^c}(\sigma_{i_j} - \sigma)\|_1 \leq \eta$, then:

$$\mathbb{E}[\widetilde{\sigma}_A] = \frac{1}{t}\sum_{j=1}^t \operatorname{tr}_{A^c}(\sigma_{i_j}) \qquad \Rightarrow \qquad \|\mathbb{E}[\widetilde{\sigma}_A] - \operatorname{tr}_{A^c}(\sigma)]\|_1 \leq \eta\,.$$

Hence,

$$\mathbb{P}\Big(\|\widetilde{\sigma}_A - \operatorname{tr}_{A^c}(\sigma)\|_1 \geq \eta + \varepsilon\Big) \leq \mathbb{P}\Big(\|\widetilde{\sigma}_A - \mathbb{E}[\widetilde{\sigma}_A]\|_1 \geq s\Big) \leq 2^{r+1} \exp\Big(\frac{-3t\varepsilon^2}{8.12^r}\Big)\,.$$

By union bound, the result follows after choosing $\delta := n^r 2^{r+1} \exp\left(\frac{-3t\varepsilon^2}{8.12^r}\right)$.

□