The Adaptive Capacity for Two Mixed States

Lucy Xiao

Under the direction of
Andrey Boris Khesin
Massachusetts Institute of Technology

Research Science Institute
August 2, 2021

Abstract

In this paper, we investigate the one-shot capacity $C_{1,1}$, and the adaptive one-shot capacity $C_{1,A}$ in transmitting classical information through quantum channels. Capacity $C_{1,A}$ allows for an adaptive strategy where the receiver can make non-disruptive measurements and adapt using the results of previous measurements. Shor proves that $C_{1,1} = C_{1,A}$ for two pure states, and conjectures that the same is true for mixed states. We show that the accessible information is concave in the probability of the states, then conclude $C_{1,1} = C_{1,A}$ for any two mixed states in arbitrary dimensions, proving the conjecture.

Summary

In quantum computing, parties communicate through a quantum channel by sending particles exhibiting certain quantum states. The receiver measures these states to distinguish them and thus uncover information. The task of uncovering the most information possible is, however, non-trivial because quantum states are formed as a superposition of different outcomes, but only one outcome can be measured. Thus, clever protocols need to be implemented for the maximal amount of information gain, which also varies under different constraints. One set of constraints is to measure each state individually, and another is to have an additional ability to measure in small steps where the later steps are dependent on the results of previous measurements – essentially adapting along the way. Shor proves that for two pure states, in which case one knows for sure the state one's holding, the adapting option does not improve one's information gain, and he conjectures that for any two arbitrary mixed states, the same is true. In this paper, we prove Shor's conjecture, now a theorem.

1 Introduction

Early in the 20th century, quantum mechanics was introduced as a revolutionary field of physics describing nature at the atomic and subatomic scale. Although controversial at first, it has now become a staple of modern physics. Mathematicians, scientists, and engineers have developed the field of quantum computing using the seemingly absurd properties in quantum mechanics, such as the uncertainty principle, the phenomenon of entanglement, and superfluidity, to their advantage.

Classical computers have revolutionized the way we solve problems. Much evidence such as Simon's problem and Shor's algorithm [1] for fast factoring has suggested that quantum computers can solve some problems exponentially or almost exponentially faster than classical computers. It is also believed that quantum computers can do a far better job at simulating quantum systems due to their similar nature in design. Drug development, for instance, can benefit profoundly from the progress in quantum computing, as well as cybersecurity, financial modeling, batteries, and the list goes on [2].

A qubit, analogous to a classical bit, is the unit of communication in quantum information. Instead of 0 and 1 for a classical bit, a qubit can be thought of as a superposition of the two states. However, one cannot simply acquire the exact number for the mixture of 0 and 1 through measurement. To be specific, if one puts a qubit through a measurement consisting of basis 0 and 1, one would only get either 0 or 1. Thus, it is apparent that to have efficient communication, clever protocols and strategies of encodings and measurements are needed to be implemented for the sender and the receiver.

Two capacities have been studied extensively: the one-shot capacity $C_{1,1}$ and the joint measurement capacity $C_{1,\infty}$. For the former, when Alice sends a large number of signals through a quantum channel to Bob, Bob can only measure each qubit individually. This one-shot strategy used can be thought of as the most basic one since it does not take into account

any entanglement or correlation between states. The joint measurement is just the opposite, where Bob can measure multiple qubits at once, taking into account the entanglement, thus acquiring more information.

A capacity that lies between the two in terms of information gain is the adaptive one-shot capacity, $C_{1,A}$. Using an adaptive one-shot strategy, Bob is still not allowed to measure multiple qubits at once just like for the non-adaptive one-shot measurement; however, he can make a measurement on a qubit that doesn't destroy all the information, then perform another measurement on a different qubit, and depending on the result, return to the first qubit to make more measurements.

It is clear that the capacity of the adaptive strategy lies between the two, but whether the bounds are tight is an interesting question. Answering such a question will give us more insight into the most efficient quantum communications strategy within the limitations of the senders and receivers in various situations. It has been proven for two pure states that the adaptive strategy won't make a difference compared to the one-shot strategy [3].

In this paper, we prove the same claim is true for two mixed states, which is the conjecture 2 at the end of [3].

Question 1. The capacity $C_{1,1} = C_{1,A}$ for two mixed states in arbitrary dimensions.

We introduce the basics of quantum computing including quantum states, measurements, accessible information, and capacities of different strategies in Section 2. Then in Section 3, we prove a lemma regarding the concavity of the accessible information, which is essential to the generalization. In Section 4, we briefly summarize the argument in [3] to use the lemma in Section 3 to finish the proof of the conjecture, now a theorem.

2 Preliminaries

2.1 Quantum States

We describe a qubit mathematically using a quantum state. There are two types of quantum states: pure states and mixed states. A *pure state* is defined as

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle\,, \tag{1}$$

where α and β are two complex numbers. This linear combination tells us that if we were to measure such $|\psi\rangle$ in the basis of $|0\rangle$ and $|1\rangle$, we would have $|\alpha|^2$ probability of getting 0 and $|\beta|^2$ probability of getting 1. It is easy to see that

$$|\alpha|^2 + |\beta|^2 = 1$$

However, a state is only *pure* iff the state can be written in the form in Equation 1. Luckily, we can describe a mixed state using a *mixture* of pure states. The density matrix for the *mixed state* ρ is defined as

$$\rho \equiv \sum_i p_i\,|\psi_i\rangle\,\langle\psi_i|\,,$$

where $\sum_i p_i = 1$, meaning ρ has p_i probability of behaving as $|\psi_i\rangle$.

2.2 Measurement

We observe and distinguish quantum states by making measurements. We call a generalized measurement POVM, positive operator valued measure. A POVM measurement is described by a set of measurement operators M_i, where i indicates the outcomes of the measurement. In our context, measurement operators such as M can be treated as matrices, while quantum states can be treated as row ($\langle\psi|$) and column ($|\psi\rangle$) vectors. An outcome i occurs for state $|\psi\rangle$ with probability

$$p(i) = \langle\psi|\,M_i^\dagger M_i\,|\psi\rangle\,.$$

It is clear that the probability for all measurement outcomes must add up to 1 for all

state $|\psi\rangle$. Thus, we have

$$1 = \sum_i p(i) = \langle\psi| M_i^\dagger M_i |\psi\rangle ,$$

which gives us the completeness relation

$$\sum_i M_i^\dagger M_i = I.$$

After coming out of measurement i, a state $|\psi\rangle$ becomes

$$\frac{M_i |\psi\rangle}{\sqrt{\langle\psi| M_i^\dagger M_i |\psi\rangle}}.$$

A special and important class of measurement is the von Neumann measurement, also called projective measurement. Such measurements satisfy the property that M_i are *orthogonal* projector: $M_i M_j = \delta_{i,j} M_i$, where delta is the indicator function.

2.3 Accessible Information

The accessible information is the amount of classical information we can get from one use of a quantum channel using a given ensemble of quantum states with optimal measurement. It is the expected difference between the *Shannon entropy* before and after such measurement. *Shannon entropy* measures the amount of uncertainty in a given probability distribution. For example, a probability distribution of $\{\frac{1}{3}, \frac{2}{3}\}$ is less uncertain than a probability distribution $\{\frac{1}{2}, \frac{1}{2}\}$. The Shannon entropy is defined as

$$H(X) \equiv H(p_1, ..., p_n) \equiv -\sum_i p_i \log p_i.$$

where X is a random variable with a discrete probability distribution over n events with probabilities p_1, p_2, ..., p_n. Let Y_i be a random variable with probability distribution conditioned on the outcome i. If we get measurement outcome i, we have information gain $H(X) - H(X|Y_i)$, and thus our accessible information is

$$I_{acc} = \sum_i p(i)(H(X) - H(X|Y_i)) = H(X) - \sum_i p(i)H(X|Y_i),$$

where $\sum_i p(i)H(X|Y_i)$ is the expected value of the Shannon entropy after the measurement.

2.4 Capacities

Capacity is the maximum accessible information over all possible protocols given some restrictions. Here we discuss three types of capacities:

1. The $C_{1,1}$ capacity for a quantum channel is the one-shot quantum capacity, where only tensor product inputs and tensor product measurements are allowed.

2. The $C_{1,A}$ capacity is a classical capacity where each state is measured in stages that only partially reduce the quantum state, and the measurement stages followed adapt to the results of previous ones.

3. The $C_{1,\infty}$ capacity allows both tensor product inputs and joint measurements are allowed, resulting in the most information gain.

It is not difficult to see that $C_{1,1} \leq C_{1,A} \leq C_{1,\infty}$. We seek to investigate the tightness of the lower bound later in the paper.

2.5 Notations

We use the following notations.

Definition 2.1. For a random variable X, $\mathbb{E}[X]$ is the expected value of X.

Definition 2.2. For the entropy of a probability distribution among 2 items, we define $H_2(p) \equiv H(p, 1-p) = -p \log p - (1-p) \log(1-p)$.

Definition 2.3. Let $P(A)$ be the probability that event A occurs, and $P(A|B)$ be the probability of event A occurs given that event B has occurred.

3 Concavity of Accessible Information for Two Mixed States

Theorem 3.1. *For two mixed states in arbitrary dimensions $C_{1,A} = C_{1,1}$.*

To prove this theorem, we first need to show that the following lemma is true.

Lemma 3.2. *For two mixed states, the accessible information is concave in the probability of the state.*

Proof. In an n–dimensional space, define two mixed states ρ_1 and ρ_2, where

$$\rho_1 = \sum_i p_i \left|v_i\right\rangle \left\langle v_i\right|,$$
$$\rho_2 = \sum_i q_i \left|w_i\right\rangle \left\langle w_i\right|,$$

where $\langle v_i|v_j\rangle = \langle w_i|w_j\rangle = \delta_{ij}$, $\sum_i p_i = \sum_i q_i = 1$, and all $p_i, q_i \geq 0$.

Note that if we originally have non-orthonormal $|v_i'\rangle$ or $|w_i'\rangle$, using the *Spectral theorem*, we can break down the original basis to some $|v_i\rangle$ and $|w_i\rangle$ of orthonormal basis. We know that all above conditions are still satisfied, since

$$1 = \mathrm{tr}(\rho) = \mathrm{tr}\left(\sum_i p_i \left|v_i\right\rangle \left\langle v_i\right|\right) = \mathrm{tr}\left(\sum_i p_i \left\langle v_i|v_i\right\rangle\right) = \sum_i p_i.$$

Assume Alice sends ρ_1 with probability r, and thus ρ_2 with probability $1 - r$. We want to prove that the accessible information is convex in r.

Let the optimal POVM measurement have measurement operators

$$M_i = \left|\psi_i\right\rangle \left\langle \psi_i\right|,$$

where $1 \leq i \leq n$.

We can write each M_i as a rank-1 projector, because otherwise, we haven't reduced the mixed state to a known pure state, thus leaving extra information undiscovered. Therefore, to extract all the information possible, we must have M_i in the form of the outer product of $|\psi_i\rangle$ and $\langle\psi_i|$.

We define a_i to be the probability of getting result i from ρ_1, and similarly we define b_i to be the probability of getting result i from ρ_2. Thus, we have

$$\begin{aligned} a_i \equiv P(i|\rho_1) &= \mathrm{tr}(M_i\rho_1) \\ &= \mathrm{tr}\left(|\psi_i\rangle\langle\psi_i| \sum_i p_i |v_i\rangle\langle v_i|\right) \\ &= \mathrm{tr}\left(\sum_i p_i \langle\psi_i|v_i\rangle\langle v_i|\psi_i\rangle\right) \\ &= \sum_j p_j |\langle\psi_i|v_j\rangle|^2. \end{aligned}$$

Analogously, we have

$$b_i \equiv P(i|\rho_2) = \sum_j q_j |\langle\psi_i|w_j\rangle|^2.$$

We know that a_i and b_i are real numbers such that $\sum_i a_i = \sum_i b_i = 1$ and a_i, $b_i > 0$.

Now we look at the accessible information. We know that $I_{acc} = H_2(r) - \mathbb{E}[H']$, where H' denotes the Shannon entropy after the measurement. The expected value of the Shannon entropy after the measurement is

$$\mathbb{E}[H'] = \sum_i P(i) H_2(P(\rho_1|i)). \tag{2}$$

We first try to compute the value of $P(i)$. We have two cases, when Alice sends ρ_1 and when Alice sends ρ_2. We can simply add the probability of each getting a measurement outcome i together. Then we have

$$\begin{aligned} P(i) &= P(i|\rho_1)P(\rho_1) + P(i|\rho_2)P(\rho_2) \\ &= a_i r + b_i(1-r). \end{aligned}$$

We shift our attention to the Shannon entropy of a given measurement outcome. To get $P(\rho_1|i)$, because we know the value of $P(i|\rho_1)$, $P(\rho_1)$, and $P(i)$, we can make use of Bayes'

Rule, which relates the probability of an event to the condition related to the event.

By plugging into Bayes' Rule with A as Alice sending ρ_1 and B as getting measurement result i, we get that

$$\begin{aligned} P(\rho_1|i) &= \frac{P(i|\rho_1)P(\rho_1)}{P(i)} \\ &= \frac{a_i r}{a_i r + b_i(1-r)}. \end{aligned}$$

Thus, we now have

$$H_2(P(\rho_1|i)) = H_2\left(\frac{a_i r}{a_i r + b_i(1-r)}\right).$$

Plugging into Equation 2, our expected value of entropy after the measurement becomes

$$\mathbb{E}[H'] = \sum_{i=1}^{n-1}((a_i \cdot r + b_i\,(1-r))\,H_2\left(\frac{a_i r}{a_i r + b_i\,(1-r)}\right).$$

Our accessible information is $I_{acc} = H(r) - \mathbb{E}[H]$. To prove that this quantity is concave in the probability in the probability of states, we want to show that the second derivative of I_{acc} is always negative. Our second derivative is

$$\frac{d^2 I_{acc}}{dr^2} = C \cdot \left(\frac{1}{(r-1)r} + \sum_i \frac{a_i b_i}{(r-1)r(b_i(r-1) - a_i r)}\right),$$

where C is some positive constant. Because we only care about the sign of the derivative, we can drop C. Now we want to show that

$$\frac{1}{(r-1)r} + \sum_i \frac{a_i b_i}{(r-1)r(b_i(1-r) - a_i r)} < 0.$$

If $r = 0$ or 1, Alice sends one state to Bob with certainty, in which case there is no information to be gained. Because r is a probability, then we have that $0 < r < 1$, and $\frac{1}{(r-1)r} < 0$. Factoring out $\frac{1}{(r-1)r}$, we want to show

$$1 + \sum_i \frac{a_i b_i}{(b_i(r-1) - a_i r)} > 0,$$

which is equivalent to

$$1 - \sum_i \frac{a_i b_i}{(b_i(1-r) + a_i r)} > 0.$$

Notice that the denominator is a weighted arithmetic average of a_i and b_i, and the numerator is the product of a_i and b_i. Since r, $1-r$, a_i, and b_i are real and positive, and $(1-r)+r = 1$, we use the weighted form of AM-GM to get that

$$b_i(1-r) + a_i r \geq b_i^{1-r} a_i^r,$$

where the equality holds only when $a_i = b_i$ for all i. Thus, we know that

$$\begin{aligned} 1 - \sum_i \frac{a_i b_i}{(b_i(1-r) + a_i r)} &\geq 1 - \sum_i \frac{a_i b_i}{b_i^{1-r} a_i^r} \\ &= 1 - \sum_i a_i^{1-r} b_i^r. \end{aligned}$$

Now we can use AM-GM again, giving us

$$a_i^{1-r} b_i^r \leq (1-r)a_i + r b_i,$$

where the equality holds only when $a_i = b_i$ for all i. This means

$$\begin{aligned} 1 - \sum_i a_i^{1-r} b_i^r &\geq 1 - \sum_i \left((1-r)a_i + r b_i\right) \\ &= 1 - \left((1-r)\sum_i a_i + r \sum_i b_i\right) \\ &= 1 - (1 - r + r) \\ &= 0. \end{aligned}$$

Thus, we've proven that $\frac{d^2 I_{acc}}{dr^2} \leq 0$, where the equality only holds when for all i, $a_i = b_i$, in which case the two mixed states are identical or completely indistinguishable in our poorly chosen measurement. In either case, no information can be communicated between Alice and Bob. □

4 Completing the Proof with Shor's Argument

Now we provide a qualitative and brief summary of Shor's proof in Section 6 of [3] completing the step from the concavity of accessible information for two mixed states to the statement that $C_{1,1} = C_{1,A}$.

For a signal S sent to Bob, there are two cases where Bob's understanding of the probability distribution of S in ρ_1 and ρ_2 change: a measurement on S, which we call a direct measurement step or a measurement on some other signal S' that correlates to the state S, which we call a refinement step. It is important to notice that additional information for S is gained in the direct measurement, but not for the refinement. This is because the change in probability distribution in S is solely caused by the correlation between the two signals, and such correlation exists independent of our measurement strategy, so the information gain is already accounted for in S'. Now consider the last time where we perform a refinement step, meaning we only perform direct measurements after this, which is always possible since our strategy is finite.

After the last refinement step, the weighted average of the new probability distributions is the probability before the refinement step. Because of our Lemma 3.2 on the concavity of the I_{acc}, the weighted average of the accessible information gain of the measurement steps after the refinement would be thus no greater than the accessible information gain if the refinement step is replaced by a direct measurement step. Then, we can induct this process by replacing the last refinement step with a measurement step until no refinement step is ever used. This would give us a strategy with no less than the capacity of our original $C_{1,A}$, and because we know all the steps are measurement steps, the strategy collapses to our desired $C_{1,1}$.

From here we finish our proof and conclude that $C_{1,1} = C_{1,A}$ for two mixed states in arbitrary dimensions.

5 Conclusion

In this paper, we first proved that for any two mixed states, the accessible information is concave. Then by using this lemma, we conclude that the one-shot capacity $C_{1,1}$ is equal to the adaptive one-shot capacity $C_{1,A}$ for any mixed states in arbitrary dimensions, extending Shor's initial result for pure states. It tells us that when distinguishing any two states, the adaptive strategy in measuring each state individually doesn't help us in getting more information.

For future exploration, we can try to prove other conjectures in [3], where Shor proposed another probable sufficient condition for $C_{1,1} = C_{1,A}$, which states the following

Conjecture 2 (Shor)**.** For an arbitrary set of pure states in two dimensions, $C_{1,1} = C_{1,A}$, and in fact, this capacity is achievable by using as signal states in the ensemble with inner product closest to 0.

It might also be true that

Conjecture 3. For an arbitrary set of mixed states in two dimensions, $C_{1,1} = C_{1,A}$.

Another more general and probably more difficult task would be to find all the necessary conditions for $C_{1,1} = C_{1,A}$. Additionally, we can also investigate the tightness of the bound between $C_{1,A}$ and $C_{1,\infty}$.

6 Acknowledgments

I would like to express my deepest gratitude to my mentor Andrey Boris Khesin for his enlightening answers to every question of mine and guidance in every step of the way of this project. Thank you to Prof. Shor for his fascinating work and engaging discussion with me. I want to thank Dr. Tanya Khovanova and my tutor Peter Gaydarov for their invaluable and

detailed suggestions throughout the project. Thank you to Kenneth Choi for his comments, and Lucy Cai and Ishan Khare for teaching me Latex. I would like to thank Naomi Kenyatta and the Nahomies for supporting me. Thank you to Dr. David Jerison and Dr. Ankur Moitra for arranging the mentorship for me. Finally, thank you to CEE, MIT, and Dr. Jay Ding for allowing me to attend RSI.

References

[1] P. W. Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th annual symposium on foundations of computer science*, pages 124–134. Ieee, 1994.

[2] S. Gossett. 8 quantum computing applications examples, 2020.

[3] P. W. Shor. The adaptive classical capacity of a quantum channel, or information capacities of three symmetric pure states in three dimensions. *IBM Journal of Research and Development*, 48(1):115–137, 2004.