Properties of the Von Neumann entropy

1. **Purity.** A pure state $\rho = |\varphi\rangle\langle\varphi|$ has $S(\rho) = 0$.

2. **Invariance.** The entropy is unchanged by a unitary change of basis

   $$S(\mathbf{U}\rho\mathbf{U}^\dagger) = S(\rho),$$

   because the entropy depends only on the eigenvalues of the density matrix.

3. **Maximum.** If ρ has D non-vanishing eigenvalues, then

   $$S(\rho) \leq \log D,$$

   with equality when all nonzero eigenvalues are equal (maximum randomness).

4. **Concavity.** For $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$,

   $$S(\sum_i \lambda_i \rho_i) \geq \sum_i \lambda_i S(\rho_i).$$

   That is, the Von Neumann entropy is larger if we know less about how the state was prepared.

5. **Entropy of measurement.** If we measure $A = \sum_y a_y |a_y\rangle\langle a_y|$ in ρ, then outcome a_y occurs with probability $p(a_y) = \langle a_y|\rho|a_y\rangle$. The Shannon entropy for the ensemble of measurements outcomes $Y = \{a_y, p(a_y)\}$ satisfies

   $$H(Y) \geq S(\rho),$$

   with equality when A and ρ commute. By measuring a non-commuting observable the results would be less predictable.

6. **Entropy of preparation.** For $\rho = \sum_x p_x |\varphi_x\rangle\langle\varphi_x|$ and $X = \{|\varphi_x\rangle, p_x\}$,

$$H(X) \geq S(\rho),$$

with equality when the $|\varphi_x\rangle$'s are mutually orthogonal. When the different states are not orthogonal then information received would be less then when different characters are fully distinguishable.

7. **Subadditivity.** For a bipartite system AB in the state ρ_{AB},

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

with equality when $\rho_{AB} = \rho_A \otimes \rho_B$. Entropy is additive for independent subsystems, but for correlated subsystems total entropy is less than the sum of the entropy of the subsystems. Similarly $H(X,Y) \leq H(X) + H(Y)$.

8. **Strong subadditivity.** For any state ρ_{ABC} of a tripartite system,

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

When B is one dimensional this property reduces to subadditivity. This property may be viewed as the fact that the sum of the entropies of two systems' union and intersection does not exceed the sum of the entropies of the two systems.

9. **Triangle inequality (Araki-Lieb inequality).** For a bipartite system

$$S(\rho_{AB}) \geq |S(\rho_A) - S(\rho_B)|,$$

in contrast to Shannon entropy

$$H(X,Y) \geq H(X), H(Y)$$

or

$$H(X|Y), H(Y|X) \geq 0.$$

There exists more information in the whole classical system than any part of it. But for quantum systems and Von Neumann entropy, we could have $S(\rho_A) = S(\rho_B)$ and $S(\rho_{AB}) = 0$ in the case of a bipartite pure state. That is, for the whole system the state is completely known, yet considering only one of the subsystems the measurement result could be complete random. This is the consequence of quantum entanglement.

If we could somehow define a conditional Von Neumann entropy, then negative entropies should result, leading to insights into quantum entanglement and measurement.

Quantum Data Compression

Consider a message composed of n letters, each chosen at random from the ensemble of pure states $\{|\varphi_x\rangle, p_x\}$, where the states may not be orthogonal. Then each letter is described by the density matrix

$$\rho = \sum_x p_x |\varphi_x\rangle\langle\varphi_x|,$$

and the entire message by

$$\rho^n = \rho \otimes \rho \otimes \cdots \otimes \rho.$$

The message can be compressed to a Hilbert space of $nS(\rho)$ dimensions, without decreasing the fidelity of the message.

So the Von Neumann entropy can be seen as the number of qubits of quantum information carried per letter by the message. Analogous to the classical case, when $\rho = \frac{1}{2}\mathbf{1}$, the (completely random) message could not be compressed.

Schumacher encoding

Similar to classical compression in which we only consider typical sequences, typical subspaces are considered in quantum messages. That is, we can represent a given quantum message in the typical subspace of its Hilbert space, and throw away the orthogonal component.

Consider a quantum message $\rho^n = \rho \otimes \rho \otimes \cdots \otimes \rho$, where $\rho = \sum_x p_x |\varphi_x\rangle\langle\varphi_x|$. In the orthonormal basis that diagonalizes ρ, the message can be seen as a classical source in which each letter is chosen from ρ's eigenstates, with probability given by the eigenvalues. Then the typical sequence of ρ eigenstates appearing in the message ρ^n forms a typical subspace. That is, we need only consider the typical eigenstates of ρ^n. Specifically, the eigenstates with eigenvalue λ satisfying

$$2^{-n(S(\rho)-\delta)} \geq \lambda \geq 2^{-n(S(\rho)+\delta)}.$$

Each eigenstate of ρ^n is a sequence of eigenstates of ρ, with eigenvalues given by the product of the corresponding eigenvalues of ρ.

There are $2^{nS(\rho)}$ typical sequences, each with probability (eigenvalue) λ satisfying (for a specified δ)

$$2^{-n(S(\rho)-\delta)} \geq \lambda \geq 2^{-n(S(\rho)+\delta)}.$$

For any δ and $\epsilon > 0$ sufficiently large, the sum of the above typical eigenvalues satisfies $\mathrm{tr}\,(\rho^n \mathbf{E}) > 1 - \epsilon$, (where $\mathbf{E}$ is the projection onto the typical subspace spanned by the typical eigenstates of ρ^n) and the dimension of the typical subspace Λ satisfies

$$2^{nS(\rho)+\delta} \geq \mathrm{dim}\Lambda \geq 2^{n(S(\rho)-\delta)}.$$

The coding strategy is to send messages in the typical subspace faithfully. First the sender performs a unitary transformation that rotates the typical eigenstates of the message to the form $\mathbf{U}|\Psi_{typ}\rangle = |\Psi_{comp}\rangle|0_{rest}\rangle$, where $|\Psi_{comp}\rangle$ is a state of $n(S(\rho)+\delta)$ qubits, and $|0_{rest}\rangle$ represent $|0\rangle$'s for all remaining qubits. The $|\Psi_{comp}\rangle$ is send, and the receiver appends $|0_{rest}\rangle$ and apply $\mathbf{U}^{-1}$ to recover the original message.