One-shot Capacity Bounds via Hypothesis Testing

Ligong Wang — MIT

Renato Renner — ETH Zurich

Beyond i.i.d. Workshop
Cambridge, 9 Jan. 2013

PRL 108, 200501 (2012); arXiv:1007.5456

Outline

Goal:

- To find "good" bounds on the "one-shot capacity"
  —the maximum number of classical bits that can be transmitted through a single use of a quantum channel with average error probability ϵ

Outline

Goal:

- To find "good" bounds on the "one-shot capacity"
  —the maximum number of classical bits that can be transmitted through a single use of a quantum channel with average error probability ϵ

Both expressions and proofs (of the bounds) come from hypothesis testing.

Outline

Goal:

- To find "good" bounds on the "one-shot capacity"
  —the maximum number of classical bits that can be transmitted through a single use of a quantum channel with average error probability ϵ

Both expressions and proofs (of the bounds) come from hypothesis testing.

Also: yield a very simple proof for the HSW Theorem, as well as other capacity formulas.

Previous and Recent Work

Classical:

- Polyanskiy, Poor & Verdú, *ISIT 2008* —lower bound.
- Wang, Colbeck & Renner, *ISIT 2009* —lower bound (almost equivalent to PPV '08) and upper bound.
- Polyanskiy, Poor & Verdú, *IEEE Trans. Infor. Theory 2010* —many, many bounds, including (similar) bounds of all above.
- Many of these ideas actually date back to Strassen '62.

Previous and Recent Work

Classical:

- Polyanskiy, Poor & Verdú, *ISIT 2008* —lower bound.
- Wang, Colbeck & Renner, *ISIT 2009* —lower bound (almost equivalent to PPV '08) and upper bound.
- Polyanskiy, Poor & Verdú, *IEEE Trans. Infor. Theory 2010* —many, many bounds, including (similar) bounds of all above.
- Many of these ideas actually date back to Strassen '62.

Quantum:

- Hayashi & Nagaoka *IEEE Trans. Infor. Theory 2010* —information-spectrum approach to channel capacity; has a key lemma used in our paper.
- One-shot bounds of Mosonyi & Datta '09, Buscemi & Datta '10, and Renes & Renner '11, etc.

A Simple Hypothesis Testing Problem

When $H = 0$, state is ρ; when $H = 1$, state is σ.

Problem: minimize $\Pr[\text{Guess } H = 0 | H = 1]$ subject to

$$\Pr[\text{Guess } H = 1 | H = 0] \leq \epsilon.$$

A Simple Hypothesis Testing Problem

When $H = 0$, state is ρ; when $H = 1$, state is σ.

Problem: minimize $\Pr[\text{Guess } H = 0 | H = 1]$ subject to

$$\Pr[\text{Guess } H = 1 | H = 0] \leq \epsilon.$$

The minimum is

$$\boxed{p_\epsilon^*(\rho|\sigma) \triangleq \inf_{\substack{Q:0\leq Q\leq I,\\ \operatorname{tr}(Q\rho)\geq 1-\epsilon}} \operatorname{tr}(Q\sigma)}$$

How to find the best Q? — semidefinite program, or Holevo '72

Define $D_{\mathsf{H}}^{\epsilon}(\rho\|\sigma)$

"Hypothesis Testing Relative Entropy"

$$\boxed{D_{\mathsf{H}}^{\epsilon}(\rho\|\sigma) \triangleq \sup_{\substack{Q:0\leq Q\leq I,\\ \mathrm{tr}(Q\rho)\geq 1-\epsilon}} \{-\log \mathrm{tr}\,(Q\sigma)\}}$$

Namely,

$$D_{\mathsf{H}}^{\epsilon}(\rho\|\sigma) = -\log p_{\epsilon}^{*}(\rho|\sigma)$$

Properties of $D_{\mathsf{H}}^{\epsilon}(\rho\|\sigma)$

$$\boxed{D_{\mathsf{H}}^{\epsilon}(\rho\|\sigma) \triangleq \sup_{\substack{Q:0\leq Q\leq I,\\ \mathrm{tr}(Q\rho)\geq 1-\epsilon}} \{-\log \mathrm{tr}\,(Q\sigma)\}}$$

- Positivity: for any ρ, σ, and $\epsilon \in [0,1]$,

  $$D_{\mathsf{H}}^{\epsilon}(\rho\|\sigma) \geq 0,$$

  with equality if (but not only if) $\rho = \sigma$ and $\epsilon = 0$.

- Monotonically increasing in ϵ

- **Data Processing Inequality:**

  $$D_{\mathsf{H}}^{\epsilon}(\rho\|\sigma) \geq D_{\mathsf{H}}^{\epsilon}(\mathcal{E}(\rho)\|\mathcal{E}(\sigma))$$

Quantum Stein's Lemma

For any ρ, σ, and any $\epsilon \in (0,1)$,

$$\boxed{\lim_{n\to\infty} \frac{1}{n} D_{\mathsf{H}}^{\epsilon}(\rho^{\otimes n} \| \sigma^{\otimes n}) = D(\rho\|\sigma)}$$

- Weak ($\epsilon \downarrow 0$) version due to Hiai & Petz '91
- Strong (any ϵ) version due to Ogawa & Nagaoka '00

A Useful Nonasymptotic Relation

For any ρ, σ, and any $\epsilon \in (0, 1)$,

$$D_{\mathsf{H}}^{\epsilon}(\rho\|\sigma) \leq \frac{D(\rho\|\sigma) + H_{\mathsf{b}}(\epsilon)}{1-\epsilon}$$

Classical Information over Quantum Channel

$$M \xrightarrow{\textbf{enc}} \xi \xrightarrow{\textbf{channel}} \rho \xrightarrow{\textbf{dec}} \hat{M}$$

Classical Information over Quantum Channel

$$M \xrightarrow{\textbf{enc}} \xi \xrightarrow{\textbf{channel}} \rho \xrightarrow{\textbf{dec}} \hat{M}$$

Observation: to derive capacity formula, can ignore ξ and only look at its **label**

Classical Information over Quantum Channel

$$M \xrightarrow{\textbf{enc}} \xi \xrightarrow{\textbf{channel}} \rho \xrightarrow{\textbf{dec}} \hat{M}$$

Observation: to derive capacity formula, can ignore ξ and only look at its **label**

Notations:

- Label of inputs: $|x\rangle\langle x|$ orthogonal ("classical") states acting on Hilbert space $\mathbb{A}$
- Output: ρ_x acting on Hilbert space $\mathbb{B}$
- "Effective" channel $\mathcal{W}$: CPM mapping operators on $\mathbb{A}$ to operators on $\mathbb{B}$

$$\mathcal{W}(|x\rangle\langle x|) = \rho_x$$

More Notations

For a PMF $\{p_x\}$ on input labels, denote

$$\pi^{\mathbb{AB}} \triangleq \sum_x p_x |x\rangle\langle x|^{\mathbb{A}} \otimes \rho_x^{\mathbb{B}}$$

$$\pi^{\mathbb{A}} \triangleq \sum_x p_x |x\rangle\langle x|$$

$$\pi^{\mathbb{B}} \triangleq \sum_x p_x \rho_x$$

More Notations

For a PMF $\{p_x\}$ on input labels, denote

$$\pi^{\mathbb{AB}} \triangleq \sum_x p_x |x\rangle\langle x|^{\mathbb{A}} \otimes \rho_x^{\mathbb{B}}$$

$$\pi^{\mathbb{A}} \triangleq \sum_x p_x |x\rangle\langle x|$$

$$\pi^{\mathbb{B}} \triangleq \sum_x p_x \rho_x$$

Note: the "Holevo Information" is

$$\begin{aligned} D(\pi^{\mathbb{AB}} \| \pi^{\mathbb{A}} \otimes \pi^{\mathbb{B}}) &= I(\mathbb{A};\mathbb{B}) \\ &= H(\pi^{\mathbb{A}}) + H(\pi^{\mathbb{B}}) - H(\pi^{\mathbb{AB}}) \\ &= H\left(\sum_x p_x \rho_x\right) - \sum_x p_x H(\rho_x) \end{aligned}$$

Result

Theorem

The maximum number R^* of bits that can be transmitted through a channel with average error probability no more than ϵ satisfies

$$\sup_{\{p_x\}} D_{\mathsf{H}}^{\epsilon'}\left(\pi^{\mathbb{AB}} \| \pi^{\mathbb{A}} \otimes \pi^{\mathbb{B}}\right) - \log \frac{4\epsilon}{(\epsilon - \epsilon')^2} \leq R^* \leq \sup_{\{p_x\}} D_{\mathsf{H}}^{\epsilon}\left(\pi^{\mathbb{AB}} \| \pi^{\mathbb{A}} \otimes \pi^{\mathbb{B}}\right)$$

for any $\epsilon' < \epsilon$.

Consequences

Theorem (General Capacity Formula)

The capacity of an **arbitrary** *channel, described by a sequence of CPMs from* $\mathbb{A}^{\otimes n}$ *to* $\mathbb{B}^{\otimes n}$*, is*

$$C = \lim_{\epsilon \downarrow 0} \varliminf_{n \to \infty} \frac{1}{n} \sup_{P_{X^n}} D_{\mathsf{H}}^{\epsilon}(\pi^{\mathbb{A}^{\otimes n}\mathbb{B}^{\otimes n}} \| \pi^{\mathbb{A}^{\otimes n}} \otimes \pi^{\mathbb{B}^{\otimes n}}).$$

(Equivalent to information-spectrum formulation of Hayashi & Nagaoka '03)

Consequences

Theorem (General Capacity Formula)

The capacity of an **arbitrary** *channel, described by a sequence of CPMs from* $\mathbb{A}^{\otimes n}$ *to* $\mathbb{B}^{\otimes n}$*, is*

$$C = \lim_{\epsilon \downarrow 0} \varliminf_{n \to \infty} \frac{1}{n} \sup_{P_{X^n}} D_{\mathsf{H}}^{\epsilon}(\pi^{\mathbb{A}^{\otimes n}\mathbb{B}^{\otimes n}} \| \pi^{\mathbb{A}^{\otimes n}} \otimes \pi^{\mathbb{B}^{\otimes n}}).$$

(Equivalent to information-spectrum formulation of Hayashi & Nagaoka '03)

For a **memoryless** channel $\Longrightarrow$ **HSW Theorem**:

$$C = \sup_{b} \frac{1}{b} \max I(\mathbb{A}^b; \mathbb{B}^b).$$

- Achievability comes from IID blocks of inputs + Quantum Stein's Lemma
- Converse comes from that "useful nonasymptotic relation"

Consequences (contd.)

Optimistic capacity:

$$\overline{C} = \lim_{\epsilon \downarrow 0} \overline{\lim_{n \to \infty}} \frac{1}{n} \sup_{P_{X^n}} D_{\mathsf{H}}^{\epsilon}(\pi^{\mathbb{A}^{\otimes n}\mathbb{B}^{\otimes n}} \| \pi^{\mathbb{A}^{\otimes n}} \otimes \pi^{\mathbb{B}^{\otimes n}}).$$

Bounds on ϵ-capacities:

$$C_{\epsilon} \leq \underline{\lim_{n \to \infty}} \frac{1}{n} \sup_{P_{X^n}} D_{\mathsf{H}}^{\epsilon}(\pi^{\mathbb{A}^{\otimes n}\mathbb{B}^{\otimes n}} \| \pi^{\mathbb{A}^{\otimes n}} \otimes \pi^{\mathbb{B}^{\otimes n}}),$$

$$C_{\epsilon} \geq \lim_{\epsilon' \uparrow \epsilon} \underline{\lim_{n \to \infty}} \frac{1}{n} \sup_{P_{X^n}} D_{\mathsf{H}}^{\epsilon'}(\pi^{\mathbb{A}^{\otimes n}\mathbb{B}^{\otimes n}} \| \pi^{\mathbb{A}^{\otimes n}} \otimes \pi^{\mathbb{B}^{\otimes n}}).$$

(Equivalent to corresponding information-spectrum formulations)

Upper Bound

Theorem (Upper Bound)

If one can send R bits through one use of a quantum channel with average error probability no larger than ϵ, then

$$R \leq \sup_{\{p_x\}} D_{\mathsf{H}}^{\epsilon}\left(\pi^{\mathbb{A}\mathbb{B}} \| \pi^{\mathbb{A}} \otimes \pi^{\mathbb{B}}\right).$$

Upper Bound—the Simple Proof

A code is given: input labels $\{x_1, \ldots, x_{2^R}\}$; corresponding decoding POVM $\{E_1, \ldots, E_{2^R}\}$.

- Choose a uniform distribution on the input labels, yielding the state

$$\pi^{\mathbb{AB}} = 2^{-R} \sum_{i=1}^{2^R} |x_i\rangle\langle x_i|^{\mathbb{A}} \otimes \rho_{x_i}^{\mathbb{B}}$$

Upper Bound—the Simple Proof

A code is given: input labels $\{x_1, \ldots, x_{2^R}\}$; corresponding decoding POVM $\{E_1, \ldots, E_{2^R}\}$.

- Choose a uniform distribution on the input labels, yielding the state

$$\pi^{\mathbb{AB}} = 2^{-R} \sum_{i=1}^{2^R} |x_i\rangle\langle x_i|^{\mathbb{A}} \otimes \rho_{x_i}^{\mathbb{B}}$$

- Let

$$Q^{\mathbb{AB}} \triangleq \sum_{i=1}^{2^R} |x_i\rangle\langle x_i|^{\mathbb{A}} \otimes E_i^{\mathbb{B}}$$

Upper Bound—the Simple Proof (contd.)

$$\pi^{\mathbb{AB}} = 2^{-R} \sum_{i=1}^{2^R} |x_i\rangle\langle x_i|^{\mathbb{A}} \otimes \rho_{x_i}^{\mathbb{B}}, \qquad Q^{\mathbb{AB}} = \sum_{i=1}^{2^R} |x_i\rangle\langle x_i|^{\mathbb{A}} \otimes E_i^{\mathbb{B}}$$

Because average error probability is no larger than ϵ we have

$$\mathrm{tr}\left(Q\pi^{\mathbb{AB}}\right) = 2^{-R} \sum_{i=1}^{2^R} \mathrm{tr}\left(E_i \rho_{x_i}\right) \geq 1 - \epsilon.$$

Upper Bound—the Simple Proof (contd.)

$$\pi^{\mathbb{AB}} = 2^{-R} \sum_{i=1}^{2^R} |x_i\rangle\langle x_i|^{\mathbb{A}} \otimes \rho_{x_i}^{\mathbb{B}}, \qquad Q^{\mathbb{AB}} = \sum_{i=1}^{2^R} |x_i\rangle\langle x_i|^{\mathbb{A}} \otimes E_i^{\mathbb{B}}$$

Because average error probability is no larger than ϵ we have

$$\operatorname{tr}\left(Q\pi^{\mathbb{AB}}\right) = 2^{-R} \sum_{i=1}^{2^R} \operatorname{tr}\left(E_i \rho_{x_i}\right) \geq 1 - \epsilon.$$

On the other hand,

$$\begin{aligned} \operatorname{tr}\left(Q(\pi^{\mathbb{A}} \otimes \pi^{\mathbb{B}})\right) &= 2^{-R} \operatorname{tr}\left(\left\{\sum_{i=1}^{2^R} E_i\right\} \pi^{\mathbb{B}}\right) \\ &= 2^{-R} \operatorname{tr}\left(I\pi^{\mathbb{B}}\right) \\ &= 2^{-R}. \end{aligned}$$

Upper Bound—the Simple Proof (contd.)

So, for the above chosen $\pi^{\mathbb{AB}}$ and $Q^{\mathbb{AB}}$,

$$\begin{aligned} R &= -\log \operatorname{tr}\left(Q^{\mathbb{AB}}\left(\pi^{\mathbb{A}} \otimes \pi^{\mathbb{B}}\right)\right) \\ &\leq \sup_{\substack{Q:0\leq Q\leq I,\\ \operatorname{tr}\left(Q\pi^{\mathbb{AB}}\right)\geq 1-\epsilon}} \left\{-\log \operatorname{tr}\left(Q\left(\pi^{\mathbb{A}} \otimes \pi^{\mathbb{B}}\right)\right)\right\} \\ &= D_{\mathsf{H}}^{\epsilon}(\pi^{\mathbb{AB}} \| \pi^{\mathbb{A}} \otimes \pi^{\mathbb{B}}). \end{aligned}$$

Q.E.D.

Upper Bound—the Simpler Proof

Consider the question: Are M and $\hat{M}$ generated **jointly** or **independently**?

Upper Bound—the Simpler Proof

Consider the question: Are M and $\hat{M}$ generated **jointly** or **independently**?

Guessing rule (possibly suboptimal): Guess "jointly" if $m = \hat{m}$ and guess "independently" otherwise.

Upper Bound—the Simpler Proof

Consider the question: Are M and $\hat{M}$ generated **jointly** or **independently**?

Guessing rule (possibly suboptimal): Guess "jointly" if $m = \hat{m}$ and guess "independently" otherwise.

Error probabilities:

- When M and $\hat{M}$ are generated jointly, the guess is wrong iff the channel decoder makes an error $\longrightarrow \epsilon$
- When M and $\hat{M}$ are generated independently, the guessing rule makes an error if they happen to be the same $\longrightarrow 2^{-R}$.

Upper Bound—the Simpler Proof

Consider the question: Are M and $\hat{M}$ generated **jointly** or **independently**?

Guessing rule (possibly suboptimal): Guess "jointly" if $m = \hat{m}$ and guess "independently" otherwise.

Error probabilities:

- When M and $\hat{M}$ are generated jointly, the guess is wrong iff the channel decoder makes an error $\longrightarrow \epsilon$
- When M and $\hat{M}$ are generated independently, the guessing rule makes an error if they happen to be the same $\longrightarrow 2^{-R}$.

So

$$2^{-R} \geq p_\epsilon^* \left(P_{M\hat{M}} | P_M \otimes P_{\hat{M}} \right) = 2^{-D_{\mathsf{H}}^\epsilon \left(P_{M\hat{M}} \| P_M \otimes P_{\hat{M}} \right)}.$$

Theorem follows by the Data Processing Inequality. *Q.E.D.*

Lower Bound

Theorem (Lower Bound)

For any $\epsilon' < \epsilon$, it is possible to send R bits through one use of a quantum channel with average error probability no larger than ϵ as long as

$$R \leq \sup_{\{p_x\}} D_{\mathsf{H}}^{\epsilon'}\left(\pi^{\mathbb{A}\mathbb{B}} \| \pi^{\mathbb{A}} \otimes \pi^{\mathbb{B}}\right) - \log \frac{4\epsilon}{(\epsilon - \epsilon')^2}.$$

Proof of Lower Bound

Another look at Shannon's Joint Typicality decoder:

For each codeword, make a hypothesis test: is it drawn jointly with the output y^n? (I.e. are they jointly typical?)

Proof of Lower Bound

Another look at Shannon's Joint Typicality decoder:

For each codeword, make a hypothesis test: is it drawn jointly with the output y^n? (I.e. are they jointly typical?)

Problem in quantum case:

Can't make 2^R tests on a **quantum** state, because measurement changes the state!

Proof of Lower Bound: A Lemma

Lemma [Hayashi & Nagaoka '03]

For any positive real c and any operators $0 \leq S \leq I$ and $T \geq 0$, we have

$$I - (S+T)^{-1/2} S (S+T)^{-1/2} \leq (1+c)(I-S) + (2+c+c^{-1})T.$$

Proof of Lower Bound: A Lemma

Lemma [Hayashi & Nagaoka '03]

For any positive real c and any operators $0 \leq S \leq I$ and $T \geq 0$, we have

$$I - (S+T)^{-1/2} S (S+T)^{-1/2} \leq (1+c)(I-S) + (2+c+c^{-1})T.$$

Application: for n hypothesis tests consisting of $\{A_i, I - A_i\}$, we can construct one measurement with n outcomes:

$$E_i = \left(\sum_{j=1}^{n} A_j\right)^{-\frac{1}{2}} A_i \left(\sum_{j=1}^{n} A_j\right)^{-\frac{1}{2}},$$

with

$$I - E_i \leq (1+c)(I - A_i) + (2+c+c^{-1}) \left(\sum_{j \neq i} A_j\right)$$

Proof of Lower Bound (contd.)

Fix $c > 0$ and $0 < \epsilon' < \epsilon/(1+c)$.

For any P_X, we will to show: for any $Q \leq I$ acting on $\mathbb{AB}$ such that $\operatorname{tr}\left(Q\pi^{\mathbb{AB}}\right) \geq 1 - \epsilon'$, there exist a codebook and a decoding POVM which satisfy

$$\Pr(\text{error}) \leq (1+c)\epsilon' + (2 + c + c^{-1})(2^R - 1) \operatorname{tr}\left(Q(\pi^{\mathbb{A}} \otimes \pi^{\mathbb{B}})\right).$$

Proof of Lower Bound (contd.)

Fix $c > 0$ and $0 < \epsilon' < \epsilon/(1+c)$.

For any P_X, we will to show: for any $Q \leq I$ acting on $\mathbb{AB}$ such that $\operatorname{tr}\left(Q\pi^{\mathbb{AB}}\right) \geq 1 - \epsilon'$, there exist a codebook and a decoding POVM which satisfy

$$\Pr(\text{error}) \leq (1+c)\epsilon' + (2 + c + c^{-1})(2^R - 1) \operatorname{tr}\left(Q(\pi^{\mathbb{A}} \otimes \pi^{\mathbb{B}})\right).$$

To this end, choose

$$A_x^{\mathbb{B}} \triangleq \operatorname{tr}_{\mathbb{A}}\left(\left(|x\rangle\langle x|^{\mathbb{A}} \otimes I^{\mathbb{B}}\right) Q\right).$$

Proof of Lower Bound (contd.)

Fix $c > 0$ and $0 < \epsilon' < \epsilon/(1+c)$.

For any P_X, we will to show: for any $Q \leq I$ acting on $\mathbb{AB}$ such that $\mathrm{tr}\left(Q\pi^{\mathbb{AB}}\right) \geq 1 - \epsilon'$, there exist a codebook and a decoding POVM which satisfy

$$\Pr(\text{error}) \leq (1+c)\epsilon' + (2 + c + c^{-1})(2^R - 1)\,\mathrm{tr}\left(Q(\pi^{\mathbb{A}} \otimes \pi^{\mathbb{B}})\right).$$

To this end, choose

$$A_x^{\mathbb{B}} \triangleq \mathrm{tr}_{\mathbb{A}}\left(\left(|x\rangle\langle x|^{\mathbb{A}} \otimes I^{\mathbb{B}}\right) Q\right).$$

Randomly generate a codebook $\{x_i\}$, and form a corresponding POVM

$$E_i = \left(\sum_{j=1}^{2^R} A_{x_j}\right)^{-\frac{1}{2}} A_{x_i} \left(\sum_{j=1}^{2^R} A_{x_j}\right)^{-\frac{1}{2}}.$$

Applying the lemma and doing simple matrix manipulations yield the result.

Q.E.D.

Summary

- Capacity of quantum channel can be well understood via hypothesis testing.
- Simple proof for the HSW Theorem; as well as general capacity formulas.
- $D_{\mathsf{H}}^{\epsilon}(\rho\|\sigma)$: a useful quantity, hopefully in other one-shot scenarios, too.

Thank you!