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 $\langle E_a\rangle_\beta = \mathrm{tr}(E_a\rho_\beta)$ and proved quantitative bounds on the size and locality of the resulting polynomials in Theorem 140. We now leverage those bounds to show that $\mathcal{L}(\lambda) = \log\mathrm{tr}\Big(e^{-\beta\sum_{a\in[M]}\lambda_a E_a}\Big)$ 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 $\mathrm{dist}_{\mathfrak{G}}(a,b) \le m$, 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 will later choose polylogarithmic in $1/(\beta\varepsilon)$. Our strategy will be to find an x such that $\mathcal{F}_a(x)$ is small for all $a \in [M]$; we will argue that if we can do then, then x is guaranteed to be closed to the true couplings λ by a convexity argument.

Here we will articulate our basic proof strategy. Let $J(x) = d\mathcal{F}(x)$ be the Jacobian of $\mathcal{F}$, namely

$$J_{ab}(x) = \frac{\partial}{\partial x_b}\mathcal{F}_a(x)\,.$$

Recall that the norm $\|\cdot\|_{\infty\to\infty}$ is defined by $\|A\|_{\infty\to\infty} = \max_{x\neq 0}\frac{\|Ax\|_\infty}{\|x\|_\infty}$. Then the idea is to use Newton iteration to find a x such that $\|x-\lambda\|_\infty \leq O(\varepsilon)$; doing so will involve bounding the size of the inverse Jacobian J^{-1} which plays an important role in Newton iteration, as well as the size of $\mathcal{F}(\lambda)$ which is the target value of $\mathcal{F}$.

To prepare for our Newton's method procedure, we will want to first establish the following facts:

(1) For suitable conditions on β and $\mathfrak{d}$, we have $\|J(x)^{-1}\|_{\infty\to\infty} \leq 2\beta^{-1}$ for all $m_{\max} \geq 1$.
(2) For any $\varepsilon > 0$, we can choose $m_{\max}$ sufficiently large (with suitable conditions on β and $\mathfrak{d}$) such that $\|\mathcal{F}(\lambda)\|_\infty \leq O(\beta\varepsilon)$.

For the first condition, we really only need the condition to hold for $m_{\max}$ sufficiently large, but in fact we will show that it holds for all $m_{\max} \geq 1$.

We will begin by establishing the first condition, and then treat the second. To this end, we have the following lemma.

Lemma 141. *Suppose that*

$$100e^6(\mathfrak{d}+1)^8\beta \leq 1\,. \tag{51}$$

Then for any $x \in [-1,1]^M$, we have $\|\mathbb{1}+\beta^{-1}J(x)\|_{\infty\to\infty} \leq \frac{1}{2}$ and $\|J(x)^{-1}\|_{\infty\to\infty} \leq 2\beta^{-1}$ for any $m_{\max} \geq 1$.

Proof. We note that if $\|\mathbb{1}+\beta^{-1}J\|_{\infty\to\infty} \leq \frac{1}{2}$, then since

$$J^{-1} = -\frac{1}{\beta}\frac{\mathbb{1}}{\mathbb{1}-(\mathbb{1}-\beta^{-1}J)} = -\frac{1}{\beta}\sum_{k=0}^{\infty}(\mathbb{1}+\beta^{-1}J)^k\,,$$

we would have

$$\|J^{-1}\|_{\infty\to\infty} \leq \beta^{-1}\sum_{k=0}^{\infty}\|\mathbb{1}+\beta^{-1}J\|_{\infty\to\infty}^k \leq 2\beta^{-1}\,.$$

Thus it suffices to show $\|\mathbb{1}+\beta^{-1}J(x)\|_{\infty\to\infty} \leq \frac{1}{2}$, or equivalently $\|\beta\mathbb{1}+J(x)\|_{\infty\to\infty} \leq \frac{\beta}{2}$, for our stated domain of β.

We observe that the Jacobian takes the form

$$J_{ab} = -\beta\,\delta_{ab} + O(\beta^2)\,,$$

and so $\beta\mathbb{1} + J = O(\beta^2)$. As such, we would like to bound the $O(\beta^2)$ remainder. Let $u = (u_1, ..., u_M)$ satisfy $|u_b| \le 1$ for all b, i.e. $\|u\|_\infty \le 1$. Then we have

$$\begin{aligned}
((J+\beta\mathbb{1})u)_a &= \sum_b (J+\beta\mathbb{1})_{ab} u_b \\
&= \sum_b u_b \left(\beta^2 \frac{\partial}{\partial x_b} p_2^{(a)}(x) + \cdots + \beta^{m_{\max}} \frac{\partial}{\partial x_b} p_{m_{\max}}^{(a)}(x) \right) \\
&= \sum_{k=2}^{m_{\max}} \beta^k \sum_{b\,:\,\mathrm{dist}_{\mathfrak{G}}(a,b)\le k} u_b \frac{\partial}{\partial x_b} p_k^{(a)}(x)\,,
\end{aligned}$$

where in going to the last line we have used Item (2) of Theorem 140. We observe that in the last sum, at each fixed k, the index b ranges over at most $1+\mathfrak{d}+\cdots+\mathfrak{d}^k \le (\mathfrak{d}+1)^k$ vertices of $\mathfrak{G}$. Now recall that each $p_k^{(a)}$ is a homogeneous polynomial of degree k, and that the sum of the absolute values of the coefficients is bounded by c_k in (49). Therefore $\left|\frac{\partial}{\partial x_b} p_k^{(a)}\right| \le k c_k$ in the domain of $\mathcal{F}$, and we have

$$\begin{aligned}
|((J+\beta\mathbb{1})u)_a| &\le \sum_{k=2}^{\infty} \beta^k (\mathfrak{d}+1)^k k c_k \\
&\le 2e^2(\mathfrak{d}+1)^2(\beta(\mathfrak{d}+1)\tau)^2 \sum_{k=2}^{\infty} (\beta(\mathfrak{d}+1)\tau)^{k-2} k(k+1) \\
&= 2e^2(\mathfrak{d}+1)^4\beta^2\tau^2 \left(\frac{6-6r+2r^2}{(1-r^3)}\bigg|_{r=\beta(\mathfrak{d}+1)\tau} \right) \\
&\le \frac{25}{2}\, e^2(\mathfrak{d}+1)^4\beta^2\tau^2\,.
\end{aligned}$$

In going from the second line to the third line we used $\beta(\mathfrak{d}+1)\tau < 1$, and in going to the last line we used $\beta(\mathfrak{d}+1)\tau \le \frac{1}{100}$. Since our u satisfying $\|u\|_\infty \le 1$ was arbitrary, we have obtained the bound $\|J+\beta\mathbb{1}\|_{\infty\to\infty} \le \frac{25}{2}\, e^2(\mathfrak{d}+1)^4\beta^2\tau^2$. Using $\tau \le 2e^2(\mathfrak{d}+1)^2$ from (48) and $100e^6(\mathfrak{d}+1)^8\beta \le 1$ from (51), we find our desired bound $\|J+\beta\mathbb{1}\|_{\infty\to\infty} \le \frac{\beta}{2}$. □

A nice consequence of the above lemma is the following convexity result:

Lemma 142. *If (51) holds, then $\nabla^{\otimes 2}\mathcal{L} \succeq \frac{\beta^2}{2}\mathbb{1}$, namely $\mathcal{L}$ is $(\frac{\beta^2}{2})$-strongly convex.*

Proof. Take $m_{\max} = \infty$ so that $\nabla^{\otimes 2}\mathcal{L} = -\beta J$, where we note that the Jacobian J is Hermitian. For a Hermitian matrix X, we have $\|X\| \le \|X\|_{\infty\to\infty}$, and so $\|\mathbb{1}+\beta^{-1}J\| \le \|\mathbb{1}+\beta^{-1}J\|_{\infty\to\infty} \le \frac{1}{2}$, implying that $\mathbb{1}+\beta^{-1}J \preceq \mathbb{1}/2$ and thus $\beta^{-1}J \preceq -\mathbb{1}/2$, which is equivalent to $\nabla^{\otimes 2}\mathcal{L} \succeq \frac{\beta^2}{2}\mathbb{1}$. □

Next let us show that if $m_{\max}$ is chosen to scale at least logarithmically in $1/(\beta\varepsilon)$, then we can arrange for $\|\mathcal{F}(\lambda)\|_\infty \le O(\beta\varepsilon)$. First we require the following lemma.

Lemma 143 (Estimating thermal expectations in parallel)**.** *For any $\varepsilon, \delta \in (0,1)$ there is a measurement procedure that (given independent copies of ρ_β) produces estimators $\widehat{E}_a$ such that*

$$\left|\widehat{E}_a - \langle E_a\rangle_\beta\right| \le \beta\,\varepsilon \qquad \textit{for all } a \in [M]$$

simultaneously with probability at least $1-\delta$*, using*

$$O\left(\frac{\mathfrak{d}}{\beta^2\varepsilon^2}\,\log\frac{M}{\delta}\right)$$

copies of ρ_β *and with time complexity*

$$O\left(\frac{N\,\mathfrak{d}}{\beta^2\varepsilon^2}\,\log\frac{M}{\delta}\right).$$

Proof. We first recall a standard fact: given a quantum state ρ and a Hermitian observable E with $\|E\|\le 1$, one can estimate $\mathrm{tr}(E\rho)$ to additive error ε_0 with success probability at least $1-\delta_0$ using $O\big(\log(1/\delta_0)/\varepsilon_0^2\big)$ independent copies of ρ. Indeed, measuring ρ in the eigenbasis of E yields an i.i.d. random variable in $[-1,1]$ whose expectation is $\mathrm{tr}(E\rho)$; Hoeffding bounds then give the stated sample complexity.

We now apply this in parallel to the family $\{E_a\}_{a\in[M]}$. Color the vertices of the dual interaction graph $\mathfrak{G}$ using at most $\mathfrak{d}+1$ colors (a greedy coloring suffices). By definition of $\mathfrak{G}$, all E_a belonging to a fixed color class act on disjoint sets of qubits. Consequently, on a single copy of ρ_β we can measure *all* E_a in that color class simultaneously: since each E_a is a Pauli string, it suffices to measure each qubit once in the appropriate single-qubit Pauli basis and multiply outcomes to obtain the eigenvalue of each E_a in the class.

Fix a color class and set the target accuracy per observable to $\varepsilon_0 := \beta\varepsilon$. By the single-observable estimate and a union bound over all a in the class, $O\big(\log(1/\delta_0)/\varepsilon_0^2\big)$ copies of ρ_β suffice to ensure that every $\widehat{E}_a$ in that class satisfies $|\widehat{E}_a-\langle E_a\rangle_\beta|\le\varepsilon_0$ with probability at least $1-\delta_0$. Repeating independently for each of the at most $\mathfrak{d}+1$ color classes, the total number of copies is

$$(\mathfrak{d}+1)\,O\Big(\frac{\log(1/\delta_0)}{\varepsilon_0^2}\Big)=O\left(\frac{\mathfrak{d}}{\beta^2\varepsilon^2}\,\log\frac{1}{\delta_0}\right).$$

Choosing $\delta_0:=\delta/M$ and applying a union bound across all M observables yields simultaneous accuracy $|\widehat{E}_a-\langle E_a\rangle_\beta|\le\beta\varepsilon$ for every $a\in[M]$ with probability at least $1-\delta$, and the stated copy complexity $O\big(\frac{\mathfrak{d}}{\beta^2\varepsilon^2}\log\frac{M}{\delta}\big)$ follows.

For the time complexity, note that each copy used in a given color round requires at most N single-qubit Pauli measurements (one per qubit), and there are $(\mathfrak{d}+1)\,O(\log(1/\delta_0)/\varepsilon_0^2)$ such copies overall. This gives time

$$O\left(N\,\frac{\mathfrak{d}}{\beta^2\varepsilon^2}\,\log\frac{M}{\delta}\right),$$

as claimed. ∎

This lemma tells us that we can set $|\widehat{E}_a-\langle E_a\rangle_\beta|\le O(\beta\varepsilon)$ for all a. With this in mind, we have the following.

Lemma 144. *Assume the high-temperature condition* (51)*. Let* τ *be as in* (48) *and set* $r:=\beta\tau$*. Suppose the empirical means obey* $\big|\widehat{E}_a-\langle E_a\rangle_\beta\big|\le\beta\varepsilon$ *for all* $a\in[M]$*. If the truncation order* $m_{\max}$ *in* (50) *satisfies*

$$(2r)^{\,m_{\max}+1}\le\frac{\beta\varepsilon}{4\,e^2\,\mathfrak{d}\,(\mathfrak{d}+1)}\,, \tag{52}$$

then $\|\mathcal{F}(\lambda)\|_\infty \le 2\beta\varepsilon$. *Equivalently, it suffices to take*

$$m_{\max} \ge \left\lceil \frac{\log\left(\dfrac{4\,e^2\,\mathfrak{d}\,(\mathfrak{d}+1)}{\beta\varepsilon}\right)}{\log\left(\dfrac{1}{2\beta\tau}\right)} \right\rceil - 1\,. \tag{53}$$

In particular, for constant $\mathfrak{d}$ *we have* $m_{\max} = O\big(\log(1/(\beta\varepsilon))\big)$.

Proof. By the definition (50) and the triangle inequality,

$$\big|\mathcal{F}_a(\lambda)\big| \le \big|\widehat{E}_a - \langle E_a\rangle_\beta\big| + \sum_{m>m_{\max}} \beta^m\,\big|p_m^{(a)}(\lambda)\big| \le \beta\varepsilon + \sum_{m>m_{\max}} \beta^m\,c_m\,.$$

Thus, with $r = \beta\tau$,

$$\sum_{m>m_{\max}} \beta^m\,c_m = 2e^2\,\mathfrak{d}(\mathfrak{d}+1) \sum_{m>m_{\max}} (m+1)\,r^{\,m}.$$

For $m \ge 1$ we may use $(m+1) \le 2^m$, where (since $2r < 1$)

$$\sum_{m>m_{\max}} (m+1)\,r^m \le \sum_{m>m_{\max}} (2r)^m = \frac{(2r)^{m_{\max}+1}}{1-2r}.$$

The high-temperature hypothesis (51) implies $r \ll 1$ (and hence $2r < 1$); in particular, $1/(1-2r) \le 2$. Therefore

$$\sum_{m>m_{\max}} \beta^m\,c_m \le 4e^2\,\mathfrak{d}(\mathfrak{d}+1)\,(2r)^{\,m_{\max}+1}.$$

Imposing (52) makes the right-hand side at most $\beta\varepsilon$, and hence $\big|\mathcal{F}_a(\lambda)\big| \le 2\beta\varepsilon$ for all a. Taking the maximum over a yields $\|\mathcal{F}(\lambda)\|_\infty \le 2\beta\varepsilon$.

Finally, solving (52) for $m_{\max}$ gives (53); since $2\beta\tau < 1$ under (51), the denominator is a positive constant when $\mathfrak{d}$ is constant, proving the claimed $O(\log(1/(\beta\varepsilon)))$ scaling. ☐

Finally, we show that we can efficiently find an x such that $\|x-\lambda\|_\infty \le 18\,\varepsilon$.

Theorem 145 (High-temperature learning via projected Newton-Raphson)**.** *Assume the high-temperature condition* (51)*. Suppose we are given estimates* $\{\widehat{E}_a\}_{a\in[M]}$ *of the thermal expectations* $\langle E_a\rangle_\beta$ *obeying* $\big|\widehat{E}_a - \langle E_a\rangle_\beta\big| \le \beta\,\varepsilon$ *for all* $a \in [M]$*. Moreover let us take* $\varepsilon \le \frac{1}{12}$*. Then there is a classical algorithm (a projected Newton-Raphson scheme with a truncated Neumann-series inverse) that outputs* $x \in [-1,1]^M$ *such that* $\|x-\lambda\|_\infty \le 18\,\varepsilon$ *in time* $O(\frac{ML}{\varepsilon}\mathrm{poly}(\mathfrak{d}, \log\frac{1}{\beta\varepsilon}))$*, where* L *is the maximum number of qubits on which any Hamiltonian term acts.*

Proof Sketch. Let us choose the judicious bound

$$m_{\max} \ge \left\lceil \frac{e}{e-1}\,\frac{1}{\log\left(\frac{1}{\beta\tau}\right)}\,\log\left(\frac{12e^2(\mathfrak{d}+1)^2}{\beta\varepsilon\log\left(\frac{1}{\beta\tau}\right)}\right)\right\rceil$$

which is compatible with our previous one. The **Newton-Raphson method** ordinarily entails an iteration like $x^{(t+1)} = x^{(t)} - (J^{-1}\mathcal{F})(x^{(t)})$, although to avoid computing the inverse of J we will instead consider an approximation $J(x)^{-1} \approx$

$\beta^{-1}\sum_{k=0}^{K-1}(\mathbb{1}+\beta^{-1}J(x))^k$ for a sufficiently large K that we will specify. Specifically, we consider the iteration

$$x^{(0)} = \vec{0}\,, \quad x^{(t+1)} = \mathrm{Proj}_{[-1,1]^M}\left[x^{(t)} + \beta^{-1}\sum_{k=1}^{K}(\mathbb{1}+\beta^{-1}J(x^{(t)}))^k \mathcal{F}(x^{(t)})\right]$$

where we have used

$$\mathrm{Proj}(u) := \begin{cases} 1 & \text{if } u \in (1,\infty) \\ u & \text{if } u \in [-1,1] \\ -1 & \text{if } u \in (-\infty,1) \end{cases}\,,$$

and take $K = \lceil \log_2\left(\frac{3}{2\varepsilon}\right)\rceil$.

Before analyzing the convergence of the iterations, let us examine the error $e^{(t)}$ between $J(x^{(t)})^{-1}\mathcal{F}(x^{(t)})$ and $\beta^{-1}\sum_{k=0}^{K-1}(\mathbb{1}+\beta^{-1}J(x^{(t)}))^k\mathcal{F}(x^{(t)})$. Specifically, we have

$$\begin{aligned} e^{(t)} &:= \left(-J(x^{(t)})^{-1} + \frac{1}{\beta}\sum_{k=0}^{K-1}(\mathbb{1}+\beta^{-1}J(x^{(t)}))^k\right)\mathcal{F}(x^{(t)}) \\ &= -\frac{1}{\beta}\sum_{k=K}^{\infty}(\mathbb{1}+\beta^{-1}J(x^{(t)})^k\mathcal{F}(x^{(t)}) \\ &= J^{-1}(x^{(t)})(\mathbb{1}+\beta^{-1}J(x^{(t)}))^K\mathcal{F}(x^{(t)})\,, \end{aligned}$$

which by Lemma 141 decays exponentially in K in the $\|\cdot\|_\infty$ norm. This will be useful for us shortly.

With the error $e^{(t)}$ under control, let us examine the convergence of $x^{(t)}$ under our Newton-Raphson iteration. Let $\mathcal{F}_a(s) : [0,1] \to \mathbb{R}$ by defined by $\mathcal{F}_a(s) := \mathcal{F}_a(x + s(\lambda - x))$. By Taylor's remainder theorem, there exists an $s' \in [0,1]$ such that

$$\underbrace{\mathcal{F}_a(1)}_{=\mathcal{F}_a(\lambda)} = \underbrace{\mathcal{F}_a(0)}_{=\mathcal{F}_a(x)} + (\partial_s\mathcal{F}_a)(0) + \frac{1}{2}(\partial_s^2\mathcal{F}_a)(s')\,.$$

Using $\partial_s = \sum_b(\lambda_b - x_b)\partial_b$ and setting $y^{(a)} := s'\lambda + (1-s')x$, we find

$$\mathcal{F}_a(\lambda) = \mathcal{F}_a(x) + \sum_b(\lambda_b - x_b)\underbrace{(\partial_b\mathcal{F}_a)(x)}_{=J_{ab}(x)} + \frac{1}{2}\sum_{b,c}(\lambda_b - x_b)(\lambda_c - x_c)(\partial_b\partial_c\mathcal{F}_a)(y^{(a)})\,.$$

Letting $\Delta^{(t)} := x^{(t)} - \lambda$ (and similarly $\Delta^{(t+1)} := x^{(t+1)} - \lambda$) where $\Delta_d^{(t)}$ denotes the dth coordinate, we have the following:

$$
\begin{aligned}
|\Delta_d^{(t+1)}| &= \left|\mathrm{Proj}_{[-1,1]}[(x - (J^{-1}\mathcal{F})(x) + e)_d] - \lambda_d\right| \\
&\leq \left|(x - (J^{-1}\mathcal{F})(x) + e)_d - \lambda_d\right| \\
&= \left|e_d^{(t)} + \Delta_d^{(t)} - \sum_a (J(x^{(t)})^{-1})_{da}\mathcal{F}_a(x^{(t)})\right| \\
&= \Bigg|e_d^{(t)} + \Delta_d^{(t)} - \sum_a J(x^{(t)})^{-1}_{da}\Bigg(\mathcal{F}_a(\lambda) - \sum_b (\lambda_b - x_b^{(t)})J_{ab}(x^{(t)}) \\
&\qquad\qquad - \frac{1}{2}\sum_{b,c}(\lambda_b - x_b^{(t)})(\lambda_c - x_c^{(t)})[\partial_b\partial_c\mathcal{F}_a](y^{(a)})\Bigg)\Bigg| \\
&= \Bigg|\left[e^{(t)} + \Delta^{(t)} - J(x^{(t)})^{-1}\mathcal{F}(\lambda) - J(x^{(t)})^{-1}J(x^{(t)})\Delta^{(t)}\right]_d \\
&\qquad\qquad + \frac{1}{2}\sum_{a,b,c} J(x^{(t)})^{-1}_{da}\Delta_b^{(t)}\Delta_c^{(t)}[\partial_b\partial_c\mathcal{F}](y^{(a)})\Bigg| \\
&= \Bigg|\left[J(x^{(t)})^{-1}\left((\mathbb{1} + \beta^{-1}J(x^{(t)}))^K\mathcal{F}(x^{(t)}) - \mathcal{F}(\lambda)\right)\right]_d \\
&\qquad\qquad + \frac{1}{2}\sum_{a,b,c} J(x^{(t)})^{-1}_{da}\Delta_b^{(t)}\Delta_c^{(t)}[\partial_b\partial_c\mathcal{F}_a](y^{(a)})\Bigg| . \qquad (54)
\end{aligned}
$$

We will bound each term in the last equation in turn. For the first part, we have

$$
\begin{aligned}
&\left|\left[J(x^{(t)})^{-1}\left((\mathbb{1} + \beta^{-1}J(x^{(t)}))^K\mathcal{F}(x^{(t)}) - \mathcal{F}(\lambda)\right)\right]_d\right| \\
&\leq \|J(x^{(t)})^{-1}\|_{\infty\to\infty}\left(\|\mathbb{1} + \beta^{-1}J(x^{(t)})\|^K_{\infty\to\infty}\|\mathcal{F}(x^{(t)})\|_\infty + \|\mathcal{F}(\lambda)\|_\infty\right) \\
&\leq 2\beta^{-1}\left(2^{-K}(2 + \beta\varepsilon) + 2\beta\varepsilon\right) \leq 6\varepsilon\,. \qquad (55)
\end{aligned}
$$

In going to the last line we have used Lemma 141 and Lemma 144, in conjunction with

$$
\begin{aligned}
|\mathcal{F}_a(x)| &\leq \left|\widehat{E}_a + \sum_{k=1}^{m_{\max}} \beta^k |p_k^{(a)}(x)|\right| \\
&\leq |\widehat{E}_a - \langle E_a\rangle_\beta| + \left|-\langle E_a\rangle_\beta + \sum_{k=1}^{m_{\max}} \beta^k |p_k^{(a)}(x)|\right| \\
&\leq \beta\varepsilon + 2\,.
\end{aligned}
$$

For the last term in (54), we have for all indices d the inequalities

$$
\begin{aligned}
&\left|\frac{1}{2}\sum_{a,b,c} J(x^{(t)})^{-1}_{da}\Delta^{(t)}_b\Delta^{(t)}_c[\partial_b\partial_c\mathcal{F}](y^{(a)})\right| \\
&\le \frac{1}{2}\|J(x^{(t)})^{-1}\|_{\infty\to\infty}\max_a\left|\sum_{b,c}\Delta^{(t)}_b\Delta^{(t)}_c[\partial_b\partial_c\mathcal{F}_a](y^{(a)})\right| \\
&\le \frac{1}{\beta}\max_a\sum_{k=0}^{\infty}\sum_{b,c}|\Delta^{(t)}_b\Delta^{(t)}_c|\,\beta^k\,|\partial_b\partial_c p^{(a)}_k(y)| \\
&\le \frac{1}{\beta}\sum_{k=0}^{\infty}\sum_{\substack{b,c:\\ \mathrm{dist}_{\mathfrak{G}}(b,a)\le k\\ \mathrm{dist}_{\mathfrak{G}}(c,a)\le k}}\|\Delta^{(t)}\|^2_\infty\,\beta^k k(k-1)c_k \\
&\le \frac{1}{\beta}\sum_{k=0}^{\infty}(\mathfrak{d}+1)^{2k}\|\Delta^{(t)}\|^2_\infty\,\beta^k k(k-1)c_k \\
&= \frac{12e^2}{\beta}\|\Delta^{(t)}\|^2_\infty(\mathfrak{d}+1)^2\frac{(\beta(\mathfrak{d}+1)^2\tau)^2}{(1-\beta(\mathfrak{d}+1)^2\tau)^4} \\
&\le \frac{25}{2}\,e^2\beta(\mathfrak{d}+1)^6\tau^2\|\Delta^{(t)}\|^2_\infty\,,
\end{aligned}
\tag{56}
$$

where in going to the second-to-last line we have used that $\beta(\mathfrak{d}+1)^2\tau<1$ and in going to the last line we have used that $\beta\mathfrak{d}^2\tau\le 1-\left(\frac{24}{25}\right)^{1/4}$. Putting together (55) and (56) we find

$$
\|\Delta^{(t+1)}\|_\infty \le 6\varepsilon + \frac{25}{2}\,e^2\beta(\mathfrak{d}+1)^6\tau^2\|\Delta^{(t)}\|^2_\infty\,.
$$

By solving the recursion, one can show that so long as $\|\Delta^{(0)}\|_\infty \le \frac{1}{25e^2\beta(\mathfrak{d}+1)^6\tau^2}\le 1$, we achieve $\|x^{(T)}-\lambda\|_\infty\le 18\varepsilon$ for

$$
T=\lceil -\log_2(300e^6(\mathfrak{d}+1)^{10}\beta\varepsilon)\rceil\,.
$$

Finally, let us sketch the runtime bound. For each $a\in[M]$, the truncated series $\mathcal{F}_a(x)=\sum_{k=0}^m\beta^k p^{(a)}_k(x)-\widehat{E}_a$ is a degree-m polynomial whose support is contained in the radius-k neighborhoods of a in $\mathfrak{G}$; the number of contributing terms at order k is at most $\mathrm{poly}(\mathfrak{d})\,(\mathfrak{d}+1)^k$ and each coefficient can be evaluated in time $O(L\,\mathrm{poly}(k))$. Hence, evaluating all M coordinates of $F(x)$ and forming (or applying) the nonzeros of the sparse Jacobian $J(x)$ at a given point x costs

$$
O\Big(M\,L\,\mathrm{poly}(\mathfrak{d})\sum_{k=0}^{m}(\mathfrak{d}+1)^k\Big)=O\big(M\,L\,\mathrm{poly}(\mathfrak{d})\,(\mathfrak{d}+1)^{O(m)}\big).
$$

One Newton step uses the truncated Neumann-series inverse $\beta^{-1}\sum_{k=0}^{K-1}\big(\mathbb{1}+\beta^{-1}J(x)\big)^k$, which requires K sparse matrix–vector multiplies with $J(x)$, and thus has cost $O\big(K\,M\,L\,\mathrm{poly}(\mathfrak{d})\,(\mathfrak{d}+1)^{O(m)}\big)$ at iteration $x=x^{(t)}$. The projection $\mathrm{Proj}_{[-1,1]^M}$ adds only $O(M)$ time. With T Newton iterations in total, the overall runtime is

$$
O\Big((K+1)\,TM\,L\,\mathrm{poly}(\mathfrak{d})\,(\mathfrak{d}+1)^{O(m)}\Big).
$$

Substituting in our parameter choices yields the stated time complexity:

$$O\left(\frac{M\,L}{\varepsilon}\mathrm{poly}\left(\mathfrak{d}, \log\frac{1}{\beta\varepsilon}\right)\right).$$

□

To summarize, we have succeeded in establishing that, for suitable β, $\mathfrak{d}$, and $m_{\max}$, we have $\|x-\lambda\|_\infty \le O(\varepsilon)$. Below we will put together all of our results thus far to get the final, overarching algorithm and associated bounds.

3.3. Putting the bounds together

We can combine all of the results above to get the main result of [**HKT22**]. We recapitulate some of the notation we have collected along the way.

Theorem 146 (Learning from high-temperature Gibbs states)**.** *Let $H=\sum_{a\in[M]}\lambda_a E_a$ be a* low-intersection *Hamiltonian on N qubits: each non-identity Pauli term E_a acts on at most $L=O(1)$ qubits and the dual interaction graph has maximum degree $\mathfrak{d}=O(1)$. Fix inverse temperature $\beta>0$ obeying the high-temperature condition* (51) *(equivalently $\beta<\beta_c(\mathfrak{d})$ for a universal constant $\beta_c>0$ depending only on $\mathfrak{d}$), and let $\rho_\beta=e^{-\beta H}/\mathrm{tr}(e^{-\beta H})$.*

For any $\varepsilon\in(0,\frac{1}{12})$ and failure probability $\delta\in(0,1)$, there is a classical algorithm which, given independent copies of ρ_β, outputs $\widehat{\lambda}\in[-1,1]^M$ satisfying

$$\|\widehat{\lambda}-\lambda\|_\infty \le 18\varepsilon$$

with probability at least $1-\delta$, using

$$S_\infty = O\left(\frac{\mathfrak{d}}{\beta^2\varepsilon^2}\,\log\frac{M}{\delta}\right)$$

copies of ρ_β. In particular, when $\mathfrak{d}=O(1)$ and $M=\Theta(N)$, this is

$$S_\infty = O\left(\frac{\log N}{\beta^2\varepsilon^2}\right).$$

Consequently, to achieve ℓ_2-error $\|\widehat{\lambda}-\lambda\|_2\le\varepsilon$ it suffices to use

$$S_2 = O\left(\frac{M}{\beta^2\varepsilon^2}\,\log\frac{M}{\delta}\right) = O\left(\frac{N}{\beta^2\varepsilon^2}\,\log\frac{N}{\delta}\right).$$

The total running time is linear in the sample size (i.e. $O(SN)$ where S is the number of copies used), up to polylogarithmic factors in $1/(\beta\varepsilon)$.

Proof. Assume (51) and let τ be as in (48). We can estimate all thermal expectations in parallel (via Lemma 143) to obtain $\{\widehat{E}_a\}_{a\in[M]}$ with $|\widehat{E}_a-\langle E_a\rangle_\beta|\le\beta\varepsilon$ for every a using $S_\infty=O\big(\frac{\mathfrak{d}}{\beta^2\varepsilon^2}\log\frac{M}{\delta}\big)$ copies of ρ_β, with success probability $\ge 1-\delta$.

Define $\mathcal{F}$ as in (50) and choose the truncation order $m_{\max}$ as in (53). Then Lemma 144 gives $\|\mathcal{F}(\lambda)\|_\infty \le 2\beta\varepsilon$. By the high-temperature conditioning in Lemma 141, we have

$$\|\mathbb{1}+\beta^{-1}J(x)\|_{\infty\to\infty}\le\tfrac12 \quad\text{and}\quad \|J(x)^{-1}\|_{\infty\to\infty}\le 2\beta^{-1}, \quad \text{for all } x\in[-1,1]^M.$$

We run the projected Newton–Raphson update with truncated Neumann inverse in Theorem 145 from $x^{(0)}=\vec{0}$ and with $K=\big\lceil \log_2\big(\frac{3}{2\varepsilon}\big)\big\rceil$. The one-step analysis yields the recursion $\|\Delta^{(t+1)}\|_\infty\le 6\varepsilon + C\,\beta\,\|\Delta^{(t)}\|_\infty^2$ with $C=\frac{25}{2}e^2(\mathfrak{d}+1)^6\tau^2$.

Solving this recursion with $T = \lceil -\log_2 \big(300e^6(\mathfrak{d}+1)^{10}\beta\varepsilon\big)\rceil$ gives $\|x^{(T)}-\lambda\|_\infty \le 18\varepsilon$. We set $\widehat{\lambda} := x^{(T)}$ to obtain the claimed accuracy with probability $\ge 1-\delta$.

The sample bound is exactly that of Lemma 143, and for $\mathfrak{d} = O(1)$ and $M = \Theta(N)$ it simplifies to $S_\infty = O\big(\frac{\log N}{\beta^2\varepsilon^2}\big)$. The ℓ_2 statement follows by targeting $\|\widehat{\lambda}-\lambda\|_\infty \le \varepsilon/\sqrt{M}$, which replaces ε by $\varepsilon/\sqrt{M}$ in Lemma 143, yielding $S_2 = O\big(\frac{M}{\beta^2\varepsilon^2}\log\frac{M}{\delta}\big)$. The runtime is $O(S_\infty N)$ for data collection plus $O\big(\frac{ML}{\varepsilon}\,\mathrm{poly}(\mathfrak{d}, \log\frac{1}{\beta\varepsilon})\big)$ for classical postprocessing, which is linear in the sample size up to polylogarithmic factors in $1/(\beta\varepsilon)$. $\square$