Von Neumann Entropy and Mutual Information

Quantum Foundations · Module 7 — Quantum Information

Node 7.4

Position in Knowledge Graph

System: Quantum Foundations **Structural Domain:** Quantum Information **Node:** 7.4

The **von Neumann entropy** $S(\rho) = -\operatorname{Tr}(\rho \log \rho)$ is the quantum analogue of Shannon entropy. It measures the information content (or equivalently, the mixedness, or the uncertainty) of a quantum state and is the foundational information-theoretic quantity for quantum systems. From it are derived quantum relative entropy, quantum mutual information, and the entropies that control channel capacities, entanglement measures, and thermodynamic bounds.

The quantum theory differs from the classical theory in two important ways. First, entropies can exhibit strict subadditivity with surprising strength — joint entropy can be *less* than the entropy of a single subsystem, a phenomenon impossible in classical Shannon theory. Second, quantum mutual information can exceed twice the corresponding classical mutual information, reflecting the extra correlations available through entanglement. Together these give von Neumann entropy a character distinct from its classical cousin, even though it reduces to Shannon entropy on diagonal states.

Formal Definition

Von Neumann entropy

For a density operator ρ on a finite-dimensional Hilbert space $\mathcal{H}$,

$$S(\rho) = -\operatorname{Tr}(\rho \log \rho) = -\sum_i \lambda_i \log \lambda_i,$$

where $\{\lambda_i\}$ are the eigenvalues of ρ (with $0 \log 0 \equiv 0$ by continuity). The logarithm is base 2 for entropies measured in bits, base e for nats.

Basic properties:

- **Non-negativity.** $S(\rho) \geq 0$, with equality iff ρ is a pure state.
- **Upper bound.** $S(\rho) \leq \log d$, where $d = \dim \mathcal{H}$, with equality iff $\rho = I/d$.
- **Unitary invariance.** $S(U\rho U^\dagger) = S(\rho)$ for any unitary U.
- **Concavity.** $S(\lambda\rho_1 + (1-\lambda)\rho_2) \geq \lambda S(\rho_1) + (1-\lambda)S(\rho_2)$.
- **Additivity over products.** $S(\rho_A \otimes \rho_B) = S(\rho_A) + S(\rho_B)$.

Relative entropy

The **quantum relative entropy** (Umegaki) is

$$S(\rho\|\sigma) = \operatorname{Tr}(\rho \log \rho) - \operatorname{Tr}(\rho \log \sigma),$$

defined when $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$ and $+\infty$ otherwise. Klein's inequality: $S(\rho\|\sigma) \geq 0$, with equality iff $\rho = \sigma$. Relative entropy is the fundamental "distance" in quantum information, underlying hypothesis testing, resource theories, and thermodynamics.

Joint, conditional, and mutual information

For a bipartite state ρ_{AB} with marginals $\rho_A = \operatorname{Tr}_B \rho_{AB}$ and $\rho_B = \operatorname{Tr}_A \rho_{AB}$:

- **Joint entropy:** $S(AB) = S(\rho_{AB})$.
- **Conditional entropy:** $S(A|B) = S(AB) - S(B)$.
- **Mutual information:** $I(A : B) = S(A) + S(B) - S(AB) = S(\rho_{AB}\|\rho_A \otimes \rho_B)$.

A striking quantum fact: $S(A|B)$ can be *negative*. For example, if ρ_{AB} is a pure entangled state, then $S(AB) = 0$ but $S(B) > 0$, so $S(A|B) = -S(B) < 0$. Negative conditional entropy has no classical analogue and is closely related to the ability to perform quantum protocols like teleportation.

Intuition

Von Neumann entropy measures "how mixed" a state is, equivalently "how much classical uncertainty is encoded in ρ." A pure state has $S = 0$ — it is a definite quantum state, and although measurements give probabilistic outcomes, the state itself is maximally known. A maximally mixed state $\rho = I/d$ has $S = \log d$ — it is the state of maximum ignorance, corresponding to the uniform classical distribution over a basis of size d.

What makes von Neumann entropy non-classical is the behavior of bipartite entropy under entanglement. For a pure entangled state $|\psi\rangle_{AB}$, the whole system has $S(AB) = 0$, but each subsystem has $S(A) = S(B) > 0$. This is structurally impossible in classical information theory: a classical joint distribution with zero entropy is a delta function, which has marginals with zero entropy. Quantum joint purity does not imply marginal purity because the whole system is more than the sum of its parts — the entanglement is "in between" the subsystems, not localized in either.

Mutual information is the natural measure of how much two subsystems are correlated. Classically, $I(A : B) \leq \min(H(A), H(B))$. Quantumly, $I(A : B) \leq 2\min(S(A), S(B))$ — a factor of 2 more, achieved by maximally entangled states. This extra factor is one of the structural signatures of quantum correlations and is tightly connected to why quantum channels can transmit more classical information per qubit (when assisted by entanglement) than classical channels can transmit per bit.

The practical slogan: entropy measures the number of (qu)bits needed on average to describe the state classically, given some reference basis. A pure state needs zero bits (you already know it). The maximally mixed state needs $\log d$ bits (you know nothing). An arbitrary state lies somewhere in between, and its position is exactly the entropy.

Structural Role

- **Foundational quantity.** Von Neumann entropy is *the* fundamental information-theoretic quantity for quantum systems, analogous to Shannon entropy in the classical case. Every capacity theorem, compression theorem, and information inequality in quantum information theory is formulated in terms of $S(\rho)$ or related quantities.
- **Entanglement entropy.** For a bipartite pure state, $S(\rho_A) = S(\rho_B)$ is the *entanglement entropy* — the unique measure of entanglement for pure states (Node 4.5). Entanglement entropy is the operational measure used in condensed matter and holography (area laws, entanglement entropy of ground states).
- **Channel capacities.** Classical and quantum capacities of quantum channels are expressed in terms of entropic quantities: the Holevo capacity $\chi = S(\bar{\rho}) - \sum_i p_i S(\rho_i)$ (Node 7.5) and the coherent information $I_c = S(B) - S(AB)$ are the fundamental capacity formulas.

- **Thermodynamic bridge.** Thermal states minimize free energy $F = E - TS$ at fixed temperature, where S is the von Neumann entropy. This gives the quantum-statistical-mechanics link between information and heat, developed in Module 9.
- **Data processing.** Mutual information cannot increase under local channels: $I(A:B)_{(\mathrm{id}\otimes\mathcal{E})(\rho)} \le I(A:B)_\rho$. This is the data-processing inequality, the quantum generalization of the classical statement that post-processing cannot create information. It is equivalent to monotonicity of quantum relative entropy under CPTP maps, which in turn is one of the deepest theorems in quantum information (Lindblad 1975, Uhlmann).
- **Strong subadditivity.** Lieb–Ruskai (1973): $S(ABC) + S(B) \le S(AB) + S(BC)$ for any tripartite state. A nontrivial and highly structural inequality; equivalent to monotonicity of relative entropy and to the data-processing inequality for quantum mutual information.

Mathematical Development

Klein's inequality

For density operators ρ, σ,

$$S(\rho\|\sigma) = \mathrm{Tr}(\rho(\log\rho - \log\sigma)) \ge 0,$$

with equality iff $\rho = \sigma$. Proof: use the operator inequality $x\log x - x\log y \ge x - y$ (from concavity of log) on eigenvalues. Klein's inequality is the quantum analogue of non-negativity of classical KL divergence and underlies most entropic inequalities.

Subadditivity

For any bipartite state ρ_{AB},

$$S(AB) \le S(A) + S(B),$$

with equality iff $\rho_{AB} = \rho_A \otimes \rho_B$. This follows from Klein's inequality applied to ρ_{AB} and $\rho_A \otimes \rho_B$:

$$0 \le S(\rho_{AB}\|\rho_A\otimes\rho_B) = -S(AB) + S(A) + S(B).$$

Subadditivity implies mutual information is non-negative: $I(A:B) = S(A) + S(B) - S(AB) \ge 0$.

Strong subadditivity (Lieb–Ruskai)

For any tripartite state ρ_{ABC},

$$S(ABC) + S(B) \le S(AB) + S(BC).$$

This is the quantum version of the classical inequality $H(ABC) + H(B) \le H(AB) + H(BC)$, but it is vastly harder to prove in the quantum case. It was conjectured by Lanford and Robinson and proved by Lieb and Ruskai in 1973 using a key operator convexity theorem of Lieb. Strong subadditivity implies monotonicity of mutual information under local channels, Markov chain conditions for quantum states, and many other central results.

Data processing inequality

For any CPTP map $\mathcal{E}$ on system B and any bipartite state ρ_{AB},

$$I(A:B)_{(\mathrm{id}_A \otimes \mathcal{E})(\rho_{AB})} \leq I(A:B)_{\rho_{AB}}.$$

Equivalently: quantum relative entropy is monotone under CPTP maps,

$$S(\mathcal{E}(\rho) \| \mathcal{E}(\sigma)) \leq S(\rho \| \sigma).$$

This is the fundamental information-theoretic bound on how much correlation between systems can survive processing. It implies the second law of thermodynamics (when applied to thermal reference states) and constrains all channel capacities.

Bounds on mutual information

For any bipartite state ρ_{AB} on $\mathcal{H}_A \otimes \mathcal{H}_B$,

$$0 \leq I(A:B) \leq 2\min(\log d_A, \log d_B).$$

The upper bound is achieved by maximally entangled pure states, where $S(A) = S(B) = \log d$ and $S(AB) = 0$, giving $I(A:B) = 2\log d$. Classical mutual information is bounded by $\min(H(A), H(B))$ (a factor of 2 smaller), reflecting the non-classical correlations available through entanglement.

Joint concavity and convexity

- **Concavity of entropy:** $S(\sum_i p_i \rho_i) \geq \sum_i p_i S(\rho_i)$. Mixing states increases entropy on average.
- **Joint convexity of relative entropy:** $S(\sum_i p_i \rho_i \| \sum_i p_i \sigma_i) \leq \sum_i p_i S(\rho_i \| \sigma_i)$.
- **Araki–Lieb triangle inequality:** $|S(A) - S(B)| \leq S(AB)$. This is the lower bound complementing subadditivity.

Worked Example

Entropies of basic states

- **Pure state:** $\rho = |\psi\rangle\langle\psi|$ has eigenvalues $(1, 0, 0, \ldots)$, so $S(\rho) = 0$.
- **Maximally mixed qubit:** $\rho = I/2$ has eigenvalues $(1/2, 1/2)$, so $S(\rho) = 1$ bit.
- **Maximally mixed d-dim state:** $S(I/d) = \log d$. For a qubit register of n qubits, $S(I/2^n) = n$ bits.
- **Mixed state $p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|$:** $S = h(p) = -p\log p - (1-p)\log(1-p)$, the binary entropy function.
- **Thermal state at temperature T:** $\rho_{\mathrm{th}} = e^{-\beta H}/Z$ with $\beta = 1/k_B T$. Then $S(\rho_{\mathrm{th}}) = \beta\langle H\rangle + \log Z$, which is the standard thermodynamic entropy formula.

Bell state

The Bell state $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ has $S(\rho_{AB}) = 0$ (pure). The reduced density matrix of either qubit is $\rho_A = I/2$, with $S(A) = S(B) = 1$ bit. Therefore:

- $I(A:B) = 1 + 1 - 0 = 2$ bits — the maximum possible mutual information for a qubit pair.
- $S(A|B) = S(AB) - S(B) = 0 - 1 = -1$ bit — negative conditional entropy, a purely quantum phenomenon.

For comparison, the maximum classical mutual information for two bits is 1 bit (perfectly correlated). The Bell state doubles this.

Mixed Bell state

Consider $\rho_{AB} = p|\Phi^+\rangle\langle\Phi^+| + (1-p)I/4$ (a Werner-like state). One finds:

- $\rho_A = \rho_B = I/2$, so $S(A) = S(B) = 1$ bit independent of p.
- ρ_{AB} has eigenvalues $(p + (1-p)/4, (1-p)/4, (1-p)/4, (1-p)/4)$, giving

$$S(AB) = -\left(p + \frac{1-p}{4}\right)\log\left(p + \frac{1-p}{4}\right) - 3 \cdot \frac{1-p}{4}\log\frac{1-p}{4}.$$

- $I(A:B) = 2 - S(AB)$. For $p = 1$, $S(AB) = 0$, $I = 2$. For $p = 0$, $S(AB) = 2$, $I = 0$. The transition is smooth and nonlinear.

Data processing on a Bell state

Apply a dephasing channel $\mathcal{E}(\rho) = (\rho + Z\rho Z)/2$ to qubit B of a Bell state. The resulting state is

$$\rho'_{AB} = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|),$$

a classical-quantum correlated state with $S(AB) = 1$ (not zero anymore) and $I(A:B) = 1 + 1 - 1 = 1$ bit. The dephasing channel has halved the mutual information, consistent with the data processing inequality: we lost exactly the "coherent" part of the correlation, retaining only the "classical" part.

Cross-Links

- **Module 5.1 (Density operators):** The object on which entropy acts. Pure states have zero entropy; mixing always increases it.
- **Module 4.5 (Entanglement measures):** Entanglement entropy $S(\rho_A)$ is the canonical entanglement measure for pure bipartite states, and is the special case of von Neumann entropy applied to reduced density matrices.
- **Module 7.3 (Quantum channels):** Data-processing inequality expresses the monotonicity of entropic quantities under CPTP maps.
- **Module 7.5 (Holevo bound):** Bounds classical information extractable from quantum states in terms of entropic quantities.
- **Module 9.4 (Quantum thermodynamics):** Von Neumann entropy is the thermodynamic entropy of quantum states; equilibrium and non-equilibrium thermodynamics are built on its behavior.
- **Module 10.5 (Quantum resource theories):** Entropies quantify resource content (entanglement, coherence, athermality).
- **Statistics:** Shannon entropy is the classical restriction. Quantum relative entropy generalizes KL divergence; quantum mutual information generalizes classical mutual information.

Structural Compression

Invariant: $S(\rho) = -\operatorname{Tr}(\rho\log\rho)$ is the unique (up to normalization) entropy functional satisfying concavity, additivity on products, and continuity. It is zero for pure states, maximized at $\log d$ for the maximally mixed state, invariant under unitaries, non-decreasing under unital channels, and monotone-decreasing for relative entropy under any CPTP map.

Mechanism: Eigendecomposition of the density matrix: entropy is Shannon entropy of the eigenvalue spectrum. All quantum-specific phenomena (subadditivity with negative conditional entropy, doubling of mutual information, entanglement entropy of pure states) arise from the non-commuting structure of the density operators involved.

Cross-System Echo: Shannon entropy in classical information theory is the diagonal restriction; KL divergence is the classical analogue of relative entropy; classical mutual information sits inside quantum mutual information. The factor-of-2 gap in mutual information upper bounds quantifies the specifically quantum contribution from entanglement.

Exercises

1. Compute $S(\rho)$ for a qubit state $\rho = (I + \vec{r} \cdot \vec{\sigma})/2$ with Bloch vector of magnitude r. Show that $S = h((1+r)/2)$.
2. Show that $S(AB) \leq S(A) + S(B)$ using Klein's inequality applied to ρ_{AB} and $\rho_A \otimes \rho_B$.
3. Verify the Araki–Lieb inequality $|S(A) - S(B)| \leq S(AB)$ for a Bell state.
4. Compute the quantum mutual information of a Werner state $\rho = p|\Phi^+\rangle\langle\Phi^+| + (1-p)I/4$ as a function of p. At what value of p does it become zero?
5. State the strong subadditivity inequality and show how it implies monotonicity of quantum mutual information under local channels.
6. For a thermal state $\rho_{\text{th}} = e^{-\beta H}/Z$, derive the formula $S = \beta\langle H\rangle + \log Z$. How does this connect to the thermodynamic identity $S = (E - F)/T$?
7. Argue, structurally, why negative conditional entropy is possible quantumly but not classically. What does $S(A|B) < 0$ mean operationally?

Quantum Foundations · Module 7 — Quantum Information · Node 7.4
Concepts: von-neumann-entropy, shannon-entropy, mutual-information, density-matrix
Prerequisites: 5.1, 7.3
Enables: 7.5, 9.4

hbar.university — own.your.cognition