Efficient learning of commuting hamiltonians on lattices

Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, Mehdi Soleimanifar

March 23, 2021

Background: In a recent work [AAKS20] we constructed an algorithm to learn the hamiltonian from a Gibbs state at any constant temperature. The algorithm is sample-efficient (polynomially tight) when the learning is required for small $\ell_2$ error. It is time-efficient above critical temperatures and for stoquastic hamiltonians.

In this note, we consider the Gibbs state of a commuting hamiltonian and provide an algorithm that is both sample-efficient and time-efficient at any constant temperature (and works for small $\ell_\infty$ error).

TL;DR: Effective hamiltonian of the reduced state of a ‘commuting Gibbs state’ is also local. Thus, learning can be performed locally.

0.1 Notation and effective hamiltonian

Fix a $D$-dimensional lattice and let each spin have dimension $d$. Consider a $k$-local hamiltonian

$$H = \sum_{\ell=1}^m h_\ell$$

with $\|h_\ell\| \leq 1$ ($\forall \ell$, where $\|\cdot\|$ denotes the $\ell_\infty$ norm) and assume that $[h_\ell, h_{\ell'}] = 0$ ($\forall \ell, \ell'$). Let $h_R$ denote the hamiltonian restricted to a region $R$. Let $\rho_\beta = \frac{e^{-\beta H}}{\text{Tr}(e^{-\beta H})}$ be the Gibbs state. For any region $R$ on the lattice, define the effective hamiltonian $H_R = -\frac{1}{\beta} \log \text{Tr}_{R^c}(\rho_\beta)$. Let $\partial R$ be the boundary of $R$, and $\partial_- R$ be the inner boundary of $R$. The following lemma says that the effective Hamiltonian is local. It is not known to hold in the general case, except above critical temperatures [KKBa20 Theorem 2].

Lemma 1. It holds that

$$H_R = \alpha R I + h_R + \Phi,$$

where $\Phi$ is only supported on $\partial_- R$ and $[\Phi, h_R] = 0$. Here, $\alpha_r$ is some real number and $\|\Phi\| \leq 2|\partial R|$.

Proof. We can write $H = h_R + h_{\partial R} + h_{R^c}$. Consider

$$\text{Tr}_{R^c}(e^{-\beta H}) = e^{-\beta h_R} \text{Tr}_{R^c}(e^{-\beta(h_{\partial R} + h_{R^c})}).$$

Define $e^{-\beta \Phi} := \text{Tr}_{R^c}(e^{-\beta(h_{\partial R} + h_{R^c})})$. It is clear that $[\Phi, h_R] = 0$ and hence $H_R$ has the form as stated in the lemma. In order to bound the norm of $\Phi$, we proceed as follows. Consider,

$$h_R + h_{R^c} - |\partial R| I \leq H \leq h_R + h_{R^c} + |\partial R| I.$$

Since every term commutes, we can exponential the Lowner inequality to obtain

$$e^{-\beta |\partial R|} e^{-\beta h_R} \otimes e^{-\beta h_{R^c}} \leq e^{-\beta H} \leq e^{\beta |\partial R|} \otimes e^{-\beta h_R} e^{-\beta h_{R^c}}.$$
Tracing out the region $R_c$, this means that
\[
e^{-\beta |\partial R|} \text{Tr} \left( e^{-\beta h_R} \right) \leq \text{Tr}_{R^c} \left( e^{-\beta H} \right) \leq e^{\beta |\partial R|} \text{Tr} \left( e^{-\beta h_R} \right) e^{-\beta h_R}.
\]
Thus, the ratio between largest and smallest eigenvalues of $e^{\beta h_R} \text{Tr}_{R^c} \left( e^{-\beta H} \right) = e^{-\beta \Phi}$ is upper bounded by $e^{2\beta |\partial R|}$. This completes the proof.

The above lemma ensures the following identity
\[
\text{Tr}_{R^c} (\rho_\beta) = \frac{e^{-\beta (h_R + \Phi)}}{\text{Tr} (e^{-\beta (h_R + \Phi)})}.
\]

### 0.2 Learning algorithm

For every $\ell$, let $R_\ell$ be the smallest region that contains $\text{supp} (h_\ell)$ in its strict interior (that is, it does not overlap with $\partial R_\ell$). We have $|R_\ell| \leq (3k)^D$. Then $\text{Tr}_{R_\ell^c} (\rho_\beta) = \frac{e^{-\beta (h_{R_\ell} + \Phi_\ell)}}{\text{Tr} (e^{-\beta (h_{R_\ell} + \Phi_\ell)})}$, where $\Phi_\ell$ is only supported $\partial R_\ell$. Since $|\Phi_\ell| \leq 2 |\partial R_\ell|$, the smallest eigenvalue of $\frac{e^{-\beta (h_{R_\ell} + \Phi_\ell)}}{\text{Tr} (e^{-\beta (h_{R_\ell} + \Phi_\ell)})}$ is at least
\[
e^{-\beta (|R_\ell| + |\partial R_\ell|)} \geq e^{-\beta \log d (3k)^D}.
\]
The algorithm is as follows. We divide $\{ R_\ell \}_{\ell=1}^m$ into different batches, such that within each batch the $R_\ell$’s don’t overlap. Number of batches needed is $(kD)^D$ (a constant). Within each batch, we perform tomography to obtain the classical description of $\frac{e^{-\beta (h_{R_\ell} + \Phi_\ell)}}{\text{Tr} (e^{-\beta (h_{R_\ell} + \Phi_\ell)})}$ up to an error of $\epsilon e^{-(\beta + \log d (3k)^D}$. This gives us a classical description of an operator $h'$ satisfying $\| h' - h_{R_\ell} - \Phi_\ell \| \leq \epsilon$. From this, $h_\ell$ can be computed by evaluating
\[
h'_\ell := \frac{1}{d|R_\ell|} \sum_{j=1}^{d^{2k}} \sigma_\ell^{(j)} \text{Tr} \left( \sigma_\ell^{(j)} h' \right),
\]
where $\{ \sigma_\ell^{(j)} \}_{j=1}^{d^{2k}}$ are the Pauli operators in the support of $h_\ell$. It can be seen that
\[
\| h'_\ell - h_\ell \| \leq d^{2k} \epsilon.
\]

In order to perform the tomography in each batch with probability of success $1 - \frac{\delta}{\text{number of batches}}$, the number of samples needed is $[\text{CW20, BMBO20, HKP20}]$

\[
\frac{\epsilon^2}{\epsilon^2} \log \left( m \frac{\text{number of batches}}{\delta} \right) \leq \frac{\epsilon^2}{\epsilon^2} \log \frac{m (kD)^D}{\delta}.
\]
Thus, setting $d = \mathcal{O}(1)$, total sample complexity is (accounting for all the batches)
\[
\frac{\epsilon^2}{d} \log \frac{m}{\delta}.
\]

Time complexity is roughly
\[
m \cdot \frac{\epsilon^2}{d} \log \frac{m}{\delta},
\]
as the time complexity for processing the data from each sample is roughly $m \cdot \epsilon^2 \Omega(\beta D)$. 

\[2\]
References


