Optimal learning of quantum Hamiltonians from high-temperature Gibbs states

Jeongwan Haah[1] Robin Kothari[1] **Ewin Tang**[2]

[1]Microsoft Quantum

[2]University of Washington

July 12, 2021

The problem: Hamiltonian learning

Let H be an N-qubit ℓ-local Hamiltonian in dimension d, with $d, \ell = O(1)$:[1]

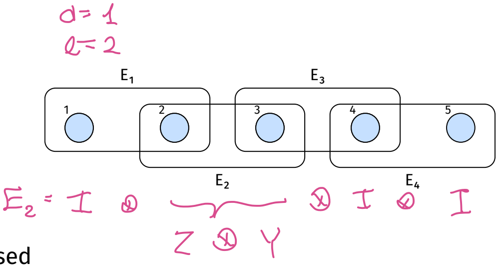
$$H = \sum_{a=1}^{M} \lambda_a E_a \qquad \text{for} \qquad \begin{array}{l} \lambda_a \in [-1, 1] \\ E_a \in \mathbb{C}^{2^N \times 2^N} \text{ such that } |\mathrm{Supp}\, E_a| \leq \ell \text{ and } \|E_a\| \leq 1 \end{array}$$

We further assume that the E_a's are distinct products of Paulis. For a known inverse temperature β, suppose we can prepare copies of the *Gibbs state*

$$\rho = \rho(\lambda) = \frac{\exp(-\beta H)}{\mathrm{Tr}\exp(-\beta H)}.$$

$\beta \to 0 \rightarrow \frac{I}{2^N}$

$\beta \to \infty \rightarrow uu^*$



The goal is to **learn the coefficients** λ_a to ε error.

1. *sample complexity*: number of copies of ρ used
2. *time complexity*: time to compute the λ_a's

We expect hardness as $\beta \to 0$ and $\beta \to \infty$.

---

[1] This implies that $M = O(N)$.

The current state of things

	sample lower bound	sample upper bound	time upper bound
classical[2]	$\frac{e^{\Omega(\beta)}}{\beta^2\varepsilon^2}\log N$ *	$\frac{e^{O(\beta)}}{\beta^2\varepsilon^2}\log N$	$\frac{e^{O(\beta)}}{\beta^2\varepsilon^2}N\log N$
quantum[3]	$\frac{1}{\beta\varepsilon}$	$\frac{e^{O(\beta^{4+c})}}{\beta^{4+c'}\varepsilon^2}N^2\log N$	

[2]Results shown for the Ising model; sample complexity results assume the terms are also unknown [Santhanam and Wainwright 2012; Vuffray, Misra, Lokhov, and Chertkov 2016]

[3]Anshu, Arunachalam, Kuwahara, and Soleimanifar 2021

The current state of things

	sample lower bound	sample upper bound	time upper bound
classical[2]	$\frac{e^{\Omega(\beta)}}{\beta^2\varepsilon^2}\log N$ *	$\frac{e^{O(\beta)}}{\beta^2\varepsilon^2}\log N$	$\frac{e^{O(\beta)}}{\beta^2\varepsilon^2}N\log N$
quantum[3]	$\frac{1}{\beta\varepsilon}$	$\frac{e^{O(\beta^{4+c})}}{\beta^{4+c'}\varepsilon^2}N^2\log N$	$\frac{1}{\beta^{4+c'}\varepsilon^2}N^{2+c''}$

Blue indicates the assumption that $\beta < \beta_c = \Theta(1)$

[2]Results shown for the Ising model; sample complexity results assume the terms are also unknown [Santhanam and Wainwright 2012; Vuffray, Misra, Lokhov, and Chertkov 2016]

[3]Anshu, Arunachalam, Kuwahara, and Soleimanifar 2021

The current state of things

	sample lower bound	sample upper bound	time upper bound
classical[2]	$\frac{e^{\Omega(\beta)}}{\beta^2\varepsilon^2}\log N$ *	$\frac{e^{O(\beta)}}{\beta^2\varepsilon^2}\log N$	$\frac{e^{O(\beta)}}{\beta^2\varepsilon^2}N\log N$
quantum[3]	$\frac{1}{\beta\varepsilon}$	$\frac{e^{O(\beta^{4+c})}}{\beta^{4+c'}\varepsilon^2}N^2\log N$	$\frac{1}{\beta^{4+c'}\varepsilon^2}N^{2+c''}$
our work	$\frac{e^{2\beta}}{\beta^2\varepsilon^2}\log N$	$\frac{1}{\beta^2\varepsilon^2}\log N$	$\frac{1}{\beta^2\varepsilon^2}N\log N$

Blue indicates the assumption that $\beta < \beta_c = \Theta(1)$

[2] Results shown for the Ising model; sample complexity results assume the terms are also unknown [Santhanam and Wainwright 2012; Vuffray, Misra, Lokhov, and Chertkov 2016]

[3] Anshu, Arunachalam, Kuwahara, and Soleimanifar 2021

The classical setting

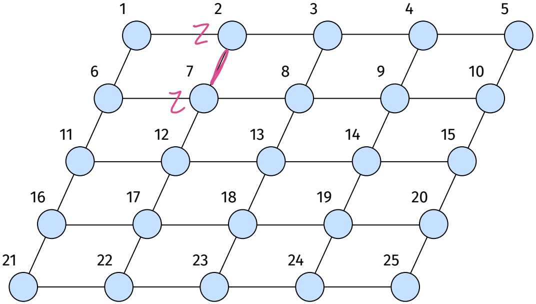
	sample lower bound	sample upper bound	time upper bound
classical		$\frac{e^{O(\beta)}}{\beta^2\varepsilon^2}\log N$	$\frac{e^{O(\beta)}}{\beta^2\varepsilon^2}N\log N$
quantum			
our work			

Classical Gibbs states are samples from Markov Random Fields (MRFs)

Restrict to diagonal Hamiltonians H (i.e. Paulis Z, I). Then the Gibbs state ρ is also diagonal, and so is a sample from a classical probability distribution. This is a *Markov Random Field* (MRF).

Consider the *Ising model*: H is 2-local on the graph $G = ([N], E)$. Interpret the basis states $|0\rangle$ and $|1\rangle$ as ± 1, so ρ is an element of $\{+1, -1\}^N$, where

$$\Pr[\rho = x] \propto \exp\left(-\beta \sum_{(a,b) \in E} \lambda_{ab} x_a x_b\right)$$



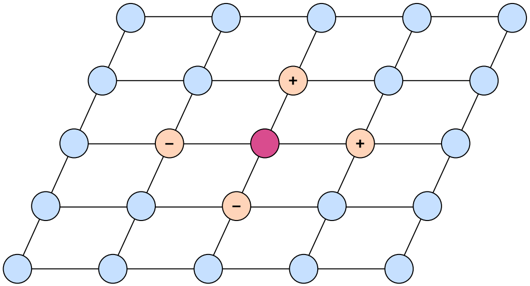
MRFs satisfy the Markov property

Conditioned on the neighborhood of i, x_i is independent of the rest of the bits.

$$\Pr[\rho = x] \propto \exp\Big(-\beta \sum_{(a,b)\in E} \lambda_{ab} x_a x_b\Big) = \prod_{j \in n(i)} \exp(-\beta \lambda_{ij} x_i x_j) \prod_{\substack{(a,b)\in E \\ a,b \neq i}} \exp(-\beta \lambda_{ab} x_a x_b)$$

+1

−1



Optimal classical Hamiltonian learning using the Markov property

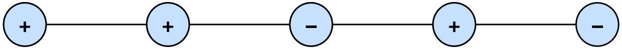
To learn a λ_{ij}:

1. Get $\frac{e^{O(\beta)}\log M}{\beta^2\varepsilon^2}$ copies of ρ.
2. Consider only samples where $x_k = +1$ for all k adjacent to $\{i, j\}$;

$$\Pr[\rho = x] \propto \exp\Big(-\beta \sum_{(a,b)\in E} \lambda_{ab} x_a x_b\Big)$$

$$\Pr[\rho_i, \rho_j = x_i, x_j \mid \rho_k = +1] \propto \exp(-\beta\lambda_{ij} x_i x_j - \beta\eta_i x_i - \beta\eta_j x_j)$$

3. Learn the conditional distribution on $\{x_i, x_j\}$ enough to infer λ_{ij};



Optimal classical Hamiltonian learning using the Markov property

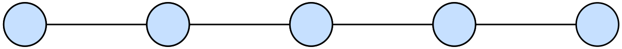
To learn a λ_{ij}:

1. Get $\frac{e^{O(\beta)} \log M}{\beta^2 \varepsilon^2}$ copies of ρ.
2. Consider only samples where $x_k = +1$ for all k adjacent to $\{i, j\}$;

$$\Pr[\rho = x] \propto \exp\Big(-\beta \sum_{(a,b) \in E} \lambda_{ab} x_a x_b\Big)$$

$$\Pr[\rho_i, \rho_j = x_i, x_j \mid \rho_k = +1] \propto \exp(-\beta \lambda_{ij} x_i x_j - \beta \eta_i x_i - \beta \eta_j x_j)$$

3. Learn the conditional distribution on $\{x_i, x_j\}$ enough to infer λ_{ij};



−	−	−	−	+
+	−	+	+	+
+	−	+	−	−
+	+	+	+	+
−	+	+	+	−
+	+	−	+	−

Optimal classical Hamiltonian learning using the Markov property

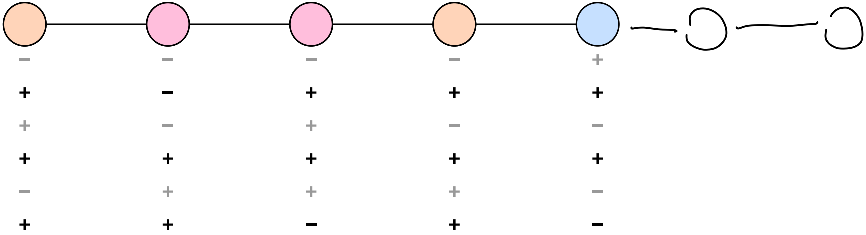
To learn a λ_{ij}:

1. Get $\frac{e^{O(\beta)} \log M}{\beta^2 \varepsilon^2}$ copies of ρ.
2. Consider only samples where $x_k = +1$ for all k adjacent to $\{i, j\}$;

$$\Pr[\rho = x] \propto \exp\Big(-\beta \sum_{(a,b)\in E} \lambda_{ab} x_a x_b\Big)$$

$$\Pr[\rho_i, \rho_j = x_i, x_j \mid \rho_k = +1] \propto \exp(-\beta\lambda_{ij}x_i x_j - \beta\eta_i x_i - \beta\eta_j x_j)$$

3. Learn the conditional distribution on $\{x_i, x_j\}$ enough to infer λ_{ij};



Sketching our main result

	sample lower bound	sample upper bound	time upper bound
classical			
quantum			
our work		$\frac{1}{\beta^2\varepsilon^2}\log N$	$\frac{1}{\beta^2\varepsilon^2}N\log N$

Trying the natural approach: generalizing the Markov property

In the classical setting, we used the Markov property to *restrict to a constant-sized subsystem*. However, the Markov property does not hold for general Hamiltonians.

An approximate version *does hold* for sufficiently small β (high temperature)[4]

- conditional mutual information bounds:

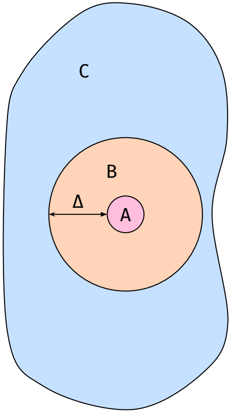
$$I(A : C \mid B) \leq e^{-\Omega(\Delta)}$$

- effective Hamiltonian bounds: let

$$\tilde{H}_A := -\beta^{-1} \log \mathrm{Tr}_{A^c}(e^{-\beta H})$$

$$H_A := \textstyle\sum_{a:\mathrm{Supp}(E_a)\subseteq A} \lambda_a E_a.$$

then $\|\tilde{H}_{A\cup B} - \tilde{H}_B - H_A\| \leq e^{-\Omega(\Delta)}$



Can restrict to a subsystem: $A \cup B$, for $\Delta = \log \frac{1}{\varepsilon}$.
However, this is $\log^d \frac{1}{\varepsilon}$ qubits: *not small enough*.

[4] Kuwahara, Kato, and Brandão 2020

Looking closer: Anshu, Arunachalam, Kuwahara, and Soleimanifar 2021

Claim

The values of $\mathrm{Tr}(E_a \rho(x))$ for all $a \in [M]$ determines $x \in [-1, 1]^M$.[5]

[5] Recall $\rho(x) = \exp(-\beta H)/\mathrm{Tr}\exp(-\beta H)$ for $H = \sum_a x_a E_a$.

Looking closer: Anshu, Arunachalam, Kuwahara, and Soleimanifar 2021

Claim

The values of $\mathrm{Tr}(E_a \rho(x))$ for all $a \in [M]$ determines $x \in [-1, 1]^M$.[5]

Strategy: get estimates $\tilde{e}_a$ of $\mathrm{Tr}(E_a \rho)$ for all $a \in [M]$, then deduce estimates $\tilde{\lambda}_a$ of λ_a.

The strategy works (information-theoretically) provided that

$$\left|\mathrm{Tr}(E_a \rho(\tilde{\lambda})) - \mathrm{Tr}(E_a \rho(\lambda))\right| \leq \varepsilon \text{ for all } a \in [M] \implies \|\tilde{\lambda} - \lambda\|_\infty \leq L\varepsilon$$

for some bound L.

$\{\mathrm{Tr}(E_a \rho(x))\}_{a \in [M]} \mapsto x$

bounded Jacobian

$x \mapsto \{\mathrm{Tr}(E_a \rho(x))\}$

bounded $(\text{Jacobian})^{-1}$

[5] Recall $\rho(x) = \exp(-\beta H)/\mathrm{Tr}\exp(-\beta H)$ for $H = \sum_a x_a E_a$.

Looking closer: Anshu, Arunachalam, Kuwahara, and Soleimanifar 2021

Claim

The values of $\operatorname{Tr}(E_a \rho(x))$ for all $a \in [M]$ determines $x \in [-1, 1]^M$.[5]

Strategy: get estimates $\tilde{e}_a$ of $\operatorname{Tr}(E_a \rho)$ for all $a \in [M]$, then deduce estimates $\tilde{\lambda}_a$ of λ_a.

The strategy works (information-theoretically) provided that

$$\left|\operatorname{Tr}(E_a \rho(\tilde{\lambda})) - \operatorname{Tr}(E_a \rho(\lambda))\right| \leq \varepsilon \text{ for all } a \in [M] \implies \|\tilde{\lambda} - \lambda\|_\infty \leq L\varepsilon$$

for some bound L.

More precisely, consider the Jacobian of the map $x \mapsto \operatorname{Tr}(E_a \rho(x))$. That is, $J_{ab}(x) = \partial_b \operatorname{Tr}(E_a \rho(x))$. Then if we show

$$\|J^{-1}(x)\|_{\infty\to\infty} \leq L \text{ for all } x \in [-1, 1]^M,$$

this implies that the strategy has sample complexity $O(L^2 \frac{\log M}{\varepsilon^2})$.

[AAKS21] prove an equivalent statement (strong concavity of the log-partition function).

---

[5] Recall $\rho(x) = \exp(-\beta H)/\operatorname{Tr}\exp(-\beta H)$ for $H = \sum_a x_a E_a$.

Exploiting high temperature: cluster expansion

Main idea: We use the structural results used by [KKB20] to achieve the high-temperature approximate Markov properties to understand how $\{\mathrm{Tr}(E_a\rho)\}_a$ and λ relate.

Cluster expansion[6]

The multivariate Taylor series expansion for the log-partition function $\mathscr{L}$,

$$\mathscr{L} = \log \mathrm{Tr} \exp(-\beta \textstyle\sum_{a=1}^{M} \lambda_a E_a) = \displaystyle\sum_{\mathbf{V}} \frac{\lambda^{\mathbf{V}}}{\mathbf{V}!} \left(\partial_{\mathbf{V}} \mathscr{L} \Big|_{\lambda=0} \right),$$

converges for $\beta < \beta_c = \Theta(1)$. So, we can write

$$\mathrm{Tr}(E_a\rho) = -\frac{1}{\beta}\left(\frac{\partial}{\partial \lambda_a}\mathscr{L}\right) = \sum_{m=0}^{\infty} \beta^m p_m^{(a)}(\lambda),$$

where $p_m^{(a)}(\lambda)$ is some degree-m homogeneous polynomial (which we can compute).

[6] Kuwahara and Saito 2019

Getting a sample complexity bound

Now, we have a system of polynomial equations,

$$\mathrm{Tr}(E_a \rho) = 0 + \beta \lambda_a + \beta^2 p_2^{(a)}(\lambda) + \beta^2 p_3^{(a)}(\lambda) + \cdots$$

In the $\beta \to 0$ regime, we have $\lambda_a = \frac{1}{\beta} \mathrm{Tr}(E_a \rho)$. So, all we need are estimates of the $\mathrm{Tr}(E_a \rho)$'s to $\beta \varepsilon$ error (which takes $O(\frac{\log M}{\beta^2 \varepsilon^2})$ samples as desired).

Formally, we show that the Jacobian J with $J_{ab} = \partial_a \mathrm{Tr}(E_b \rho)$ satisfies

$$\|J - \beta I\|_{\infty \to \infty} \leq \frac{\beta}{2} \implies \|J^{-1}\|_{\infty \to \infty} \leq \frac{2}{\beta}$$

for sufficiently small β.

$J = \beta I + O(\beta^2)$

The full algorithm

Quantum part

Given copies of $\rho(\lambda) = \frac{\exp(-\beta H)}{\operatorname{Tr}\exp(-\beta H)}$, get estimates $\tilde{e}_a$ of $\operatorname{Tr}(E_a \rho(\lambda))$ up to $\beta\varepsilon$ error, for all $a \in [M]$.

Classical part

We want to find an x such that, for all $a \in [M]$,

$$\operatorname{Tr}(E_a \rho(\lambda)) \approx \tilde{e}_a \approx \beta x_a + \beta^2 p_2^{(a)}(x) + \cdots + \beta^{\mathfrak{m}} p_{\mathfrak{m}}^{(a)}(x) \approx \sum_{m=1}^{\infty} \beta^m p_m^{(a)}(x) = \operatorname{Tr}(E_a \rho(x)).$$

We *truncate* at $\mathfrak{m} = O(\log \frac{1}{\varepsilon})$.

The full algorithm

Quantum part

Given copies of $\rho(\lambda) = \frac{\exp(-\beta H)}{\operatorname{Tr}\exp(-\beta H)}$, get estimates $\tilde{e}_a$ of $\operatorname{Tr}(E_a \rho(\lambda))$ up to $\beta\varepsilon$ error, for all $a \in [M]$.

Classical part

We want to find an x such that, for all $a \in [M]$,

$$\operatorname{Tr}(E_a \rho(\lambda)) \approx \tilde{e}_a \approx \beta x_a + \beta^2 p_2^{(a)}(x) + \cdots + \beta^{\mathfrak{m}} p_{\mathfrak{m}}^{(a)}(x) \approx \sum_{m=1}^{\infty} \beta^m p_m^{(a)}(x) = \operatorname{Tr}(E_a \rho(x)).$$

We *truncate* at $\mathfrak{m} = O(\log \frac{1}{\varepsilon})$. In other words, we want that $\|\mathcal{F}(x)\|_\infty \le 50\beta\varepsilon$, for

$$\mathcal{F}(x)_a := -\tilde{e}_a + \beta x_a + \beta^2 p_2^{(a)}(x) + \cdots + \beta^{\mathfrak{m}} p_{\mathfrak{m}}^{(a)}(x).$$

Use the *Newton–Raphson* method for root-finding: $x^{(0)} = \vec{0}$, and

$x^{(t+1)} = x^{(t)} - (J^{-1}\mathcal{F})(x^{(t)})$ until convergence ($O(\log \frac{1}{\beta\varepsilon})$ iterations).

Thank you!

$$\begin{bmatrix} 1+\varepsilon & & \\ & 1-\varepsilon & \\ & & -2 \end{bmatrix} \text{ or } \begin{bmatrix} 1 & & \\ & 1 & \\ & & -2 \end{bmatrix}$$

	sample lower bound	sample upper bound	time upper bound
classical[7]	$\frac{e^{\Omega(\beta)}}{\beta^2\varepsilon^2}\log N$ *	$\frac{e^{O(\beta)}}{\beta^2\varepsilon^2}\log N$	$\frac{e^{O(\beta)}}{\beta^2\varepsilon^2}N\log N$
quantum[8]	$\frac{1}{\beta\varepsilon}$	$\frac{e^{O(\beta^{4+c})}}{\beta^{4+c'}\varepsilon^2}N^2\log N$	$\frac{1}{\beta^{4+c'}\varepsilon^2}N^{2+c''}$
our work	$\frac{e^{2\beta}}{\beta^2\varepsilon^2}\log N$	$\frac{1}{\beta^2\varepsilon^2}\log N$	$\frac{1}{\beta^2\varepsilon^2}N\log N$

Blue indicates the assumption that $\beta < \beta_c = \Theta(1)$

[7] Results shown for the Ising model; sample complexity results assume the terms are also unknown [Santhanam and Wainwright 2012; Vuffray, Misra, Lokhov, and Chertkov 2016]

[8] Anshu, Arunachalam, Kuwahara, and Soleimanifar 2021

Why is this algorithm optimal?

The runtime is dominated by the time to compute the polynomial approximation

$$\mathrm{Tr}(E_a \rho(x)) \approx \beta x_a + \beta^2 p_2^{(a)}(x) + \cdots + \beta^{\mathfrak{m}} p_{\mathfrak{m}}^{(a)}(x),$$

for all $a \in [M]$: $M \exp(c\mathfrak{m})$. *This gets easier as β gets smaller.*

With our analysis, β_c is small enough that $\mathfrak{m} < \frac{1}{c} \log \frac{1}{\varepsilon} + c'$. This gives a runtime of $O(M \frac{1}{\varepsilon} \operatorname{polylog}(\frac{1}{\varepsilon\beta}))$ for the Newton–Raphson method, which is smaller than the $O(\frac{N \log M}{\beta^2 \varepsilon^2})$ time needed to get the estimates.

Bounding the Taylor series expansion

The multivariate Taylor series expansion for the log-partition function $\mathscr{L}$,

$$\mathscr{L} = \log \operatorname{Tr} \exp(-\beta H) = \sum_{\mathbf{V}} \frac{\lambda^{\mathbf{V}}}{\mathbf{V}!} \left(\partial_{\mathbf{V}} \mathscr{L} \Big|_{\lambda=0} \right), \text{ where } \mathbf{V} = \{(a, \mu(a)) : a \in [M], \mu(a) \in \mathbb{Z}_{\geq 0}\}$$

Key observation

Define the *dual graph* $\mathfrak{G}$ to have vertices $[M]$ and an edge (a, b) iff $|\operatorname{Supp}(E_a) \cap \operatorname{Supp}(E_b)| \neq 0$. This graph is degree $O(1)$.

Then $\partial_{\mathbf{V}} \mathscr{L}|_{\lambda=0} = 0$ if $\mathbf{V}$ is *not connected* in $\mathfrak{G}$.

Bounding the Taylor series expansion

The multivariate Taylor series expansion for the log-partition function $\mathscr{L}$,

$$\mathscr{L} = \log \operatorname{Tr} \exp(-\beta H) = \sum_{\text{connected } \mathbf{V}} \frac{\lambda^{\mathbf{V}}}{\mathbf{V}!} \left(\partial_{\mathbf{V}} \mathscr{L} \Big|_{\lambda=0} \right).$$

Lemma

There are $\exp(O(m))$ *many* $\mathbf{V}$*'s that are connected and weight* m*.*

Lemma

For a $\mathbf{V}$ *of weight* m*,* $\frac{1}{\mathbf{V}!} |\partial_{\mathbf{V}} \mathscr{L}|_{\lambda=0}| = \exp(O(m)) \beta^m$*.*