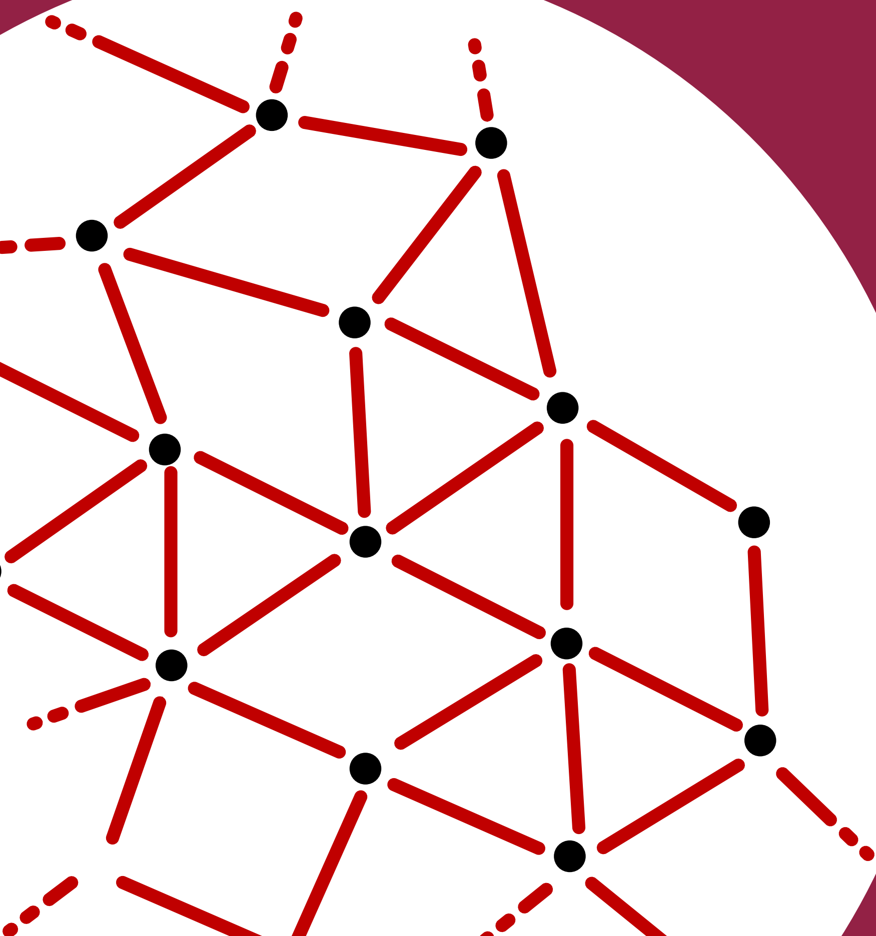
Sample-efficient learning of quantum many-body systems

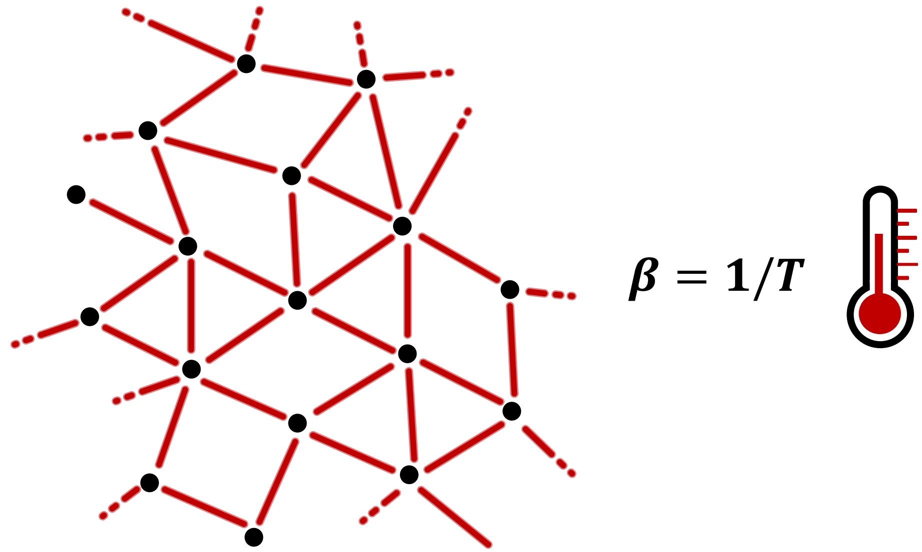


Mehdi Soleimanifar (MIT)

(arxiv: 2004.07266, FOCS'20)

Joint work with
Anurag Anshu (UC Berkeley)
Srinivasan Arunachalam (IBM)
Tomotaka Kuwahara (RIKEN)

Setup and problem statement



Hamiltonian $H(\mu) = \sum_{k=1}^{m} \mu_k E_k$, E_k **local basis,** $[E_k, E_\ell] \neq 0$

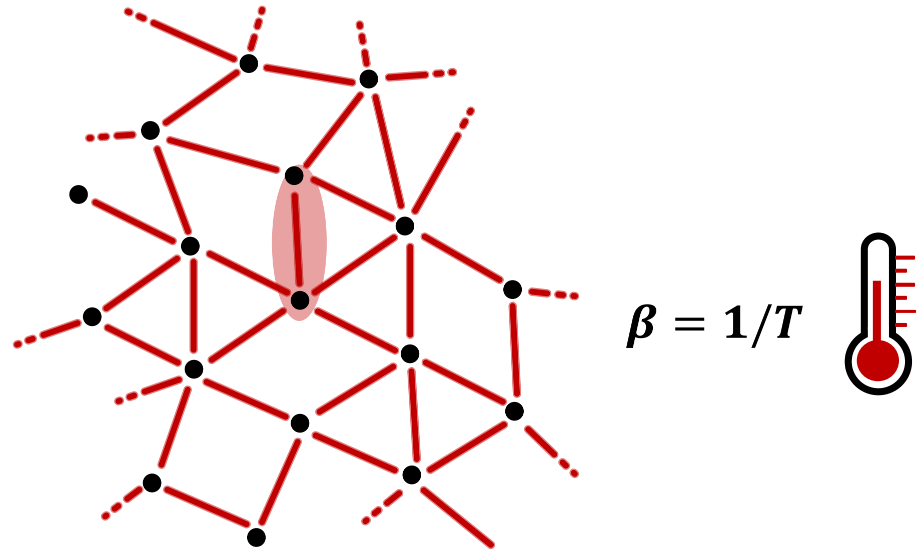
e.g., $\otimes$ of Pauli operators

Interaction coefficients $\mu = (\mu_1, \mu_2, \ldots, \mu_m)$, $|\mu_k| \leq 1$, $m = O(n)$

Gibbs state $\rho_\mu = \frac{1}{Z(\mu)} \exp\left(-\beta H(\mu)\right)$

Partition function $Z(\mu) = \mathrm{Tr}[e^{-\beta H(\mu)}]$

This talk:
Learn Hamiltonian $H(\mu)$ from local measurements



Hamiltonian $H(\mu) = \sum_{k=1}^{m} \mu_k E_k$, $\quad$ E_k **local basis,** $[E_k, E_\ell] \neq 0$

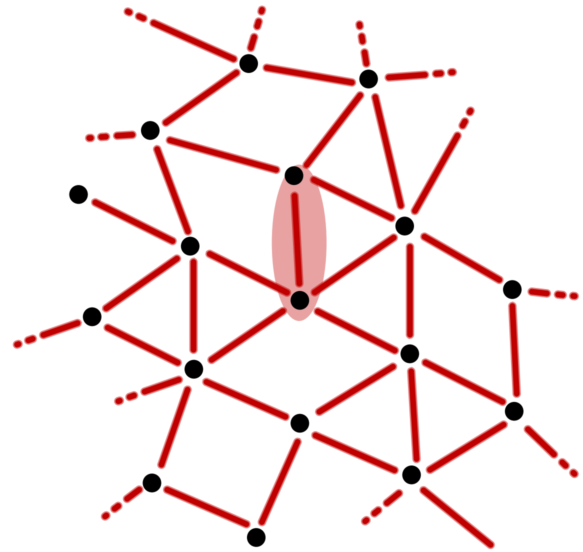
e.g., $\otimes$ *of Pauli operators*

Interaction coefficients $\quad \mu = (\mu_1, \mu_2, \ldots, \mu_m)$, $\quad |\mu_k| \leq 1,\ m = O(n)$

Gibbs state $\quad \rho_\mu = \frac{1}{Z(\mu)} \exp\left(-\beta H(\mu)\right)$

Partition function $\quad Z(\mu) = \mathrm{Tr}\left[e^{-\beta H(\mu)}\right]$

This talk:
Learn $(\mu_1, \mu_2, \dots, \mu_m)$ from local measurements



$\beta = 1/T$

Hamiltonian $H(\mu) = \sum_{k=1}^{m} \mu_k E_k$, E_k **local basis,** $[E_k, E_\ell] \neq 0$

e.g., $\otimes$ of Pauli operators

Interaction coefficients $\mu = (\mu_1, \mu_2, \dots, \mu_m)$, $|\mu_k| \leq 1$, $m = O(n)$

Gibbs state $\rho_\mu = \frac{1}{Z(\mu)} \exp(-\beta H(\mu))$

Partition function $Z(\mu) = \mathrm{Tr}[e^{-\beta H(\mu)}]$

Why care about Quantum Hamiltonian learning?

Many previous results, but no rigorous performance guarantee
[BAL19, BGP+20, WGFC14, EHF19, WPS+17, ...]

Verification of quantum devices:

How to verify quantum algorithms involving Gibbs states?

e.g., quantum SDP solvers, quantum annealing, finite T simulations

Many-body physics:

Can we test our theories for interacting quantum systems?

e.g., interactions in newly synthesized materials or cold atom setups

Quantum Machine Learning:

Can ML techniques help learn quantum data?

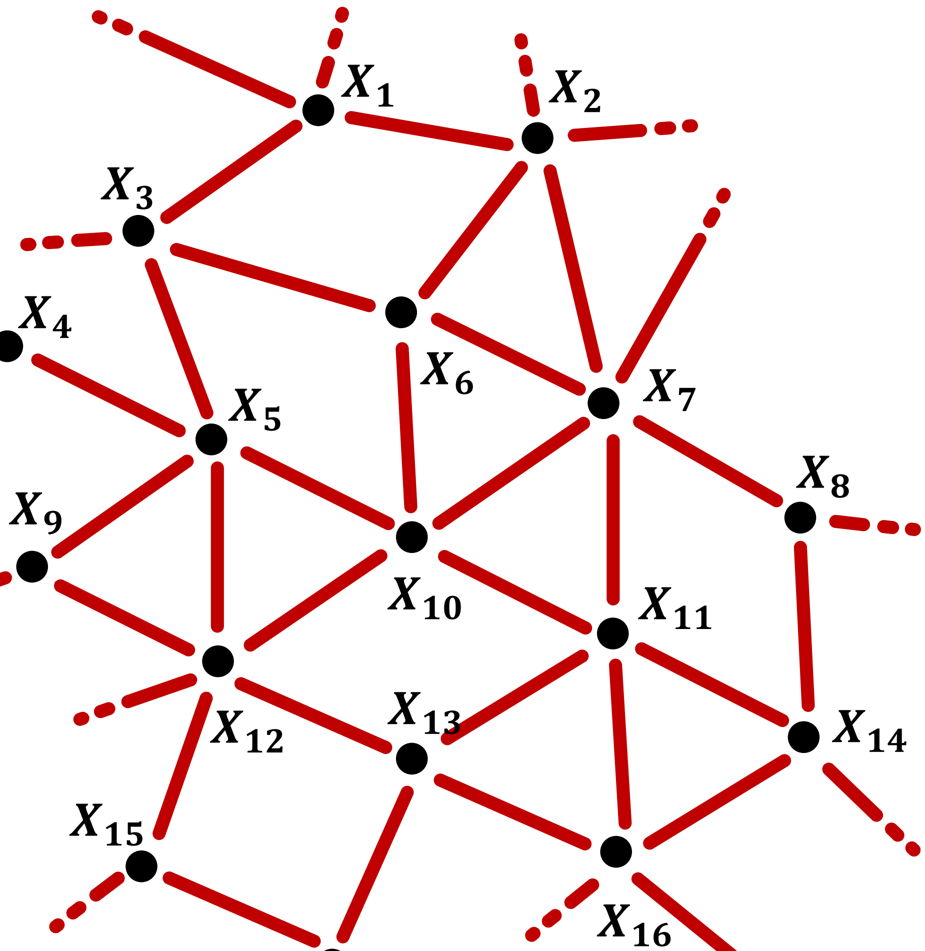
Connection to Machine Learning

$$X = (X_1, X_2, \ldots, X_n) \in \{+1, -1\}^n$$

Vertices of a bounded degree graph

$$X \sim p_\theta$$

Learn θ from *i.i.d.* samples of p_θ



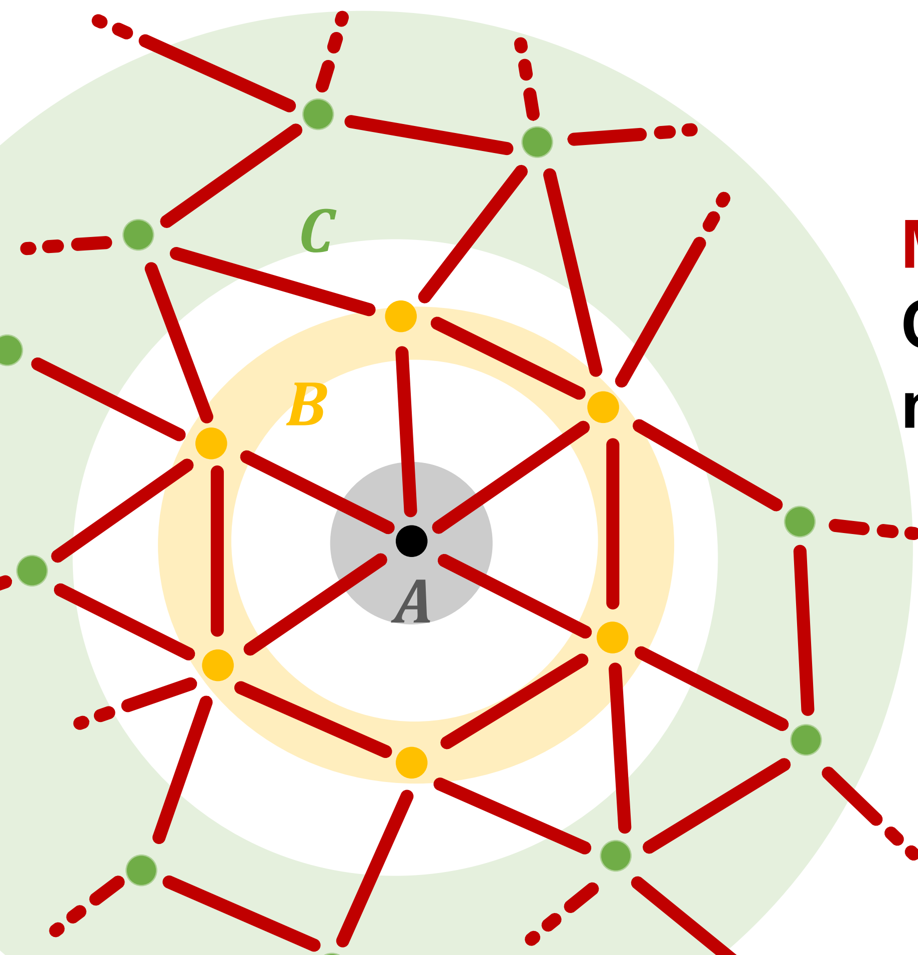
Ising Model
Boltzmann Machine

$X = (X_1, X_2, \dots, X_n) \in \{+1, -1\}^n$

Vertices of a bounded degree graph

Gibbs Distribution $p_\theta(X = x) = \frac{1}{Z} \exp(\sum_{k\sim\ell} \theta_{k\ell} x_k x_\ell)$

Hammersley-Clifford Theorem



Markov Property
Correlations mediated through neighboring variables

$I(A : C | B) = 0$

Learning algorithms for these models with efficient time/sample complexity

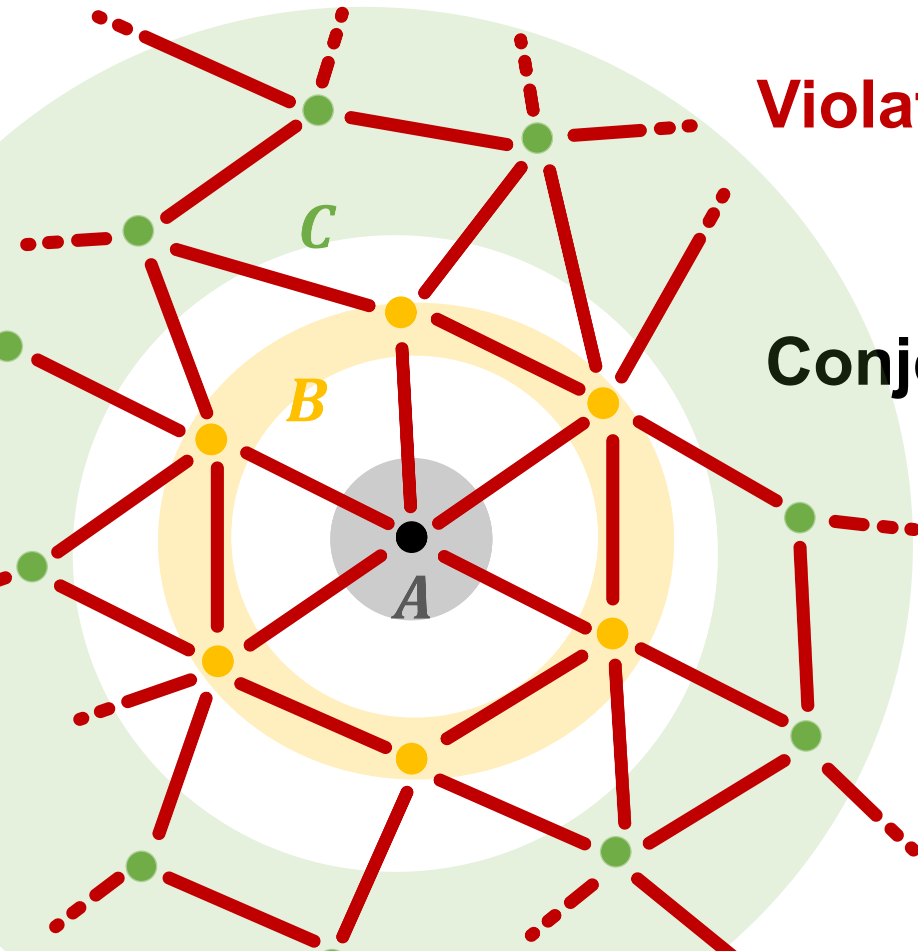
[Bresler15, KM17, VMLC16,…]

Learning Quantum Interactions

Quantum state $\rho \geq 0, \mathrm{Tr}[\rho] = 1$

Quantum Gibbs state $\rho_\mu = \frac{1}{Z(\mu)} \exp(-\beta \sum_k \mu_k E_k)$

Spatially Local Hamiltonian $H(\mu) = \sum_{k=1}^{m} \mu_k E_k$



Violates exact Markov property

$I(A:C|B) \neq 0$ [LP08]

Conjectured to obey [KB19, KKB19]

$I(A:C|B) \leq e^{-O(\mathrm{dist}(A,C))}$

Can we obtain unconditional algorithms for learning quantum Hamiltonians?

Efficient Sample Complexity

Our main result:

$\widetilde{O}\left(\frac{e^{\mathrm{poly}(\beta)}}{\mathrm{poly}(\beta)}\frac{m^3}{\varepsilon^2}\right)$ *i.i.d.* copies of ρ_μ suffices to obtain estimate $\hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_m)$ s.t. $\|\hat{\mu} - \mu\|_2 \le \varepsilon$

Hamiltonian $H(\mu) = \sum_{k=1}^{m} \mu_k E_k$ $\sqrt{\sum_{k=1}^{m} (\hat{\mu}_k - \mu_k)^2}$

We also show a lower bound of $\widetilde{\Omega}\left(\sqrt{m}/\varepsilon\right)$ for the # of samples

Main ideas in our proof

Sufficient statistics

A function of the input data that contains all the information about the unknown parameters

e.g., sample mean and variance in Gaussian distributions

(Thermal averages)

Local expectations $e_k = \mathrm{Tr}[\rho_\mu E_k], \quad k \in [m]$

uniquely determine $\rho_\mu = \frac{1}{Z(\mu)} \exp(-\beta \sum_k \mu_k E_k)$

[Jaynes57, KS14, BKL+17]

Maximum entropy optimization

$$\rho_\mu = \underset{\text{states } \sigma}{\mathrm{argmax}} \quad S(\sigma) = -\mathrm{Tr}[\sigma \log \sigma] \text{ von Neumann entropy}$$

$$\text{s.t.} \quad \mathrm{Tr}[\sigma E_k] = e_k, \qquad k \in [m]$$

Sufficient statistics

A function of the input data that contains all the information about the unknown parameters

e.g., sample mean and variance in Gaussian distributions

(Thermal averages)

Local expectations $e_k = \mathrm{Tr}[\rho_\mu E_k]$, $k \in [m]$

uniquely determine $\rho_\mu = \frac{1}{Z(\mu)} \exp(-\beta \sum_k \mu_k E_k)$

[Jaynes57, KS14, BKL+17]

Maximum entropy optimization

$$\hat{\rho} = \operatorname*{argmax}_{\text{states } \sigma} \quad S(\sigma) = -\mathrm{Tr}[\sigma \log \sigma] \text{ von Neumann entropy}$$

$$\text{s.t.} \quad \mathrm{Tr}[\sigma E_k] = \hat{e}_k, \qquad k \in [m]$$

Empirical Values

Dual program

$$\max_{\text{states } \sigma} \quad S(\sigma)$$

$$\text{s.t.} \quad \mathrm{Tr}[\sigma\, E_k] = e_k, \qquad k \in [m]$$

Dual

Partition Function

$$Z(\lambda) = \mathrm{Tr}[\mathrm{e}^{-\beta \sum_k \lambda_k E_k}]$$

$$\mu = \mathrm{argmin}_{(\lambda_1,\ldots,\lambda_m)} \quad \log Z(\lambda) + \beta \sum_k \lambda_k\, e_k \longrightarrow \mathrm{Tr}[\rho_\mu\, E_k]$$

Dual program

$$\max_{\text{states } \sigma} \quad S(\sigma)$$

$$\text{s.t.} \quad \mathrm{Tr}[\sigma\, E_k] = \hat{e}_k, \qquad k \in [m]$$

Dual

Empirical Values

$$\hat{\mu} = \mathrm{argmin}_{(\lambda_1,\ldots,\lambda_m)} \quad \log Z(\lambda) + \beta \textstyle\sum_k \lambda_k\, \hat{e}_k$$

Error in interactions vs **Error in local expectations**

$$\|\hat{\mu} - \mu\|_2 \qquad\qquad \|\hat{e} - e\|_2$$

Dual program

$$\max_{\text{states } \sigma} \quad S(\sigma)$$

$$\text{s.t.} \quad \mathrm{Tr}[\sigma\, E_k] = \hat{e}_k, \qquad k \in [m]$$

⇕ Dual

$$\hat{\mu} = \mathrm{argmin}_{(\lambda_1,\dots,\lambda_m)} \quad \log Z(\lambda) + \beta \textstyle\sum_k \lambda_k\, \hat{e}_k$$

Toy model

$$-\frac{b}{2a} = \mathrm{argmin} \quad ax^2 + bx + c \qquad a > 0$$

Dual program

$$\max_{\text{states } \sigma} \quad S(\sigma)$$

$$\text{s.t.} \quad \mathrm{Tr}[\sigma E_k] = \hat{e}_k, \qquad k \in [m]$$

Dual

$$\hat{\mu} = \mathrm{argmin}_{(\lambda_1,\ldots,\lambda_m)} \quad \log Z(\lambda) + \beta \sum_k \lambda_k \hat{e}_k$$

Toy model

$$-\frac{\hat{b}}{2a} = \mathrm{argmin} \quad ax^2 + \hat{b}x + c \qquad a > 0$$

$$|\hat{b} - b| \ll 2a$$

Convexity determines allowed statistical error

Dual program

$$\max_{\text{states } \sigma} \quad S(\sigma)$$

$$\text{s.t.} \quad \mathrm{Tr}[\sigma\, E_k] = \hat{e}_k, \qquad k \in [m]$$

Dual

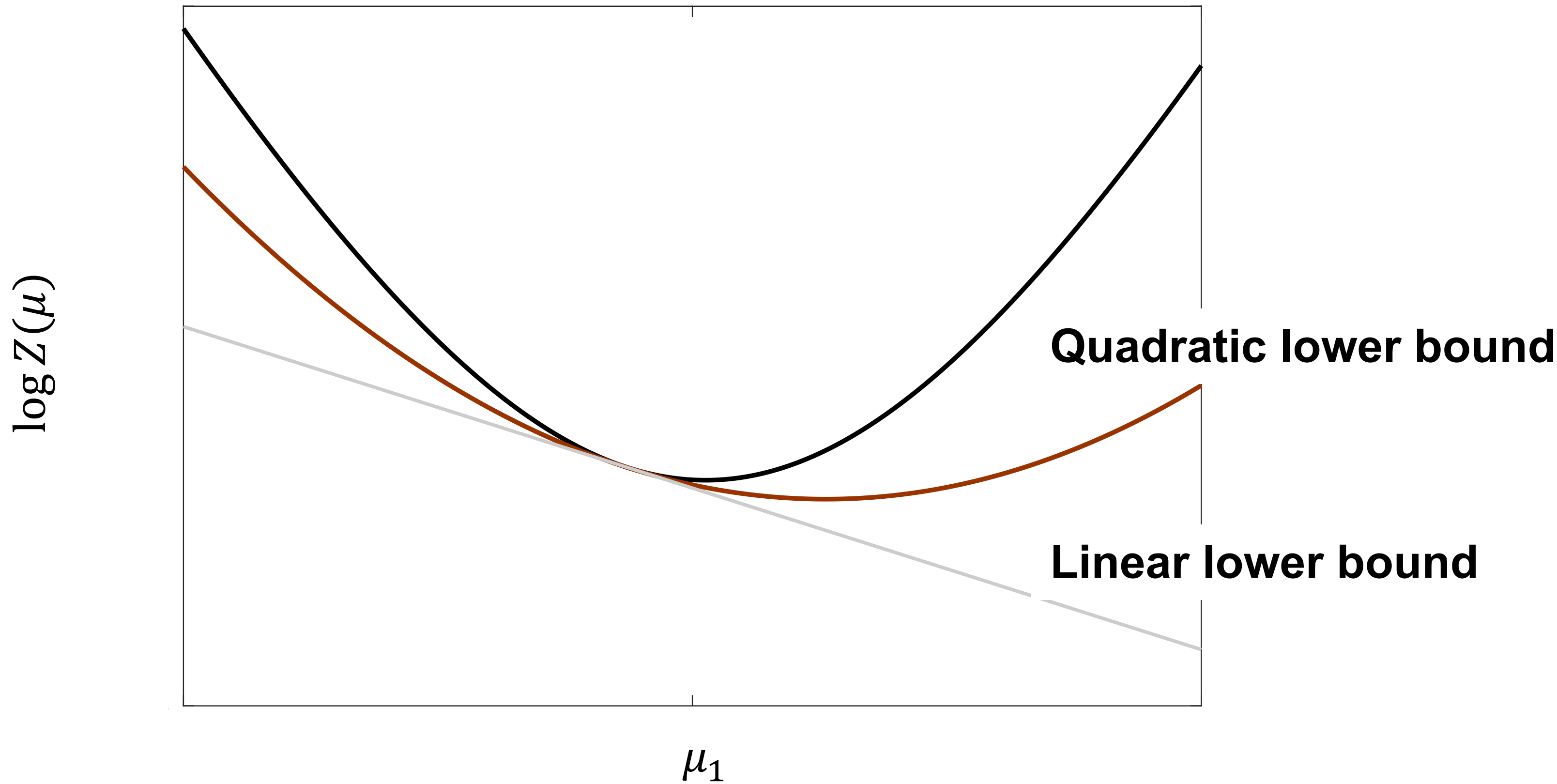
$$\hat{\mu} = \mathrm{argmin}_{(\lambda_1,\ldots,\lambda_m)} \quad \boxed{\log Z(\lambda)} + \beta \sum_k \lambda_k\, \hat{e}_k$$

Stochastic Convex Optimization

$f(x)$ is α-strongly convex if

$$f(x) \geq f(x_0) + \nabla f(x_0)(x - x_0) + \frac{1}{2}\alpha \|x - x_0\|_2^2$$

Strong Convexity



Our main contribution: $\lambda_{\min}\left(\nabla^2 \log Z\right) \geq \Omega(1/m)$

This implies: $\|\hat{\mu} - \mu\|_2 \leq O(m)\ \|\hat{e} - e\|_2 \ \rightarrow N = \tilde{O}\left(m^3/\varepsilon^2\right)$

Error in interactions **Error in local expectations**

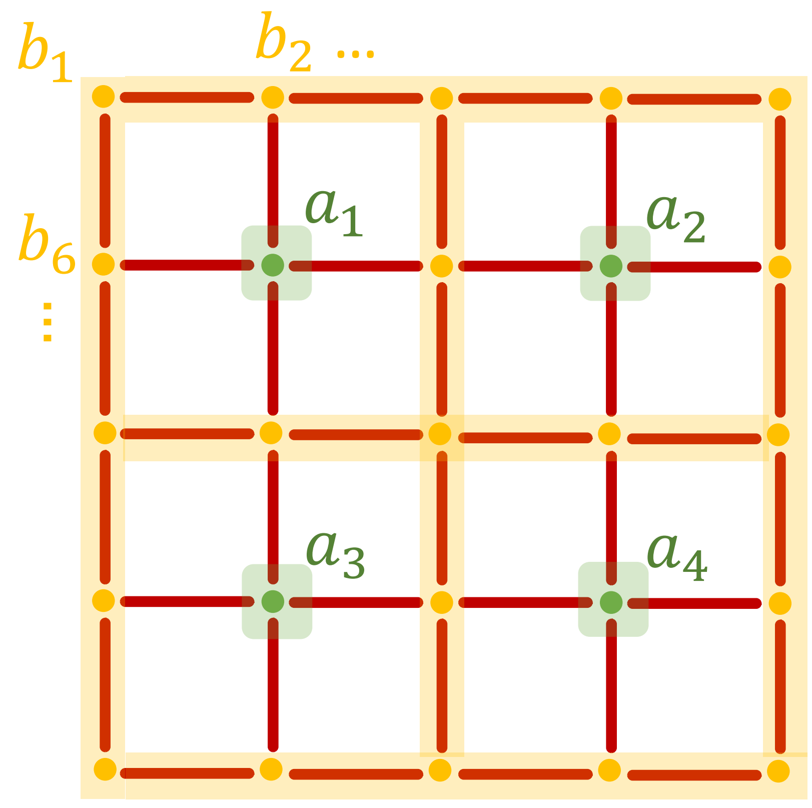
Strong convexity of $\log Z$ (classical case)

$$v^T \cdot \nabla^2 \log Z \cdot v \geq \Omega(1) \cdot \|v\|_2^2$$

Proof

$$\sum_{k,\ell} v_k v_\ell \underbrace{\frac{\partial^2}{\partial \mu_k \partial \mu_\ell} \log Z}_{= \mathrm{Cov}[E_k, E_\ell]} = \mathrm{Var}[\textstyle\sum_k v_k E_k]$$

$$\mathrm{Var}[\textstyle\sum_k v_k E_k]$$
$$\geq \mathbb{E}_B[\,\mathrm{Var}[\textstyle\sum_k v_k E_k | B]\,]$$
$$\geq \textstyle\sum_{a \in A} \mathbb{E}_B[\,\underbrace{\mathrm{Var}[\textstyle\sum_{k \sim a} v_k E_k | B]}_{\text{Local terms}}\,]$$
$$\geq \Omega(1) \|v\|_2^2$$



Relies on Markov property
and $[E_k, E_\ell] = 0$

Proof of strong convexity of $\log Z$ (quantum case)

$$\sum_{k,\ell} v_k v_\ell \frac{\partial^2}{\partial \mu_k \partial \mu_\ell} \log Z \geq \mathrm{Var}\left[\sum_{k=1}^{m} v_k \widetilde{E}_k\right]$$

$$\gtrapprox \max_k (v_k^2) \cdot \mathrm{Var}\left[\widetilde{W}_{k_0}\right]^{O(1)}$$

$$\geq \Omega(1/m) \|v\|_2^2$$

$\widetilde{W}_{k_0}, \widetilde{E}_k$ quasi-local operator

Relies on
Quantum belief propagation [Hastings'07]

Connecting global and local properties
of Quantum systems [AKL'16]

Results:

Sample complexity of quantum Hamiltonian learning $O\left(\frac{m^3}{\varepsilon^2}\right)$

We also show a lower bound of $\Omega\left(\sqrt{m}/\varepsilon\right)$

Time complexity

Maximum entropy optimization

$$\rho_\mu = \operatorname*{argmax}_{\text{states } \sigma} \; S(\sigma)$$

$$\text{s.t.} \quad \mathrm{Tr}[\sigma E_k] = e_k, \qquad k \in [m]$$

NP-hard in the worst case
Requires computing $\log Z(\mu)$ [Montanari15]

Approximate versions exist via
mean field approximation, or pseudo-likelihood

Efficient algorithms at **high temperatures** $T > T_c$

[Harrow, Mehraban, Soleimanifar'19], [KKB'19], [MH'20]

Open questions

1) Close the upper and lower bound for sample complexity?

2) Can we learn μ in ℓ_∞ distance with $\mathbf{polylog}(n)$ samples?

3) Can we obtain time-efficient algorithms?

   (possible for commuting Hamiltonians in ℓ_∞ distance)

4) Practical implementations

References:

[AKL16] Itai Arad, Tomotaka Kuwahara, and Zeph Landau. Connecting global and local energy distributions in quantum spin models on a lattice. **arxiv: 1406.3898**

[BGP+ 20] Eyal Bairey, Chu Guo, Dario Poletti, Netanel H Lindner, and Itai Arad. Learning the dynamics of open quantum systems from their steady states. **arxiv:1907.11154**

[BKL+17] Fernando Brandão, et al. Quantum SDP solvers: Large speed-ups, optimality, and applications to quantum learning. ***arXiv:1710.02581***.

[Bre15] Guy Bresler. Efficiently learning Ising models on arbitrary graphs. **arxiv.org:1411.6156**

[EHF19] Tim J Evans, Robin Harper, and Steven T Flammia. Scalable bayesian Hamiltonian learning. **arXiv:1912.07636**

[Hastings07] Matthew B Hastings. Quantum belief propagation: An algorithm for thermal quantum systems. **arxiv:0706.4094**

[Jay57a] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957

[KB19] Kato, Kohtaro, and Fernando GSL Brandao. Quantum approximate Markov chains are thermal. **arxiv:1609.06636**

[KKB19] Tomotaka Kuwahara, Kohtaro Kato, and Fernando GSL Brandão. Clustering of conditional mutual information for quantum Gibbs states above a threshold temperature. **arXiv:1910.09425**

[KM17] Adam R. Klivans and Raghu Meka. Learning graphical models using multiplicative weights. **arxiv:1706.06274**

[LP08] Matthew S Leifer and David Poulin. Quantum graphical models and belief propagation. **arxiv: 0708.1337**

[Montnari15] Computational implications of reducing data to sufficient statistics **arXiv: 1409.3821**

[VMLC16] Marc Vuffray, Sidhant Misra, Andrey Lokhov, and Michael Chertkov. Interaction screening: Efficient and sample-optimal learning of Ising models. **arxiv:1605.07252**

[WPS+ 17] Jianwei Wang, et al. Experimental quantum hamiltonian learning. **arxiv:1703.05402**

[WGFC14] Nathan Wiebe, Christopher Granade, Christopher Ferrie, and David Cory. Quantum hamiltonian learning using imperfect quantum resources. **arxiv: 1311.5269**

Sample-efficient learning of quantum many-body systems

Mehdi Soleimanifar (MIT)

(arxiv: 2004.07266, FOCS'20)

Joint work with
Anurag Anshu (UC Berkeley)
Srinivasan Arunachalam (IBM)
Tomotaka Kuwahara (RIKEN)