Additivity Conjectures in Quantum Shannon Theory

Alex Kazachek

Abstract

Quantum channels allow two parties to communicate by sending quantum states. Quantum Shannon theory seeks an understanding of the amount of information which may be efficiently and reliably communicated across these channels. At the heart of such questions often lies some functional f, taking in a channel and returning a number quantifying its propensity for communication. Additivity conjectures then ask if $f(\mathcal{N} \otimes \mathcal{S}) = f(\mathcal{N}) + f(\mathcal{S})$ for all channels $\mathcal{N}$ and $\mathcal{S}$, as this would often greatly simplify computations if true. In this note we examine two such f – Holevo capacity and minimum output p-entropy. We establish the connection between the two and explore some modern approaches to tackling the latter. On the way, we show how quantum channels may be viewed as subspaces and introduce the basic tools of asymptotic geometric analysis. We use those tools to prove additivity does not hold for $p > 1$.

19 December 2023

1 Introduction

A quantum channel between Alice and Bob establishes a way for Alice to send Bob a quantum state. A fundamental question is how "well" Alice and Bob may communicate using a given channel – and a second is what "well" may even mean. Shannon answered these questions for classical channels in his seminal work [Sha48], introducing the entropy of a distribution and showing the capacity of the channel is the mutual information between the distributions of Alice's inputs and Bob's outputs. Quantum Shannon theory aims to extend these classical results to quantum channels.

A short review during the formative era of these concepts was given by [Hol77]. Eventually, the notion of classical capacity fell out – the rate at which classical information which may be reliably sent through a quantum channel. For a quantum channel $\mathcal{N}$, its Holevo capacity $\chi(\mathcal{N})$ is intimately connected with the classical capacity. Namely, classical capacity is exactly equal to the regularized Holevo capacity:

$$\lim_{n\to\infty} \frac{1}{n}\chi(\mathcal{N}^{\otimes n}),$$

This is the content of the celebrated Holevo-Schumacher-Westmoreland theorem [Hol73; SW97].

Unfortunately, the presence of the limit makes computing the regularized Holevo capacity entirely intractable. This would not be a problem, however, if for any two channels $\mathcal{N}, \mathcal{S}$ we had $\chi(\mathcal{N} \otimes \mathcal{S}) = \chi(\mathcal{N}) + \chi(\mathcal{S})$. This is referred to as the additivity of the Holevo capacity, and was long suspected to be true.

Its additivity was shown in [Sho04] to be equivalent to additivity of another quantity, the minimum output (von Neumann) entropy of channel. Since [Has09] later showed the minimum output entropy is not additive, we know Holevo capacity is not either. So, we are forever burdened by this regularization.

Statements about the minimum output entropy may be rephrased into structural statements

about norms on certain subspaces, allowing [ASW10] to provide an alternative (dis)proof of its additivity by employing tools from asymptotic geometric analysis. We will not examine this exact proof, but instead a similar (though much easier) one for minimum output Rényi p-entropy. Here, $p > 1$ and the von Neumann case is approached as $p \to 1$.

1.1 Classical Information Theory

We start with a primer on classical information theory, as developed by Shannon, using [Pre16] as reference. Throughout, let X and Y be finitely-supported discrete random variables, with probability masses p_X and p_Y (with the subscript dropped if the context is clear).

Recall the Shannon entropy $H(X) = -\sum_x p(x) \log_2 p(x)$, where the sum is over the support. The base of the logarithm does not really matter, but in the classical case it is convenient to use base 2 since this corresponds to bits.

A common theme in information theory is to only worry about the "typical" messages, as in the asymptotic case all messages will be satisfactory.

Definition 1 (δ-Typical Sequence)**.** *Let $\{X_i\}_{i=1}^n$ i.i.d copies of X. Denote by $p(x_1, \ldots, x_n)$ their joint distribution. We say the sample $(x_1, \ldots, x_n)$ is δ-typical for $\delta > 0$ if*

$$H(X) - \delta \leq -\frac{1}{n} \log_2 p(x_1, \ldots, x_n) \leq H(X) + \delta.$$

Proposition 1. *For any $\varepsilon, \delta > 0$, we can find some n such that any n-symbol sequence is δ-typical with probability no less than $1 - \varepsilon$.*

Definition 2 (Rate)**.** *A compression rate R is associated with a protocol compressing some messages comprising n-symbols to nR-symbol codewords. We say R is achievable if there is asymptotically vanishing error (i.e. the probability of decompression error goes to 0 as $n \to \infty$).*

Theorem 1 (Shannon Source Coding)**.** *All compression rates greater than $H(X)$ are achievable.*

Our goal is communication across distance, not local compression. However, source coding provides a very illustrative demonstration of the role entropy plays in defining "information". Also note that $H(X) \leq 1$, only saturated if X is uniform, so this theorem tells us any non-uniform message can be compressed somehow.

A channel C allows Alice to receive some message in X and send it to Bob, who receives a message in Y. It is characterized by the conditional distribution $Y|X$. We need only the conditional distribution because Alice, in principle, gets to choose what she sends to Bob.

Definition 3 (Jointly δ-Typical)**.** *Let $\delta, \varepsilon > 0$, and let $\{X_i Y_i\}_{i=1}^n\}$ be i.i.d. according to $XY \sim p_{XY}$. We say an n-symbol message $(x_1 y_1, \ldots, x_n y_n)$ is jointly δ-typical if*

$$\begin{aligned}
2^{-n(H(X)+\delta)} &\leq p_X(x_1, \ldots, x_n) \leq 2^{-n(H(X)-\delta)} \\
2^{-n(H(Y)+\delta)} &\leq p_Y(y_1, \ldots, y_n) \leq 2^{-n(H(Y)-\delta)} \\
2^{-n(H(XY)+\delta)} &\leq p_{XY}(x_1 y_1, \ldots, x_n y_n) \leq 2^{-n(H(XY)-\delta)}.
\end{aligned}$$

Note that above we write p_{XY} for $p_{(XY)^n}$.

Exactly the same argument as for one variable will show most messages are jointly typical in the limit. For a jointly δ-typical message $(xy, \ldots, xy) = (x^n, y^n)$, we see

$$p(x^n|y^n) = \frac{p_{XY}(x^n, y^n)}{p_Y(y^n)} \leq \frac{2^{-n(H(XY)-\delta)}}{2^{-n(H(Y)+\delta)}} = 2^{-n(H(XY)-H(Y)-2\delta)}$$

and

$$p(x^n|y^n) = \frac{p_{XY}(x^n, y^n)}{p_Y(y^n)} \geq \frac{2^{-n(H(XY)+\delta)}}{2^{-n(H(Y)-\delta)}} = 2^{-n(H(XY)-H(Y)+2\delta)}.$$

This prompts the following definitions.

Definition 4 (Conditional Entropy and Mutual Information)**.** *The conditional entropy of X given Y and the mutual information between X and Y are respectively defined as*

$$H(X|Y) = H(XY) - H(Y) \text{ and } I(X:Y) = H(X) - H(X|Y).$$

Conditional entropy answers the following question: If I know Y, what is my remaining ignorance about X? In the context of jointly δ-typical messages, we see that knowing Y lets us achieve a compression rate of $H(X|Y)$, asymptotically.

The inverse is the answered by mutual information: If I know Y, by how much has my compression burden been reduced? It is precisely the difference between $H(X)$ and $H(X|Y)$, since these are the best achievable rates.

Definition 5 (Rate and Capacity)**.** *A communication rate R is associated with a protocol encoding n-symbol output messages in Y to nR-symbol codewords in X. We say R is achievable if there is asymptotically vanishing error (i.e. the probability of a decoding error goes to 0 as $n \to \infty$).*

The capacity of a channel C is the supremum over all achievable rates R with respect to all codeword distributions X.

Theorem 2 (Shannon Noisy Coding)**.** *Channel capacity is given by* $\sup_X I(X:Y)$.

1.2 Functional Analysis

Now, let us recall some facts and definitions from functional analysis, which may all be found in a standard text such as [RS81]. On a Hilbert space $\mathcal{H}$ we denote its inner product by $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ with associated norm $\| \cdot \|_{\mathcal{H}}$, and identity operator $\mathbb{1}_{\mathcal{H}}$ (dropping subscripts if meaning remains clear). If $\mathcal{H} = \mathbb{C}$, we write the standard Euclidean inner product as $| \cdot |$. We will always assume $\mathcal{H}$ is separable and that its underlying field is $\mathbb{C}$.

Let $B(\mathcal{H})$ be (bounded!) linear operators $\mathcal{H} \to \mathcal{H}$, of which T will be a generic element. We know $B(\mathcal{H})$ is Banach under the operator norm

$$\|T\|_{\text{op}} = \sup\{\|Tx\|_{\mathcal{H}} : \|x\|_{\mathcal{H}} = 1\}.$$

Self-adjoint means $\langle Tx, y \rangle = \langle x, Ty \rangle$ and positive means $\langle Tx, x \rangle \geq 0$, for all $x, y \in \mathcal{H}$.

Recall that self-adjoint T admit a (continuous!) functional calculus, defining $f(T) \in B(\mathcal{H})$ for any f continuous (with respect to the maximum norm) on the spectrum $\sigma(T)$. We will only need $f(x) = |x|, \sqrt{x}, x \log x$. If T is also compact, then it admits a spectral decomposition

$$T = \sum_{\lambda \in \sigma(T)} \lambda |\lambda\rangle\langle\lambda|.$$

Here we introduce Dirac notation, where $|\lambda\rangle$ is a vector and $\langle\lambda| = \langle \cdot, \lambda \rangle$ is the dual (so that $|\lambda\rangle\langle\lambda| = \langle \cdot, \lambda \rangle \lambda$ is the projection onto $\lambda = |\lambda\rangle$).

We call T trace-class if its trace is finite and independent of the orthonormal basis $\{e_i\}$:

$$\operatorname{tr} T = \sum_i \langle Te_i, e_i \rangle.$$

The set of all such operators is $B_1(\mathcal{H})$ (which is Banach under $\|T\|_1 = \operatorname{tr}|T|$). For $p \in [1, \infty)$, we then define the Schatten p-norm $\|T\|_p = (\operatorname{tr}|T|^p)^{1/p}$. Essentially by definition, the Schatten p-norm is the ℓ^p norm on the singular values $\sigma(|T|)$ (if T is compact).

1.3 Quantum Channels

Lastly, we will introduce the basics of quantum information theory, with a good reference being [BŻ17]. Define the set of states $S(\mathcal{H}) \subseteq B_1(\mathcal{H})$ to be positive self-adjoint operators with unit trace. Note we may identify unit vectors $|x\rangle \in \mathcal{H}$ with states $|x\rangle\langle x| \in S(\mathcal{H})$, called pure states, the set of which is $S_1(\mathcal{H})$. Also establish two Hilbert spaces – Alice $\mathcal{A}$ and Bob $\mathcal{B}$.

Definition 6 (Quantum Channel)**.** *A quantum channel is a completely-positive trace preserving (CPTP) map $\mathcal{N}\colon S(\mathcal{A}) \to S(\mathcal{B})$. Trace preserving means $\operatorname{tr}\mathcal{N}(\rho) = 1$ for all states ρ, and completely-positive means $\mathcal{N} \otimes \mathbb{1}_n \geq 0$ for all integers $n \geq 0$ (note $n = 0$ means $\mathcal{N}$ alone is positive!).*

We write $\mathcal{N}^{\mathcal{A}\to\mathcal{B}}$ as shorthand to specify its input and output spaces. These two conditions reasonably describe sending a state from one place to another. We want to output states, so trace must be preserved. We also want positivity not just in isolation, but also whenever we append some extra operators acting on ancillary spaces.

Theorem 3 (Steinspring Dilation)**.** *Take a channel $\mathcal{N}^{\mathcal{A}\to\mathcal{B}}$. Then, there exists some larger Hilbert space $\mathcal{E}$ and an isometry $V\colon \mathcal{A} \to \mathcal{B} \otimes \mathcal{E}$ such that*

$$\mathcal{N}(\rho) = \operatorname{tr}_{\mathcal{E}}(V\rho V^*).$$

This gives an operational interpretation of channels. When Alice sends a state to Bob, there may be some noise involved – the state interacts with the environment $\mathcal{E}$. Of course, we do not have access to the environment, and so must trace it out. As a reminder, the partial trace $\operatorname{tr}_{\mathcal{E}}\colon B_1(\mathcal{B} \otimes \mathcal{E}) \to B_1(\mathcal{B})$ is the unique linear operator satisfying $\operatorname{tr}_{\mathcal{E}}(T \otimes S) = \operatorname{tr}(S)R$.

Theorem 4 (Schmidt Decomposition)**.** *Let $x \in \mathcal{A} \otimes \mathcal{B}$. Then, there exist orthonormal bases $\{e_i\}$ and $\{f_i\}$, respectively of $\mathcal{A}$ and $\mathcal{B}$, along with scalars $\lambda_i \geq 0$ so*

$$x = \sum_i \lambda_i e_i \otimes f_i.$$

2 Quantum Shannon Theory

Quantum Shannon theory broadly encompasses the study of quantum channels and entanglement measures. The most straightforward generalization of classical information theory is to ask how Alice and Bob can use a quantum channel to communicate classical information (such as by encoding strings via the outcomes of measurements on quantum states). We focus on the capacity for this classical communication, but several other notions of capacity exist too, such as quantum capacity or capacity under privacy restrictions (see [GIN18] for a review).

2.1 Classical Capacity

Henceforth, we assume $\mathcal{A}$ and $\mathcal{B}$ are finite-dimensional for simplicity, and follow the presentation of [Pre16]. In principle, though, nothing is wrong with infinite-dimensional channels – CPTP maps still make perfect sense and their the classical capacity has also been studied [HS13].

Definition 7 (Positive Operator-Valued Measurement)**.** *A positive operator-valued measurement (POVM) comprises a set* $\{E_i\}$ *of positive operators such that* $\mathbb{1} = \sum_i E_i$*.*

Say Alice wishes to send some message in X. To each element of its support x she may associate a state $\rho(x)$, sent through her quantum channel $\mathcal{N}$ with probability $p_X(x)$. Equivalently, Bob receives the ensemble $\rho(X) = \sum_i p_X(x)\rho(x)$.

On his end, to recover the classical information he may construct a POVM E, where each element $E(y)$ is associated with some classical information y. By performing the measurement $E\mathcal{N}(\rho(X))$, we interpret the analogue of Bob's classical distribution Y to be the distribution whose outcomes are y with probabilities $p_{Y_E}(y) = \operatorname{tr}(E(y)\mathcal{N}(\rho(X)))$. The conditional distribution $Y_E|X = x$ has masses $p(y|x) = \operatorname{tr}(E(y)\mathcal{N}(\rho(x)))$.

Unlike the classical case, Bob's distribution does not solely depend on Alice's distribution X and the channel $\mathcal{N}$. Instead, Bob now gets to decide which POVM he uses, hence our labelling of his distribution by Y_E.

Definition 8 (Accessible Information)**.** *With notation as above, the accessible information of* $\rho(X)$ *is*

$$I(\rho(X)) = \sup_E I(X : Y_E).$$

This quantity is not at all easy to compute. However, there is a concrete bound due to [Hol73] in terms of the von Neumann entropy $S(\rho) = -\operatorname{tr}\rho\log\rho$ of a state ρ.

Definition 9 (Holevo Capacity)**.** *For the ensemble* $\rho(X) = \sum p_X(x)\rho(x)$*, its Holevo chi is*

$$\chi(\rho(X)) = S(\rho(X)) - \sum_x p_X(x)S(\rho(x))$$

and the Holevo capacity of a quantum channel $\mathcal{N}$ *is*

$$\chi(\mathcal{N}) = \sup_\rho \chi(\mathcal{N}(\rho)).$$

Theorem 5 (Holevo)**.** *For every ensemble* $\rho(X)$ *and quantum channel* $\mathcal{N}$*,* $I(\rho(X)) \leq \chi(\mathcal{N}(\rho(X)))$*.*

The name Holevo capacity is suggestive. Indeed, we now define channel capacity and recount the result from [SW97] we mentioned in the introduction.

Definition 10 (Rate and Capacity)**.** *A quantum channel* $\mathcal{N}$ *communicates classically at a rate* R *if associated with Bob's* n*-symbol measurement outcomes in* Y *are* 2^{nR} *n-symbol codewords in* X *(which Alice sends as* n*-register quantum states). We say* R *is achievable if it may be accomplished with asymptotically vanishing error.*

The classical capacity of the channel $C(\mathcal{N})$ *is the supremum over achievable rates with respect to all input distributions* X*.*

Theorem 6 (Holevo-Schumacher-Westmoreland)**.** *For a quantum channel* $\mathcal{N}$*, its classical capacity is*

$$C(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n}\chi(\mathcal{N}^{\otimes n}).$$

Additivity of the Holevo capacity, the property that all quantum channels $\mathcal{N}, \mathcal{S}$ satisfy

$$\chi(\mathcal{N} \otimes \mathcal{S}) = \chi(\mathcal{N}) + \chi(\mathcal{S}),$$

would indeed be wonderful if true. But is it? A partial answer was also given in [SW97], that additivity holds if Alice is restricted to product states – i.e. she can only feed $\mathcal{N}^{\otimes n}$ states of the form $\rho_1 \otimes \cdots \otimes \rho_n$. However, what if Alice may input entangled states?

2.2 Minimum Output Entropy

Definition 11 (Rényi p-Entropy)**.** *For $p \in (0, \infty)$ we define the Rényi p-entropy of a state by*

$$S_p(\rho) = \frac{1}{1-p} \log \operatorname{tr}(\rho^p).$$

The case $p = 1$ is given by the limit, where $S_1 = S$.

Rényi p-entropy is somehow nicer than von Neumann entropy, as for $p > 1$ it is naturally related to the Schatten p-norm by $S_p(\rho) = p \log \|\rho\|_p / (1-p)$.

Definition 12 (Minimum Output p-Entropy)**.** *The minimum output p-entropy of a channel $\mathcal{N}^{\mathcal{A}\to\mathcal{B}}$ is*

$$S_p^{min}(\mathcal{N}) = \inf_{x \in S_1(\mathcal{A})} S_p(\mathcal{N}(|x\rangle\langle x|)).$$

This is a quantity related to channels, and so it makes sense to speak of its additivity too! For two channels $\mathcal{N}, \mathcal{S}$, is it always true that

$$S_p^{\min}(\mathcal{N} \otimes \mathcal{S}) = S_p^{\min}(\mathcal{N}) + S_p^{\min}(\mathcal{S})?$$

Shor established a crucial connection between the minimum output 1-entropy (i.e. von Neumann entropy) and the Holevo capacity in [Sho04], showing their additivities are equivalent (alongside many other quantities, like entanglement of formation). And this is how we know Holevo capacity is not additive.

The first step was taken by [WH02], with numerical experiments showing minimum output p-entropy is not additive for $p > 4.79$. However, it was still suspected that not only does additivity hold for $p = 1$, but in fact all $p \in [1, 2]$ (due to the convexity of the map $\rho \mapsto \rho^p$). However, two classes of counterexamples in [HW08] showed additivity is false for all $p > 1$. Shortly after, [Cub+08] gave numerical evidence of non-additivity for $p < 0.11$. In spite of all this, the von Neumann case was still believed to be true – until Hastings showed it is false [Has09].

3 Existence of Non-Additive Channels

We know minimum output von Neumann entropy is not additive. Yet, we somehow lack any examples, as all proofs are probabilistic, showing that with a positive probability some random construction yields a channel violating additivity. The construction is a subspace, and is why asymptotic geometric analysis appears – this field concerns itself with which regularity conditions hold for random subspaces with large probability as the dimension grows.

3.1 Asymptotic Geometry Analysis

The main result we need is due to Dvoretzky alone [Dvo64], however Milman contributed a novel proof of it [Mil71], spawning the field of asymptotic geometric analysis. We will adapt the approach of [AS17] in our cursory overview of it.

We require the notion of the Haar measure, the unique (translation-invariant regular) probability measure on a compact group – for us it will be on the unitary group $\mathrm{U}(n)$. When we speak of a random k-dimensional subspace $E \subseteq \mathbb{C}^n$, we mean taking a Haar-distributed unitary $U \in \mathrm{U}(n)$ and setting $E = U\imath(\mathbb{C}^k)$, where $\imath(\mathbb{C}^k) \subseteq \mathbb{C}^n$ is inclusion. Denote the unit sphere in $\mathbb{C}^n$ by $S_{\mathbb{C}}^{n-1}$, and recall the Haar measure also induces a uniform probability measure σ on $S_{\mathbb{C}}^{n-1}$.

Definition 13 (Circled Convex Body)**.** *We say $K \subseteq \mathbb{C}^n$ is a convex body if it is convex, compact, and has non-empty interior. If for every $x \in K$ we have $\exp(i\vartheta)x \in K$ for all $\vartheta \in \mathbb{R}$, as well as $0 \in K$, then we say K is circled.*

Circled convex bodies K exactly correspond to norms on $\mathbb{C}^n$, since each defines a norm by the Minkowski functional

$$\|\cdot\|_K \colon \mathbb{C}^n \to \mathbb{R} \text{ by } \|x\|_K \mapsto \inf\{t > 0 : x \in tK\},$$

and each norm has a unit ball (which is a circled convex body).

Definition 14 (Inradius and Mean Width)**.** *Define the inradius of a circled convex body K by*

$$\operatorname{inrad} K = \sup\{r > 0 : B(0, r) \subseteq K\},$$

where $B(0, r)$ is the open ball of radius r about the origin. Also define the mean width

$$w(K) = \int_{S_{\mathbb{C}}^{n-1}} \|x\|_K \, \mathrm{d}\sigma(x).$$

Definition 15 (Dvoretzky Dimension)**.** *The Dvoretzky dimension of a circled convex body $K \subseteq \mathbb{C}^n$ is*

$$k_*(K) = (w(K) \operatorname{inrad}(K))^2 n.$$

Theorem 7 (Dvoretzky-Milman)**.** *There exist universal constants $c, c' > 0$ such that the following always holds. Take $K \subseteq \mathbb{C}^n$ to be a circled convex body. Fix some $\varepsilon \in (0, 1]$ and define $k = c\varepsilon^2 k_*(K)$. Then, any random subspace $E \subseteq \mathbb{C}^n$ with $\dim E \leq k$ satisfies*

$$(1 - \varepsilon)w(K)|x| \leq \|x\|_K \leq (1 + \varepsilon)w(K)|x|$$

for all $x \in E$ with probability at least $1 - \exp(-c'\varepsilon^2 k_*(K))$.

A moral interpretation of the theorem is that convex bodies are "locally approximately Euclidean" in high dimensions. The convex body norm on random low-dimensional subspaces (where low is with respect to the Dvoretzky dimension) is typically within an epsilon of error of the standard Euclidean norm, after scaling by the mean width of the body.

3.2 Channels as Subspaces

Consider a channel $\mathcal{N}^{\mathcal{A}\to\mathcal{B}}$ and its Steinspring isometry $V : \mathcal{A} \to \mathcal{B} \otimes \mathcal{E}$. Take some pure state $|x\rangle \in \mathcal{A}$ so that

$$\mathcal{N}(|x\rangle\langle x|) = \operatorname{tr}_{\mathcal{E}}(V|x\rangle\langle x|V^*).$$

As V is an isometry, we know $V|x\rangle = |y\rangle$ is pure and thus admits some Schmidt decomposition

$$|y\rangle = \sum_i \lambda_i |e_i\rangle \otimes |f_i\rangle.$$

Tracing out the environment yields

$$\operatorname{tr}_{\mathcal{E}}(|y\rangle\langle y|) = \sum_i \lambda_i^2 |e_i\rangle\langle e_i|,$$

and so

$$S_p(\mathcal{N}(|x\rangle\langle x|)) = \frac{1}{1-p}\log\sum_i \lambda_i^2.$$

Thus, the (squares of the) Schmidt coefficients of $V|x\rangle$ entirely determine the output p-entropy.

Now, V is an isometry and therefore a bijection onto its range $V(\mathcal{A}) \subseteq \mathcal{B}\otimes\mathcal{E}$. Importantly, we have an isometric isomorphism between subspaces: $\mathcal{A} \cong V(\mathcal{A})$. This allows us to uniquely identify $\mathcal{N}$ with a subspace of $\mathcal{B}\otimes\mathcal{E}$.

Furthermore, V gives a bijection between pure states $S_1(\mathcal{A})$ and $S_1(V(\mathcal{A}))$. Recalling that minimum output p-entropy optimizes over all pure states, we have

$$S_p^{\min}(\mathcal{N}) = \inf_{x\in S_1(\mathcal{A})} S_p(\mathcal{N}(|x\rangle\langle x|)) = \inf_{y\in S_1(V(\mathcal{A}))} S_p(|y\rangle\langle y|).$$

Our conclusion is that the study of the minimum output p-entropy of $\mathcal{N}$ reduces to the study of the Schmidt coefficients of pure states in $V(\mathcal{A})$. This goes the other way too – given any subspace $\mathcal{V} \subseteq \mathcal{B}\otimes\mathcal{E}$, we know that there is some isometric isomorphism $V : \mathbb{C}^{\dim\mathcal{V}} \to \mathcal{V}$, and that $\rho \mapsto \mathrm{tr}_{\mathcal{E}}(V\rho V^*)$ defines a channel $\mathcal{N}^{\mathbb{C}^{\dim\mathcal{V}}\to\mathcal{B}}$.

This is the modern approach to tackling channel additivity questions. Dealing with subspaces and the Schmidt coefficients of their unit vectors is generally a much more tractable and computationally concrete problem than dealing with the Holevo capacity itself.

3.3 Failure of Additivity for $p > 1$

The original proof that minimum output von Neumann entropy is not additive [Has09] indeed took the approach of estimating Schmidt coefficients. However, it did not do so by looking at subspaces in isolation (instead at outputs of random unital channels). This proof was restated by [BH10] in terms of concentration of measure results, and the full extension to the Dvoretzky-Milman theorem was ultimately presented in [ASW10].

We will not present this proof, but a morally similar (and much easier) proof of additivity failing for $p > 1$, filling in the Dvoretzky-pertinent gaps of [Wu23]. We use the notation $\approx$ to hide universal constants, and let M_{kd} be the space of k-by-d complex matrices. Recall there is a natural way to associate tensors with matrices by lexicographical ordering, and for $\mathcal{V} \subseteq \mathbb{C}^k\otimes\mathbb{C}^d$ we will write $\mathrm{M}(\mathcal{V}) \subseteq \mathrm{M}_{kd}$ to denote this.

Lemma 1. *Let $\mathcal{N}$ be a quantum channel with associated subspace $\mathcal{V} \subseteq \mathbb{C}^k\otimes\mathbb{C}^d$. For all $p > 1$ we have*

$$S_p^{min}(\mathcal{N}) = \frac{2p}{1-p}\log \sup_{\substack{M\in\mathrm{M}(\mathcal{V})\\ \|M\|_2=1}} \|M\|_{2p}.$$

Proposition 2. *Let $B_p^{mn} \subseteq \mathrm{M}_{mn}$ be the unit ball with respect to the Schatten p-norm. Its inradius is*

$$\mathrm{inrad}(B_p^{mn}) = \begin{cases} m^{1/2-1/p} & 1 \le p \le 2 \\ 1 & 2 \le p \le \infty \end{cases}$$

and the mean width is $w(B_p^{mn}) \approx m^{1/p-1/2}$.

Proposition 3. *Fix $p > 1$. Then, there exist sufficiently large k, d and some subspace $\mathcal{V} \subseteq \mathbb{C}^k \otimes \mathbb{C}^d$ such that for the associated channel $\mathcal{N}$ we have*

$$\log k - c_1(p) \leq S_p^{min}(\mathcal{N}) \leq \log k - c_2(p).$$

Here, $c_1(p) > c_2(p) > 0$ are constants depending only on p, and $\dim \mathcal{V} \approx k^{1/p} d$.

Proof. Let $K = B_{2p}^{kd}$ be the Schatten $2p$-ball in M_{kd}. By PROPOSITION 2, it has Dvoretzky dimension

$$k_*(K) \approx \begin{cases} kd & 1 \leq 2p \leq 2 \\ k^{1/p} d & 2 \leq 2p \leq \infty \end{cases}$$

and mean width $w(K) \approx k^{1/2p-1/2}$. For the dimension, we are always in the latter case since $p > 1$. Therefore, we apply THEOREM 7 and for any $\varepsilon > 0$ obtain with high probability some subspace $\mathcal{V} \subseteq \mathrm{M}_{kd}$ of dimension $m \approx \varepsilon^2 k_*(K)$ such that

$$(1-\varepsilon)k^{1/2p-1/2}\|M\|_2 \leq \|M\|_{2p} \leq (1+\varepsilon)k^{1/2p-1/2}\|M\|_2$$

for all $M \in \mathcal{V}$. This inequality is actually hiding a constant (which ε can eat anyway).

From LEMMA 1 we have

$$S_p^{\min}(\mathcal{N}) = \frac{2p}{1-p} \log \sup_{\substack{M \in \mathrm{M}(\mathcal{V}) \\ \|M\|_2 = 1}} \|M\|_{2p}$$

and so our bounds tell us

$$\frac{2p}{1-p} \log\Big((1-\varepsilon)k^{1/2p-1/2}\Big) \geq S_p^{\min}(\mathcal{N}) \geq \frac{2p}{1-p} \log\Big((1+\varepsilon)k^{1/2p-1/2}\Big),$$

where the supremum is immaterial since $\|M\|_2 = 1$ always. Notice too the inequality flips as $p > 1$ and so we are decreasing in the logarithm. We may rewrite this as

$$\log(k) + \frac{2p}{1-p}\log(1-\varepsilon) \geq S_p^{\min}(\mathcal{N}) \geq \log(k) + \frac{2p}{1-p}\log((1+\varepsilon)).$$

The $1 \pm \varepsilon$ terms define the constants $c_1(p)$ and $c_2(p)$. Though these technically depend on ε, we do not actually care about the probability of getting such a channel, so any fixed ε will work. Setting this constant is exactly how we get $m \approx k^{1/p} d$. □

Proposition 4. *Let $\mathcal{V} \subseteq \mathbb{C}^k \otimes \mathbb{C}^d$ have corresponding channel $\mathcal{N}$ with isometry $V \colon \mathbb{C}^m \to \mathbb{C}^k \otimes \mathbb{C}^d$, where $m = \dim \mathcal{V}$. Denote by $\bar{\mathcal{N}}$ the conjugate channel given by $\bar{V}$. Then,*

$$S_p^{min}(\mathcal{N} \otimes \bar{\mathcal{N}}) \leq \frac{p}{1-p} \log \frac{m}{kd}.$$

Theorem 8. *Minimum output p-entropy is not additive for all $p > 1$.*

Proof. Start by fixing $p > 1$, then applying PROPOSITION 3 to obtain a channel $\mathcal{N}^{\mathbb{C}^m \to \mathbb{C}^k}$ such that $S_p^{\min}(\mathcal{N}) \geq \log k - c(p)$ for some constant $c(p) > 0$ and $m \approx k^{1+1/p}$. Without loss of generality

assume $\mathcal{N}$ extends into an ancilliary environment also of dimension k. Applying PROPOSITION 4 then tells us

$$S_p^{\min}(\mathcal{N} \otimes \bar{\mathcal{N}}) \leq \frac{p}{1-p} \log \frac{m}{k^2} \leq \log k + c'(p),$$

where $c'(p) > 0$ is some positive constant (previously hidden by universality). So,

$$\begin{aligned} S_p^{\min}(\mathcal{N}) + S_p^{\min}(\bar{\mathcal{N}}) &\geq 2 \log k - 2c(p) \\ &> \log k - 2c(p) + \log k - c'(p) \geq S_p^{\min}(\mathcal{N} \otimes \bar{\mathcal{N}}) + \log k - 2c(p). \end{aligned}$$

Since p is fixed, we may simply take k as large as necessary to ensure $\log k - 2c(p) > 0$, and then we have the desired violation of additivity. □

4 Conclusion

To date, no constructive counterexamples exist violating additivity for minimum output von Neumann entropy. All arguments are probabilistic, similar to the above. However, there is an ongoing search. Constructive counterexamples for all $p > 2$ were found by [GHP10], using the antisymmetric subspace, for an orthonormal basis $\{e_i\}$ of $\mathbb{C}^k$ defined by

$$W_{\text{as}} = \left\{ \frac{1}{\sqrt{2}} \left(e_i \otimes e_j - e_j \otimes e_i \right) : i < j \right\} \subseteq \mathbb{C}^k \otimes \mathbb{C}^k.$$

This required $k \to \infty$ as $p \to 2$, and since $\dim W_{\text{as}} = \binom{k}{2}$ the channel size explodes. This subspace was recently revisited by [SS23], who found that with m fixed, there exist n-dimensional subspaces $X_n \subseteq \mathbb{C}^k \otimes \mathbb{C}^k$ for all $n = 1, 2, \ldots, \lfloor k/2 \rfloor$ such that $W_n = W_{\text{as}} \oplus X_n$ violates additivity. Though this tames the size of the channel, it comes at the cost of losing control of p, and so cannot approach the von Neumann entropy at all.

Take a set of arbitrary scalars $G = \{\lambda_i\}_{i=1}^{2d-1} \subseteq \mathbb{C}$ and define $u_\lambda = \sum_{i=0}^{d-1} \lambda^i e_i$. Then, the so-called completely entangled subspace

$$W_{\text{ce}} = \text{span}\{u_\lambda \otimes u_\lambda : \lambda \in G\}^\perp \subseteq \mathbb{C}^k \otimes \mathbb{C}^k$$

contains no decomposable tensors [Par04]. In the same work, [SS23] found numerical evidence that subspaces $S \subseteq W_{\text{ce}}$ can violate additivity. With $k = 5$, when $\dim W_{\text{ce}} = (5-1)^2 = 16$, there exist some 14-dimensional S which violate additivity for $p \approx 1.844$, the first constructive counterexample with $1 < p < 2$.

These types of "pathological" subspaces, which heavily constrain the Schmidt coefficients of their vectors, are promising areas to search for counterexamples. The difficulty in moving from numerical to analytic expressions is the optimization problem inherent in minimum output entropy. Until a precise expression for the Schmidt coefficients is found, the problem can be numerically intractable even for low dimensions – let alone solvable by hand.

References

[Sha48] Claude Shannon. "A Mathematical Theory of Communication". In: *Bell System Technical Journal* 27 (3 1948), pp. 379–423.

[Dvo64] Aryeh Dvoretzky. "Some Results on Convex Bodies and Banach Spaces". In: *Matematika* 8 (1 1964), pp. 73–102.

[Mil71] Vitali Milman. "A New Proof of A. Dvoretzky's Theorem on Intersections of Convex Bodies". In: *Funktsionalnyi Analiz i Ego Prilozheniya* 5 (4 1971), pp. 28–37.

[Hol73] Alexander Holevo. "Bounds for the Quantity of Information Transmitted by a Quantum Communication Channel". In: *Problems of Information Transmission* 9 (1973), pp. 177–183.

[Hol77] Alexander Holevo. "Problems in the Mathematical Theory of Quantum Communication Channels". In: *Reports on Mathematical Physics* 12 (2 1977), pp. 273–278.

[RS81] Michael Reed and Barry Simon. *Functional Analysis*. Academic Press, 1981.

[SW97] Benjamin Schumacher and Michael Westmoreland. "Sending Classical Information via Noisy Quantum Channels". In: *Physical Review A* 56 (1997), pp. 131–138.

[WH02] Reinhard Werner and Alexander Holevo. "Counterexample to an Additivity Conjecture for Output Purity of Quantum Channels". In: *Journal of Mathematical Physics* 43 (9 2002), pp. 4353–4357.

[Par04] K. R. Parthasarathy. "On the Maximal Dimension of a Completely Entangled Subspace for Finite Level Quantum Systems". In: *Proceedings Mathematical Sciences* 114 (2004), pp. 365–374.

[Sho04] Peter Shor. "Equivalence of Additivity Questions in Quantum Information Theory". In: *Communications in Mathematical Physics* 246 (2004), pp. 453–472.

[Cub+08] Toby Cubitt et al. "Counterexamples to Additivity of Minimum Output p-Rényi Entropy for p Close to 0". In: *Communications in Mathematical Physics* 284 (2008), pp. 281–290.

[HW08] Patrick Hayden and Andreas Winter. "Counterexamples to the Maximal p-Norm Multiplicativity Conjecture for all $p > 1$". In: *Communications in Mathematical Physics* 284 (2008), pp. 263–280.

[Has09] Matthew Hastings. "Superadditivity of Communication Capacity Using Entangled Inputs". In: *Nature Physics* 5 (2009), pp. 255–257.

[ASW10] Guillaume Aubrun, Stanisław Szarek, and Elisabeth Werner. "Hasting's Additivity Counterexample via Dvoretzky's Theorem". In: *Communications in Mathematical Physics* 305 (2010), pp. 85–97.

[BH10] Fernando Brandão and Michał Horodecki. "On Hastings' Counterexamples to the Minimum Output Entropy Additivity Conjecture". In: *Open Systems & Information Dynamics* 17 (1 2010), pp. 31–52.

[GHP10] Andrzej Grudka, Michał Horodecki, and Łukasz Pankowski. "Constructive Counterexamples to the Additivity of the Minimum Output Rényi Entropy of Quantum Channels for all $p > 2$". In: *Journal of Physics A* 43 (2010).

[HS13] Alexander Holevo and Maksim Shirokov. "On Classical Capacities of Infinite-Dimensional Quantum Channels". In: *Problems of Information Transmission* 49 (2013), pp. 15–31.

[Pre16] John Preskill. "Quantum Shannon Theory". In: *arXiv:1604.07450* (2016).

[AS17] Guillaume Aubrun and Stanisław Szarek. *Alice and Bob Meet Banach. The Interface of Asymptotic Geometric Analysis and Quantum Information Theory*. American Mathematical Society, 2017.

[BŻ17] Ingemar Bengtsson and Karol Życzkowski. *Geometry of Quantum States*. Cambridge University Press, 2017.

[GIN18] Laszlo Gyongyosi, Sandor Imre, and Hung Viet Nguyen. "A Survey on Quantum Channel Capacities". In: *IEEE Comunnications Surveys & Tutorials* 20 (2 2018), pp. 1149–1205.

[SS23] K. Szczygielski and M. Studziński. "New Constructive Counterexamples to Additivity of Minimum Output Rényi Entropy of Quantum Channels". In: *arXiv:2301.07428* (2023).

[Wu23] Peixue Wu. *Generic Nonadditivity of Minimum Output Entropy of Quantum Channels*. Personal Communication. 2023.