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 $Z(\beta) = \operatorname{tr}(e^{-\beta H})$. 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 $Z_{\mathbf{V}} := \operatorname{tr}\exp(-\beta H_{\mathbf{V}})$, 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

a). For $w \geq 1$, let $C_{\mathfrak{G}}(a, w)$ be the number of connected clusters $\mathbf{V}$ of total weight w rooted at a. Our goal will be to upper bound

$$\max_{a \in [M]} C_{\mathfrak{G}}(a, w)$$

which will give us a bound on the number of clusters with weight w containing a. For this, two elementary observations will be useful:

(1) If a support S has $|S| = k$ vertices, the number of ways to assign positive multiplicities summing to w is

$$\#\{\mu : \ \textstyle\sum_{b \in S} \mu(b) = w, \ \mu(b) \in \mathbb{Z}_{\geq 1}\} = \binom{w-1}{k-1}.$$

In particular, the multiplicity factor depends only on k (not on the geometry of S).

(2) Consequently,

$$C_{\mathfrak{G}}(a, w) = \sum_{k=1}^{w} N_{\mathfrak{G}}(a, k) \binom{w-1}{k-1}. \tag{40}$$

Hence, if we want to upper bound $C_{\mathfrak{G}}(a, w)$, it suffices to first upper bound each $N_{\mathfrak{G}}(a, k)$.

Our strategy will be to upper bound $N_{\mathfrak{G}}(a, k)$ by

$$N_{\mathfrak{G}}(a, k) \leq \max_{\mathfrak{H} \text{ with degree } \leq \mathfrak{d}} N_{\mathfrak{H}}(a, k)\,,$$

namely to maximize over all graphs $\mathfrak{H}$ with degree at most $\mathfrak{d}$. The next proposition establishes the desired maximization.

Proposition 133 (Tree maximizes rooted supports)**.** *For every $k \geq 1$ and every graph $\mathfrak{G}$ with $\Delta(\mathfrak{G}) \leq \mathfrak{d}$,*

$$N_{\mathfrak{G}}(a, k) \leq N_{T_{\mathfrak{d}}}(r, k), \tag{41}$$

where $T_{\mathfrak{d}}$ is the infinite $\mathfrak{d}$-regular tree and r is its root. Equivalently, among all degree–$\leq \mathfrak{d}$ graphs, the number of connected supports of size k rooted at a is maximized by $T_{\mathfrak{d}}$.

Proof. We construct an injective map from the family of connected supports $S \subseteq V(\mathfrak{G})$ of size k with $a \in S$ to the family of rooted subtrees of $T_{\mathfrak{d}}$ with k vertices containing r. Fix once and for all a total order on $V(\mathfrak{G})$. Given S, run breadth-first search on the induced subgraph $\mathfrak{G}[S]$ starting at a, breaking ties by the fixed order. This yields a *canonical* rooted spanning tree T_S of S. In T_S the root has at most $\mathfrak{d}$ children and each nonroot has at most $\mathfrak{d} - 1$ children.

For each vertex $v \in V(\mathfrak{G})$, fix an injective labeling $\alpha_v : \Gamma_{\mathfrak{G}}(v) \hookrightarrow \{1, 2, \ldots, \mathfrak{d}\}$ of its neighbors. Direct the edges of T_S away from the root and label each parent$\to$child edge $u \to v$ by the *port* number $\ell(u \to v) := \alpha_u(v)$. By construction, siblings of a vertex use distinct port labels.

Now label the $\mathfrak{d}$ edges incident to every vertex of $T_{\mathfrak{d}}$ with the symbols $\{1, \ldots, \mathfrak{d}\}$. Starting at r, read the labeled rooted tree (T_S, ℓ) as instructions: from any vertex in $T_{\mathfrak{d}}$, for each child edge of T_S bearing label j, follow the unique incident edge labeled j. Distinct child labels ensure that the image is a well-defined rooted subtree of size k. Denote the resulting subtree by $\Phi(S)$.

From $\Phi(S)$ one can recover the labeled rooted tree (T_S, ℓ) (reading off port labels along edges), and then recover S itself level-by-level: the children of $u \in S$ are $\alpha_u^{-1}(\{\text{child-labels at } u\})$. Hence Φ is injective, and the claim follows. $\square$

Remark 134. *The proposition formalizes the intuition that "unrolling cycles cannot reduce the number of rooted connected substructures" under a local degree cap; $T_{\mathfrak{d}}$ is the universal cover of any degree–$\leq \mathfrak{d}$ graph. In our cluster expansion,* (40) *then shows that, for fixed w and $\mathfrak{d}$, the total number of rooted clusters is maximized on $T_{\mathfrak{d}}$.*

We now provide a quantitative bound on $N_{T_{\mathfrak{d}}}(r,k)$, which is the number of rooted subtrees of $T_{\mathfrak{d}}$ with exactly k vertices. This number does not depend on the root since $T_{\mathfrak{d}}$ is self-similar, and so we write $N_{T_{\mathfrak{d}}}(r,k) = N_{T_{\mathfrak{d}}}(k)$. We have the lemma:

Lemma 135. *For $k \in \mathbb{Z}_{\geq 0}$, let $N_{T_{\mathfrak{d}}}(k)$ be the number of all connected rooted subtrees with k nodes in the infinite $\mathfrak{d}$-regular tree. Then*

$$N_{T_{\mathfrak{d}}}(k) = \binom{k(\mathfrak{d}-1)+1}{k-1}\frac{\mathfrak{d}}{k(\mathfrak{d}-1)+1} \leq e\,\mathfrak{d}\,(e(\mathfrak{d}-1))^{k-1}\,.$$

This lemma follows from some standard generating function manipulations in analytic combinatorics, which are carried out in [**HKT22**].

By the above lemma, in conjunction with (40) and (41), we have

$$\begin{aligned}
C_{\mathfrak{G}}(a,w) &= \sum_{k=1}^{w} N_{\mathfrak{G}}(a,k)\binom{w-1}{k-1} \\
&\leq \sum_{k=1}^{w} N_{T_{\mathfrak{d}}}(k)\binom{w-1}{k-1} \\
&\leq \sum_{k=1}^{w} e\,\mathfrak{d}(e(\mathfrak{d}-1))^{k-1}\binom{w-1}{k-1} \\
&= e\,\mathfrak{d}(1+e(\mathfrak{d}-1))^{w-1},
\end{aligned}$$

and so we have obtained the following result:

Proposition 136. *Let $\mathfrak{G}$ be any graph with degree $\mathfrak{d} \geq 2$, and fix $a \in V(\mathfrak{G})$. For every $w \in \mathbb{Z}_{>0}$, the number of connected clusters $\boldsymbol{V}$ of total weight w rooted at a satisfies*

$$C_{\mathfrak{G}}(a,w) \leq e\,\mathfrak{d}\,\big(1+e(\mathfrak{d}-1)\big)^{w-1}.$$

In the degenerate case $\mathfrak{d}=1$, a trivial estimate gives $C_{\mathfrak{G}}(a,w) \leq w$.

3.1.3. *Estimating the size of cluster derivatives*

Having estimated the number of (rooted) clusters with fixed weight, we now turn to bounding the size of $\frac{1}{\mathbf{V}!}\mathcal{D}_{\mathbf{V}}\mathcal{L}$ for fixed $\mathbf{V}$. For $|\mathbf{V}| = m$, we will ultimately find a bound

$$\left|\frac{1}{\mathbf{V}!}\mathcal{D}_{\mathbf{V}}\mathcal{L}\right| \leq (2e(\mathfrak{d}+1)\beta)^{m+1}$$

which only depends on the degree of the graph $\mathfrak{d}$, inverse temperature β, and weight m. In order to establish this bound, we will first prove an intermediate lemma which bounds $|\mathcal{D}_{\mathbf{V}}\mathcal{L}|$ in terms of a graph constructed from the data of $\mathbf{V}$ and $\mathfrak{G}$.

To construct such an ancillary graph, consider a fixed $\mathbf{V}$. We define a graph $\mathrm{Gra}(\mathbf{V})$ from $\mathbf{V}$ as follows. The set of vertices of $\mathrm{Gra}(\mathbf{V})$ is taken to be

$$\mathrm{Mar}(\mathbf{V}) := \{(a,i) \in (\mathrm{Supp}\mathbf{V}) \times \mathbb{Z}_{>0} \,:\, 1 \leq i \leq \mu(a)\}\,,$$

where 'mar' stands for 'marked vertices'. Thus, $\mathrm{Gra}(\mathbf{V})$ has $\mu(a)$ vertices corresponding to each $a \in \mathbf{V}$, giving $|\mathbf{V}|$ vertices in total. We impose that in $\mathrm{Gra}(\mathbf{V})$, there is an edge between (a,i) and (a',i') if and only if $a = a'$ or $\mathrm{Supp}(E_a) \cap \mathrm{Supp}(E_{a'}) \neq \emptyset$ in the Hamiltonian. We have the following lemma from [**WA23**]:

Lemma 137 ([**WA23**]). *Letting* $\deg(v)$ *denote the number of neighbors of a vertex* $v \in \mathrm{Gra}(\boldsymbol{V})$, *we have the bound*

$$|\mathcal{D}_{\boldsymbol{V}}\mathcal{L}| \leq |\beta|^{|\boldsymbol{V}|} \prod_{v \in \mathrm{Mar}\,\boldsymbol{V}} (2\deg(v))\,.$$

This lemma follows by an elaborate graph coloring argument, which is explicated in a comprehensive manner in [**HKT22**]. For our purposes, this lemma is the main ingredient in our proposition of interest:

Proposition 138. *Let* $\boldsymbol{V}$ *be a cluster with weight* $|\boldsymbol{V}| = m+1 \geq 1$. *Then*

$$\left|\frac{1}{\boldsymbol{V}!}\mathcal{D}_{\boldsymbol{V}}\mathcal{L}\right| \leq (2e(\mathfrak{d}+1)\beta)^{m+1}\,.$$

To prove this, we will need one more elementary algebraic lemma:

Lemma 139. *Let* $\mu_1, ..., \mu_n \in \mathbb{R}_{>0}$ *and* $y_1, ..., y_n \in \mathbb{R}_{\geq 0}$. *Then*

$$\left(\frac{y_1}{\mu_1}\right)^{\mu_1} \cdots \left(\frac{y_n}{\mu_n}\right)^{\mu_n} \leq \left(\frac{y_1 + \cdots + y_n}{\mu_1 + \cdots + \mu_n}\right)^{\mu_1 + \cdots + \mu_n},$$

where equality holds when $\frac{y_j}{\mu_j} = \frac{\sum_i y_i}{\sum_j \mu_j}$ *for all* j.

Proof. The inequality holds trivially if any $y_i = 0$, so let us assume $y_i > 0$ for all i. If we take the log of both sides of the inequality and divide by $\sum_i \mu_i$ we find

$$\sum_{i=1}^{n} \frac{\mu_i}{\sum_j \mu_j} \log\left(\frac{y_i}{\mu_i}\right) \leq \log\left(\frac{y_1 + \cdots + y_n}{\mu_1 + \cdots + \mu_n}\right),$$

which is just Jensen's inequality applied to a concave function of the logarithm. $\square$

Now we turn to proving Proposition 138.

Proof of Proposition 138. From the definition of $\mathrm{Gra}(\mathbf{V})$ we have that for any $b \in \mathrm{Supp}\mathbf{V}$,

$$\deg((b,i)) = (\mu(b) - 1) + \sum_{a \in \Gamma(b)} \mu(a)\,, \tag{42}$$

where $\Gamma(b)$ is the set of neighbors of b in $\mathfrak{G}$ that appear in the cluster $\mathbf{V}$. Then we have the simple bound

$$\begin{aligned}\sum_{b\in\operatorname{Supp}\mathbf{V}}\deg((b,1)) &= \sum_{b\in\operatorname{Supp}\mathbf{V}}\left((\mu(b)-1)+\sum_{a\in\Gamma(b)}\mu(a)\right)\\ &\le m+\sum_{b\in\operatorname{Supp}\mathbf{V}}\sum_{a\in\Gamma(b)}\mu(a)\\ &\le m+\mathfrak{d}(m+1)\,, \end{aligned}\tag{43}$$

where in going from the first line to the second line we used $\sum_{b\in\operatorname{Supp}\mathbf{V}}(\mu(b)-1)\le\left(\sum_{b\in\operatorname{Supp}\mathbf{V}}\mu(b)\right)-1=(m+1)-1=m$, and in going from the second line to the third line we used $\sum_{b\in\operatorname{Supp}\mathbf{V}}\sum_{a\in\Gamma(b)}\mu(a)=\sum_{a\in\mathbf{V}}\mu(a)\,|\{b\in\operatorname{Supp}\mathbf{V}:b\in\Gamma(a)\}|\le\mathfrak{d}\sum_{a\in\mathbf{V}}\mu(a)=\mathfrak{d}(m+1)$. Using Lemma 137 we have

$$\begin{aligned}\frac{1}{\mathbf{V}!}|\mathcal{D}_{\mathbf{V}}\mathcal{L}| &\le \frac{(2\beta)^{m+1}}{\mathbf{V}!}\prod_{b\in\operatorname{Supp}\mathbf{V}}\prod_{i=1}^{\mu(b)}\deg((b,i))\\ &= (2\beta)^{m+1}\prod_{b\in\operatorname{Supp}\mathbf{V}}\frac{1}{\mu(b)!}\left(\mu(b)-1+\sum_{a\in\Gamma(b)}\mu(a)\right)^{\mu(b)}\\ &\le (2e\beta)^{m+1}\prod_{b\in\operatorname{Supp}\mathbf{V}}\left(\frac{\mu(b)-1+\sum_{a\in\Gamma(b)}\mu(a)}{\mu(b)}\right)^{\mu(b)}\\ &\le (2e\beta)^{m+1}\left(\frac{(1+\mathfrak{d})(m+1)}{m+1}\right)^{m+1}=(2e(\mathfrak{d}+1)\beta)^{m+1}, \end{aligned}$$

where in going from the first line to the second line we used (42), in going from the second line to the third line we used $u!\ge u^u e^{-u}$, and in going to the last line we used (43) and Lemma 139. $\square$

3.1.4. *Bounds on the sizes of polynomials*

Proposition 138 gives us a nice bound on the size of $\frac{1}{\mathbf{V}!}\mathcal{D}_{\mathbf{V}}\mathcal{L}$. Looking back to (39), we see that this should allow us to bound the sizes of the polynomials arising in the expansion of $\mathcal{L}$. We put the pieces together below to achieve such a bound, which comes from [**HKT22**]:

Theorem 140 (High-temperature Taylor expansion and size bounds)**.** *Let $H=\sum_{a\in[M]}\lambda_a E_a$ be a Hamiltonian with known traceless Hermitian terms E_a, $\|E_a\|\le 1$, and unknown coefficients $\lambda_a\in[-1,1]$. Let $\mathfrak{G}$ be its dual interaction graph of maximum degree $\mathfrak{d}$, and write $\rho_\beta=e^{-\beta H}/Z(\beta)$ with $Z(\beta)=\operatorname{tr}e^{-\beta H}$. Then for each $a\in[M]$ we have a (formal) β–series*

$$\langle E_a\rangle_\beta:=\operatorname{tr}(E_a\rho_\beta)=\frac{\operatorname{tr}(E_a)}{d}+\sum_{m=1}^{\infty}\beta^m\,p_m^{(a)}(\lambda_1,\dots,\lambda_M),\tag{44}$$

which holds as an identity whenever the series converges absolutely. Moreover, for every $m\in\mathbb{Z}_{>0}$ the coefficient $p_m^{(a)}$ satisfies:

(1) $p_m^{(a)} \in \mathbb{R}[\lambda_1, \ldots, \lambda_M]$ *is a homogeneous polynomial of degree m in the Hamiltonian coefficients.*
(2) (Locality of dependence) $p_m^{(a)}$ *can involve λ_b only if* $\mathrm{dist}_{\mathfrak{G}}(a,b) \le m$*.*
(3) (Number of monomials) $p_m^{(a)}$ *contains at most* $e\,\mathfrak{d}\,\big(1+e(\mathfrak{d}-1)\big)^m$ *monomials.*
(4) (Coefficient size) *The magnitude of the coefficient in front of any monomial of $p_m^{(a)}$ is at most* $(2e(\mathfrak{d}+1))^{m+1}(m+1)$*.*

If, in addition, each E_a is a Pauli string supported on at most L qubits, then after an $O(LM\mathfrak{d}\log\mathfrak{d})$ pre-processing (basis bookkeeping) the following algorithmic statements hold for every $m \ge 1$:

(A) The list of monomials appearing in $p_m^{(a)}$ can be enumerated *in time $O(m\,\mathfrak{d}\,C)$ where C is the number of monomials; in particular in time* $O\big(m\,\mathfrak{d}^2\big(1+e(\mathfrak{d}-1)\big)^m\big)$*.*
(B) The coefficient of any specific monomial can be computed exactly *(as a rational number) in time* $O\big(Lm^3 + 8^m m^5 \log^2 m\big) = (8^m + L)\,\mathrm{poly}(m)$*.*

A nice consequence of the theorem is as follows. Letting

$$\tau := \big(1+e(\mathfrak{d}-1)\big)\,\big(2e(\mathfrak{d}+1)\big) \le 2e^2(\mathfrak{d}+1)^2,$$

Items 3 and 4 above imply that the series (44) converges absolutely whenever $\beta < 1/\tau$; in particular it suffices that $\beta < \beta_c = \frac{1}{2e^2(\mathfrak{d}+1)^2}$.

We will provide a proof Items (1)-(4) of Theorem 140 using the ingredients we have previously derived, and then discuss (A) and (B) which are proved in [**HKT22**].

Proof of Items (1)-(4) in Theorem 140. Recall that $\mathcal{L}(\lambda) := \log \mathrm{tr}\exp\big(-\beta\sum_b \lambda_b E_b\big)$ so that $-\frac{1}{\beta}\,\partial_{\lambda_a}\mathcal{L}(\lambda) = \mathrm{tr}(E_a\rho_\beta)$. We recall that by analyticity of $\mathcal{L}$ in a neighborhood of the origin and the multivariate Taylor formula, we have the cluster (multi-index) expansion

$$\mathcal{L}(\lambda) = \sum_{m\ge 0}\sum_{\mathbf{V}:|\mathbf{V}|=m} \frac{\lambda^{\mathbf{V}}}{\mathbf{V}!}\mathcal{D}_{\mathbf{V}}\mathcal{L}\,. \tag{45}$$

Crucially, only *connected* clusters contribute by virtue of Lemma 132. Using $\langle E_a\rangle_\beta = -\frac{1}{\beta}\,\partial_{\lambda_a}\mathcal{L}(\lambda)$ and differentiating (45) termwise on a domain of absolute convergence, we obtain

$$\langle E_a\rangle_\beta = -\frac{1}{\beta}\sum_{m\ge 0}\sum_{\substack{\mathbf{V}:|\mathbf{V}|=m+1\\ a\in\mathbf{V}}} \frac{\partial_{\lambda_a}\lambda^{\mathbf{V}}}{\mathbf{V}!}\mathcal{D}_{\mathbf{V}}\mathcal{L}\,. \tag{46}$$

Every $\mathcal{D}_{\mathbf{V}}\mathcal{L}$ carries a factor $\beta^{|\mathbf{V}|}$, so after accounting for the overall factor $1/\beta$ in (46) we can regroup terms by β^m with $m = |\mathbf{V}|-1$, arriving at (44) with

$$p_m^{(a)}(\lambda) = (-1)^{m+1}\sum_{\substack{\mathbf{V}:|\mathbf{V}|=m+1\\ a\in\mathbf{V}}} \frac{\partial_{\lambda_a}\lambda^{\mathbf{V}}}{\mathbf{V}!}\,\frac{\mathcal{D}_{\mathbf{V}}\mathcal{L}}{\beta^{m+1}}. \tag{47}$$

Since each $\partial_{\lambda_a}\lambda^{\mathbf{V}}$ is a monomial of total degree m, Item (1) follows. Moreover, because only *connected* clusters $\mathbf{V}$ contribute, any cluster counted in (47) must lie within graph distance $\le m$ of a, giving Item (2).

The number of connected clusters of total weight $w = m + 1$ that contain a is at most $e\,\mathfrak{d}\,(1 + e(\mathfrak{d} - 1))^{w-1}$ by Proposition 136, hence we have Item (3). Since Proposition 138 gives the uniform bound

$$\left|\frac{1}{\mathbf{V}!}\,\mathcal{D}_{\mathbf{V}}\mathcal{L}\right| \le \big(2e(\mathfrak{d}+1)\,\beta\big)^{m+1},$$

and differentiating the monomial $\lambda^{\mathbf{V}}$ contributes at most a factor $(m + 1)$ since $|\partial_{\lambda_a}\lambda^{\mathbf{V}}| \le \mu(a) \le m + 1$. Dividing by β^{m+1} as in (47) yields Item (4): each monomial coefficient in $p_m^{(a)}$ has size at most $(2e(\mathfrak{d} + 1))^{m+1}(m + 1)$. □

Now let us briefly discuss Items (A) and (B) in Theorem 140. For Item (A), to enumerate all contributing monomials, one enumerates connected clusters of weight m rooted at a by a breadth-first, layer-by-layer procedure (see Algorithm 1, i.e. "`tails`" in [**HKT22**]). Given random-access to neighbors in $\mathfrak{G}$, the total time is $O(m\,\mathfrak{d}\,C)$ where C is the number of clusters (hence monomials), giving Item (A).

For Item (B), to compute an individual coefficient exactly, [**HKT22**] shows how to evaluate the needed cluster derivatives $\mathcal{D}_{\mathbf{V}}\mathcal{L}$ symbolically using faithful Pauli representations, in time $O(Lm^3 + 8^m m^5 \log^2 m)$.

3.2. Finding a solution using convexity

In the previous subsection we established a high-temperature expansion for the observables $\langle E_a\rangle_\beta = \operatorname{tr}(E_a\rho_\beta)$ and proved quantitative bounds on the size and locality of the resulting polynomials in Theorem 140. We now leverage those bounds to show that $\mathcal{L}(\lambda) = \log\operatorname{tr}\Big(e^{-\beta\sum_{a\in[M]}\lambda_a E_a}\Big)$ is *locally strongly convex* in the high-temperature regime. This convexity will be the key ingredient that lets us robustly invert the map from Hamiltonian coefficients to thermal expectations, and thereby learn the coefficients.

Fix a vector $x = (x_1, \ldots, x_M) \in [-1, 1]^M$. By Theorem 140 we may write

$$\langle E_a\rangle_\beta(x) = \sum_{m=1}^{\infty} \beta^m\, p_m^{(a)}(x)\,, \quad p_m^{(a)} \text{ homogeneous of degree } m,$$

where $p_m^{(a)}$ only depends on entries x_b with $\operatorname{dist}_{\mathfrak{G}}(a, b) \le m$, and its number and size of coefficients obey the bounds from Theorem 140 (Items (3)–(4)). In particular, letting

$$\tau := (1 + e(\mathfrak{d} - 1))\,(2e(\mathfrak{d} + 1)) \le 2e^2(\mathfrak{d} + 1)^2 \tag{48}$$

as before, the sum of absolute coefficients of $p_m^{(a)}$ is bounded by

$$\begin{aligned} c_m &= e\,\mathfrak{d}\,(1 + e(\mathfrak{d} - 1))^m\,(2e(\mathfrak{d} + 1))^{m+1}(m + 1) \\ &= 2e^2\mathfrak{d}(\mathfrak{d} + 1)\,\tau^{\,m}(m + 1)\,. \end{aligned} \tag{49}$$

For the learning task we will work with a *shifted, truncated* map $\mathcal{F} : [-1, 1]^M \to \mathbb{R}^M$ whose a-th component is

$$\mathcal{F}_a(x) := \sum_{m=0}^{m_{\max}} \beta^m\, p_m^{(a)}(x) = -\widehat{E}_a - \beta\, x_a + \beta^2 p_2^{(a)}(x) + \cdots + \beta^m p_{m_{\max}}^{(a)}(x)\,, \tag{50}$$

where $\widehat{E}_a$ is an estimate of $\langle E_a\rangle_\beta(\lambda)$ obtained from measurements (so we set $p_0^{(a)} := -\widehat{E}_a$), $p_1^{(a)}(x) = -x_a$ by a short computation, and $m_{\max}$ is a truncation order we