Complexity of Reconstructing Quantum States and Green's Functions

PhD Thesis Defense

Alexander Meiburg

Department of Physics, UC Santa Barbara

Jun 5 2023

Part I

Quantum State Tomography

You have a procedure for preparing a quantum state, and you want to know what that state is.

Quantum State Tomography

You have a procedure for preparing a quantum state, and you want to know what that state is.

- Identifying the output of a quantum circuit
- Characterizing the result of some experiment
- Calibrating a quantum device (photonics, superconducting qubits, etc.)

Tomography of the quantum state of photons entangled in high dimensions

Megan Agnew, Jonathan Leach, Melanie McLaren, F. Stef Roux, and Robert W. Boyd
Phys. Rev. A **84**, 062101 – Published 2 December 2011

Scalable on-chip quantum state tomography

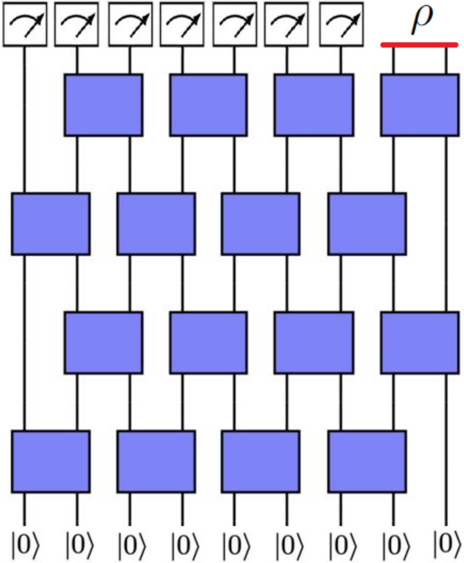
James G. Titchener, Markus Gräfe, René Heilmann, Alexander S. Solntsev, Alexander Szameit & Andrey A. Sukhorukov

npj Quantum Information **4**, Article number: 19 (2018) | Cite this article

5373 Accesses | **43** Citations | **3** Altmetric | Metrics

Experimental Single-Setting Quantum State Tomography

Roman Stricker, Michael Meth, Lukas Postler, Claire Edmunds, Chris Ferrie, Rainer Blatt, Philipp Schindler, Thomas Monz, Richard Kueng, and Martin Ringbauer
PRX Quantum **3**, 040310 – Published 21 October 2022


ρ
$|0\rangle$ $|0\rangle$ $|0\rangle$ $|0\rangle$ $|0\rangle$ $|0\rangle$ $|0\rangle$ $|0\rangle$ $|0\rangle$

We have some unknown quantum state $|\psi\rangle$ (or a mixed state ρ).
We have a method to repeatedly prepare $|\psi\rangle$:

We have some unknown quantum state $|\psi\rangle$ (or a mixed state ρ).
We have a method to repeatedly prepare $|\psi\rangle$:

- Running the quantum circuit
- Doing our experiment to prepare the state
- Run our device to calibrate from a known state

Quantum State Tomography

We have some unknown quantum state $|\psi\rangle$ (or a mixed state ρ). We have a method to repeatedly prepare $|\psi\rangle$:

- Running the quantum circuit
- Doing our experiment to prepare the state
- Run our device to calibrate from a known state

Repeatedly prepare $|\psi\rangle$ and measure it (in *some* basis), take these measurements to estimate $|\psi\rangle$.

Hilbert space dimension small d, $|\psi\rangle \in \mathbb{C}^d$. Demand a full picture of ψ

Quantum State Tomography: Usually Works?

Plenty of methods known to almost surely *converge* to correct answer:

Plenty of methods known to almost surely *converge* to correct answer:

- Pick an informationally complete basis and measure lots of times

Quantum State Tomography: Usually Works?

Plenty of methods known to almost surely *converge* to correct answer:

- Pick an informationally complete basis and measure lots of times
- Converges with $O(d^3/\epsilon^2)$ many samples (central limit theorem)
  - Can do $O(d^2/\epsilon^2)$ with a quantum computer
  - $O(d/\epsilon^2)$ if pure

Quantum State Tomography: Usually Works?

Plenty of methods known to almost surely *converge* to correct answer:

- Pick an informationally complete basis and measure lots of times
- Converges with $O(d^3/\epsilon^2)$ many samples (central limit theorem)
  - Can do $O(d^2/\epsilon^2)$ with a quantum computer
  - $O(d/\epsilon^2)$ if pure
- Many extensions and specializations, e.g. shadow tomography, Pauli strings...

Quantum State Tomography: Usually Works?

Plenty of methods known to almost surely *converge* to correct answer:

- Pick an informationally complete basis and measure lots of times
- Converges with $O(d^3/\epsilon^2)$ many samples (central limit theorem)
  - Can do $O(d^2/\epsilon^2)$ with a quantum computer
  - $O(d/\epsilon^2)$ if pure
- Many extensions and specializations, e.g. shadow tomography, Pauli strings...

The reconstruction step:

Given the measurement data, find $|\psi\rangle$

Quantum State Tomography: Reconstruction Step

Given the measurement data, find $|\psi\rangle$

Given the measurement data, find $|\psi\rangle$

$|\psi\rangle$ is a unit vector in $\mathbb{C}^d$. I have taken some number n of measurements, and now I would like to estimate ψ as accurately as possible.

- Measurement outcomes $|\gamma_i\rangle$ are (wlog) unit vectors in $\mathbb{C}^d$, each ψ has a likelihood $|\langle\psi|\gamma_i\rangle|^2$

Given the measurement data, find $|\psi\rangle$

$|\psi\rangle$ is a unit vector in $\mathbb{C}^d$. I have taken some number *n* of measurements, and now I would like to estimate ψ as accurately as possible.

- Measurement outcomes $|\gamma_i\rangle$ are (wlog) unit vectors in $\mathbb{C}^d$, each ψ has a likelihood $|\langle\psi|\gamma_i\rangle|^2$
- Measurements are already performed. Not a question of picking what to measure

Given the measurement data, find $|\psi\rangle$

$|\psi\rangle$ is a unit vector in $\mathbb{C}^d$. I have taken some number n of measurements, and now I would like to estimate ψ as accurately as possible.

- Measurement outcomes $|\gamma_i\rangle$ are (wlog) unit vectors in $\mathbb{C}^d$, each ψ has a likelihood $|\langle\psi|\gamma_i\rangle|^2$
- Measurements are already performed. Not a question of picking what to measure
- State space d is *not* big. Physically, only $\log(d)$ many qubits

Given the measurement data, find $|\psi\rangle$

$|\psi\rangle$ is a unit vector in $\mathbb{C}^d$. I have taken some number n of measurements, and now I would like to estimate ψ as accurately as possible.

- Measurement outcomes $|\gamma_i\rangle$ are (wlog) unit vectors in $\mathbb{C}^d$, each ψ has a likelihood $|\langle\psi|\gamma_i\rangle|^2$
- Measurements are already performed. Not a question of picking what to measure
- State space d is *not* big. Physically, only $\log(d)$ many qubits
- Estimating ψ equivalent to estimating a complete basis of its observables

Given the measurement data, find $|\psi\rangle$

$|\psi\rangle$ is a unit vector in $\mathbb{C}^d$. I have taken some number n of measurements, and now I would like to estimate ψ as accurately as possible.

- Measurement outcomes $|\gamma_i\rangle$ are (wlog) unit vectors in $\mathbb{C}^d$, each ψ has a likelihood $|\langle\psi|\gamma_i\rangle|^2$
- Measurements are already performed. Not a question of picking what to measure
- State space d is *not* big. Physically, only $\log(d)$ many qubits
- Estimating ψ equivalent to estimating a complete basis of its observables
- Also equivalent to estimating overall probability of this set of measurement outcomes (a partition function Z)

Given the measurement data, find $|\psi\rangle$

$|\psi\rangle$ is a unit vector in $\mathbb{C}^d$. I have taken some number n of measurements, and now I would like to estimate ψ as accurately as possible.

- Measurement outcomes $|\gamma_i\rangle$ are (wlog) unit vectors in $\mathbb{C}^d$, each ψ has a likelihood $|\langle\psi|\gamma_i\rangle|^2$
- Measurements are already performed. Not a question of picking what to measure
- State space d is *not* big. Physically, only $\log(d)$ many qubits
- Estimating ψ equivalent to estimating a complete basis of its observables
- Also equivalent to estimating overall probability of this set of measurement outcomes (a partition function Z)

Main result: this is NP-hard to approximate, even within an exponential factor!

Existing approaches all have some drawback

Existing approaches all have some drawback

- Produce a non-normalized $|\psi\rangle$ (hoping for ok $\langle\psi|\mathcal{O}|\psi\rangle$)

Quantum State Tomography: Reconstruction Step

Existing approaches all have some drawback

- Produce a non-normalized $|\psi\rangle$ (hoping for ok $\langle\psi|\mathcal{O}|\psi\rangle$)
- ρ is normalized with $\mathrm{Tr}[\rho] = 1$, but has negative eigenvalues

Quantum State Tomography: Reconstruction Step

Existing approaches all have some drawback

- Produce a non-normalized $|\psi\rangle$ (hoping for ok $\langle\psi|\mathcal{O}|\psi\rangle$)
- ρ is normalized with $\mathrm{Tr}[\rho] = 1$, but has negative eigenvalues
- Don't take into account different possible states

Existing approaches all have some drawback

- Produce a non-normalized $|\psi\rangle$ (hoping for ok $\langle\psi|\mathcal{O}|\psi\rangle$)
- ρ is normalized with $\mathrm{Tr}[\rho] = 1$, but has negative eigenvalues
- Don't take into account different possible states

Principled approach:

Quantum State Tomography: Reconstruction Step

Existing approaches all have some drawback

- Produce a non-normalized $|\psi\rangle$ (hoping for ok $\langle\psi|\mathcal{O}|\psi\rangle$)
- ρ is normalized with $\mathrm{Tr}[\rho] = 1$, but has negative eigenvalues
- Don't take into account different possible states

Principled approach: **Bayesian** statistics!

Quantum State Tomography: Reconstruction Step

Existing approaches all have some drawback

- Produce a non-normalized $|\psi\rangle$ (hoping for ok $\langle\psi|\mathcal{O}|\psi\rangle$)
- ρ is normalized with $\mathrm{Tr}[\rho] = 1$, but has negative eigenvalues
- Don't take into account different possible states

Principled approach: **Bayesian** statistics!

Integrate over all possible $|\psi\rangle$, weighted by the likelihood of observed data.

The classical case

- Unknown probability distribution P over d elements

The classical case

- Unknown probability distribution P over d elements
- A weighted die with d sides, a bag with d different colors of marble in it

The classical case

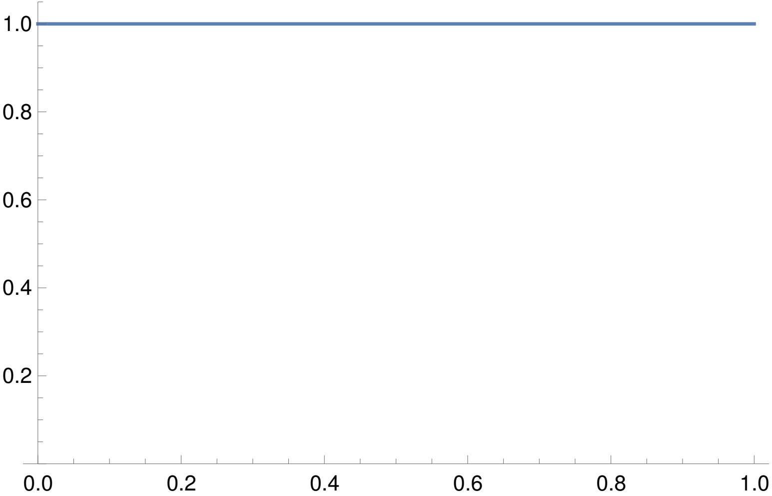
- Unknown probability distribution P over d elements
- A weighted die with d sides, a bag with d different colors of marble in it
- With $d = 3$, I have counts k_1, k_2, k_3 of my observations

The classical case

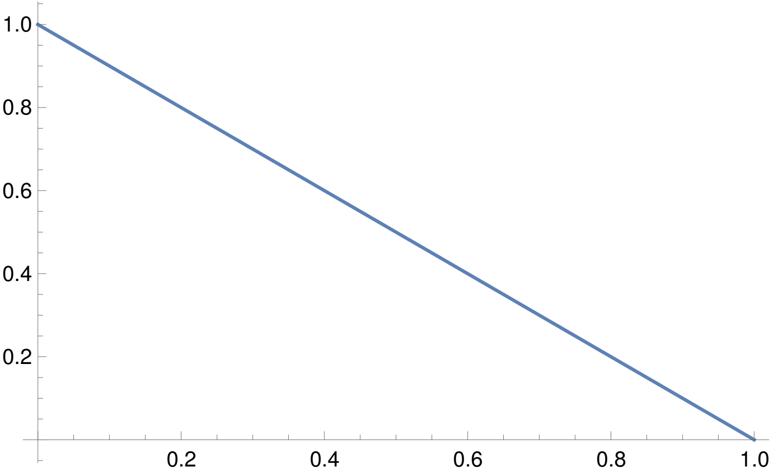
- Unknown probability distribution P over d elements
- A weighted die with d sides, a bag with d different colors of marble in it
- With $d = 3$, I have counts k_1, k_2, k_3 of my observations
- Want to know p_1, p_2, p_3.

Quantum State Tomography: Classical Analog

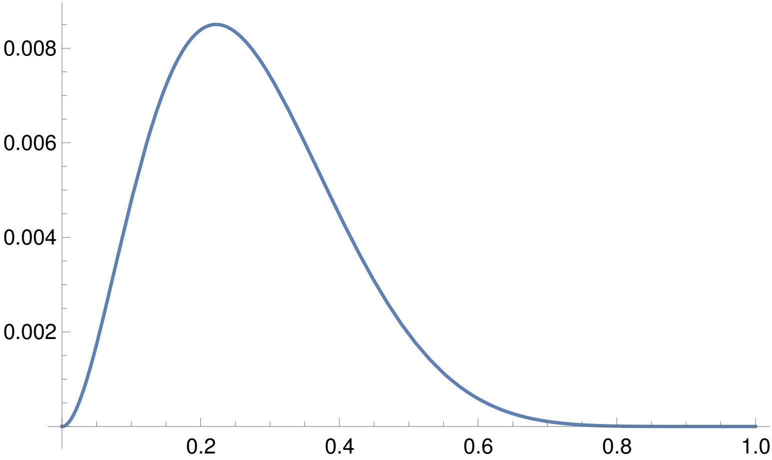
With $d = 2$ weighted coin, just trying to estimate one number: p_{Heads}. Initial distribution over possible p's is flat:



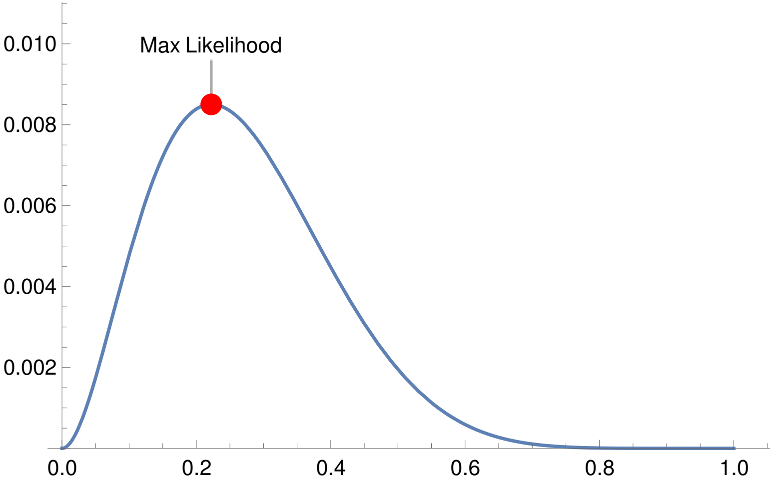
After flipping the coin and getting tails once, the likelihoods update. I can rule out $p_{Heads} = 1.0$.



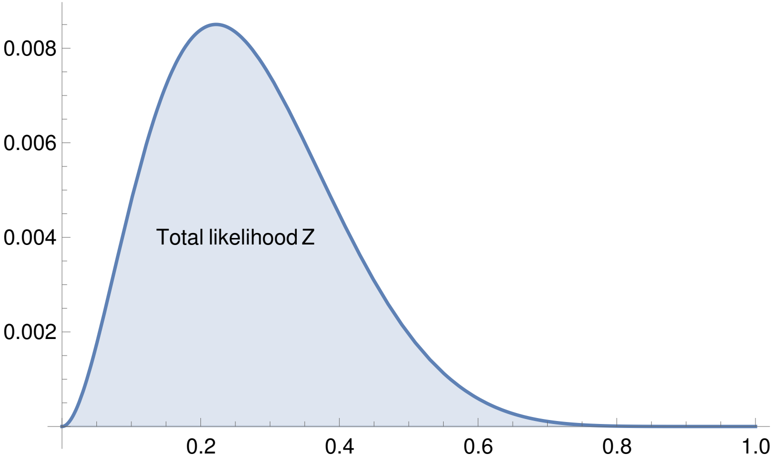
After 2 heads and 9 tails, the possible probabilities begin to concentrate:



After 2 heads and 9 tails, the possible probabilities begin to concentrate:

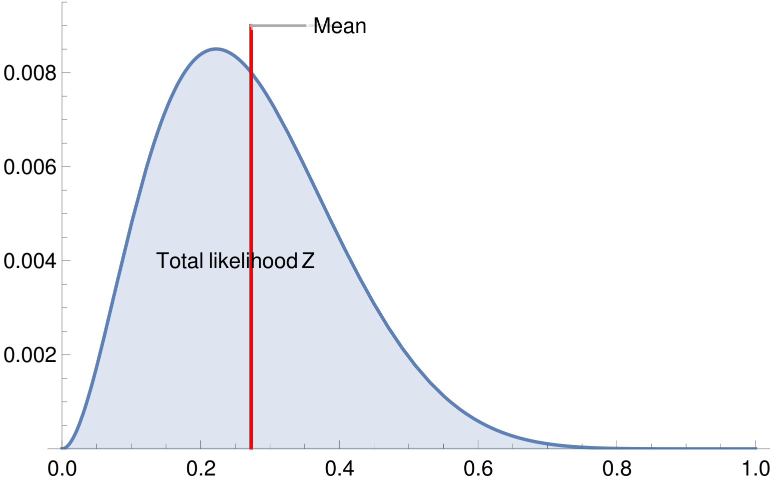


After 2 heads and 9 tails, the possible probabilities begin to concentrate:

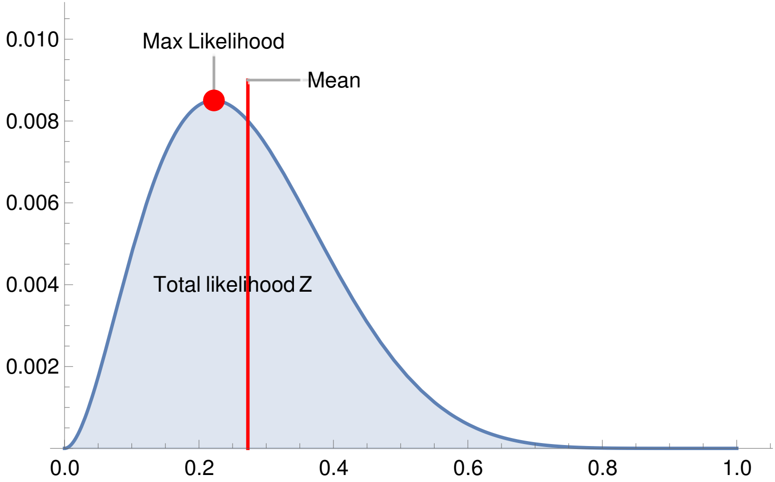


Quantum State Tomography: Classical Analog

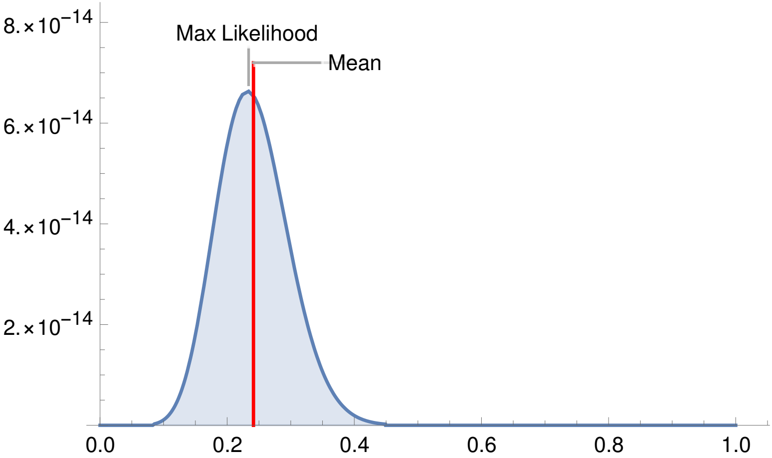
After 2 heads and 9 tails, the possible probabilities begin to concentrate:



After 2 heads and 9 tails, the possible probabilities begin to concentrate:



After 2 heads and 9 tails, the possible probabilities begin to concentrate:



Likelihood of a hypothetical p:

$$L(p) = p^{\#\text{ of Heads}}(1-p)^{\#\text{ of Tails}}$$

e.g. after 10 heads and 20 tails,

$$L(p) = p^{10}(1-p)^{20}$$

Quantum State Tomography: Classical Analog

Likelihood of a hypothetical p:

$$L(p) = p^{\#\text{ of Heads}}(1-p)^{\#\text{ of Tails}}$$

e.g. after 10 heads and 20 tails,

$$L(p) = p^{10}(1-p)^{20}$$

Maximum likelihood estimator (MLE) asks for the p that maximizes $L(p)$

Quantum State Tomography: Classical Analog

Likelihood of a hypothetical p:

$$L(p) = p^{\#\text{ of Heads}}(1-p)^{\#\text{ of Tails}}$$

e.g. after 10 heads and 20 tails,

$$L(p) = p^{10}(1-p)^{20}$$

Maximum likelihood estimator (MLE) asks for the p that maximizes $L(p)$

Chance of getting heads next time, is $\mathbb{E}[p]$, which is integrating p across possible coins:

$$\mathbb{E}[p] = \int_{\mathfrak{p}=0}^{1} \mathfrak{p}\, L(\mathfrak{p})\, d\mathfrak{p} = \int_{\mathfrak{p}=0}^{1} \mathfrak{p}\, \mathfrak{p}^{10}(1-\mathfrak{p})^{20}\, d\mathfrak{p}$$

With $d = 3$, I have counts k_1, k_2, k_3 of my observations:

$$Z = \iint\limits_{\mathbf{p}\in\Delta_3} L(\mathbf{p})\, d\mathbf{p} = \iint\limits_{\mathbf{p}\in\Delta_3} \mathfrak{p}_1^{k_1}\, \mathfrak{p}_2^{k_2}\, \mathfrak{p}_3^{k_3}\, d\mathbf{p}$$

With $d = 3$, I have counts k_1, k_2, k_3 of my observations:

$$Z = \iint\limits_{\mathbf{p}\in\Delta_3} L(\mathbf{p})\, d\mathbf{p} = \iint\limits_{\mathbf{p}\in\Delta_3} \mathfrak{p}_1^{k_1}\, \mathfrak{p}_2^{k_2}\, \mathfrak{p}_3^{k_3}\, d\mathbf{p}$$

$$\mathbb{E}[p_1] = \iint\limits_{\mathbf{p}\in\Delta_3} \mathfrak{p}_1 \left(\mathfrak{p}_1^{k_1}\, \mathfrak{p}_2^{k_2}\, \mathfrak{p}_3^{k_3}\right) d\mathbf{p}$$

$\implies$ chance of getting outcome “1” on another sample.

Quantum State Tomography: Classical Analog

With $d = 3$, I have counts k_1, k_2, k_3 of my observations:

$$Z = \iint_{\mathbf{p}\in\Delta_3} L(\mathbf{p})\,d\mathbf{p} = \iint_{\mathbf{p}\in\Delta_3} \mathfrak{p}_1^{k_1}\,\mathfrak{p}_2^{k_2}\,\mathfrak{p}_3^{k_3}\,d\mathbf{p}$$

$$\mathbb{E}[p_1] = \iint_{\mathbf{p}\in\Delta_3} \mathfrak{p}_1 \left(\mathfrak{p}_1^{k_1}\,\mathfrak{p}_2^{k_2}\,\mathfrak{p}_3^{k_3}\right) d\mathbf{p}$$

$\Longrightarrow$ chance of getting outcome “1” on another sample.

With $d = 3$, I have counts k_1, k_2, k_3 of my observations:

$$Z = \iint_{\mathbf{p} \in \Delta_3} L(\mathbf{p})\, d\mathbf{p} = \iint_{\mathbf{p} \in \Delta_3} \mathfrak{p}_1^{k_1}\, \mathfrak{p}_2^{k_2}\, \mathfrak{p}_3^{k_3}\, d\mathbf{p}$$

$$\mathbb{E}[p_1] = \iint_{\mathbf{p} \in \Delta_3} \mathfrak{p}_1 \left(\mathfrak{p}_1^{k_1}\, \mathfrak{p}_2^{k_2}\, \mathfrak{p}_3^{k_3} \right) d\mathbf{p}$$

$\implies$ chance of getting outcome "1" on another sample.

Not easy immediately, really this is

$$\mathbb{E}[p_1] = \int_{\mathfrak{p}_1 = 0}^{1} \int_{\mathfrak{p}_2 = 0}^{1 - \mathfrak{p}_1} \mathfrak{p}_1^{k_1}\, \mathfrak{p}_2^{k_2}\, (1 - \mathfrak{p}_1 - \mathfrak{p}_2)^{k_3}\, d\mathbf{p}$$

Quantum State Tomography: Classical Analog

With $d = 3$, I have counts k_1, k_2, k_3 of my observations:

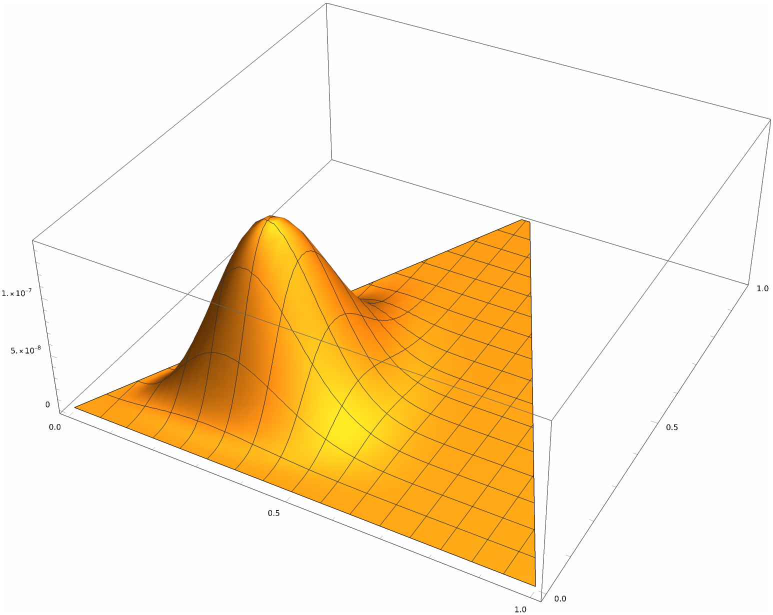
$$Z = \iint_{\mathbf{p} \in \Delta_3} L(\mathbf{p})\, d\mathbf{p} = \iint_{\mathbf{p} \in \Delta_3} \mathfrak{p}_1^{k_1}\, \mathfrak{p}_2^{k_2}\, \mathfrak{p}_3^{k_3}\, d\mathbf{p}$$

$$\mathbb{E}[p_1] = \iint_{\mathbf{p} \in \Delta_3} \mathfrak{p}_1 \left(\mathfrak{p}_1^{k_1}\, \mathfrak{p}_2^{k_2}\, \mathfrak{p}_3^{k_3} \right) d\mathbf{p}$$

$\implies$ chance of getting outcome "1" on another sample.

But! Integrand is **convex**, and so can be computed efficiently! Picture:

Quantum State Tomography: Classical Analog



Quantum Case

- Bayesian statistics requires integrating over all possible states

Quantum Case

- Bayesian statistics requires integrating over all possible states
- *Approximate* answer: integrating over all *plausible* states
- Classically tractable, in $O(nd^3)$ time

Quantum Case

- Bayesian statistics requires integrating over all possible states
- *Approximate* answer: integrating over all *plausible* states
- Classically tractable, in $O(nd^3)$ time
- Quantum case is **surprisingly hard**, exponentially(!) hard in d

Quantum Case

- Bayesian statistics requires integrating over all possible states
- *Approximate* answer: integrating over all *plausible* states
- Classically tractable, in $O(nd^3)$ time
- Quantum case is **surprisingly hard**, exponentially(!) hard in d

Going to prove problem is difficult in general, by exhibiting a particular set of measurement outcomes where this integral must be hard to compute (reducing to an NP-hard problem)

Going to prove problem is difficult in general, by exhibiting a particular set of measurement outcomes where this integral must be hard to compute (reducing to an NP-hard problem)

Integral

$$Z = \int_\psi L(\psi)\, d\psi = \int_\psi \prod_i |\langle\psi|\gamma_i\rangle|^2 \; d\psi$$

The integrand:

$$\prod_i |\langle\psi|\gamma_i\rangle|^2$$

is a polynomial in the coordinates of ψ. Each observation γ_i adds a zero hyperplane to this polynomial: zero chance that ψ is perpendicular to γ_i.

Going to prove problem is difficult in general, by exhibiting a particular set of measurement outcomes where this integral must be hard to compute (reducing to an NP-hard problem)

Integral
$$Z = \int_{\psi} L(\psi)\, d\psi = \int_{\psi} \prod_i |\langle \psi | \gamma_i \rangle|^2 \; d\psi$$

The integrand:
$$\prod_i |\langle \psi | \gamma_i \rangle|^2$$

is a polynomial in the coordinates of ψ. Each observation γ_i adds a zero hyperplane to this polynomial: zero chance that ψ is perpendicular to γ_i.

Lots of zeros $\rightarrow$ highly oscillatory function $\rightarrow$ hard to maximize.

Hardness of state estimation

Suppose we have measured in the computational basis (or any fixed basis) severalt times, and observed each basis state once:

Hardness of state estimation

Suppose we have measured in the computational basis (or any fixed basis) severalt times, and observed each basis state once:

- $|\gamma_1\rangle = |1\rangle = (1, 0, 0, \dots)$
- $|\gamma_2\rangle = |2\rangle = (0, 1, 0, \dots)$
- etc.

Suppose we have measured in the computational basis (or any fixed basis) severalt times, and observed each basis state once:

- $|\gamma_1\rangle = |1\rangle = (1, 0, 0, \dots)$
- $|\gamma_2\rangle = |2\rangle = (0, 1, 0, \dots)$
- etc.

ψ can't have any zero (or small) entries. If kth entry is zero, then $\langle\psi|\gamma_k\rangle$ is zero, an impossible observation

Hardness of state estimation

Suppose we have measured in the computational basis (or any fixed basis) severalt times, and observed each basis state once:

- $|\gamma_1\rangle = |1\rangle = (1, 0, 0, \dots)$
- $|\gamma_2\rangle = |2\rangle = (0, 1, 0, \dots)$
- etc.

ψ can't have any zero (or small) entries. If kth entry is zero, then $\langle\psi|\gamma_k\rangle$ is zero, an impossible observation

By taking many copies of each basis vector (say, poly(d) many), we ensure that each entry of ψ is roughly equal in magnitude.

Only significant terms in the integral are:

$$\psi \approx \frac{1}{\sqrt{d}}(e^{i\theta_1}, e^{i\theta_2}, \ldots e^{i\theta_d})$$

By symmetry, we can fix $\theta_1 = 0$. Not physical anyway

Only significant terms in the integral are:

$$\psi \approx \frac{1}{\sqrt{d}}(e^{i\theta_1}, e^{i\theta_2}, \ldots e^{i\theta_d})$$

By symmetry, we can fix $\theta_1 = 0$. Not physical anyway
Assume we have measurement outcomes

$$\gamma_{+,2} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0, 0, \ldots\right)$$

$$\gamma_{-,2} = \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, 0, 0, 0, \ldots\right)$$

Then $e^{i\theta_2}$ cannot be close to -1 or $+1$. Probability is maximized with $+i$ and $-i$.

Only significant terms in the integral are:

$$\psi \approx \frac{1}{\sqrt{d}}(e^{i\theta_1}, e^{i\theta_2}, \dots e^{i\theta_d})$$

By symmetry, we can fix $\theta_1 = 0$. Not physical anyway
Assume we have measurement outcomes

$$\gamma_{+,2} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0, 0, \dots\right)$$

$$\gamma_{-,2} = \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, 0, 0, 0, \dots\right)$$

Then $e^{i\theta_2}$ cannot be close to -1 or $+1$. Probability is maximized with $+i$ and $-i$.

By taking many copies of $\gamma_{+,k}$ and $\gamma_{-,k}$, ensure that all $e^{i\theta_k}$ are close to $+i$ or $-i$.

Hardness of state estimation: Qubit example

For $d = 2$ qubit, this looks like:

1. Many Z basis measurements, getting both $|0\rangle$ and $|1\rangle$ many times. $\implies$ Must be near uniform superposition
2. Many X basis measurements, getting both $|+\rangle$ and $|-\rangle$ many times. $\implies$ Must be a $\pm Y$ eigenstate, but *we don't know which*

Hardness of state estimation: Qubit example

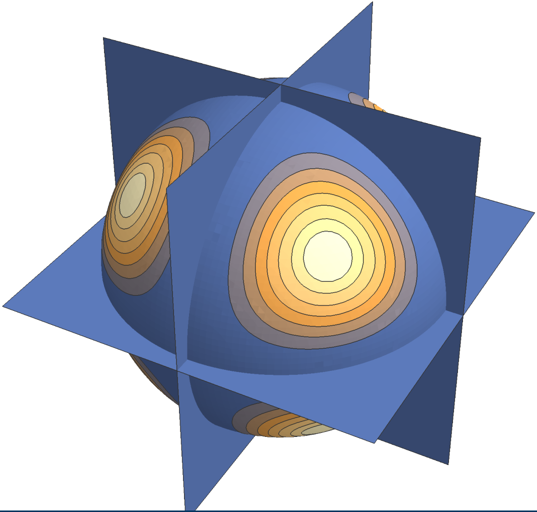
For $d = 2$ qubit, this looks like:

1. Many Z basis measurements, getting both $|0\rangle$ and $|1\rangle$ many times. $\implies$ Must be near uniform superposition
2. Many X basis measurements, getting both $|+\rangle$ and $|-\rangle$ many times. $\implies$ Must be a $\pm Y$ eigenstate, but *we don't know which*

For higher d, we get exponentially many different options, 2^{d-1} many

Illustration for d=3

$$\psi \approx \frac{1}{\sqrt{d}}(1, \pm i, \cdots \pm i)$$



Hardness of state estimation

Now integral concentrates on these 2^{d-1} discrete points: total integral is proportional to sum of likelihood of these points, plus an exponentially smaller additive error (the other implausible points).

Hardness of state estimation

Now integral concentrates on these 2^{d-1} discrete points: total integral is proportional to sum of likelihood of these points, plus an exponentially smaller additive error (the other implausible points).

Cut out some of that list of points.
The state

$$\gamma_{(234)} = \left(0, \frac{-2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, 0, 0, 0 \dots \right)$$

is perpendicular to $(0, 1, 1, 1, 0, 0, 0 \dots)$, and eliminates the possibility that all three signs are equal.

Now integral concentrates on these 2^{d-1} discrete points: total integral is proportional to sum of likelihood of these points, plus an exponentially smaller additive error (the other implausible points).

Cut out some of that list of points.
The state

$$\gamma_{(234)} = \left(0, \frac{-2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, 0, 0, 0 \dots\right)$$

is perpendicular to $(0, 1, 1, 1, 0, 0, 0 \dots)$, and eliminates the possibility that all three signs are equal.

$$\gamma_{(234),B} = \left(0, \frac{1}{\sqrt{6}}, \frac{-2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, 0, 0, 0 \dots\right)$$

$$\gamma_{(234),C} = \left(0, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{-2}{\sqrt{6}}, 0, 0, 0 \dots\right)$$

to keep the probability symmetric across which of the three signs should differ.

Hardness: Main Result

Reduce from NOT-ALL-EQUAL-3SAT: given some triples of variables, finding an assignment of Boolean variables such that no specified triple has all equal values. NP-complete.

- "Set $\{v_1, v_2, v_3, v_4, v_5\}$ so that each of (v_1, v_2, v_4), (v_1, v_3, v_5), (v_2, v_4, v_5), (v_2, v_3, v_5) have at least one TRUE and one FALSE"
- "Set the phases in $|\psi\rangle = \frac{1}{\sqrt{6}}(1, s_1, s_2, s_3, s_4, s_5)\dots$"

Hardness: Main Result

Reduce from NOT-ALL-EQUAL-3SAT: given some triples of variables, finding an assignment of Boolean variables such that no specified triple has all equal values. NP-complete.

- "Set $\{v_1, v_2, v_3, v_4, v_5\}$ so that each of (v_1, v_2, v_4), (v_1, v_3, v_5), (v_2, v_4, v_5), (v_2, v_3, v_5) have at least one TRUE and one FALSE"
- "Set the phases in $|\psi\rangle = \frac{1}{\sqrt{6}}(1, s_1, s_2, s_3, s_4, s_5)\ldots$"

Given a NAE-3SAT problem on v variables, can write down a set of $n = \text{poly}(v)$ measurements Γ on $d = v + 1$ variables, such that:

- For each solution to the original problem, there is exactly one $|\psi\rangle$ with high likelihood, at least $f(n)$.
- If no solutions to the original, all $|\psi\rangle$ are exponentially unlikely, at most $f(n)2^{-\text{poly}(d)}$.

Widely believed that P $\neq$ NP, that you cannot solve NAE-3SAT efficiently. If you had an algorithm to find a good $|\psi\rangle$ given the measurement data, you could use it to solve NAE-3SAT, so we conclude that this should be impossible.

Widely believed that P $\neq$ NP, that you cannot solve NAE-3SAT efficiently. If you had an algorithm to find a good $|\psi\rangle$ given the measurement data, you could use it to solve NAE-3SAT, so we conclude that this should be impossible.

Informal: NP-hard to find $|\psi\rangle$, whose likelihood is within exponential factor of the optimal answer.

Widely believed that P $\neq$ NP, that you cannot solve NAE-3SAT efficiently. If you had an algorithm to find a good $|\psi\rangle$ given the measurement data, you could use it to solve NAE-3SAT, so we conclude that this should be impossible.

Informal: NP-hard to find $|\psi\rangle$, whose likelihood is within exponential factor of the optimal answer.

Formally: for any $C < 1$, NP-hard to find $|\psi\rangle$ with likelihood within a factor 2^{n^C} of the optimal answer.

Partition function Z

Partition function Z:

- Approximately counts good "solutions" ψ, each one contributes a similar amount
- Can be used to recover a solution ψ
- Must be hard to approximate within exponential factor too!
- Turns out to mathematically take the form of a **positive semidefinite permanent**

People care about PSD permanents

A series of works from the mathematical side:

1. Marcus '63: $n!$ factor approximation
2. Rahimi-Keshari/Lund/Ralph '17: Stockmeyer counting approach
3. Anari/Gurvits/Gharan/Saberi '17: 4.85^n approximation
4. Grier & Schaeffer '18: Hard to compute exactly
5. Barvinok '20: Fast algorithm for $\lambda_{\max}/\lambda_{\min} \leq 2$
6. Yuan & Parrilo '21: Fast algorithm, also requires close eigenvalues

Conjectured by several to be easy to approximate

People care about PSD permanents

Connections to thermal BosonSampling devices:

1. *Connection between BosonSampling with quantum and classical input states*, Kim et al.
2. *Multiboson correlation interferometry with multimode thermal sources*, Tamma et al.
3. Chakhmakhchyan et al. '17 Quantum inspired(!) algorithm for estimation
4. Mukerji & Yang '22, Lim & Oh '22: More quantum algorithms
5. Chabaud et al. '22: "Quantum-inspired identities"

People care about PSD permanents

Connections to thermal BosonSampling devices:

1. *Connection between BosonSampling with quantum and classical input states*, Kim et al.
2. *Multiboson correlation interferometry with multimode thermal sources*, Tamma et al.
3. Chakhmakhchyan et al. '17 Quantum inspired(!) algorithm for estimation
4. Mukerji & Yang '22, Lim & Oh '22: More quantum algorithms
5. Chabaud et al. '22: "Quantum-inspired identities"

The hardness of finding $|\psi\rangle \implies$ hardness of approximation $Z \implies$ hardness of approximating these permanents! Thermal inputs should not lose much computational power

Wrapping up...

- **Not** hard because quantum has exponentially big Hilbert space
- **Easy** for classical case, $O(nd^3)$
- Hard for logarithmically many qubits / particles
- Classical: probabilities are positive. Quantum: sign problem.

Wrapping up...

- **Not** hard because quantum has exponentially big Hilbert space
- **Easy** for classical case, $O(nd^3)$
- Hard for logarithmically many qubits / particles
- Classical: probabilities are positive. Quantum: sign problem. **Can't guess the signs**

Wrapping up...

- **Not** hard because quantum has exponentially big Hilbert space
- **Easy** for classical case, $O(nd^3)$
- Hard for logarithmically many qubits / particles
- Classical: probabilities are positive. Quantum: sign problem. **Can't guess the signs**
- There is a $O(n^d)$ algorithm for computing it exactly

Wrapping up...

- **Not** hard because quantum has exponentially big Hilbert space
- **Easy** for classical case, $O(nd^3)$
- Hard for logarithmically many qubits / particles
- Classical: probabilities are positive. Quantum: sign problem. **Can't guess the signs**
- There is a $O(n^d)$ algorithm for computing it exactly
- For a fixed Hilbert space dimension d, it's tractable in n

Wrapping up...

- **Not** hard because quantum has exponentially big Hilbert space
- **Easy** for classical case, $O(nd^3)$
- Hard for logarithmically many qubits / particles
- Classical: probabilities are positive. Quantum: sign problem. **Can't guess the signs**
- There is a $O(n^d)$ algorithm for computing it exactly
- For a fixed Hilbert space dimension d, it's tractable in n
- Proved that $O(n^d)$ is essentially optimal (parameterized complexity)

Wrapping up...

- **Not** hard because quantum has exponentially big Hilbert space
- **Easy** for classical case, $O(nd^3)$
- Hard for logarithmically many qubits / particles
- Classical: probabilities are positive. Quantum: sign problem. **Can't guess the signs**
- There is a $O(n^d)$ algorithm for computing it exactly
- For a fixed Hilbert space dimension d, it's tractable in n
- Proved that $O(n^d)$ is essentially optimal (parameterized complexity)
- PSD permanent hardness

Wrapping up...

- **Not** hard because quantum has exponentially big Hilbert space
- **Easy** for classical case, $O(nd^3)$
- Hard for logarithmically many qubits / particles
- Classical: probabilities are positive. Quantum: sign problem. **Can't guess the signs**
- There is a $O(n^d)$ algorithm for computing it exactly
- For a fixed Hilbert space dimension d, it's tractable in n
- Proved that $O(n^d)$ is essentially optimal (parameterized complexity)
- PSD permanent hardness
- State estimation requires some other assumption (about "typical" measurement outcomes)

Part II

Green's Function Estimation

Green's Function Estimation

- We would like to use quantum computers to calculate physical quantities of interest

Green's Function Estimation

- We would like to use quantum computers to calculate physical quantities of interest
  - Ground state properties

- We would like to use quantum computers to calculate physical quantities of interest
  - Ground state properties
  - Spectral functions $A(k, \omega)$

Green's Function Estimation

- We would like to use quantum computers to calculate physical quantities of interest
  - Ground state properties
  - Spectral functions $A(k, \omega)$
  - Scattering amplitudes

Green's Function Estimation

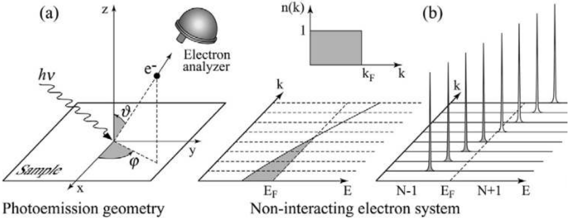
- We would like to use quantum computers to calculate physical quantities of interest
  - Ground state properties
  - Spectral functions $A(k, \omega)$
  - Scattering amplitudes
- Given a state $|\psi\rangle$ and observable $\mathcal{O}$, we can directly sample from $\langle\psi|\mathcal{O}|\psi\rangle$ - but this is very noisy

Green's Function Estimation

- We would like to use quantum computers to calculate physical quantities of interest
  - Ground state properties
  - Spectral functions $A(k, \omega)$
  - Scattering amplitudes
- Given a state $|\psi\rangle$ and observable $\mathcal{O}$, we can directly sample from $\langle\psi|\mathcal{O}|\psi\rangle$ - but this is very noisy
- Depending on the observable(s) we care about, we can likely do much better!

Green's Function Estimation

- We would like to use quantum computers to calculate physical quantities of interest
  - Ground state properties
  - Spectral functions $A(k, \omega)$
  - Scattering amplitudes
- Given a state $|\psi\rangle$ and observable $\mathcal{O}$, we can directly sample from $\langle\psi|\mathcal{O}|\psi\rangle$ - but this is very noisy
- Depending on the observable(s) we care about, we can likely do much better!
- Green's function: $G(t) = \langle\psi| \exp(iHt)\mathcal{O}^\dagger \exp(-iHt)\mathcal{O}|\psi\rangle$



$$I(\mathbf{k},\omega) \propto A(\mathbf{k},\omega)$$

$$A(\mathbf{k},\omega) = -\frac{1}{\pi}\Im\hat{G}(\mathbf{k},\omega)$$

Here $\hat{G}(\omega)$ is the Fourier transform of $G(t)$, where $\mathcal{O}$ is $a_{\mathbf{k}}^{\dagger}$. G also gives linear response theory (electrical or thermal conductivity) via Kubo relations, etc.

Source: Probing the Electronic Structure of Complex Systems by ARPES, Damascelli 2004

Green's Function Estimation

Usual Green's function has $\mathcal{O}$ as creation operator,

$$G(t) = \langle\psi| \exp(iHt) a \exp(-iHt) a^\dagger |\psi\rangle$$

- Our discussion will apply to, generally, any correlation functions
- We restrict to unitary $\mathcal{O}$
  - Can recover $\mathcal{O}$: take unitary $\mathcal{O}_\pm = a^\dagger \pm a$
  - Obtain associated correlation functions $G_\pm$
  - Standard G is $\frac{G_+ + G_-}{2}$
- Unitarity makes the math and quantum circuits much simpler

Green's Function Estimation

Usual Green's function has $\mathcal{O}$ as creation operator,

$$G(t) = \langle\psi| \exp(iHt) a \exp(-iHt) a^\dagger |\psi\rangle$$

- Our discussion will apply to, generally, any correlation functions
- We restrict to unitary $\mathcal{O}$
  - Can recover $\mathcal{O}$: take unitary $\mathcal{O}_\pm = a^\dagger \pm a$
  - Obtain associated correlation functions $G_\pm$
  - Standard G is $\frac{G_+ + G_-}{2}$
- Unitarity makes the math and quantum circuits much simpler
- I'll probably keep calling them all "Green's functions" ☺

In principle, all our techniques could apply to $\langle \mathcal{O}(t) \rangle$, or even $\rho(t)$

In principle, all our techniques could apply to $\langle\mathcal{O}(t)\rangle$, or even $\rho(t)$

Compared with just getting a single $\langle\mathcal{O}\rangle$ or ρ, this is harder, because we're trying to do across all time

In principle, all our techniques could apply to $\langle\mathcal{O}(t)\rangle$, or even $\rho(t)$

Compared with just getting a single $\langle\mathcal{O}\rangle$ or ρ, this is harder, because we're trying to do across all time

$G(t)$ is just a single complex scalar, so the dimension is very low (two). We focus our attention will be on dealing with time-dependence

What We Sample

$$G(t) = \langle\psi| \exp(iHt)\mathcal{O}^\dagger \exp(-iHt)\mathcal{O}|\psi\rangle$$

- At each t, $|G(t)| \leq 1$
- One $n = 2$ quantum phase estimation gives a random bit, $p = \frac{1+\Re[G(t)]}{2}$
- Slight modification, can also get $p = \frac{1+\Im[G(t)]}{2}$

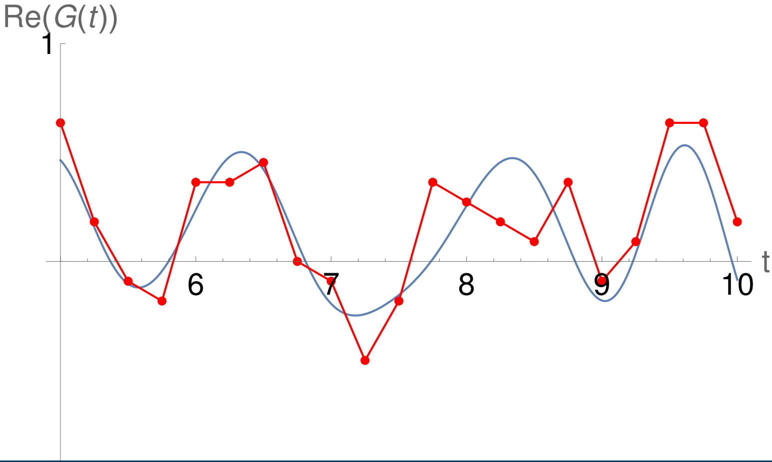
What We Sample

$$G(t) = \langle\psi| \exp(iHt)\mathcal{O}^\dagger \exp(-iHt)\mathcal{O}|\psi\rangle$$

- At each t, $|G(t)| \leq 1$
- One $n = 2$ quantum phase estimation gives a random bit, $p = \frac{1+\Re[G(t)]}{2}$
- Slight modification, can also get $p = \frac{1+\Im[G(t)]}{2}$
- After N runs, can measure $G(t)$ to $1/\sqrt{N}$ accuracy
- ... but we want to know $G(t)$ across all (or at least a full interval) of time!
- Simplest approach: linear interpolation

$S = 1$ XXX Heisenberg model, ground state excitation

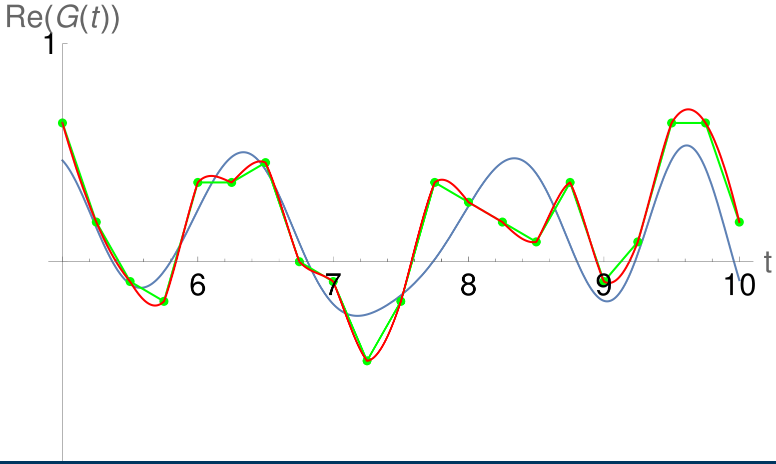
$$H = \sum_{i=1}^{L} \vec{S}_i \cdot \vec{S}_{i+1}, \qquad L = 6$$



Cubic interpolation

$S = 1$ XXX Heisenberg model, ground state excitation

$$H = \sum_{i=1}^{L} \vec{S}_i \cdot \vec{S}_{i+1}, \qquad L = 6$$



Linear interpolation *looks* wrong - we have other prior knowledge of the function!

- Space of all continuous bounded functions $C([0, t]) \cap B([0, t])$

Linear interpolation *looks* wrong - we have other prior knowledge of the function!

- Space of all continuous bounded functions $C([0,t]) \cap B([0,t])$
- Smoothness: $C^\infty([0,t])$

Linear interpolation *looks* wrong - we have other prior knowledge of the function!

- Space of all continuous bounded functions $C([0,t]) \cap B([0,t])$
- Smoothness: $C^{\infty}([0,t])$
- Mixture of Fourier components (frequencies excited by $\mathcal{O}$)

Linear interpolation *looks* wrong - we have other prior knowledge of the function!

- Space of all continuous bounded functions $C([0,t]) \cap B([0,t])$
- Smoothness: $C^\infty([0,t])$
- Mixture of Fourier components (frequencies excited by $\mathcal{O}$)
- Other facts, like $G(0) = 1$, limited energy, etc.

Linear interpolation *looks* wrong - we have other prior knowledge of the function!

- Space of all continuous bounded functions $C([0,t]) \cap B([0,t])$
- Smoothness: $C^\infty([0,t])$
- Mixture of Fourier components (frequencies excited by $\mathcal{O}$)
- Other facts, like $G(0) = 1$, limited energy, etc.

Most principled approach:

Integrate over all possible functions, weight by their likelihood (prior and observations), and take pointwise mean

Linear interpolation *looks* wrong - we have other prior knowledge of the function!

- Space of all continuous bounded functions $C([0,t]) \cap B([0,t])$
- Smoothness: $C^\infty([0,t])$
- Mixture of Fourier components (frequencies excited by $\mathcal{O}$)
- Other facts, like $G(0) = 1$, limited energy, etc.

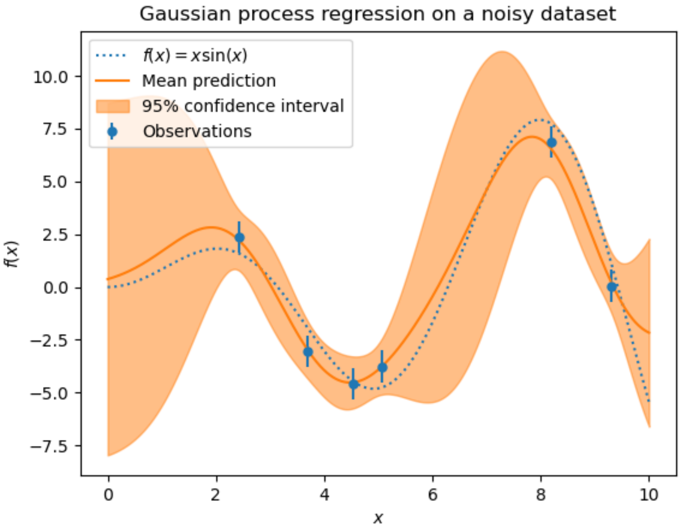
Most principled approach:

Integrate over all possible functions, weight by their likelihood (prior and observations), and take pointwise mean

$\implies$ Totally intractable

Gaussian Process (GP)

- Approximate your prior knowledge by an infinite-dimensional Gaussian distribution
  - Gaussian in the vector space of functions, not that the functions themselves look Gaussian
- Each finite set of points $\{G(t_0), G(t_1), G(t_2), \ldots\}$ is a multivariate Gaussian distribution



Source: Scikit-learn Gaussian Process

Gaussian Process (GP)

- Approximate your prior knowledge by an infinite-dimensional Gaussian distribution
  - Gaussian in the vector space of functions, not that the functions themselves look Gaussian
- Each finite set of points $\{G(t_0), G(t_1), G(t_2), \ldots\}$ is a multivariate Gaussian distribution
- Can be efficiently evaluated exactly, $\sim O(N^3)$ time
- All marginals are Gaussian, prediction is the mean
- Prior is specified by "kernel", $K(x) = \langle G(t)G(t+x)\rangle$
- Appropriate kernels ensure smoothness

Gaussian Process with Link Function

- Normal GP: never bounded

Gaussian Process with Link Function

- Normal GP: never bounded
  - Always some nonzero probability that $G(t) = 100$, bad

Gaussian Process with Link Function

- Normal GP: never bounded
  - Always some nonzero probability that $G(t) = 100$, bad
- Solution: GP $f(t)$, map $G(t) = \sigma(f(t))$ for some nonlinear σ

Gaussian Process with Link Function

- Normal GP: never bounded
  - Always some nonzero probability that $G(t) = 100$, bad
- Solution: GP $f(t)$, map $G(t) = \sigma(f(t))$ for some nonlinear σ
  - Hope that some "hidden" function f is "linear" like a Gaussian is

Gaussian Process with Link Function

- Normal GP: never bounded
  - Always some nonzero probability that $G(t) = 100$, bad
- Solution: GP $f(t)$, map $G(t) = \sigma(f(t))$ for some nonlinear σ
  - Hope that some "hidden" function f is "linear" like a Gaussian is
  - Compress f from $\mathbb{R}$ down to our interval $[-1, 1]$

Gaussian Process with Link Function

- Normal GP: never bounded
  - Always some nonzero probability that $G(t) = 100$, bad
- Solution: GP $f(t)$, map $G(t) = \sigma(f(t))$ for some nonlinear σ
  - Hope that some "hidden" function f is "linear" like a Gaussian is
  - Compress f from $\mathbb{R}$ down to our interval $[-1, 1]$
- Well-studied problem in ML, $G(t) \in [0, 1]$ is a true/false label

Gaussian Process with Link Function

- Normal GP: never bounded
  - Always some nonzero probability that $G(t) = 100$, bad
- Solution: GP $f(t)$, map $G(t) = \sigma(f(t))$ for some nonlinear σ
  - Hope that some "hidden" function f is "linear" like a Gaussian is
  - Compress f from $\mathbb{R}$ down to our interval $[-1, 1]$
- Well-studied problem in ML, $G(t) \in [0, 1]$ is a true/false label
  - Underlying function $f(t)$ is the logit or "log-odds"

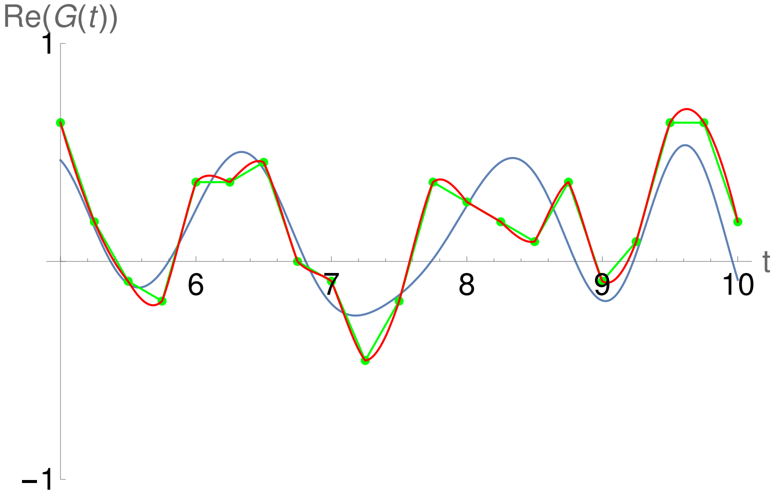
Gaussian Process with Link Function

- Normal GP: never bounded
  - Always some nonzero probability that $G(t) = 100$, bad
- Solution: GP $f(t)$, map $G(t) = \sigma(f(t))$ for some nonlinear σ
  - Hope that some "hidden" function f is "linear" like a Gaussian is
  - Compress f from $\mathbb{R}$ down to our interval $[-1, 1]$
- Well-studied problem in ML, $G(t) \in [0, 1]$ is a true/false label
  - Underlying function $f(t)$ is the logit or "log-odds"
- No longer exact, but saddle-point approximation very good (convexity!)

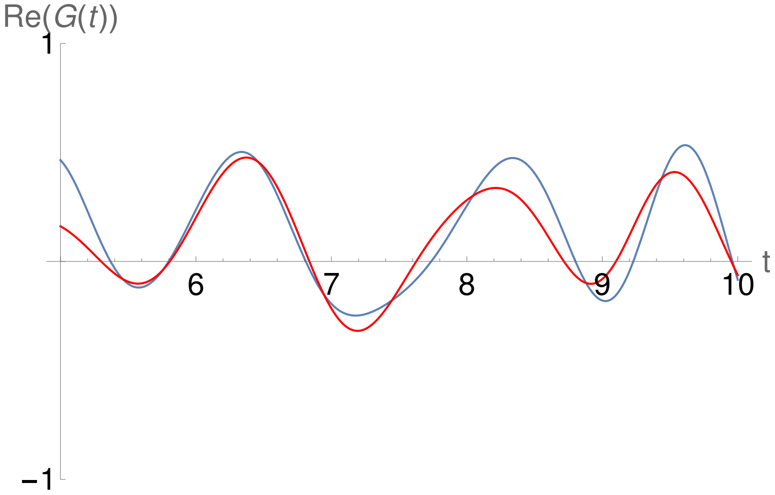
Gaussian Process with Link Function

- Normal GP: never bounded
  - Always some nonzero probability that $G(t) = 100$, bad
- Solution: GP $f(t)$, map $G(t) = \sigma(f(t))$ for some nonlinear σ
  - Hope that some "hidden" function f is "linear" like a Gaussian is
  - Compress f from $\mathbb{R}$ down to our interval $[-1, 1]$
- Well-studied problem in ML, $G(t) \in [0, 1]$ is a true/false label
  - Underlying function $f(t)$ is the logit or "log-odds"
- No longer exact, but saddle-point approximation very good (convexity!)
- Extended this to 2D, $G(t)$ in complex unit disk, $(f_{\Re}, f_{\Im})$

Cubic interpolation



GP interpolation



- At low sampling (pictures above, 500 runs), already 27% better (RMS error on sampled interval)
- Moderately better scaling - not just constant factor reduction

- At low sampling (pictures above, 500 runs), already 27% better (RMS error on sampled interval)
- Moderately better scaling - not just constant factor reduction
- Downside: starts to run very slowly at $N > 1000$
- Could be improved with careful linear algebra routines, matrix sparsity, etc.

Bayesian Statistics: Gaussian Process

- At low sampling (pictures above, 500 runs), already 27% better (RMS error on sampled interval)
- Moderately better scaling - not just constant factor reduction
- Downside: starts to run very slowly at $N > 1000$
- Could be improved with careful linear algebra routines, matrix sparsity, etc.
- Good kernel function, link function is somewhat system dependent

Fact:

$$G(t) = \sum_k a_k \exp(i\omega_k t)$$

$$a_k \geq 0, \quad \sum a_k = 1$$

If $|\psi\rangle$ is ground state, $\omega_k \geq 0$ as well.

Fact:

$$G(t) = \sum_k a_k \exp(i\omega_k t)$$

$$a_k \geq 0, \quad \sum a_k = 1$$

If $|\psi\rangle$ is ground state, $\omega_k \geq 0$ as well.

$\implies$ Big statement about Fourier transform! All phases are zero, and L^1 distribution

Find:

$$G(t) = \sum_k a_k \exp(i\omega_k t)$$

$$a_k \geq 0, \quad \sum a_k = 1$$

that best fits the observed data.

Find:

$$G(t) = \sum_k a_k \exp(i\omega_k t)$$

$$a_k \geq 0, \quad \sum a_k = 1$$

that best fits the observed data.

Difficult, highly nonlinear in ω_k. ☹

Find:

$$G(t) = \sum_k a_k \exp(i\omega_k t)$$

$$a_k \geq 0, \quad \sum a_k = 1$$

that best fits the observed data. Fix some dense set of ω_k, say, $\{-10, -9.9, -9.8 \cdots + 9.9, +10\}$.

Find:

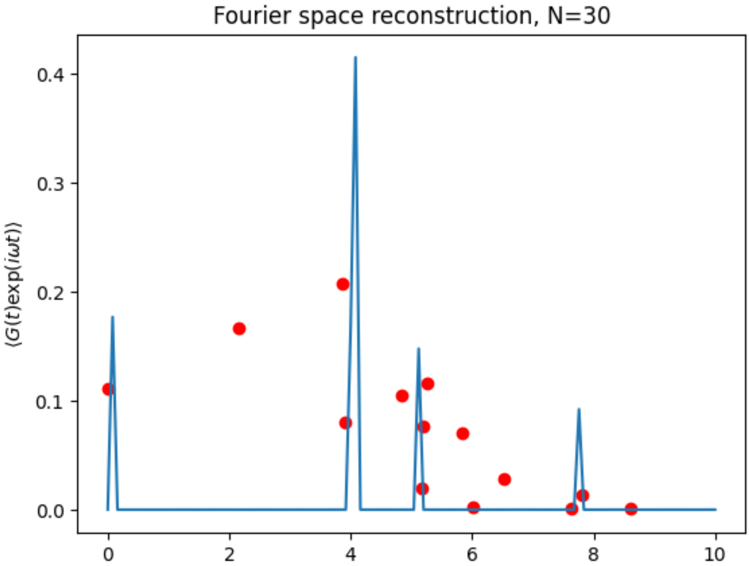
$$G(t) = \sum_k a_k \exp(i\omega_k t)$$

$$a_k \geq 0, \quad \sum a_k = 1$$

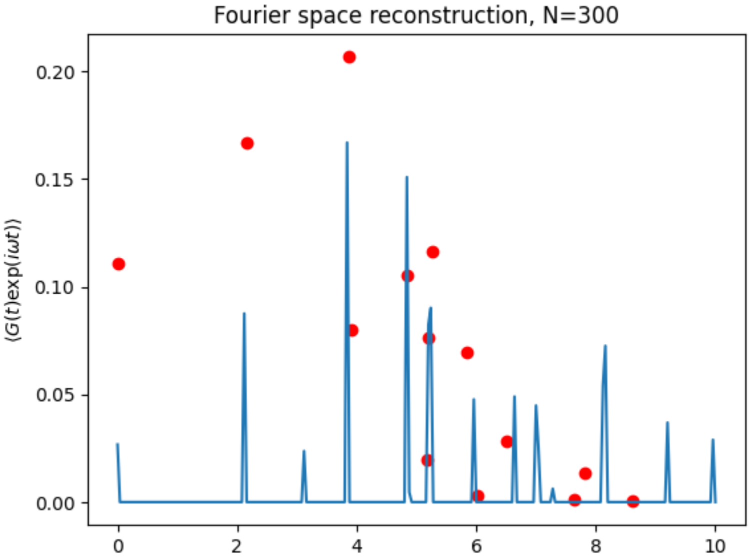
that best fits the observed data. Fix some dense set of ω_k, say, $\{-10, -9.9, -9.8 \cdots + 9.9, +10\}$.

Model is linear in a_k, linear constraint on a_k, convex likelihood function ☺

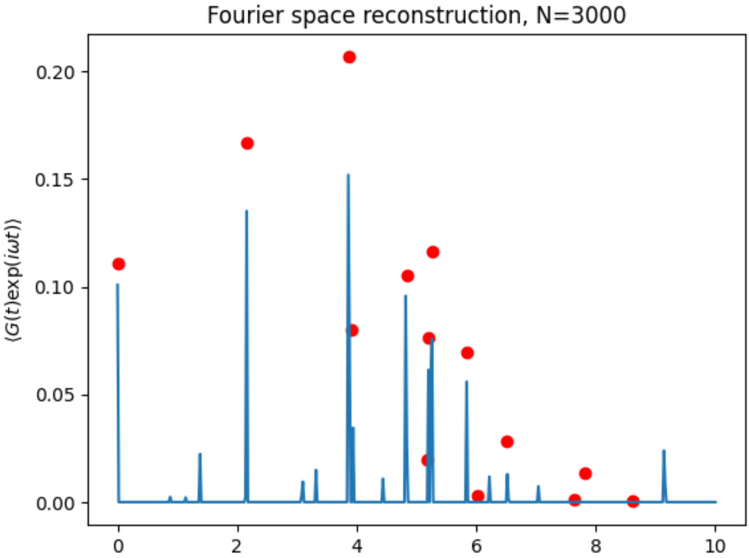
Maximum likelihood estimator, 30 samples:

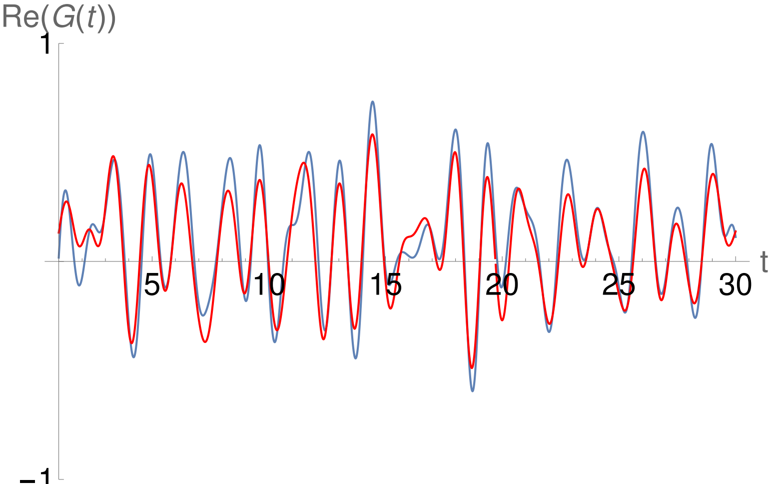


Maximum likelihood estimator, 300 samples:

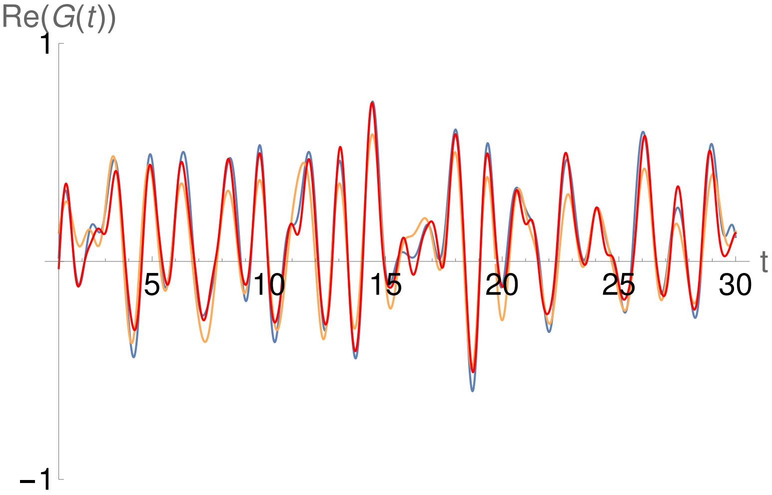


Maximum likelihood estimator, 3000 samples:




Re($G(t)$)
1
−1
5
10
15
20
25
30
t

Bayesian Statistics: GP vs Fourier



Details

- Maximum Likelihood Estimation (MLE): L^2 vs Bayes
- Going beyond MLE
- Adaptive sampling

Details

- Maximum Likelihood Estimation (MLE): L^2 vs Bayes
  - Bayes: find the true maximum likelihood estimator. Converges fast
- Going beyond MLE
- Adaptive sampling

Details

- Maximum Likelihood Estimation (MLE): L^2 vs Bayes
  - Bayes: find the true maximum likelihood estimator. Converges fast
  - L^2 approximation: linearize the problem, fast to find the optimum
  - Didn't observe any significant improvement between them
- Going beyond MLE
- Adaptive sampling

Details

- Maximum Likelihood Estimation (MLE): L^2 vs Bayes
- Going beyond MLE
  - Finding the mean estimator: averaging a_k's weighted by likelihood
- Adaptive sampling

Details

- Maximum Likelihood Estimation (MLE): L^2 vs Bayes
- Going beyond MLE
  - Finding the mean estimator: averaging a_k's weighted by likelihood
  - Polynomial time in theory - integrating over convex likelihood function
  - In practice, slow-ish but workable, but not much benefit
- Adaptive sampling

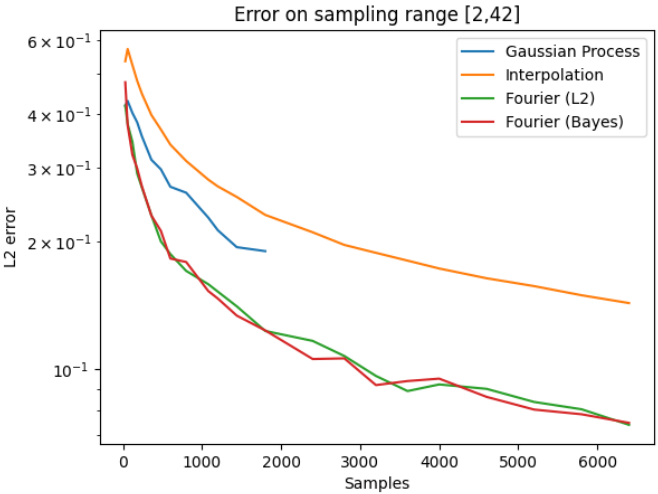
Details

- Maximum Likelihood Estimation (MLE): L^2 vs Bayes
- Going beyond MLE
- Adaptive sampling
  - Choose points to sample based on what gives the most "information"
  - e.g. if $G(t) \approx 0.99$, further samples of $\Re[G(t)]$ are not useful
  - ... but $\Im[G(t)]$ is less certain, that could be useful to sample
  - Nonlinear interaction of Fourier terms means that uncertainty varies considerably

Details

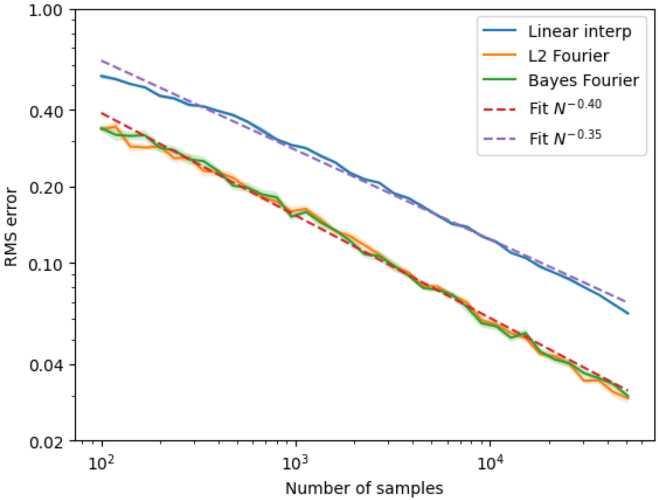
- Maximum Likelihood Estimation (MLE): L^2 vs Bayes
- Going beyond MLE
- Adaptive sampling
  - Choose points to sample based on what gives the most "information"
  - e.g. if $G(t) \approx 0.99$, further samples of $\Re[G(t)]$ are not useful
  - ... but $\Im[G(t)]$ is less certain, that could be useful to sample
  - Nonlinear interaction of Fourier terms means that uncertainty varies considerably
  - Hard to quantify "information" well
  - There is a good answer, but it requires integrating over likelihood function again
  - Impractical for now, requires fast integrals of Gaussians over simplex

Performance Comparison



At an accuracy of $\approx 1\%$ in $G(t)$, roughly 100x sample efficiency

Performance Comparison



At an accuracy of $\approx 1\%$ in $G(t)$, roughly 100x sample efficiency

Performance Comparison

- Interpolation: $N \approx \epsilon^{-2.9}$

Performance Comparison

- Interpolation: $N \approx \epsilon^{-2.9}$
- Gaussian Process: something better, but slow

Performance Comparison

- Interpolation: $N \approx \epsilon^{-2.9}$
- Gaussian Process: something better, but slow
- Fourier space: $N \approx \epsilon^{-2.5}$

Performance Comparison

- Interpolation: $N \approx \epsilon^{-2.9}$
- Gaussian Process: something better, but slow
- Fourier space: $N \approx \epsilon^{-2.5}$

Thank you!

Some facts about permanents

- Defn:

$$\text{Perm}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} A_{i,\sigma(i)} \tag{1}$$

...similar to determinant of A, but without the $(-1)^\sigma$.

Some facts about permanents

- Defn:

$$\mathrm{Perm}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} A_{i,\sigma(i)} \tag{1}$$

...similar to determinant of A, but without the $(-1)^\sigma$.

- Can be computed in $O(2^n)$ [Ryser, 63]

Some facts about permanents

- Defn:
$$\text{Perm}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} A_{i,\sigma(i)} \quad (1)$$
...similar to determinant of A, but without the $(-1)^\sigma$.
- Can be computed in $O(2^n)$ [Ryser, 63]
- Hard to compute exactly [Valiant, 79]

Some facts about permanents

- Defn:

$$\text{Perm}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} A_{i,\sigma(i)} \quad (1)$$

...similar to determinant of A, but without the $(-1)^\sigma$.

- Can be computed in $O(2^n)$ [Ryser, 63]
- Hard to compute exactly [Valiant, 79]
- Easy to estimate if all entries are nonnegative [JSV, 01]

Some facts about permanents

- Defn:

$$\mathrm{Perm}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} A_{i,\sigma(i)} \tag{1}$$

  ...similar to determinant of A, but without the $(-1)^\sigma$.
- Can be computed in $O(2^n)$ [Ryser, 63]
- Hard to compute exactly [Valiant, 79]
- Easy to estimate if all entries are nonnegative [JSV, 01]

If A is positive-semidefinite (PSD):

- Can be written as an an integral of a nonnegative function, so $\mathrm{Perm}(A) \geq 0$

Some facts about permanents

- Defn:

$$\mathrm{Perm}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} A_{i,\sigma(i)} \quad (1)$$

...similar to determinant of A, but without the $(-1)^\sigma$.

- Can be computed in $O(2^n)$ [Ryser, 63]
- Hard to compute exactly [Valiant, 79]
- Easy to estimate if all entries are nonnegative [JSV, 01]

If A is positive-semidefinite (PSD):

- Can be written as an an integral of a nonnegative function, so $\mathrm{Perm}(A) \geq 0$
- Still hard to compute exactly

Some facts about permanents

- Defn:

$$\text{Perm}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} A_{i,\sigma(i)} \tag{1}$$

  ...similar to determinant of A, but without the $(-1)^\sigma$.
- Can be computed in $O(2^n)$ [Ryser, 63]
- Hard to compute exactly [Valiant, 79]
- Easy to estimate if all entries are nonnegative [JSV, 01]

If A is positive-semidefinite (PSD):

- Can be written as an an integral of a nonnegative function, so $\text{Perm}(A) \geq 0$
- Still hard to compute exactly
- Can efficiently compute a 4.85^n approximation [Anari+, 17]

Some facts about permanents

- Defn:

$$\mathrm{Perm}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} A_{i,\sigma(i)} \tag{1}$$

...similar to determinant of A, but without the $(-1)^\sigma$.
- Can be computed in $O(2^n)$ [Ryser, 63]
- Hard to compute exactly [Valiant, 79]
- Easy to estimate if all entries are nonnegative [JSV, 01]

If A is positive-semidefinite (PSD):

- Can be written as an an integral of a nonnegative function, so $\mathrm{Perm}(A) \geq 0$
- Still hard to compute exactly
- Can efficiently compute a 4.85^n approximation [Anari+, 17]
- Easy to estimate if $\lambda_{\max}/\lambda_{\min} \leq 2$ [Barvinok, 20]

Some facts about permanents

If A is positive-semidefinite (PSD):

- Can be written as an an integral of a nonnegative function, so $\mathrm{Perm}(A) \geq 0$
- Still hard to compute exactly

Some facts about permanents

If A is positive-semidefinite (PSD):

- Can be written as an an integral of a nonnegative function, so $\mathrm{Perm}(A) \geq 0$
- Still hard to compute exactly
- **Represent output probabilities of Boson sampling experiments when inputs are thermal (as opposed to coherent beams)**

Some facts about permanents

If A is positive-semidefinite (PSD):

- Can be written as an an integral of a nonnegative function, so $\text{Perm}(A) \geq 0$
- Still hard to compute exactly
- **Represent output probabilities of Boson sampling experiments when inputs are thermal (as opposed to coherent beams)**
- This quantum connection inspired other algorithms, that also work better when spectral radius is small [CCG, 17]

Some facts about permanents

If A is positive-semidefinite (PSD):

- Can be written as an an integral of a nonnegative function, so $\mathrm{Perm}(A) \geq 0$
- Still hard to compute exactly
- **Represent output probabilities of Boson sampling experiments when inputs are thermal (as opposed to coherent beams)**
- This quantum connection inspired other algorithms, that also work better when spectral radius is small [CCG, 17]

Question remains: are these PSD permanents hard to approximate?

Connection to permanents

Measurements γ_i form an $n \times d$ matrix Γ. Partition function Z is a function only of Γ.

$$Z = \int_{\mathbb{C}_1^d} \prod_i P(\gamma_i|\psi)\, d\psi = \int_{\mathbb{C}_1^d} \prod_i (\psi^\dagger \gamma_i)(\gamma_i^\dagger \psi)\, d\psi$$

Connection to permanents

Measurements γ_i form an $n \times d$ matrix Γ. Partition function Z is a function only of Γ.

$$Z = \int_{\mathbb{C}_1^d} \prod_i P(\gamma_i|\psi)\, d\psi = \int_{\mathbb{C}_1^d} \prod_i (\psi^\dagger \gamma_i)(\gamma_i^\dagger \psi)\, d\psi$$

Invariant under permutations of n rows. Order of observations doesn't matter, each was from a fresh $|\psi\rangle$.

Invariant under a unitary transformation acting on the d-dimensional space. Just a change of basis.

Linear in each γ_i and its adjoint $\gamma_i^\dagger$. Enough to establish:

$$Z = C\, \mathrm{Perm}(\Gamma^\dagger \Gamma)$$

$$Z = C\,\mathrm{Perm}(\Gamma^\dagger \Gamma)$$

This matrix $\Gamma^\dagger\Gamma$ is $n \times n$ PSD. Constant C is easily computed as

$$C = \frac{2\pi^n}{(d+n-1)!}$$

Hardness of quantum state estimation $\to$ hardness of PSD permanents.

Z as an integral over unit sphere is very similar to other formulations (Barvinok) of PSD permanents as a spherical integral

Consequences, Future Work

- No APX for PSD permanents (unless $\mathrm{P} = \mathrm{NP}$)
- Haven't ruled out $(1 + \epsilon)^n$ approximation algorithms
- These PSD matrices are always rank $d \ll n$. Likely to be more improvements in terms of spectral radius, $\lambda_{\min} > 0$
- Only showed NP-hardness (0 solutions or ≥ 1?). Can likely improve to approximately counting solutions
- Doesn't mean quantum state tomography is *typically* hard: these types of measurements are unlikely
- Would be nice to show that some efficient algorithms for state reconstructions converge with high probability as more measurements are taken (from any basis)

Consequences, Future Work

- $O(n^d)$ algorithm means that this is in the XP complexity class, *slicewise polynomial*
- Could hope that the d part becomes some constant factor of difficulty, e.g. $O(2^d n^2)$
- Would be called *fixed-parameter tractable*, or FPT

Consequences, Future Work

- $O(n^d)$ algorithm means that this is in the XP complexity class, *slicewise polynomial*
- Could hope that the d part becomes some constant factor of difficulty, e.g. $O(2^d n^2)$
- Would be called *fixed-parameter tractable*, or FPT
- More involved construction lets relate this to MAX-CLIQUE in graph, which is W[1]-hard
- Low-rank PSD permanent is W[1]-hard as well
- Parameterized complexity theory: if P $\neq$ NP, then W[1] $\neq$ FPT
- Proves that we can't do better than $O(n^{f(d)})$

Consequences, Future Work

- Thermal Boson sampling is *sort of roughly* as hard as coherent Boson sampling, with maybe an $n \to n^3$ type of slowdown

- Thermal Boson sampling is *sort of roughly* as hard as coherent Boson sampling, with maybe an $n \to n^3$ type of slowdown... in the sense of how many modes you need to encode a SAT-type problem
- Can't make this statement rigorous, because no one has actually shown either one to be hard (in a sampling sense).

Relation to Boson sampling

Linear optical circuit mixes modes with some unitary, e.g.

$$U = \begin{bmatrix} u_{1,1} & u_{1,2} & u_{1,3} \\ u_{2,1} & u_{2,2} & u_{3,3} \\ u_{3,1} & u_{3,2} & u_{3,3} \end{bmatrix}$$

Relation to Boson sampling

Linear optical circuit mixes modes with some unitary, e.g.

$$U = \begin{bmatrix} u_{1,1} & u_{1,2} & u_{1,3} \\ u_{2,1} & u_{2,2} & u_{3,3} \\ u_{3,1} & u_{3,2} & u_{3,3} \end{bmatrix}$$

Rows = Input modes, columns = Output modes

Relation to Boson sampling

Linear optical circuit mixes modes with some unitary, e.g.

$$U = \begin{bmatrix} u_{1,1} & u_{1,2} & u_{1,3} \\ u_{2,1} & u_{2,2} & u_{3,3} \\ u_{3,1} & u_{3,2} & u_{3,3} \end{bmatrix}$$

Rows = Input modes, columns = Output modes
If I put two Bosons at mode 2-in and one at mode 3-in, what's the probability of observing one excitation at mode 1-out and two at mode 3-out?

Linear optical circuit mixes modes with some unitary, e.g.

$$U = \begin{bmatrix} u_{1,1} & u_{1,2} & u_{1,3} \\ u_{2,1} & u_{2,2} & u_{3,3} \\ u_{3,1} & u_{3,2} & u_{3,3} \end{bmatrix}$$

Rows = Input modes, columns = Output modes
If I put two Bosons at mode 2-in and one at mode 3-in, what's the probability of observing one excitation at mode 1-out and two at mode 3-out?

$$U = \begin{bmatrix} u_{1,1} & u_{1,2} & u_{1,3} \\ u_{2,1} & u_{2,2} & u_{3,3} \\ u_{3,1} & u_{3,2} & u_{3,3} \end{bmatrix}$$

Relation to Boson sampling

Linear optical circuit mixes modes with some unitary, e.g.

$$U = \begin{bmatrix} u_{1,1} & u_{1,2} & u_{1,3} \\ u_{2,1} & u_{2,2} & u_{2,3} \\ u_{3,1} & u_{3,2} & u_{3,3} \end{bmatrix}$$

Rows = Input modes, columns = Output modes
If I put two Bosons at mode 2-in and one at mode 3-in, what's the probability of observing one excitation at mode 1-out and two at mode 3-out?

$$U = \begin{bmatrix} \color{red}{u_{2,1}} & \color{red}{u_{2,2}} & \color{red}{u_{2,3}} \\ u_{2,1} & u_{2,2} & u_{2,3} \\ u_{3,1} & u_{3,2} & u_{3,3} \end{bmatrix}$$

Relation to Boson sampling

Linear optical circuit mixes modes with some unitary, e.g.

$$U = \begin{bmatrix} u_{1,1} & u_{1,2} & u_{1,3} \\ u_{2,1} & u_{2,2} & u_{2,3} \\ u_{3,1} & u_{3,2} & u_{3,3} \end{bmatrix}$$

Rows = Input modes, columns = Output modes
If I put two Bosons at mode 2-in and one at mode 3-in, what's the probability of observing one excitation at mode 1-out and two at mode 3-out?

$$U = \begin{bmatrix} u_{2,1} & \color{red}{u_{2,1}} & u_{2,3} \\ u_{2,1} & \color{red}{u_{2,1}} & u_{2,3} \\ u_{3,1} & \color{red}{u_{3,1}} & u_{3,3} \end{bmatrix}$$

If I put two Bosons at mode 2-in and one at mode 3-in, what's the probability of observing one excitation at mode 1-out and two at mode 3-out?

$$U = \begin{bmatrix} u_{2,1} & u_{2,1} & u_{2,3} \\ u_{2,1} & u_{2,1} & u_{2,3} \\ u_{3,1} & u_{3,1} & u_{3,3} \end{bmatrix}$$

$$\mathrm{Prob}((2,2,3) \to (1,1,3)) \propto \mathsf{Perm}(U)^2$$

If I put two Bosons at mode 2-in and one at mode 3-in, what's the probability of observing one excitation at mode 1-out and two at mode 3-out?

$$U = \begin{bmatrix} u_{2,1} & u_{2,1} & u_{2,3} \\ u_{2,1} & u_{2,1} & u_{2,3} \\ u_{3,1} & u_{3,1} & u_{3,3} \end{bmatrix}$$

$$\mathrm{Prob}((2,2,3) \to (1,1,3)) \propto \mathsf{Perm}(U)^2$$

If permanents are hard to approximate, and these experiments read out the permanents of arbitrary matrices, then the experiments are doing something computationally powerful... maybe?

If permanents are hard to approximate, and these experiments read out the permanents of arbitrary matrices, then the experiments are doing something computationally powerful... maybe?

If permanents are hard to approximate, and these experiments read out the permanents of arbitrary matrices, then the experiments are doing something computationally powerful... maybe?
Not quite – you don't control which results occur! You just see relative frequencies. Would need to design a matrix that has one output that is very likely or very unlikely, and its occurrence or not tells you something useful.

If permanents are hard to approximate, and these experiments read out the permanents of arbitrary matrices, then the experiments are doing something computationally powerful... maybe?
Not quite – you don't control which results occur! You just see relative frequencies. Would need to design a matrix that has one output that is very likely or very unlikely, and its occurrence or not tells you something useful.
Permanent approximation is #P-Hard, which is expected to be much more than what quantum computers can achieve.