Learning the closest product state

John Bostanci

Joint work with Ainesh Bakshi, William Kretschmer, Ryan O’Donnell, Zeph Landau, Jerry Li, Allen Liu, and Ewin Tang

Motivation

Let’s imagine the following scenario:

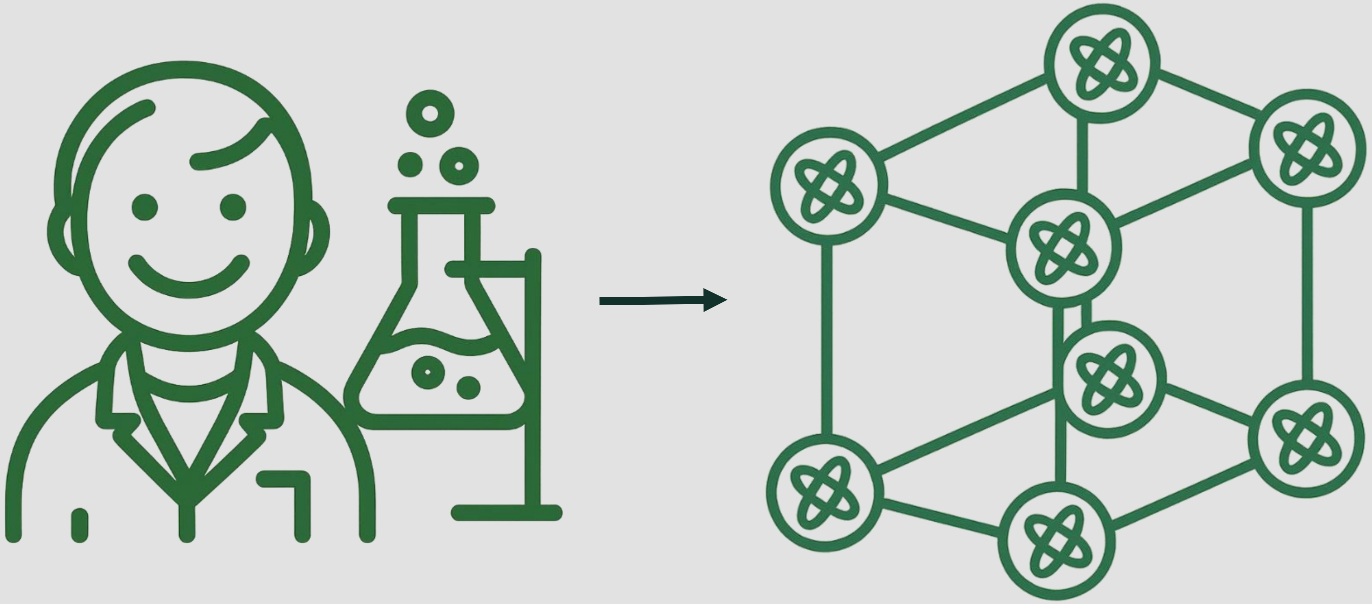
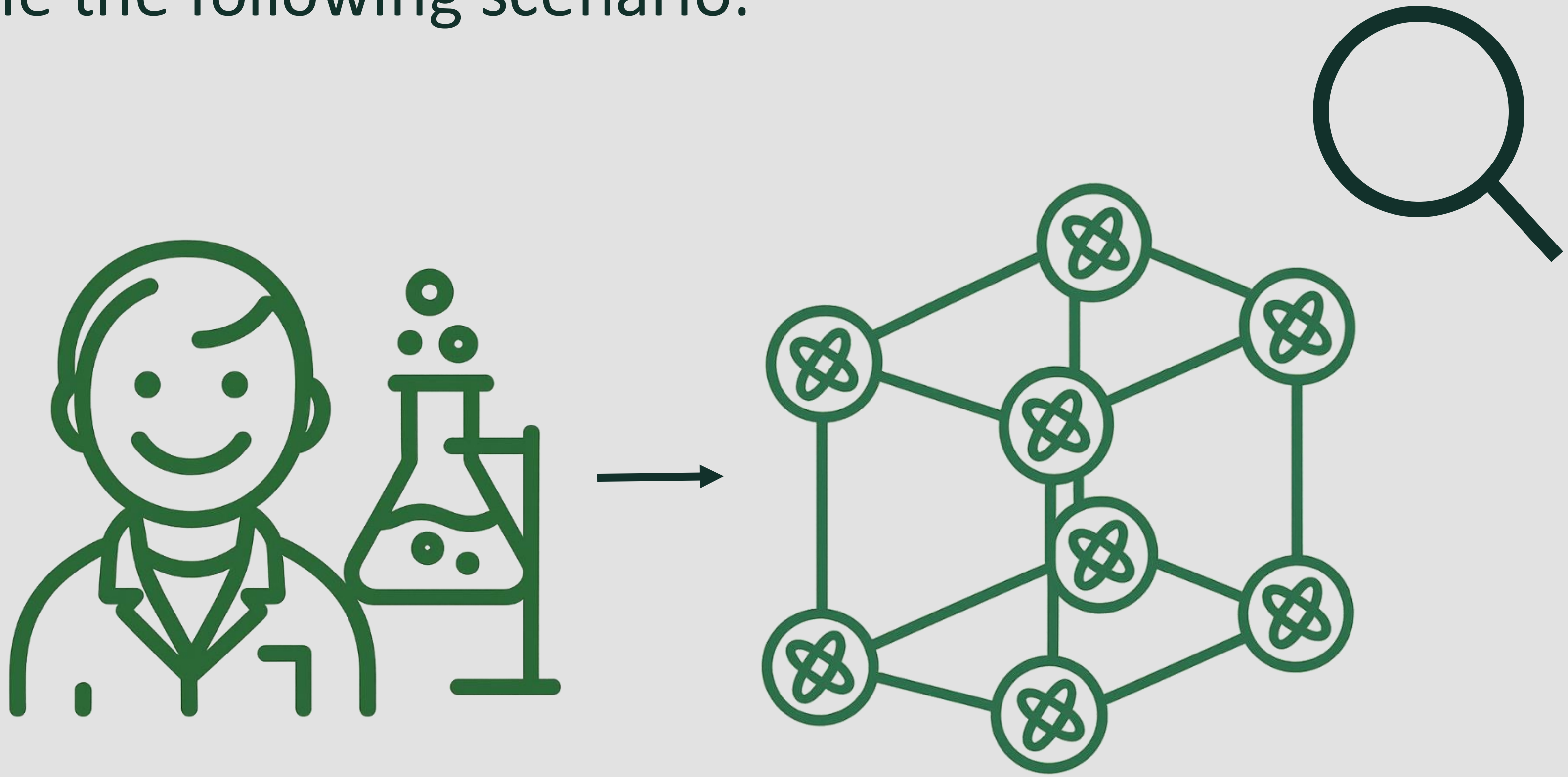


Image credits: OpenAI

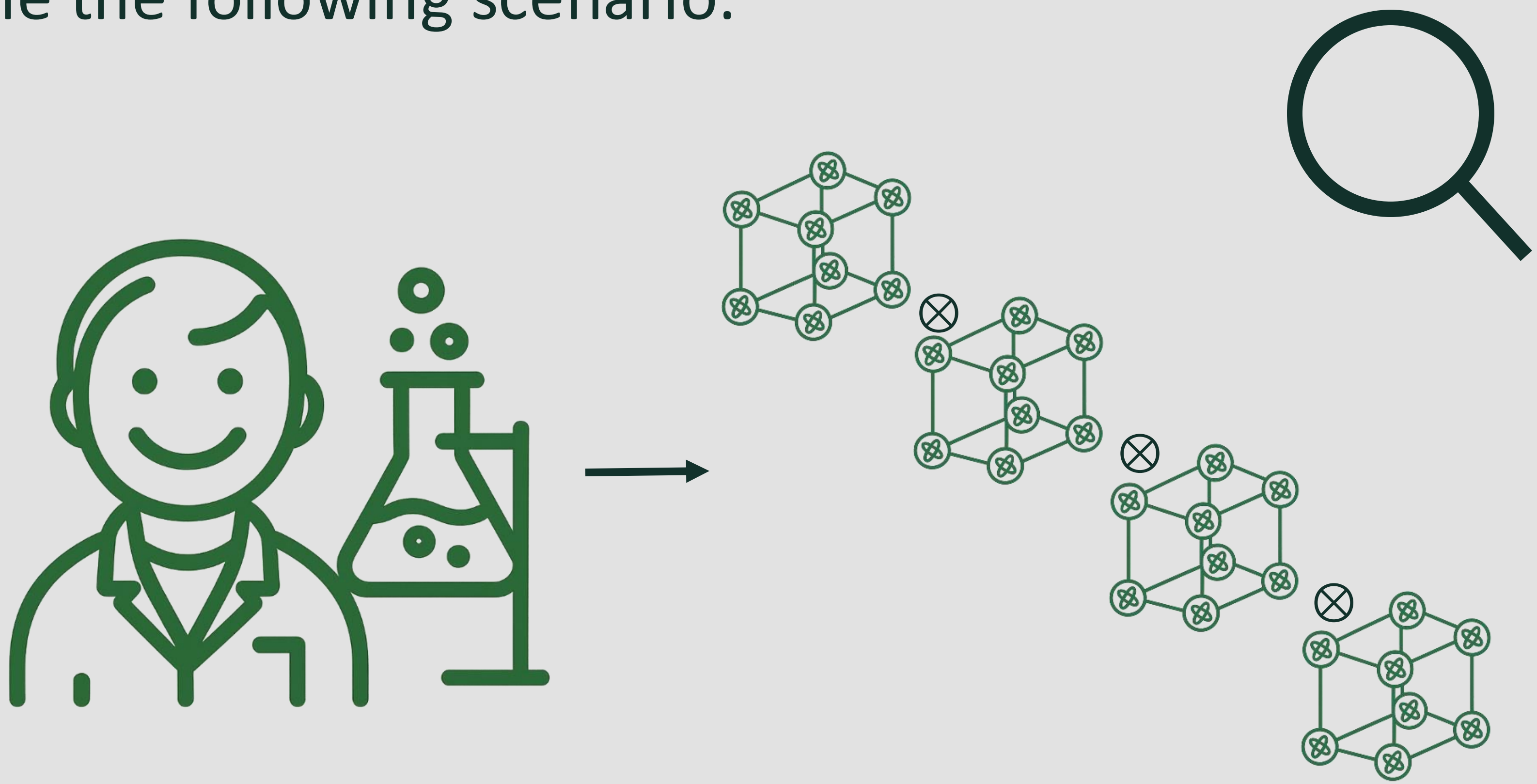
Motivation

Let’s imagine the following scenario:



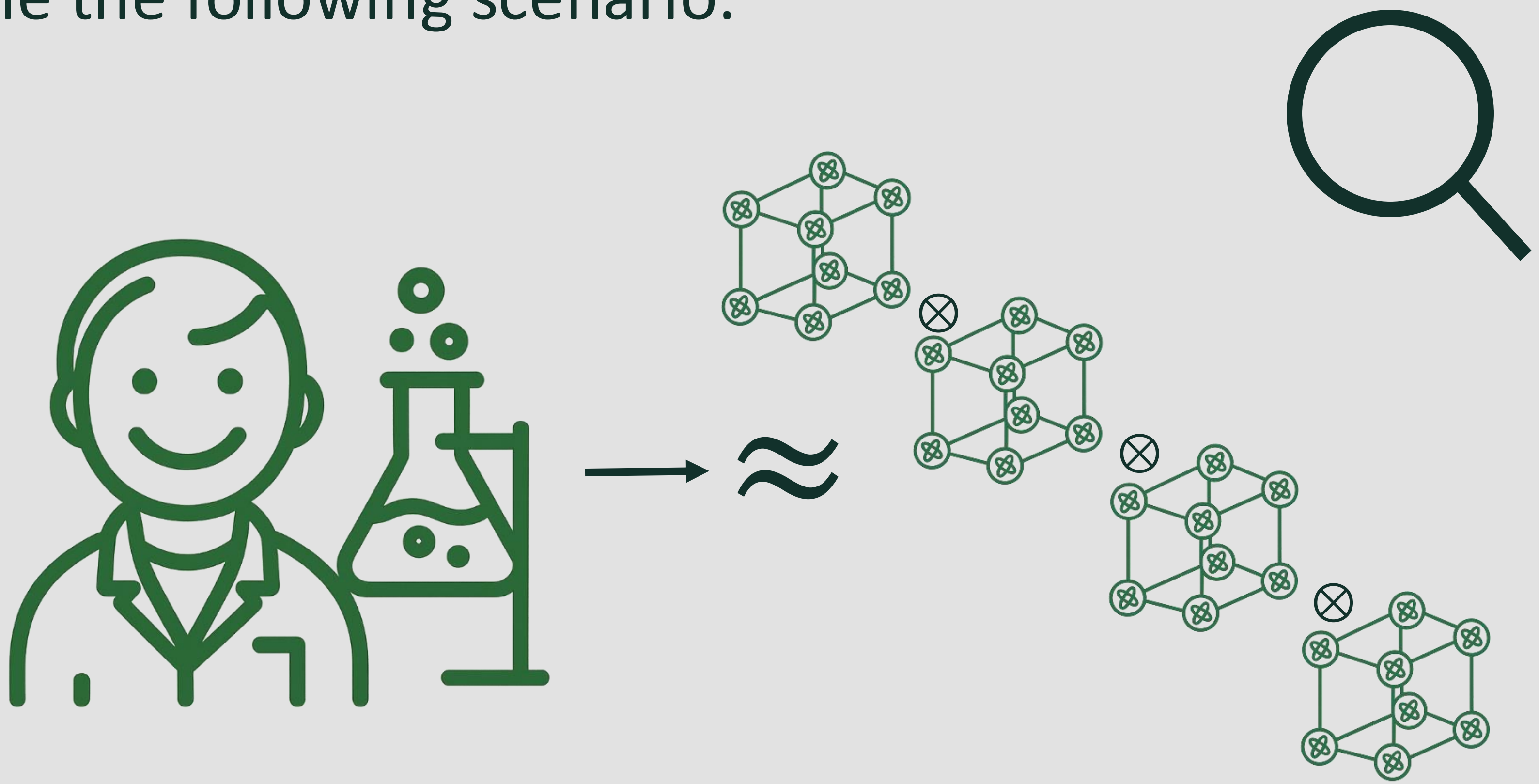
Motivation

Let's imagine the following scenario:



Motivation

Let’s imagine the following scenario:



Agnostic quantum tomography

Given a model class C and copies of an arbitrary quantum state ρ, output a description of the closest state in C to ρ.

$$(C, \rho^{\otimes n}) \rightarrow \underset{|\psi\rangle \in C}{\operatorname{argmax}} \langle\psi|\rho|\psi\rangle$$

Agnostic quantum tomography

Given a model class C and copies of an arbitrary quantum state ρ, output a description of the closest state in C to ρ.

$$(C, \rho^{\otimes n}) \to \underset{|\psi\rangle \in C}{\operatorname{argmax}} \langle\psi|\rho|\psi\rangle$$

OPT

Agnostic quantum tomography

Given a model class C and copies of an arbitrary quantum state ρ, output a description of the closest state in C to ρ.

$$\left(C, \rho^{\otimes n}\right) \to |\phi\rangle \in C \colon \mathrm{OPT}\; - \langle \phi | \rho | \phi \rangle \leq \epsilon$$

Agnostic quantum tomography

Given a model class C and copies of an arbitrary quantum state ρ, output a description of the closest state in C to ρ.

$$\left(C, \rho^{\otimes n}\right) \rightarrow |\phi\rangle \in C: \mathrm{OPT} - \langle\phi|\rho|\phi\rangle \leq \epsilon$$

Shadow tomography provides a sample efficient, but not computationally efficient, solution.

Agnostic tomography is hard

Even for “simple” model classes like product states, agnostic tomography isn’t immediate.

Agnostic tomography is hard

Even for "simple" model classes like product states, agnostic tomography isn't immediate. Consider the following state:

$$\sqrt{1-\epsilon}|0^n\rangle + \sqrt{\epsilon}|+^n\rangle$$

Agnostic tomography is hard

Even for “simple” model classes like product states, agnostic tomography isn’t immediate. Consider the following state:

$$\sqrt{1-\epsilon}|0^n\rangle + \sqrt{\epsilon}|+^n\rangle$$

The closest product state is $|0^n\rangle$

Agnostic tomography is hard

Even for “simple” model classes like product states, agnostic tomography isn’t immediate. Consider the following state:

$$\sqrt{1-\epsilon}|0^n\rangle + \sqrt{\epsilon}|+^n\rangle$$

The closest product state is $|0^n\rangle$

Every marginal is $\approx \sqrt{1-\epsilon}|0\rangle + \sqrt{\epsilon}|1\rangle$.
Taking the tensor product is bad!

Our results (1)

Main result: there is an algorithm for agnostic tomography of product states that has sample and time complexity

$$\mathrm{poly}(n^{poly\left(\frac{1}{\epsilon}\right)})$$

Our results (2)

Suppose there was an polynomial-time algorithm for agnostic tomography of product states when ϵ is inverse polynomial in n, i.e. outputting a state satisfying

$$\langle\phi|\rho|\phi\rangle \geq \mathrm{OPT} - \frac{1}{\mathrm{poly}(n)}$$

Then BQP contains NP.

Our results (3)

In some settings, we can get fully polynomial run-times!

- When OPT $\geq \frac{5}{6}$, we give an algorithm that uses $O\left(\frac{n}{\epsilon}\right)$ copies and runs in time $O(n^2 \log n)$.

Our results (3)

In some settings, we can get fully polynomial run-times!

- When OPT $\geq \frac{5}{6}$, we give an algorithm that uses $O\left(\frac{n}{\epsilon}\right)$ copies and runs in time $O(n^2 \log n)$.
- When we have the promise that each qudit can only be one of s states, and all of the states have fidelity at most 1 - δ, we give an algorithm that runs in time $\mathrm{poly}(ns)^{\log(1/\epsilon)/\delta}$.

Our results (3)

In some settings, we can get fully polynomial run-times!

- When OPT $\geq \frac{5}{6}$, we give an algorithm that uses $O\left(\frac{n}{\epsilon}\right)$ copies and runs in time $O(n^2 \log n)$.
- When we have the promise that each qudit can only be one of s states, and all of the states have fidelity at most 1 - δ, we give an algorithm that runs in time $\mathrm{poly}(ns)^{\log(1/\epsilon)/\delta}$.
- We give an improper tomography algorithm for matrix product states that runs in time $\mathrm{poly}(n, 1/\epsilon)$.

Rest of the talk

The goal of the rest of the talk is to give an overview and intuition for the main algorithm.

- Going from local to global

Rest of the talk

The goal of the rest of the talk is to give an overview and intuition for the main algorithm.

- Going from local to global
- Reduction to polynomial optimization

Rest of the talk

The goal of the rest of the talk is to give an overview and intuition for the main algorithm.

- Going from local to global
- Reduction to polynomial optimization
- Solving the polynomial optimization

Rest of the talk

The goal of the rest of the talk is to give an overview and intuition for the main algorithm.

- Going from local to global
- Reduction to polynomial optimization
- Solving the polynomial optimization
- Putting everything together

Going from local to global

We saw before that simply taking the single qubit marginals of the state and taking their tensor product is a bad idea.

Going from local to global

We saw before that simply taking the single qubit marginals of the state and taking their tensor product is a bad idea.

Lets try to improve our approach a little bit!

Naïve strategy: Let $|\pi\rangle$ have $\mathrm{OPT} - \epsilon$ fidelity with the first k-1 qubits. Take an ϵ-net over states $\{|\pi_k\rangle\}$, and test if any of the following states have high fidelity ρ:

$$|\pi\rangle \otimes |\pi_k\rangle$$

Naïve strategy: Let $|\pi\rangle$ have $\mathrm{OPT} - \epsilon$ fidelity with the first k-1 qubits. Take an ϵ-net over states $\{|\pi_k\rangle\}$, and test if any of the following states have high fidelity ρ:

$$|\pi\rangle \otimes |\pi_k\rangle$$

However: If $|\pi\rangle$ has fidelity $\mathrm{OPT} - \epsilon$, and we're taking an ϵ-net, every state of this form could have fidelity $\mathrm{OPT} - 2\epsilon$.

Naïve strategy: Let $|\pi\rangle$ have $\mathrm{OPT} - \epsilon$ fidelity with the first k-1 qubits. Take an ϵ-net over states $\{|\pi_k\rangle\}$, and test if any of the following states have high fidelity ρ:

$$|\pi\rangle \otimes |\pi_k\rangle$$

However: If $|\pi\rangle$ has fidelity $\mathrm{OPT} - \epsilon$, and we're taking an ϵ-net, every state of this form could have fidelity $\mathrm{OPT} - 2\epsilon$.

We lose a little on every step!
Can we fix this loss?

Main problem: Let $|\pi\rangle$ be a product state such that $\langle\pi|\rho|\pi\rangle \geq \mathrm{OPT} - 10\,\epsilon$, can we find a nearby $|\pi'\rangle$ such that $\langle\pi'|\rho|\pi'\rangle \geq \mathrm{OPT} - \epsilon$?

We found $|\pi\rangle$ via a **local** update (appending $|\pi_k\rangle$)

Main problem: Let $|\pi\rangle$ be a product state such that $\langle\pi|\rho|\pi\rangle \geq \mathrm{OPT} - 10\,\epsilon$, can we find a nearby $|\pi'\rangle$ such that $\langle\pi'|\rho|\pi'\rangle \geq \mathrm{OPT} - \epsilon$?

We found $|\pi\rangle$ via a **local** update (appending $|\pi_k\rangle$)

Main problem: Let $|\pi\rangle$ be a product state such that $\langle\pi|\rho|\pi\rangle \geq \mathrm{OPT} - 10\,\epsilon$, can we find a nearby $|\pi'\rangle$ such that $\langle\pi'|\rho|\pi'\rangle \geq \mathrm{OPT} - \epsilon$?

$|\pi'\rangle$ should incorporate something **global** around ρ

We found $|\pi\rangle$ via a **local** update (appending $|\pi_k\rangle$)

Main problem: Let $|\pi\rangle$ be a product state such that $\langle\pi|\rho|\pi\rangle \geq \mathrm{OPT} - 10\,\epsilon$, can we find a nearby $|\pi'\rangle$ such that $\langle\pi'|\rho|\pi'\rangle \geq \mathrm{OPT} - \epsilon$?

$|\pi'\rangle$ should incorporate something **global** around ρ

Searching only near $|\pi\rangle$ helps with **efficiency**

Main problem: Let $|\pi\rangle$ be a product state such that $\langle\pi|\rho|\pi\rangle \geq \mathrm{OPT}\ - 10\ \epsilon$, can we find a nearby $|\pi'\rangle$ such that $\langle\pi'|\rho|\pi'\rangle \geq \mathrm{OPT}\ - \epsilon$?

We can naturally parametrize a product state using a vector $\vec{z} \in \mathbb{C}^n$:

$$|\pi_{\vec{z}}\rangle = \bigotimes \frac{|0\rangle + z_i|1\rangle}{\sqrt{1 + |z_i|^2}}$$

We can naturally parametrize a product state using a vector $\vec{z} \in \mathbb{C}^n$:

$$|\pi_{\vec{z}}\rangle = \bigotimes \frac{|0\rangle + z_i|1\rangle}{\sqrt{1 + |z_i|^2}}$$

If we assume that our starting product state $|\pi\rangle = |0^n\rangle$, then when we say close to $|\pi\rangle$, we mean that $||\vec{z}||_2$ is small.

Observation: When $\|\vec{z}\|_2$ is small, $|\pi_{\vec{z}}\rangle$ is mostly supported on low Hamming weight strings.

Observation: When $\|\vec{z}\|_2$ is small, $|\pi_{\vec{z}}\rangle$ is mostly supported on low Hamming weight strings.

Lemma: Assume that $\|\vec{z}\|_2 \leq C$, and let $d = O(C + \log(1/\epsilon))$. Let Π_d be the projection onto all strings $|\mathrm{x}\rangle$ with Hamming weight at most d, then

$$\|\Pi_d \, |\pi_{\vec{z}}\rangle\|_2 \geq 1 - \epsilon$$

Observation: When $\|\vec{z}\|_2$ is small, $|\pi_{\vec{z}}\rangle$ is mostly supported on low Hamming weight strings.

Lemma: Assume that $\|\vec{z}\|_2 \leq C$, and let $d = O(C + \log(1/\epsilon))$. Let Π_d be the projection onto all strings $|\mathrm{x}\rangle$ with Hamming weight at most d, then

$$\|\Pi_d \, |\pi_{\vec{z}}\rangle\|_2 \geq 1 - \epsilon$$

In other words, if we want to (approximately) optimize $\vec{z}$, we just have to learn ρ on the small subspace Π_d.

Since $\dim(\Pi_d) = \binom{n}{d} = O(n^{C+\log(1/\epsilon)})$, we can do subspace tomography with copy and time complexity $\mathrm{poly}(n^{C+\log(1/\epsilon)})$.

Since $\dim(\Pi_d) = \binom{n}{d} = O(n^{C+\log(1/\epsilon)})$, we can do subspace tomography with copy and time complexity $\mathrm{poly}(n^{C+\log(1/\epsilon)})$.

The rest of our algorithm will be completely classical!

Recall: Let $|\pi\rangle$ be a product state such that $\langle\pi|\rho|\pi\rangle \geq \mathrm{OPT}\ - 10\,\epsilon$, can we find a nearby $|\pi'\rangle$ such that $\langle\pi'|\rho|\pi'\rangle \geq \mathrm{OPT}\ - \epsilon$?

Recall: Let $|\pi\rangle$ be a product state such that $\langle\pi|\rho|\pi\rangle \geq \mathrm{OPT} - 10\,\epsilon$, can we find a nearby $|\pi'\rangle$ such that $\langle\pi'|\rho_d|\pi'\rangle \geq \mathrm{OPT} - \epsilon$?

Restrict ρ to a low Hamming weight subspace

Recall: Let $|\pi\rangle$ be a product state such that $\langle\pi|\rho|\pi\rangle \geq \mathrm{OPT} - 10\,\epsilon$, can we find a nearby $|\pi'\rangle$ such that $\langle\pi'|\rho_d|\pi'\rangle \geq \mathrm{OPT} - \epsilon$?

We want to optimize over $||\vec{z}||_2 \leq C$:

$$\langle\pi_{\vec{z}}|\rho_d|\pi_{\vec{z}}\rangle = \frac{1}{\prod(1+|z_i|^2)} \sum_{|x|,|x'|\leq d} \langle x|\rho_d|x'\rangle \left(\prod_{x_i=1} \bar{z_i} \prod_{x_i'=1} z_i\right)$$

Recall: Let $|\pi\rangle$ be a product state such that $\langle\pi|\rho|\pi\rangle \geq \mathrm{OPT} - 10\,\epsilon$, can we find a nearby $|\pi'\rangle$ such that $\langle\pi'|\rho_d|\pi'\rangle \geq \mathrm{OPT} - \epsilon$?

We want to optimize over $||\vec{z}||_2 \leq C$:

$$\langle\pi_{\vec{z}}|\rho_d|\pi_{\vec{z}}\rangle = \frac{1}{\prod(1+|z_i|^2)} \underbrace{\sum_{|x|,|x'|\leq d} \langle x|\rho_d|x'\rangle \left(\prod_{x_i=1} \bar{z_i} \prod_{x_i'=1} z_i\right)}_{\text{A polynomial of degree 2d in } \vec{z} \text{ ☺}}$$

Recall: Let $|\pi\rangle$ be a product state such that $\langle\pi|\rho|\pi\rangle \geq \mathrm{OPT} - 10\,\epsilon$, can we find a nearby $|\pi'\rangle$ such that $\langle\pi'|\rho_d|\pi'\rangle \geq \mathrm{OPT} - \epsilon$?

We want to optimize over $||\vec{z}||_2 \leq C$:

$$\langle\pi_{\vec{z}}|\rho_d|\pi_{\vec{z}}\rangle = \underbrace{\frac{1}{\prod(1+|z_i|^2)}}_{\text{This term is annoying ☹}} \sum_{|x|,|x'|\leq d} \langle x|\rho_d|x'\rangle \left(\prod_{x_i=1} \bar{z_i} \prod_{x'_i=1} z_i\right)$$

How do we deal with this annoying term in front?

$$\frac{1}{\prod(1 + |z_i|^2)}$$

Idea 1: When all z_i are small, $\frac{1}{(1+|z_i|^2)} \approx e^{-|z_i|^2}$, so if we knew they were all small, we would have:

$$\frac{1}{\prod(1 + |z_i|^2)} \approx e^{-\|z_i\|_2^2}$$

How do we deal with this annoying term in front?

$$\frac{1}{\prod(1+|z_i|^2)}$$

Idea 1: When all z_i are small, $\frac{1}{(1+|z_i|^2)} \approx e^{-|z_i|^2}$, so if we knew they were all small, we would have:

$$\frac{1}{\prod(1+|z_i|^2)} \approx e^{-\|z_i\|_2^2}$$

How do we deal with this annoying term in front?

Idea 2: Not too many z_i can be large if $||z_i||_2 \leq C$, just guess which z_i are large and guess their values (i.e. just brute force over them), and optimize over the remaining terms (with smaller norm).

How do we deal with this annoying term in front?

After guessing, up to re-scaling, we can say that maximizing over close product states is equivalent to finding:

$$\max_{\substack{\|\vec{z}\|_2 \leq 1 \\ \|\vec{z}\|_\infty \leq \mu}} p(\vec{z})$$

Where p is a degree 2d polynomial.

But… solving a polynomial optimization is not easy in general.

Luckily, our polynomial only takes on small values (since it corresponds to a fidelity between two quantum states). So we can prove the following:

Lemma (informal): There exists a subspace V of dimension $O(d^2/\epsilon)$ such that

$$|p(\vec{z}) - p(V\,\vec{z})| \leq \epsilon.$$

Luckily, our polynomial only takes on small values (since it corresponds to a fidelity between two quantum states). So we can prove the following:

Lemma (informal): There exists a subspace V of dimension $O(d^2/\epsilon)$ such that

$$|p(\vec{z}) - p(V\,\vec{z})| \leq \epsilon.$$

What does it mean: To optimize the polynomial *without ℓ_∞ constaints*, we just have to brute force over V, taking time $O((1/\epsilon)^{\mathrm{poly}(d,\epsilon)})$.

Luckily, our polynomial only takes on small values (since it corresponds to a fidelity between two quantum states). So we can prove the following:

Lemma (informal): There exists a subspace V of dimension $O(d^2/\epsilon)$ such that

$$|p(\vec{z}) - p(V\ \vec{z})| \leq \epsilon.$$

What does it mean: To optimize the polynomial *without ℓ_∞ constaints*, we just have to brute force over V, taking time $O((1/\epsilon)^{\text{poly}(d,\epsilon)})$.

Again, we can just guess the coodinates that saturate the constraints and remove them!

Putting everything together

Full algorithm (for growing a single candidate):

1. Use subspace tomography to get a description of ρ_d.
2. Guess the coordinates of z_i that are large and their values.
3. Guess which coordinates saturate ℓ_∞ constraints exactly.
4. For the remaining coordinates, solve the ℓ_2 constrained polynomial optimization problem.

Putting everything together

Full algorithm (for growing a single candidate):

1. Use subspace tomography to get a description of ρ_d.
2. Guess the coordinates of z_i that are large and their values.
3. Guess which coordinates saturate ℓ_∞ constraints exactly.
4. For the remaining coordinates, solve the ℓ_2 constrained polynomial optimization problem.

In reality, we try to maintain a set of “all” good candidate states, and grow them all!

Lower bound

Lower bound

The lower bound comes from basically inverting the construction for the $d = 2$ subspace. Given a 4-tensor (think: degree 4 polynomial),

Lower bound

The lower bound comes from basically inverting the construction for the d = 2 subspace. Given a 4-tensor (think: degree 4 polynomial),

$$T \mapsto \sum_{i,j,k,\ell} T_{i,j,k,\ell} |e_i, e_j, e_k, e_\ell\rangle$$

Lower bound

The lower bound comes from basically inverting the construction for the d = 2 subspace. Given a 4-tensor (think: degree 4 polynomial),

$$T \mapsto \sum_{i,j,k,\ell} T_{i,j,k,\ell} |e_i, e_j, e_k, e_\ell\rangle$$

The best product state approximation, $\vec{z}$, to this* yields a solution to the tensor optimization $T(\vec{z}, \vec{z}, \vec{z}, \vec{z})$, which is NP-hard.

Open questions

- Agnostic tomography for other states?
  - Free-Fermionic states?
  - Low-degree circuits?
  - QAC_0?
- Tolerant testing for these models?
- Agnostic learning for other models?
  - Unitaries/Channels?
  - Hamiltonians evolution?

Thanks for listening!

Extra slides

Lemma: Assume that $\|\vec{z}\|_2 \leq C$, and let $d = O(C + \log(1/\epsilon))$. Let Π_d be the projection onto all strings $|\mathrm{x}\rangle$ with Hamming weight at most d, then

$$\|\Pi_d \, |\pi_{\vec{z}}\rangle\|_2 \geq 1 - \epsilon$$

Proof: The mass on any string x is given by

$$|c_x|^2 = \frac{1}{\prod(1 + |z_i|^2)} \prod_{x=1} |z_i|^2$$

Lemma: Assume that $||\vec{z}||_2 \leq C$, and let $d = O(C + \log(1/\epsilon))$. Let Π_d be the projection onto all strings $|\mathrm{x}\rangle$ with Hamming weight at most d, then

$$||\Pi_d \; |\pi_{\vec{z}}\rangle||_2 \geq 1 - \epsilon$$

Proof: The mass on any string x is given by

$$|c_x|^2 = \frac{1}{\prod(1 + |z_i|^2)} \prod_{x=1} |z_i|^2$$

As if the i'th coin has probability $\frac{|z_i|^2}{1+|z_i|^2}$

Lemma: Assume that $\|\vec{z}\|_2 \leq C$, and let $d = O(C + \log(1/\epsilon))$. Let Π_d be the projection onto all strings $|\mathrm{x}\rangle$ with Hamming weight at most d, then

$$\|\Pi_d \,|\pi_{\vec{z}}\rangle\|_2 \geq 1 - \epsilon$$

Proof: The mass on any string x is given by

$$|c_x|^2 = \frac{1}{\prod(1 + |z_i|^2)} \prod_{x=1} |z_i|^2$$

As if the i'th coin has probability $\frac{|z_i|^2}{1+|z_i|^2}$

Chernoff bound says this is small when |x| is too large!