the geometric structure of physical Hamiltonians, and the particular exponential form of the quantum Gibbs state (35). We proceed with this task below.

3. A strategy for learning Gibbs states at high temperatures

We will present the quantum learning algorithm of [**HKT22**] for learning quantum Gibbs states at high temperature. For this we set up some notation and then sketch the proof before delving into details. Suppose we write our many-body Hamiltonian as

$$H = \sum_{a \in [M]} \lambda_a \, E_a \, ,$$

where each $E_a \in \mathbb{C}^{d \times d}$ is a distinct, non-identity, traceless Hermitian operator with $\|E_a\| \leq 1$, and we take each Hamiltonian coefficient to be a real $\lambda_a \in [-1, 1]$. The list of coefficients is $\lambda = (\lambda_1, ..., \lambda_M)$. We will sometimes denote the data of the Hamiltonian by (a, E_a, λ), indicating the index set $a \in [M]$, the set of operators E_a, and the list of coefficients λ. The following definition is useful.

Definition 128 (Dual interaction graph, using the notation of [**HKT22**])**.** *For any Hamiltonian in the set* $\{(a, E_a, \lambda) : a \in [M]\}$ *there is an* ***dual interaction graph*** $\mathfrak{G}$ *with vertex set* $[M]$ *where an edge connects vertices* a *and* b *if and only if* $a \neq b$ *and*

$$\mathrm{Supp}(E_a) \cap \mathrm{Supp}(E_b) \neq \emptyset \, .$$

We let $\mathfrak{d}$ *denote the maximum degree of the graph* $\mathfrak{G}$*.*

We will consider the setting in which $\mathfrak{d}$ is a constant independent of M, e.g. if each E_a acts on a constant number of qubits and each qubit participates in a constant number of terms. This covers a large class of physical Hamiltonians. Henceforth we will take, without loss of generality, each E_a to be a non-identity Pauli string acting on constant number of qubits.

The basic architecture of the proof of [**HKT22**] is as follows. We consider expectation values of each E_a in the thermal state of interest, and perform the high-temperature (i.e. small β) expansion

$$\langle E_a \rangle_\beta = \frac{\mathrm{tr}(e^{-\beta H} E_a)}{\mathrm{tr}(e^{-\beta H})} = \frac{\mathrm{tr}(E_a)}{d} + \sum_{m=1}^{\infty} \beta^m p_m^{(a)}(\lambda_1, ..., \lambda_M) \tag{36}$$

where the term $\frac{\mathrm{tr}(E_a)}{d} = 0$ since each E_a is a non-identity Pauli and thus traceless, and each $p_m^{(a)}$ is a degree m homogeneous polynomial in the Hamiltonian coefficients. Moreover, we can determine the form of any particular $p_m^{(a)}$ via an efficient classical computation. We first find a constant β_c below which the above series converges, i.e. a temperature above which our expansion makes sense. For this we find the radius of convergence of the series in the complex β-plane, which involves constraining the maximum 'sizes' of the polynomials p_m (recalling that λ_a's are at most magnitude 1) using the locality structure of the Hamiltonian and a so-called **cluster expansion**, to be explicated shortly.

Having argued that (36) makes sense above some constant temperature, the basic strategy is to argue that we can truncate the sum over m at some finite order;

then we have

$$\langle E_a \rangle_\beta \approx \sum_{m=1}^{m_{\max}} \beta^m p_m^{(a)}(\lambda_1, ..., \lambda_M) \tag{37}$$

for each a. (We will of course quantify how the approximation '≈' depends on $m_{\max}$ later on.) The left-hand side is measurable, and since the p_m's on the right-hand side are efficiently classically computable, we can feasibly try to 'solve the polynomial system' given by (36) for $a \in [M]$, and obtain the coefficients $\lambda_1, ..., \lambda_M$. This last part seems tricky; for instance, there could be many spurious solutions to the equations which do not give the true Hamiltonian, or possibly many 'near'-solutions which are hard to distinguish from true solutions. Remarkably, using some nice properties of a generating function for correlation functions of the Gibbs state, we can formulate the solution of the system in (37) as a *minimization problem* which is guaranteed to be *convex* in our high-temperature regime of interest. As such, we can efficiently land on the correct λ_a's within a small approximation error.

We will segment our description of the proof into four parts accordingly. First we will explain the high-temperature cluster expansion which allows us to write (36) for all $\beta < \beta_c = O(1)$. Then we will show how solving the system given by (37) for $a \in [M]$ can be formulated as a minimization problem which is convex in our regime of interest. Next we explain a useful and efficient algorithm for solving said optimization problem. Finally we put all of the bounds together and formulate the full algorithm, in its full complexity-theoretic glory.

3.1. High-temperature cluster expansion

As advertised, we begin by justifying the series expansion in (36), and moreover in particular providing a bound on its radius of convergence β_c.

3.1.1. *Generating functions for Gibbs states*

First we need a useful mathematical object, namely the generating function of correlation functions of our Gibbs state. We write

$$\mathsf{F}(\beta, \lambda_1, ..., \lambda_M) := -\frac{1}{\beta} \log \operatorname{tr} \exp(-\beta H) = -\frac{1}{\beta} \log \operatorname{tr} \exp\left(-\beta \sum_{a \in [M]} \lambda_a \, E_a\right) ,$$

which is called the **Helmholtz free energy**, which we will call by its nickname, the 'free energy'. Note that using our notation from before $Z(\beta) = \operatorname{tr}(e^{-\beta H})$, the free energy can be written as $\mathsf{F} = -\frac{1}{\beta} \log Z(\beta)$. The free energy will serve as a generating function due to the following lemma.

Lemma 129 (The free energy is a generating function for the Gibbs state)**.** *Consider a Hamiltonian $H = \sum_{a \in [M]} \lambda_a \, E_a$ where the λ_a are regarded as formal variables. For non-zero $\beta \in \mathbb{C}$, we have*

$$\operatorname{tr}(E_a \rho_\beta) = \frac{\partial}{\partial \lambda_a} \mathsf{F}(\beta, \lambda_1, ..., \lambda_M)$$

for all $a \in [M]$.

PROOF. Let us write

$$
\begin{aligned}
-\frac{1}{\beta}\frac{\partial}{\partial\lambda_a}\operatorname{tr}\exp(-\beta H) &= -\frac{1}{\beta}\sum_{m=0}^{\infty}\frac{1}{m!}\operatorname{tr}\left[\frac{\partial}{\partial\lambda_a}(-\beta H)^m\right] \\
&= -\frac{1}{\beta}\sum_{m=1}^{\infty}\frac{1}{m!}\sum_{k=1}^{m}\operatorname{tr}[(-\beta H)^{k-1}(-\beta E_a)(-\beta H)^{m-k}] \\
&= \sum_{m=1}^{\infty}\frac{1}{m!}\sum_{k=1}^{m}\operatorname{tr}[E_a(-\beta H)^{m-1}] \\
&= \operatorname{tr}[E_a\exp(-\beta H)]\,,
\end{aligned}
$$

where we have used the linearity of the trace to move β's outside of it, and the cyclicity of the trace in going from the second line to the third line. We complete the proof by observing that

$$
-\frac{1}{\beta}\frac{\partial}{\partial\lambda_a}\log\operatorname{tr}\exp(-\beta H) = -\frac{1}{\beta}\frac{1}{\operatorname{tr}\exp(-\beta H)}\frac{\partial}{\partial\lambda_a}\operatorname{tr}\exp(-\beta H) = \frac{\operatorname{tr}(E_a\exp(-\beta H))}{\operatorname{tr}\exp(-\beta H)}\,.
$$

□

Thus to study series expansions of the form (36), it is natural to leverage the free energy F. One minor annoyance of the free energy is that it does not converge as $\beta \to 0$, going as $\sim 1/\beta$. This is not a problem of course, and motivates us to define the ancillary quantity

$$
\mathcal{L}(\beta,\lambda_1,...,\lambda_M) := (-\beta)\,\mathsf{F}(\beta,\lambda_1,...,\lambda_M) = \log\operatorname{tr}\exp\left(-\beta\sum_{a\in[M]}\lambda_a\,E_a\right),
$$

which goes to a constant as $\beta \to 0$. We will mostly use $\mathcal{L}$ henceforth.

Taking a step back, let us get a sense of what we want to prove. Consider the toy function

$$
f(\beta) = \sum_{m=1}^{\infty} c_m\,\beta^m\,.
$$

We would like to understand under what conditions on the c_m's is there a non-zero radius of convergence. We recall from complex analysis that the radius of convergence β_c of a function of the form $f(\beta)$ is given by

$$
\frac{1}{\beta_c} = \limsup_{m\to\infty}|c_m|^{1/m}\,,
$$

and so β_c is non-zero when the c_m's grow *at most* exponentially in m. In our more complicated Gibbs setting, we will show that $\max_{\lambda_1,...,\lambda_M\in[-1,1]}|p_m^{(a)}(\lambda_1,...,\lambda_M)|$ indeed grows at most exponentially in m, which will do the job. Moreover, our β_c will not depend on M and thus on the number of sites n, which is desirable since we would like our temperature bound to be system size independent. In particular, we will show that each $p_m^{(a)}$ satisfies two properties:

(1) Each $p_m^{(a)}$ is a sum of at most $e\mathfrak{d}(1+e(\mathfrak{d}-1))^m$ monomials.
(2) The coefficient in front of any monomial of $p_m^{(a)}$ has magnitude at most $(2e(\mathfrak{d}+1))^{m+1}(m+1)$.

Putting these together, we will have

$$\max_{\lambda_1,...,\lambda_M\in[-1,1]} |p_m^{(a)}(\lambda_1,...,\lambda_M)| \leq e(m+1)\mathfrak{d}(1+e(\mathfrak{d}-1))^m(2e(\mathfrak{d}+1))^{m+1}$$

which manifestly grows at most exponentially in m, and is independent of M. The key to getting M-independence will be to use the spatial locality structure of the Hamiltonian.

To clarify the structure of the series expansion of $\mathcal{L}$ and thus F, it is useful to introduce some notation. We first opt to write

$$\mathcal{L} = \log \operatorname{tr} \exp\left(-\sum_{a\in[M]} z_a\, E_a\right) \quad \text{where } z_a := \beta\lambda_a\,.$$

We will take $z = (z_1, ..., z_M) \in \mathbb{C}^M$, i.e. considering complexified couplings for purposes of assessing convergence. By the chain rule

$$\frac{\partial \mathcal{L}}{\partial \beta} = \sum_{a\in[M]} \frac{\partial z_a}{\partial \beta}\frac{\partial \mathcal{L}}{\partial z_a} = \sum_{a\in[M]} \lambda_a \frac{\partial \mathcal{L}}{\partial z_a}$$

which gives us the multivariable series expansion

$$\begin{aligned}\mathcal{L} &= \sum_{m=0}^{\infty} \frac{\beta^m}{m!}\left(\frac{\partial^m \mathcal{L}}{\partial \beta^m}\bigg|_{\beta=0}\right)\\ &= \sum_{m=0}^{\infty} \frac{\beta^m}{m!} \sum_{a_1,a_2,...,a_m\in[M]} \lambda_{a_1}\cdots\lambda_{a_m}\left(\frac{\partial^m \mathcal{L}}{\partial z_{a_1}\cdots\partial z_{a_m}}\bigg|_{z=(0,...,0)}\right). \qquad (38)\end{aligned}$$

In the last equation, for each fixed m, we have an inner sum over m variables $a_1, ..., a_m \in [M]$. This, of course, is the standard structure for a multivariable series expansion; it behooves us to write this in a more compact notation so that it is more intelligible. To this end, we have the definition:

Definition 130 (Clusters of multivariate indices)**.** *A* cluster $\boldsymbol{V}$ *is a set of tuples* $\{(a,\mu(a)) : a \in [M]\}$ *where* $\mu : [M] \to \mathbb{Z}_{\geq 0}$ *counts the multiplicity of each* a*. Then the total weight* $|\boldsymbol{V}|$ *of* $\boldsymbol{V}$ *is* $\sum_a \mu(a)$*. We will write* $a \in \boldsymbol{V}$ *if* $\mu(a) \geq 1$*, and define the* support *of* $\boldsymbol{V}$ *as* $\operatorname{Supp}\boldsymbol{V} := \{a \in [M] : \mu(a) \geq 1\}$*. Finally, we define the combinatorial factor* $\boldsymbol{V}! := \prod_{a\in[M]} \mu(a)!$*.*

This is sometimes called **multi-index notation**, where $\mathbf{V}$ is the multi-index. With this notation in mind, let us rewrite (38) in a more compact manner, and define a few more pieces of notation along the way. In particular, (38) can be written as

$$\begin{aligned}\mathcal{L} &= \sum_{m=0}^{\infty} \beta^m \sum_{\mathbf{V}:|\mathbf{V}|=m} \frac{1}{\mathbf{V}!} \prod_{a\in\operatorname{Supp}\mathbf{V}} \lambda_a^{\mu(a)} \left(\prod_{b\in\operatorname{Supp}\mathbf{V}} \frac{\partial^{\mu(b)}}{\partial z_b^{\mu(b)}}\right)\Bigg|_{z=(0,...,0)} \mathcal{L}\\ &= \sum_{m=0}^{\infty} \sum_{\mathbf{V}:|\mathbf{V}|=m} \frac{1}{\mathbf{V}!} \underbrace{\prod_{a\in\operatorname{Supp}\mathbf{V}} \lambda_a^{\mu(a)}}_{=:\,\lambda^{\mathbf{V}}} \underbrace{\left(\prod_{b\in\operatorname{Supp}\mathbf{V}} \frac{\partial^{\mu(b)}}{\partial \lambda_b^{\mu(b)}}\right)\Bigg|_{\lambda=(0,...,0)}}_{=:\,\mathcal{D}_{\mathbf{V}}} \mathcal{L}\\ &= \sum_{m=0}^{\infty} \sum_{\mathbf{V}:|\mathbf{V}|=m} \frac{\lambda^{\mathbf{V}}}{\mathbf{V}!}\mathcal{D}_{\mathbf{V}}\mathcal{L} \qquad (39)\end{aligned}$$

where we have introduced the notation $\lambda^{\mathbf{V}}$ and $\mathcal{D}_{\mathbf{V}}$. The final form (39) is compactly expressed and readily intelligible, and so was worth our efforts in notational wrangling. The form of (39) makes the source of our difficulty clearer. For $|\mathbf{V}| = m$ and any fixed $\mathbf{V}$, the product rule expansion $\mathcal{D}_{\mathbf{V}}\mathcal{L}$ naïvely has $m!$ terms; then if we supposed that each term has size "1" (in fact, size can be larger), we would have the back-of-the-envelope estimate

$$\sum_{\mathbf{V}:|\mathbf{V}|=m} \frac{1}{\mathbf{V}!}\, m! = M^m\,,$$

which is the number of unique length-m strings of M symbols. This type of M^m growth is exponential in m so would have a finite radius of convergence, but that radius of convergence would go as $\sim \frac{1}{M}$ which gets worse as M (or accordingly, the system size) gets larger, which we do not want. So we need to more cleverly exploit the structure of derivatives of $\mathcal{L}$ and the locality of the Hamiltonian.

To proceed, we first show $\sum_{\mathbf{V}:|\mathbf{V}|=m} \frac{\lambda^{\mathbf{V}}}{\mathbf{V}!}\mathcal{D}_{\mathbf{V}}\mathcal{L}$ contains much fewer than M^m terms. Specifically, we show that $\mathcal{D}_{\mathbf{V}}\mathcal{L}$ is non-zero only when $\mathbf{V}$ is *connected*, in the following sense:

Definition 131 (Connected clusters)**.** *A cluster $\boldsymbol{V} = \{(a, \mu(a)) : a \in [M]\}$ is* connected *if the subgraph of $\mathfrak{G}$ induced by the support of $\boldsymbol{V}$ is connected.*

Then, as advertised, we have the following lemma:

Lemma 132. *Recall that $Z(\beta) = \mathrm{tr}(e^{-\beta H})$. If $\boldsymbol{V}'$ and $\boldsymbol{V}''$ are nonempty and mutually disjoint and if there is no edge in $\mathfrak{G}$ connecting $\boldsymbol{V}'$ and $\boldsymbol{V}''$, then $\mathcal{D}_{\boldsymbol{V}'\cup\boldsymbol{V}''}Z = (\mathcal{D}_{\boldsymbol{V}'}Z)(\mathcal{D}_{\boldsymbol{V}''}Z)$. Thus if a cluster $\boldsymbol{V}$ is not connected, then we have $\mathcal{D}_{\boldsymbol{V}}\mathcal{L} = 0$.*

Proof. Let $H_{\mathbf{V}} := \sum_{a\in\mathrm{Supp}\mathbf{V}} \lambda_a E_a$. Then $H_{\mathbf{V}'}$ and $H_{\mathbf{V}''}$ commute since the supports of their constituent operators do not overlap. Moreover letting $Z_{\mathbf{V}} := \mathrm{tr}\exp(-\beta H_{\mathbf{V}})$, we evidently have $Z_{\mathbf{V}'\cup\mathbf{V}''} = Z_{\mathbf{V}'}Z_{\mathbf{V}''}$, and so we find

$$\mathcal{D}_{\mathbf{V}'\cup\mathbf{V}''}Z = \mathcal{D}_{\mathbf{V}'\cup\mathbf{V}''}Z_{\mathbf{V}'\cup\mathbf{V}''} = (\mathcal{D}_{\mathbf{V}'}Z_{\mathbf{V}'})(\mathcal{D}_{\mathbf{V}''}Z_{\mathbf{V}''}) = (\mathcal{D}_{\mathbf{V}'}Z)(\mathcal{D}_{\mathbf{V}''}Z)$$

as we claimed.

Letting $\mathcal{L}_{\mathbf{V}} := \log Z_{\mathbf{V}}$, we see that $\mathcal{L}_{\mathbf{V}'\cup\mathbf{V}''} = \mathcal{L}_{\mathbf{V}'} + \mathcal{L}_{\mathbf{V}''}$. Then we have

$$\mathcal{D}_{\mathbf{V}'\cup\mathbf{V}''}\mathcal{L} = \mathcal{D}_{\mathbf{V}'\cup\mathbf{V}''}\mathcal{L}_{\mathbf{V}'\cup\mathbf{V}''} = \mathcal{D}_{\mathbf{V}'\cup\mathbf{V}''}(\mathcal{L}_{\mathbf{V}'} + \mathcal{L}_{\mathbf{V}''}) = 0\,,$$

which is zero because the $\mathbf{V}''$ part of the derivative annihilates $\mathcal{L}_{\mathbf{V}'}$ and the $\mathbf{V}'$ part of the derivative annihilates $\mathcal{L}_{\mathbf{V}''}$. □

We have thus shown that a term $\mathcal{D}_{\mathbf{V}}\mathcal{L}$ only contributes to (39) if $\mathbf{V}$ is a connected cluster. Next, it will be useful to count the number of connected clusters $\mathbf{V}$ such that $\mathbf{V}$ contains some particular vertex a, and $|\mathbf{V}| = w$, i.e. we want to count the number of clusters with weight w containing a. We do this below.

3.1.2. *Counting the number of connected clusters of fixed weight*

Recall that the dual interaction graph of our Hamiltonian is a graph $\mathfrak{G}$ of maximum degree at most $\mathfrak{d}$. For convenience, let us distinguish a 'root' vertex $a \in V(\mathfrak{G})$. We say that a cluster $\mathbf{V} = \{(a, \mu(a)) : a \in [M]\}$ is *rooted at a* if $a \in \mathrm{Supp}\mathbf{V}$.

For $k \geq 1$ let $N_{\mathfrak{G}}(a, k)$ be the number of connected vertex sets $S \subseteq V(\mathfrak{G})$ with $|S| = k$ and $a \in S$ (we will refer to such an S as a "connected support" of size k at