Title: Sample-efficient learning of quantum many-body hamiltonians

Speakers: Anurag Anshu

Series: Perimeter Institute Quantum Discussions

Date: September 23, 2020 - 4:00 PM

URL: http://pirsa.org/20090017

Abstract: We study the problem of learning the Hamiltonian of a quantum many-body system given samples from its Gibbs (thermal) state. The classical analog of this problem, known as learning graphical models or Boltzmann machines, is a well-studied question in machine learning and statistics. In this work, we give the first sample-efficient algorithm for the quantum Hamiltonian learning problem. In particular, we prove that polynomially many samples in the number of particles (qudits) are necessary and sufficient for learning the parameters of a spatially local Hamiltonian in l_2-norm.

Our main contribution is in establishing the strong convexity of the log-partition function of quantum many-body systems, which along with the maximum entropy estimation yields our sample-efficient algorithm. Classically, the strong convexity for partition functions follows from the Markov property of Gibbs distributions. This is, however, known to be violated in its exact form in the quantum case. We introduce several new ideas to obtain an unconditional result that avoids relying on the Markov property of quantum systems, at the cost of a slightly weaker bound. In particular, we prove a lower bound on the variance of quasi-local operators with respect to the Gibbs state, which might be of independent interest.

Joint work with Srinivasan Arunachalam, Tomotaka Kuwahara, Mehdi Soleimanifar

Sample efficient learning of quantum many-body systems

Anurag Anshu, Srinivasan Arunachalam (IBM TJ Watson), Tomotaka Kuwahara (RIKEN AIP), Mehdi Soleimanifar (MIT)

Institute for Quantum Computing,
Department of Combinatorics and Optimization, UWaterloo,
Perimeter Institute for Theoretical Physics
arxiv 2004.07266, FOCS 2020

September 23, 2020

Upcoming section

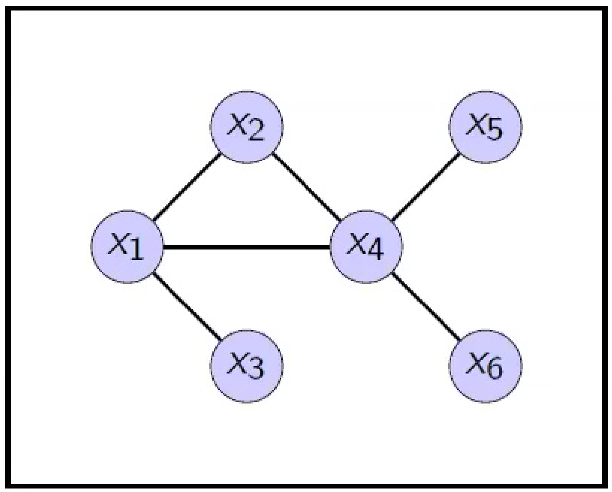
Learning graphical models

Learning quantum Hamiltonians

Techniques involved

Classical graphical models

- Also known as Markov random fields, Boltzmann Machines, etc.
- Fix a graph, which defines conditional independence.



Classical graphical models

- If the distribution is positive, then Hammersley-Clifford theorem ensures that the distribution is a *Gibbs state* given by

$$P_\beta(x_1, x_2, \dots x_6) = \frac{e^{-\beta \sum_{i\sim j} h_{i,j}(x_i, x_j)}}{Z_\beta}.$$

  Here Z_β is the partition function.
- The classical hamiltonian respects the graphical structure.
- A widely studied example is the Ising model, where the hamiltonian is

$$\sum_{i\sim j} \mu_{i,j} x_i x_j + \sum_i \theta_i x_i.$$

Importance of graphical models

- Bioinformatics (protein-protein interaction graph).

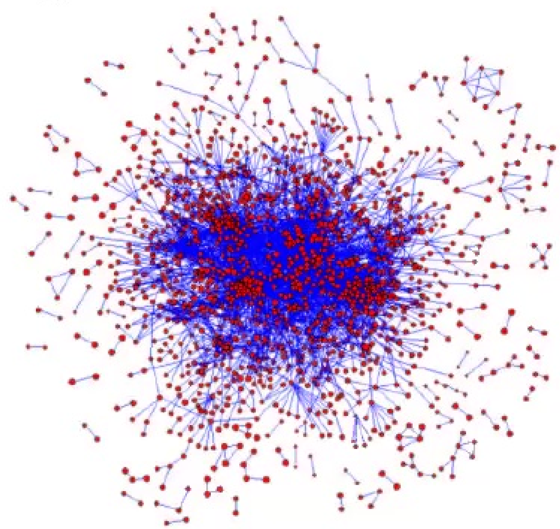


Image source: Knowledge discovery in Proteomics, Natasa Przulj (ed. Igor Jurisica, Dennis Wigle)

Importance of graphical models

- Statistics and machine learning (high dimensional statistics).
  - Can several parameters be learned with a few samples? The promise is of sparse graphs.

Importance of graphical models

- Statistics and machine learning (high dimensional statistics).
  - Can several parameters be learned with a few samples? The promise is of sparse graphs.
- Statistical physics (inverse Ising problem): infer the interaction based on observations (Chayes, Chayes, Lieb [Comm. Math. Phys. 1984]).

Inverse statistical problems: from the inverse Ising problem to data science

H. Chau Nguyen, Riccardo Zecchina & **Johannes Berg**

Pages 197-261 | Received 02 Aug 2016, Accepted 03 Jun 2017, Published online: 29 Jun 2017

Learning graphical models

- Imagine samples are being given from the distribution P_β, for the Ising model.
- We are promised that $-1 \leq \mu_{i,j}, \theta_i \leq 1$ (natural assumption in practise) and that the underlying graph has small degree (graph structure unknown).

Learning graphical models

- Imagine samples are being given from the distribution P_β, for the Ising model.
- We are promised that $-1 \leq \mu_{i,j}, \theta_i \leq 1$ (natural assumption in practise) and that the underlying graph has small degree (graph structure unknown).
- How many samples are needed to output real numbers $\{\mu'_{i,j}, \theta'_i\}$ such that:
  - ℓ_∞ **guarantee:** we have $|\mu'_{i,j} - \mu_{i,j}| \leq \varepsilon$ and $|\theta'_i - \theta_i| \leq \varepsilon$.
  - ℓ_2 **guarantee:** we have $\sqrt{\sum_{i\sim j} |\mu'_{i,j} - \mu_{i,j}|^2} \leq \varepsilon$ and $\sum_i \sqrt{|\theta'_i - \theta_i|^2} \leq \varepsilon$.

A series of classical results

- Several old works have tackled the learning problem in special settings.
  - Chow, Liu [IEEE IT 1963]; Ackley, Hinton, Sejnowski [Cognitive Science 1985]; Toshiyuki [PRE 1998], etc.
- Successful line of work for graphs with ℓ_∞ learning.
  - Bresler [STOC 2015]; Vuffray, Misra, Lokhov, Chertkov [NeurIPS, 2016]; Klivans, Meka [FOCS, 2017], etc.
  - Sample complexity is near optimal, roughly $e^{\mathcal{O}(\beta)} \log |G|$.
  - Time complexity is $\mathcal{O}\left(|G|^2\right)$.

The talk so far...

- Learning graphical models is an important task.
- Several classical algorithms achieve this for learning in ℓ_∞ norm. Sufficient to infer the structure of the graph.
- Perspective: number of unknown graphs of constant degree on $|G|$ vertices is $\sim e^{\Omega(|G| \log |G|)}$.
- Thus, a sample complexity of $e^{\mathcal{O}(\beta)} \log |G|$ is a massive reduction.

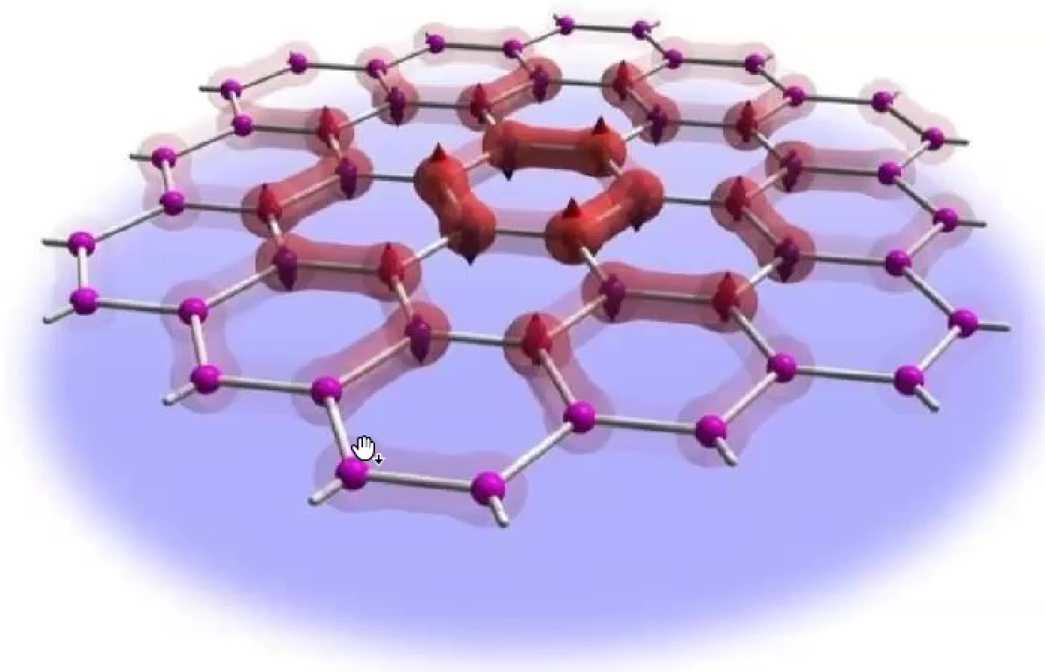
Upcoming section

Learning graphical models

Learning quantum Hamiltonians

Techniques involved

Quantum many-body systems



Quantum many-body systems

- A local hamiltonian describes physical interactions:

$$H(\mu) = \sum_{\ell=1}^{m} \mu_\ell E_\ell,$$

where $-1 \leq \mu_\ell \leq 1$, $\|E_\ell\|_\infty \leq 1$ (spins don't interact too strongly).

- E_ℓ are local operators, they act on closeby spins. Example, pauli operators with small weight.
- Gibbs state is gives by

$$\rho_\beta(\mu) = \frac{e^{-\beta H(\mu)}}{Z_\beta(\mu)},$$

where $Z_\beta(\mu) = \mathrm{Tr}[e^{-\beta H(\mu)}]$ is the quantum partition function.

Motivation to learn a quantum Hamiltonian

- Several recent learning results in varying domains.
- Sample complexity of supervised learning: Arunachalam, de Wolf [CCC, 2017].
- Shadow tomography: Aaronson [STOC 2018]; Huang, Kueng, Preskill [Nat. Phys. 2020].
- Quantum Boltzmann machines: Amin, Andriyash, Rolfe, Kulchytskyy, Melko [Phys Rev X, 2018].

Motivation to learn a quantum Hamiltonian

- The 'inverse problem': finding the complex interactions in real-world materials.
  - The lattice structure can play a crucial role in the physics of the system.
  - Type of the interaction is important for classifying the phase of matter.
  - Experimental push: Zhao et. al. [Science, 2020], Sibille et. al. [Nat. Phys., 2020], etc.
- Gibbs sampling is taking a central stage in quantum algorithms and learning (quantum SDP solver of Brandao, Svore [FOCS 2017] & subsequent works; quantum simulated annealing Montanaro [Proc. Roy. Soc. 2015]; Harrow, Wei [SODA, 2020]; so on).
  - Verification of the Gibbs sampler is an inevitable problem.

Prior works

- Set up a linear system of equations, which can be inverted to obtain the parameters μ.
  - Bairey, Arad, Lindner [PRL, 2019]; Evans, Harper, Flammia [2019]; Qi, Ranard [Quantum, 2019].
  - Applies to Gibbs states or ground states or high energy eigenstates.
  - But rigorous guarantee on invertibility of the linear system not present.

Our results

- We consider a hamiltonian $H(\mu)$ on a geometrically local graph in finite dimensional space. Recall that $H(\mu) = \sum_{\ell=1}^{m} \mu_\ell E_\ell$ and E_ℓ is local.

Theorem
We give an algorithm that fails with tiny probability δ. Whenever it succeeds, it outputs a μ' that approximates μ in ℓ_2 norm ε. The number of samples needed is

$$\mathcal{O}\left(\frac{e^{\beta c}}{\beta^{c'}\varepsilon^2} \cdot m^3 \log \frac{m}{\delta}\right).$$

Our results

- We consider a hamiltonian $H(\mu)$ on a geometrically local graph in finite dimensional space. Recall that $H(\mu) = \sum_{\ell=1}^{m} \mu_\ell E_\ell$ and E_ℓ is local.

Theorem

We give an algorithm that fails with tiny probability δ. Whenever it succeeds, it outputs a μ' that approximates μ in ℓ_2 norm ε. The number of samples needed is

$$\mathcal{O}\left(\frac{e^{\beta^c}}{\beta^{c'}\varepsilon^2} \cdot m^3 \log \frac{m}{\delta}\right).$$

Theorem

In any algorithm with ℓ_2 guarantee ε, sample complexity must be at least $\frac{\sqrt{m-\delta}}{\beta\varepsilon}$. We provide a classical example for this.

Our results

- Algorithm consists of quantum measurements on small number of qubits, followed by classical postprocessing.
- Time efficiency in some cases (discussed later).

Comparison with prior work

- Classical results:
  - are time efficient and sample near-optimal. We are time efficient conditionally and polynomially optimal in sample.
  - They recover the graph, as long as it is low degree. We need the graph to be geometrically local (such as a lattice or a combination of lattices).
  - They work in ℓ_∞ guarantee, as it is enough to determine the graph. ℓ_2 norm is costly for them: $\sqrt{m}$ samples needed even classically.

Comparison with prior work

- Quantum results.
  - Sample efficiency in other results is conditional: it depends on properties of the linear system. We are provably sample efficient.

The talk so far...

- Learning quantum many-body Hamiltonian is a natural generalization of learning graphical models.
- We give an algorithm that learns a lattice Hamiltonian under ℓ_2- guarantee. Sample complexity is $\approx m^3$ and one cannot do better than $\approx \sqrt{m}$.
- Time efficient in some regime. Earlier works were time efficient, but sample complexity not guaranteed.

Upcoming section

Learning graphical models

Learning quantum Hamiltonians

Techniques involved

Sufficient statistics

- The marginals of a Gibbs state uniquely determine the Hamiltonian.
- Let $H(\mu) = \sum_{\ell=1}^{m} \mu_\ell E_\ell$.
- Consider the expectation values $e_\ell = \mathrm{Tr}[E_\ell \rho_\beta(\mu)]$, for all ℓ.
- If there is some μ' with

$$\mathrm{Tr}[E_\ell \rho_\beta(\mu)] = \mathrm{Tr}[E_\ell \rho_\beta(\mu')], \quad \forall \ell,$$

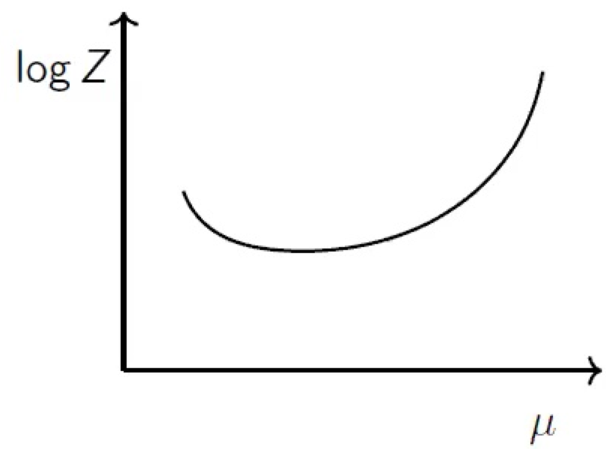
then $\mu = \mu'$.

Sufficient statistics

- The idea goes back to Jaynes [1957].
- Used extensively in the classical literature. See Kato, Brandao [Comm. Math. Phys, 2019] or Brandao et. al. [ICALP 2019] for applications to quantum Markov chains or quantum algorithms.
- Learning method: compute the expectation values e_ℓ exactly. Then obtain μ (see Kim, Swingle [PRL 2014]).

Quantum partition function

- Sufficient statistics can be visualized using the partition function $Z_\beta(\mu) = \mathrm{Tr}[e^{-\beta H(\mu)}]$.
- The log-partition function $\log Z_\beta(\mu)$ is a convex function of μ.

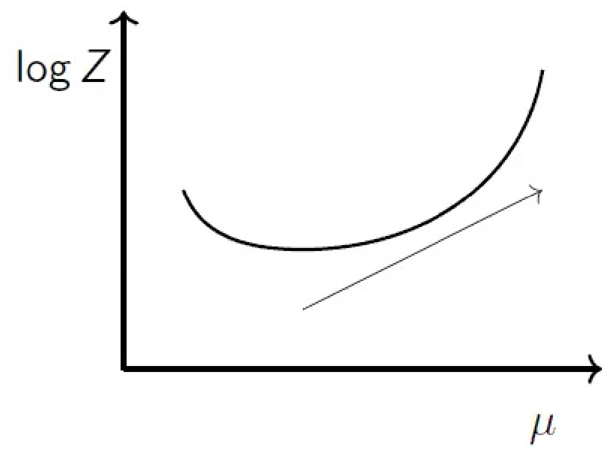


Quantum partition function

- It can be shown that

$$-\beta e_\ell = \frac{\partial \log Z_\beta(\mu)}{\partial \mu_\ell},$$

for all ℓ (recall statistical mechanics course... partition function contains all the information).

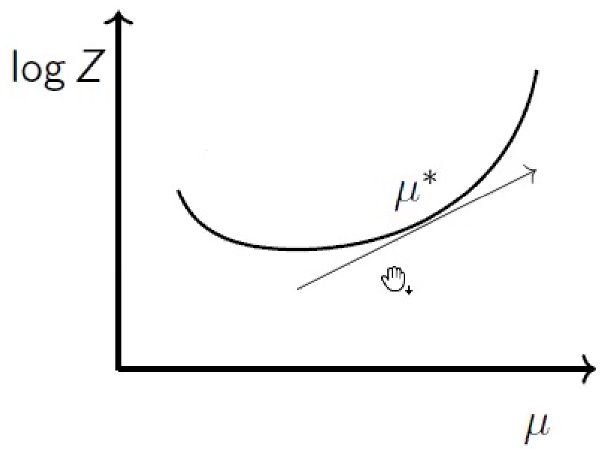


Quantum partition function

- It can be shows that

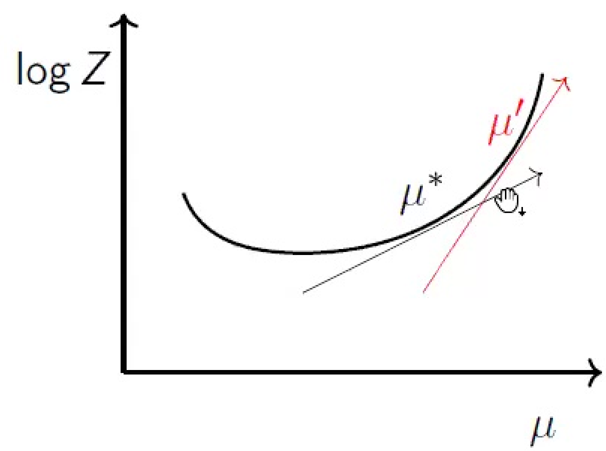
$$-\beta e_\ell = \frac{\partial \log Z_\beta(\mu)}{\partial \mu_\ell},$$

for all ℓ (recall statistical mechanics course... partition function contains all the information).



What happens under error?

- Suppose we estimate e_ℓ and obtain e'_ℓ.
- Then more "curved" the curve is, closer our estimate μ' is to μ^*.



Strong convexity of log-partition function

- We show that the log-partition function is strongly convex, which says its second derivate at every point, along every direction pair, is large.

Theorem (Main technical result)

It holds that

$$\nabla^2 \log Z_\beta \succeq \frac{e^{-\beta^c}\beta^{c'}}{m}\mathrm{I}.$$

Strong convexity of log-partition function

- We show that the log-partition function is strongly convex, which says its second derivate at every point, along every direction pair, is large.

Theorem (Main technical result)

It holds that

$$\nabla^2 \log Z_\beta \succeq \frac{e^{-\beta^c} \beta^{c'}}{m} \mathrm{I}.$$

Classically, a stronger result is true: $\nabla^2 \log Z_\beta \succeq e^{-\beta} \beta^{c'} \mathrm{I}$. This still leads to polynomial sample complexity, with $m^3 \to m$.

Interlude: strong convexity in classical world

- Essential tool in algorithms based on convex optimization.
- Many of the efficient classical algorithms need to establish strong convexity of some function.
- For graphical models, this reduces to proving that variance of local operators with respect to Gibbs state is large.

Interlude: variance of local operators

- Instructive example: consider the Ising model with $H = -\sum_{i=1}^{m-1} \sigma_i^z \sigma_{i+1}^z$.
- Magnetization is of interest: $M = \sum_{i=1}^{m} \sigma_i^z$.
- Computation reveals that the variance of the magnetization $\mathrm{Tr}[M^2 \rho_\beta] - \mathrm{Tr}[M\rho_\beta]^2$ is proportional to m.
- In other words, with high probability, the value of magnetization concentrates around $\mathrm{Tr}[M\rho_\beta] \pm \Omega(\sqrt{m})$.
- We do not expect the value to concentrate any better. For instance, $\mathrm{Tr}[M\rho_\beta] \pm 10.23$ does not happen.

What is known about the variance?

- Several works have shown (Araki [Comm. Math. Phys. 1967]; Frohlich, Ueltschi [Comm. Math. Phys. 2015]; Kliesch et. al. [Phys Rev X, 2014]) that for small β and $M = \sum_i G_i$, where G_i is local,

$$\mathrm{Tr}[M^2\rho_\beta] - \mathrm{Tr}[M\rho_\beta]^2 = \mathcal{O}(m).$$

Proving strong convexity: variance of quasi-local operators

- Classically, given a hamiltonian $H(\mu)$ and some other local operator $L(\nu) = \sum_{\ell=1}^{m} \nu_\ell E_\ell$, we can prove that

$$\mathrm{Tr}[L(\nu)^2 \rho_\beta(\mu)] - \mathrm{Tr}[L(\nu)\rho_\beta(\mu)]^2 \geq e^{-\beta}\beta^{c'} \cdot \sum_\ell \nu_\ell^2.$$

- This directly leads to strong convexity of log-partition function.
- This is much harder quantumly and we further need to show this for quasi-local operators $L(\nu)$, which have decaying locality.
- We prove that

$$\mathrm{Tr}[L(\nu)^2 \rho_\beta(\mu)] - \mathrm{Tr}[L(\nu)\rho_\beta(\mu)]^2 \geq \frac{e^{-\beta^c}\beta^{c'}}{m} \cdot \sum_\ell \nu_\ell^2.$$

Proving strong convexity: variance of quasi-local operators

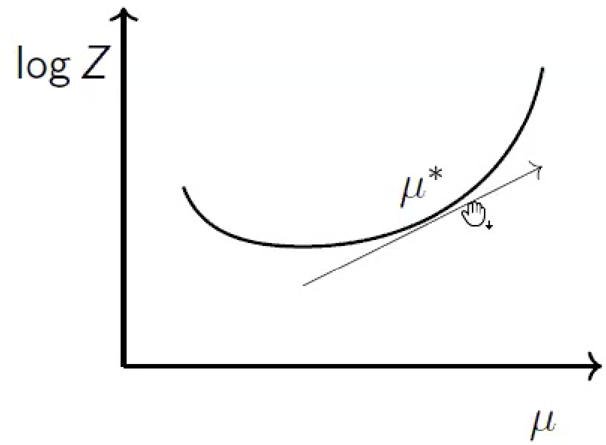
- Step 1: Use quantum belief propagation (Hastings [PRB 2007], Kim [2017], Kato and Brandao [Comm Math Phys, 2018]) to reduce strong convexity to proving variance lower bound of quasi-local operators.
- Step 2: Use local unitaries to make the problem local:

$$O \to O_i = O - \frac{1}{d}\mathrm{tr}_i[O]$$

(Fourier analysis, anyone?)

- Prove variance lower bound for this local operator O_i and use known results (Arad, Kuwahara, Landau [J. Stat. Mech., 2016]) to make statements about O.

Time complexity



- Finding μ^* is same as minimizing the function

$$\log Z_\beta(\mu) + \beta \sum_\ell \mu_\ell e_\ell.$$

Time complexity

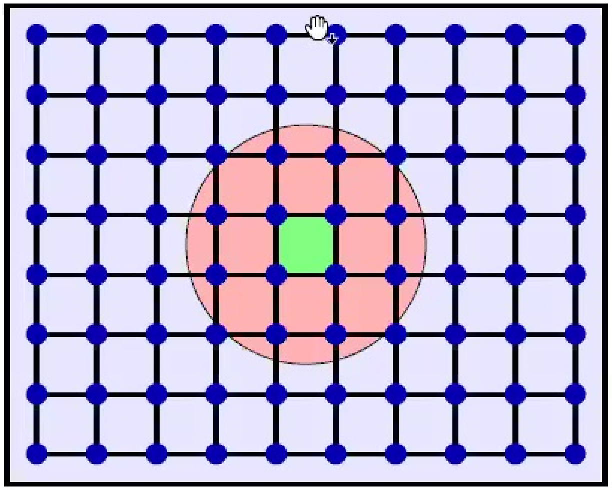
- Minimize $\log Z_\beta(\mu) + \beta \sum_\ell \mu_\ell e_\ell$, which has a unique minimum μ^*.
- One can use the gradient descent algorithm.
- Time efficient if the partition function $Z_\beta(\mu')$ at every μ' in the region can be computed efficiently.
- Thus, time efficient at high temperature (Harrow, Mehraban, Soleimanifar [STOC 2020]; Kuwahara, Kato, Brandao [PRL, 2020]) and in the stoquastic case (Bravyi, Terhal [SIAM J Comp, 2008]).

The talk so far...

- Sufficient statistics gives a natural approach to the learning problem.
- We prove a robust version of sufficient statistics approach.
- Reduces to proving variance lower bound on quantum many-body local operators.
  - Similar to classical case, we expect this to have further applications.

Future directions: sufficient local statistics

- Can we modify sufficient statistics such that μ_ℓ can be inferred by computing $e_{\ell'}$, where ℓ' is geometrically close to ℓ.
- Then estimation of μ_ℓ will be a local procedure.



Future directions: sufficient local statistics

- This may be useful in obtaining time efficient results.
  - Can be established in the commuting case.
- Time efficient learning algorithm with near optimal sample complexity in general case?

Future directions

- Improving the strong convexity result to be independent of m. A central problem in condensed matter physics, on variance of operators.
- Our procedure reduces the variance problem to a local region, which leads to a loss. Can this be avoided?

Thank you for your attention!