Understanding the complexity of learning from Gibbs states

Sriram Gopalakrishnan

October 11, 2022

1 Introduction

In a remarkable paper [1], Anurag *et. al.* rigorously proved that the coefficients of very general quantum local Hamiltonians can be learnt to arbitrary precision given polynomially many copies of the system's Gibbs states on an experimental device. In subsequent work by Haah *et. al.* [2], optimal sample and time complexity bounds were obtained in a high-temperature regime. While being of general interest to a broader scientific community, this avenue of research poses interesting problems from complexity-theoretic and physics perspectives [3].

1.1 An important result in Anurag *et. al.*'s work was the strong convexity of quantum log-partition functions. Does the strong convexity (α) of log-partition function have a physical meaning? For commuting Hamiltonians, a simple calculation yields that the largest lower bound on α attainable by any strategy is the minimum variance of the Hamiltonian over a unit sphere in parameter space. The generalization to non-commuting Hamiltonians is complicated, however we take preliminary steps in the right direction.

1.2 In a lengthy derivation, the authors obtained an $O(1/n)$ lower bound on α at finite temperature for non-commuting Hamiltonians. Unfortunately, this bound would vanish for sufficiently large system-size, thereby badly affecting complexity guarantees as well as sensitivity to noise. Is there a larger bound on α, ideally independent of system size (n)? Does this lead to a sample complexity linear in n? How does it improve time complexity and sensitivity to noise?

2. Computing the log-partition function is efficient for Hamiltonians that are sign-problem-free due to a quantum-to-classical mapping. Is there a provably time-efficient and practical learning algorithm for sign-free Hamiltonians?

3. Is there a much better learning algorithm for low-temperature Gibbs states? Is there a phase transition in computational complexity analogous to a divergence in correlation length?

4. Mathematically, the inverse problem considered by Anurag *et. al.* is well-posed in the sense that there's a bijective map between the space of traceless local Hamiltonians and the space of local marginals induced by each of them. In an ill-posed scenario, suppose we have access only to a subset of the local marginals (say in a small sub-region), can the full Hamiltonian still be recovered using sparsity constraints? This question is motivated from a compressed-sensing viewpoint.

We try to tackle some of these questions systematically.

2 Setup and Notation

We are interested in learning from Gibbs states of local Hamiltonians prepared on synthetic many-qubit devices. It is worth noting that although the Gibbs state is a matrix of the same dimension as the Hamiltonian, it has qualitatively different tensor structure than the Hamiltonian itself. There are quantum algorithms capable of preparing the Gibbs state on synthetic devices [4, 5]. Such algorithms rely on the ability to approximate arbitrary smooth functions of a Hermitian matrix by a linear combination of unitaries (LCU) amenable to circuit constructions [6, 7].

Concretely, we are interested in probing copies of Gibbs states of κ-local lattice Hamiltonians H on n-qubits, with $m = O(n)$ local terms that are: mutually orthogonal, traceless, and in general non-commuting (the last criterion being a key hurdle). For simplicity, we assume that local operators have exactly κ-sized support, as opposed to usual studies that allow $\leq \kappa$ support. For convenience of thinking in the energy basis, we assume that the spectrum of H and its local basis are gapped and non-degenerate. Generalizations may be non-trivial, left for future work.

With the above mental picture, we begin our analysis with some basic definitions.

Definition 1 (qubit Hilbert space). *The state space of a single-qubit is a two-dimensional complex euclidean space [c.e.s aka Hilbert space] denoted* $\mathcal{X} = \mathbb{C}^2$.

The state space of n-qubits is $\mathcal{X}^{\otimes n}$. A valid basis for $\mathcal{X}^{\otimes n}$ is $\{|0\rangle, |1\rangle, |2\rangle \ \cdots \ |2^n - 1\rangle\}$ when we allow complex linear combinations. We also call it the *standard basis*.

Definition 2 (graph). *A graph denoted* $G = (V, E)$ *is an indexed set of vertices* $V(G)$*, and a set of edges* $E(G)$*. Each edge is a pair of vertices.*

For n-qubits placed on the vertices of a graph, $|V| = n$. The graph basically restricts the allowed set of gates in a quantum computation. Detail: Our graph is finite, undirected, and loopless. Any finite graph has an embedding in $\mathbb{R}^3$, and we restrict visualization to this basic embedding. Next we define a special graph - the lattice

Definition 3 (lattice). *A lattice denoted* $\Lambda = (V, E)$ *is a sparse connected graph with constant interior-degree. Each interior vertex has a constant neighborhood:* $|E(v)| = c \quad \forall v \in int(V)$*. This means* $|E| = O(n)$ *for an* n*-qubit lattice. We interpret the neighborhood induced by any vertex as being* ***graphically local****.*

Detail: Degree can change at the boundary of Λ - denoted $\partial\Lambda$. Consequently, $\Lambda = \text{int}(\Lambda) \cup \partial\Lambda$. All constant-degree graphs are sparse, but not vice versa.

Observe that our notion of a lattice only requires graphical locality, going beyond its subset - geometric locality. There is nothing "unphysical" about graphical locality in the context of synthetic devices (See for example [8] that realizes long-range-connectivity in a ring-resonator architecture for superconducting qubits). Small changes in euclidean distances can change the interaction strength but **not** the connectivity graph itself. In this lens, the locality of a QC is robust to deformations that preserve the connectivity graph. 2-local or larger measurements are allowed with graphically local qubits even when they are geometrically distant. As such, we will freely use the phrase "spatially contiguous qubits" in the sense of both graphical and geometric locality. We note that Haah *et. al.*'s [2] intro comment about constant-degree expander graphs being low-intersection (graphical) but non-local (geometric) is sensible, however our definition also counts these graphs as lattices, since "distance" is generally well understood purely in terms of number of edges [9].

Perhaps Lieb-Robinson bounds may not hold, but we aren't worried about it at this point.

Definition 4 (κ-local operator (LO)). *A Hermitian matrix O is called a κ-LO if it acts non-trivially on κ spatially contiguous qubits. Its tensor structure is supposed to be understood via the operator space:* $\mathrm{Herm}(\mathcal{X}^{\otimes\kappa}) \otimes \mathbb{1}^{\otimes(n-\kappa)}$. *Its support set has size κ:* $|\mathrm{Supp}(O)| = \kappa$

Each element of $\left(\mathrm{Herm}(\mathcal{X}^{\otimes\kappa}) \otimes \mathbb{1}^{\otimes(n-\kappa)}\right)$ is a κ-LO that acts non-trivially on a chosen subset $V_0 \subset V(\Lambda)$ of κ spatially contiguous qubits. A note on our notation for Identity operators: when we use $\mathbb{1}$ without a subscript, it is a single-qubit Identity operator (in $\mathbb{C}^{2\times 2}$). When we do use a subscript like in $\mathbb{1}_m$, m is the matrix dimension of I. For example, $\mathbb{1} \equiv \mathbb{1}_2$ and $\mathbb{1}^{\otimes 2} \equiv \mathbb{1}_4$. Notation such as $\mathrm{D}(\cdot)$, $\mathrm{Pos}(\cdot)$, $\mathrm{Herm}(\cdot)$ for operator spaces and $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ for Hilbert spaces are adapted from Watrous [10]. $\mathrm{Conv}(\cdot)$ denotes convex hull. $\mathrm{Spec}(\cdot)$ denotes spectrum.

Worth refreshing some facts (good mental exercise, nothing more):

(i) $\mathrm{D}(\mathcal{X}) \subset \mathrm{Pos}(\mathcal{X}) \subset \mathrm{Herm}(\mathcal{X}) \subset \mathrm{Normal}(\mathcal{X})$. Other relevant subspaces of $\mathrm{Normal}(\mathcal{X})$ are $\mathrm{U}(\mathcal{X})$ and $\mathrm{Proj}(\mathcal{X})$. Operator $\widehat{O} \in \mathrm{Normal}(\mathcal{X})$ *iff* it permits a unitary diagonalization.

(ii) If we choose $f(\theta) = e^{i\theta}$ for $\theta \in \mathbb{R}$, then $f(\mathrm{Herm}(\mathcal{X})) = U(\mathcal{X})$.

(iii) $D(\mathcal{X}^{\otimes n})$ is the union of a continuum of standard $(2^n - 1)$-simplices $\mathrm{Conv}(\{|v\rangle \langle v|\})$ where $\{|v\rangle\}$ is a chosen ON basis in $\mathcal{X}$. A continuum of ON bases in $\mathcal{X}$ are generated by unitary transforms on $\{|v\rangle\}$.

(iv) An n-qubit "quantum computation" $U_0 \in U(\mathcal{X}^{\otimes n})$ does two things as a map:

- It leaves the unit-sphere in $\mathcal{X}^{\otimes n}$ invariant while "moving patches on its surface". The "fixed points" on the unit-sphere are eigenstates of U_0.
- It leaves $D(\mathcal{X}^{\otimes n})$ invariant while "moving one simplex to another". There is exactly one "fixed simplex", which is the one spanned by the eigenbasis of U_0.

Definition 5 (κ-local Hamiltonian (LH)). *Given an n-qubit lattice $\Lambda = (V, E)$, a κ-LH on it is a sum of fixed-size LOs supported on different neighborhoods -*

$$H(x) = \sum_{j=1}^{m} x_j L_j \quad \in \mathrm{Herm}(\mathcal{X}^{\otimes n}) \tag{1}$$

where $m = O(n)$, parameters $x = (x_1\ x_2\ \ldots\ x_m)^\top \in \mathbb{R}^m$, and $\{L_j\}$ are basis LOs satisfying some important constraints -

$$\begin{array}{rll}
\kappa\text{-local:} & |\mathrm{Supp}(L_j)| = \kappa & j \in [m] \\
\text{mutually orthogonal:} & \mathrm{Tr}(L_i L_j) = 2^n \delta_{ij} & i, j \in [m] \\
\text{non-commuting in general:} & [L_i, L_j] \neq 0 \quad \text{whenever} \quad \mathrm{Supp}(L_i) \cap \mathrm{Supp}(L_j) \neq \varnothing & i, j \in [m] \\
\text{Hermitian \& traceless:} & L_j^\dagger = L_j \ \ \& \ \ \mathrm{Tr}(L_j) = 0 & j \in [m]
\end{array}$$

All the above properties are characteristic of many local Pauli bases, however in an generalized view, the properties themselves matter more than a specific choice of Pauli basis.

Recall: Any complex Hermitian H is characterized by a spectral decomposition $H = \sum_{j=1}^{2^n} \lambda_j \Pi_j$ where λ_j are *real* and $\Pi_j \in \mathrm{Proj}(\mathcal{X}^{\otimes n})$ are *complex* valued. λ_j can be positive, negative or zero.

Definition 6 (Gibbs state of a LH). *The Gibbs state ($\beta > 0$) of a κ-LH $H(x) \in \mathrm{Herm}(\mathcal{X}^{\otimes n})$ denoted $\rho(x)$ is defined as*

$$\rho(x) = \frac{e^{-\beta H(x)}}{\mathrm{Tr}(e^{-\beta H(x)})} \quad \in \mathrm{D}(\mathcal{X}^{\otimes n}) \tag{2}$$

Henceforth $\rho(\cdot)$ is used exclusively for Gibbs states. Interpretations:

- In the energy basis, the Gibbs state represents a Boltzmann distribution over the eigenstates of H familiar from stat-mech. $Z(x) = \mathrm{Tr}(e^{-\beta H(x)})$ is referred to as partition function.
- We have $Z(x) > 0$ always. If $H(x)$ has even one $\lambda \leq 0$, then $Z(x) \geq 1$. Hence, low-energy states ramp-up $Z(x)$, while high-energy states ramp-down $Z(x)$.
- Gibbs states are not just positive semi-definite, but positive definite. Hence they have a full Image and empty Kernel: $\mathrm{Im}(\rho(x)) = \mathcal{X}^{\otimes n}$ while $\mathrm{Ker}(\rho(x)) = \varnothing$.

However, note that $e^{-\beta\lambda_H}$ can be arbitrarily small under appropriate limits. Suppose we define an "ϵ-approximate Kernel" of $\rho \in \mathrm{D}(\mathcal{X}^{\otimes n})$ as follows:

$$\mathrm{Ker}_\epsilon(\rho) = \{|v\rangle \in \mathcal{X}^{\otimes n} : \|\rho\,|v\rangle\|_2 \leq \epsilon\} \tag{3}$$

Then even for exponentially small ϵ, $\mathrm{Ker}_\epsilon(\rho(x))$ is non-empty for some choices of $H(x)$. For our problem, it is desirable that Gibbs states have *empty approximate Kernels* as well. Non-empty approximate Kernels of $\rho(x)$ are likely when $\lambda_H \gg k_B T$, that is when $H(x)$ has high-energy states that far surpass the thermal energy scale. Hence we desire that $\mathbf{T} > \lambda^*/\mathbf{k_B}$ [where $\lambda^* = \max_{\mathbf{x}} \|\mathbf{H}(\mathbf{x})\|_\infty$] in an experiment for accurate Hamiltonian learning. Full-rank Gibbs states $\rho(x)$ with empty approximate Kernels favor lower sensitivity to noise in learning from their marginals.

We propose a distinction between *stable recovery* and *unique recovery* in the learning problem. One does not guarantee the other. Unique recovery is self-explanatory: the inverse map must be injective. Stable recovery means: reconstructions are robust against small perturbations/noise in their pre-image. In QHLP, this means noise in local marginals must not get amplified in what we learn: LH parameters. It appears that very-low-temperature states forbid stable recovery, while very-high-temperature states forbid unique recovery. We will see this heuristically in the proof of Lemma 8 and rigorously in Lemma 14

Problem (Quantum Hamiltonian Learning Problem (QHLP)). *Consider characters Alice and Bob. Given an n-qubit lattice $\Lambda = (V, E)$, Alice chooses a κ-LH (strictly obeying Definition 5)*

$$H(\mu) = \sum_{j=1}^{m} \mu_j L_j \qquad \mu \in \mathbb{R}^m$$

with basis $\{L_j\}$ known to Bob but coefficients $\{\mu_j\}$ unknown to him. She tells him: I can give you polynomially many copies of $H(\mu)$'s Gibbs state $\rho(\mu)$ on a QC, can you learn $\{\mu_j\}$ to ϵ-precision in 2-norm? What is your sample and time complexity? Formally, we require Bob to learn an estimate $\widehat{\mu}$ of μ such that $\|\widehat{\mu} - \mu\|_2 \leq \epsilon$, for arbitrary $\epsilon > 0$ provided by Alice.

Remarkably, Anurag *et. al.* provided a rigorous solution to QHLP with a polynomial sample complexity [1]. As we discussed earlier, they did not investigate time complexity, however, only in a high-temperature regime, sample and time optimal bounds utilizing cluster expansions are now known [2].

3 Existence of bijective map

The foundational principle underlying QHLP is the existence of a bijective map between the space of LHs and the space of local marginals induced by their Gibbs states. We define local marginals in Definition 7.

That such a map exists for quantum LHs should point to a fundamental info-theoretic property, a possible manifestation of short-range-entanglement (SRE). I believe SRE of quantum Gibbs states is in direct analogy to conditional-independence (CI) of **classical** Gibbs distributions on graphs. CI in any multivariate probability distribution manifests as a simple yet fundamental info-theoretic property: $I(A:C|B)=0$ (zero CMI) for all tri-partitions (A,B,C) of Λ that have **no direct edges** from A to C. However, I'm not aware of an equivalent info-theoretic manifestation of SRE, since $I(A:C|B)\neq 0$ (non-zero QCMI) in general for similar tri-partitions of Λ for quantum Gibbs states. Intuitively, zero QCMI means that for every pair of vertices $\{v_1,v_2\}$ with no direct edge, the reduced density matrix $\rho_{v_1v_2}$ is separable. Even without zero QCMI, quantum Gibbs states are locally reconstructible.

Definition 7 (κ-Local Marginal (LM)). *For each local basis operator L_j, there is an associated LM $l_j(\mu)\in\mathbb{R}$, given by*

$$l_j(\mu)=\mathrm{Tr}(\rho(\mu)L_j)\equiv\langle L_j\rangle_\mu \tag{4}$$

These LMs can be stacked into a column vector like: $l(\mu)=(l_1(\mu)\ l_2(\mu)\ \ldots l_m(\mu))^T\in\mathbb{R}^m$.

Lemma 8 (Bijective map from LH to LMs). *The non-linear mapping $\mathcal{T}:\mathbb{R}^m\to\mathbb{R}^m$ that takes LH parameters μ to LMs $l(\mu)$ is bijective. Notation: $\mathcal{T}[\mu]=l(\mu)$, and $\mathcal{T}^{-1}[l(\mu)]=\mu$*

Proof. Since the domain and co-domain of $\mathcal{T}$ are the same set, if $\mathcal{T}$ is injective, it is also surjective. Hence, it suffices to prove that $\mathcal{T}$ is injective. Injectivity requires us to prove: For $\mu,\lambda\in\mathbb{R}^m$, $\mu\neq\lambda\Rightarrow l(\mu)\neq l(\lambda)$. Equivalently, we require: $l(\mu)=l(\lambda)\Rightarrow\mu=\lambda$ ("consistent marginals implies identical LHs")

Claim 8.1: The von-Neumann entropy of a Gibbs state $\rho(\lambda)$ is given by

$$S(\rho(\lambda))=\beta\langle H(\lambda)\rangle_\lambda+\log Z(\lambda) \tag{5}$$

Notice how it is consistent with the thermodynamic relation $F=E-TS$ if $\log Z=-\beta F$.

Claim 8.2: The relative entropy of $\rho(\mu)$ w.r.t. $\rho(\lambda)$ is given by

$$S(\rho(\mu)\|\rho(\lambda))=\beta[\langle H(\lambda)\rangle_\mu-\langle H(\mu)\rangle_\mu]+[\log Z(\lambda)-\log Z(\mu)] \tag{6}$$

Both the above claims follow straightforwardly from the definition of quantum entropies applied to the Gibbs state. Now consider the following chain of thought

$$\begin{aligned}
l(\mu)=l(\lambda)&\Longrightarrow\langle L_j\rangle_\mu=\langle L_j\rangle_\lambda\quad\forall j\in[m] && \text{(elementwise comparison)}\\
&\Longrightarrow\langle H(x)\rangle_\mu=\langle H(x)\rangle_\lambda\quad\forall x\in\mathbb{R}^m && \text{(linearity of expectation)}\\
&\Longrightarrow\langle H(\lambda)\rangle_\mu=\langle H(\lambda)\rangle_\lambda && \text{(choosing } x=\lambda)\\
&\Longrightarrow S(\rho(\mu)\|\rho(\lambda))=S(\rho(\lambda))-S(\rho(\mu)) && \text{(substituting into eq 6)}\\
&\Longrightarrow S(\rho(\lambda))\geq S(\rho(\mu)) && \text{(non-negativity of relative entropy)}
\end{aligned}$$

Recall that the relative entropy is an asymmetric yet non-negative function. Had we swapped the arguments of $S(\cdot\|\cdot)$ in the same chain of thought, we would get $S(\rho(\mu))\geq S(\rho(\lambda))$. Hence

$S(\rho(\mu)) = S(\rho(\lambda))$, implies the relative entropy of two Gibbs states with consistent marginals is exactly zero, implies the Gibbs states would have to be identical to begin with:

$$l(\mu) = l(\lambda) \Rightarrow S(\rho(\mu)) = S(\rho(\lambda)) \Rightarrow S(\rho(\mu)\|\rho(\lambda)) = 0 \Rightarrow \rho(\mu) = \rho(\lambda) \tag{7}$$

The subtlety in this argument is that $S(\rho\|\sigma) = \mathrm{Tr}(\rho\log\rho - \rho\log\sigma)$ holds only when $\mathrm{Im}(\rho) \subseteq \mathrm{Im}(\sigma)$ by definition. Our ability to swap ρ and σ and still use the same formula is strictly contingent on $\mathrm{Im}(\rho) = \mathrm{Im}(\sigma)$. Hence, $\rho(\mu)$ **and** $\rho(\lambda)$ **must have identical Kernel and Image spaces** for equation 7 to be considered valid.

Lucky for us, the *exact* Kernel of all finite-temp Gibbs states are empty: $\mathrm{Ker}(\rho(x)) = \varnothing \quad \forall x \in \mathbb{R}^m$. But we have also seen in the last section that *approximate* Kernels may be non-empty if $H(x)$ has high-energy states that surpass the thermal energy scale k_BT. We can preclude this possibility by assuming a sufficiently large temperature $\mathbf{T} > \lambda^*/\mathbf{k_B}$ where $\lambda^* = \max_x \lambda_{max}(H(x))$

Aside: We expect sensitivity $\frac{\|\rho(\mu)-\rho(\lambda)\|_2}{\|l(\mu)-l(\lambda)\|_2}$ to be an increasing function of the constant $\beta\lambda^*$.

Given that, we now have to prove: $\rho(\mu) = \rho(\lambda) \Rightarrow \mu = \lambda$.

$$\begin{aligned}
\rho(\mu) = \rho(\lambda) &\Longrightarrow \log\rho(\mu) = \log\rho(\lambda) \\
&\Longrightarrow \beta H(\mu) + \log Z(\mu)\mathbb{1}^{\otimes n} = \beta H(\lambda) + \log Z(\lambda)\mathbb{1}^{\otimes n} \\
&\Longrightarrow \log\left(\frac{Z(\mu)}{Z(\lambda)}\right)\mathbb{1}^{\otimes n} = \beta\sum_{j=1}^{m}(\mu_j - \lambda_j)L_j \\
&\Longrightarrow \log\left(\frac{Z(\mu)}{Z(\lambda)}\right)\mathrm{Tr}(L_k) = \beta(\mu_k - \lambda_k)2^n \quad \forall k \in [m] \qquad \text{(Ortho: } \mathrm{Tr}(L_jL_k) = 2^n\delta_{jk}\text{)} \\
&\Longrightarrow \mu = \lambda \quad \text{for } \beta > 0 \qquad\qquad\qquad\qquad\qquad\qquad \text{(Traceless: } \mathrm{Tr}(L_k) = 0\text{)}
\end{aligned}$$

Hence, fixing basis $\{L_j\}$, each finite-temperature Gibbs state is generated by a unique LH

- When $\beta \to 0$ ($T \to \infty$), $\mu \neq \lambda$ is allowed $\Rightarrow$ unique recovery forbidden. This conforms to the statement: *nothing is learnable from a maximally mixed state.*
- When $\beta \to \infty$ ($T \to 0$), $\mu = \lambda$ if basis $\{L_j\}$ is traceless and $\log(Z(\mu)/Z(\lambda))$ scales as $O(\beta)$. Since $\frac{\lambda_{max}}{k_BT} \to \infty$, the ground state is extremely sensitive to noise in local marginals, and so are the LH parameters $\Rightarrow$ stable recovery forbidden. This conforms to the statement: *learning from ground states is highly sensitive to noise*

We see that $\mathcal{T}$ is certainly injective at finite-temperature. As explained in the start, this also means $\mathcal{T}$ is bijective. This concludes our proof. □

PS: The existence of a bijective map is solely concerned with *uniqueness*. As discussed, it does not by itself account for the lack of *stability* at low temperatures.

4 Explicit form of inverse map & sensitivity to noise

Thus far, we have shown the existence of a bijective non-linear map $\mathcal{T}$ from κ-LHs to the κ-LMs induced by their Gibbs states. We also have an explicit form for the forward map as computing local expectation values. We would now like to present an explicit form for the inverse map $\mathcal{T}^{-1}(\cdot)$ going from κ-LMs back to κ-LH parameters, as well as understand the sensitivity of $\mathcal{T}^{-1}$ to noise

in the marginals, which is understood via a bound on the condition number of $\mathcal{T}$. We present a small factor-of-2 improvement over the condition number obtained by Anurag *et. al.*

Let us recall what we know. We have a synthetic many-qubit device on which we prepare copies of the Gibbs states $\rho(\mu)$ of a chosen local Hamiltonian $H(\mu)$. We then make local measurements on a certain number of copies (sample complexity) of $\rho(\mu)$ to obtain the full set of local marginals denoted $l(\mu) = (l_1(\mu) \quad l_2(\mu) \dots l_m(\mu))^\top$. The "inputs" to the inverse map are the local marginals $l(\mu) \in \mathbb{R}^m$, and the "outputs" of the inverse map are the LH parameters $\mu \in \mathbb{R}^m$. Without loss of generality, we can assume that μ lies in a 2-norm ball $\|x\|_2 \leq 1$ to reduce our search space.

Lemma 9 (Entropy-Free Energy duality). *Consider a primal optimization task as follows:*

$$\begin{aligned} p^* = \max_{\sigma \in \mathrm{Pos}(\mathcal{X}^{\otimes n})} \quad & S(\sigma) = -\operatorname{Tr}(\sigma \log \sigma) \\ s.t. \quad & \operatorname{Tr}(\sigma L_j) = l_j(\mu) \quad \forall j \in [m] \\ & \operatorname{Tr}(\sigma) = 1 \end{aligned}$$

Its Lagrange dual is the following minimization

$$\begin{aligned} d^* = \min_{x \in \mathbb{R}^m} \quad & g(x) = \log Z(x) + \beta x^\top l(\mu) \\ s.t. \quad & \|x\|_2 \leq 1 \end{aligned}$$

For fixed temperature $\beta > 0$, the unique solution to the above problem is: $(\sigma^*, x^*) = (\rho(\mu), \mu)$. Additionally, strong duality holds: $p^* = d^* = S(\rho(\mu))$

Proof. For clarity, view the primal objective as minimizing negative-entropy $\operatorname{Tr}(\sigma \log \sigma)$, which is convex. We then construct a Lagrangian and find its pointwise infimum to be $-g(x)$, which should be concave [pointwise infimum of affine functions]. Maximizing $-g(x)$ then is same as minimizing $g(x)$, a convex function. Also note that the primal form only has equality constraints, so dual variables are unconstrained (except the artificial $\|x\|_2 \leq 1$)

One example of a state that qualifies the marginal constraints $\operatorname{Tr}(\sigma L_j) = l_j(\mu) \quad \forall j \in [m]$ while maximizing entropy is the Gibbs state $\rho(\mu) = e^{-\beta H(\mu)}/Z(\mu)$. This makes the primal form strictly feasible, hence strong duality holds [Slater condition].

The primal feasible set has many states, however *the only Gibbs state in the primal feasible set is $\rho(\mu)$*. This follows from Lemma 8, where we proved a bijection between Gibbs states and their LMs. Since maximum-entropy states with linear constraints are Gibbs, we expect that $\rho(\mu)$ is in fact the unique primal solution. This intuition will be made rigorous.

Use dual variables $\beta x_j \in \mathbb{R}$ for marginal constraints and $\nu \in \mathbb{R}$ for the trace constraint to construct the following Lagrangian

$$\mathcal{L}(\sigma, x, \nu) = \operatorname{Tr}(\sigma \log \sigma) + \beta \left[\operatorname{Tr}(\sigma H(x)) - x^\top l(\mu)\right] + \nu \left[\operatorname{Tr}(\sigma) - 1\right] \tag{8}$$

The domain of $\mathcal{L}$ is: $\sigma \in \mathrm{Pos}(\mathcal{X}^{\otimes n})$, $x \in \mathbb{R}^m_+$, $\nu \in \mathbb{R}_+$. By the KKT theorem, σ^* is the primal solution iff (σ^*, x^*, ν^*) is a saddle point of $\mathcal{L}(\cdot)$. The necessary (and often sufficient) conditions for the same are:

(i) Stationarity w.r.t. σ: $\nabla_\sigma \mathcal{L}(\sigma^*, x^*, \nu^*) = 0$

(ii) Primal feasibility: $\operatorname{Tr}(\sigma^* L_j) = l_j(\mu) \quad \forall j \in [m]$, and $\operatorname{Tr}(\sigma^*) = 1$

(iii) Dual feasibility: $x^* \in \mathbb{R}^m$ and $\nu^* \in \mathbb{R}$

(iv) Complementary slackness: Void for all-equality-constraint primal form

Dual feasibility and complementary slackness are trivial in our problem.

The stationarity condition yields the following

$$\nabla_\sigma \mathcal{L}(\sigma^*, x^*, \nu^*) = 0 \implies \log \sigma^* + \beta H(x^*) + (\nu^* + 1)\mathbb{1}_{2^n} = 0 \tag{9}$$

$$\implies \sigma^* = \frac{e^{-\beta H(x^*)}}{e^{(\nu^*+1)}} = \rho(x^*) \tag{10}$$

We see that σ^* is indeed a Gibbs state. We have seen that $\rho(\mu)$ is the unique Gibbs state in the primal feasible set. Hence

$$(\sigma^*, x^*, \nu^*) = (\rho(\mu),\ \mu,\ \log Z(\mu) - 1) \tag{11}$$

Hence σ^* is primal feasible. The more general stationarity condition away from the saddle point: $\nabla_\sigma \mathcal{L}(\sigma^*, x, \nu) = 0$ yields the solution

$$\sigma^*(x, \nu) = \frac{e^{-\beta H(x)}}{e^{(\nu+1)}} = \rho(x) \tag{12}$$

Hence the dual objective for minimization is given by

$$\begin{aligned} g(x) = -\inf_\sigma \mathcal{L}(\sigma, x, \nu) &= -\mathcal{L}(\sigma^*(x,\nu), x, \nu) \\ &= S(\rho(x)) - \beta\left[\langle H(x)\rangle_x - x^\top l(\mu)\right] \\ &= \cancel{\beta\langle H(x)\rangle_x} + \log Z(x) - \cancel{\beta\langle H(x)\rangle_x} + \beta x^\top l(\mu) \\ &= \log Z(x) + \beta x^\top l(\mu) \end{aligned}$$

To verify strong duality, note that

$$\begin{aligned} d^* = g(\mu) &= \log Z(\mu) + \beta \mu^\top l(\mu) \\ &= \log Z(\mu) + \beta\langle H(\mu)\rangle_\mu && (\langle H(\mu)\rangle_\mu = \mu^\top l(\mu) \text{ using Definition 7}) \\ &= S(\rho(\mu)) = p^* && (\text{using Claim 8.1}) \end{aligned}$$

This concludes our proof □

Corollary 10 (Inverse map). *As a result of Lemma 9, we have*

$$\begin{aligned} \mu = \mathcal{T}^{-1}(l(\mu)) = \arg\min_{x \in \mathbb{R}^m} \quad & \left[\log Z(x) + \beta x^\top l(\mu)\right] \\ s.t. \quad & \|x\|_2 \le 1 \end{aligned}$$

Provided we choose a μ inside the Unit 2-norm Ball (U2B): $\|x\|_2 \le 1$ in $\mathbb{R}^m$, the above inverse map gives a clear prescription for learning μ: Measure LMs $l(\mu)$, and minimize $\left[\log Z(x) + \beta x^\top l(\mu)\right]$ numerically inside U2B. So far, we proved the existence of a bijection between LHs and LMs in Lemma 8. The forward map $\mathcal{T}$ was trivially known by Definition 7. The inverse map $\mathcal{T}^{-1}$ is now known due to Lemma 9 and Corollary 10. Before worrying about sample/time complexity, the last piece in this analysis is understanding the sensitivity of $\mathcal{T}^{-1}$ to noise in LMs.

We will see that the sensitivity of $\mathcal{T}^{-1}$ to noise in LMs depends exactly on two factors:

– Temperature: We observed heuristically in Lemma 8 that *lower the temperature, larger the sensitivity to noise in learning*. We will confirm this rigorously.

– Curvature in $\log Z(x)$: The curvature in $\log Z(x)$ is a positive constant α that depends only on temperature and the local basis $\{L_j\}$ chosen in Definition 5. The *lower this curvature, larger the sensitivity to noise in learning*.

Fix inverse temperature β. Then for commuting LHs, we will see that the curvature in $\log Z(x)$ is proportional to a Minimum-Energy-Variance (MEV) over U2B Gibbs states. Also recall from our discussion about empty-approximate-Kernels the quantity $\lambda^* = \max_{x \in \text{U2B}} \|H(x)\|_\infty$. It would be an interesting exercise to study the relationships between α, MEV, and λ^* numerically. Theoretically, for commuting LHs, we expect

$$\alpha \propto \text{MEV} \propto \frac{1}{\lambda^*} \tag{13}$$

The sensitivity to noise of $\mathcal{T}^{-1}$ is defined by a bound on its condition number. Central to bounding the condition number of $\mathcal{T}^{-1}$ is a sharp understanding of strongly convex multivariate functions, in our case applied to $\log Z(x)$. For this reason, it is worth taking a small interlude into some general results from convex optimization to guide our thinking

Definition 11 (Convexity characterizations). *A twice-differentiable function $f : \mathbb{R}^m \to \mathbb{R}$ is convex in $\mathcal{D} \subset \mathbb{R}^m$ under the following equivalent criteria -*

(i) $f(y) \geq f(x) + \nabla f(x)^\top (y - x) \quad \forall x, y \in \mathcal{D}$ (linear lower bound)

(ii) $(\nabla f(x) - \nabla f(y))^\top (x - y) \geq 0 \quad \forall x, y \in \mathcal{D}$ (gradient monotone)

(iii) $\nabla^2 f(x) \succeq 0 \quad \forall x \in \mathcal{D}$ (positive semidefinite Hessian)

Definition 12 (Strong convexity characterizations). *A twice-differentiable function $f : \mathbb{R}^m \to \mathbb{R}$ is said to be α-strongly convex ($\alpha > 0$) in $\mathcal{D} \subset \mathbb{R}^m$ under the following equivalence -*

(i) $g(x) = f(x) - \frac{\alpha}{2}\|x\|_2^2$ is convex in $\mathcal{D}$

(ii) $f(y) \geq f(x) + \nabla f(x)^\top (y - x) + \frac{\alpha}{2}\|x - y\|_2^2 \quad \forall x, y \in \mathcal{D}$ (quadratic lower bound)

(iii) $(\nabla f(x) - \nabla f(y))^\top (x - y) \geq \alpha \|x - y\|_2^2 \quad \forall x, y \in \mathcal{D}$ (strong gradient monotone)

(iv) $\nabla^2 f(x) \succeq \alpha \mathbb{1}_m \quad \forall x \in \mathcal{D}$ (positive definite Hessian)

To a reader familiar with convex optimization, these are fairly standard results [11, 12]. A somewhat non-trivial characterization is the gradient monotone. It follows by simply swapping x and y in the linear lower bound and adding the two inequalities. A rough geometric interpretation is - if the input to f is changed along some direction, the gradient of f should also change in the same/similar direction. The strong gradient monotone in Definition 12 is understood similarly.

Lemma 13 (Variant of strong gradient monotone). *Let $f : \mathbb{R}^m \to \mathbb{R}$ be an α-strongly convex function ($\alpha > 0$). Then the following is true -*

$$\|x - y\|_2 \leq \frac{1}{\alpha} \|\nabla f(x) - \nabla f(y)\|_2 \quad \forall x, y \in \mathcal{D} \tag{14}$$

Proof. Write the strong gradient monotone condition in reverse

$$0 \leq \alpha \|x - y\|_2^2 \leq (\nabla f(x) - \nabla f(y))^\top (x - y)$$

Notice that the RHS is a non-negative inner product, can be upper bounded by Cauchy-Schwarz

$$(\nabla f(x) - \nabla f(y))^\top (x - y) \leq \|\nabla f(x) - \nabla f(y)\|_2 \|x - y\|_2$$

Hence the result follows

$$\begin{aligned} & \alpha \|x - y\|_2^2 \leq \|\nabla f(x) - \nabla f(y)\|_2 \|x - y\|_2 \\ \implies \quad & \|x - y\|_2 \leq \frac{1}{\alpha} \|\nabla f(x) - \nabla f(y)\|_2 \quad \forall x, y \in \mathcal{D} \quad \square \end{aligned}$$

We are now prepared to present the sensitivity to noise bound for QHLP. Recall that the inverse map $\mathcal{T}^{-1}$ takes LMs $l(\mu) \in \mathbb{R}^m$ back to LH parameters $\mu \in \mathbb{R}^m$

Lemma 14 (Bound on condition number of $\mathcal{T}^{-1}$). *Let $\mathcal{D} \subset \mathbb{R}^m$ and $\nabla^2 \log Z(x) \geq \alpha \mathbb{1}_m \quad \forall x \in \mathcal{D}$. For any pair of LH parameters $\mu, \lambda \in \mathcal{D}$, let $l(\mu), l(\lambda) \in \mathbb{R}^m$ be the corresponding local marginals. Then we have -*

$$\|\mu - \lambda\|_2 \leq \frac{\beta}{\alpha} \|l(\mu) - l(\lambda)\|_2 \tag{15}$$

Proof. The proof follows by applying Lemma 13 to $f(x) = \log Z(x)$ evaluated at μ, λ.

We first note that since μ, λ minimize inverse problem objectives as described by Corollary 10, they obey first derivative conditions -

$$\nabla \log Z(\mu) = -\beta l(\mu) \qquad \nabla \log Z(\lambda) = -\beta l(\lambda) \tag{16}$$

We then have

$$\|\mu - \lambda\|_2 \leq \frac{1}{\alpha} \|\log Z(\mu) - \log Z(\lambda)\|_2 = \frac{\beta}{\alpha} \|l(\mu) - l(\lambda)\|_2 \quad \square$$

Denote the absolute condition number of $\mathcal{T}^{-1}$ as $\mathrm{CoNum}(\mathcal{T}^{-1})$. For each $l(\mu) \in \mathbb{R}^m$, consider a small variation $\delta l(\mu)$ about $l(\mu)$, which results in a variation $\delta\mu$ about μ. It follows that

$$\mathrm{CoNum}(\mathcal{T}^{-1}) = \lim_{\Delta \to 0} \sup_{\|\delta l(\mu)\|_2 \leq \Delta} \frac{\|\delta\mu\|_2}{\|\delta l(\mu)\|_2} \leq \frac{\beta}{\alpha} \tag{17}$$

Anurag *et. al.* [1] obtained a sensitivity bound of $\frac{2\beta}{\alpha}$. In the above simpler proof, we see that the sensitivity bound can be improved by a factor-of-2 to $\frac{\beta}{\alpha}$

Whether this factor-of-2 improvement in the sensitivity bound changes sample or time complexity meaningfully is not clear yet (working on it).

However, we can now rigorously see the distinction between stable recovery and unique recovery we claimed earlier.

References

[1] Anurag Anshu et al. "Sample-efficient learning of quantum many-body systems". In: *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2020, pp. 685–691.

[2] Jeongwan Haah, Robin Kothari, and Ewin Tang. "Optimal learning of quantum Hamiltonians from high-temperature Gibbs states". In: *arXiv preprint arXiv:2108.04842* (2021).

[3] Anurag Anshu. "Some Recent Progress in Learning Theory: The Quantum Side". In: *Harvard Data Science Review* 4.1 (Jan. 27, 2022). https://hdsr.mitpress.mit.edu/pub/3x2sd8nq. DOI: 10.1162/99608f92.9c47e402. URL: https://hdsr.mitpress.mit.edu/pub/3x2sd8nq.

[4] Fernando GSL Brandão et al. "Quantum SDP solvers: Large speed-ups, optimality, and applications to quantum learning". In: *arXiv preprint arXiv:1710.02581* (2017).

[5] Anirban Narayan Chowdhury and Rolando D Somma. "Quantum algorithms for Gibbs sampling and hitting-time estimation". In: *arXiv preprint arXiv:1603.02940* (2016).

[6] Sathyawageeswar Subramanian, Stephen Brierley, and Richard Jozsa. "Implementing smooth functions of a Hermitian matrix on a quantum computer". In: *Journal of Physics Communications* 3.6 (2019), p. 065002.

[7] Andrew M Childs and Nathan Wiebe. "Hamiltonian simulation using linear combinations of unitary operations". In: *arXiv preprint arXiv:1202.5822* (2012).

[8] Sumeru Hazra et al. "Ring-Resonator-Based Coupling Architecture for Enhanced Connectivity in a Superconducting Multiqubit Network". In: *Physical Review Applied* 16.2 (2021), p. 024018.

[9] Matthew B Hastings. "Locality in quantum systems". In: *Quantum Theory from Small to Large Scales* 95 (2010), pp. 171–212.

[10] John Watrous. *The theory of quantum information*. Cambridge university press, 2018.

[11] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[12] Xingyu Zhou. "On the fenchel duality between strong convexity and lipschitz continuous gradient". In: *arXiv preprint arXiv:1803.06573* (2018).

[13] Isaac H Kim. "Entropy scaling law and the quantum marginal problem". In: *Physical Review X* 11.2 (2021), p. 021039.

[14] Xiao-Liang Qi and Daniel Ranard. "Determining a local Hamiltonian from a single eigenstate". In: *Quantum* 3 (2019), p. 159.

[15] Matthew B Hastings. "Quantum belief propagation: An algorithm for thermal quantum systems". In: *Physical Review B* 76.20 (2007), p. 201102.

[16] Sergey Bravyi et al. "On the complexity of quantum partition functions". In: *arXiv preprint arXiv:2110.15466* (2021).

[17] Álvaro M Alhambra. "Quantum many-body systems in thermal equilibrium". In: *arXiv preprint arXiv:2204.08349* (2022).

[18] 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*. 2020, pp. 378–386.