Comments on Hastings' Additivity Counterexamples

C. King
Department of Mathematics
Northeastern University

Tucson, Arizona 2009

Punchline

The longstanding conjecture that *quantum channel capacity is additive* was finally demolished this past year by Matt Hastings: his paper is

http://arxiv.org/abs/0809.3972

The goal of this talk is to explain some aspects of Hastings' paper. Emphasis in on main ideas, so estimates are rough, factors and constants omitted.

Quantum Channels.

A *quantum channel* $\mathcal{F}$ is a map which describes the dynamical evolution of states in a quantum system, including the effects of entanglement with the environment:

$$\mathcal{F} \,:\, \mathcal{S}(\mathcal{H}_{in}) \to \mathcal{S}(\mathcal{H}_{out})$$

$\mathcal{H}_{in},\ \mathcal{H}_{out}$ are input and output state spaces, $\mathcal{S}(\mathcal{H}) \subset \mathcal{B}(\mathcal{H})$ is the set of states on $\mathcal{H}$. The map $\mathcal{F}$ must be linear and completely positive. Examples:

- $\mathcal{F}(\rho) = U\rho U^*$ where $U = e^{-iHt}$ (unitary noiseless evolution)
- $\mathcal{F}(\rho) = \sum_i p_i U_i \rho U_i^*$ where $\sum_i p_i = 1$ (random unitary channel).

Why the name 'Channel'?

Any assignment of labels $1, 2, \ldots$ to states $|\psi_1\rangle, |\psi_2\rangle, \ldots$ can be viewed as a *storage of classical information* in a quantum system. Similarly any (generalized) measurement on a quantum system produces a map from states to the labels $1, 2, \ldots$ of the measurement operators $E_1, E_2, \ldots$, and so can be viewed as a *retrieval of classical information*.
If a quantum channel acts on the system between storage and retrieval we can view it as the quantum analog of a classical discrete memoryless channel.

$$X \in \{1, 2, \ldots\} \to \rho_X \to \mathcal{F}(\rho_X) \to Y \in \{1, 2, \ldots\}$$

with $P(Y = j | X = i) = \mathrm{Tr}\, \mathcal{F}(\rho_i) E_j$

Classical Capacity

Holevo derived the following generalization of Shannon's capacity formula for a quantum channel $\mathcal{F}$:

$$\chi^*(\mathcal{F}) = \sup_{p_i,\sigma_i} \left\{ S\Big(\sum_i p_i\, \mathcal{F}(\sigma_i)\Big) - \sum_i p_i\, S(\mathcal{F}(\sigma_i)) \right\}$$

where the sup runs over ensembles $\{p_i, \sigma_i\}$ of input states, and where $S(\sigma) = -\mathrm{Tr}\,\sigma \log \sigma$ is the von Neumann entropy.

Holevo-Schumacher-Westmoreland Theorem (1998): $\chi^*(\mathcal{F})$ is the maximum rate at which information can be transmitted through the channel $\mathcal{F}$.

In order to reach the maximum, we must allow multiple uses of the channel with block coding over the input states, and allow measurements at the output using operators entangled over many uses.

Classical Capacity

Holevo derived the following generalization of Shannon's capacity formula for a quantum channel $\mathcal{F}$:

$$\chi^*(\mathcal{F}) = \sup_{p_i,\sigma_i} \left\{ S\Big(\sum_i p_i\, \mathcal{F}(\sigma_i)\Big) - \sum_i p_i\, S(\mathcal{F}(\sigma_i)) \right\}$$

where the sup runs over ensembles $\{p_i, \sigma_i\}$ of input states, and where $S(\sigma) = -\mathrm{Tr}\,\sigma \log \sigma$ is the von Neumann entropy.

Holevo-Schumacher-Westmoreland Theorem (1998): $\chi^*(\mathcal{F})$ is the maximum rate at which information can be transmitted through the channel $\mathcal{F}$.

In order to reach the maximum, we must allow multiple uses of the channel with block coding over the input states, and allow measurements at the output using operators entangled over many uses.

The First Conjecture

However for the HSW Theorem we are restricted to using ensembles of *product states* at the input.

The question remained: can we increase the rate by using entangled input states? Dropping the restriction to product inputs leads to the following formula for the capacity:

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

Conjecture

$C(\mathcal{F}) = \chi^*(\mathcal{F})$ for all channels $\mathcal{F}$.

The Additivity Conjectures

The first conjecture is equivalent to additivity of χ^* for multiple products of the same channel. So a natural generalization is

Conjecture

For all channels $\mathcal{F}$ and $\mathcal{G}$,

$$\chi^*(\mathcal{F} \otimes \mathcal{G}) = \chi^*(\mathcal{F}) + \chi^*(\mathcal{G})$$

Define the minimal output entropy:

$$S_{min}(\mathcal{F}) = \inf_{\rho} S(\mathcal{F}(\rho))$$

Conjecture

For all channels $\mathcal{F}$ and $\mathcal{G}$,

$$S_{min}(\mathcal{F} \otimes \mathcal{G}) = S_{min}(\mathcal{F}) + S_{min}(\mathcal{G})$$

The Additivity Conjectures

The first conjecture is equivalent to additivity of χ^* for multiple products of the same channel. So a natural generalization is

Conjecture

For all channels $\mathcal{F}$ and $\mathcal{G}$,

$$\chi^*(\mathcal{F} \otimes \mathcal{G}) = \chi^*(\mathcal{F}) + \chi^*(\mathcal{G})$$

Define the minimal output entropy:

$$S_{min}(\mathcal{F}) = \inf_{\rho} S(\mathcal{F}(\rho))$$

Conjecture

For all channels $\mathcal{F}$ and $\mathcal{G}$,

$$S_{min}(\mathcal{F} \otimes \mathcal{G}) = S_{min}(\mathcal{F}) + S_{min}(\mathcal{G})$$

Hastings Results

In 2002 Peter Shor showed that the last two conjectures are equivalent. In 2008 Hastings showed that S_{min} is not additive, and hence that both are false.

The Hastings counterexamples concern finite-dimensional channels:

$$\mathcal{H}_{in} = \mathcal{H}_{out} = \mathbb{C}^N$$

The channels are random unitary channels:

$$\mathcal{E} : \mathbb{C}^{N\times N} \to \mathbb{C}^{N\times N}, \qquad \mathcal{E}(\rho) = \sum_{i=1}^{D} p_i \, U_i \rho U_i^*$$

- $1 << D << N$
- $p_i \geq 0, \quad \sum_i p_i = 1$
- $U_i U_i^* = I_{N\times N}$

Winter's product channel

To construct the counterexample, Hastings made use of Andreas Winter's idea to use the product $\mathcal{E} \otimes \overline{\mathcal{E}}$: recall the maximally entangled state in $\mathbb{C}^N \otimes \mathbb{C}^N$:

$$|ME\rangle = \frac{1}{N} \sum_{a=1}^{N} |a\rangle \otimes |a\rangle$$

Note that for any matrix A we have $(A \otimes I)|ME\rangle = (I \otimes A^T)|ME\rangle$. Hence for a unitary matrix U we get

$$(U \otimes \overline{U})|ME\rangle = |ME\rangle$$

So define the complex conjugate map

$$\overline{\mathcal{E}}(\rho) = \sum_{i=1}^{D} p_i \, \overline{U_i} \rho \overline{U_i^*}$$

Hastings easy bound

Then the product channel $\mathcal{E} \otimes \overline{\mathcal{E}}$ acts as follows:

$$
\begin{aligned}
(\mathcal{E} \otimes \overline{\mathcal{E}})|ME\rangle\langle ME| &= \sum_{i=1}^{D} p_i^2 \; |ME\rangle\langle ME| \\
&+ \sum_{i \neq j} p_i p_j \; (U_i \otimes \overline{U_j})|ME\rangle\langle ME|(U_i^* \otimes \overline{U_j^*})
\end{aligned}
$$

The presence of the first term on the right side implies that the state $(\mathcal{E} \otimes \overline{\mathcal{E}})|ME\rangle\langle ME|$ cannot be too noisy. Hastings derived from this the bound

$$
S_{min}(\mathcal{E} \otimes \overline{\mathcal{E}}) \leq 2 \log D - \frac{\log D}{D}
$$

Hastings hard bound

Theorem

There is $h_0 < \infty$ such that for all $h \geq h_0$, and all D, $\frac{N}{D}$ sufficiently large, there is a random unitary channel $\mathcal{E}$ with

$$S_{min}(\mathcal{E}) \geq \log D - \frac{h}{D}$$

Easy to show that $S_{min}(\overline{\mathcal{E}}) = S_{min}(\mathcal{E})$. Hence the two bounds together show that for D, $\frac{N}{D}$ sufficiently large, for the channel $\mathcal{E}$ of the Theorem,

$$\begin{aligned} S_{min}(\mathcal{E} \otimes \overline{\mathcal{E}}) &\leq 2 \log D - \frac{\log D}{D} \\ &< 2 \log D - 2\frac{h}{D} \\ &\leq S_{min}(\mathcal{E}) + S_{min}(\overline{\mathcal{E}}) \end{aligned}$$

Full Hastings result

Recall $\mathcal{E}(\rho) = \sum p_i U_i \rho U_i^*$. Let

$$\begin{aligned} \Omega_{N,D} &= \{\text{random unitary channels}\} \\ &= \{p_1, \ldots, p_d, U_1, \ldots, U_D\} \\ &= \mathcal{S}_D \times U(N)^D \end{aligned}$$

where $\mathcal{S}_D$ is the simplex of discrete D-dimensional distributions. Define product probability measure on Ω:

$$\begin{aligned} \mathbb{P}_{\mathcal{E}} &= \nu \times \text{Haar} \\ d\nu(p_1, \ldots, p_D) &= C_{N,D} \prod_{i=1}^{D} p_i^{N-1} \delta(1 - \sum_i p_i)[dp] \end{aligned}$$

The measure $d\nu$ is used because it is the marginal of the uniform measure on the sphere S^{2ND-1}.

Full Hastings result

Then for D sufficiently large

$$\mathbb{P}_{\mathcal{E}}\left(S_{min}(\mathcal{E}) \geq \log D - \frac{h}{D}\right) \to 1$$

as $N \to \infty$.

So a random choice of random unitary channel will almost certainly satisfy the Hastings bound.

Step 1: switch to Conjugate Channel

Define the $N \times D$ matrix

$$A = (\sqrt{p_1} U_1 |\psi\rangle \cdots \sqrt{p_D} U_D |\psi\rangle)$$

then

$$\mathcal{E}(|\psi\rangle\langle\psi|) = \sum_{i=1}^{D} p_i \, U_i |\psi\rangle\langle\psi| U_i^* = AA^*$$

Since $D << N$ this has many zero eigenvalues. It has the same nonzero spectrum as

$$A^* A = \sum_{i,j=1}^{D} \sqrt{p_i p_j} \, \langle\psi| U_j^* U_i |\psi\rangle \, |i\rangle\langle j|$$

So define the *conjugate channel*:

$$\begin{aligned} \mathcal{E}^c &: \mathbb{C}^{N\times N} \to \mathbb{C}^{D\times D} \\ \mathcal{E}^c(|\psi\rangle\langle\psi|) &= A^*A = \sum_{i,j=1}^{D} \sqrt{p_i p_j}\, \langle\psi|U_j^* U_i|\psi\rangle\, |i\rangle\langle j| \end{aligned}$$

It follows that for any pure state $|\psi\rangle$

$$S(\mathcal{E}(|\psi\rangle\langle\psi|)) = S(\mathcal{E}^c(|\psi\rangle\langle\psi|))$$

and hence

$$S_{min}(\mathcal{E}) = S_{min}(\mathcal{E}^c)$$

More convenient to work with $\mathcal{E}^c$.

Step 2: Random unit vectors

Recall $\mathcal{E}(|\psi\rangle\langle\psi|) = AA^*$ where

$$A = (\sqrt{p_1}U_1|\psi\rangle \cdots \sqrt{p_D}U_D|\psi\rangle)$$

$$\mathrm{Tr}\,\mathcal{E}(|\psi\rangle\langle\psi|) = 1 \Rightarrow \sum_{i,j} |A_{ij}|^2 = 1$$

and so the components of A form a *unit vector* in $\mathbb{C}^{ND}$.

The *columns* of A are vectors in $\mathbb{C}^N$: their lengths are $\{\sqrt{p_1}, \ldots, \sqrt{p_D}\}$.

$\mathbb{P}_{\mathcal{E}} = \nu \times \text{Haar} \quad \Rightarrow$ for any state $|\psi\rangle$, A is a random unit vector w.r.t. uniform measure on the sphere S^{2ND-1}

Hence for any state $|\psi\rangle$, $\mathcal{E}^c(|\psi\rangle\langle\psi|) = A^*A$ is the $D \times D$ reduced density matrix of a random ND-dimensional unit vector, and its distribution is known exactly:

$$A^*A = V \ \mathrm{diag}(\lambda_1 \ldots \lambda_D) \ V^*$$

$$d\mu(\lambda_1, \ldots, \lambda_D) = Z_{N,D}^{-1} \, \delta(1 - \sum \lambda_i) \prod_{i<j} (\lambda_i - \lambda_j)^2 \prod_{i=1}^{D} \lambda_i^{N-D} \theta(\lambda_i) \, [d\lambda]$$

where $\theta(\lambda) = 1$ if $\lambda > 0$, $= 0$ if $\lambda < 0$.

Random unit vectors: summary

- W.r.t. the probability measure $\mathbb{P}_{\mathcal{E}}$ for random unitary channels, the eigenvalues of $\mathcal{E}^c(|\psi\rangle\langle\psi|)$ have the distribution $d\mu$ for any state $|\psi\rangle$.
- If $\mathcal{E}$ is selected randomly with distribution $\mathbb{P}_{\mathcal{E}}$ and $|\psi\rangle$ is selected randomly and uniformly on the unit sphere, then the eigenvalues of $\mathcal{E}^c(|\psi\rangle\langle\psi|)$ have the distribution $d\mu$.

Why might the 'hard' result be true?

The 'hard' result says that for a typical channel $\mathcal{E}$ all output states have entropy at least $\log D - h/D$. Similar kinds of results have been derived before.

Example 1) [D. Page (1993), S. Sen (1996)] Expected entropy of reduced density matrix:

$$\mathbb{E}_{\mathcal{E},|\psi\rangle}[S(\mathcal{E}^c(|\psi\rangle\langle\psi|))] \geq \log D - \frac{D}{2N}$$

So on average a random channel and random state produce an output with entropy very close to the maximal value $\log D$.

Levy's Lemma (concentration of measure): for any map

$$f : S^k \to \mathbb{R}, \qquad |f(x) - f(y)| \leq \eta \, ||x - y||_2,$$

$$\mathbb{P}(|f(x) - \mathbb{E}[f]| > \alpha) \leq e^{-C(k-1)\alpha^2/\eta^2}$$

Apply this to $S(\mathcal{E}^c(|\psi\rangle\langle\psi|))$: $\eta \leq C \log D$,

Example 2

$$\mathbb{P}_{\mathcal{E},|\psi\rangle}\left(S\Big(\mathcal{E}^c(|\psi\rangle\langle\psi|)\Big) < \log D - \frac{h}{D}\right) \leq e^{-C'N/D(\log D)^2}$$

Hence for each N, D there exists a channel $\mathcal{E}$ such that this bound holds for a randomly chosen input state $|\psi\rangle$.

Are we done yet?

These bounds show that for a typical channel $\mathcal{E}$, as $N \to \infty$ the probability that a randomly chosen input state will violate Hastings' bound goes to zero.

But it is necessary to prove that there exists a channel such that **NO** input states will violate the bound.

Main Idea

Suppose that $\mathcal{E}$ is a typical channel, then for 'most' input states $|\psi\rangle$ the output $\mathcal{E}^c(|\psi\rangle\langle\psi|)$ will be close to a maximally mixed state. This is implied by the second result above.

Hastings makes a stronger statement: for a typical channel there are constants γ, δ such that

$$\mathbb{P}_\psi\left(\left|\left|\mathcal{E}^c(|\psi\rangle\langle\psi|) - \frac{1}{D}I\right|\right|_\infty > \gamma D\sqrt{\frac{\log N}{N}}\right) < e^{-\delta D^2 \log N}$$

This bound is derived using the explicit expression for the marginal distribution $d\mu$. It applies to the 'typical' random unitary channels (the measure of the atypical channels goes to zero as $N \to \infty$).

Main Idea

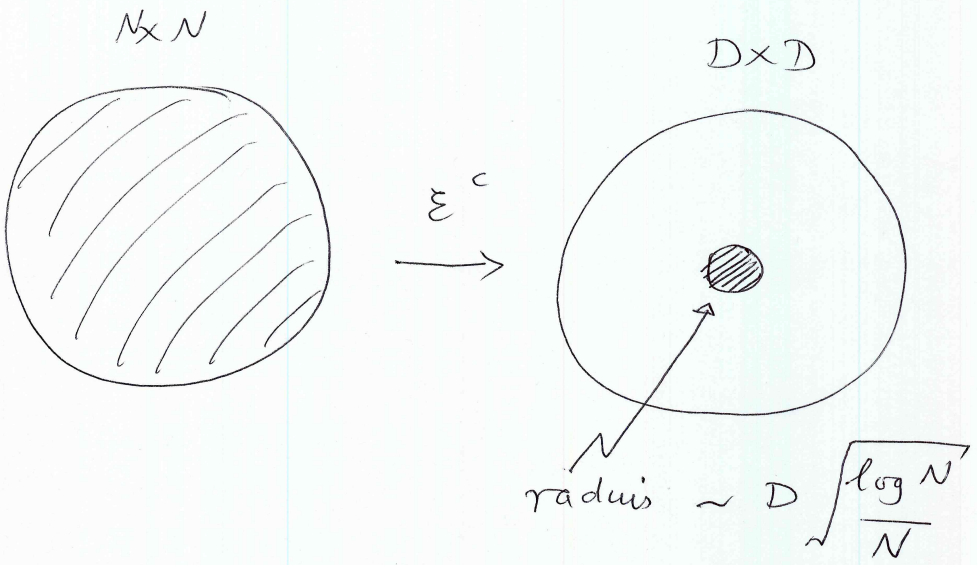
Suppose that $\mathcal{E}$ is a typical channel, then for 'most' input states $|\psi\rangle$ the output $\mathcal{E}^c(|\psi\rangle\langle\psi|)$ will be close to a maximally mixed state. This is implied by the second result above.

Hastings makes a stronger statement: for a typical channel there are constants γ, δ such that

$$\mathbb{P}_\psi\left(\left|\left|\mathcal{E}^c(|\psi\rangle\langle\psi|) - \frac{1}{D}I\right|\right|_\infty > \gamma D\sqrt{\frac{\log N}{N}}\right) < e^{-\delta D^2 \log N}$$

This bound is derived using the explicit expression for the marginal distribution $d\mu$. It applies to the 'typical' random unitary channels (the measure of the atypical channels goes to zero as $N \to \infty$).

Typical channel maps most states to maximally mixed state



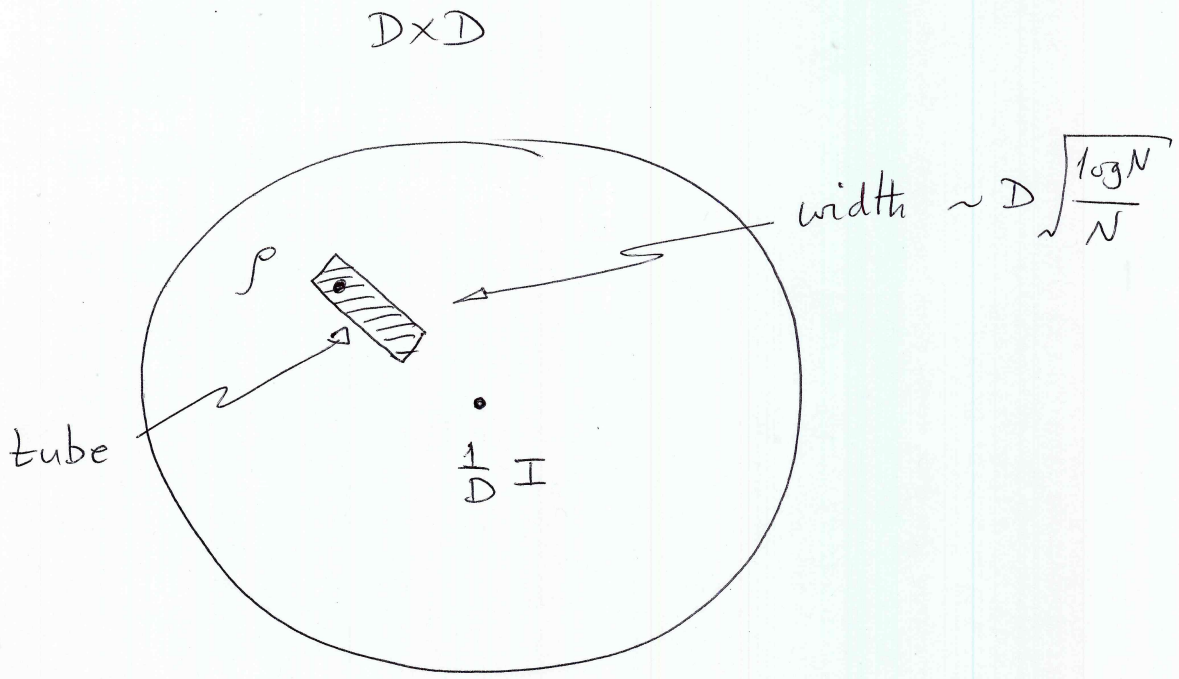
Tubular neighborhood

Now suppose that $|\psi\rangle$ is an 'exceptional state' with small entropy which violates the bound. Hastings shows that it cannot be isolated, and there must be a neighborhood of such states.

For any state $\rho \in \mathbb{C}^{D\times D}$ define the 'tube':

$$\text{tube}(\rho) = \left\{ y\rho + (1-y)\frac{1}{D}I + \sigma \,\middle|\, y \geq 1/2,\ ||\sigma||_\infty \leq \gamma D\sqrt{\frac{\log N}{N}} \right\}$$

The tube around a state ρ in the image of $\mathcal{E}^c$



Theorem

For a typical channel $\mathcal{E}$,

$$\mathbb{P}_\theta\left(\mathcal{E}^c(|\theta\rangle\langle\theta|) \in \textit{tube}(\rho) \,\middle|\, \exists\, |\psi\rangle \ \textit{s.t.}\ \rho = \mathcal{E}^c(|\psi\rangle\langle\psi|)\right) \geq e^{-kN}$$

Why is the bound true?

Fix a state $|\psi\rangle \in \mathbb{C}^N$ so that $\rho = \mathcal{E}^c(|\psi\rangle\langle\psi|)$, then for a random input state $|\theta\rangle$ we can write

$$|\theta\rangle = \langle\psi|\theta\rangle\,|\psi\rangle + |v\rangle$$

where $\langle\psi|v\rangle = 0$. Since $|\theta\rangle$ is random in $\mathbb{C}^N$, the state $|v\rangle$ is random in $\mathbb{C}^{N-1}$. Furthermore, most random states in $\mathbb{C}^N$ are almost orthogonal to $|\psi\rangle$ anyway:

$$\mathbb{P}_\theta\left(|\langle\psi|\theta\rangle|^2 > O(\frac{1}{\sqrt{N}})\right) \leq e^{-k\sqrt{N}}$$

So we get a small error by replacing $|v\rangle$ by an independent random state $|w\rangle \in \mathbb{C}^N$:

$$|\theta\rangle \simeq \langle\psi|\theta\rangle\,|\psi\rangle + |w\rangle$$

Why is the bound true?

Fix a state $|\psi\rangle \in \mathbb{C}^N$ so that $\rho = \mathcal{E}^c(|\psi\rangle\langle\psi|)$, then for a random input state $|\theta\rangle$ we can write

$$|\theta\rangle = \langle\psi|\theta\rangle\,|\psi\rangle + |v\rangle$$

where $\langle\psi|v\rangle = 0$. Since $|\theta\rangle$ is random in $\mathbb{C}^N$, the state $|v\rangle$ is random in $\mathbb{C}^{N-1}$. Furthermore, most random states in $\mathbb{C}^N$ are almost orthogonal to $|\psi\rangle$ anyway:

$$\mathbb{P}_\theta\left(|\langle\psi|\theta\rangle|^2 > O(\frac{1}{\sqrt{N}})\right) \leq e^{-k\sqrt{N}}$$

So we get a small error by replacing $|v\rangle$ by an independent random state $|w\rangle \in \mathbb{C}^N$:

$$|\theta\rangle \simeq \langle\psi|\theta\rangle\,|\psi\rangle + |w\rangle$$

Now since $\mathcal{E}$ is a typical channel, with high probability the state $|w\rangle\langle w|$ is mapped close to the state $\frac{1}{D}I$, and also with high probability $|\langle\psi|w\rangle|^2$ is small. So this leads to the following result:

$$\mathcal{E}^c(|\theta\rangle\langle\theta|) = |\langle\psi|\theta\rangle|^2\,\mathcal{E}^c(|\psi\rangle\langle\psi|) + \left(1 - |\langle\psi|\theta\rangle|^2\right)\frac{1}{D}I + \sigma$$

where

$$\mathbb{P}_\theta\left(||\sigma||_\infty > \gamma D\sqrt{\frac{\log N}{N}}\right) < e^{-\delta D^2 \log N}$$

The picture is that most states $|\theta\rangle$ are mapped into a small ball of radius $\sim D\sqrt{\frac{\log N}{N}}$ around the point

$$y\mathcal{E}^c(|\psi\rangle\langle\psi|) + (1-y)\frac{1}{D}I, \qquad y = |\langle\psi|\theta\rangle|^2$$

Restrict to states with $y = |\langle\psi|\theta\rangle|^2 \geq 1/2$: these are mapped into tube(ρ).

Note that

$$\mathbb{P}_\theta\left(|\langle\psi|\theta\rangle|^2 > 1/2\right) = \left(\frac{1}{2}\right)^{(2N-3)}$$

and this leads directly to the bound: for any state $|\psi\rangle$,

$$\mathbb{P}_\theta\left(\mathcal{E}^c(|\theta\rangle\langle\theta|) \in \text{tube}(\rho) \,\middle|\, \exists\, |\psi\rangle\ \rho = \mathcal{E}^c(|\psi\rangle\langle\psi|)\right) \geq e^{-kN}$$

Apply to minimum output entropy

Now suppose that $S_{min}(\mathcal{E}) < \log D - \frac{h}{D}$, and let $|\psi_0\rangle$ be a state for which

$$S\Big(\mathcal{E}^c(|\psi_0\rangle\langle\psi_0|)\Big) < \log D - \frac{h}{D}$$

Then we have the bound

$$\mathbb{P}_\theta\left(\mathcal{E}^c(|\theta\rangle\langle\theta|) \in \text{tube}(\mathcal{E}^c(|\psi_0\rangle\langle\psi_0|)) \,\middle|\, S_{min}(\mathcal{E}) < \log D - \frac{h}{D}\right) \geq e^{-kN}$$

Define

$$\begin{aligned} A &= \left\{ S_{min}(\mathcal{E}) < \log D - \frac{h}{D} \right\} \\ B &= \{ \mathcal{E}^c(|\theta\rangle\langle\theta|) \in \text{tube}(\mathcal{E}^c(|\psi_0\rangle\langle\psi_0|)) \} \end{aligned}$$

Then

$$\mathbb{P}_{\mathcal{E}}(A) = \mathbb{P}_{\mathcal{E},\theta}(A) = \frac{\mathbb{P}_{\mathcal{E},\theta}(B \textbf{ and } A)}{\mathbb{P}_{\mathcal{E},\theta}(B \mid A)}$$

The denominator is bounded below by first conditioning on the event that $\mathcal{E}$ is typical, then using the bounds above which hold for a typical channel. This leads to

$$\mathbb{P}_{\mathcal{E}}\left(S_{min}(\mathcal{E}) < \log D - \frac{h}{D} \right) \leq c\ e^{kN}\ \mathbb{P}_{\mathcal{E},\theta}(B \textbf{ and } A)$$

Final hard estimate

This last joint probability is estimated by using the fact that when $\mathcal{E}$ and θ are chosen randomly, the state $\mathcal{E}^c(|\psi_0\rangle\langle\psi_0|)$ has the same distribution as the reduced density matrix of a randomly chosen pure state in $\mathbb{C}^{N\times D}$.

Proposition

For $D, N/D$ sufficiently large,

$$\mathbb{P}_{\mathcal{E},\theta}\left(\mathcal{E}^c(|\theta\rangle\langle\theta|) \in \mathit{tube}(\mathcal{E}^c(|\psi_0\rangle\langle\psi_0|)) \textbf{ and } S_{min}(\mathcal{E}) < \log D - \frac{h}{D}\right) \leq K\, e^{-N\sqrt{h}/\log h}$$

Put this together with previous:

$$\mathbb{P}_{\mathcal{E}}\left(S_{min}(\mathcal{E}) < \log D - \frac{h}{D}\right) \leq c\, e^{kN}\, K\, e^{-N\sqrt{h}/\log h} < 1$$

for h, N sufficiently large.

A few words about the last bound ...

Use explicit formulas for distribution μ of eigenvalues of reduced density matrix.

The density of μ is

$$d\mu(\lambda_1, \ldots, \lambda_D) = Z_{N,D}^{-1}\, \delta(1 - \sum \lambda_i) \prod_{i<j} (\lambda_i - \lambda_j)^2 \prod_{i=1}^{D} \lambda_i^{N-D} \theta(\lambda_i)\, [d\lambda]$$

The normalization is bounded by

$$Z_{N,D}^{-1} \leq e^{2D^2 \log N + ND \log D}$$

Combining with the factor $\prod_{i=1}^{D} \lambda_i^{N-D}$ leads to the bound

$$\int f\, d\mu \leq e^{kD^2 \log N} \int f(\lambda)\, e^{N(\sum_{i=1}^{D} \log(D\lambda_i))}\, \delta(1 - \sum \lambda_i)\, [d\lambda]$$

Applying this to the estimation of the probability that $\mathcal{E}^c(|\theta\rangle\langle\theta|)$ belongs to tube(ρ) where $S(\rho) < \log D - h/D$ leads to this optimization problem: find

$$\sup\{\sum_{i=1}^{D} \log(D\lambda_i)\}$$

subject to the constraints

$$\lambda_i = y\, p_i + (1-y)\,\frac{1}{D}, \qquad 1/2 \le y \le 1,$$

$$\sum_{i=1}^{D} p_i \log(Dp_i) > \frac{h}{D}, \qquad \sum_{i=1}^{D} p_i = 1, \quad \log(D\lambda_i) \ge -\sqrt{h}$$

The solution is

$$\sum_{i=1}^{D} \log(D\lambda_i) \leq \frac{C}{\sqrt{h}} \sum_{i=1}^{D} \log(Dp_i)$$

where

$$p_1 \simeq \frac{h}{D \log h}, \qquad p_2 = \cdots = p_D \simeq \frac{1}{D} - \frac{h}{D^2 \log h}$$

leading to the value

$$\sup\{\sum_{i=1}^{D} \log(D\lambda_i)\} \simeq -C\frac{\sqrt{h}}{\log h}$$

Summary

This is a broad outline of the construction of the counterexamples. The proof is non-constructive, and does not provide any explicit examples.

Result is quite delicate and relies on detailed comparisons of quantities of the same order of magnitude. It would be interesting to find classes of counterexamples where the gap is larger.