Quantum Entropy and Trace Inequalities

Relevant Bibliography

1) Eric Carlen, "Trace Inequalities and Quantum Entropy: An Introductory Course", 73-140 Contemp. Math. 529, Amer. Math. Soc. (2010).

2) Mark Wilde, "Quantum Information Theory", Cambridge University Press.

3) Marco Tomamichel, "Quantum Information Processing with Finite Resources", Springer Briefs in Mathematical Physics.

1. Basic Definitions and Notations

*) $\mathcal{M}_n$: Space of $n \times n$ matrices (also referred to as operators on $\mathbb{C}^n$)

*) $\langle \cdot, \cdot \rangle$: Inner product on $\mathbb{C}^n$, or other Hilbert spaces $\mathcal{H}, \mathcal{K}, \mathcal{L}$.

*) A^* : Hermitian conjugate or adjoint of A.

*) $\mathcal{B}(\mathcal{H})$: Bounded linear operators on $\mathcal{H}$.

*) $\mathcal{A}(\mathcal{H})$: Hermitian operators on $\mathcal{H}$.

*) $\mathcal{D}(\mathcal{H}) := \{ \rho \in \mathcal{B}(\mathcal{H}) \,/\, \rho = \rho^*, \ \rho \geq 0, \ \mathrm{Tr}[\rho] = 1 \}$

Density matrices on $\mathcal{H}$.

2 Possible Formulations To Describe Quantum States: Heisenberg ($|\psi\rangle \in \mathcal{H}$) vs. Schrödinger ($\rho \in \mathcal{D}(\mathcal{H})$)

Remark The set of density matrices on $\mathbb{C}^n$ is a convex set.
The extreme points of $\mathcal{D}(\mathbb{C}^n)$ are rank-one orthogonal projections on $\mathbb{C}^n$, called pure states. $|\psi\rangle\langle\psi|$, $|\psi\rangle \in \mathcal{H}$.

Remark Observables in quantum mechanical systems are, in general, operators on infinite-dimensional, separable Hilbert spaces.

Here we restrict to finite-dimensional spaces.

Reason? Density matrices are compact operators + Approximation property

For this to be effective, we need continuity of the entropic quantities $\longrightarrow$ See Nilanjana Datta's talk!

2. ENTROPIES

They are fundamental in classical and quantum information theory.

Classically: Given a probability distribution $p = \{p_1, \dots p_d\}$ with $0 \le p_x \le 1$ $\forall x \in \{1 \dots d\}$ and $\sum_{x=1}^{d} p_x = 1$,

Shannon Entropy: $$\boxed{H(\{p_x\}) := -\sum_{x=1}^{d} p_x \log p_x}$$

Note that "log" here means logarithm in base 2!

Quantumly: Given $\rho \in \mathcal{D}(\mathcal{H})$,

2.1) VON NEUMANN ENTROPY: $$\boxed{S(\rho) := -\mathrm{Tr}[\rho \log \rho]}$$

Interpretation? Since $\rho \geq 0 \Rightarrow \rho$ can be diagonalized (Spectral theorem)

$\Rightarrow \exists\, U$ unitary, D diagonal, such that $\rho = U D U^{-1}$

with $D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{pmatrix}$, $\{\lambda_1, \dots \lambda_d\}$ set of eigenvalues of ρ

Thus,

$$\boxed{\log \rho = U \begin{pmatrix} \log \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \log \lambda_d \end{pmatrix} U^{-1}}$$

Remark. This can be extended to more general functions f when defining $f(\rho) \Rightarrow$ Functional calculus.

$$\boxed{\begin{aligned} S(\rho) &= -\mathrm{Tr}[\rho \log \rho] = -\mathrm{Tr}[UDU^{-1} \log UDU^{-1}] \\ &= -\mathrm{Tr}[UDU^{-1}U \log D U^{-1}] = -\mathrm{Tr}[D \log D] = -\sum_{x=1}^{d} \lambda_x \log \lambda_x \end{aligned}}$$

Operational interpretation. Quantum information content per letter (minimum number of qubits per letter that are necessary to encode a message, i.e. data compression limit for a memoryless quantum source).

$[\log(x+1) \leq x]$

Other entropies

Rényi entropies $\boxed{\alpha \in (0,1) \cup (1,+\infty)}$ **Tsallis entropies**

$$\boxed{S_\alpha(\rho) := \frac{1}{1-\alpha} \log \mathrm{Tr}[\rho^\alpha]} \quad \boxed{\lim_{\alpha \to 1} S_\alpha(\rho) = S(\rho) = \lim_{\alpha \to 1} T_\alpha(\rho)} \quad \boxed{T_\alpha(\rho) := \frac{1}{\alpha - 1}\left(\mathrm{Tr}[\rho^\alpha] - 1\right)}$$

Before proving some fundamental properties of the von Neumann entropy, we need to introduce some notions / tools.

Definition: (Partial trace)

$\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ bipartite Hilbert space.

Given $O_{AB} \in \mathcal{B}(\mathcal{H}_{AB})$,

$O_A := \mathrm{Tr}_B[O_{AB}]$ is the unique operator on $\mathcal{B}(\mathcal{H}_A)$ such that

$$\boxed{\mathrm{Tr}[O_{AB}(Q_A \otimes \mathbb{1}_B)] = \mathrm{Tr}[O_A Q_A]}$$

for every $Q_A \in \mathcal{B}(\mathcal{H}_A)$. Note $\mathrm{Tr}[O_{AB}] = \mathrm{Tr}[O_A]$.

Technical tool: Schmidt decomposition, purity and purification.

If $|\psi\rangle \in \mathcal{H}_{AB}$, there exist orthonormal bases $\{|i_A\rangle\} \subseteq \mathcal{H}_A$, $\{|i_B\rangle\} \subseteq \mathcal{H}_B$ and coefficients $\lambda_i \geq 0$, $\sum_i \lambda_i^2 = 1$ such that:

$$\boxed{|\psi\rangle = \sum_i \lambda_i \, |i_A\rangle \otimes |i_B\rangle}$$

The λ_i are called the Schmidt coefficients, and the number of λ_i, with multiplicity, is the Schmidt rank.

Purification Given $\rho_A \in \mathcal{D}(\mathcal{H}_A)$, we can introduce an auxiliary system $\mathcal{H}_R$ and construct a pure state $|\psi_{AR}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_R$ such that

$$\boxed{\rho_A = \mathrm{Tr}_R\left[|\psi_{AR}\rangle\langle\psi_{AR}|\right]}$$

Connection between mixed states and pure states in a larger Hilbert space. $\dim(\mathcal{H}_R) = \dim(\mathcal{H}_A)$

PROPERTIES OF VON NEUMANN ENTROPY.

1) Purity $S(|\psi\rangle\langle\psi|) = 0$.

2) Maximum $S(\rho) \leq \log d$

$$\boxed{0 \leq S(\rho) \leq \log d} \quad \forall \rho \in \mathcal{D}(\mathcal{H})$$

$\left[S(\rho) = -\sum \lambda_x \log \lambda_x\right]$ "$= \log d$" $\iff \lambda_x = \frac{1}{d}$ $\forall x = 1 \dots d$.

Proof: Consequence of the fact that $t \mapsto t\log t$ is convex.

3) Unitary invariance $S(U\rho U^{-1}) = S(\rho)$

4) Concavity $p_i \geq 0$, $\sum_i p_i = 1$, $\rho_1 \dots \rho_n$ states:

$$\boxed{S\left(\sum_i p_i \rho_i\right) \geq \sum_i p_i S(\rho_i)}$$

5) <u>Additivity</u> $\rho_{AB} \in \mathcal{D}(\mathcal{H}_{AB})$, $\boxed{\rho_{AB} = \rho_A \otimes \rho_B}$:

$$S(\rho_{AB}) = S(\rho_A) + S(\rho_B)$$

6) <u>Subadditivity</u> For a general $\rho_{AB} \in \mathcal{D}(\mathcal{H}_{AB})$,

$$\boxed{S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B)}$$

(classical and quantum)

7) <u>Araki-Lieb-Thirring inequality</u> Using purification above:

$$S(\rho_C) \leq S(\rho_{BC}) + S(\rho_B)$$

and hence,

$$|S(\rho_A) - S(\rho_B)| \leq S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B)$$

(closest to the classical monotonicity: $S(\rho_{AB}) \geq \max\{S(\rho_A), S(\rho_B)\}$)

8) <u>Strong Subadditivity</u> (Lieb-Ruskai) $\mathcal{H}_{ABC} = \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$.

$$\boxed{S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC})}$$

<u>Tool in the proof</u> For Hermitian operators X and Y,

<u>Golden-Thompson</u>: $\boxed{\mathrm{Tr}[e^{X+Y}] \leq \mathrm{Tr}[e^X e^Y]}$

The direct extension to three operators is false!!

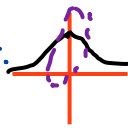
<u>Lieb</u>: $$\mathrm{Tr}[e^{X+Y+Z}] \leq \mathrm{Tr}\left[\int_0^\infty e^X \frac{1}{e^Z + t} e^Y \frac{1}{e^Z + t}\, dt\right]$$

Beautiful, but difficult to handle in general.

<u>Sutter-Berta-Tomamichel</u>: (Multivariate Trace Inequalities)

$$\mathrm{Tr}[e^{X+Y+Z+W}] \leq \mathrm{Tr}\left[\int_{-\infty}^{+\infty} dt\, p_0(t)\, e^X e^{Z\left(\frac{1+it}{2}\right)} e^Y e^{Z\left(\frac{1-it}{2}\right)}\right]$$

$e^{W\left(\frac{1+it}{2}\right)}$ $e^{W\left(\frac{1-it}{2}\right)}$

$p_0(t)$: 

Quantum Channels.

$\mathcal{H}, K$ Hilbert spaces. Let $\Phi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(K)$ be a linear map. Φ is a quantum channel if it is a completely positive, trace preserving map:

- Positive: $\boxed{\Phi(A) \geq 0}$ whenever $A \geq 0$.
- Trace preserving: $\boxed{\mathrm{Tr}[\Phi(A)] = \mathrm{Tr}[A]}$ $\forall A \in \mathcal{B}(\mathcal{H})$.
- Completely positive: $\boxed{\Phi \otimes \mathbb{1}_{\mathcal{B}(\mathcal{L})}}$ is positive for any other Hilbert space $\mathcal{L}$.

Physical maps.

Examples

1) Unitary evolution: $\rho \mapsto U \rho U^*$
2) Adding an ancilla: $\rho \mapsto \rho \otimes \rho_E$.
3) Partial trace.

Equivalent formulations.

$T : \underset{\mathbb{C}^d}{\mathcal{B}(\mathcal{H})} \to \underset{\mathbb{C}^{d'}}{\mathcal{B}(K)}$ linear. TFAE:

1) T is a quantum channel.
2) Choi-Jamiołkowski: For $C := (T \otimes \mathbb{1}_d)(|\phi\rangle\langle\phi|)$, $|\phi\rangle = \frac{1}{\sqrt{d}} \sum_{k=1}^{d} |kk\rangle$ max. entangled state.

   $C \geq 0$ and $\mathrm{Tr}_1[C] = \frac{\mathbb{1}}{d}$.
3) Kraus decomposition:

$$T(\rho) = \sum_{k=1}^{dd'} A_k \rho A_k^* ; \quad \text{with} \quad \sum_{k=1}^{dd'} A_k^* A_k = \mathbb{1}.$$

4) Stinespring dilation:

$$T(\rho) = \mathrm{Tr}_2\left[U(\rho \otimes |\psi\rangle\langle\psi|)U^*\right]$$

with U a unitary on $\mathbb{C}^d \otimes \mathbb{C}^{dd'}$ and $|\psi\rangle$ a state (for the environment).

2.2. Relative Entropy (Umegaki)

Given $\rho, \sigma \in \mathcal{D}(\mathcal{H})$, their (Umegaki) relative entropy is:

$$D(\rho \| \sigma) := \begin{cases} \mathrm{Tr}[\rho(\log\rho - \log\sigma)] & \text{if } \mathrm{Ker}(\sigma) \subseteq \mathrm{Ker}(\rho) \\ +\infty & \text{otherwise.} \end{cases}$$

Remark It is one of the possible quantum extensions of the classical Kullback-Leibler divergence:

$\{p_x\}_{x=1}^d$, $\{q_x\}_{x=1}^d$, $0 \le p_x \le 1$, $0 \le q_x \le 1$, $\sum p_x = 1$, $\sum q_x = 1$

$$KL(\{p_x\}, \{q_x\}) = \sum_x p_x \log \frac{p_x}{q_x}$$

$\log \rho^{1/2}\sigma^{-1}\rho^{1/2}$

It is a measure of distinguishability between quantum states.

Properties

1) Unitary invariance: $D(\rho\|\sigma) = D(U\rho U^* \| U\sigma U^*)$

2) Non-negativity: $D(\rho\|\sigma) \geq 0$ and "$=0$" iff $\rho = \sigma$.

3) Continuity: $\rho \mapsto D(\rho\|\sigma)$ is continuous.

4) Additivity: ρ_{AB}, σ_{AB}. $D(\rho_A \otimes \rho_B \| \sigma_A \otimes \sigma_B) = D(\rho_A\|\sigma_A) + D(\rho_B\|\sigma_B)$

5) Superadditivity: $D(\rho_{AB} \| \sigma_A \otimes \sigma_B) \geq D(\rho_A \| \sigma_A) + D(\rho_B \| \sigma_B)$

6) Data-processing inequality For T a quantum channel, "=" (Petz)

$$D(\rho \| \sigma) \geq D(T(\rho) \| T(\sigma))$$

(Lindblad)

$D(\rho \| \sigma) - D(T(\rho) \| T(\sigma)) \geq \underset{\geq 0}{f(\rho, \sigma)}$

Axiomatic Characterization

Properties 3) — 6) characterize the relative entropy (Matsumoto, Wilming-Gallego-Eisert)

Equivalence

TFAE:

1) Data-processing inequality

$$D(\rho \| \sigma) \geq D(T(\rho) \| T(\sigma))$$

2) Joint convexity of the RE.

$$D\left(\sum_x \lambda_x \rho_x \,\Big\|\, \sum_x \lambda_x \sigma_x\right) \leq \sum_x \lambda_x D(\rho_x \| \sigma_x).$$

3) Strong Subadditivity of vNE.

$$S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC}).$$

Operational interpretation

In the task of asymmetric hypothesis testing, we consider $\rho^{\otimes n}$, $\sigma^{\otimes n}$, T_n a hypothesis test (channel).

$\beta_n(T_n; \sigma) := \mathrm{Tr}[\sigma^{\otimes n} T_n]$ is the second kind error (wrongly concluding that the state is $\rho^{\otimes n}$)

Then (quantum Stein lemma), $D(\rho \| \sigma) = \lim_{n \to \infty} -\frac{1}{n} \log \beta_n$

2.3. DIVERGENCES.

Definition: A functional $D: S \times P \to \mathbb{R}$ is called a generalized divergence if it satisfies the DPI:

$$D(\mathcal{N}(\rho) \,\|\, \mathcal{N}(\sigma)) \leq D(\rho \,\|\, \sigma)$$

Standard f-divergences $D_f(\rho\|\sigma) = \mathrm{Tr}[\sigma^{1/2} f(L_\rho R_{\sigma^{-1}}) \sigma^{1/2}]$	Maximal f-divergences. $\hat{D}_f(\rho\|\sigma) = \mathrm{Tr}[\sigma f(\sigma^{-1/2} \rho \sigma^{-1/2})]$
(Sandwiched) Rényi divergences $\tilde{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1} \log \mathrm{Tr}\left[\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}}\right]^\alpha$	Maximal Rényi divergences (or geometric) $\hat{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1} \log \mathrm{Tr}\left[\sigma^{1/2} (\sigma^{-1/2} \rho \sigma^{-1/2})^\alpha \sigma^{1/2}\right]$
$\downarrow \alpha \to 1$	$\downarrow \alpha \to 1$
(Umegaki) Relative entropy $D(\rho\|\sigma) = \mathrm{Tr}[\rho(\log\rho - \log\sigma)]$	Belavkin-Staszewski relative entropy $D_{BS}(\rho\|\sigma) = \mathrm{Tr}[\rho \log(\rho^{1/2} \sigma^{-1} \rho^{1/2})]$

$\uparrow \alpha \to 1$

Petz Rényi divergence

$$\bar{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1} \log \mathrm{Tr}[\rho^\alpha \sigma^{1-\alpha}]$$

$\uparrow$

It provides optimal exponential rate for the error commited in the task of binary hypothesis testing when considering errors of kinds first and second jointly.

⚠ Geometric Rényi divergence because it is defined from the weighted matrix geometric mean:

$$\hat{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1} \log \mathrm{Tr}\, G_{1-\alpha}(\rho\|\sigma)$$

$$G_\alpha(\rho\|\sigma) = \rho^{1/2} (\rho^{-1/2} \sigma \rho^{-1/2})^\alpha \rho^{1/2}$$

Inequalities:

*) BS- Relative entropies:

$$\boxed{D(\rho\|\sigma) \le D_{BS}(\rho\|\sigma)}$$

*) Rényi divergences: $\forall \alpha \in (1,\infty)$

$$\boxed{\tilde{D}_\alpha(\rho\|\sigma) \le \overline{D}_\alpha(\rho\|\sigma)}$$

*) Sandwiched Rényi divergences: $\forall \alpha \in (1,\infty)$

$$\boxed{D(\rho\|\sigma) = \lim_{\alpha\to 1} \tilde{D}_\alpha(\rho\|\sigma) \le \tilde{D}_\alpha(\rho\|\sigma) \le \lim_{\alpha\to\infty} \tilde{D}_\alpha(\rho\|\sigma) = D_{max}(\rho\|\sigma)}$$

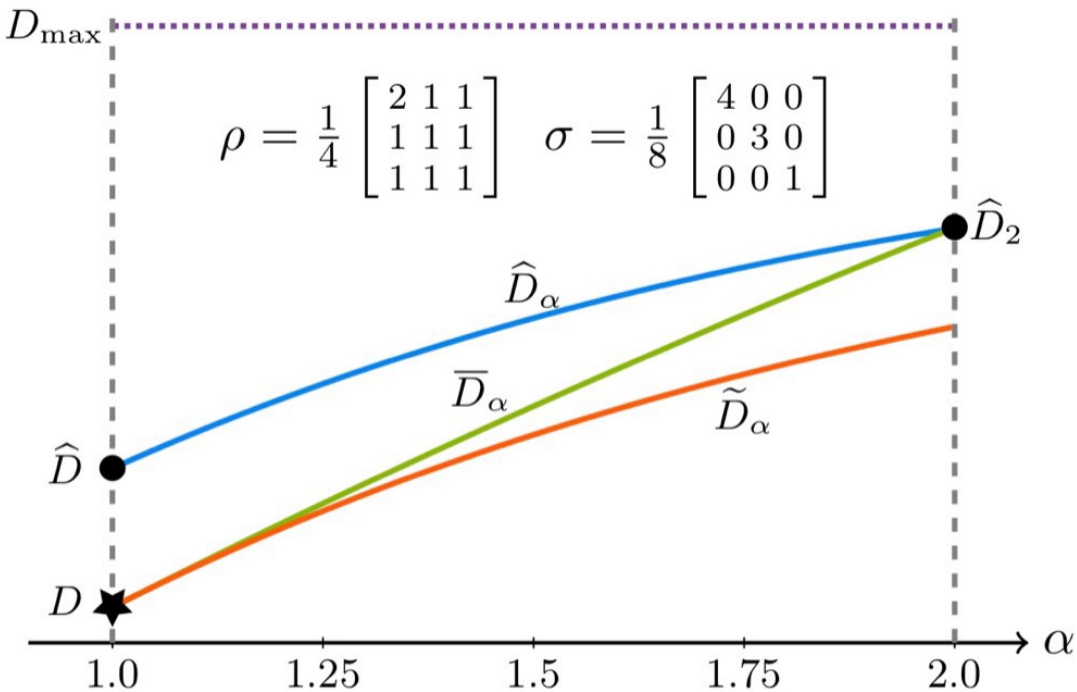
where $D_{max}(\rho\|\sigma) = \min\{\log t \mid \rho \le t\sigma\}$.

*) Geometric Rényi divergences: $\forall \alpha \in (1,2]$

$$\boxed{D_{BS}(\rho\|\sigma) \le \hat{D}_\alpha(\rho\|\sigma)}$$

*) Comparison with D and Dmax: $\forall \alpha \in (1,2]$

$$\boxed{D(\rho\|\sigma) \le \hat{D}_\alpha(\rho\|\sigma) \le D_{max}(\rho\|\sigma)}$$



3. Application to Quantum Capacities.

Quantum capacity: Maximum rate at which a noisy quantum channel can reliably transmit quantum information over asymptotically many uses of the channel.

Two different quantum capacities:

*) (Unassisted) quantum capacity Q.

*) Two-way assisted quantum capacity $Q^{\leftrightarrow}$.

Quantum capacity theorem:

Quantum capacity = Regularized channel coherent information.

$$\boxed{Q(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n} I_c(\mathcal{N}^{\otimes n}) = \sup_{n\in\mathbb{N}} \frac{1}{n} I_c(\mathcal{N}^{\otimes n})}$$

where $I_c(\mathcal{N}) = \max_{\rho\in S} \left[S(\mathcal{N}(\rho)) - S(\mathcal{N}^c(\rho)) \right]$, for $\mathcal{N}^c$ the complementary channel of $\mathcal{N}$.

⚠ Because of the regularization, the quantum capacity is difficult to evaluate.

3.1. Unassisted Quantum Capacity.

Converse (upper) bounds on the unassisted quantum capacity.

Definition For any divergence D, the generalized Rains bound of a quantum state ρ_{AB} is defined as:

$$\boxed{R(\rho_{AB}) = \min_{\sigma_{AB}\in PPT'(A:B)} D(\rho_{AB} \| \sigma_{AB})}$$

where $PPT'(A:B) = \{ \sigma_{AB} \mid \sigma_{AB} \geq 0,\ \| \sigma_{AB}^{T_B} \|_1 \leq 1 \}$.

Definition The generalized Rains Information is defined as:

$$\boxed{R(\mathcal{N}) = \max_{\rho_A \in S(\mathcal{H}_A)} \min_{\sigma_{AB} \in PPT'(A:B)} D(\mathcal{N}_{A' \to B}(\phi_{AA'}) \| \sigma_{AB})}$$

where $\phi_{AA'}$ is a purification of ρ_A.

Definition The max-Rains Information is induced by the geometric Rényi divergence.

$$\boxed{R_{max}(\mathcal{N}) = \max_{\rho_A \in S(\mathcal{H}_A)} \min_{\sigma_{AB} \in PPT'(A:B)} D_{max}(\mathcal{N}_{A' \to B}(\phi_{AA'}) \| \sigma_{AB})}$$

We denote by $\hat{R}_\alpha$ the generalized Rains Information induced by the geometric Rényi divergence:

$$\boxed{\hat{R}_\alpha(\mathcal{N}) = \max_{\rho_A \in S(\mathcal{H}_A)} \min_{\sigma_{AB} \in PPT'(A:B)} \hat{D}_\alpha(\mathcal{N}_{A' \to B}(\phi_{AA'}) \| \sigma_{AB})}$$

Theorem For any quantum channel $\mathcal{N}$ and $\alpha \in (1,2]$:

$$\boxed{Q(\mathcal{N}) \leq Q^+(\mathcal{N}) \leq R(\mathcal{N}) \leq \hat{R}_\alpha(\mathcal{N}) \leq R_{max}(\mathcal{N})}$$

where $Q^+(\mathcal{N})$ is the strong converse capacity of $\mathcal{N}$.