Search for an analogue of the sub-additive–doubling–rotation proof of Gaussian optimality in quantum systems

Jinpei Zhao

Dept. of Information Engg., The Chinese University of Hong Kong

25 September 2025

Basic notation

Shannon Entropy: Given $X \sim p_X$,

$$H(X) := -\sum_x p_X \log p_X.$$

Differential Entropy: Given $X \sim f_X(x)$,

$$h(X) := -\int f_X(x) \log f_X(x) dx$$

Mutual Information: Given $(X, Y) \sim p_{XY}$,

$$I(X; Y) := \sum_{x,y} p_{X,Y} \log \frac{p_{XY}}{p_X p_Y}$$

Conditional Mutual Information: Given $(X, Y, Z) \sim p_{XYZ}$,

$$I(X; Y|Z) := \sum_z p_Z \sum_{x,y} p_{X,Y|Z} \log \frac{p_{XY|Z}}{p_{X|Z} p_{Y|Z}}$$

Basic notation

Markov chain: Let X, Y, Z be random variables, we say $X \to Y \to Z$ if

$$p(z|x, y) = p(z|y) \text{ whenever } p(y, z) > 0.$$

An equivalent condition for $X \to Y \to Z$ is $I(X; Z|Y) = 0$.

Gaussian distribution: Let X be $\mathbb{R}^n$-valued Gaussian random variable with first and second moments μ, K, then the density function is

$$f_X(x) = \frac{1}{(\sqrt{2\pi})^n \det(K)} e^{-\frac{1}{2}(x-\mu)^T K^{-1}(x-\mu)},$$

and

$$h(X) = \frac{1}{2} \log(2\pi e)^n \det(K)$$

The sub-additive–doubling–rotation proof of Gaussian optimality

Let f be some functional defined on probability distributions.
Using an abuse of notation, let $f(X)$ be a *short-hand* notation for $f(p_X)$.

- (Strict) Sub-additivity: $f(X_1, X_2) \leq f(X_1) + f(X_2)$ with equality *if and only if* $X_1 \perp X_2$;
- Rotation invariant: $f(QX) = f(X)$, where Q is a unitary matrix.
  Running example: $X_+ = \frac{1}{\sqrt{2}}(X_1 + X_2), X_- = \frac{1}{\sqrt{2}}(X_1 - X_2)$.
- f is continuous with respect to weak convergence.

Goal: Maximize $f(X)$ subject to $E[XX^T] \preceq K$, where f satisfies above properties.
Suppose p_{X^*} is a maximizer for f with corresponding maximum value v, then take two independent copies of X^*, say X_1 and X_2.
Note that

$$\begin{aligned}
2v &= f(X_1) + f(X_2) \\
&= f(X_1, X_2) \\
&= f(X_+, X_-) \\
&\leq f(X_+) + f(X_-) \leq 2v,
\end{aligned}$$

thus, each inequality must have been an equality in the above derivation. Hence, we have $X_+ \perp X_-$.

The sub-additive proof of Gaussian optimality

The final step of the proof relies on an important observation:

Theorem (Kac-Bernstein)

Let X_1, X_2 be independent random variables such that $X_1 + X_2 \perp X_1 - X_2$, then they must be Gaussian distributed with the same covariance.

Denote the characteristic function of a random variable X by $\phi_X(t) := E[e^{itX}]$. The proof of the theorem is done in the space of characteristic functions.

Proof sketch.

By using the independence conditions $X_1 \perp X_2$ and $X_1 + X_2 \perp X_1 - X_2$, we can write the characteristic function of $(X_1 + X_2, X_1 - X_2)$ in two ways:

$$\phi_{(X_1+X_2, X_1-X_2)}(t_1, t_2) = \phi_{X_1}(t_1 + t_2) \cdot \phi_{X_2}(t_1 - t_2),$$
$$\phi_{(X_1+X_2, X_1-X_2)}(t_1, t_2) = \phi_{X_1}(t_1) \cdot \phi_{X_1}(t_2) \cdot \phi_{X_2}(t_1) \cdot \phi_{X_2}(-t_2).$$

Solving the equation, one gets ϕ_{X_1} and ϕ_{X_2} must in the form

$$\phi_{X_1}(t) = e^{-ct^2 + iat}, \phi_{X_2} = e^{-ct^2 + ibt}.$$

This proves the theorem. □

The sub-additive proof of Gaussian optimality

The above proof technique has its origins in functional analysis in the work of Carlen, Lieb, etc. in the early 90s

It was discovered in the context of information theory (as explained in this talk) for solving the capacity region of a Gaussian broadcast channel with private and common messages by Geng and Nair (2012).

The use of it to prove the Entropy power inequality (as I will describe later) was developed in a paper by Anantharam, Jog, and Nair (2019).

Example

Theorem

Let X be a random variable with second moment σ^2, then $h(X) \leq \frac{1}{2}\log(2\pi e\sigma^2)$, equality holds if and only if X is Gaussian.

Proof.

Let X_1, X_2 be independent copies of the maximizer X^* and let $v = h(X^*)$. Then

$$\begin{aligned}
2v &= h(X_1) + h(X_2) \\
&= h(X_1, X_2) && \text{(subadditivity)} \\
&= h(X_+, X_-) && \text{(rotation invariant)} \\
&= h(X_+) + h(X_-) - I(X_+; X_-) && \text{(subadditivity)} \\
&\leq 2v - I(X_+; X_-). && \text{(optimality)}
\end{aligned}$$

Therefore, $I(X_+; X_-) = 0$, and X_1, X_2 are Gaussians by the Kac-Berstein's theorem. □

The entropy power inequality

[Shannon '48] proposed that the following inequality was true

Entropy Power Inequality (EPI)

Let X and Y be two independent continuous-valued random vectors taking values in $\mathbb{R}^d$, with finite second moment and well-defined differential entropies for $h(X), h(Y)$, and $h(X+Y)$ (base of the logarithm is 2), then

$$2^{\frac{2}{d}h(X+Y)} \geq 2^{\frac{2}{d}h(X)} + 2^{\frac{2}{d}h(Y)}.$$

The equality holds if and only if X and Y are Gaussians with proportional covariance matrices.

The entropy power inequality

[Shannon '48] proposed that the following inequality was true

Entropy Power Inequality (EPI)

Let X and Y be two independent continuous-valued random vectors taking values in $\mathbb{R}^d$, with finite second moment and well-defined differential entropies for $h(X), h(Y)$, and $h(X+Y)$ (base of the logarithm is 2), then

$$2^{\frac{2}{d}h(X+Y)} \geq 2^{\frac{2}{d}h(X)} + 2^{\frac{2}{d}h(Y)}.$$

Theorem (Lieb's Equivalent Form)

Suppose X and Y are independent $\mathbb{R}^d$-valued random variables. For any $\lambda \in [0,1]$, we have

$$\inf_{X,Y:X\perp Y} h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h(X) - (1-\lambda)h(Y) \geq 0,$$

where the equality holds if and only if X and Y are Gaussians with identical covariance matrices.

The entropy power inequality

[Shannon '48] proposed that the following inequality was true

Entropy Power Inequality (EPI)

Let X and Y be two independent continuous-valued random vectors taking values in $\mathbb{R}^d$, with finite second moment and well-defined differential entropies for $h(X), h(Y)$, and $h(X+Y)$ (base of the logarithm is 2), then

$$2^{\frac{2}{d}h(X+Y)} \geq 2^{\frac{2}{d}h(X)} + 2^{\frac{2}{d}h(Y)}.$$

- It was proven by Stam (1958), building on the work of de Bruijn that constructs the link between differential entropy and Fisher information.
- Several proofs of this inequality have subsequently been discovered
  - Almost all of them follow a variational approach and prove an associated Fisher Information inequality

The entropy power inequality

[Shannon '48] proposed that the following inequality was true

Entropy Power Inequality (EPI)

Let X and Y be two independent continuous-valued random vectors taking values in $\mathbb{R}^d$, with finite second moment and well-defined differential entropies for $h(X), h(Y)$, and $h(X+Y)$ (base of the logarithm is 2), then

$$2^{\frac{2}{d}h(X+Y)} \geq 2^{\frac{2}{d}h(X)} + 2^{\frac{2}{d}h(Y)}.$$

- It was proven by Stam (1958), building on the work of de Bruijn, which established the link between differential entropy and Fisher information.
- Several proofs of this inequality have subsequently been discovered
  - Almost all of them follow a variational approach and prove an associated Fisher information inequality
- The sub-additive–doubling–rotation proof introduces a different approach without using any Fisher information inequalities.

Conditional EPI

Instead of showing EPI directly, we prove a more general form of EPI, called the conditional EPI, formulated by the following theorem.

Theorem (Conditional EPI)

Let $X, Y \in \mathbb{R}^d$ be continuous-valued random vectors and U be an arbitrary random variable satisfying $X \to U \to Y$. X, Y have finite second moment and well-defined conditional differential entropies for $h(X|U), h(Y|U)$, and $h(X+Y|U)$ (base of the logarithm is 2), then

$$2^{\frac{2}{d}h(X+Y|U)} \geq 2^{\frac{2}{d}h(X|U)} + 2^{\frac{2}{d}h(Y|U)}.$$

The Conditional EPI reduces to EPI by taking U to be constant.

Similarly, Lieb's equivalent form of conditional EPI says that

$$\inf_{X,Y: X \to U \to Y} h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y|U) - \lambda h(X|U) - (1-\lambda)h(Y|U) \geq 0,$$

Double Markovity

Before we present the proof of the entropy power inequality using the sub-additive–doubling–rotation technique, we need the following result [Ahlswede Köner '74] as a lemma:

Theorem (Double Markovity)

Let A, B, C be real-valued random vectors. Then $A \to B \to C$ and $A \to C \to B$ hold if and only if there exist functions $g, h : \mathbb{R} \to \mathbb{R}$ such that $g(B) = h(C)$ with probability one and $A \perp (B, C)$ conditional on $f(B)$.

In a special case where $f(b, c) > 0$ for all b, c, observe that $f(a|b, c) = f(a|b) = f(a|c)$, then

$$f(a|b, c) = f(a|c) = \int f(a|c)dF(b) = \int f(a|b)dF(b) = f(a).$$

Therefore, A is independent of (B, C).

Proof of EPI using sub-additivity–doubling-rotation

Part 1

Note by Lieb's equivalent form, it suffices to show for all $a, b > 0$,

$$h(X + Y|U) - ah(X|U) - bh(Y|U) \geq 0.$$

However, in order to utilize the double Markovity theorem as well as assist later analysis, we define the following functional:

$$g_{\delta,\epsilon}(X, Y|U) := h(\hat{X} + \hat{Y}|U) - ah(\hat{X}|U) - bh(\hat{Y}|U) + \epsilon I(\hat{X}, \hat{Y}; \hat{X} + Z_1, \hat{Y} + Z_2|U)$$

where Z_1, Z_2 are independent standard Gaussians, $\hat{X} = X + \delta N_1, \hat{Y} = Y + \delta N_2$, N_1, N_2 are also independent standard Gaussian noises.

- Adding Gaussian noise allows us to work on random variables with full support.
- Once we prove that $g_{\delta,\epsilon}$ is optimized by Gaussian, we can take a sequence of functions $g_{\delta,\epsilon}$ with $\epsilon, \delta \downarrow 0$ to argue the desired functional is also optimized by Gaussian.

Proof of EPI using sub-additivity–doubling-rotation

Part 2

Fix $\delta, \epsilon > 0$. Let (X^*, Y^*, U^*) be a minimizer of $g_{\delta,\epsilon}$ and v be the minimum value. Take two copies of (X^*, Y^*), say $(X_1, Y_1), (X_2, Y_2)$, such that $(X_1, Y_1|U^* = u) \perp (X_2, Y_2|U^* = u)$ for all u. Then

$$
\begin{aligned}
2v &= h(\hat{X}_1 + \hat{Y}_1, \hat{X}_2 + \hat{Y}_2|U^*) - ah(\hat{X}_1, \hat{X}_2|U^*) - bh(\hat{Y}_1, \hat{Y}_2|U^*) \\
&\quad + \epsilon I(\hat{X}_1, \hat{X}_2, \hat{Y}_1, \hat{Y}_2; \hat{X}_1 + Z_{11}, \hat{X}_2 + Z_{12}, \hat{Y}_1 + Z_{21}, \hat{Y}_2 + Z_{22}|U^*) \\
&= h(\hat{X}_+ + \hat{Y}_+, \hat{X}_- + \hat{Y}_-|U^*) - ah(\hat{X}_+, \hat{X}_-|U^*) - bh(\hat{Y}_+, \hat{Y}_-|U^*) \\
&\quad + \epsilon I(\hat{X}_+, \hat{X}_-, \hat{Y}_+, \hat{Y}_-; \hat{X}_+ + Z_{1+}, \hat{X}_- + Z_{1-}, \hat{Y}_+ + Z_{2+}, \hat{Y}_- + Z_{2-}|U^*) \\
&= h(\hat{X}_+ + \hat{Y}_+|U^*) - ah(\hat{X}_+|U^*) - bh(\hat{Y}_+|U^*) + \epsilon I(\hat{X}_+, \hat{Y}_+; \hat{X}_+ + Z_{1+}, \hat{Y}_+ + Z_{2+}|U^*) \\
&\quad + h(\hat{X}_- + \hat{Y}_-|U^*, \hat{X}_+, \hat{Y}_+) - ah(\hat{X}_-|U^*, \hat{X}_+, \hat{Y}_+) - bh(\hat{Y}_-|U^*, \hat{X}_+, \hat{Y}_+) \\
&\qquad + \epsilon I(\hat{X}_-, \hat{Y}_-; \hat{X}_- + Z_{1-}, \hat{Y}_- + Z_{2-}|U^*, \hat{X}_+, \hat{Y}_+) \\
&\quad + I(\hat{X}_- + \hat{Y}_-; \hat{X}_+, \hat{Y}_+|U^*, \hat{X}_+ + \hat{Y}_+) \\
&\qquad + \epsilon I(\hat{X}_+, \hat{Y}_+; \hat{X}_- + Z_{1-}, \hat{Y}_- + Z_{2-}|U^*, \hat{X}_+ + Z_{1+}, \hat{Y}_+ + Z_{2+}) \\
&\qquad + \epsilon I(\hat{X}_-, \hat{Y}_-; \hat{X}_+ + Z_{1+}, Y_+ + Z_{2+}|U^*, \hat{X}_+, \hat{Y}_+, \hat{X}_- + Z_{1-}, \hat{Y}_- + Z_{2-}) \\
&\geq v + v
\end{aligned}
$$

Proof of EPI using sub-additivity–doubling-rotation

Part 3

Again, it implies that the gap of inequalities must be 0, in particular,

$$I(\hat{X}_+, \hat{Y}_+; \hat{X}_- + Z_{1-}, \hat{Y}_- + Z_{2-} | U^*, \hat{X}_+ + Z_{1+}, \hat{Y}_+ + Z_{2+}) = 0.$$

This implies, conditioned on U^*, the following two Markov chains hold:

- $I(\hat{X}_+, \hat{Y}_+; \hat{X}_- + Z_{1-}, \hat{Y}_- + Z_{2-} | U^*, \hat{X}_+ + Z_{1+}, \hat{Y}_+ + Z_{2+}) = 0.$
- $I(\hat{X}_+ + Z_{1+}, \hat{Y}_+ + Z_{2+}; \hat{X}_- + Z_{1-}, \hat{Y}_- + Z_{2-} | U^*, \hat{X}_+, \hat{Y}_+) = 0.$

Therefore, we obtain, from Double Markovity, that

$$I(\hat{X}_+, \hat{Y}_+, \hat{X}_+ + Z_{1+}, \hat{Y}_+ + Z_{2+}; \hat{X}_- + Z_{1-}, \hat{Y}_- + Z_{2-} | U^*) = 0$$

From here, we obtain that

$$I(\hat{X}_+, \hat{Y}_+; \hat{X}_- + Z_{1-}, \hat{Y}_- + Z_{2-} | U^*) = 0,$$

and further (using the positivity of the characteristic function of the Gaussian) that

$$I(\hat{X}_+, \hat{Y}_+; \hat{X}_-, \hat{Y}_- | U^*) = 0,$$

as desired; establishing the Gaussianity of the optimizers.

Basic notations for quantum system

von Neumann Entropy: Given state ρ_X,

$$S(X) := -\operatorname{tr}(\rho \log \rho).$$

Weyl displacement operators: Let $R = (R_1, \ldots, R_{2n}) = (\hat{x}_1, \hat{p}_1, \ldots, \hat{x}_n, \hat{p}_n)$. The Weyl displacement operator is given by

$$D(\xi) = e^{i\xi JR} \text{ where } J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^{\oplus n}.$$

Characteristic function: Given state ρ_X,

$$\chi_X(\xi) = \operatorname{tr}(\rho D(\xi)).$$

Basic notations for quantum system

First and second moments: The first moment is a vector $d = (d_1, \ldots, d_{2n}) \in \mathbb{R}^{2n}$ given by

$$d_k = \mathrm{tr}(\rho R_k).$$

The second moment is a $2n \times 2n$ positive matrix $\gamma = [\gamma_{kl}]$ given by

$$\gamma_{kl} = \mathrm{tr}(\rho\{R_k - d_k, R_l - d_l\}).$$

Gaussian state: A state is Gaussian if its characteristic function is in the form

$$\chi_X(\xi) = \exp\left(i\xi^T J d - \frac{1}{4}\xi^T J^T \gamma J \xi\right),$$

with first and second moment d, γ.

The entropy power inequalities for quantum systems

Let ρ_X, ρ_Y be n-mode quantum states, [König Smith, '12] showed that

Theorem (Quantum EPI)

Let the map $\mathcal{E}_\lambda$ be associated with a beamsplitter of transmissivity λ, then

$$e^{S(\mathcal{E}_{1/2}(\rho_X \otimes \rho_Y))/n} \geq \frac{1}{2} e^{S(\rho_X)/n} + \frac{1}{2} e^{S(\rho_Y)/n}, \qquad \textit{(Quantum EPI)}$$

$$S(\mathcal{E}_\lambda(\rho_X \otimes \rho_Y)) \geq \lambda S(\rho_X) + (1-\lambda) S(\rho_Y). \qquad \textit{(Lieb's equivalent form)}$$

- The proof of this theorem was based on Stam's proof for the classical case.
- The authors built analogues of random variable addition, additive Gaussian noise, Fisher information, etc., and proved de Bruijn identity and Stam's inequality in quantum settings.

Quantum addition using beamsplitter

The authors introduced the following analogue for the addition $\sqrt{\lambda}X + \sqrt{1-\lambda}Y$ of classical random variables:

Let ρ_X and ρ_Y be two n-mode states, $0 < \lambda < 1$.

$$(\rho_X, \rho_Y) \mapsto \mathcal{E}_\lambda(\rho_X \otimes \rho_Y) := \mathrm{tr}_Y\left(U_\lambda(\rho_X \otimes \rho_Y)U_\lambda^\dagger\right),$$

where U_λ is a Gaussian unitary whose action is given by

$$S = \begin{pmatrix} \sqrt{\lambda}I_n & \sqrt{1-\lambda}I_n \\ \sqrt{1-\lambda}I_n & -\sqrt{\lambda}I_n \end{pmatrix} \otimes I_2.$$

This construction leads to the following observations:

- Covariance of $\mathcal{E}_\lambda(\rho_X \otimes \rho_Y)$ is given by $\lambda\gamma_X + (1-\lambda)\gamma_Y$
-

$$\chi_{U_\lambda(\rho_X\otimes\rho_Y)U_\lambda^\dagger}(\xi_1, \xi_2) = \chi_X(\sqrt{\lambda}\xi_1 + \sqrt{1-\lambda}\xi_2) \cdot \chi_Y(\sqrt{1-\lambda}\xi_1 - \sqrt{\lambda}\xi_2),$$
$$\chi_{\mathcal{E}_\lambda(\rho_X\otimes\rho_Y)} = \chi_X(\sqrt{\lambda}\xi_1) \cdot \chi_Y(\sqrt{1-\lambda}\xi_1).$$

Searching for analogue of the sub-additive-doubling-rotation proof

Inspired by the quantum addition, we may extend Kac-Bernstein's theorem to the quantum setting.

Theorem

Define $\rho_+ = \mathrm{tr}_Y(U_\lambda(\rho_X \otimes \rho_Y)U_\lambda^\dagger)$ *and* $\rho_- = \mathrm{tr}_X(U_\lambda(\rho_X \otimes \rho_Y)U_\lambda^\dagger)$*. Suppose that* $\rho_+ \otimes \rho_- = U_\lambda(\rho_X \otimes \rho_Y)U_\lambda^\dagger$*, then* ρ_X, ρ_Y *are Gaussian states.*

Note that

$$\chi_+(\xi) = \chi_X(\sqrt{\lambda}\xi) \cdot \chi_Y(\sqrt{1-\lambda}\xi), \qquad \chi_- = \chi_X(\sqrt{1-\lambda}\xi) \cdot \chi_Y(-\sqrt{\lambda}\xi).$$

Then, similar to the classical proof, we can write the characteristic function in two ways:

$$\begin{aligned}\chi_{U_\lambda(\rho_X \otimes \rho_Y)U_\lambda^\dagger}(\xi_1, \xi_2) &= \chi_X(\sqrt{\lambda}\xi_1 + \sqrt{1-\lambda}\xi_2) \cdot \chi_Y(\sqrt{1-\lambda}\xi_1 - \sqrt{\lambda}\xi_2) \\ &= \chi_X(\sqrt{\lambda}\xi_1) \cdot \chi_Y(\sqrt{1-\lambda}\xi_1) \cdot \chi_X(\sqrt{1-\lambda}\xi_2) \cdot \chi_Y(-\sqrt{\lambda}\xi_2).\end{aligned}$$

The remaining proof is identical to the classical proof.

Example: Gaussian maximizes entropy

Recall the following result:

Theorem

Let ρ_G be a Gaussian state with covariance γ. Then $S(\rho) \leq S(\rho_G)$ for all ρ with covariance γ.

Let ρ_1, ρ_2 be copies of maximizer of $S(\rho)$ subject to fixed second moment γ. Let $v = S(\rho_1) = S(\rho_2)$. Then

$$\begin{aligned} 2v &= S(\rho_1) + S(\rho_2) \\ &= S(\rho_1 \otimes \rho_2) \\ &= S(U_\lambda(\rho_1 \otimes \rho_2)U_\lambda^\dagger) \\ &= S(\rho_+) + S(\rho_-) - I(\rho_+;\rho_-) \leq 2v - I(\rho_+;\rho_-), \end{aligned}$$

implying that $U_\lambda(\rho_1 \otimes \rho_2)U_\lambda^\dagger = \rho_+ \otimes \rho_-$. Applying the Kac-Bernstein's theorem, ρ_1 and ρ_2 are Gaussian.

Goal of the research

Develop tool ideas of the other components of the proof so that one can employ them in quantum settings.

In classical setting: Assume

$$X, Y, Z, U$$

have full support (for simplicity).
Then, the following double Markovity result holds,

$$I(X;Z|U,Y) = 0, \quad I(Y;Z|U,X) = 0$$

implies that

$$I(X,Y;Z|U) = 0.$$

Question: Is there a quantum equivalent to this result?

Summary and future work

The sub-additive–doubling–rotation technique provides proofs for Gaussian optimality in various problem settings, sometimes going beyond the traditional perturbation-based techniques.
It relies on the sub-additivity of the underlying functional.

The extension from the classical setting to the quantum setting becomes challenging when the strict sub-additive functional becomes more complicated, involving conditional entropy terms and Markov structures.

My current research aims to develop extensions of the tools in the quantum case.

Summary and future work

The sub-additive–doubling–rotation technique provides proofs for Gaussian optimality in various problem settings, sometimes going beyond the traditional perturbation-based techniques.
It relies on the sub-additivity of the underlying functional.

The extension from the classical setting to the quantum setting becomes challenging when the strict sub-additive functional becomes more complicated, involving conditional entropy terms and Markov structures.

My current research aims to develop extensions of the tools in the quantum case.

Thank you for your time!