Lower bounds for quantum-inspired classical algorithms via communication complexity

Changpeng Shao

Academy of Mathematics and Systems Science, CAS
based on joint work with Nikhil S. Mande (U. Liverpool)
arXiv:2402.15686

2024-03-21
Institute of Computing Technology, CAS

Background

- Quantum computers can solve some problems much faster than classical computers.

Background

- Quantum computers can solve some problems much faster than classical computers.
- For example, solving linear systems: given $A, \mathbf{b}$, output the quantum state $|A^{-1}\mathbf{b}\rangle$. The complexity is
  - $\widetilde{O}(\|A\|\|A^{-1}\|)$ if A is sparse [Ambainis '12]
  - $\widetilde{O}(\|A\|_F\|A^{-1}\|)$ if A is stored in QRAM [Gilyén, Su, Low, Wiebe '18]

Background

- Quantum computers can solve some problems much faster than classical computers.
- For example, solving linear systems: given $A, \mathbf{b}$, output the quantum state $|A^{-1}\mathbf{b}\rangle$. The complexity is
  - $\widetilde{O}(\|A\|\|A^{-1}\|)$ if A is sparse [Ambainis '12]
  - $\widetilde{O}(\|A\|_F\|A^{-1}\|)$ if A is stored in QRAM [Gilyén, Su, Low, Wiebe '18]
- Classically, the complexity is $O(n\|A\|\|A^{-1}\|)$ with output $A^{-1}\mathbf{b}$ by conjugate gradient, where n is the dimension.

Background

- Quantum computers can solve some problems much faster than classical computers.
- For example, solving linear systems: given $A, \mathbf{b}$, output the quantum state $|A^{-1}\mathbf{b}\rangle$. The complexity is
  - $\widetilde{O}(\|A\|\|A^{-1}\|)$ if A is sparse [Ambainis '12]
  - $\widetilde{O}(\|A\|_F\|A^{-1}\|)$ if A is stored in QRAM [Gilyén, Su, Low, Wiebe '18]
- Classically, the complexity is $O(n\|A\|\|A^{-1}\|)$ with output $A^{-1}\mathbf{b}$ by conjugate gradient, where n is the dimension.
- Exponential quantum speedup wrt n!? The comparison maybe not fair because the outputs are different!

Background

- Quantum computers can solve some problems much faster than classical computers.
- For example, solving linear systems: given $A, \mathbf{b}$, output the quantum state $|A^{-1}\mathbf{b}\rangle$. The complexity is
  - $\widetilde{O}(\|A\|\|A^{-1}\|)$ if A is sparse [Ambainis '12]
  - $\widetilde{O}(\|A\|_F\|A^{-1}\|)$ if A is stored in QRAM [Gilyén, Su, Low, Wiebe '18]
- Classically, the complexity is $O(n\|A\|\|A^{-1}\|)$ with output $A^{-1}\mathbf{b}$ by conjugate gradient, where n is the dimension.
- Exponential quantum speedup wrt n!? The comparison maybe not fair because the outputs are different!
- Need more attempts to convince classical/quantum computer scientists.

Background

- From the quantum viewpoint, need to consider what can we do with a quantum state?

Background

- From the quantum viewpoint, need to consider what can we do with a quantum state?
- For a quantum state $|\mathbf{x}\rangle = \sum_i x_i |i\rangle$, we can
  - measure it: obtain i with probability $|x_i|^2$, a sampling problem.
  - query it: approximate x_i.
  - approximate norm: $\sum_i |x_i|^2$ for some problems.
  - $\ldots$.

Background

- From the quantum viewpoint, need to consider what can we do with a quantum state?
- For a quantum state $|\mathbf{x}\rangle = \sum_i x_i|i\rangle$, we can
  - measure it: obtain i with probability $|x_i|^2$, a sampling problem.
  - query it: approximate x_i.
  - approximate norm: $\sum_i |x_i|^2$ for some problems.
  - $\ldots$.
- So classically, we can similarly ask: what is the complexity of sampling/querying for $A^{-1}\mathbf{b}$?
  (a weaker problem than outputting a vector solution)

Background

- From the quantum viewpoint, need to consider what can we do with a quantum state?
- For a quantum state $|\mathbf{x}\rangle = \sum_i x_i |i\rangle$, we can
  - measure it: obtain i with probability $|x_i|^2$, a sampling problem.
  - query it: approximate x_i.
  - approximate norm: $\sum_i |x_i|^2$ for some problems.
  - $\ldots$.
- So classically, we can similarly ask: what is the complexity of sampling/querying for $A^{-1}\mathbf{b}$?
  (a weaker problem than outputting a vector solution)
- Call these algorithms quantum-inspired classical algorithms [Tang '18].

Quantum-inspired classical algorithms: assumptions

Definition 1 (Sampling and query access to a vector, SQ)

For a vector $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{C}^n$, *we have* $SQ(\mathbf{v})$, *sampling and query access to* $\mathbf{v}$, *if we can*

1. *obtain independent samples of indices* $i \in [n]$, *each distributed as* $\Pr(i) = |v_i|^2 / \|\mathbf{v}\|^2$;
2. *query for entries of* $\mathbf{v}$, *i.e., given* i, *we can query* v_i;
3. *query for* $\|\mathbf{v}\|$.

Quantum-inspired classical algorithms: assumptions

Definition 1 (Sampling and query access to a vector, SQ)

For a vector $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{C}^n$, *we have* $SQ(\mathbf{v})$, *sampling and query access to* $\mathbf{v}$, *if we can*

1. *obtain independent samples of indices* $i \in [n]$, *each distributed as* $\Pr(i) = |v_i|^2/\|\mathbf{v}\|^2$;
2. *query for entries of* $\mathbf{v}$, *i.e., given* i, *we can query* v_i;
3. *query for* $\|\mathbf{v}\|$.

Remarks.

- Similar to the tasks that can be done by having $|\mathbf{v}\rangle$.
- We have a similar definition for matrices: SQ for each row + $SQ(\|A_1\|, \ldots, \|A_n\|)$, where A_i is the i-th row.
- SQ can be achieved easily if $\mathbf{v}$ is stored in a dynamic data structure [Chia et al '19 arXiv:1910.06151].

Quantum-inspired classical algorithms: definition

By a QIC algorithm of (query) complexity T for a matrix-related problem, e.g., $\arg\min \|A\mathbf{x} - \mathbf{b}\|$, we mean we obtain $SQ(\mathbf{x})$ with T applications of $SQ(A), SQ(\mathbf{b})$ and an arbitrary number of other arithmetic operations that are independent of "SQ".

Some known results for linear regression

For the linear regression: given $SQ(A), SQ(\mathbf{b})$, output $SQ(\mathbf{x})$, where $\|\mathbf{x} - \mathbf{x}_*\| \leq \varepsilon \|\mathbf{x}_*\|$, where $\mathbf{x}_* = A^{-1}\mathbf{b}$.

Some known results for linear regression

For the linear regression: given $SQ(A), SQ(\mathbf{b})$, output $SQ(\mathbf{x})$, where $\|\mathbf{x} - \mathbf{x}_*\| \leq \varepsilon \|\mathbf{x}_*\|$, where $\mathbf{x}_* = A^{-1}\mathbf{b}$.

Progress on this problem: $\kappa_F = \|A\|_F \|A^{-1}\|$, $\kappa = \|A\| \|A^{-1}\|$,

- $\widetilde{O}(\kappa_F^{24})$ [Tang '18]
- $\widetilde{O}(\kappa_F^{6} \kappa^{16})$ [Chia et al. '19]
- $\widetilde{O}(\kappa_F^{6} \kappa^{6})$ [Gilyén, Song, Tang '20]
- $\widetilde{O}(\kappa_F^{4} \kappa^{2})$ [Montanaro, Shao '21] (consistent systems)
- $\widetilde{O}(\kappa_F^{2})$ [Montanaro, Shao '21] (row sparse consistent systems)
- $\widetilde{O}(\kappa_F^{4} \kappa^{11})$ [Bakshi, Tang '23]

Some known results for linear regression

For the linear regression: given $SQ(A), SQ(\mathbf{b})$, output $SQ(\mathbf{x})$, where $\|\mathbf{x} - \mathbf{x}_*\| \leq \varepsilon \|\mathbf{x}_*\|$, where $\mathbf{x}_* = A^{-1}\mathbf{b}$.

Progress on this problem: $\kappa_F = \|A\|_F \|A^{-1}\|$, $\kappa = \|A\| \|A^{-1}\|$,

- $\widetilde{O}(\kappa_F^{24})$ [Tang '18]
- $\widetilde{O}(\kappa_F^{6}\kappa^{16})$ [Chia et al. '19]
- $\widetilde{O}(\kappa_F^{6}\kappa^{6})$ [Gilyén, Song, Tang '20]
- $\widetilde{O}(\kappa_F^{4}\kappa^{2})$ [Montanaro, Shao '21] (consistent systems)
- $\widetilde{O}(\kappa_F^{2})$ [Montanaro, Shao '21] (row sparse consistent systems)
- $\widetilde{O}(\kappa_F^{4}\kappa^{11})$ [Bakshi, Tang '23]

A nice survey paper: Many QIC algorithms for many different problems [Chia et al. '19, arXiv.1910.06151]

So what is the limit? No lower bounds known so far!

Why we care about this?

- Quantum-inspired classical algorithms show that, assuming certain data structures, there is no exponential speedup for linear systems, and indeed for many matrix-relevant problems.

  $\Rightarrow$ Quantum-classical separation is polynomial! e.g., at most quartic for linear regression.

Why we care about this?

- Quantum-inspired classical algorithms show that, assuming certain data structures, there is no exponential speedup for linear systems, and indeed for many matrix-relevant problems.

  $\Rightarrow$ Quantum-classical separation is polynomial! e.g., at most quartic for linear regression.

- Lower bounds can help:
  1. better understand quantum-classical separations
  2. see how far these classical algorithms are far from optimal

Why we care about this?

- Quantum-inspired classical algorithms show that, assuming certain data structures, there is no exponential speedup for linear systems, and indeed for many matrix-relevant problems.

  $\Rightarrow$ Quantum-classical separation is polynomial! e.g., at most quartic for linear regression.

- Lower bounds can help:
  1. better understand quantum-classical separations
  2. see how far these classical algorithms are far from optimal

The goal of this talk: we provide the first approach (?) of computing lower bounds.

Why we care about this?

- Quantum-inspired classical algorithms show that, assuming certain data structures, there is no exponential speedup for linear systems, and indeed for many matrix-relevant problems.

  $\Rightarrow$ Quantum-classical separation is polynomial! e.g., at most quartic for linear regression.

- Lower bounds can help:
  1. better understand quantum-classical separations
  2. see how far these classical algorithms are far from optimal

The goal of this talk: we provide the first approach (?) of computing lower bounds.

We are more concerned about the dependence on κ_F, because $\|A\|_F > \|A\|$. So it is usually the dominating term.

Summary of the main results for linear regression

Recall: $\kappa_F = \|A\|_F \|A^{-1}\|$, $\kappa = \|A\| \|A^{-1}\|$, $\gamma = \|A\mathbf{x}_*\| / \|\mathbf{b}\|$

	Algs	Upper bounds	Lower bounds
Row sparse	Q	$\widetilde{O}(\kappa_F/\gamma)$	$\widetilde{\Omega}(\kappa_F/\gamma)$
		$\widetilde{O}(\kappa/\gamma)$ column is sparse too	$\widetilde{\Omega}(\kappa/\gamma)$
	QIC	$\widetilde{O}(\kappa_F^2)$ assumes $\gamma = 1$	$\widetilde{\Omega}(\kappa_F^2 + 1/\gamma^2)$
Dense	Q	$\widetilde{O}(\kappa_F/\gamma)$	$\widetilde{\Omega}(\kappa_F/\gamma)$
	QIC	$\widetilde{O}(\kappa_F^4 \kappa^{11}/\varepsilon^2\gamma^2)$	$\widetilde{\Omega}(\kappa_F^4 + 1/\gamma^2)^*$
		$\widetilde{O}(\kappa_F^4 \kappa^2/\varepsilon^2)$ assumes $\gamma = 1$	

Summary of the main results for other problems

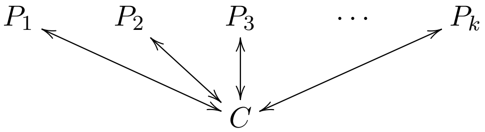
Problems	Upper bounds (Q)	Upper bounds (QIC)	Lower bounds (QIC)
Clustering	$O(\frac{\|A\|_F^2\|\mathbf{b}\|^2}{\varepsilon})$	$O(\frac{\|A\|_F^4\|\mathbf{b}\|^4}{\varepsilon^2})$	$\Omega(\frac{\|A\|_F^4\|\mathbf{b}\|^4}{\varepsilon})$
PCA	$O(\|A\|_F)$	$O(\|A\|_F^6)$	$\Omega(\|A\|_F^2)$
RecSys	$O(\|A\|_F)$	$O(\|A\|_F^4)$	$\Omega(\|A\|_F^2)$
HS	$O(\|A\|_F)$	$O(\|A\|_F^4)$	$\Omega(\|A\|_F^2)$

Key ingredient: communication complexity

- Introduced by Yao in 1979.
- Useful in many applications, especially lower bounds analysis.
- DISJ problem: Alice has $(x_1, \ldots, x_n) \in \{0,1\}^n$ and Bob has $(y_1, \ldots, y_n) \in \{0,1\}^n$, they want to determine if $\exists j$ such that $x_j = y_j = 1$ using as little communication as possible.
- CC $=$ number of bits used in the communication.
- For DISJ, CC $= \Theta(n)$.[Razborov, '90].

Multi-player coordinator model (number in hand)

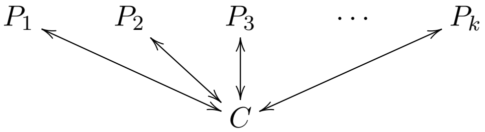
There are $k \geq 2$ players $P_1, \ldots, P_k$ and a coordinator C:



Each player holds some private information, and their goal is to solve some problem using as little communication as possible.

Multi-player coordinator model (number in hand)

There are $k \geq 2$ players $P_1, \ldots, P_k$ and a coordinator C:



Each player holds some private information, and their goal is to solve some problem using as little communication as possible.

For example, k-player DISJ problem:
P_i has $x_i = (x_{i1}, \ldots, x_{in}) \in \{0,1\}^n$. Their goal is to determine if there is a $j \in \{2, \ldots, k\}$ such that $\mathsf{DISJ}(P_1, P_j) = 1$.

CC $= \Theta(kn)$ [Phillips, Verbin, Zhang, '11].

Key theorem: A connection between QIC and CC

Theorem 1

Assume P_i *holds a matrix* $A^{(i)} \in \mathbb{R}^{\ell_i \times n}$ *and a vector* $\mathbf{b}^{(i)} \in \mathbb{R}^{m_i}$ *with* $m := \sum_i \ell_i = \sum_i m_i$. *Assume that all entries are specified by* $O(\log q)$ *bits. Let*

$$A = \begin{pmatrix} A^{(1)} \\ \vdots \\ A^{(k)} \end{pmatrix}_{m \times n}, \quad \mathbf{b} = \begin{pmatrix} \mathbf{b}^{(1)} \\ \vdots \\ \mathbf{b}^{(k)} \end{pmatrix}_{m \times 1}.$$

Then we have the following:

- *The coordinator* C *can use* $SQ(A)$ $O(T)$ *times, using* $O((T+k)\log(qmn))$ *bits of communication.*
- *The coordinator* C *can use* $SQ(\mathbf{b})$ $O(T)$ *times, using* $O((T+k)\log(qm))$ *bits of communication.*

Implications

Theorem 1 implies that a QIC algorithm of complexity T induces a communication protocol for the same problem with complexity $\widetilde{O}(T + k)$.

Implications

Theorem 1 implies that a QIC algorithm of complexity T induces a communication protocol for the same problem with complexity $\widetilde{O}(T + k)$.

- Prove lower bounds of QIC from CC. ($\checkmark$)
- Propose efficient communication protocols from QIC.

A short proof for $SQ(\mathbf{b})$

Assume $k = 2$ and $\mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix}$, where Alice has $\mathbf{b}_1 \in \mathbb{C}^m$ and Bob has $\mathbf{b}_2 \in \mathbb{C}^n$.

- Query the i-th entry: if $i \leq m$, then the coordinator C asks Alice return the i-th entry, otherwise asks Bob return the $(i - m)$-th entry

A short proof for $SQ(\mathbf{b})$

Assume $k = 2$ and $\mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix}$, where Alice has $\mathbf{b}_1 \in \mathbb{C}^m$ and Bob has $\mathbf{b}_2 \in \mathbb{C}^n$.

- Query the i-th entry: if $i \leq m$, then the coordinator C asks Alice return the i-th entry, otherwise asks Bob return the $(i - m)$-th entry
- Query the norm: Alice and Bob send $\|\mathbf{b}_1\|, \|\mathbf{b}_2\|$ to C.

A short proof for $SQ(\mathbf{b})$

Assume $k = 2$ and $\mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix}$, where Alice has $\mathbf{b}_1 \in \mathbb{C}^m$ and Bob has $\mathbf{b}_2 \in \mathbb{C}^n$.

- Query the i-th entry: if $i \le m$, then the coordinator C asks Alice return the i-th entry, otherwise asks Bob return the $(i - m)$-th entry
- Query the norm: Alice and Bob send $\|\mathbf{b}_1\|, \|\mathbf{b}_2\|$ to C.
- Sampling: C samples an index $i \in \{1, 2\}$ from

$$\left\{ \frac{\|\mathbf{b}_1\|^2}{\|\mathbf{b}\|^2}, \quad \frac{\|\mathbf{b}_2\|^2}{\|\mathbf{b}\|^2} \right\}$$

  If receives 1 then C asks Alice return a sample from $\mathbf{b}_1$, otherwise C asks Bob return a sample from $\mathbf{b}_2$.

Lower bounds for linear regressions

Recall: we are given $SQ(A)$ and $SQ(\mathbf{b})$, the goal is to output $SQ(\tilde{\mathbf{x}}_*)$ such that $\|\tilde{\mathbf{x}}_* - \mathbf{x}_*\| \leq \varepsilon \|\mathbf{x}_*\|$, where $\mathbf{x}_* = A^+\mathbf{b}$.

There are 3 tasks in $SQ(\mathbf{x})$:

- Sampling: obtain i with probability $x_i^2/\|\mathbf{x}\|^2$
- Query entries: given i output x_i
- Norm: output $\|\mathbf{x}\|^2$

Lower bounds for sampling

Proposition 1 (row sparse case)

Assume that A is row sparse and $\varepsilon \in (0,1)$ is a constant. Then ε-approximately sampling from $A^+\mathbf{b}$ requires making

$$\widetilde{\Omega}((\kappa^2 + \kappa_F)/\gamma)$$

calls to $SQ(A), SQ(\mathbf{b})$.

Lower bounds for sampling: proof sketch

Based on k-DISJ: P_i has $T_i = (T_{i1}, \ldots, T_{in}) \in \{0,1\}^n$. Their goal is to determine if $\exists j \in \{2, \ldots, k\}$ such that DISJ$(P_1, P_j) = 1$.

Lower bounds for sampling: proof sketch

Based on k-DISJ: P_i has $T_i = (T_{i1}, \ldots, T_{in}) \in \{0,1\}^n$. Their goal is to determine if $\exists j \in \{2, \ldots, k\}$ such that $\mathrm{DISJ}(P_1, P_j) = 1$.

Construction: let

$$\mathbf{t}_j = \sum_{\ell=1}^{n} T_{j\ell} |\ell\rangle_n, \quad |\mathbf{t}_j\rangle = \mathbf{t}_j / \|\mathbf{t}_j\|,$$

and

$$A = \begin{pmatrix} \beta|1\rangle_n & & & \\ & |\mathbf{t}_2\rangle & & \\ & & \ddots & \\ & & & |\mathbf{t}_k\rangle \end{pmatrix}_{kn\times k}, \quad \mathbf{b} = \begin{pmatrix} \beta|1\rangle_n \\ n|\mathbf{t}_1\rangle \\ \vdots \\ n|\mathbf{t}_1\rangle \end{pmatrix}_{kn\times 1},$$

where β is a free parameter.

Lower bounds for sampling: proof sketch

If no such j:

$$\mathbf{x}_* = |1\rangle_k.$$

If $\mathrm{DISJ}(P_1, P_j) = 1$, then

$$\mathbf{x}_* = |1\rangle_k + \frac{n}{\|\mathbf{t}_1\|\|\mathbf{t}_j\|}|j\rangle_k \approx |1\rangle_k + |j\rangle_k$$

because k-DISJ is still hard when $\|\mathbf{t}_j\| = \Theta(n)$ for all j.

Lower bounds for sampling: proof sketch

If no such j:

$$\mathbf{x}_* = |1\rangle_k.$$

If $\mathsf{DISJ}(P_1, P_j) = 1$, then

$$\mathbf{x}_* = |1\rangle_k + \frac{n}{\|\mathbf{t}_1\|\|\mathbf{t}_j\|}|j\rangle_k \approx |1\rangle_k + |j\rangle_k$$

because k-DISJ is still hard when $\|\mathbf{t}_j\| = \Theta(n)$ for all j.

Now

$$\kappa^2 = \frac{\max(\beta^2, 1)}{\min(\beta^2, 1)}, \quad \kappa_F^2 = \frac{\beta^2 + k - 1}{\min(\beta^2, 1)}, \quad \gamma^2 = \frac{\beta^2 + O(1)}{\beta^2 + (k-1)n^2}.$$

Lower bounds for sampling: proof sketch

If no such j:

$$\mathbf{x}_* = |1\rangle_k.$$

If DISJ$(P_1, P_j) = 1$, then

$$\mathbf{x}_* = |1\rangle_k + \frac{n}{\|\mathbf{t}_1\|\|\mathbf{t}_j\|}|j\rangle_k \approx |1\rangle_k + |j\rangle_k$$

because k-DISJ is still hard when $\|\mathbf{t}_j\| = \Theta(n)$ for all j.

Now

$$\kappa^2 = \frac{\max(\beta^2, 1)}{\min(\beta^2, 1)}, \quad \kappa_F^2 = \frac{\beta^2 + k - 1}{\min(\beta^2, 1)}, \quad \gamma^2 = \frac{\beta^2 + O(1)}{\beta^2 + (k-1)n^2}.$$

- If $\beta^2 = k$, then $\kappa^2, \kappa_F^2 = \Theta(k), \gamma^2 = \Theta(1/n^2) \Rightarrow$ a lower bound is $\Omega(\kappa^2/\gamma)$ because CC(k-DISJ)=$\Theta(kn)$.

Lower bounds for sampling: proof sketch

If no such j:

$$\mathbf{x}_* = |1\rangle_k.$$

If DISJ$(P_1, P_j) = 1$, then

$$\mathbf{x}_* = |1\rangle_k + \frac{n}{\|\mathbf{t}_1\|\|\mathbf{t}_j\|}|j\rangle_k \approx |1\rangle_k + |j\rangle_k$$

because k-DISJ is still hard when $\|\mathbf{t}_j\| = \Theta(n)$ for all j.

Now

$$\kappa^2 = \frac{\max(\beta^2, 1)}{\min(\beta^2, 1)}, \quad \kappa_F^2 = \frac{\beta^2 + k - 1}{\min(\beta^2, 1)}, \quad \gamma^2 = \frac{\beta^2 + O(1)}{\beta^2 + (k-1)n^2}.$$

- If $\beta^2 = k$, then $\kappa^2, \kappa_F^2 = \Theta(k), \gamma^2 = \Theta(1/n^2) \Rightarrow$ a lower bound is $\Omega(\kappa^2/\gamma)$ because CC(k-DISJ)=$\Theta(kn)$.
- If $\beta^2 = 1$, then $\kappa = 1, \kappa_F^2 = \Theta(k), \gamma^2 = \Theta(1/kn^2) \Rightarrow$ a lower bound is $\Omega(\kappa_F/\gamma)$.

Lower bounds for sampling

Proposition 2 (dense case)

Assume that $\varepsilon \in (0,1)$ *is a constant. Then* ε*-approximately sampling from* $A^+\mathbf{b}$ *requires making*

$$\widetilde{\Omega}(\kappa_F^2)$$

calls to $SQ(A), SQ(\mathbf{b})$.

Lower bounds for sampling: proof sketch

Based on distributed sampling problem: Alice has $f : \{0,1\}^n \to \{\pm 1\}$ and Bob has $g : \{0,1\}^n \to \{\pm 1\}$. Their goal is to approximately sample from the distribution defined by

$$\Pr(y) := \left(\frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x) g(x) (-1)^{x \cdot y} \right)^2 .$$

CC $= \Omega(2^n)$ [Montanaro '19]

Lower bounds for sampling: proof sketch

Based on distributed sampling problem: Alice has $f : \{0,1\}^n \to \{\pm 1\}$ and Bob has $g : \{0,1\}^n \to \{\pm 1\}$. Their goal is to approximately sample from the distribution defined by

$$\Pr(y) := \left(\frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x) g(x) (-1)^{x \cdot y} \right)^2 .$$

CC $= \Omega(2^n)$ [Montanaro '19]

Let $D_f = \sum_x f(x) |x\rangle\langle x|$ and $|g\rangle = \frac{1}{\sqrt{2^n}} \sum_x g(x) |x\rangle$, then the distribution is defined by the state $(D_f H^{\otimes n})^{-1} |g\rangle$, where $H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$.

Lower bounds for sampling: proof sketch

Based on distributed sampling problem: Alice has $f:\{0,1\}^n \to \{\pm 1\}$ and Bob has $g:\{0,1\}^n \to \{\pm 1\}$. Their goal is to approximately sample from the distribution defined by

$$\Pr(y) := \left(\frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x) g(x) (-1)^{x \cdot y} \right)^2 .$$

CC $= \Omega(2^n)$ [Montanaro '19]

Let $D_f = \sum_x f(x)|x\rangle\langle x|$ and $|g\rangle = \frac{1}{\sqrt{2^n}} \sum_x g(x)|x\rangle$, then the distribution is defined by the state $(D_f H^{\otimes n})^{-1}|g\rangle$, where $H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$.

Now we have $\kappa_F^2 = 2^n$, $\kappa = \gamma = 1$. So we obtain a lower bound of $\widetilde{\Omega}(\kappa_F^2)$.

Lower bounds for norm approximation

Proposition 3 (Sparse case)

Assume that A is row sparse and $\varepsilon \in (0,1)$. Then approximating $\|A^{+}\mathbf{b}\|$ up to relative error ε requires making

$$\widetilde{\Omega}(\kappa_F^2 + 1/\gamma^2)$$

calls to $SQ(A), SQ(\mathbf{b})$.

Lower bounds for norm approximation: proof sketch

Gap-Hamming problem: Alice has $\mathbf{a} = (a_1, \ldots, a_n) \in \{\pm 1\}^n$ and Bob has $\mathbf{b} = (b_1, \ldots, b_n) \in \{\pm 1\}^n$. The goal is to determine if $\mathbf{a} \cdot \mathbf{b} \geq \sqrt{n}$ or $\leq -\sqrt{n}$. CC $= \Theta(n)$. [Chakrabarti, Regev '11]

Lower bounds for norm approximation: proof sketch

Gap-Hamming problem: Alice has $\mathbf{a} = (a_1, \ldots, a_n) \in \{\pm 1\}^n$ and Bob has $\mathbf{b} = (b_1, \ldots, b_n) \in \{\pm 1\}^n$. The goal is to determine if $\mathbf{a} \cdot \mathbf{b} \geq \sqrt{n}$ or $\leq -\sqrt{n}$. CC $=$ $\Theta(n)$. [Chakrabarti, Regev '11]

Construction:

$$A = \begin{pmatrix} 1 & & & a_1/\sqrt{n} \\ & \ddots & & \vdots \\ & & 1 & a_n/\sqrt{n} \\ & & & 1 \end{pmatrix}, \quad \tilde{\mathbf{b}} = \begin{pmatrix} b_1/\sqrt{n} \\ \vdots \\ b_n/\sqrt{n} \\ 1 \end{pmatrix}.$$

The solution of $A\mathbf{x} = \tilde{\mathbf{b}}$ is $\mathbf{x} = |n+1\rangle + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (b_i - a_i)|i\rangle$. So $\|\mathbf{x}\|^2 = 3 - \frac{2}{n} \mathbf{a} \cdot \mathbf{b}$.

Lower bounds for norm approximation: proof sketch

Gap-Hamming problem: Alice has $\mathbf{a} = (a_1, \ldots, a_n) \in \{\pm 1\}^n$ and Bob has $\mathbf{b} = (b_1, \ldots, b_n) \in \{\pm 1\}^n$. The goal is to determine if $\mathbf{a} \cdot \mathbf{b} \geq \sqrt{n}$ or $\leq -\sqrt{n}$. CC $= \Theta(n)$. [Chakrabarti, Regev '11]

Construction:

$$A = \begin{pmatrix} 1 & & & a_1/\sqrt{n} \\ & \ddots & & \vdots \\ & & 1 & a_n/\sqrt{n} \\ & & & 1 \end{pmatrix}, \quad \tilde{\mathbf{b}} = \begin{pmatrix} b_1/\sqrt{n} \\ \vdots \\ b_n/\sqrt{n} \\ 1 \end{pmatrix}.$$

The solution of $A\mathbf{x} = \tilde{\mathbf{b}}$ is $\mathbf{x} = |n+1\rangle + \frac{1}{\sqrt{n}} \sum_{i=1}^n (b_i - a_i)|i\rangle$. So $\|\mathbf{x}\|^2 = 3 - \frac{2}{n}\mathbf{a} \cdot \mathbf{b}$.

- If $\mathbf{a} \cdot \mathbf{b} \geq \sqrt{n}$, then $\|\mathbf{x}\|^2 \leq 3 - 2/\sqrt{n}$
- If $\mathbf{a} \cdot \mathbf{b} \leq -\sqrt{n}$, then $\|\mathbf{x}\|^2 \geq 3 + 2/\sqrt{n}$

Lower bounds for norm approximation: proof sketch

Gap-Hamming problem: Alice has $\mathbf{a} = (a_1, \ldots, a_n) \in \{\pm 1\}^n$ and Bob has $\mathbf{b} = (b_1, \ldots, b_n) \in \{\pm 1\}^n$. The goal is to determine if $\mathbf{a} \cdot \mathbf{b} \geq \sqrt{n}$ or $\leq -\sqrt{n}$. CC $= \Theta(n)$. [Chakrabarti, Regev '11]

Construction:

$$A = \begin{pmatrix} 1 & & & a_1/\sqrt{n} \\ & \ddots & & \vdots \\ & & 1 & a_n/\sqrt{n} \\ & & & 1 \end{pmatrix}, \quad \tilde{\mathbf{b}} = \begin{pmatrix} b_1/\sqrt{n} \\ \vdots \\ b_n/\sqrt{n} \\ 1 \end{pmatrix}.$$

The solution of $A\mathbf{x} = \tilde{\mathbf{b}}$ is $\mathbf{x} = |n+1\rangle + \frac{1}{\sqrt{n}} \sum_{i=1}^n (b_i - a_i)|i\rangle$. So $\|\mathbf{x}\|^2 = 3 - \frac{2}{n}\mathbf{a} \cdot \mathbf{b}$.

- If $\mathbf{a} \cdot \mathbf{b} \geq \sqrt{n}$, then $\|\mathbf{x}\|^2 \leq 3 - 2/\sqrt{n}$
- If $\mathbf{a} \cdot \mathbf{b} \leq -\sqrt{n}$, then $\|\mathbf{x}\|^2 \geq 3 + 2/\sqrt{n}$

So choose $\varepsilon \approx 1/\sqrt{n}$. Fortunately, the dependence on ε is $\log(1/\varepsilon)$ for QIC. Now $\kappa_F^2 = \Theta(n), \kappa = \Theta(1), \gamma = 1$. So we obtain a lower bound of $\widetilde{\Omega}(\kappa_F^2)$.

Lower bounds for querying an entry

Proposition 4

Assume that A is sparse and $\varepsilon \in (0,1)$. Then there is an index i such that computing x_i with $|x_i - (A^+\mathbf{b})_i| \le \varepsilon$ requires making $\widetilde{\Omega}(\kappa_F^2 + 1/\gamma^2)$ calls to $SQ(A), SQ(\mathbf{b})$.

Still based on Gap-Hamming problem. Let

$$A = \begin{pmatrix} 1/a_1 & -t & & \\ & a_1/a_2 & \ddots & \\ & & \ddots & -t \\ & & & a_{n-1}/a_n \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2/t \\ \vdots \\ b_n/t^{n-1} \end{pmatrix}.$$

The first entry of the solution of $A\mathbf{x} = \mathbf{b}$ is $\sum_i a_i b_i$.
If we choose $t = 1/2$, then $\kappa = \Theta(1), \kappa_F^2 = \Theta(n)$ and $\gamma = 1$. This leads to a lower bound of $\widetilde{\Omega}(\kappa_F^2)$.

Lower bounds for norm approximation

Proposition 5 (Dense case)

Assume that $\varepsilon \in (0,1)$. Then approximating $\|A^{+}\mathbf{b}\|^2$ up to additive error ε requires making $\widetilde{\Omega}(\kappa_F^4)$ calls to $SQ(A), SQ(\mathbf{b})$.

Lower bounds for norm approximation

Proposition 5 (Dense case)

Assume that $\varepsilon \in (0,1)$. Then approximating $\|A^{+}\mathbf{b}\|^2$ up to additive error ε requires making $\widetilde{\Omega}(\kappa_F^4)$ calls to $SQ(A), SQ(\mathbf{b})$.

Based on k-Gap-Hamming problem: For $i \in \{1, \ldots, k+1\}$, player P_i has $\mathbf{a}_i \in \{\pm 1\}^n$ with the promise that $\mathbf{a} := \sum_{i=1}^{k} \mathbf{a}_i \in \{\pm 1\}^n$ and $\mathbf{a} \cdot \mathbf{a}_{k+1} \in [-c_2\sqrt{n}, c_2\sqrt{n}]$. Their goal is to determine if $\mathbf{a} \cdot \mathbf{a}_{k+1} \geq c_1\sqrt{n}$ or $\leq -c_1\sqrt{n}$. Here $0 < c_1 < c_2$ are constants.

CC $= \Omega(kn)$. [Li, Lin, Woodruff, '23]

Lower bounds for norm approximation: proof sketch

Let $\mathbf{e} = \frac{1}{m}(1, \ldots, 1) \in \mathbb{R}^n$, where m is an integer such that $\mathbf{e} \cdot \mathbf{a}_{k+1} = O(1)$.

Let $M \in \mathbb{R}^{k \times n}$ be the matrix such that the i-th row is $\mathbf{a}_i + \mathbf{e}$ and

$$A = \begin{pmatrix} I_n & \mathbf{0} \\ M/2n & I_k \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} -2n\mathbf{a}_{k+1} \\ \mathbf{0} \end{pmatrix}.$$

The solution is

$$\mathbf{x} = \begin{pmatrix} -2n\mathbf{a}_{k+1} \\ (\mathbf{a}_1 + \mathbf{e}) \cdot \mathbf{a}_{k+1} \\ \vdots \\ (\mathbf{a}_1 + \mathbf{e}) \cdot \mathbf{a}_{k+1} \end{pmatrix}.$$

Lower bounds for norm approximation: proof sketch

So

$$\|\mathbf{x}\|^2 = 4n^3 + \sum_{i=1}^{k} \Big((\mathbf{a}_i \cdot \mathbf{a}_{k+1})^2 + (\mathbf{a}_{k+1} \cdot \mathbf{e})^2 + 2(\mathbf{a}_i \cdot \mathbf{a}_{k+1})(\mathbf{a}_{k+1} \cdot \mathbf{e}) \Big).$$

Lower bounds for norm approximation: proof sketch

So

$$\|\mathbf{x}\|^2 = 4n^3 + \sum_{i=1}^{k} \Big((\mathbf{a}_i \cdot \mathbf{a}_{k+1})^2 + (\mathbf{a}_{k+1} \cdot \mathbf{e})^2 + 2(\mathbf{a}_i \cdot \mathbf{a}_{k+1})(\mathbf{a}_{k+1} \cdot \mathbf{e}) \Big).$$

Similarly, if we set M to be the matrix such that the i-th row is $\mathbf{a}_i - \mathbf{e}$, then

$$\|\mathbf{x}'\|^2 = 4n^3 + \sum_{i=1}^{k} \Big((\mathbf{a}_i \cdot \mathbf{a}_{k+1})^2 + (\mathbf{a}_{k+1} \cdot \mathbf{e})^2 - 2(\mathbf{a}_i \cdot \mathbf{a}_{k+1})(\mathbf{a}_{k+1} \cdot \mathbf{e}) \Big).$$

Lower bounds for norm approximation: proof sketch

So

$$\|\mathbf{x}\|^2 = 4n^3 + \sum_{i=1}^{k} \Big((\mathbf{a}_i \cdot \mathbf{a}_{k+1})^2 + (\mathbf{a}_{k+1} \cdot \mathbf{e})^2 + 2(\mathbf{a}_i \cdot \mathbf{a}_{k+1})(\mathbf{a}_{k+1} \cdot \mathbf{e}) \Big).$$

Similarly, if we set M to be the matrix such that the i-th row is $\mathbf{a}_i - \mathbf{e}$, then

$$\|\mathbf{x}'\|^2 = 4n^3 + \sum_{i=1}^{k} \Big((\mathbf{a}_i \cdot \mathbf{a}_{k+1})^2 + (\mathbf{a}_{k+1} \cdot \mathbf{e})^2 - 2(\mathbf{a}_i \cdot \mathbf{a}_{k+1})(\mathbf{a}_{k+1} \cdot \mathbf{e}) \Big).$$

Now

$$\|\mathbf{x}\|^2 - \|\mathbf{x}'\|^2 = 4(\mathbf{a}_{k+1} \cdot \mathbf{e}) \sum_{i=1}^{k} (\mathbf{a}_i \cdot \mathbf{a}_{k+1}).$$

Since $\mathbf{a}_{k+1} \cdot \mathbf{e} = \Theta(1)$, if we can approximate $\|\mathbf{x}\|^2 - \|\mathbf{x}'\|^2$, we then can solve the k-Gap-Hamming problem.

Lower bounds for norm approximation: proof sketch

So

$$\|\mathbf{x}\|^2 = 4n^3 + \sum_{i=1}^{k} \Big((\mathbf{a}_i \cdot \mathbf{a}_{k+1})^2 + (\mathbf{a}_{k+1} \cdot \mathbf{e})^2 + 2(\mathbf{a}_i \cdot \mathbf{a}_{k+1})(\mathbf{a}_{k+1} \cdot \mathbf{e}) \Big).$$

Similarly, if we set M to be the matrix such that the i-th row is $\mathbf{a}_i - \mathbf{e}$, then

$$\|\mathbf{x}'\|^2 = 4n^3 + \sum_{i=1}^{k} \Big((\mathbf{a}_i \cdot \mathbf{a}_{k+1})^2 + (\mathbf{a}_{k+1} \cdot \mathbf{e})^2 - 2(\mathbf{a}_i \cdot \mathbf{a}_{k+1})(\mathbf{a}_{k+1} \cdot \mathbf{e}) \Big).$$

Now

$$\|\mathbf{x}\|^2 - \|\mathbf{x}'\|^2 = 4(\mathbf{a}_{k+1} \cdot \mathbf{e}) \sum_{i=1}^{k} (\mathbf{a}_i \cdot \mathbf{a}_{k+1}).$$

Since $\mathbf{a}_{k+1} \cdot \mathbf{e} = \Theta(1)$, if we can approximate $\|\mathbf{x}\|^2 - \|\mathbf{x}'\|^2$, we then can solve the k-Gap-Hamming problem.

Now $\kappa_F^2 = k + n$, $\kappa = \Theta(1)$ and $\gamma = 1$. Yields a lower bound of $\Omega(\kappa_F^4)$ by setting $k \approx n$.

Lower bounds for querying an entry

Proposition 6 (Dense case)

Assume that $\varepsilon \in (0,1)$. *Then there is an index* i *such that computing* x_i *with* $|x_i - (A^+\mathbf{b})_i| \le \varepsilon$ *requires making* $\widetilde{\Omega}(\kappa_F^4)$ *calls to* $SQ(A), SQ(\mathbf{b})$.

Lower bounds for querying an entry

Proposition 6 (Dense case)

Assume that $\varepsilon \in (0,1)$. Then there is an index i such that computing x_i with $|x_i - (A^+\mathbf{b})_i| \leq \varepsilon$ requires making $\widetilde{\Omega}(\kappa_F^4)$ calls to $SQ(A), SQ(\mathbf{b})$.

Still based on k-Gap-Hamming problem. Now

$$A = \begin{pmatrix} I_n & \mathbf{0} \\ M/2n & I_k \end{pmatrix} \begin{pmatrix} I_n & \mathbf{0} \\ \mathbf{0} & H^{\otimes l} \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} -2n\mathbf{a}_{k+1} \\ \mathbf{0} \end{pmatrix}.$$

The solution is

$$\mathbf{x} = \begin{pmatrix} -2n\mathbf{a}_{k+1} \\ \frac{1}{\sqrt{k}} \sum_{i=1}^k \mathbf{a}_i \cdot \mathbf{a}_{k+1} \\ \vdots \end{pmatrix}.$$

The $(n+1)$-th is $\sum_i \mathbf{a}_i \cdot \mathbf{b}/\sqrt{k}$. In particular, if $k = c_1 n$, then $\frac{1}{\sqrt{k}} \sum_{i=1}^k \mathbf{a}_i \cdot \mathbf{a}_{k+1} \geq c_1\sqrt{n}$ or $\leq -c_1\sqrt{n}$.

Summary

- Provide an approach to prove lower bounds of quantum-inspired classical algorithms.
- We only made some partial progress using this approach.
- Not covered:
  - lower bounds for other problems
  - reduction from quantum query to communication complexity for matrix-relevant problems

Open questions

- Prove tighter lower bounds in the dense case for linear regression.

Open questions

- Prove tighter lower bounds in the dense case for linear regression.
  The main difficulty:
  - There are many parameters to be considered: κ_F, κ, γ.
  - Find the right reduction from linear regression to some problems in communication complexity (number in hand model).
  - In the reduction, we need to estimate κ_F, κ, γ efficiently.

Open questions

- Prove tighter lower bounds in the dense case for linear regression.
  The main difficulty:
  - There are many parameters to be considered: κ_F, κ, γ.
  - Find the right reduction from linear regression to some problems in communication complexity (number in hand model).
  - In the reduction, we need to estimate κ_F, κ, γ efficiently.
- Prove tighter lower bounds for other problems.

Open questions

- Prove tighter lower bounds in the dense case for linear regression.
  The main difficulty:
  - There are many parameters to be considered: κ_F, κ, γ.
  - Find the right reduction from linear regression to some problems in communication complexity (number in hand model).
  - In the reduction, we need to estimate κ_F, κ, γ efficiently.
- Prove tighter lower bounds for other problems.
- Propose efficient communication protocols from quantum-inspired classical algorithms.

Open questions

- Prove tighter lower bounds in the dense case for linear regression.
  The main difficulty:
  - There are many parameters to be considered: κ_F, κ, γ.
  - Find the right reduction from linear regression to some problems in communication complexity (number in hand model).
  - In the reduction, we need to estimate κ_F, κ, γ efficiently.
- Prove tighter lower bounds for other problems.
- Propose efficient communication protocols from quantum-inspired classical algorithms.

Thanks very much for your attention!