Quantum communication complexity of linear regressions

Changpeng Shao

School of Mathematics, University of Bristol, UK
based on joint work with Ashley Montanaro
arXiv:2210.01601

IRIF AlgoComp Seminar
6th December 2022

Background

Quantum computers can solve some problems much faster than classical computers, e.g., Shor's algorithm, quantum simulation, Grover's algorithm, etc.

For example, quantum computers could be good at solving linear algebra problems.

A fundamental problem:

Solving linear systems.
Given some quantum access to a matrix A, and the quantum state $|\mathrm{b}\rangle$, output $|A^{-1}\mathrm{b}\rangle$. Here A^{-1} means the pseudoinverse.

A brief history of quantum algorithms for linear systems

- 2008, Harrow, Hassidim, Lloyd: if A is s-sparse, then the cost is $\widetilde{O}(s\kappa^2/\varepsilon)$.
- 2010, Ambainis: $\widetilde{O}(s\kappa/\varepsilon^3)$.
- 2015, Childs, Kothari, Somma: $\widetilde{O}(s\kappa)$.
- 2017, Wossnig, Zhao, Prakash: dense case using QRAM, $\widetilde{O}(\kappa^2\|A\|_F/\varepsilon)$.
- 2018, Chakraborty, Gilyén, Jeffery: using α-block-encoding $\widetilde{O}(\alpha\kappa)$. E.g., $\alpha = s$ if sparse, and $\alpha = \|A\|_F$ if assuming QRAM.
- .....

Leads to wide applications, especially in machine learning, e.g., recommendation systems [Kerenidis and Prakash, arXiv:1603.08675].

Quantum-inspired classical algorithms

In 2018, Tang showed that assuming a similar data structure to QRAM, there is a classical algorithm that can solve the recommendation systems in polylog time. [Tang, arXiv:1807.04271].

So no quantum exponential speedups for many machine learning problems in the low-rank case, this includes solving linear systems $A\mathbf{x} = \mathbf{b}$. [Chia et al, arXiv:1910.06151].

No exponential speedups is with respect to time and query complexities.

The main takeaway

How about the quantum speedups in terms of communication complexity?

The main takeaway

How about the quantum speedups in terms of communication complexity?

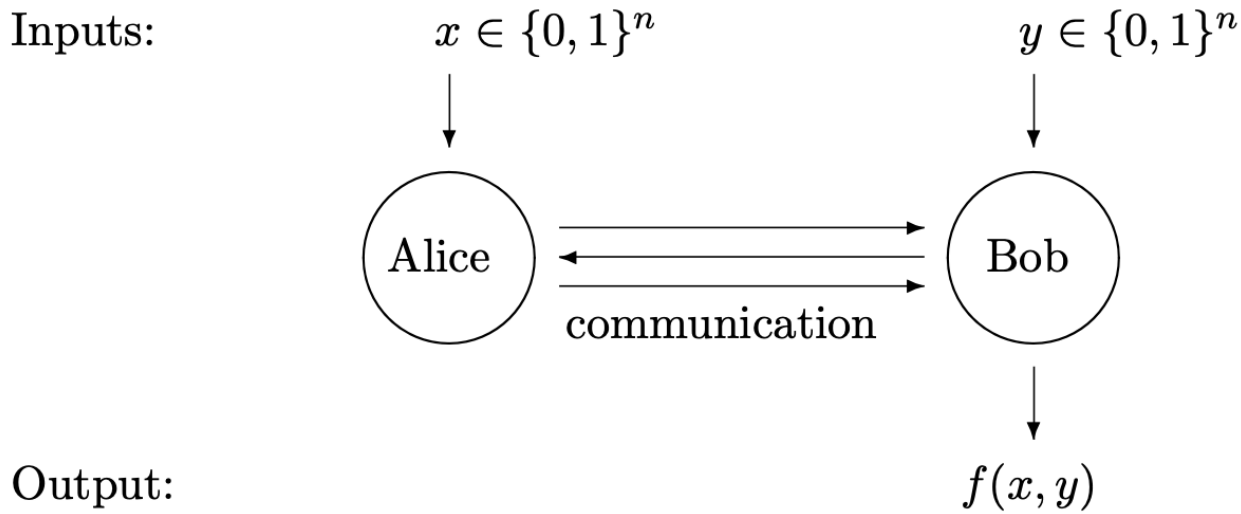
For linear repressions, quantum computers can have exponential speedups in terms of communication complexity.

Communication complexity: classical case

Let f be a function, say from $\{0,1\}^n \times \{0,1\}^n$ to $\{0,1\}$. Alice receives $x \in \{0,1\}^n$ and Bob receives $y \in \{0,1\}^n$, they want to compute $f(x,y)$ together.

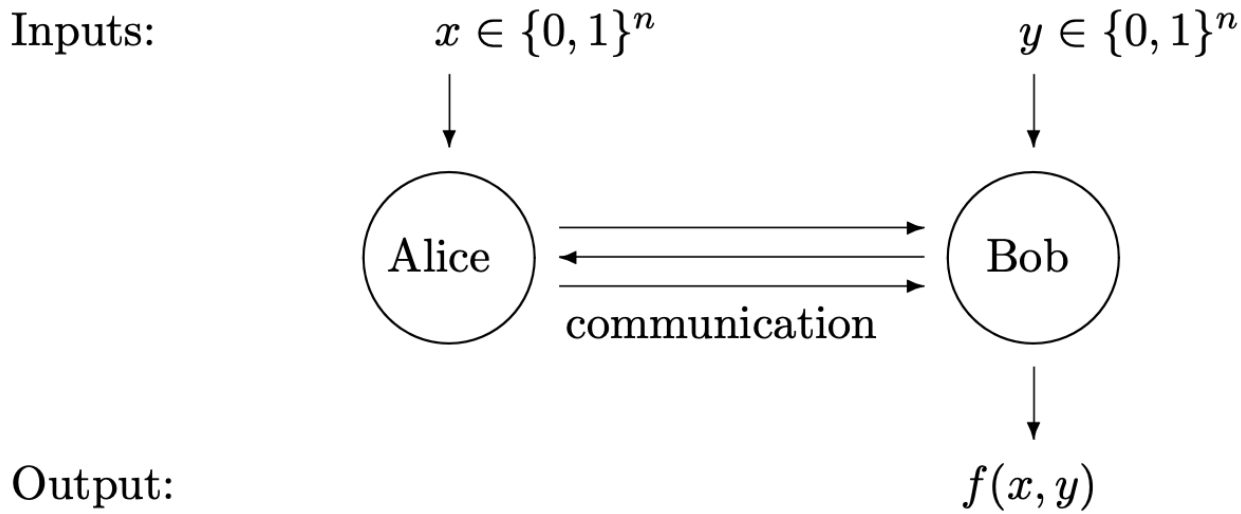
Communication complexity: classical case

Let f be a function, say from $\{0,1\}^n \times \{0,1\}^n$ to $\{0,1\}$. Alice receives $x \in \{0,1\}^n$ and Bob receives $y \in \{0,1\}^n$, they want to compute $f(x,y)$ together.



Communication complexity: classical case

Let f be a function, say from $\{0,1\}^n \times \{0,1\}^n$ to $\{0,1\}$. Alice receives $x \in \{0,1\}^n$ and Bob receives $y \in \{0,1\}^n$, they want to compute $f(x,y)$ together.



The communication complexity is counted by the number of bits used in the communication.

Communication complexity: classical case

Local costs are not considered in communication complexity, and we usually assume Alice and Bob have unlimited computational powers.

Communication complexity: classical case

Local costs are not considered in communication complexity, and we usually assume Alice and Bob have unlimited computational powers.

The communication can be 1-way or 2-way:

1. 1-way: Only Alice can send information to Bob, or the other way around.
2. 2-way: Alice and Bob can send information to each other.

Examples

Disjointness problem:

$$f(x,y) = \begin{cases} 1 & x_i = y_i = 1 \text{ for some } i, \\ 0 & \text{otherwise.} \end{cases}$$

The naive protocol is optimal, Alice has to send all her bits to Bob. So the cost is $\Theta(n)$.[1]

[1] Razborov, On the distributional complexity of disjointness, 1992

[2] Kremer, Nisan, Ron, On randomized one-round communication complexity, 1999.

Examples

Disjointness problem:

$$f(x,y) = \begin{cases} 1 & x_i = y_i = 1 \text{ for some } i, \\ 0 & \text{otherwise.} \end{cases}$$

The naive protocol is optimal, Alice has to send all her bits to Bob. So the cost is $\Theta(n)$.[1]

Index problem: Alice has $(x_1, \ldots, x_n) \in \{0,1\}^n$ and Bob has an index $j \in \{1, \ldots, n\}$, they want to compute x_j.

- If Bob can send information to Alice, then $O(\log n)$.
- Hard case is 1-way (only Alice can speak): $\Theta(n)$.[2]

[1] Razborov, On the distributional complexity of disjointness, 1992

[2] Kremer, Nisan, Ron, On randomized one-round communication complexity, 1999.

Communication complexity: quantum case

In the quantum case, Alice and Bob can send quantum states to each other. The complexity is counted by the number of qubits used.

[3] Aaronson and Ambainis, Quantum search of spatial regions, FOCS'03.

Communication complexity: quantum case

In the quantum case, Alice and Bob can send quantum states to each other. The complexity is counted by the number of qubits used.

Disjointness problem: We can use Grover's algorithm $O(\sqrt{n}\log n)$. It can be improved to $\Theta(\sqrt{n})$.[3]

[3] Aaronson and Ambainis, Quantum search of spatial regions, FOCS'03.

Communication complexity: quantum case

In the quantum case, Alice and Bob can send quantum states to each other. The complexity is counted by the number of qubits used.

Disjointness problem: We can use Grover's algorithm $O(\sqrt{n}\log n)$. It can be improved to $\Theta(\sqrt{n})$.[3]

- Define $z = x \wedge y$, they need to determine if $z_i = 1$ for some i.
- To apply Grover's algorithm, they need an oracle

$$|i\rangle \mapsto (-1)^{x_i \wedge y_i} |i\rangle.$$

- Alice tags on x_i in an extra qubit, and sends $|i\rangle|x_i\rangle$ to Bob, then Bob can apply the unitary $|i\rangle|x_i\rangle \mapsto (-1)^{x_i \wedge y_i}|i\rangle|x_i\rangle$. He sends the state $(-1)^{x_i \wedge y_i}|i\rangle|x_i\rangle$ back to Alice, who can set the second register to $|0\rangle$.

[3] Aaronson and Ambainis, Quantum search of spatial regions, FOCS'03.

Solving linear regressions: 2-party case

Problem statement.

Alice: $A \in \mathbb{R}^{m \times n}$
Bob: $\mathbf{b} \in \mathbb{R}^m$
Goal: Solve $\arg\min \|A\mathbf{x} - \mathbf{b}\|$

More precisely,

- **in the quantum case: output** $|A^{-1}\mathbf{b}\rangle$
- **in the classical case: output a probability distribution defined by** $|A^{-1}\mathbf{b}\rangle$

Solving linear regressions: 2-party case

Problem statement.

Alice: $A \in \mathbb{R}^{m \times n}$
Bob: $\mathbf{b} \in \mathbb{R}^m$
Goal: Solve $\arg\min \|A\mathbf{x} - \mathbf{b}\|$

More precisely,

- **in the quantum case: output** $|A^{-1}\mathbf{b}\rangle$
- **in the classical case: output a probability distribution defined by** $|A^{-1}\mathbf{b}\rangle$

For simplicity, we assume that the entries of $A, \mathbf{b}$ are specified by $\mathrm{polylog}(mn)$ bits.

Classically, the task of outputting a vector solution was studied by Vempala, Wang, and Woodruff [SODA'20].

Our results

Reminder: Alice has matrix $A \in \mathbb{R}^{m \times n}$, Bob has vector $\mathbf{b} \in \mathbb{R}^m$.

	Alice $\to$ Bob	Bob $\to$ Alice	Alice $\leftrightarrow$ Bob
Q	$\widetilde{O}(\kappa^2 \min(m,n))$ $\Omega(\min(m,n))$	$\widetilde{O}(\kappa^2)$ $\Omega(\kappa^2)$	$\widetilde{O}(\kappa)$ $\Omega(\kappa)$
C	$\widetilde{O}(mn)$ $\Omega(\min(m,n))$	$\widetilde{O}(m)$ $\Omega(\min(m,n))$	$\widetilde{O}(m)$ $\Omega(\min(m,n))$
	at most quadratic	can be exponential	can be exponential

$\to$ 1-way communication; $\leftrightarrow$ 2-way communication.

$\kappa =$ condition number of A.

All the lower bounds hold even if $m = n, \kappa = O(1)$.

Quantum protocols: 1-way from Bob to Alice

- Bob sends the state $|0\rangle|\mathbf{b}\rangle$ to Alice.
- Alice constructs a unitary U such that

$$U = \begin{pmatrix} A^{-1}/\|A^{-1}\| & * \\ * & * \end{pmatrix}.$$

She then applies U to $|0\rangle|\mathbf{b}\rangle$:

$$\frac{1}{\|A^{-1}\|}|0\rangle \otimes A^{-1}|\mathbf{b}\rangle + |1\rangle|G\rangle.$$

The success probability is

$$\frac{\|A^{-1}|\mathbf{b}\rangle\|^2}{\|A^{-1}\|^2} \geq \frac{1}{\kappa^2}.$$

- So Bob has to send $O(\kappa^2)$ copies of $|0\rangle|\mathbf{b}\rangle$ to Alice.

Quantum protocols: 1-way from Alice to Bob

- Alice sends $|A^{-1}\rangle = \frac{1}{\|A^{-1}\|_F} \sum_{i,j} (A^{-1})_{i,j} |i,j\rangle$ to Bob.
- Bob constructs a unitary V such that $V|0\rangle = |\mathbf{b}\rangle$. He then applies $I \otimes V^\dagger$ to $|A^{-1}\rangle$:

$$\frac{1}{\|A^{-1}\|_F} A^{-1}|\mathbf{b}\rangle \otimes |0\rangle + |G'\rangle|1\rangle.$$

The success probability is

$$\frac{\|A^{-1}|\mathbf{b}\rangle\|^2}{\|A^{-1}\|_F^2} \geq \frac{1}{\kappa^2 \min(m,n)}.$$

- So Alice has to send $O(\kappa^2 \min(m,n))$ copies of $|A^{-1}\rangle$ to Bob.

Lower bound analysis

We use the hardness of disjointness problem.

Recall: In this 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 $x_i = y_i = 1$ for some i.

Quantum: $\Theta(n)$ (1-way)[4], $\Theta(\sqrt{n})$ (2-way)[5]
Classical: $\Theta(n)$ (1-, 2-way)

[4] Buhrman and de Wolf, Communication complexity lower bounds by polynomials, 2001

[5] Razborov, Quantum communication complexity of symmetric predicates, 2003.

Lower bound analysis

Alice constructs a diagonal matrix A by

$$A_{ii} = \begin{cases} 1 & x_i = 1, \\ 1/\varepsilon & x_i = 0. \end{cases}$$

Bob constructs a vector $\mathbf{b}$ by

$$b_i = \begin{cases} 1 & y_i = 1, \\ \varepsilon & y_i = 0. \end{cases}$$

So the solution of $A\mathbf{x} = \mathbf{b}$ satisfies

$$(A^{-1}\mathbf{b})_i = \begin{cases} 1 & x_i = y_i = 1, \\ \varepsilon & x_i = 1 \text{ xor } y_i = 1, \\ \varepsilon^2 & x_i = y_i = 0. \end{cases}$$

Lower bound analysis

Let

$$P = \{i : x_i = y_i = 1\}, \quad Q = \{i : x_i = 1 \text{ xor } y_i = 1\},$$

then the quantum state of the solution is

$$\frac{1}{\|A^{-1}\mathbf{b}\|} \left(\sum_{i \in P} |i\rangle + \varepsilon \sum_{i \in Q} |i\rangle + \varepsilon^2 \sum_{i \notin P \cup Q} |i\rangle \right).$$

Lower bound analysis

Let

$$P = \{i : x_i = y_i = 1\}, \quad Q = \{i : x_i = 1 \text{ xor } y_i = 1\},$$

then the quantum state of the solution is

$$\frac{1}{\|A^{-1}\mathbf{b}\|} \left(\sum_{i \in P} |i\rangle + \varepsilon \sum_{i \in Q} |i\rangle + \varepsilon^2 \sum_{i \notin P \cup Q} |i\rangle \right).$$

Choose $\varepsilon = 1/\sqrt{n}$, then

- If $|P| = 1$, the probability of seeing $i \in P$ is as large as a constant. $\rightarrow$ see the same index many times
- If $|P| = 0$, the state is close to a uniform superposition of indices from Q. $\rightarrow$ see different indices uniformly

Multi-party case

There are s players $P_1, \ldots, P_s$, each P_i has a matrix $A_i \in \mathbb{R}^{m_i \times n}$ and a vector $\mathbf{b}_i \in \mathbb{R}^{m_i}$, they want to solve

$$\arg\min_{\mathbf{x}} \quad \|A\mathbf{x} - \mathbf{b}\|,$$

where

$$A = \begin{pmatrix} A_1 \\ \vdots \\ A_s \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_s \end{pmatrix}.$$

Multi-party case

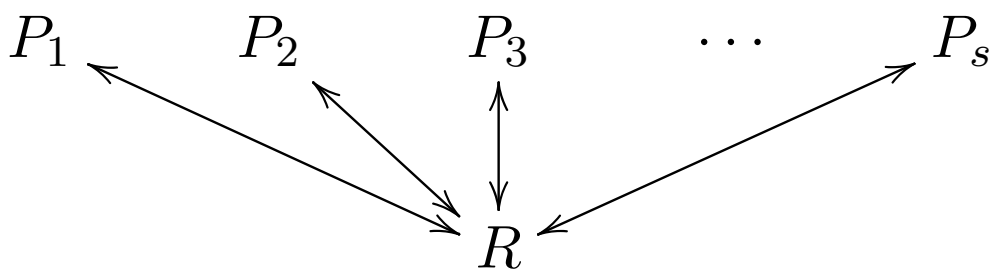
There are s players $P_1, \ldots, P_s$, each P_i has a matrix $A_i \in \mathbb{R}^{m_i \times n}$ and a vector $\mathbf{b}_i \in \mathbb{R}^{m_i}$, they want to solve

$$\arg\min_{\mathbf{x}} \quad \|A\mathbf{x} - \mathbf{b}\|,$$

where

$$A = \begin{pmatrix} A_1 \\ \vdots \\ A_s \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_s \end{pmatrix}.$$

We assume there is a referee R, such that the communication is between each P_i and R.



The model we use

In the simultaneous message passing model, the communication is 1-way, i.e., only P_i can send information to R, R is not allowed to send information to each P_i.

By the result in the Alice-Bob model, $\Omega(n)$ qubits communication are required to prepare the state $|A^{-1}\mathbf{b}\rangle$ in this model.

The model we use

In the simultaneous message passing model, the communication is 1-way, i.e., only P_i can send information to R, R is not allowed to send information to each P_i.

By the result in the Alice-Bob model, $\Omega(n)$ qubits communication are required to prepare the state $|A^{-1}\mathbf{b}\rangle$ in this model.

So similar to the classical coordinator model, we assume that the communication is 2-way between each player and the referee, we call this the quantum coordinator model. The referee outputs

- the quantum state $|A^{-1}\mathbf{b}\rangle$ in the quantum case
- a distribution defined by $|A^{-1}\mathbf{b}\rangle$ in the classical cse

I show you the result first

Quantum	$\widetilde{O}(s^{1.5}\kappa)$	$\Omega(s\kappa)$
Classical	$O(sn^2)$	$\Omega(sn)$

Exponential speedups exists when $s, \kappa \ll n$.

Main technique: Quantum singular value transformation (QSVT).[6]

[6] Gilyén, Su, Low, Wiebe, Quantum singular value transformation, STOC'19

The quantum singular value transformation

Suppose A is Hermitian and there is a unitary

$$U = \begin{pmatrix} A/\alpha & * \\ * & * \end{pmatrix}.$$

Let $f(x)$ be a polynomial bounded by $1/2$ in $[-1, 1]$, then there is a quantum circuit that implements the unitary

$$\widetilde{U} = \begin{pmatrix} f(A/\alpha) & * \\ * & * \end{pmatrix}.$$

It uses $O(\deg(f))$ applications of $U, U^\dagger$ and some one- and two-qubit gates.

An example

Suppose $f(x) \approx \delta/x$ in $[-1,1]\backslash[-\delta,\delta]$, then

$$\widetilde{U} \approx \begin{pmatrix} \alpha\delta A^{-1} & * \\ * & * \end{pmatrix}.$$

Apply $\widetilde{U}$ to $|0\rangle|\mathbf{b}\rangle$, we obtain

$$\alpha\delta|0\rangle \otimes A^{-1}|\mathbf{b}\rangle + |1\rangle|G\rangle.$$

This solves the linear regression problem.

A quantum protocol

To apply QSVT to design an efficient quantum communication protocol for linear regression, the referee needs

- a polynomial $f(x) \approx 1/x$. (easy)
- a unitary

$$U = \begin{pmatrix} A/\alpha & * \\ * & * \end{pmatrix}, \quad \text{where} \quad A = \begin{pmatrix} A_1 \\ \vdots \\ A_s \end{pmatrix}.$$

This is achieved by the technique of linear combination of unitaries.

What is U?

Each P_i constructs a unitary

$$U_i = \begin{pmatrix} A_i/\|A_i\| & * \\ * & * \end{pmatrix}.$$

Then

$$U = \left(\sum_{i=1}^{s} |i\rangle\langle i| \otimes U_i \right) (V \otimes I_2 \otimes I_n) = \begin{pmatrix} a_1 U_1 & * \\ a_2 U_2 & * \\ \cdots & \cdots \\ a_s U_s & * \end{pmatrix}.$$

is a claimed unitary, where $\alpha = \sqrt{\sum_i \|A_i\|^2} = O(\sqrt{s}\|A\|)$, and

$$V|0\rangle = \sum_{i=1}^{s} a_i |i\rangle, \quad a_i = \frac{\|A_i\|}{\alpha}.$$

How to use U?

Thanks to the 2-way communication, the referee can use U with $\widetilde{O}(s)$ qubits communication. For example, to apply U to a state $|\psi\rangle = \sum_{i=1}^{s} |i\rangle|\psi_i\rangle$,

How to use U?

Thanks to the 2-way communication, the referee can use U with $\widetilde{O}(s)$ qubits communication. For example, to apply U to a state $|\psi\rangle = \sum_{i=1}^{s} |i\rangle|\psi_i\rangle$,

1. all P_i send $\|A_i\|$ to the referee, so the referee can find V
2. the referee applies $V \otimes I_2 \otimes I_n$ to $|\psi\rangle$
3. the referee sends the resulting state to P_1 and asks P_1 to apply a controlled U_1 to it if the first register is $|1\rangle$
4. P_1 sends the state back to the referee
5. The referee sends the new state to P_2 and asks P_2 to apply a controlled U_2 to the new state if the first register is $|2\rangle$
6. P_2 sends the state back to the referee
7. .....

How to use U?

Thanks to the 2-way communication, the referee can use U with $\widetilde{O}(s)$ qubits communication. For example, to apply U to a state $|\psi\rangle = \sum_{i=1}^{s} |i\rangle|\psi_i\rangle$,

1. all P_i send $\|A_i\|$ to the referee, so the referee can find V
2. the referee applies $V \otimes I_2 \otimes I_n$ to $|\psi\rangle$
3. the referee sends the resulting state to P_1 and asks P_1 to apply a controlled U_1 to it if the first register is $|1\rangle$
4. P_1 sends the state back to the referee
5. The referee sends the new state to P_2 and asks P_2 to apply a controlled U_2 to the new state if the first register is $|2\rangle$
6. P_2 sends the state back to the referee
7. .....

The quantum state $|\mathbf{b}\rangle \propto \sum_i \|\mathbf{b}_i\| \, |i\rangle|\mathbf{b}_i\rangle$ can be prepared similarly.

The last step

By QSVT, the referee knows the implementation of the unitary

$$\widetilde{U} = \begin{pmatrix} f(A/\alpha) & * \\ * & * \end{pmatrix}.$$

Indeed, the referee also knows how to use $\widetilde{U}$ since

$$\widetilde{U} = W_0 U W_1 U^\dagger W_2 U \cdots W_d U,$$

where $W_0, W_1, \ldots, W_d$ are generated by one- and two-qubit gates depending on the polynomial f and some other public gates, and $d = \deg(f)$.

The final result

Recall: when considering time complexity of using QSVT to solve linear regression, the time complexity is

$$\widetilde{O}((T_U + T_b)\alpha\kappa),$$

where T_U is the cost of implementing U on a quantum computer and T_b is the cost of preparing the quantum state of $\mathbf{b}$.

The final result

Recall: when considering time complexity of using QSVT to solve linear regression, the time complexity is

$$\widetilde{O}((T_U + T_b)\alpha\kappa),$$

where T_U is the cost of implementing U on a quantum computer and T_b is the cost of preparing the quantum state of $\mathbf{b}$.

For communication complexity, we indeed proved that

$$\alpha = O(\sqrt{s}\|A\|), \quad T_U, T_b = \widetilde{O}(s).$$

So the communication complexity is

$$\widetilde{O}((T_U + T_b)\alpha\kappa) = \widetilde{O}(s^{1.5}\kappa).$$

The lower bounds

The lower bounds is proved by using the hardness of multi-player set-disjointness problem:

> P_i **has a subset** $T_i \subseteq [n]$**, and the goal is to determine if** $T_1 \cap T_i \neq \emptyset$ **for some** $i \geq 2$**.**

Quantum: $\Omega(s\sqrt{n})$, Classical: $\Omega(sn)$[7]

[7] Phillips, Verbin, Zhang, Lower bounds for number-in-hand multiparty communication complexity, made easy, 2012.

The lower bounds

Let the diagonal matrix D and vector $\mathbf{b}_j$ be

$$D_j = \begin{cases} 1 & j \in T_1, \\ 1/\varepsilon & j \notin T_1. \end{cases} \qquad \mathbf{b}_i(j) = \begin{cases} 1 & j \in T_i, \\ \eta & j \notin T_i. \end{cases}$$

The hope is to find a linear regression such that the optimal solution is close to

$$D^{-1}(\mathbf{b}_2 + \cdots + \mathbf{b}_s).$$

The lower bounds

Let the diagonal matrix D and vector $\mathbf{b}_j$ be

$$D_j = \begin{cases} 1 & j \in T_1, \\ 1/\varepsilon & j \notin T_1. \end{cases} \quad \mathbf{b}_i(j) = \begin{cases} 1 & j \in T_i, \\ \eta & j \notin T_i. \end{cases}$$

The hope is to find a linear regression such that the optimal solution is close to

$$D^{-1}(\mathbf{b}_2 + \cdots + \mathbf{b}_s).$$

The construction:

$$A = \begin{pmatrix} D \\ \xi I_n \\ \vdots \\ \xi I_n \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} \mathbf{0} \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_s \end{pmatrix}.$$

Conclusions

QSVT can be dequantized $\rightarrow$ no exponential speedup in terms of time and query complexity.

When the communication is 2-way, we can use QSVT $\rightarrow$ efficient quantum communication protocols.

$\Rightarrow$ For many problems where quantum computers lose exponential speedups in terms of time and query complexity, it is possible to have exponential speedups in terms of communication complexity.

$\Rightarrow$ It is interesting to explore more, but the hard part is the lower bound analysis.

Conclusions

QSVT can be dequantized $\rightarrow$ no exponential speedup in terms of time and query complexity.

When the communication is 2-way, we can use QSVT $\rightarrow$ efficient quantum communication protocols.

$\Rightarrow$ For many problems where quantum computers lose exponential speedups in terms of time and query complexity, it is possible to have exponential speedups in terms of communication complexity.

$\Rightarrow$ It is interesting to explore more, but the hard part is the lower bound analysis.

Thank you very much for your attention!