Quantum Speedups of Continuous Sampling and Optimization Problems

Ruizhe Zhang
Simons Institute

CQC 2023 Workshop I
Complexity of Sampling

Given **black-box** access to a function $f: \mathbb{R}^d \rightarrow \mathbb{R}$, what is the minimum number of queries required to approximately sample from the distribution with density $\pi(x) \propto e^{-f(x)}$ in $\mathbb{R}^d$?

A fundamental problem with wide applications:

- **Statistical physics**
  - $f(x)$ represents the energy of a state $x$ and the equilibrium distribution over states is the Gibbs distribution whose density $\propto e^{-f(x)/T}$ ($T$ is the temperature of the system).

- **Bayesian inference**
  \[ p(\theta|x) = \frac{p(x|\theta)p(\theta)}{\int_{\mathbb{R}^d} p(x|\theta')p(\theta')d\theta'} \]

- **Convex body volume estimation**
  - Given access to the membership oracle of a convex body $\mathcal{K} \subseteq \mathbb{R}^d$, estimate $\text{vol}(\mathcal{K})$.
  - Reduce to **uniformly** sample a point inside some convex body.
**Sampling** \(( \pi \propto e^{-f} )\)

**Log-concave**

- **Classical:** easy (Langevin diffusion)
- **Quantum:**
  - Volume estimation: (Chakrabarti et al.'23)
  - General case: (Childs et al.'22)

**Non-log-concave**

- **Classical:** hard (Langevin diffusion takes exponential time), efficient algorithms for some family of distributions
- **Quantum:** open

**Optimization** \(( \min_x f(x) )\)

**Convex**

- **Classical:** easy (gradient descent)
- **Quantum:**
  - General case: (Chakrabarti et al.'19, van Apeldoorn et al.'20)
  - Quantum LP/SDP: (Brandão-Svore'17, van Apeldoorn et al.'20, ...)

**Non-convex**

- **Classical:** NP-hard in general, algorithms works well in practice.
- **Quantum:** Some recent works (Liu et al.'22, Gong et al.'22, ...) show quantum advantages over specific classical algorithms (e.g., SGD).

---


Li, Z. Quantum Speedups of Optimizing Approximately Convex Functions with Applications to Logarithmic Regret Stochastic Convex Bandits. (*NeurIPS* 22)
- The glued tree problem (Rolando’s talk on Mon)
- State preparation (Mario’s talk on Mon)
- Markov chains (Open-system day)
- Quantum walk circuit implementation (Jingbo’s talk on Thu)

### Log-concave sampling

Polling quantum speedup

### Quantum walk

- The glued tree problem (Rolando’s talk on Mon)
- State preparation (Mario’s talk on Mon)
- Markov chains (Open-system day)
- Quantum walk circuit implementation (Jingbo’s talk on Thu)

### Mean estimation w/ source code (Rabin’s talk on Mon)

### Normalizing constant estimation

### Quantum mean estimators

### Stochastic convex optimization

### Stochastic bandit problem

### Exponential quantum advantage

### Polynomial quantum speedup
Log-Concave Distribution

**Definition (Log-concave distribution)**
A probability distribution \( \pi(dx) \propto e^{-f(x)} \) is log-concave if \( f: \mathbb{R}^d \rightarrow \mathbb{R} \) is a convex function. We further assume that \( f \) is \( \mu \)-strongly convex and \( L \)-smooth:

\[
\frac{\mu}{2} \|x - y\|^2 \leq f(y) - f(x) - \nabla f(x)^\top (y - x) \leq \frac{L}{2} \|x - y\|^2 \quad \forall x, y \in \mathbb{R}^d.
\]

Let \( \kappa := \frac{L}{\mu} \) be the condition number.

**Examples**

1. High-dimensional Gaussian distribution \( \mathcal{N}(\theta, \Sigma) \) for positive definite \( \Sigma \).
2. Uniform distribution \( \pi(x) \propto 1_{\mathcal{K}}(x) \) for a convex and compact \( \mathcal{K} \subset \mathbb{R}^d \).

**Langevin diffusion:**

\[
\text{d}X_t = -\nabla f(X_t) \text{d}t + \sqrt{2} \text{d}B_t
\]

- gradient flow
- Brownian motion

\[\text{Stationary distribution is } \pi\]
Metropolis Adjusted Langevin Algorithm (MALA)

To sample from the log-concave distribution, we need to simulate the Langevin diffusion.

\[
\frac{dX_t}{dt} = -\nabla f(X_t) + \sqrt{2} dB_t \quad \text{discretize} \quad X_{i+1} = X_i - h\nabla f(X_i) + \sqrt{2h} z_i \\
X_0 \sim N(0, I)
\]

However, the stationary distribution of the discretized process is not \( \pi \).

**MALA** combines the Langevin dynamics with the **Metropolis–Hastings accept/reject mechanism**:

1. Initialize \( x_0 \sim \mu_0 \)
2. For \( i = 0, 1, 2, \ldots \):
   1. Propose \( z_{i+1} \sim N(x_i - h\nabla f(x_i), 2hI) \)
   2. Accept \( x_{i+1} \leftarrow z_{i+1} \) with probability

\[
\min \left\{ 1, \frac{\exp(-f(z_{i+1}) - \|x_i - z_{i+1} + h\nabla f(z_{i+1})\|^2/(4h))}{\exp(-f(x_i) - \|z_{i+1} - x_i + h\nabla f(x_i)\|^2/(4h))} \right\}
\]

- Stationary distribution \( \pi \)
- polylog\((1/\epsilon)\)-dependence
- Gradient oracle query

Quantum speedup for MALA?
Quantum sampling

Ideally, we want to generate a quantum state ($\text{qsamp}le$) to represent a classical distribution:

$$\{\pi(x)\}_{x \in \Omega} \leftrightarrow |\pi\rangle = \int_{\Omega} \sqrt{\pi(x)} |x\rangle \, dx.$$ 

<table>
<thead>
<tr>
<th>Reference</th>
<th>Complexity</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Zalka'98, Grover-Rudolph'02, Kaye-Mosca'01)</td>
<td>$O(\log 1/\epsilon)$</td>
<td>controlled rotations only for efficiently integrable density functions</td>
</tr>
<tr>
<td>(Aharanov-Ta-Shma'03)</td>
<td>$O(1/\delta)$</td>
<td>Qsampling is hard in general unless $\text{SZK} \subseteq \text{BQP}$</td>
</tr>
<tr>
<td>(Wocjan-Abeyesinghe'08)</td>
<td>$O\left(1/\sqrt{\delta}\right)$</td>
<td>adiabatic evolution for Markov chains</td>
</tr>
<tr>
<td>(Low-Yoder-Chuang'14, Ozols-Roetteler-Roland'13, Wiebe-Granade'15)</td>
<td>$O^*(1/\sqrt{\epsilon})$</td>
<td>Szegedy's quantum walks + amplitude amplification</td>
</tr>
</tbody>
</table>

$\delta$ is the spectral gap of Markov chain and $\epsilon$ is the approximation error.
Discrete-Time Quantum Walk (DTQW)

• A classical Markov chain over $\Omega$ can be represented by stochastic transition operator $P$ such that

$$\sum_{y \in \Omega} P(x, y) = 1 \quad \forall x \in \Omega.$$ 

• A probability distribution $\pi$ is stationary if

$$\sum_{y \in \Omega} \pi(x)P(x, y) = \pi(y) \quad \forall y \in \Omega.$$ 

Transition operator $P$

Acting on two registers $(x, y) \in \Omega \times \Omega$

• Step 1:

$$(x, y) \rightarrow (x, y_*) \quad y_* \in N(x)$$

• Step 2:

$$(x, y_*) \rightarrow (y_*, x)$$

Quantum walk operator $W$

Acting on two quantum registers $|x\rangle|y\rangle$

• Step 1:

reflect $|y\rangle$ through $\sum_{z \in \Omega} \sqrt{P(x, z)} |z\rangle$

• Step 2:

$|x\rangle|y\rangle \rightarrow |y\rangle|x\rangle$
Szegedy’s Quantum Walk Operator

- Define $|\psi_x\rangle := |x\rangle \sum_{y \in \Omega} \sqrt{P(x, y)} |y\rangle$ for any $x \in \Omega$.
- $\Pi = \sum_{x \in \Omega} |\psi_x\rangle \langle \psi_x|$ is the projection to the subspace $\text{span}\{|\psi_x\rangle\}_{x \in \Omega}$.
- $S = \sum_{x \in \Omega} \sum_{y \in \Omega} |y, x\rangle \langle x, y|$ is the swap operator for the two quantum registers.
- The quantum walk operator can be defined as
  \[ W := S(2\Pi - I) = S \cdot U \cdot (2(I \otimes |0\rangle\langle 0|) - I) \cdot U^\dagger, \]
  where $U$ implements the QW update:
  \[ U|x, 0\rangle = |\psi_x\rangle \quad \forall x \in \Omega. \]
- **Connection to QSVT:** Let $W' := U^\dagger \cdot W \cdot U$. Then, $W'^k$ is a block-encoding of $T_k(P)$, the $k$-th Chebyshev polynomial, i.e.,
  \[ (I \otimes \langle 0|)W'^k (I \otimes |0\rangle) = T_k(P). \]

We assume that $P$ is symmetric, i.e.,
\[ P(x, y) = P(y, x) \quad \forall x, y \in \Omega. \]
In general, we should consider
\[ D(x, y) = \sqrt{P(x, y)P(y, x)}. \]
Spectrum of Quantum Walk Operator

- $W$ has phase gap $\Delta = \Theta(\sqrt{\delta})$, where $\delta$ is the spectral gap of $P$.

Eigenvalues of $P$: \{\cos \theta_i\}

Eigenvalues of $W$: \{e^{\pm i\theta_i}\}
DTQW for Searching

DTQW can quadratically speed up the hitting time of a reversible MC.

- Hitting time: the expected time to hit a marked vertex starting from the stationary distribution.
- Reversible: $P$ satisfies the detailed balance condition: $\pi(x) \cdot P(x, y) = \pi(y) \cdot P(y, x) \ \forall x, y \in \Omega$, which is required by the spectral analysis of $P$.

Examples of gapped systems:

- Johnson graph $J(n, m)$: $\delta = \frac{n}{m(n-m)}$.
- Ising model with Glauber dynamics:
  \[
  \pi(x) \propto \exp(x^T J x + h^T x) \ \forall x \in \{\pm 1\}^n.
  \]

There are numerous classical papers studying the spectral gaps in different parameter regimes, e.g., (Dobrushin’68, Jerrum-Sinclair’93, Mossel-Sly’13, Chen et al.’21, Eldan et al.’21, Jain et al.’22,…).
**DTQW for Sampling**

**Question:** how to generate a sample from the stationary distribution $\pi$?

- Classically, the #steps needed in the worst-case is the *mixing time* of the Markov chain.

- For a reversible MC, the mixing time is bounded by $\frac{1}{\delta} \cdot \log \left( \frac{1}{\min_{x \in \Omega} \pi(x)} \right)$.

- DTQW can be used to prepare the quantum sample ($\text{qsample}$) of the stationary distribution:

$$|\pi\rangle = \int_{\Omega} \sqrt{\pi(x)} |x\rangle dx.$$ 

In the most general case, the cost is $1/\sqrt{\delta} \cdot 1/\sqrt{\min \pi(x)}$.

Can we do better under some assumptions?
Speedup for Slowly-Varying Markov Chains

Theorem (Wocjan-Abeyesinghe’08)

Let $M_0, M_1, ..., M_r$ be classical reversible Markov chains with stationary distribution $\pi_0, \pi_1, ..., \pi_r$ such that

1. Each chain has spectral gap $\geq \delta$.
2. $|\langle \pi_i | \pi_{i+1} \rangle|^2 \geq p$ for all $i \in \{0, 1, ..., r - 1\}$ (Quantum Simulated Annealing (QSA) condition).
3. $|\pi_0\rangle$ is easy to prepare.

Then $|\pi_r\rangle$ can be approximately prepared using $\tilde{O}\left(\frac{1}{\sqrt{\delta} \cdot \frac{r}{p}}\right)$ calls to the quantum walk operators.

Remark

To implement the quantum walk operator $W$, it suffices to implement $U$:

$$U x_0 = x_0 \int P(x, y) dy \forall x \in \Omega.$$ 

For MALA, you can do it with $O(1)$ queries to $O_\#$ and $O_\nabla \#$.

$$O_2 x, y = x, y + f(x) \quad O_\nabla 2 x, y = x, y + \nabla f(x)$$
In continuous space, the spectral gap of “useful” Markov chains (e.g., MALA) are difficult to bound, since it characterizes the mixing behavior in the worst-case (i.e., for any initial distribution).

Classically, there are several techniques to overcome the spectral gap barrier:

- Discounting the ill-effect of small (and problematic) sets in measuring mixing time (in the average-case).
  
  \[ s\text{-conductance (Lovász-Simonovits'93), average conductance (Lovász-Kannan'99), blocking conductance (Kannan-Lovász-Montenegro'06), approximate spectral gap (Atchadé'19)} \]

- Only focusing on “good” distributions with some warmness \( \beta \geq \sup \left\{ \frac{\pi_0(A)}{\pi(A)} : A \subseteq \Omega \right\} \).
  
  \[ \text{For MALA with a “warm-start”, see e.g. (Lee-Shen-Tian'20, Wu-Schmidler-Chen'22)} \]

Can we adapt these techniques to the quantum walk?
**Effective Spectral Gap for Warm-Start**

**Lemma** (Childs-Li-Liu-Wang-Zhang’22, Chakrabarti et al.’23)

Let $M$ be a Markov chain with stationary distribution $\pi$. Let $\pi_0$ be an initial distribution mixing in $t$ steps. Furthermore, assuming $\pi_0$ is a warm-start with respect to $\pi$. Then, those “bad” eigenvalues in $[1 - t^{-1}, 1)$ will not be effective during the quantum walk on $|\pi_0\rangle$.

---


Quantum MALA with Warm-Start

**Theorem** *(Childs-Li-Liu-Wang-Zhang’22)*

Let $\pi_0$ be a warm start for the log-concave distribution $\pi \propto e^{-f}$. Given access to a unitary $U_I$ that prepares the initial state $|\pi_0\rangle$, there is a quantum algorithm that outputs a state $|\tilde{\pi}\rangle$ that is $\epsilon$-close to $|\pi\rangle$ with query complexity to the evaluation oracle $O_f$ and gradient oracle $O_{\nabla f}$:

$$\tilde{O}(\sqrt{\kappa d^{1/4}}).$$

- Classically, $t_{\text{mix}} = \tilde{O}(\kappa \sqrt{d})$ for MALA with a warm-start *(Wu-Schmidler-Chen’22)*.

- A special instance of state preparation with large initial overlap. The (query) cost of our algorithm is sublinear in $\log$(system size).

---

Quantum MALA without Warm-Start

A warm-start MALA is not always accessible. What about starting from a Gaussian distribution?

- \( \beta = \kappa^{d/2} \) and \( t = \tilde{O}(\kappa d) \) (Lee et al.'21, Chen et al.'21).
- We cannot directly apply our theorem since the overlap \(|\langle \pi_0 | \pi \rangle| \sim \kappa^{-d/4} \) is too small!

**Idea:** using a simulated annealing process to construct a slowly-varying MCs.

- \( |\pi_0 \rangle \) is easy to prepare. Then, we use quantum walk to evolve \( |\pi_i \rangle \rightarrow |\pi_{i+1} \rangle \) for \( i = 0,1, \ldots, M \).
- The overlaps \(|\langle \pi_i | \pi_{i+1} \rangle| \) should be large for all stages.
Quantum MALA without Warm-Start

**Theorem (Childs-Li-Liu-Wang-Zhang’22)**

If we take $\sigma_{i+1}^2 = \sigma_i^2 \cdot \left(1 + \frac{1}{\sqrt{d}}\right)$ and $M = \tilde{O}(\sqrt{d})$, Quantum MALA (without warm-start) can approximately prepare the state $|\pi\rangle$ for $\pi \propto e^{-f}$ with query complexity:

$$\tilde{O}(\sqrt{d} \times \sqrt{\kappa d}) = \tilde{O}(\sqrt{\kappa d}).$$

- Classical query complexity of MALA is $\tilde{O}(\kappa d)$.  
Problem (Normalizing constant estimation)

Let $\pi \propto e^{-f}$ be a $d$-dimensional log-concave distribution. Define the normalizing constant:

$$Z := \int_{\mathbb{R}^d} e^{-f(x)} \, dx.$$ 

Given black-box access to $f$, output $\tilde{Z} \in \mathbb{R}$ such that $\tilde{Z} \in (1 \pm \epsilon)Z$.

This problem is also called the partition function estimation in statistical physics and has been studied in both classical (Dyer et al.’91, Gelman-Meng’98, Brosse et al.’18, Ge-Lee-Lu’21, …) and quantum (Montanaro’15, Harrow-Wei’20, Arunachalam et al.’21, Cornelissen-Hamoudi’23).

- Prior quantum algorithms mainly focused on discrete systems.
- We focus on the continuous version of this problem.
Simulated Annealing + Log-Concave Sampling

**Annealing schedule:**

\[
\begin{align*}
\pi_0 & \quad \quad \quad e^{-\frac{\|x\|^2}{2\sigma_1^2}} \\
\pi_1 & \quad \quad \quad e^{-\frac{\|x\|^2}{2\sigma_1^2}} \\
\pi_2 & \quad \quad \quad e^{-\frac{\|x\|^2}{2\sigma_2^2}} \\
\cdots & \quad \quad \quad \cdots \\
\pi_M & \quad \quad \quad e^{-\frac{\|x\|^2}{2\sigma_M^2}} \\
\pi_{M+1} & = \pi \\
\end{align*}
\]

**Normalizing constants:**

\[
\begin{align*}
Z_1 & \\
Z_2 & \\
Z_M & \\
Z_{M+1} & = Z
\end{align*}
\]

We can rewrite the normalizing constant as: \( Z = Z_1 \cdot \prod_{i=1}^{M} \frac{Z_{i+1}}{Z_i} \).

- Sample \( X_i^{(1)}, \ldots, X_i^{(K)} \) from distribution \( \pi_i = Z_i^{-1} \cdot \exp \left( -f - \frac{\|x\|^2}{2\sigma_i^2} \right) \).
- \( \frac{Z_{i+1}}{Z_i} = \mathbb{E}_{\pi_i} [g_i] \), where \( g_i = \exp \left( \frac{1}{2} \left( \sigma_i^{-2} - \sigma_{i+1}^{-2} \right) \|x\|^2 \right) \).
  - Estimator: \( Z_{i+1}/Z_i \approx \frac{1}{K} \sum_{j=1}^{K} g_i \left( X_i^{(j)} \right) \).
  - This annealing schedule has bounded relative variance, i.e., \( \frac{\mathbb{E}_{\pi_i} [g_i^2]}{\mathbb{E}_{\pi_i} [g_i]^2} = O(1) \) (Ge-Lee-Lu’21).

---

Quantum MALA for Estimating Normalizing Constant

**Theorem** (Childs-Li-Liu-Wang-Zhang’22)

Let $Z$ be the normalizing constant. There is a quantum algorithm which outputs an estimate $\tilde{Z}$, such that $\tilde{Z} \in (1 \pm \epsilon)Z$ with high probability using $\tilde{O}(d^{3/2}k^{1/2}\epsilon^{-1})$ queries to the evaluation oracle $O_f$ and gradient oracle $O_{\nabla f}$.

*Proof idea:*

It suffices to estimate each ratio $\frac{Z_{i+1}}{Z_i} = \mathbb{E}_{\pi_i}[g_i]$ within error $\frac{\epsilon}{M}$ with $M = \tilde{O}(\sqrt{d})$.

i. By the non-destructive mean estimation (Harrow-Wei’21, Chakrabarti et al.’21), we need $\tilde{O}\left(\frac{M}{\epsilon}\right)$ copies of $|\tilde{\pi}_{i-1}\rangle$ and $\tilde{O}(\sqrt{kd}M/\epsilon)$ calls of the quantum walk operator $W_i$.

ii. We need to apply $W_i$ for $\tilde{O}(\sqrt{kd})$ times to evolve each state $|\tilde{\pi}_{i-1}^{(j)}\rangle$ to $|\tilde{\pi}_i^{(j)}\rangle$.

Query complexity: $\tilde{O}(\sqrt{d}) \times \tilde{O}(M/\epsilon) \times \tilde{O}(\sqrt{kd}) \times O(1) = \tilde{O}(d^{3/2}k^{1/2}\epsilon^{-1})$. 

| #stages | #qsamples | Q-MALA | cost of $W_i$ |
Further Improvements?

Langevin dynamics can also be simulated by the randomized midpoint method for underdamped Langevin diffusion (ULD-RMM) (Shen-Lee’19, Durmus-Moulines’17).

<table>
<thead>
<tr>
<th>Methods</th>
<th>Sampling</th>
<th>Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALA</td>
<td>$\kappa d$</td>
<td>$\kappa d^2\epsilon^{-2}$</td>
</tr>
<tr>
<td>ULD-RMM</td>
<td>$\kappa^{7/6}d^{1/6}\epsilon^{-1/3} + \kappa d^{1/3}\epsilon^{-2/3}$</td>
<td>$\kappa^{7/6}d^{7/6}\epsilon^{-2} + \kappa d^{4/3}\epsilon^{-2}$</td>
</tr>
</tbody>
</table>

Classical query complexities. Log factors are omitted.

- In the log-concave sampling problem, ULD-RMM has only $\log(1/\epsilon)$ complexity while MALA has only $\log(1/\epsilon)$.

\[
\begin{align*}
\frac{dv_t}{dt} &= -\gamma v_t dt - \nabla f(x_t) dt + \sqrt{2\gamma} dW_t, \\
\frac{dx_t}{dt} &= v_t dt,
\end{align*}
\]

- Multi-level Monte-Carlo method is used by ULD-RMM to achieve a nearly optimal $\epsilon$-dependence for ULD-RMM.
Multi-Level Monte Carlo (MLMC)

- Consider estimating $\frac{Z_{i+1}}{Z_i} = E_{\pi_i}[g_i]$. We can express it as a telescoping sum:

$$E[g_i(X)] = E[g_i(X_0)] + E[g_i(X_1) - g_i(X_0)] + E[g_i(X_2) - g_i(X_1)] + \cdots + E[g_i(X_i) - g_i(X_{i-1})]$$

The variance $\text{var}[g_i(X_j) - g_i(X_{j-1})]$ is decreasing.

The sampling cost of $X_j$ is increasing.

- $X_j$ is sampled by simulating the Langevin dynamics with time step size $\eta_j$. MLMC chooses different number of samples $N_j$ to balance the total cost.

- (An et al.’21) developed a quantum-accelerated MLMC (QA-MLMC), which can quadratically reduce the $\epsilon$-dependence of the sample complexity of MLMC.

**Theorem (Childs-Li-Liu-Wang-Zhang’22)**

There exist quantum algorithms for estimating $Z$ with relative error $\epsilon$ using the quantum inexact ULD-RMM with $\tilde{O}(\kappa^{7/6} d^{7/6} \epsilon^{-1} + \kappa d^{4/3} \epsilon^{-1})$ queries to $O_f$.

# Quantum Log-Concave Sampling and Estimation

<table>
<thead>
<tr>
<th>Problem</th>
<th>Method</th>
<th>Complexity</th>
<th>Oracle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-concave sampling</td>
<td>MALA</td>
<td>$\kappa d / \kappa \sqrt{d}$ (warm)</td>
<td>$O_f, O_{\nabla f}$</td>
</tr>
<tr>
<td></td>
<td>Q-MALA</td>
<td>$\sqrt{\kappa d} / \sqrt{\kappa d^{1/4}}$ (warm)</td>
<td>$O_f, O_{\nabla f}$</td>
</tr>
<tr>
<td>Normalizing constant estimation</td>
<td>MALA</td>
<td>$\kappa d^2 \epsilon^{-2}$</td>
<td>$O_f, O_{\nabla f}$</td>
</tr>
<tr>
<td></td>
<td>Q-MALA</td>
<td>$\kappa^{1/2} d^{3/2} \epsilon^{-1}$</td>
<td>$O_f, O_{\nabla f}$</td>
</tr>
<tr>
<td></td>
<td>ULD-RMM</td>
<td>$\kappa^{7/6} d^{7/6} \epsilon^{-2} + \kappa d^{4/3} \epsilon^{-2}$</td>
<td>$O_f, O_{\nabla f}$</td>
</tr>
<tr>
<td></td>
<td>Q-ULD-RMM</td>
<td>$\kappa^{7/6} d^{7/6} \epsilon^{-1} + \kappa d^{4/3} \epsilon^{-1}$</td>
<td>$O_f$</td>
</tr>
</tbody>
</table>

Quantum query complexity lower bound: $\epsilon^{1-o(1)}$

Log factors are omitted.
Quantum Query Complexity Lower Bound

**Theorem** (Childs-Li-Liu-Wang-Zhang’22)

Given query access to a function $f: \mathbb{R}^d \to \mathbb{R}$ that is $1.5$-smooth and $0.5$-strongly convex, the quantum query complexity of estimating the normalizing constant $Z$ with relative error $\epsilon$ with probability at least $2/3$ is $\epsilon^{-1+o(1)}$.

*Proof idea:*

- The construction of $f$ is motivated by (Ge-Lee-Lu’21).
- A hypercube is partitioned into $n$ cells with two types (blue and yellow). Estimating normalizing constant is reduced to approximately counting the number of blue cells.
- Then, we apply the quantum lower bound on the Hamming weight problem (Nayak-Wu’99): given $x \in \{0,1\}^n$, decide whether $|x|$ is $\ell_1$ or $\ell_2$.

$$\begin{array}{cccccc} 0 & 1 & 1 & 0 & 1 & \ldots \\ \hline \text{#1's} & = & (1 - \delta)^{n/2} & \text{or} & (1 + \delta)^{n/2} \end{array}$$

$\Omega(1/\delta)$ queries!
Recent Progress in Log-Convex Sampling

- Very recently, (Fan-Yuan-Chen’23) and (Altschuler-Chewi’23) concurrently improved the classical query complexity of log-concave sampling to $\tilde{O}(\kappa \sqrt{d})$, without a warm-start.

- Our Q-MALA has query complexity $\tilde{O}(\sqrt{\kappa d})$.
  - Can be improved to $\tilde{O}(\sqrt{\kappa d}^{3/4})$ by directly quantizing (Fan-Yuan-Chen’23).
  - The extra $\sqrt{d}$ factor comes from the length of the annealing schedule.

**Open question 1:** is there a quantum log-concave sampling algorithm that beats classical algorithms in both $\kappa$ and $d$?

**Open question 2:** can ULD or ULD-RMM, which are irreversible MCs, be quantumly sped up?

- (Chewi et al.’23) proved an $\tilde{\Omega}(\log \kappa)$ query complexity lower bound for log-concave sampling.

**Open question 3:** quantum query complexity lower bound? Tighter classical lower bound?
Problem (Approximately convex optimization)

We say $F: \mathbb{R}^d \to \mathbb{R}$ is approximately convex over a convex set $\mathcal{K}$ if there is a convex function $f: \mathcal{K} \to \mathbb{R}$ such that

$$\sup_{x \in \mathcal{K}} |F(x) - f(x)| \leq \frac{\epsilon}{d}.$$ 

Given access to the evaluation oracle of $F$, find an $x^* \in \mathcal{K}$ such that

$$F(x^*) - \min_{x \in \mathcal{K}} F(x) \leq \epsilon.$$
**Problem (Stochastic convex optimization)**

We say $F: \mathcal{K} \to \mathbb{R}$ is a stochastic convex function if

$$F(x) = f(x) + \epsilon_x \quad \forall x \in \mathcal{K}$$

for some convex function $f: \mathcal{K} \to \mathbb{R}$ and $\epsilon_x$ is a sub-Gaussian random variable.

Given access to the stochastic evaluation oracle $O_f^{stoc}$, find an $x^*$ such that

$$f(x^*) - \min_{x \in \mathcal{K}} f(x) \leq \epsilon.$$

**Applications:**

- Optimization with private data (Belloni et al.’15)
- Stochastic programming (Dyer et al.’13)
- Online learning (Rakhlin et al.’12, Lattimore’20, … )
Overview of Our Results

**Approximately convex optimizer**
- The best classical algorithm due to (Belloni et al.’15) has query complexity $\tilde{O}(d^{4.5})$.
- (Li-Zhang’22) gives a quantum algorithm with query complexity $\tilde{O}(d^3)$.

**Stochastic convex optimizer**
- The best classical algorithm uses $\tilde{O}(d^{7.5}/\epsilon^2)$ queries.
- We show a quantum algorithm with $\tilde{O}(d^5/\epsilon)$ queries to the quantum stochastic oracle:

$$\mathcal{O}_f^{\text{stoc}}|x\rangle|0\rangle = |x\rangle \int_{\mathbb{R}} \sqrt{g_x(\xi)}|f(x) + \xi\rangle d\xi$$

where $g_x$ is the density of sub-gaussian random variable $\epsilon_x$.

---

Belloni, Liang, Narayanan, Rakhlin. Escaping the local minima via simulated annealing: Optimization of approximately convex functions. (COLT 15)

Li, Z. Quantum Speedups of Optimizing Approximately Convex Functions with Applications to Logarithmic Regret Stochastic Convex Bandits. (NeurIPS 22)

October 5, 2023
We consider the **quantum version** of the zeroth-order stochastic convex bandit problem:

**Definition:** Let $f: \mathcal{K} \to [0,1]$ be a convex function over $\mathcal{K} \subseteq \mathbb{R}^d$. An online quantum learner and environment interact alternatively over $T$ rounds. In each round:

\[
|\psi_t\rangle = \sum_x c_x |x\rangle |0\rangle
\]

\[
O_f^{\text{stoc}} |\psi_t\rangle
\]

\[
x_t \text{ (guess of the minimizer)}
\]

The goal is to minimize the regret: $R_T = \mathbb{E}\left[\sum_{i=1}^{T} (f(x_i) - f^*)\right]$, where $f^* = \min_{x \in \mathcal{K}} f(x)$.

- Classically, the regret has an upper bound $\tilde{O}(d^{4.5} \sqrt{T})$ and a lower bound $\Omega(d \sqrt{T})$.
- We show a quantum algorithm with regret $d^5 \text{poly}(\log(T))$, achieving an **exponential quantum advantage** in terms of $T$. 

Application: Stochastic Bandit Problem
How to Achieve Logarithmic Regret

1. Quantum stochastic bandit
2. Quantum approximately convex optimizer
3. Quantum stochastic convex optimizer
Quantum Bandit Algorithm

- $T$ rounds:
- Log($T$) intervals: $T_1$, $T_2$, $T_3$, ..., $T_i$
- Log($T$) blocks:

- Run the quantum stochastic optimizer with $\frac{2^{i-1}}{\log(T)}$ queries:
- In the next interval $T_{i+1}$ (round-$2^i$ to round-$2^{i+1} - 1$), quantum learner always outputs $X_i$.
- Each interval accumulates regret: $2^i \cdot \tilde{O}(d^5 / 2^{i-1}) = \tilde{O}(d^5)$.

\[ \Rightarrow \text{Total regret: } d^5 \cdot \text{poly} \log(T). \]

**Quantum stochastic optimizer guarantees:**

\[ f(x_i) - \min_{x \in \mathcal{X}} f(x) \leq \tilde{O} \left( \frac{d^5 \log(T)}{2^{i-1}} \right) \]

Classically, here is $2^{(i-1)/2}$, resulting in a $\sqrt{T}$ factor.

**Take-home message:** The exponential improvement comes from the quadratically faster error-decay rate in quantum.
Our goal is to find $x \in \mathcal{K}$ that minimizes $F$.

Define a distribution $\pi$ in $\mathcal{K}$ with density

$$\pi(dx) \propto e^{-F(x)/T}$$

for the approximately convex function $F$ and $T \in \mathbb{R}_+$. If we can sample from $\pi$ with small enough $T$, then

$$\mathbb{E}_{\pi}[F(X)] \approx \min_{x \in \mathcal{K}} F(x).$$
Quantum Approximately Convex Optimizer

Quantum three-level framework

- **High-level**: Perform a simulated annealing with $K = \tilde{O}(\sqrt{d})$ stages. At the $i$-th stage, the target distribution $\pi_i$ has density $\propto g_i(x) = e^{-F(x)/T_i}$, where $T_i := (1 - 1/\sqrt{d})^i$.

  $\rightarrow$ The same annealing schedule also satisfies the QSA condition.

- **Middle-level**: Use $N = \tilde{O}(d)$ samples from $\pi_i$ to construct a linear transformation $\Sigma_i$, rounding the distribution to near-isotropic position.

  $\rightarrow$ Maintain $N$ copies of the qsample $|\tilde{\pi}_i\rangle$, and apply a non-destructive rounding procedure.

- **Low-level**: Run the hit-and-run walk to evolve from $\pi_i$ to $\pi_{i+1}$ with mixing time $\tilde{O}(d^3)$.

  $\rightarrow$ Quantum walk with $\tilde{O}(d^{1.5})$ queries to obtain $|\tilde{\pi}_{i+1}\rangle$.

- Finally, measure the $N$ copies of $|\tilde{\pi}_K\rangle$ to obtain $N$ classical samples and output the best one.

Total quantum query complexity: $\tilde{O}(\sqrt{d}) \times \tilde{O}(d) \times \tilde{O}(d^{1.5}) = \tilde{O}(d^3)$. 
Hit-and-run walk

In each iteration,

1. Pick a uniformly distributed random line $\ell$ through the current point.
2. Move to a random point $y$ along the line $\ell$ chosen from the restricted distribution $\pi_\ell$. 
Quantum Speedup for Stochastic Convex Optimization

- Classically, (Belloni et al.'15) gave an algorithm with $\tilde{O}(d^{7.5}/\epsilon^2)$ queries:
  $$\tilde{O}(d^3/\epsilon^2) \times \tilde{O}(d^{4.5}) = \tilde{O}(d^{7.5}/\epsilon^2)$$
  - Reduction cost
  - Approx. convex optimization cost

- (Li-Zhang'22) gives a quantum algorithm with $\tilde{O}(d^5/\epsilon)$ queries to the quantum stochastic oracle:
  $$O_f^{\text{stoc}}|x\rangle|0\rangle = |x\rangle \int_{\mathbb{R}} \sqrt{g_x(\xi)}|f(x) + \xi\rangle d\xi,$$
  where $g_x$ is the density of sub-gaussian random variable $\epsilon_x$.

Proof idea:
- We use the quantum sub-gaussian mean estimator (Hamoudi'21) to improve the reduction cost to $\tilde{O}(d^2/\epsilon)$ queries.
- Quantum approximately convex optimization costs $\tilde{O}(d^3)$ queries.
Open Questions

1. Is there a quantum log-concave sampling algorithm that beats classical algorithms in both $\kappa$ and $d$?
2. Can ULD or ULD-RMM, which are irreversible MCs, be quantumly sped up?
3. Quantum query complexity lower bound for log-concave sampling? Tighter classical lower bound?
4. Is it possible to achieve exponential quantum advantages in some sampling problems?
5. Apply classical techniques (e.g., warm-start, average-conductance,…) to analyze the mixing time of some Lindbladians?
6. Quantum algorithm for stochastic differential equations (SDEs)?
7. More applications of provable quantum algorithms for reinforcement learning or online learning?
8. Near-term or early fault-tolerant quantum algorithm for sampling? End-to-end cost analysis for quantum algorithms for sampling problems in practice?

Thank you! Questions?