University of Waterloo

Quantum Information Complexity and Communication Complexity

Jason Hu
zs2hu@uwaterloo.ca
December 20, 2018

Abstract

this is a course project report of CS860, Advanced Topics in Algorithms and Complexity, Fall 2018. In this project, we explore the basic concepts in classical and quantum information theory, classical and quantum information complexity, and communication complexity.

Additionally, we particularly focus on the bound of communication complexity by information complexity. Roughly, the randomized communication complexity of a function f has following relation with information complexity: $R(f) = 2^{O(IC(f))}$. On the other hand, the quantum variant is also proved: $QCC(f) = 2^{O(QIC(f))}$. In this report, we try to explain these two proofs and explore the differences.

1 Problem Definition

Informally speaking, communication complexity measures the amount of the communication needed to solve a problem while information complexity measures the amount of information exchange. The question is how these two quantities relate to each other. In [5, 6], the two relations are given respectively, one for the classical case, and the other one for the quantum case, stating the communication is bounded by information exponentially.

$$R(f) = 2^{O(IC(f))}$$
$$QCC(f) = 2^{O(QIC(f))}$$

In this report, we present and make the proof arguments more verbose, and therefore easier to understand. Additionally, despite their similar forms, their proofs employ two different approaches. In the classical case, the proof quite straightforwardly compresses a protocol into a different one that has communication upper bound. On the other hand, in the quantum case, the proof is somewhat indirect, by utilizing the property of generalized discrepancy method, and only relies on protocol compression in an intermediate relation. We will compare these two different proof techniques and attempt to explain why the proof techniques in the classical case are not directly applicable in the quantum case.

Symbols	Meaning
Σ	(default) alphabet
$\mathcal{X}, \mathcal{Y}, \mathcal{Z}$	domains / spaces
X, Y, Z	random variables
x, y, z	elements (from matching spaces, i.e. $x \in \mathcal{X}, y \in \mathcal{Y}$)
p, q, μ	probability distributions / probability vectors
π	communication protocol between two spaces
Π	random variable over a communication protocol π, where the inputs are themselves random variables
ϵ	error

Table 2.1: Convention table

2 Preliminaries: the Classical Case

In this section, we overview a number of basic definitions in order to define the problem.

2.1 Conventions

Before entering the technical discussion, we want to first outline the a number of notational conventions in Table 2.1, in order to make sure the writing is consistent.

2.2 Communication Complexity

Consider two players, Alice and Bob, and their input spaces, $\mathcal{X}$ and $\mathcal{Y}$. Their goal is to compute a function $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$. The way they compute f for a given $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ held by Alice and Bob respectively, is to follow a *communication protocol*, π, which gives instructions to Alice and Bob on when and who to send which bit to the other, so that $f(x, y)$ can be computed. We primarily use $\pi(x, y)$ to denote a *transcript* of $f(x, y)$, which is a record of communication between Alice and Bob, but also overload $\pi(x, y)$ to represent the result of $f(x, y)$ sometimes. The meaning should be distinguishable according to the context. In this model, we assume Alice and Bob have infinite computational power, and the cost is merely the number of bits they have to exchange.

In this report, we exclusively discuss the communication settings with public randomness, or public coin. Under this settings, before the communication begins, there is a publicly available infinitely long random string that Alice and Bob have access to for free. There are models where such random string doesn't exist, or Alice and Bob have their own random string. These models will not be discussed here.

Definition 2.1. *The communication cost of a protocol π, is*

$$CC(\pi) = \max_{x \in \mathcal{X}, y \in \mathcal{Y}} |\pi(x, y)|$$

where $\pi(x, y)$ here means the transcript, and $|\pi(x, y)|$ is the number of bits transmitted according to this transcript.

Definition 2.2. *The distributional communication complexity of $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ with input distribution $\mu : \mathcal{X} \times \mathcal{Y} \to [0, 1]$ and error $\epsilon > 0$, is*

$$D^{\mu}_{\epsilon}(f) = \min_{\pi : P\{\pi(x,y) \neq f(x,y)\} \leq \epsilon} CC(\pi)$$

where π is a deterministic protocol.

Notice that even if it's a deterministic protocol, the protocol can rely on the public randomness. The distributional communication complexity measures the cost of the best deterministic protocol for the worst inputs.

Definition 2.3. *The randomized communication complexity of $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ with public randomness and error $\epsilon > 0$, is*

$$R_\epsilon(f) = \min_{\pi : P\{\pi(x,y) \neq f(x,y)\} \leq \epsilon} CC(\pi)$$

where π is a randomized protocol.

Contrast to deterministic protocol, a randomized protocol can be considered as a result of selecting a deterministic protocol based on some probability distribution. The randomized communication complexity measures the cost of the best randomized protocol for the worst inputs.

Clearly, $R_\epsilon(f) \geq D^\mu_\epsilon(f)$. This is because in the randomized case, the communication cost cannot get smaller than deterministically optimal solution, by averaging with other sub-optimal protocols. These two definitions are further related in the following way.

Theorem 2.1. *[20, Yao's min-max]*

$$R_\epsilon(f) = \max_\mu D^\mu_\epsilon(f)$$

This theorem states that in the worst distribution, selecting the best deterministic protocol essentially works the same as running an optimal randomized protocol.

2.3 Information Theory

In 1948, C.E. Shannon published a celebrated paper [13], which describes a well-established mathematical theory on data transmission over channels, and defines fundamental concepts in today's information theory.

Definition 2.4. *The Shannon entropy, or entropy, of a nonnegative real vector $v \in [0,\infty)^\Sigma$ is*

$$H(v) = -\sum_{x \in \Sigma} v(x) \log v(x)$$

Throughout, we assume the base of logarithm is always 2, so we omit it by default. A special case is when v is a probability vector of a random variable X. An interpretation of this formula is that entropy is the expectation of number of bits required to record one sample of the random variable X. Due to this statistical nature, Shannon entropy defines the amount of information of a random variable in an amortized sense.

Following are definitions of a few more (classical) information theoretic concepts.

Definition 2.5. *Conditional entropy of X given Y is*

$$H(X|Y) = H(XY) - H(Y)$$

where $H(XY)$ is joint entropy of X and Y.

In particular, if X and Y are unrelated, then $H(X|Y) = H(X)$, since their joint entropy $H(XY) = H(X) + H(Y)$. Clearly, $H(X|Y) \leq H(X)$ in general, because putting condition cannot generate more information.

Definition 2.6. *Mutual information of X and Y is*

$$I(X;Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(XY)$$

Mutual information measures the amount of information shared between X and Y. If X and Y are unrelated, then $I(X;Y) = 0$; if X and Y are identical, then $I(X;Y) = H(X) = H(Y)$.

Definition 2.7. *Divergence / relative entropy of two nonnegative real vectors p and q is*

$$D(p \parallel q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)}$$

where $supp(p) \subseteq supp(q)$.

Divergence measures how far away q is from p.

It's critical to realize that information theory measures one-way data transmission in an amortized sense, so it's an approximation, and converges as more and more data is transmitted. This can be seen in the following theorem of Shannon entropy.

Theorem 2.2. *(Shannon's source coding theorem [13, Theorem 4])*

$$\lim_{n \to \infty} \frac{C_n(X)}{n} = H(X)$$

where $C_n(X)$ means the number of bits used to transmit n (independent) samples of X.

Theorem 2.2 is probably the most important theorem which turns information theory into highly practical use. It states a way to actually approximate Shannon entropy in a manner similar to the law of large numbers and hence assigns an operational meaning to Shannon entropy. Here are some interpretations of this theorem.

1. X can be transmitted losslessly provided a channel with at least $H(X)$ bit bandwidth.
2. A compression of X so that the result requires less than $H(X)$ bits to describe the information must be lossy.

2.4 Information Complexity

In [3, 5], Braverman defines the interactive information complexity. Compared to information theory, which models one-way communication, information complexity theory models the amount of information exchange between Alice and Bob, when they execute a communication protocol, π. In order to define communication complexity, we first need to define communication cost.

Definition 2.8. *The (internal) information cost of a protocol π over $\mathcal{X} \times \mathcal{Y}$ is*

$$IC_\mu(\pi) = I(\Pi; X|Y) + I(\Pi; Y|X)$$

where μ is a distribution on the inputs X and Y.

This definition is explicitly introduced in [2]. Another common notation for the same concept is $IC^i_\mu(\pi)$, in which the superscript i denotes the information cost is *internal*. Prior to [2], information cost referred to a similar concept, which now is called *external* information cost.

Definition 2.9. *The external information cost of a protocol π over $\mathcal{X} \times \mathcal{Y}$ is*

$$IC^{ext}_\mu(\pi) = I(XY; \Pi)$$

Another common notation is $IC^o_\mu(\pi)$.

The information cost of a protocol π measures how much additional information Alice and Bob can learn about the input of the other in total, by knowing the communication protocol. On the other hand, the external information cost of a protocol π measures how much the communication

protocol is related to both inputs held by Alice and Bob in terms of the amount of information. This measure is *external*, because it captures the mutual information as an external observer, to whom all of the inputs and the protocol are visible.

Since [2], the information cost has become more interesting concept than the external information cost, due to a property similar to Theorem 2.2. Before stating the concrete theorem, we first need to define information complexity.

Definition 2.10. *The information complexity of a function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ with error ϵ and probability distribution of inputs μ is*

$$IC_\mu(f,\epsilon) = \inf_{\pi : P\{\pi(x,y) \neq f(x,y)\} \leq \epsilon} IC_\mu(\pi)$$

In this definition, the *infimum* is taken over all communication protocols, which can compute f with no more than ϵ error. This infimum is necessary, because there are problems the optimal information complexities of which are not reachable, but nonetheless can be approximated by (infinitely long) sequences of communication protocols. As protocols appear later in the sequences, the information costs converge to $IC_\mu(f,\epsilon)$ but can never reach it. The infimum in the definition is handled situations of this kind. Details can be found in [19, 2.3.4, Theorem 2.3.2], which shows the AND function is one of the examples.

At this point, the information complexity still depends on a prior distribution μ of inputs. The worst such distribution defines the information complexity of the function.

Definition 2.11. *The information complexity of a function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ with error ϵ is*

$$IC(f,\epsilon) = \inf_{\pi : P\{\pi(x,y) \neq f(x,y)\} \leq \epsilon} \max_\mu IC_\mu(\pi)$$

If we simply factor out μ from $IC_\mu(f,\epsilon)$, we can obtain a closely related concept.

Definition 2.12. *The max-distributional information complexity of a function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ with error ϵ is*

$$IC_D(f,\epsilon) = \max_\mu IC_\mu(f,\epsilon) = \max_\mu \inf_{\pi : P\{\pi(x,y) \neq f(x,y)\} \leq \epsilon} IC_\mu(\pi)$$

IC is the inf-max of the information cost and IC_D is the max-inf of the information cost. They are closely related, but not necessarily equal. They are related by the following two theorems.

Theorem 2.3. *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ and $\epsilon \geq 0$, then for $\alpha \in (0,1)$,*

$$IC(f, \frac{\epsilon}{\alpha}) \leq \frac{IC_D(f,\epsilon)}{1-\alpha}$$

Theorem 2.4. *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$,*

$$IC(f,0) = IC_D(f,0)$$

The following theorem captures a desired property we want to have for information complexity.

Theorem 2.5. *(additivity) For $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$,*

$$IC_D(f^n,\epsilon) = n \cdot IC_D(f,\epsilon)$$
$$IC(f^n,\epsilon) = n \cdot IC(f,\epsilon)$$

where f^n is a function which computes f for n (independent) pairs of inputs.

This theorem captures the intuition that computing the same function on n inputs one by one should cost the same as computing at the same time. This draws the attention to amortized cost of communications. The next theorem addresses this and resembles Theorem 2.2, which makes the (internal) information complexity a more interesting definition than external information complexity.

Definition 2.13. *$D_\epsilon^{\mu,n}(f^n)$ is the distributional communication complexity of computing $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ on n (independent) pairs of inputs drawn from distribution μ, each with no more than $\epsilon \in (0,1)$ error.*

We can have a similar definition for the randomized case.

Definition 2.14. *$R_\epsilon^n(f^n)$ is the randomized communication complexity of computing $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ on n (independent) pairs of inputs, each with no more than $\epsilon \in (0,1)$ error.*

The following theorem connects information complexity with the amortized cost of repeating the same function.

Theorem 2.6.

$$IC_\mu(f,\epsilon) = \lim_{n\to\infty} \frac{D_\epsilon^{\mu,n}(f^n)}{n}$$
$$IC(f,\epsilon) = \lim_{n\to\infty} \frac{R_\epsilon^n(f^n)}{n}, \epsilon > 0$$

Finally, following theorem connects information complexity and communication complexity.

Theorem 2.7. *[5, Theorem 5.3] For $\rho \in (0, 1/16)$,*

$$D^\mu_{\rho+\epsilon}(f) = 2^{O(1/\rho + IC_\mu(f,\epsilon)/\rho^2)}$$
$$R_{\rho+\epsilon}(f) = 2^{O(1/\rho + IC(f,\epsilon)/\rho^2)}$$

These two bounds are rather weak, stating that communication complexity and information complexity are related exponentially. Following theorem shows that for some problems, this relation is in fact tight.

Theorem 2.8. *[9, 8, Theorem 1, Theorem 2]*
There is a function that separate communication complexity and information complexity exponentially.

3 Proof of Theorem 2.7

3.1 Proof Idea

In this section, we review the proof of Theorem 2.7. We aim at reviewing the proof in greater details in order to make the proof more understandable for audience that are less familiar with the background. Before expanding the proof, the proof can be roughly divided into following steps.

1. Begin with a protocol, π, so that its information complexity $\approx IC(f,\epsilon)$.
2. Discover another protocol, π', which approximates π with a little extra error (compression), and has a bound of communication complexity exponential to information complexity.
3. Since the error between π and $IC(f,\epsilon)$ can be arbitrarily small, the theorem is concluded.

The construction of the π' is somewhat convoluted, and is the core step of the proof. The actual construction is shown in Lemma 3.6.

3.2 Useful Theorems and Lemmas

Following are a number of lemmas that will be used in the proofs.

Lemma 3.1. *For all p, q nonnegative real vectors,*

$$D(p \parallel q) \geq 0$$

Theorem 3.2. *(Markov's inequality)*

$$P(X \geq a) \leq \frac{E(X)}{a}$$
$$P(X \leq a) \geq 1 - \frac{E(X)}{a}$$

We need a specialized Chernoff bound for the sum of independent variables obeying Bernoulli distribution.

Theorem 3.3. *[10] (Chernoff bound) Let X_i is random variable obeying Bernoulli distribution with probability p to take value 1. Let $X = \sum_i X_i$, then*

$$Pr\{X \geq (1+\delta)E(X)\} \leq e^{-\frac{\delta^2}{2+\delta}E(X)}, \delta > 0$$

Lemma 3.4. *(Conditional mutual information as divergence)*

$$I(X;Y|Z) = E_{XZ}(D((Y|XZ) \parallel (Y|Z)))$$

Proof. Let p be the joint distribution of all X, Y and Z. We denote the marginal distributions using subscripts, e.g. $p_X(x) = \sum_{y\in\mathcal{Y}, z\in\mathcal{Z}} p(x,y,z)$.

$$\begin{aligned}
I(X;Y|Z) =& H(X|Z) - H(X|YZ) \\
=& -\sum_{z\in\mathcal{Z}} p_Z(z) \sum_{x\in\mathcal{X}} p_{X|Z}(x|z) \log p_{X|Z}(x|z) \\
&+ \sum_{y\in\mathcal{Y}, z\in\mathcal{Z}} p_{YZ}(y,z) \sum_{x\in\mathcal{X}} p_{X|YZ}(x|y,z) \log p_{X|YZ}(x|y,z) \\
=& -\sum_{x,y,z} p(x,y,z) \log \frac{p_{XZ}(x,z)p_{YZ}(y,z)}{p_Z(z)p(x,y,z)} \\
=& \sum_{x,y,z} p(x,y,z) \log \frac{p_Z(z)p(x,y,z)}{p_{XZ}(x,z)p_{YZ}(y,z)} \\
=& \sum_{x,z} p_{XZ}(x,z) \sum_y p_{Y|XZ}(y|x,z) \log \frac{p_{Y|XZ}(y|x,z)}{p_{Y|Z}(y|z)} \\
=& \sum_{x,z} p_{XZ}(x,z) D(p_{Y|XZ} \parallel p_{Y|Z}) \\
=& E_{XZ}(D((Y|XZ) \parallel (Y|Z)))
\end{aligned}$$

□

Corollary 3.4.1. *(Information cost as divergence) For all $\mu : \mathcal{X} \times \mathcal{Y} \to [0,1]$ a distribution and π a protocol,*

$$IC_\mu(\pi) = E_{(x,y)\sim\mu}(D(\pi(x,y) \parallel \pi(X,y)) + D(\pi(x,y) \parallel \Pi(x,Y)))$$

Proof. This lemma is a specialization of Lemma 3.4.

$$\begin{aligned} IC_\mu(\pi) =& I(X;\Pi|Y) + I(Y;\Pi|X) \\ =& E_{XY}(D((\Pi|XY) \parallel (\Pi|Y)) + D((\Pi|XY) \parallel (\Pi|X))) \end{aligned}$$

The conclusion is obtained by expanding the definition. □

To state further lemmas, we first want to explicitly define *total variation* of two probability distributions.

Definition 3.1. [1] *The total variation of $p, q : \mathcal{X} \to [0,1]$, two distributions, denoted by $|p-q|$, is*

$$|p-q| = \frac{1}{2}\sum_x |p(x) - q(x)|$$

The following helper lemma is shown in [5, 19, 3].

Lemma 3.5. *For p, q distributions, so that $D(p \parallel q) \leq I$ for some I. For any ϵ,*

$$Pr\{x : 2^{(I+1)/\epsilon} q(x) < p(x)\} < \epsilon$$

Proof. Consider two sets, $\mathcal{A} = \{x : p(x) < q(x)\}$ and $\mathcal{B} = \{x : 2^{(I+1)/\epsilon} q(x) < p(x)\}$. Clearly, $A \cap B = \emptyset$. Note that only $x \in \mathcal{A}$ contribute negative values to $D(p \parallel q)$.

Therefore,

$$\begin{aligned} D(p \parallel q) &= \sum_x p(x) \log \frac{p(x)}{q(x)} \\ &\geq \sum_{x\in\mathcal{A}} p(x) \log \frac{p(x)}{q(x)} + \sum_{x\in\mathcal{B}} p(x) \log \frac{p(x)}{q(x)} \qquad (1) \\ &> \sum_{x\in\mathcal{A}} p(x) \log \frac{p(x)}{q(x)} + \sum_{x\in\mathcal{B}} p(x) \frac{I+1}{\epsilon} \qquad (2) \end{aligned}$$

ineq. (2) holds, because all negative terms have been captured by $\mathcal{A}$, so whatever dropped have to be positive. ineq. (2) is due to $\mathcal{B}$'s definition, which has $p(x)/q(x) > 2^{(I+1)/\epsilon}$.

For the first half, we have $p(x)/q(x) \in (0,1)$ (recall that $supp(p) \subseteq supp(q)$, so $p(x)/q(x) \neq 0$, for x under consideration). Consider $f(x) = x \log x$, we know $x \in (0,1), f(x) \in (-1, 0)$. Therefore

$$\begin{aligned} \sum_{x\in\mathcal{A}} p(x) \log \frac{p(x)}{q(x)} &= \sum_{x\in\mathcal{A}} q(x) \frac{p(x)}{q(x)} \log \frac{p(x)}{q(x)} \\ &> \sum_{x\in\mathcal{A}} q(x) \cdot (-1) \\ &\geq -1 \qquad \text{because } q \text{ is a distribution} \end{aligned}$$

So overall,

$$\begin{aligned} I \geq D(p \parallel q) &> -1 + \sum_{x\in\mathcal{B}} p(x) \frac{I+1}{\epsilon} \\ &\geq -1 + Pr\{\mathcal{B}\} \frac{I+1}{\epsilon} \end{aligned}$$

By transforming the inequality, we have $Pr\{\mathcal{B}\} < \epsilon$. □

[1] The version presented here is normalized. There is an unnormalized version which does not have $\frac{1}{2}$ factor.

3.3 Main Proof

The following big auxiliary lemma is used in the proof of Theorem 2.7, which construct a protocol π' statistically closely related to $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, and the task is then used to calculate and bound the randomized communication complexity.

Essentially, the proof idea is to consider a statistically equivalently difficult problem with public randomness, and a protocol that approximates the answer with some extra error. The sampling here means Alice and Bob only "sample" a subset from the public random string ($\mathcal{A}$ and $\mathcal{B}$ in the proof), but the sample sets are well engineered to make sure the goal is achieved.

Lemma 3.6. *(Sampling) Let $\mu, p_A, p_B : \mathcal{X} \to [0,1]$ distributions, $I \geq 0$, error $\epsilon \in (0, 1/16)$, so that $D(\mu \parallel p_A) \leq I, D(\mu \parallel p_B) \leq I$.*

Let four other real functions $s_A, s_B, u_A, u_B : \mathcal{X} \to [0,1]$. s_A and u_A are given to Alice, and s_B and u_B are given to Bob. They satisfy

$$\begin{aligned} \mu(x) &= s_A(x)s_B(x) \\ p_A(x) &= s_A(x)u_A(x) \\ p_B(x) &= s_B(x)u_B(x) \end{aligned}$$

There exists a sampling protocol π' with public randomness achieves the following with $2^{O((I+1)/\epsilon)}$ bits of communication.

1. *Alice and Bob output $x_A, x_B \in \mathcal{X}$ respectively.*
2. *there is an event $\mathcal{E}$ so that $\neg\mathcal{E} \Rightarrow x_A = x_B$ and $Pr\{\mathcal{E}\} < \epsilon$.*
3. *Let $\mu' = Pr\{x_A \in \mathcal{X} : \neg\mathcal{E}\}$, then the total variation $|\mu - \mu'| < \frac{\epsilon}{2}$.*

Proof. From the public randomness, Alice and Bob draw $(x_i, \alpha_i, \beta_i) \in \mathcal{X} \times [0,1] \times [0,1]$ uniformly. Here the subscript $i \in [1, T]$ is the index of the instance drawn from the public randomness, where $T = 2|\mathcal{X}| \ln \frac{1}{\epsilon}$. Note that i has upper bound T, so Alice and Bob stop consider any further instances at that point. On the other hand, they try to find the first index satisfying

$$\alpha_i \leq s_A(x_i), \beta_i \leq s_B(x_i)$$

To compute the chance for an $x \in \mathcal{X}$ which can satisfy this requirement,

$$\begin{aligned} Pr\{\alpha \leq s_A(x), \beta \leq s_B(x)\} &= Pr\{\alpha \leq s_A(x)\} \times Pr\{\beta \leq s_B(x)\} && (1) \\ &= s_A(x)s_B(x) && (2) \\ &= \mu(x) \end{aligned}$$

eq. (1) holds because (x_i, α_i, β_i) is drawn uniformly and therefore α and β are independent. eq. (2) holds is again because α and β are drawn uniformly, so the chance for $\alpha \leq s_A(x)$ is just $s_A(x)$. β works in the same way.

Therefore, for any draw, indicated by the index i, the chance for it to succeed is

$$\begin{aligned} Pr\{\alpha_i \leq s_A(x_i), \beta_i \leq s_B(x_i)\} &= \sum_{x \in \mathcal{X}} Pr\{x \in \mathcal{X}\} Pr\{\alpha \leq s_A(x), \beta \leq s_B(x)\} \\ &= \sum_{x \in \mathcal{X}} \frac{1}{|\mathcal{X}|} \mu(x) = \frac{1}{|\mathcal{X}|} \end{aligned}$$

Let τ be the first such index Alice and Bob encounter. Clearly, $\tau \leq T$. If we generalize τ's range, we can conveniently represent the case where no such τ is found as $\tau > T$. Namely, Alice

and Bob fail.

$$\begin{aligned} Pr\{\tau > T\} &= (1 - \frac{1}{|\mathcal{X}|})^{2|\mathcal{X}| \ln 1/\epsilon} \\ &< e^{2 \ln \epsilon} = \epsilon^2 < \frac{\epsilon}{16} \end{aligned}$$

Next, we will need to discover event $\mathcal{E}$ in the conclusion and give its probabilistic bound. This part is quite vague in [5], so we are trying to make it clearer here. We let $\mathcal{A}$ and $\mathcal{B}$ two sets of indices

$$\begin{aligned} \mathcal{A} &= \{i \leq T : \alpha_i \leq s_A(x_i), \beta_i \leq 2^{8(I+1)/\epsilon} u_A(x_i)\} \\ \mathcal{B} &= \{i \leq T : \beta_i \leq s_B(x_i), \alpha_i \leq 2^{8(I+1)/\epsilon} u_B(x_i)\} \end{aligned}$$

And there are two subsets of $\mathcal{X}$:

$$\begin{aligned} \mathcal{F}_A &= \{x : 2^{8(I+1)/\epsilon} p_A(x) < \mu(x)\} \\ \mathcal{F}_B &= \{x : 2^{8(I+1)/\epsilon} p_B(x) < \mu(x)\} \end{aligned}$$

Then if $x_\tau \notin \mathcal{F}_A$, we have

$$\begin{aligned} 2^{8(I+1)/\epsilon} p_A(x_\tau) &\geq \mu(x_\tau) \\ 2^{8(I+1)/\epsilon} s_A(x_\tau) u_A(x_\tau) &\geq s_A(x_\tau) s_B(x_\tau) \\ 2^{8(I+1)/\epsilon} u_A(x_\tau) &\geq s_B(x_\tau) \end{aligned}$$

Since τ satisfies the requirements, we have $\beta_\tau \leq s_B(x_\tau) \leq 2^{8(I+1)/\epsilon} u_A(x_\tau)$. Together with $\alpha_\tau \leq s_A(x_\tau)$, we know $\tau \in \mathcal{A}$. Similarly, $\tau \in \mathcal{B}$. So $x_\tau \notin \mathcal{F}_A \cup \mathcal{F}_B \Rightarrow \tau \in \mathcal{A} \cap \mathcal{B}$.

Due to an argument similar to a previous one, from the uniformity of the distribution, we know

$$\begin{aligned} Pr\{i \in \mathcal{A}\} &= Pr\{\alpha_i \leq s_A(x_i), \beta_i \leq 2^{8(I+1)/\epsilon} u_A(x_i)\} \\ &= Pr\{\alpha_i \leq s_A(x_i)\} \cdot Pr\{\beta_i \leq 2^{8(I+1)/\epsilon} u_A(x_i)\} \\ &= \sum_i \frac{1}{|\mathcal{X}|} s_A(x_i) 2^{8(I+1)/\epsilon} u_A(x_i) \\ &= \frac{2^{8(I+1)/\epsilon}}{|\mathcal{X}|} \end{aligned}$$

Notice that, $i \in \mathcal{A}$ obeys Bernoulli distribution. Therefore, we can see for each $i \in [1, T]$ as an independent sampling, with $Pr\{i \in \mathcal{A}\}$ probability to get 1, or otherwise 0. We can obtain that

$$E(|\mathcal{A}|) = \sum_i Pr\{i \in \mathcal{A}\} = TPr\{i \in \mathcal{A}\} = 2^{8(I+1)/\epsilon} \cdot 2 \ln \frac{1}{\epsilon}$$

Apply Theorem 3.3, with $\delta = 1$,

$$\begin{aligned} Pr\{|\mathcal{A}| \geq 2E(|\mathcal{A}|)\} &\leq e^{-E(|A|)/3} \\ &= e^{2^{8(I+1)/\epsilon} \cdot 2 \ln \epsilon/3} \\ &< \epsilon^{2^{8/\epsilon} \cdot 2/3} \\ &< \epsilon^2 < \frac{\epsilon}{16} \end{aligned} \tag{3}$$

ineq. (3) holds because $I \geq 0$, and we know that $2^{8/\epsilon} > 3$. In [5], it's unclear which form of Chernoff bound is used and the paper seems overshoot by a lot for some reason, so it's worth to make it explicit here.

On the other hand, $2E(|\mathcal{A}|) = 2^{8(I+1)/\epsilon} \cdot 4 \ln \frac{1}{\epsilon} < 2^{9(I+1)/\epsilon}$ is clear by comparing $2^{1/\epsilon}$ with $4 \ln 1/\epsilon$. So $Pr\{|\mathcal{A}| \geq 2^{9(I+1)/\epsilon}\} < \frac{\epsilon}{16}$ holds. Similarly, $Pr\{|\mathcal{B}| \geq 2^{9(I+1)/\epsilon}\} < \frac{\epsilon}{16}$ holds.

Now consider $\mathcal{F}_A$ and $\mathcal{F}_B$. Applying Lemma 3.5, we obtain $\mu(\mathcal{F}_A) < \epsilon/8$, $\mu(\mathcal{F}_B) < \epsilon/8$. Therefore

$$Pr\{x_\tau \in \mathcal{F}_A \cup \mathcal{F}_B\} < 2\frac{\epsilon}{8} = \frac{\epsilon}{4}$$

Consider the following event, $\mathcal{E}' = \{x_\tau \in \mathcal{F}_A \cup \mathcal{F}_B \vee \tau > T \vee |\mathcal{A}| > 2^{9(I+1)/\epsilon} \vee |\mathcal{B}| > 2^{9(I+1)/\epsilon}\}$. Its probability

$$\begin{aligned} Pr\{\mathcal{E}'\} &\leq Pr\{x_\tau \in \mathcal{F}_A \cup \mathcal{F}_B\} + Pr\{\tau > T\} + Pr\{|\mathcal{A}| > 2^{9(I+1)/\epsilon}\} + Pr\{|\mathcal{B}| > 2^{9(I+1)/\epsilon}\} \\ &= \frac{\epsilon}{4} + \frac{3\epsilon}{16} < \frac{\epsilon}{2} \end{aligned}$$

Looking at $\mathcal{E}'$, $\neg\mathcal{E}'$ requires $\tau \leq T$, so Alice and Bob should discover τ. Note that, the sizes of $\mathcal{A}$ and $\mathcal{B}$ have been restricted to $2^{O(I+1)/\epsilon}$, which gives an upper bound on the number of bits to communicate. Note that to completely state $\mathcal{E}$, we still need another small disjunction in addition to $\mathcal{E}'$. We will only be able to state this disjunction after we describe the protocol. However, $\mathcal{E}'$ has covered most of the characteristics that we can use to analyze $\mathcal{E}$.

Since $\neg\mathcal{E}'$ only restricts sampling space in $\mathcal{X}/\mathcal{F}_A \cup \mathcal{F}_B$. We then know how to adjust probability and compute the total variation between μ and μ'

$$\begin{aligned} |\mu - \mu'| &= \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \mu'(x)| \\ &= \frac{1}{2} \sum_{x \in \mathcal{F}_A \cup \mathcal{F}_B} |\mu(x) - 0| + \frac{1}{2} \sum_{x \notin \mathcal{F}_A \cup \mathcal{F}_B} |\mu(x) - \mu'(x)| \\ &< \frac{\epsilon}{8} + \frac{1}{2} \sum_{x \notin \mathcal{F}_A \cup \mathcal{F}_B} |\mu(x) - \frac{\mu(x)}{1 - \epsilon/4}| \\ &< \frac{\epsilon}{8} + \frac{\epsilon}{8 - 2\epsilon} < \frac{\epsilon}{2} \end{aligned}$$

So condition 3 is satisfied.

The actual protocol is given in [5], as follows.

1. Alice and Bob computes $\mathcal{A}$ and $\mathcal{B}$ respectively, and abort if their sizes go beyond $2^{9(I+1)/\epsilon}$.
2. Alice computes $d = \lceil 18\frac{I+1}{\epsilon} + \log 1/\epsilon + 2 \rceil$ hash bits for each $x_i, i \in \mathcal{A}$, and send them all to Bob.
3. Bob computes the hashes as well and find the first index τ, and send it to Alice. Bob output x_B.
4. Alice outputs x_A according to τ.

Given I and ϵ, d is linearly bounded. The communication complexity from Alice is bound by $|\mathcal{A}| \cdot d = O(|\mathcal{A}|)$. Bob's communication is ignorable. The chance for hash clashes for two different value is

$$2^{-d} < 2^{-(18(I+1)/\epsilon + \log 1/\epsilon + 2)} \leq \frac{\epsilon}{4|\mathcal{A}||\mathcal{B}|}$$

So the overall hash clashes chances are

$$|\mathcal{A}||\mathcal{B}|2^{-d} < \frac{\epsilon}{4}$$

Recall that $\mathcal{E}'$ misses a disjunction to define $\mathcal{E}$. Now, what we need to do is to simply make no hash clashes as this disjunction, so $\neg\mathcal{E}$ can guarantee hash code equality implies element equality. We add all the probability up

$$Pr\{\mathcal{E}\} < \frac{\epsilon}{2} + \frac{\epsilon}{4} < \epsilon$$

So the probability for this protocol to fail is bounded. Together with the inequality above, condition 1 and 2 are satisfied. □

The proof technique used here is to use public randomness to compress the original protocol to a protocol that has communication bound. This technique is best explained in [2]. Having defined the sampling lemma Lemma 3.6, we can now proceed to the proof of Theorem 2.7.

Proof of Theorem 2.7. Recall that $IC_\mu(f,\epsilon) = \inf_{\pi:P\{\pi(x,y)\neq f(x,y)\}\leq\epsilon} IC_\mu(\pi)$, so for some fixed constant, $\delta > 0$, we now there must exist a protocol π, so that $I_\mu = IC_\mu(\pi) = IC_\mu(f,\epsilon) + \delta$. We have

$$\begin{aligned} I_\mu &= IC_\mu(f,\epsilon) + \delta \\ &= IC_\mu(\pi) \\ &= E_{(x,y)\sim\mu}(D(\pi(x,y) \parallel \pi(X,y)) + D(\pi(x,y) \parallel \Pi(x,Y))) && \text{(Corollary 3.4.1)} \end{aligned}$$

By Lemma 3.1 and Theorem 3.2, we have

$$\begin{aligned} &Pr\{D(\pi(x,y) \parallel \pi(X,y)) + D(\pi(x,y) \parallel \Pi(x,Y)) \leq \frac{2I_\mu}{\rho}\} \\ &\geq 1 - \frac{E(D(\pi(x,y) \parallel \pi(X,y)) + D(\pi(x,y) \parallel \Pi(x,Y)))}{2I\mu/\rho} && \text{(Theorem 3.2)} \\ &= 1 - \frac{I_\mu}{2I_\mu/\rho} = 1 - \frac{\rho}{2} \end{aligned}$$

Note that divergence is nonnegative, so both $D(\pi(x,y) \parallel \pi(x,Y)) \leq \frac{2I_\mu}{\rho}$ and $D(\pi(x,y) \parallel \pi(x,Y)) \leq \frac{2I_\mu}{\rho}$ with confidence of $1 - \rho/2$.

Now consider π as a protocol tree. Due to randomness on inputs, each node in π as a tree must have some Bernoulli distribution on which bit to send next. An example tree is shown in Figure 3.1. Notice that, in a protocol tree, at each node, its probability to go left or right is automatically conditioned on all the bits sent before it, and therefore chain rule for probabilities applies.

For example, consider a node, v, it's probability to be reached is the product of all the probabilities along the path from the root to it, namely,

$$\mu(v) = \prod_{t\in\{\text{path from root to } v\}} p_{\{0,1\}}(t)$$

where $p_{\{0,1\}}(u)$ means $p_0(u)$ or $p_1(u)$ depending on the actual bit being sent. Again, this is the consequence of the chain rule.

Following this line, we can define following functions, acting on each leaf v in the protocol tree

$$\begin{aligned}
s_A(v) &= \prod_{t\in\{\text{path from root to } v\}\wedge t \text{ is Alice's nodes}} p_{\{0,1\}}(t) \\
s_B(v) &= \prod_{t\in\{\text{path from root to } v\}\wedge t \text{ is Bob's nodes}} p_{\{0,1\}}(t) \\
u_A(v)_{|x} &= \prod_{t\in\{\text{path from root to } v\}\wedge t \text{ is Bob's nodes}} p_{\{0,1\}}(t)_{|x} \\
u_B(v)_{|y} &= \prod_{t\in\{\text{path from root to } v\}\wedge t \text{ is Alice's nodes}} p_{\{0,1\}}(t)_{|y}
\end{aligned}$$

where $u_A(v)_{|x}$ measures the product of all probabilities of Bob's nodes, given Alice holds x. The same holds for $u_B(v)_{|y}$.

Since the nodes in the tree are either Alice's or Bob's. We can easily see $\mu(v) = s_A(v)s_B(v)$. On the other hand, given any x held by Alice, the only thing varies is y held by Bob. In that case, in each Bob's node, the probability of sending 1 is just the expectation for him to send 1 over all possible y's, and this is what Alice can see. Therefore, given any x, $p_A(v)_{|x} = s_A(v)u_A(v)$, where $p_A(v)_{|x}$ is the probability to each v given x in Alice. Similarly, we have $p_B(v)_{|y} = s_B(v)u_b(v)$.

Seeing μ, p_A, p_B, and their relation with s_A, s_B, u_A and u_B, and considering Lemma 3.6, we can let $I = \frac{2I_\mu}{\rho}$, and error ρ. From Lemma 3.6, we obtain another protocol π', which is $\rho/2$ away from π in total variation from condition 3, and it's communication complexity is

$$\begin{aligned}
2^{O((1+2I_\mu/\rho)/\rho)} &= 2^{O(1/\rho+I_\mu/\rho^2)} \\
&= 2^{O(1/\rho+(IC_\mu(f,\epsilon)+\delta)/\rho^2)}
\end{aligned}$$

We can ignore δ because it can be arbitrarily small.

Recall that all this is under the assumption of $D(\pi(x,y) \parallel \pi(x,Y)) \leq 2I_\mu/\rho$ and $D(\pi(x,y) \parallel \pi(X,y)) \leq 2I_\mu/\rho$. The confidence for this is $1-\rho/2$. Let such x, y be good. π' only works with good x, y. To show π''s error bound, we have

$$\begin{aligned}
Pr\{\pi'(x,y) \neq f(x,y)\} \leq & Pr\{(x,y) \text{ not good}\} + Pr\{\pi'(x,y) \neq \pi(x,y) | (x,y) \text{ good}\} \\
& + Pr\{\pi(x,y) \neq f(x,y)\} \\
< & \rho/2 + \rho/2 + \epsilon = \rho + \epsilon
\end{aligned}$$

Therefore $D^\mu_{\rho+\epsilon}(f) = 2^{O(1/\rho+IC_\mu(f,\epsilon)/\rho^2)}$ holds.

Next, we just need to apply Theorem 2.1, we can obtain the same bound for randomized information complexity, by setting μ to be the optimal choice.

$$R_\epsilon(f) = D^\mu_\epsilon(f) = 2^{O(1/\rho+IC_\mu(f,\epsilon)/\rho^2)} \leq 2^{O(1/\rho+IC(f,\epsilon)/\rho^2)}$$

□

4 Preliminaries: the Quantum Case

In this section, we briefly overview the related concepts but in the quantum settings.

4.1 Conventions

As in classical case, we also need some additional quantum related notations, which are shown in Table 4.1.

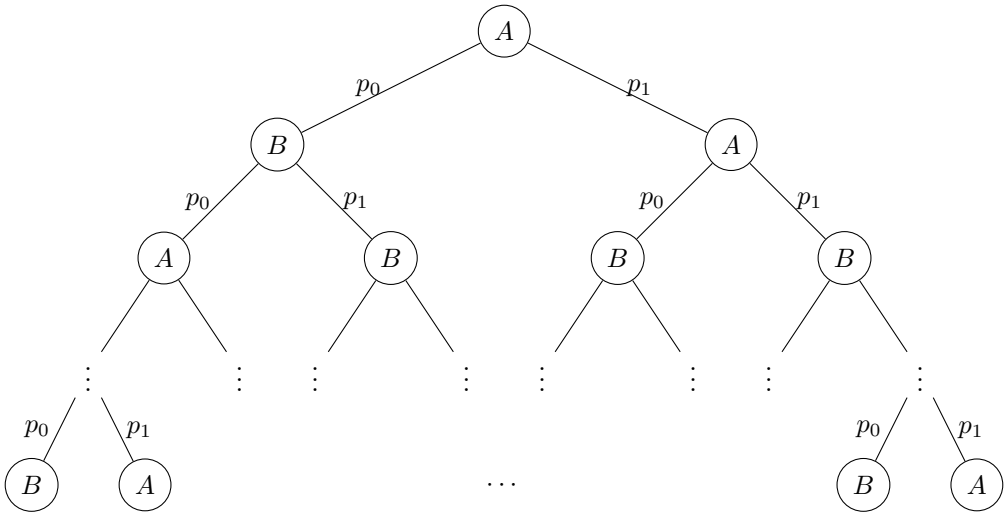

Figure 3.1: An example protocol tree. A represents Alice and B represents Bob. Going to left child means sending 0, and right means sending 1. Each node (except leaf) has certain probability to send 0 or 1. These probabilities do not have to be the same.

Symbols	Meaning
$\mathcal{X}, \mathcal{Y}, \mathcal{Z}$	complex Euclidean spaces
v, u	vectors
A, B	linear operators
ρ, σ	quantum states / density operators
P, Q	positive semidefinite matrices
A^*	adjoint of A
Φ, Ψ	quantum channels

Table 4.1: Convention table

4.2 Quantum Information Theory

Quantum information theory is a generalized information theory in the quantum settings. A more complete discussion can be found in [18]. Note that all quantum related notions in this report respect the convention in [18], which does not agree with some other references.

Definition 4.1. *We use the following notions represent sets of linear operators / matrices.*

$$\begin{aligned} L(\mathcal{X}, Y) &= \{\textit{linear operators from } \mathcal{X} \textit{ to } \mathcal{Y}\} \\ L(\mathcal{X}) &= L(\mathcal{X}, \mathcal{X}) \\ Herm(\mathcal{X}) &= \{H \in L(\mathcal{X}) : H = H^*\} && \text{Hermitian matrcies} \\ Pos(\mathcal{X}) &= \{A^*A : A \in L(\mathcal{X})\} && \text{positive semidefinite matrices} \\ Pd(\mathcal{X}) &= \{P \in Pos(\mathcal{X}) : \det(P) \neq 0\} && \text{positive definite matrices} \\ D(\mathcal{X}) &= \{\rho \in Pos(\mathcal{X}) : Tr(\rho) = 1\} && \text{density matrices} \end{aligned}$$

where Tr is the trace operator.

Definition 4.2. *The von Neumann entropy, or entropy, of $P \in Pos(\mathcal{X})$ is*

$$H(P) = H(\lambda(P))$$

where $\lambda(P)$ is the vector of eigenvalues of P, and the second H denotes Shannon entropy.

Similarly, we can define quantum conditional entropy and quantum mutual information in the same way as the classical case.

Definition 4.3. *Quantum divergence / quantum relative entropy of P w.r.t. Q, where $P, Q \in Pos(\mathcal{X})$ is*

$$D(P \parallel Q) = \begin{cases} Tr(P \log P) - Tr(P \log Q), & \text{if } im(P) \subseteq im(Q) \\ \infty, & \text{otherwise} \end{cases}$$

where $im(P)$ defines the imagine of P as an operator. Logarithm of P is defined by taking logarithm on eigenvalues in P's spectral decomposition.

$$\log P = \sum_i \log(\lambda_i)\Pi_i$$
$$\text{where } P = \sum_i \lambda_i \Pi_i, \Pi_i \text{ are projection matrices.}$$

Next, we define quantum channels.

Definition 4.4. *A quantum channel is a linear map,*

$$\Phi : L(\mathcal{X}) \to L(\mathcal{Y})$$

so that it

1. *preserves trace.*
2. *is completely positive. Namely, it preserves positive semi-definiteness of the inputs when tensored with identity map in any space.*

The set of channels is denoted by $\mathcal{C}(\mathcal{X}, \mathcal{Y})$, or $\mathcal{C}(\mathcal{X})$ when $\mathcal{X} = \mathcal{Y}$.

Complete positivity of a channel says if a channel tensors with an identity channel, then the resulting linear map is still a channel. Therefore, if the space is known, it's unambiguous to implicitly augment channels with identity channel. Namely, for $\Phi : C(\mathcal{X}, \mathcal{Y})$, we write $\Phi : C(\mathcal{X} \otimes \mathcal{Z}, \mathcal{X} \otimes \mathcal{Z})$ instead of $\Phi \otimes I_{\mathcal{Z}}$.

We might want to compose channels.

$$\Psi\Phi(X) = \Psi(\Phi(X))$$

Namely, we write the channels right to left in the order they are applied. We sometimes might need to take partial trace of a matrix $A \in \mathcal{X} \otimes \mathcal{Y}$, we might use any of these notations

$$Tr_{\mathcal{X}}(A) = A[Y] = A_Y$$

Partial and full traces are examples of channels.

Definition 4.5. *The Kraus representation of a channel $\Phi \in \mathcal{C}(\mathcal{X}, \mathcal{Y})$ is*

$$\Phi(X) = \sum_a A_a X A_a^*,$$
$$s.t. \sum_a A_a^* A_a = \mathbb{1}_X$$

We can define the adjoint of a channel via Kraus representation.

Definition 4.6. *The adjoint of a channel $\Phi \in \mathcal{C}(\mathcal{X}, \mathcal{Y})$, $\Phi^* \in \mathcal{C}(\mathcal{Y}, \mathcal{X})$, is*

$$\Phi^*(Y) = \sum_a A_a^* Y A_a$$

Definition 4.7. *A unitary channel, or just unitary, $U \in \mathcal{C}(\mathcal{X})$, is a channel, such that*

$$U^*U(X) = X, \text{for all } X \in L(\mathcal{X})$$

The set of all unitary channel is denoted by $\mathcal{U}(\mathcal{X})$. In Kraus representation, a unitary channel can be represented using a unitary matrix.

$$U(X) = U_a X U_a^*, \text{where } U \in \mathcal{U}(\mathcal{X}), U_a \text{ unitary matrix}$$

We then define norms on operators.

Definition 4.8. *The norms of A are defined as follows.*

$$\|A\| = \|A\|_1 = Tr(\sqrt{AA^*}) \qquad \text{(trace norm)}$$
$$\|A\|_2 = \sqrt{\langle A, A\rangle} \qquad \text{(2-norm, Frobenius norm)}$$
$$\|A\|_\infty = \max\{s(A)\} \qquad \text{(spectral norm)}$$

where $s(A)$ is the set of singular values of A.

A quantum state is a special kind of operator.

Definition 4.9. *A quantum state ρ is a density operator.*

$$\rho \in D(\mathcal{X})$$

Following definitions and theorem involve pure states and purification of states.

Definition 4.10. *A state $\rho \in D(\mathcal{X})$ is pure, when there exists $v \in \mathcal{X}$ satisfying*

$$\rho = vv^*$$

Definition 4.11. *For a state $\rho \in D(\mathcal{X})$, a vector $v \in \mathcal{X} \otimes \mathcal{Y}$ purifies ρ, when*

$$\rho = Tr_{\mathcal{Y}}(vv^*)$$

Theorem 4.1. *For any $\rho \in D(\mathcal{X})$, there exists $v \in \mathcal{X} \otimes \mathcal{Y}$ purifies ρ, iff $dim(\mathcal{Y}) \geq dim(\mathcal{X})$.*

4.3 Quantum Communication Complexity

A quantum communication model is presented in Figure 4.1. For each state, their scripted fonts represents their space. For example, B_2 lives in $\mathcal{B}_2$. The model has following components.

1. $\rho \in D(\mathcal{A}_I \otimes \mathcal{B}_I)$, the input state.
2. U_i, each is a unitary channel with appropriate input and output spaces.
3. ψ, a shared pure state.
4. U_i communicates with U_{i+1} using register C_i.
5. A_I and B_I are input registers, and they together represent ρ. Alice uses A_i registers to pass the state from previous channel to the next. Similarly, Bob uses B_i for the same purpose.

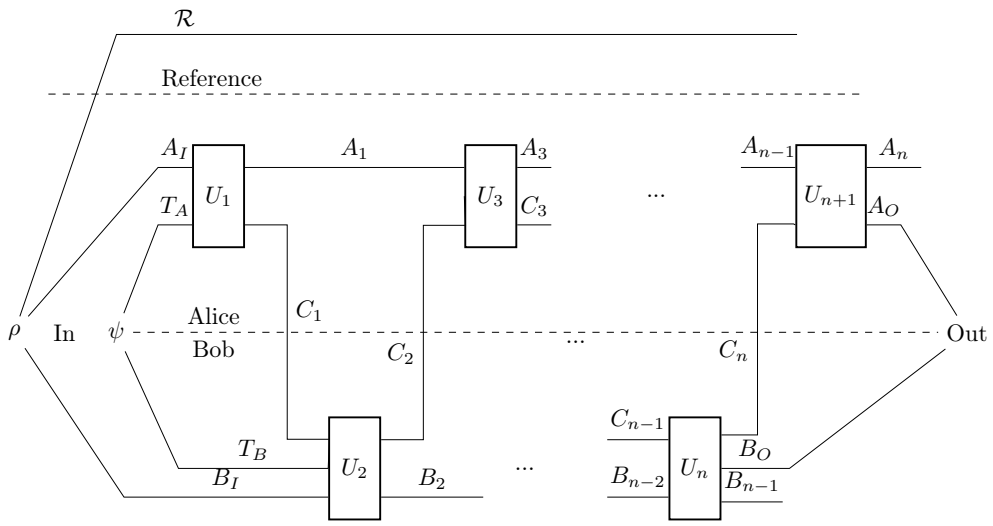

Figure 4.1: Illustration of quantum communication model [17]

6. Together, A_O and B_O represent the output of the overall protocol.

7. $\mathcal{R}$ serves as a reference space.

Therefore, the overall communication protocol can be represent as follows.

$$\Pi(\rho) = Tr_{\mathcal{A}_n\mathcal{B}_{n-1}} U_{n+1}U_n \cdots U_1(\rho \otimes \psi)$$

Note that we put implicit identity channels wherever appropriate to make the dimension correct.

$\mathcal{R}$ is a space introduced so that $\rho_{A_IB_IR}$ is a purification of ρ.

Definition 4.12. *For an intended channel $\Phi \in \mathcal{C}(\mathcal{A}_I \otimes \mathcal{B}_I, \mathcal{A}_O \otimes \mathcal{B}_O)$, Π implements Φ on $\rho \in D(\mathcal{A}_I \otimes \mathcal{B}_I)$, if the error $\epsilon \in [0, 2]$ has*

$$\|\Pi \otimes I_{\mathcal{R}}(\rho_{A_IB_IR}) - \Phi \otimes I_{\mathcal{R}}(\rho_{A_IB_IR})\| \leq \epsilon$$

where $\rho_{A_IB_IR}$ denotes a purification of ρ for suitable $\mathcal{R}$.

We denote the set of all Π with error bounded by ϵ as $\mathcal{T}(\Phi, \rho, \epsilon)$.

Definition 4.13. *For a protocol Π, the quantum communication cost is*

$$QCC(\Pi) = \sum_i \log \dim(C_i)$$

Definition 4.14. *For a channel $\Phi \in \mathcal{C}(\mathcal{A}_I \otimes \mathcal{B}_I, \mathcal{A}_O \otimes \mathcal{B}_O)$ and input state $\rho \in D(\mathcal{A}_I \otimes \mathcal{B}_I)$, and error $\epsilon \in [0, 2]$, the quantum communication complexity is*

$$QCC(\Phi, \rho, \epsilon) = \min_{\Pi \in \mathcal{T}(\Phi,\rho,\epsilon)} QCC(\Pi)$$

Similarly, we can define protocols of n-fold channels and the definitions of corresponding quantum communication cost and communication complexity.

Definition 4.15. *A protocol Π_n implements n-fold product channel $\Phi^{\otimes n} \in \mathcal{C}((\mathcal{A}_I \otimes \mathcal{B}_I)^{\otimes n}, (\mathcal{A}_O \otimes \mathcal{B}_O)^{\otimes n})$ on input $\rho^{\otimes n} \in D((\mathcal{A}_I \otimes \mathcal{B}_I)^{\otimes n})$ with error ϵ, if for all $i \in [n]$,*

$$\|\Pi_n(\rho^{\otimes n}_{A_I B_I R})_{A_O^i B_O^i R^i} - \Phi \otimes I_{\mathcal{R}}(\rho_{A_I B_I R})\| \leq \epsilon$$

where the superscript of i denotes the i-th copy of the space.

Namely, the definition requires each computation performed by Π_n has bounded error.

Definition 4.16. *The n-fold quantum communication complexity is*

$$QCC(\Phi^{\otimes n}, \rho^{\otimes n}, \epsilon) = \min_{\Pi_n \in \mathcal{T}_n(\Phi^{\otimes n}, \rho^{\otimes n}, \epsilon)} QCC(\Pi_n)$$

Definition 4.17. *The amortized quantum communication complexity is*

$$AQCC(\Phi, \rho, \epsilon) = \limsup_{n \to \infty} \frac{1}{n} QCC_n(\Phi^{\otimes n}, \rho^{\otimes n}, \epsilon)$$

4.4 Quantum Information Complexity

In [17], Touchette motivates a definition which generalizes the classical definition of information complexity to the quantum case. Following the communication diagram in Figure 4.1, we can give the definition of quantum information cost.

Definition 4.18. *The quantum information cost of a protocol Π on an input ρ is*

$$QIC(\Pi, \rho) = \sum_{i \ odd} \frac{1}{2} I(C_i; R|B_{i-1}) + \sum_{i \ even} \frac{1}{2} I(C_i; R|A_{i-1})$$

where $B_0 = B_I \otimes T_B$.

Intuitively, the quantum information cost is an accumulation of quantum information relative to the reference space exchanged between Alice and Bob, conditioned on the register they are holding. Following this intuition, indeed this definition gives a sense of generalization of the classical case.

The definition of quantum information complexity follows quite similarly.

Definition 4.19. *The quantum information complexity of Φ on ρ with error $\epsilon \in [0, 2]$ is*

$$QIC(\Phi, \rho, \epsilon) = \inf_{\Pi \in \mathcal{T}(\Phi, \rho, \epsilon)} QIC(\Pi, \rho)$$

Similar to Theorem 2.6, there is also a theorem which operationalizes this definition of quantum information complexity.

Theorem 4.2.

$$QIC(\Phi, \rho, \epsilon) = AQCC(\Phi, \rho, \epsilon)$$

This says that the amortized cost of communication of the best protocol is essentially as much as the least information exchange. This operationalization theorem indicates that this definition of quantum information complexity is the correct definition in the quantum settings.

In Section 2.4, the definitions of information complexity are based on worst-case distribution over inputs. In quantum settings, since quantum state itself has encoded randomness, we can just pick the worst input for the same purpose.

Definition 4.20. *The max-distributional quantum information complexity is*

$$QIC_D(\Phi, \epsilon) = \max_{\rho \in D(\mathcal{X} \otimes \mathcal{Y})} QIC(\Phi, \rho, \epsilon) = \max_{\rho \in D(\mathcal{X} \otimes \mathcal{Y})} \inf_{\Pi \in \mathcal{T}(\Phi, \rho, \epsilon)} QIC(\Pi, \rho)$$

Definition 4.21. *The quantum information complexity is*

$$QIC(\Phi, \epsilon) = \inf_{\Pi \in \mathcal{T}(\Phi, \epsilon)} \max_{\rho \in D(\mathcal{X} \otimes \mathcal{Y})} QIC(\Pi, \rho)$$

where $\mathcal{T}(\Phi, \epsilon)$ denotes the set of all protocols implementing Φ with the worst case error ϵ.

It would be very desirable if the prior-free definition of quantum information complexity, $QIC(\Phi, \epsilon)$, also has operationalization property as Theorem 4.2. However, such property is not proved in [6], nor a proof is discovered during our literature review.

From the definition, we can already see some similarity between the quantum and the classical case. [17, 6] more thoroughly explores many properties of this definition quantum information complexity. Following theorem resembles the classical version.

Theorem 4.3. *[6, Theorem 4.13]*

$$QIC(f, \frac{\epsilon}{\alpha}) \leq \frac{QIC_D(f, \epsilon)}{1 - \alpha}$$

[6] proves an analogue of Theorem 2.7 in quantum settings, which is stated in the following theorem.

Theorem 4.4. *[6, Theorem D, Corollary 5.8] Let f be a boolean function,*

$$QCC(f, 1/3) \leq 2^{O(QIC(f, 1/3)+1)}$$

In this statement, the notion overloads QCC and QIC to take a function, where in the formal definitions, these notions take a channel. A classical function can be made into a quantum channel, by considering each input and output. More formally, for $f \in A \times B \to X \times Y$, $a \in A, b \in B$, if $f(a, b) = (x, y)$, then

$$\Phi(|a\rangle\langle a| \otimes |b\rangle\langle b|) = |x\rangle\langle x| \otimes |y\rangle\langle y|$$

5 Proof Sketch of Theorem 4.4

In this section, we attempt to establish a high level proof sketch of Theorem 4.4. The actual proof of Theorem 4.4 is rather complex, and a thorough proof as in Section 3 would be very long. Therefore, it's much better off to provide a clear proof outline instead of getting into the details.

From very high level, the proof utilizes the fact that, generalized discrepancy method, GDM, is a lower bound of quantum information complexity. Therefore, the relation is relatively straightforwardly established by using the relation between discrepancy and randomized communication complexity. This method is not the same as the proof in Section 3, where the original protocol is compressed to another protocol by sampling public randomness. More verbosely, the proof can be roughly divided into the following steps.

1. Show that there is always a protocol which computes a function f by n-fold, such that the probability for computing at least some ratio of correct results is bounded from below.
2. Use result from [15] to establish the relation between generalized discrepancy methods and quantum information complexity.
3. Finalize the proof by chaining a series of inequalities in communication complexity.

In particular, the observation is generalized discrepancy method, GDM, has already been a strong lower bound, and the work pending to be done, is to actually demonstrate a protocol that witness the relation between GDM and QIC_D (which is done in Item 1). The concrete definition of GDM is defined as follows.

Definition 5.1. *For boolean function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ and error ϵ,*

$$\begin{aligned} GDM_\epsilon(f) &= \max_{\mu \text{ a distribution over } \mathcal{X}\times\mathcal{Y}} \{GDM_\epsilon^\mu(f)\} \\ GDM_\epsilon^\mu(f) &= \max_{g:\mathcal{X}\times\mathcal{Y}\to\{0,1\}, \Pr_{x,y\sim\mu}\{f(x,y)\neq g(x,y)\}\leq\epsilon\}} \{\log \frac{1}{disc^\mu(g)}\} \\ disc^\mu(g) &= \max_R |\sum_{(x,y)\in R} (-1)^{g(x,y)}\mu(x,y)| \end{aligned}$$

where R is a rectangle in $\mathcal{X} \times \mathcal{Y}$.

Definition 5.2. *A rectangle R in $\mathcal{X} \times \mathcal{Y}$ satisfies*

1. *$R \subseteq \mathcal{X} \times \mathcal{Y}$*
2. *For some $\mathcal{A} \subseteq \mathcal{X}, \mathcal{B} \subseteq \mathcal{Y}$, $R = \mathcal{A} \times \mathcal{B}$.*

It happens that discrepancy is a lower bound of classical communication complexity.

Theorem 5.1. *For boolean function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, error ϵ and distribution $\mu : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$,*

$$D^\mu_{1/2-\epsilon}(f) \geq \log \frac{2\epsilon}{disc^\mu(f)}$$

To state the significance of GDM, we fist need the following definition.

Definition 5.3. *For positive integer n, percentage η and error ϵ, n-fold quantum communication complexity $QCC(f^n, \eta n, \epsilon) \leq C$, if there exists a protocol Π which computes f^n, such that*

1. *$QCC(\Pi) \leq C$*
2. *$\Pr\{\Pi$ computes at least ηn coordinates correctly$\} \geq 1 - \epsilon$*

Note that this definition requires Π successfully computes ηn coordinates with certain probability (but not necessarily high) for all inputs. Then, the following theorem is shown in [15].

Theorem 5.2. *[15, Theorem 1] There exists an absolute constant $\epsilon_{sh} > 0$, such that for $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$,*

$$\Omega(nGDM_{1/5}(f)) \leq QCC(f^n, (1-\epsilon_{sh})n, 1 - 2^{-\epsilon_{sh}n})$$

with arbitrary entanglement.

Notice that the meaning of $QCC(f^n, (1-\epsilon_{sh})n, 1-2^{-\epsilon_{sh}n})$ is subtle. First, since ϵ_{sh} is a small absolute constant, the theorem essentially requires most of the coordinates to success, and only in this situation, the quantum has a lower bound. This quantity is more specific than, say, an probably natural definition of a prior-free accumulated quantum communication complexity, similar to Definition 4.17. Second, the error $1 - 2^{-\epsilon_{sh}n}$ grows exponentially as n grows, so the chances for success is exponentially decreasing. This implies the success case is unlikely to happen, and therefore amortization of this definition is unlikely approximating $QIC(f, \epsilon)$. This definition essentially prevents a proof plan similar to Theorem 2.7, as we will expand further. Despite that, we can show the following relation.

Theorem 5.3.

$$QCC(f^n, (1-\epsilon_{sh})n, 1-2^{-\epsilon_{sh}n}) \leq O(n(QIC(f,\epsilon)+2)) + o(n)$$

Following two previous theorems and by transitivity, we will have the following theorem.

Theorem 5.4. [2] *[6, Theorem 5.7]*

$$\Omega(GDM_{1/5}(f)) \leq QIC(f,\epsilon) + O(1)$$

Essentially, the proof of Theorem 4.4 is connected by the previous inequalities.

Proof of Theorem 4.4.

$$\begin{aligned}
QCC(f,1/3) &\leq R(f,1/3) \\
&\leq \Big(\frac{1}{disc(f)}\Big)^{O(1)} && (\tfrac{1}{disc(f)} \text{ approximates number of rectangles}) \\
&\leq 2^{O(GDM_{1/5}(f))} && \text{(by definition of } GDM) \\
&\leq 2^{O(QIC(f,\epsilon)+O(1))} && \text{(by Theorem 5.4)}
\end{aligned}$$

In the second inequality, we omit the probability distribution μ. We can take the worst case μ for this purpose, and the third inequality will still hold, because GDM is maximized over all μ's.

To finalize the proof, we just need to relax the error. □

6 Comparing Proofs of Theorem 2.7 and Theorem 4.4

Recall that, in the classical and quantum case, communication complexity is bounded by information complexity exponentially.

$$\begin{aligned}
R(f) &= 2^{O(IC(f))} \\
QCC(f) &= 2^{O(QIC(f))}
\end{aligned}$$

Despite the similarity of their statements, their proof techniques are not quite similar. In particular, in the classical case, the main part of the proof is to show that any protocol can be approximated by a lossy compression, so that the communication complexity is upper bounded. However, in the quantum case, the proof essentially relies on the characteristics of general discrepancy method. The proof of the quantum case is less satisfactory, in the sense that, it's unclear which protocol can be taken to actually witness the upper bound of communication complexity. This dissatisfaction can be found in [5, Problem 4] and [1].

Reading [5, 2], one can realize that the correctness of the compression protocol relies on that in the classical settings, both Alice and Bob have persistent memory, namely they don't lose knowledge. In particular, in the compression protocol in [2, Section 7], a player can rely on an operation ([2, Lemma 4.14]) to correct his / her estimation of the overall communication protocol. This operation implicitly requires the players remembers the overall estimation to begin with. However, in the quantum case, if a player processes a quantum state using a channel and obtains another one, the player behaves as if the previous quantum state never existed. From this perspective, one can see that in the method of protocol compression, the classical communication

[2] We formulate this theorem slightly differently here. In [6], this theorem concludes with QIC_D, which seems strange, because QIC_D is a weaker bound than QIC, and we don't have theorem to go from QIC_D to QIC. However, the proof steps seem just ok with stating the conclusion with QIC directly.

model and quantum model seem to exhibit some fundamental differences, which forbids a direct transcribe from classical settings to quantum settings.

Moreover, quantum lower bound by generalized discrepancy method (Theorem 5.2) holds with arbitrary entanglement. This is another difference between two communication models which makes a direct transcription much less obvious.

7 Conclusion

In this report, we've briefly summarized basic definitions of information theory, communication complexity and information complexity in both classical and quantum settings. We discussed the theorems stating that communication complexity is bounded by information complexity exponentially in both settings. The proof for classical case is expanded with greater details and the one for quantum case is outlined. At last, we explained why indirect proof in quantum case is unsatisfactory and a number of differences between models in discussion which might have brought difficulties in a direct proof via compression argument.

References

[1] A. Anshu, D. Touchette, P. Yao, and N. Yu. Exponential separation of quantum communication and classical information. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, pages 277–288, New York, NY, USA, 2017. ACM.

[2] B. Barak, M. Braverman, X. Chen, and A. Rao. How to compress interactive communication. *SIAM Journal on Computing*, 42(3):1327–1363, 2013.

[3] M. Braverman. Interactive information complexity. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pages 505–524, New York, NY, USA, 2012. ACM.

[4] M. Braverman. Interactive information and coding theory. In *Proceedings of the International Congress of Mathematicians*, pages 535–559. Citeseer, 2014.

[5] M. Braverman. Interactive information complexity. *SIAM Journal on Computing*, 44(6):1698–1739, 2015.

[6] M. Braverman, A. Garg, Y. Kun-Ko, J. Mao, and D. Touchette. Near-optimal bounds on bounded-round quantum communication complexity of disjointness. *CoRR*, abs/1505.03110, 2015.

[7] M. Braverman and A. Rao. Information equals amortized communication. *CoRR*, abs/1106.3595, 2011.

[8] A. Ganor, G. Kol, and R. Raz. Exponential separation of information and communication. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*.

[9] A. Ganor, G. Kol, and R. Raz. Exponential separation of information and communication for boolean functions. In *Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing*, STOC '15, pages 557–566, New York, NY, USA, 2015. ACM.

[10] M. Goemans. Chernoff bounds, and some applications.

[11] I. Kerenidis, S. Laplante, V. Lerays, J. Roland, and D. Xiao. Lower bounds on information complexity via zero-communication protocols and applications. *CoRR*, abs/1204.1505, 2012.

[12] T. Pitassi. Communication complexity, information complexity and applications.

[13] C. E. Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948.

[14] A. A. Sherstov. The pattern matrix method for lower bounds on quantum communication. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08, pages 85–94, New York, NY, USA, 2008. ACM.

[15] A. A. Sherstov. Strong direct product theorems for quantum communication and query complexity. *CoRR*, abs/1011.4935, 2010.

[16] D. Touchette. Quantum information complexity and amortized communication. *CoRR*, abs/1404.3733, 2014.

[17] D. Touchette. Quantum information complexity. In *Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing*, STOC '15, pages 317–326, New York, NY, USA, 2015. ACM.

[18] J. Watrous. The theory of quantum information. 2018.

[19] O. Weinstein. *Interactive Information Complexity and Applications*. PhD thesis, Princeton University, 2015.

[20] A. C. Yao. Lower bounds by probabilistic arguments. In *24th Annual Symposium on Foundations of Computer Science (sfcs 1983)*, pages 420–428, Nov 1983.

[21] A. C.-C. Yao. Quantum circuit complexity. In *Proceedings of 1993 IEEE 34th Annual Foundations of Computer Science*, pages 352–361, Nov 1993.