IP Current Developments in Mathematics, 2005

The Additivity Conjecture in Quantum Information Theory

Peter W. Shor

Abstract. In this paper, I discuss the additivity conjecture in quantum information theory. The additivity conjecture was originally a set of at least four conjectures. These conjectures said that certain functions of quantum states and channels were additive under tensor products. While some of these conjectures were previously known to be stronger than others, they have recently all been proved equivalent. This conjecture is a very intriguing mathematical question which the best efforts of a large number of quantum information theorists have not been able to resolve for nearly a decade. It is a mathematically elegant question that is one of the most important open questions in the field of quantum information and computation. This paper then is intended to be both an exposition of the conjecture, its background, and some of the methods that have been used to yield partial results for it, as well as a plea for help in resolving this conjecture. Very recently (summer 2007), substantial progress has been made on this conjecture, in that counterexamples to a set of stronger conjectures have been found. These will be described briefly.

1. Introduction

Earlier in this decade, the title of this paper would have been ambiguous. Until 2003, there was not any one problem which could have been titled "The additivity conjecture." Rather, there were a host of different additivity questions, which we will describe in more detail later in this manuscript. Two major open additivity questions were the additivity of *classical channel capacity* and the additivity of *entanglement of formation,* but researchers had also conjectured the additivity of *the minimum entropy output of a quantum channel,* and conjectured a stronger property than additivity of entanglement of formation called *the strong superadditivity* of entanglement of formation. In 2003, it was shown that these four open conjectures were all equivalent [**29, 33**].

This unified four seemingly separate conjectures into one unified conjecture. Holevo [**17, 18**] has written recent, thorough surveys of the state of affairs of the additivity question. There, however, has been significant progress since then [**37, 13, 10**] as will be discussed later in this paper. Several other quantities had earlier been conjectured to be additive, including the *relative entropy of entanglement,* the *distillable entanglement.* These now both appear to be non-additive; the first has been rigorously proved so [**35**], and for the second there is strong evidence of non-additivity [**34**]

The question of the additivity of a quantity f, which is a function defined over quantum channels (or quantum states) can be expressed as follows. Suppose Φ and Ψ are two quantum channels (or states):

$$\text{Is } f(\Phi \otimes \Psi) = f(\Phi) + f(\Psi)?$$

where $\Phi \otimes \Psi$ is the tensor product of Φ and Ψ. In general, and certainly for all these quantities mentioned above, inequality in one direction is trivial, while the other may be quite difficult.

2. Background

Before we can describe this in more detail, we need to give some definitions. We will be dealing solely with the quantum mechanics of finite dimensional systems in this paper. This is because it is easier to deal with finite dimensions than with infinite dimensions and also because , at least for the channel capacity question, the additivity conjecture for infinite dimensional systems is implied by the conjecture for finite-dimensional systems [**31**].

In quantum mechanics, a d-dimensional system can be represented by a complex vector space $\mathbb{C}^d$. A *pure quantum state* of the system is a unit complex vector in this vector space. Quantum states are often represented as *kets*, $|\, v\rangle$, although in this paper we often omit the kets, yielding a notation which is more familiar to mathematicians. We still use $|\, 0\rangle$, $|\, 1\rangle$, $\ldots$, for basis states of a quantum system. We will be using the generally accepted convention in physics of representing quantum states by column vectors.

Two pure quantum states are equivalent if they differ only by a phase, so mathematically, a quantum state lies in a projective complex vector space. In many calculations, however, it is much more convenient to represent a state as a unit complex vector, and we will be doing this. A *mixed quantum state,* or *density matrix* is a trace 1 positive[1] $d \times d$ matrix. A pure state v is equivalent to the rank one density matrix $vv^\dagger$. If we have a finite ensemble of quantum states, where the probability of

[1] In this paper, by positive, I will mean positive semidefinite Hermitian matrix.

v_i is p_i, then the density matrix associated with this ensemble is

$$\sum_i p_i v_i v_i^\dagger$$

One example of quantum states with which many people are familiar is the polarization states of light. This is a two-dimensional quantum system (which quantum information scientists have started calling a qubit) and thus has two basis vectors. For example, we can take the vertically and horizontally polarized photons a basis. We call these states

$$|\updownarrow\rangle \quad \text{and} \quad |\leftrightarrow\rangle$$

Any other pure state of polarization is a linear combination of these states. For example,

$$|\nearrow\rangle = \frac{1}{\sqrt{2}}(|\updownarrow\rangle + |\leftrightarrow\rangle) \tag{1}$$

$$|\searrow\rangle = \frac{1}{\sqrt{2}}(|\updownarrow\rangle + |\leftrightarrow\rangle) \tag{2}$$

and right and left circular polarization states are the linear combinations with imaginary coefficients $\frac{1}{\sqrt{2}}(|\updownarrow\rangle \pm i\,|\leftrightarrow\rangle)$.

When you consider two quantum systems, their joint state space is represented by the tensor product of the two individual state spaces. This joint state space includes states which are not tensor products. These states are said to be *entangled*. For example, the state

$$\frac{1}{\sqrt{2}}\left(|\updownarrow\rangle \otimes |\leftrightarrow\rangle - |\leftrightarrow\rangle \otimes |\updownarrow\rangle\right)$$

is such a state, this one called an EPR pair (after Einstein, Podolsky and Rosen's famous thought experiment [**7**], which involved a similar physical state of two entangled particles).

Density matrices arise from pure quantum states in two ways. The first, as was described above, is if we have a quantum system about which we have incomplete information. If this quantum system is in state v_i with probability p_i, then the density matrix is

$$\rho = \sum_i p_i v_i v_i^\dagger$$

The second way is if we have an entangled system, and we discard or ignore one part. To describe this mathematically, we need to introduce the partial trace operator. Suppose we have a density matrix ρ over a tensor product vector $A \otimes B$. If we discard system B, this corresponds to taking the partial trace over B of the state ρ, written as $\mathrm{Tr}_B\rho$. The partial trace for tensor product states $\rho_A \otimes \rho_B$ is defined as

$$\mathrm{Tr}_B(\rho_A \otimes \rho_B) = \rho_A \, \mathrm{Tr}\rho_B$$

and it is defined for arbitrary states ρ by linearly extending this definition.

Any two orthogonal quantum states are distinguishable, meaning that there is some theoretical measurement which distinguishes perfectly between these two states. If two quantum states are not orthogonal, then no such measurements exist, although many measurements may give some probability of distinguishing between these two states. The most general quantum measurement is called a POVM (positive operator valued measurement). In this paper, we will restrict consideration to POVM's with a finite number of outcomes. Such POVM's can be represented by a set of positive Hermitian matrices, E_1, E_2, …, E_k, where each matrix is associated with a different outcome of the measurement. The one constraint on these matrices is that they sum to the identity, i.e., $\sum_i E_i = I$. Given a pure quantum state v, the probability of outcome i of the measurement is

$$\mathrm{Prob}(i) = v^\dagger E_i v.$$

For a mixed quantum state ρ, the probability of outcome i is

$$\mathrm{Prob}(i) = \mathrm{Tr} \rho E_i.$$

The constraint that their sum is the identity is equivalent to the constraint that the sum of the probabilities of the outcomes of a measurement is 1.

A valid map on quantum states allowed by physics is a trace preserving completely positive (TPC) map Φ, taking density matrices to density matrices. Here, trace preserving means

$$\mathrm{Tr}\Phi(\rho) = \mathrm{Tr}\rho,$$

positive means that Φ maps positive matrices to positive matrices, and completely positive means that $\Phi \otimes I_d$ is positive for the identity map I_d on a d-dimensional quantum system, for all d. There are two alternative characterizations of trace preserving completely positive maps. The first is the Krauss operator sum representation. Any trace-preserving completely positive operator can be written as

$$\Phi(\rho) = \sum_k A_k \rho A_k^\dagger, \quad \text{where} \quad \sum_k A_k^\dagger A_k = I.$$

The second is the Stinespring representation. This says that any d-dimensional TPC map can be represented as the composition of the following three operations. First, taking the tensor product of this system with a fixed state v_0 in a Hilbert space of dimension at most $d_{\mathrm{in}} \cdot d_{\mathrm{out}}$. Second, making a unitary transform U on the resulting state. Finally, taking the partial trace to leave a d_{out} dimensional subsystem. That is, Φ can be represented as the composition of operations:

$$\rho \rightarrow \rho \otimes v_0 v_0^\dagger \rightarrow U(\rho \otimes v_0 v_0^\dagger) U^\dagger \rightarrow \mathrm{Tr}_2\, U(\rho \otimes v_0 v_0^\dagger) U^\dagger.$$

3. The maximum purity conjectures

The minimun entropy output of a quantum channel Φ is defined as

$$H_{\min}(\Phi) = \min_{\rho} H(\Phi(\rho)).$$

Because of the convexity of the entropy function, and the linearity of Φ, it is immediate that this minimum is achieved on the boundary of the set of density matrices, i.e., on a rank-one density matrix $\rho = vv^{\dagger}$.

The additivity question in this case is whether

$$H_{\min}(\Phi \otimes \Psi) = H_{\min}(\Phi) + H_{\min}(\Psi)$$

The $\leq$ is easy to prove. Let the states ρ_0 and ρ_1 be those where Φ and Ψ achieve their minimum entropy. Then

$$H(\,(\Phi \otimes \Psi)(\rho_0 \otimes \rho_1)) = H(\Phi(\rho_0)) + H(\Psi(\rho_1)).$$

The minimum entropy output seems like it may be the simplest of the various equivalent additivity conjectures to prove. Indeed, the proofs that the other conjectures imply additivity of $H_{\min}$ are all quite easy, while the proofs in the other direction are substantially harder. Amosov, Holevo and Werner [**1**] have come up with a very natural conjecture on p-norms of channels that would imply the minimum entropy output conjecture. This is called the output purity conjecture. This is a means by which substantial progress has been made on the minimum entropy output conjecture for special classes of channels. Very recently, this conjecture has been shown to be false [**37, 13**]. However, the counterexample does not appear to extend to give a counterexample to the additivity conjecture for entropy.

We now describe the output purity conjecture. We first define a p-norm on a quantum channel Φ.

$$\nu_p(\Phi) = \max_{\rho} ||\Phi(\rho)||_p$$

where the optimization is over density matrices, i.e., trace 1 positive matrices, over the input space of the channel. Here $||\cdot||_p$ is the p-norm $||\sigma||_p = (\operatorname{Tr}\sigma^p)^{\frac{1}{p}}$. Again, because of convexity, the ρ yielding maximum ν_p will always be a pure state for any channel Φ. This is related to the minimum output entropy because of the following theorem, shown in [**1**]

$$\lim_{p\downarrow 1} \nu_p(\Phi) = H_{\min}(\Phi)$$

which can be proved using the fact that

$$\lim_{p\downarrow 1} \frac{x - x^p}{p-1} = -x \log x.$$

This theorem implies that if the following conjecture is true for $1 < p < 1+\epsilon$ for some positive ϵ, the minimum entropy output is additive

$$\nu_p(\Phi \otimes \Psi) = \nu_p(\Phi)\nu_p(\Psi). \tag{3}$$

This was realized by Amosov, Holevo and Werner, who then conjectured that this minimum output purity was multiplicative for all p [**1**]. This is equivalent to the additivity of minimum output Rényi entropy of order p. Unfortunately, their original conjecture was not true; two of the proposers of this conjecture found a counterexample not long after their original paper [**36**]. The counterexample shows that the conjecture is false for $p \geq 4.8$. There did not seem to be any obstacle to it holding for $1 < p \leq 2$, which was the revised version of the conjecture. However, this has recently also shown to be false. Winter [**37**] gave a counterexample which showed that the conjecture was false for $p > 2$. The structure of this counterexample led Hayden [**13**] to find a counterexample for all p, $1 < p < 2$. Shortly afterwards, Harrow [**10**] found a counterexample for the additivity of Rényi entropy at p sufficiently close to 0, which numerically seems to work for $0 \leq p < 0.12$. (This is equivalent to multiplicativity of minimum output p-norm.) Thus, if the conjecture holds, it most likely holds only for the von Neumann entropy case of $p = 1$.

We now describe the structure of Winter's counterexample [**37**], which holds for $1 < p < 2$. We take m random unitary transformations over an n-dimensional Hilbert space, $U_1, \ldots, U_m$. Now, let

$$\Phi : \rho \to \frac{1}{m} \sum_{j=1}^{m} U_m \rho U_m^\dagger$$

and

$$\Phi^* : \rho \to \frac{1}{m} \sum_{j=1}^{m} U_m^\dagger \rho U_m.$$

It can be shown that if m grows faster than $O(n \log n)$, then $\Phi(vv^\dagger)$ and $\Phi^*(vv^\dagger)$ are close to random for all v, so for Φ and Φ^*, all the eigenvalues of any output density matrix are of order $1/n$. However, if we consider the EPR state

$$w = \frac{1}{\sqrt{n}} \sum_i |\, i\rangle \,|\, i\rangle\,,$$

then $\Phi \otimes \Phi^*(ww^\dagger)$ will have one large eigenvalue. This is because for any for any unitary U,

$$\begin{aligned} w^\dagger(U \otimes U^\dagger)w &= \frac{1}{n}(\sum_j \langle j\,| \otimes \langle j\,|) \sum_i U\,|\, i\rangle \otimes U^\dagger\,|\, i\rangle) \\ &= \frac{1}{n} \sum_{i,j} \langle j\,|\, U\,|\, i\rangle \otimes \langle j\,|\, U^\dagger\,|\, i\rangle \\ &= \frac{1}{n} \sum_{i,j} |\langle j\,|\, U\,|\, i\rangle|^2 \\ &= 1, \end{aligned}$$

where the last inequality holds because $U\,|\,i\rangle$ and $|\,j\rangle$ are two different orthonomal bases for the Hilbert space.

The above equality shows that the terms where $i = j$ in the expression

$$w^\dagger \Big(\Phi \otimes \Phi^*(ww^\dagger) \Big) w = w^\dagger \left(\frac{1}{m^2} \sum_{i,j} U_i \otimes U_j^\dagger ww^\dagger U_i^\dagger \otimes U_j \right) w$$

contribute a total of $1/m$. Since the other terms are positive, this shows that $(\Phi \otimes \Phi^*)(ww^\dagger)$ has an eigenvalue larger than $1/m$. If we choose $m = n^{1+\epsilon}$ for ϵ small, this will give a counterexample to the maximum purity conjecture for any $p > 2$, since the maximum p-norms of each of Φ and Φ^* are approximately $n^{(1-p)/p}$, and the maximum p-norm of $\Phi \otimes \Phi^*$ is approximately $1/m$.

Inspired by this result, Hayden [**13**] found a counterexample to the conjecture for any $p > 1$. This is not quite so easy to understand intuitively, and the calculations are somewhat more difficult, but the basic idea behind the counterexample is the same. Here, the two channels are

$$\Phi(\rho) = \mathrm{Tr}_B\; U(\rho \otimes v_0 v_0^\dagger) U^\dagger$$

and

$$\Phi^*(\rho) = \mathrm{Tr}_B\; U^\dagger(\rho \otimes v_0 v_0^\dagger) U,$$

where U is a random unitary transform.

4. The other equivalent conjectures

In this section, we will describe several other conjectures equivalent to the additivity of $H_{\min}$ There are at least six other equivalent conjectures [**33, 29, 27, 9**], of which we will mention four. These deal with two quatitites. The first of these quantities is the *entanglement of formation* of a quantum state, and the second of these is the *classical information capacity* of a quantum channel.

A pure quantum state in a tensor product space $A \otimes B$ is said to be *entangled* if it cannot be written as a tensor product $\phi \otimes \psi$ where $\phi \in A$ and $\psi \in B$. Entanglement is a very important concept in quantum information theory. It is essentially the source of Einstein's distress over what he called "spooky action at a distance," and is also responsible for Bell's inequality, and protocols for quantum information transmission such as quantum teleportation, quantum superdense coding, and so forth. It is thus an important question as to how to quantify entanglement.

The canonical example of an entangled state is an EPR pair. This is a system of two qubits in the state

$$\psi_{EPR} = \frac{1}{\sqrt{2}}(|\,01\rangle - |\,10\rangle),$$

or any state that can be derived from this state by *local* unitary transformations. That is, if we have two unitary transformations on the two Hilbert spaces U_A acting on A and U_B acting on B, then $(U_A \otimes U_B)\psi_{EPR}$ is also an EPR pair. By convention, we take the amount of entanglement in a single EPR pair to be one bit.

It is fairly easy to quantify entanglement of pure quantum states (recall these are vectors in our tensor product Hilbert space). The amount of entanglement in a pure quantum state v consists of

$$E_P(v) = H(\mathrm{Tr}_A vv^\dagger) = H(\mathrm{Tr}_B vv^\dagger).$$

It is a theorem [**6**] that any two pure quantum states with the same amount of entanglement can be asymptotically interconverted with high efficiency, using only local unitary operations and classical communication. That is, suppose we have a pure quantum state v. We denote the quantum state of an EPR pair, ψ_{EPR} above, by ψ. Then for any fixed small $\epsilon > 0$, there is a large enough n so that, if two people, Alice and Bob, hold A and B respectively, then Alice and Bob, using only local quantum operations and classical communication, can convert $v^{\otimes n}$ to a state ρ that is very close to $\psi^{\otimes(1-\epsilon)nE_P}$, i.e., such that

$$\psi^{\dagger \otimes(1-\epsilon)nE_P} \rho \, \psi^{\otimes(1-\epsilon)nE_P} > 1-\epsilon$$

where E_P is the pure state entanglement of v. We also have the reverse theorem, where we interchange the role of v and ψ above, so that we start with nE_P EPR pairs: $\psi^{\otimes nE_P}$ and we end with a state ρ that is very close to $v^{\otimes(1-\epsilon)n}$.

We call a transformation that can be performed using local quantum operations and classical communication a LOCC operation. In this model, we are allowed to perform local quantum operations conditioned on classical variables, i.e., on the outcome of measurements or on classical information that Alice has received from Bob (or vice versa). This is thus a class of transformations that is quite difficult to characterize.

When we try to prove analogous results for mixed states, that is, we try to quantify the entanglement in a mixed state ρ, something unfortunate happens. There is no longer a single good measure of entanglement, as there was with pure states. We can look at the number of nearly perfect EPR pairs we can obtain from many copies of a mixed state ρ, analogous to the theorem we stated for pure state entanglement above. This gives a quantity is known as distillable entanglement, E_D. We can also define the amount of pure entanglement asymptotically required to create a mixed state ρ using LOCC operations. This gives a quantity called entanglement cost, known as E_C. In general, these are not the same for mixed states. There are *bound entangled* states ρ with $E_D(\rho) = 0$, so no states close to an EPR pair can be created out of any number of copies of these states, but still with the entanglement cost $E_C(\rho) > 0$ [**20**]. We do not have any nice formulas for distillable entanglement E_D, but there is a formula for entanglement cost E_C. Let us define entanglement of formation by

$$E_F(\rho) = \min_{\{p_i, v_i\}} \sum_i p_i E_P(v_i)$$

where the minimum is taken over ensembles of pure states whose density matrix is ρ, that is

$$\sum_i p_i v_i v_i^\dagger = \rho.$$

It is then known [**14**] that

$$E_C(\rho) = \lim_{n \to \infty} \frac{1}{n} E_F(\rho^{\otimes n}).$$

The question of whether entanglement of formation is equal to entanglement cost, that is, whether we really need the limit in the above expression, is equivalent to the question of whether entanglement of formation is additive, that is, whether for two density matrices ρ_1 and ρ_2,

$$E_F(\rho_1 \otimes \rho_2)) = E_F(\rho_1) + E_F(\rho_2).$$

The structure of the states may need a little more explanation here. We have four Hilbert spaces, A_1, A_2, B_1 and B_2, and we work in the tensor product $A_1 \otimes A_2 \otimes B_1 \otimes B_2$. We have $\rho_1 \in \mathcal{M}(A_1 \otimes B_1)$, $\rho_2 \in \mathcal{M}(A_2 \otimes B_2)$. The definition of the entanglement of formation assumes that the quantum space is a tensor product of two systems. In each case, this tensor product defining the entanglement of formation is between the "A" and the "B" compenents of the system.

There is a stronger conjecture about entanglement of formation than the additivity. It is called the *strong superadditivity* of entanglement of formation [**35, 3**]. In this case, we have an arbitrary entangled state ρ in a Hilbert space which is the tensor product of four systems, which we again designate as A_1, B_1, A_2, B_2. The conjecture is

$$E_F(\rho) \leq E_F(\mathrm{Tr}_2 \rho) + E_F(\mathrm{Tr}_1 \rho)$$

where the entanglement of formation is defined over the A-B split of subsystems, and the partial traces are over the 1-2 split of subsystems. Strong superadditivity of E_F is easily seen to imply additivity of E_F [**35**], whereas it is harder to prove the reverse direction.

We now turn to the definition of the classical capacity of a quantum channel. Recall the definition of the capacity of a classical noisy channel. A classical noisy channel can be thought of as a stochastic map from a random variable X representing the input of a channel to a random variable Y representing the output of the channel. Shannon's theorem says that a noisy channel N has the capacity

$$\max_{p(x)} I(X;Y) = H(Y) - H(Y|X)$$

where $I(X;Y)$ is the mutual information between the ramdom variables X and Y. We have $H(Y)$ is the entropy of the random variable Y, and $H(Y|X)$ is the conditional entropy of the random variable Y. This is defined as

$$H(Y|X) = \sum_x \text{Prob}(X = x) H(Y|X = x)$$

where $H(Y|X = x)$ is the entropy of the output Y when the input is x. One can think of this definition of the mutual information of a channel as the entropy of the average output minus the average entropy of the output.

One might think that the classical capacity of a quantum channel Φ would be the maximum over all input distributions X of quantum states, and over all measurements of the output, of the mutual information between random variable X representing the input and the random variable Y representing the measurement outcome. This quantity is called the *accessible information* for a quantum channel Φ, which we denote by $I_A(\Phi)$. This guess is not correct; while one can clearly achieve the accessible information, it is not an upper bound. The accessible information can be exceeded by using joint measurements of the output state across several uses of a quantum channel. To get the true classical capacity of a quantum, one must take

$$\lim_{n\to\infty} (\frac{1}{n} I_A \Phi^{\otimes n}),$$

the limit of $\frac{1}{n}$ times the accessible information of the tensor power of n uses of the channel. This formula allows both entangled inputs to the channel and joint measurements of the channel. To obtain this limit, one can first take the limit for the joint measurements, and then the limit for the entangled inputs. If we allow joint measurements, but only unentangled inputs, one can achieve the following capacity, which was formerly called the *Holevo bound* but is now often called the *Holevo capacity* or *Holevo formula.* This was long known merely as an upper bound for the accessible information [**15**]. It is now known to be the capacity as long as the protocols are not allowed to use entangled inputs [**16, 30**].

$$\chi(\Phi) = \max_{\{p_i, v_i\}} H(\Phi(\sum_i p_i v_i v_i^\dagger)) - \sum_i p_i H(\Phi(v_i v_i^\dagger))$$

where the maximum is taken over all probability distributions p_i on pure states v_i in the input space of the channel. This is again the entropy of the average output minus the average entropy of the output.

Recall that the classical capacity of a quantum channel is the limit of the accessible information over protocols using both entangled inputs and joint measurements on the output. Taking the limit just over joint measurements of the output gives us the Holevo capacity. To get the true classical capacity, we need to then take the limit of the Holevo capacity over n uses of the channel, as n goes to ∞. We thus get the formula

$$\text{Capacity}(\Phi) = \lim_{n\to\infty} \frac{1}{n}\chi(\Phi^{\otimes n})$$

If the Holevo capacity is additive, then this limit is the same as the Holevo capacity. This, of course, leads us to the last conjecture, whether

$$\chi(\Phi \otimes \Psi) = \chi(\Phi) + \chi(\Psi)$$

5. Partial results on the output purity and minimum entropy conjectures.

In this section, I will try to sketch the techniques used to prove partial results on these conjectures. There are several cases for which these conjectures are proved. The most imporant of these may be unital qubit channels [**21**], depolarizing channels [**8, 22**], and entanglement breaking channels [**23, 32**]. The proofs here are often fairly complicated, and have a lot of details which may obscure the essential features of these proofs.

An entanglement breaking channel is one which breaks the entanglement of any input state with a reference system. That is, Φ is an entanglement breaking channel

$$I \otimes \Phi(vv^\dagger)$$

never has entanglement between the output space of I and that of Φ. A unital qubit channel Φ is one which takes a 2-dimensional quantum space to a 2-dimensional quantum space, and for which $\Phi(I) = I$. Finally, a depolarizing channel mixes a quantum state ρ with the maximally mixed state I/d. That is, the depolarizing channel with parameter λ, which we call Δ_λ, takes

$$\Delta_\lambda(\rho) = \lambda\rho + (1-\lambda)I/d$$

where d is the dimension of the input (and output) space of the channel.

Two of the key tools for these proofs are inequalities. When dealing with von Neumann entropy, one of the most important tools is the strong subadditivity of entropy [**24**]. This says that for a quantum system in a mixed state in a tensor product of three systems, $\rho = \rho_{ABC}$, we have

$$H(\rho_{AB}) + H(\rho_{AC}) \geq H(\rho_A) + H(\rho_{ABC})$$

where the density matrices for subsystems are obtained by taking the partial trace of ρ, that is, $\rho_{AB} = \text{Tr}_C\rho$, $\rho_{AC} = \text{Tr}_B\rho$, and $\rho_A = \text{Tr}_{BC}\rho$.

The other inequality, which has been useful for proving theorems involving output purity, is the Lieb-Thirring inequality[2] [**25, 21**], which says that for two matrices,

$$\text{Tr}(A^{1/2}BA^{1/2})^p \leq \text{Tr}A^{p/2}B^pA^{p/2}$$

[2] There are actually two inequalities, quite different, which are both known as the Lieb-Thirring inequality. This is one of them.

for $p \geq 1$. These inequalities do not appear to be related at first sight, but there may be some connection between them, as the additivity of entropy for entanglement breaking channels was proved using strong subadditivity [**32**], while the output purity conjecture for p-norms in entanglement breaking channels was proved by the Lieb-Thirring inequality [**23**].

I will now try to sketch the proofs of two of these theorems. First, we show the additivity entropy for entanglement breaking channels.

We consider two channels. Let Ψ be an arbitrary channel acting on space A, and Φ be an entanglement breaking channel acting on space B. Now, consider a state ρ_{AB}. We know (by the definition of entanglement breaking channel) that $I \otimes \Phi(\rho_{AB}$ is separable, so we have

$$I \otimes \Phi(\rho_{AB}) = \sum_j q_j v_j v_j^\dagger \otimes w_j w_j^\dagger$$

Here, the v_j are in system A and the w_j are in system B. We let σ_{AB} be

$$\Psi \otimes \Phi(\rho_{AB}) = \sum_j q - j\Psi(v_j v_j^\dagger) \otimes w_j w_j^\dagger$$

We now add a third Hilbert space, C, to the mix, and work in the space $A \otimes B \otimes C$. We define

$$\sigma_{ABC} = \sum q_j \Psi(v_j v_j^\dagger) \otimes w_j w_j^\dagger \otimes e_j e_j^\dagger$$

where the e_j are orthonormal vectors in system C.

We now have a tensor product of three systems, and can apply the strong subadditivity of entropy to σ_{ABC}

$$H(\sigma_{AB}) \geq H(\sigma_{ABC}) - H(\sigma_{BC}) + H(\sigma_B).$$

But we find

$$\sigma_B = \Phi(\rho_B) \geq S_{min}(\Phi)$$

and

$$H(\sigma_{ABC}) - H(\sigma_{BC}) = \sum_j q_j S(\Phi(v_j v_j^\dagger) \geq S_{min}(\Psi)$$

which proves the theorem.

Note that this does not automatically prove the additivity of channel capacity if one of the two channels is an entanglement breaking channel; the proof that additivity of minimum entropy implies additivity of channel capacity invokes the additivity of $S_{\min}$ for a different channel. However, a more complicated calculation also proves the addivity of channel capacity if one channel is entanglement breaking.

The proof of multiplicativity of ν_p for depolarizing channels is more complicated. The key tool is the Lieb-Thirring inequality [**25**]. Recall that the depolarizing channel in d dimensions is

$$\Delta_\lambda(\rho) = \lambda\rho + (1 - \lambda)I/d$$

The proof uses the dephasing channel. Suppose that e_i, $i = 1 \dots d$ is an orthonomal basis for our vector space/ The dephasing channel (with dephasing parameter λ) is

$$\Phi_\lambda(\rho) = \lambda\rho + (1-\lambda)\sum_{i=1}^{d} e_i e_i^\dagger \rho e_i e_i^\dagger$$

It is fairly easy to express the depolarizing channel Δ_λ as a convex combination of dephasing channels (each having a different set of eigenvectors e_i). For the proof, we need a specific combination of dephasing channels. If our depolarizing channel is acting on a state ρ with orthonormal eigenvectors v_j, we can express Δ_λ as a convex combination of dephasing channels, each of which has eigenvectors e_i complementary to v_i. That is, $|e_i^\dagger v_i|^2 = \frac{1}{d}$. This lets us prove results on dephasing channels, and then apply them to the depolarizing channel.

The result that is proved goes as follows.

Lemma: For a dephasing channel,

$$||(\Phi_\lambda \otimes I)(\rho_{AB})||_p \leq d^{1-1/p}\nu_p(\Delta_\lambda)\left[\sum_{i=1}^{d} \mathrm{Tr}(\rho_B^{(i)})^p\right]^{1/p} \tag{4}$$

where

$$\rho_B^{(i)} = \mathrm{Tr}_A e_i^\dagger e_i \otimes I \rho_{AB}.$$

The clever step in the proof of the lemma is to factor $(\Phi_\lambda \otimes I)(\rho_{AB})$ into matrices $M_1^{1/2} M_2 M_1^{1/2}$ so that we can apply the Lieb-Thirring inequality. We then apply this lemma to the state

$$\rho_{AB} = (I \otimes \Psi)(\sigma_{AB})$$

to get multiplicativity. The last term in equation 4 together with $d^{-1/p}$ gives the bound on $\nu_p(\Psi)$, provided we select the proper decomposition of the depolarizing channel into dephasing channels. The second to last term gives the bound on $\nu_p(\Phi) = \nu_p(\Delta_\lambda)$, and the rest, combined with the details of the decomposition of the depolarizing channel Δ_λ into dephasing channels, result in the desired inequality.

6. Equivalence of the additivity conjectures

The equivalence of the additivity conjectures is shown by using several techniques, including one called channel extension. We will sketchily describe one case of this technique; for more details, see the papers [**33, 19, 9**]. What we will describe is the additivity of something called the *constrained Holevo capacity.* This is the problem of maximizing the Holevo capacity, when the average input is constrained to be a given density matrix ρ. More formally, we would like to maximize over all probability distributions p_i over density matrices ρ_i in the input space with $\sum_i p_i \rho_i$. That is, for a channel Ψ we would like to find

$$\begin{aligned} \chi_\rho(\Psi) &= \max H(\Psi(\rho)) - \sum_i p_i H(\Psi(\rho_i)) \\ & \quad \text{subject to} \quad \sum_i p_i \rho_i = \rho. \end{aligned}$$

Although we do not explain it in this paper, it is fairly straightforward that the Stinespring dilation theorem shows that the additivity of constrained Holevo capacity is equivalent to the entanglement of formation.

Since ρ is fixed in the above problem, we can just minimize the second term on the right-hand side, namely

$$\text{minimize} \quad \sum_i p_i H(\Psi(v_i v_i^\dagger)) \tag{5}$$

subject to the same constraint on the average input. This minimization problem is a linear program in the p_i, and as such, it has a dual linear program which has the same optimum value. This program is

$$\begin{aligned} &\text{maximize} \quad \text{Tr } \tau\rho \\ &\text{subject to} \quad v^\dagger \tau v \le H(\Psi(vv^\dagger)) \end{aligned} \tag{6}$$

where the maximization is over Hermitian matrices τ, and the constraint (6) holds for all for all vectors v in the input space. The linear programming duality theorem says that for states v_i which have non-zero probability in the optimum for probability distribution for Eq. (5), the dual constraint Eq. (6) holds with equality. We will call such states *signal states.*

Now, suppose we could find a new channel Φ such that for any vector v in the input space,

$$H(\Phi(vv^\dagger)) = H(\Psi(vv^\dagger)) + C - v^\dagger \tau v. \tag{7}$$

Then, for signal states v_i,

$$H(\Phi(v_i v_i^\dagger)) = H(\Psi(v_i v_i^\dagger)) + C - v_i^\dagger \tau v_i = C.$$

holds with equality, while for other states v,

$$H(\Phi(vv^\dagger)) \le C.$$

Thus, the signal states for this channel are exactly the minimum entropy output states.

Now, suppose we had two channels Ψ_1 and Ψ_2, and we could find a Φ_1 and Φ_2 as described above. A calculation then shows the additivity of minimum entropy output for Φ_1 and Φ_2 would imply the additivity of constrained Holevo capacity for Ψ_1 and Ψ_2.

Unfortunately, we cannot actually construct a channel Φ which satisfies Eq. (7) above. However, we can find a channel that nearly satisfies this. In fact, this is very closely related to the original channel. We choose some POVM with two elements E, $I-E$, which we will determine later. Now, we flip a coin which comes up heads with probability q. If it is heads, we send the input through the original channel Ψ. If it is tails, we measure the input with the POVM, and output either k completely random bits or a specified pure state. We also send an indicator state which tells which of the three above possibilities occured. It is straightforward to show that for an input state v, the output of this channel has entropy

$$H(\Phi(vv^{\dagger})) = qH(\Psi(vv^{\dagger})) + C + (1-q)k\mathrm{Tr}E\rho + (1-q)H_2(TrE\rho)$$

where $H_2(x) = -x\log x - (1-x)\log(1-x)$ is the binary entropy function, and C is the constant $H_2(q)$. By letting k go to ∞, and adjusting q and E appropriately, we can approach the desired channel extension (7) arbitrarily closely, and this construction lets us prove that additivity of mininimum entropy output implies additivity of constrained Holevo capacity and thus additivity of entanglement of formation.

The technique of creating the new channel Φ from Ψ is known as channel extension, and this technique, applied somewhat differently, is an important component in several of the other reductions contained in the proof of equivalence.

References

[1] G.G. Amosov, A.S. Holevo, R.F. Werner, *On some additivity problems in quantum information theory*, Problems in information transmission **36** 25–34 (2000); arXiv e-print math-ph/0003002.

[2] S. Arimoto, *An algorithm for calculating the capacity of an arbitrary discrete memoryless channel*, IEEE Trans. Info. Theory **18** 14–20 (1972).

[3] K.M.R. Audenaert and S.L. Braunstein, *On strong superadditivity of the entanglement of formation*, Comm. Math. Phys. **246** 443–452 (2004); arXiv e-print quant-ph/0303045.

[4] K. Audenaert, F. Verstraete, B. De Moor, *Variational characterizations of separability and entanglement of formation*, Phys. Rev. A **64**, art. 052304, 2001; arXiv e-print quant-ph/0006128.

[5] H. Barnum, J.A. Smolin, and B.M. Terhal, *Quantum capacity is properly defined without encodings*, Phys. Rev. A **58** 3496–3501 (1998).

[6] C.H. Bennett, H.J. Bernstein, S. Popescu, and B. Schumacher, *Concentrating partial entanglement by local operations*, Phys. Rev. A **53** 2046–2052 (1996).

[7] A. Einstein, B. Podolsky, and N. Rosen, *Can quantum-mechanical description of physical reality be considered complete?* Phys. Rev. **47** 777–780 (1935).

[8] *Additivity of the capacity of depolarizing channels*, Physics Letters A **299** 469–475 (2002).

[9] M. Fukuda, *Simplification of additivity conjecture in quantum information theory*, Quantum Information Proc. **6** 179–186 (2007); arXiv e-print quant-ph/0608010.

[10] A. Harrow, *Counterexamples to additivity of minimum output p-Renyi entropy for p close to* 0, talk at MIT, Sept. 10, 2007, manuscript in preparation.

[11] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W.K. Wootters, *Classical information capacity of a quantum channel*, Phys. Rev. A **54** 1869–1876 (1996).

[12] M. Hayashi and H. Nagaoka, *General formulas for capacity of classical-quantum channels*, IEEE Trans. Inf. Theory **49** 1753–1768 (2003); arXiv e-print quant-ph/0206186.

[13] P. Hayden, *The maximal p-norm multiplicativity conjecture is false*, arXiv e-print quant-ph/0707.3291.

[14] P.M. Hayden, M. Horodecki, and B.M. Terhal, *The asymptotic entanglement cost of preparing a quantum state*, J. Phys. A: Math. Gen. **34** 6892–6898 (2001).

[15] A.S. Holevo, *Information theoretical aspects of quantum measurements*, Probl. Info. Transm. (USSR), **9**(**2**) 31–42 (1973) (in Russian); [translation: A.S. Kholevo, Probl. Info. Transm., **9** 177–183 (1973)].

[16] A.S. Holevo, *The capacity of the quantum channel with general signal states*, IEEE Trans. Info. Theory **44** 269–273 (1998).

[17] A.S. Holevo, *The additivity problem in quantum information theory*, in 'Proceedings of the International Congress of Mathematicians', Madrid, Spain, 2006, 999–1018, EMS, Zurich, 2007.

[18] A.S. Holevo, *Multiplicativity of p-norms of completely positive maps and the additivity problem in quantum information theory*, Uspekhi Matematicheskikh Nauk **61**(**2**) 113–152 (2006); English version: Russian Math Surveys **61**(**2**) 301–339 (2006).

[19] A.S. Holevo and M.E. Shirokov, *On Shor's channel extension and constrained channels*, Comm. Math. Phys. **249** 417–430 (2004).

[20] M. Horodecki, P. Horodecki, and R. Horodecki, *Mixed-state entanglement and distillation: Is there "bound entanglement" in nature?*, Phys. Rev. Lett. **80** 5239–5242 (1998).

[21] *Additivity for unital qubit channels*, Journal of Math. Phys. **43** 4641–4653 (2002).

[22] C. King, *The capacity of the quantum depolarizing channel*, IEEE Trans. Inform. Theory **49** 221–229 (2003); arXiv e-print quant-ph/0204172.

[23] C. King, *Maximal p-norms of entanglement breaking channels*, Quantum Inf. Comput. **3** 186–190 (2003); arXiv e-print quant-ph/0212057.

[24] E.H. Lieb and M.B. Ruskai, *Proof of the strong subadditivity of quantum mechanical entropy*, J. Math. Phys. **14** 1938–1941 (1973).

[25] E.H. Lieb and W. Thirring, *Inequalities for the moments of the eigenvalues of the Schrödinger Hamiltonian and their relation to Sobolev inequalities*, in 'Studies in Mathematical Physics', E. Lieb, B. Simon, A Wieghtman eds., 269-303, Princeton University Press, 1976.

[26] K. Matsumoto, T. Shimono, and A. Winter, *Remarks on additivity of the Holevo channel capacity and of the entanglement of formation*, Comm. Math. Phys. **246** 427–442 (2004); arXiv e-print quant-ph/0206148.

[27] K. Matsumoto, *Yet another additivity conjecture*, arXiv e-print quant-ph/0506052.

[28] S. Osawa and H. Nagaoka, *Numerical experiments on the capacity of quantum channel with entangled input states*, in 'IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences', E84-A, 2583-2590 (2001); arXiv e-print quant-ph/0007115.

[29] A.A. Pomeransky, *Strong superadditivity of the entanglement of formation follows from its additivity*, Phys. Rev. A **68** 032317 (2003).

[30] B. Schumacher and Westmoreland, *Sending classical information via a noisy quantum channel*, Phys. Rev. A **56** 131–138 (1997).

[31] M.E. Shirokov, *The Holevo capacity of infinite dimensional channels and the additivity problem*, Comm. Math. Phys. **262** 137–159 (2006); arXiv e-print quant-ph/0408009.

[32] P.W. Shor, *Additivity of the classical capacity of entanglement-breaking channels*, J. Math. Physics **43** 4334–4340 (2002).

[33] P.W. Shor, *Equivalence of additivity questions in quantum information theory*, Comm. Math. Phys. **246** 453–472 (2004).

[34] P.W. Shor, J.A. Smolin, and B.M. Terhal, *Nonadditivity of bipartite distillable entanglement follows from a conjecture on bound entangled Werner states*, Phys. Rev. Lett. **86** 2681-2684 (2001).

[35] K.G.H. Vollbrecht and R.F. Werner, *Entanglement measures under symmetry*, Phys. Rev. A **64** 062307 (2001).

[36] R.F. Werner and A.S. Holevo, *Counterexample to an additivity conjecture for output purity of quantum channels*, quant-ph/0203003.

[37] A. Winter, *The maximum output p-norm of quantum channels is not multiplicative for any $p > 2$*, arXiv e-print quant-ph/0707.0402 (2007).

MIT, Cambridge, MA 02139
E-mail address: shor@math.mit.edu