Universal Bounds for the Holevo Quantity, Coherent Information, and the Jensen-Shannon Divergence

Wojciech Roga,1 Mark Fannes,2 and Karol Zyczkowski1,3

1Smoluchowski Institute of Physics, Jagiellonian University, ul. Reymonta 4, 30-059 Kraków, Poland
2Instituut voor Theoretische Fysica, Universiteit Leuven, B-3001 Leuven, Belgium
3Centrum Fizyki Teoretycznej, Polska Akademia Nauk, Al. Lotników 32/44, 02-668 Warszawa, Poland

(Rceived 7 May 2010; published 23 July 2010)

The mutual information between the sender of a classical message encoded in quantum carriers and a receiver is fundamentally limited by the Holevo quantity. Using strong subadditivity of entropy, we prove that the Holevo quantity is not larger than an exchange entropy. This implies an upper bound for coherent information. Moreover, restricting our attention to classical information, we bound the transmission distance between probability distributions by their entropic distance, which is a concave function of their Hellinger distance.

DOI: 10.1103/PhysRevLett.105.040505

PACS numbers: 03.67.Hk, 02.10.Ud

The goal of quantum information is to efficiently apply quantum resources for handling information. One of the fundamental results on transmitting classical information by quantum means is the Holevo bound: It is an upper bound for the mutual information. Mutual information quantifies the dependence of decoded on typical encoded messages and is therefore a figure of merit of the transmission process.

Assume that a classical source emits messages written in an alphabet $X$ that are sent to a receiver by using a quantum device. Each letter $a_i$ is encoded in a quantum state, i.e., in a density matrix $\rho_i$. We obtain in this way a quantum ensemble $\{q_i, \rho_i\}$, where $q_i$ is the probability that the source emits $a_i$. The receiver performs a general positive operator valued measurement (POVM) and reads classical data written in an alphabet $Y$. The mutual information is

$$H(X:Y) = H(X) + H(Y) - H(X, Y),$$

where $H(X, Y)$ is the Shannon entropy of the joint probability distribution of sending in $a_i$ and reading out $b_j$.

Holevo’s theorem [1,2] provides a fundamental upper bound on the mutual information in terms of the ensemble $\{q_i, \rho_i\}$: The bound depends not only on the statistics of the source but also on the encoding of the source alphabet in the density matrices $\{\rho_i\}$:

$$H(X:Y) \leq S\left(\sum_i q_i \rho_i\right) - \sum_i q_i S(\rho_i) =: \chi(\{q_i, \rho_i\}).$$

(1)

Here, $S(\rho)$ is the von Neumann entropy of $\rho$. The right-hand side of (1) is the Holevo $\chi$ quantity of the ensemble $\{q_i, \rho_i\}$. Note that $\chi$ is independent of the measurement, phrased differently: Quantum distinguishability between initial states cannot be increased by measuring them [3].

It is well known [3] that the Shannon entropy of the probability vector $\{q_i\}$ is an upper bound for $\chi$:

$$\chi(\{q_i, \rho_i\}) \leq H(\{q_i\}).$$

In this work we derive a finer bound for the Holevo quantity and explore its consequences for classical and quantum information theory. Our main result is Theorem 1 or the equivalent Theorem 3. The first is stated in terms of transmission of information, while the second is formulated as a bound on the concavity of von Neumann entropy, quantified by a quantum analogue of the Jensen-Shannon divergence. Our result yields a new universal entropy inequality, valid in both quantum and classical settings. We provide at the end of the paper some applications to informational distances.

A general quantum operation $\Phi$ on a system $A$ is an affine, completely positive transformation of the states on $A$. The Schrödinger version of Stinespring’s theorem characterizes such maps in terms of an ancillary system $B$ and an isometry $V: \mathcal{H}_A \rightarrow \mathcal{H}_{AB}$:

$$\Phi: \rho \mapsto \rho' = Tr_B V \rho V^\dagger.$$

(2)

Note that the density matrix $V \rho V^\dagger$ has, up to multiplicities of zero, the same eigenvalues as $\rho$. An equivalent dynamical picture is

$$V \rho V^\dagger = U(\rho \otimes |\phi\rangle \langle \phi|) U^\dagger,$$

where $|\phi\rangle \in \mathcal{H}_B$ is a pure state of the environment and $U$ a global unitary on $\mathcal{H}_{AB}$. Fixing an orthonormal basis $\{|i\rangle\}$ in $\mathcal{H}_B$, we have

$$V |\psi\rangle = \sum_i K_i |\psi\rangle \otimes |i\rangle \quad \text{and}$$

$$V \rho V^\dagger = \sum_{ij} K_{ij} \rho K_{ij}^\dagger \otimes |i\rangle \langle j|,$$

(3)

which yields the Kraus form

$$\Phi(\rho) = \sum_i K_i \rho K_i^\dagger.$$

(4)

As $V$ is an isometry, we have a resolution of the identity:
The density matrix $\sigma = \tilde{\Phi}(\rho)$ is the state of the environment after the interaction and is called a correlation matrix. If the initial state $\rho$ is pure, then $S(\sigma)$ is the entropy exchanged between the system and the environment. Therefore, $S(\sigma)$ is called the exchange entropy.

Because of the identity resolution, the set of Kraus operators describes a POVM. Such a selective measurement transforms an initial state $\rho$ into one of the output states

$$\rho'_i := \frac{1}{\text{Tr}\rho K_i^\dagger K_i} \rho K_i^\dagger$$

with probability $q_i = \text{Tr}\rho K_i^\dagger K_i$. For this setup, shown in Fig. 1(a), one defines the Holey quanty $\chi(q_i, \rho'_i)$; see (1).

Our main result is contained in the following bounds.

**Theorem 1.**—Consider a state $\rho$, a quantum operation $\Phi$, and the image of $\rho$ under $\Phi$: $\rho' = \Phi(\rho) = \sum_i \gamma_i K_i^\dagger K_i$. The complementary channel produces a correlation matrix $\sigma = \tilde{\Phi}(\rho')$ as in (6). Define the probability vector with entries $\gamma_i := \text{Tr}\rho K_i^\dagger K_i$ and quantum states $\rho_i' := K_i^\dagger \rho K_i$ so that $\rho' = \sum_i \gamma_i \rho_i'$. Then (a) the Holey quantity is bounded by the exchange entropy:

$$\chi(q_i, \rho'_i) \leq S(\sigma) \leq H(q_i)$$

and (b) the average entropy is bounded by the entropy of the initial state:

$$\sum_i q_i S(\rho_i') \leq S(\rho).$$

**Proof.**—(a) The rightmost inequality $S(\sigma) \leq H(q_i)$ is a direct consequence of the majorization theorem, which says that for any state $S(\sigma) \leq S[\text{diag}(\sigma)]$; see, e.g., [7]. To prove the left inequality, consider the isometry $W$ and the three-partite quantum state $\omega_{123}$:

$$W|\psi\rangle := \sum_i K_i |\psi\rangle \otimes |i\rangle \otimes |i\rangle \omega_{123} := W\rho W^\dagger$$

$$= \sum_{ij} K_i \rho K_i^\dagger |i\rangle \langle j| \otimes |i\rangle \langle j|.$$  

It is convenient to introduce the notation $A_{ij} := K_i \rho K_i^\dagger$, so that $q_i = \text{Tr}A_{ii}$ and $\rho'_i = A_{ii}/q_i$. One checks that

$$S(\omega_{123}) = S(\sigma), \quad S(\omega_{123}) = S \left( \sum_i q_i \rho_i' \right),$$

and

$$\sum_i q_i S(\rho_i') = \sum_i \text{Tr}A_{ii} \ln \text{Tr}A_{ii} - \sum_i \text{Tr}A_{ii} \ln A_{ii}$$

$$= S(\omega_{13}) - S(\omega_2).$$

Substituting these expressions in the strong subadditivity inequality for quantum entropy written in the form $[3,8]$

$$S(\omega_1) + S(\omega_2) \leq S(\omega_{13}) + S(\omega_{23})$$

yields the first inequality in (8). (b) Since the transformation $W$ in (10) is an isometry, the three-partite state $\omega_{123}$ has the same spectrum as $\rho$ up to multiplicities of zero. Hence, $\omega_{123}$ and $\rho$ have the same entropy. The equality (11) and the Araki-Lieb inequality $S(\omega_{13}) - S(\omega_2) \leq S(\omega_{123})$ then yield (9).

The bounds in Theorem 1 are universal; they hold for any quantum operation $\Phi$ and any initial state $\rho$. We analyze here some of their consequences. Inequality (8) is saturated for orthogonal Kraus operators, which form a projective von Neumann measurement: $K_i^\dagger K_j = \delta_{ij} K_i$. In this case all output states $\rho_i'$ are pure, so $\sum_i q_i S(\rho_i') = 0$. The state $\rho'$ is a mixture of pure and orthogonal states with probabilities $q_i = \text{Tr}\rho K_i^\dagger K_i$. The correlation matrix $\sigma$ is then diagonal and $S(\rho') = S(\sigma)$.

Note also that inequality (9) differs from $S(\sum_i q_i S(\rho_i')) \leq S(\rho')$, which is implied by the concavity of entropy. For a bistochastic map $\Phi$, the entropy does not decrease, so in this case we may write $S(\sum_i q_i S(\rho_i')) \leq S(\rho')$.

The Jamiołkowski isomorphism represents a quantum map $\Phi$ acting on an $N$-level system by a density matrix $\sigma_{\Phi}$ on an extended space: $\sigma_{\Phi} = \Phi \otimes \text{id}(|\phi^+\rangle \langle \phi^+|)$, where $|\phi^+\rangle = (1/\sqrt{N}) \sum_i |i\rangle \otimes |i\rangle$ is a maximally entangled state. The degree of nonunitarity of an operation $\Phi$ can be quantified by its entropy $[9]$, defined as the entropy of the corresponding state $S(\Phi) := S(\sigma_{\Phi})$. If the initial state $\rho$ is maximally mixed, then the exchange entropy $S(\sigma)$ is equal to the entropy of the condition $[10]$. Theorem 1 yields now a simple interpretation of the entropy of a map: It is an upper bound for the Holey quantity (1) for a transformation of the maximally mixed state $\rho_s = 1/N$. Furthermore, the entropy of a map is an upper bound for the Holey quantity associated with ensembles of Kraus maps acting on the mixed state $\rho_s$. 

![FIG. 1 (color online). (a) A dynamical picture: An initial quantum state $\rho$ is sent by a map $\Phi$ into $\rho'$, while the complementary map $\Phi$ sends it into $\sigma$. A Kraus operator $K_i$ maps $\rho$ into $\rho_i'$ with probability $q_i$, so that $\rho'$ is the barycenter of the ensemble $\{q_i, \rho_i\}$; (b) A static picture: The ensemble $\{q_i, \rho_i\}$ determines the barycenter $\bar{\rho}$.](image-url)
where the maximum is taken over sets of Kraus operators that realize the same quantum operation: \( \Phi(\rho) = \sum K_i \rho K_i^\dagger \).

By applying the Araki-Lieb triangle inequality to (3), Lindblad [11] proved

\[
S(\rho') - S(\sigma) \leq S(\rho) \leq S(\rho') + S(\sigma). \tag{14}
\]

The difference of entropies \( I^\rho_{\text{coh}} := S[\Phi(\rho)] - S(\sigma) \) is the coherent information [3], so Eq. (14) implies that \( I^\rho_{\text{coh}} \leq S(\rho) \). We are now in position to refine this bound.

Proposition 2.—Consider a state \( \rho \) and quantum operations \( \Phi_1 \) and \( \Phi_2 \), with \( \Phi_1(\rho) = \sum K_i \rho K_i^\dagger \) and a quantum ensemble \( \{q_i, \rho_i'\} \), where \( q_i := \text{Tr} K_i \rho K_i^\dagger \) and \( \rho_i' := K_i \rho K_i^\dagger / q_i \). Then (a) the coherent information for the quantum operation \( \Phi_1 \) is bounded by

\[
I^\rho_{\text{coh}}(\Phi_1) \leq \sum_i q_i S(\rho_i') \leq S(\rho). \tag{15}
\]

(b) The coherent information for the concatenation \( \Phi_2 \circ \Phi_1 \) is bounded by

\[
I^\rho_{\text{coh}}(\Phi_2 \circ \Phi_1) \leq \sum_i p_i S[\Phi_2(\rho_i')]. \tag{16}
\]

\[\text{Proof.}—\text{Relation (15) is a direct consequence of Proposition 2, as these inequalities are obtained by combining (8) and (9). Another proof of the first relation (15) was recently given by Holevo and Shirokov [12]. To show (16) we consider the four-partite state}\]

\[
\omega_{1234} := \sum_{ijkl} L_i K_j \rho K_j^\dagger L_i^\dagger \otimes |i\rangle\langle j| \otimes |k\rangle\langle \ell| \otimes |k\rangle\langle \ell|, \tag{17}
\]

where \( \Phi_2(\rho) = \sum L_i \rho L_i^\dagger \). Consider the strong subadditivity relation

\[
S(\omega_2') + S(\omega_{234}) \leq S(\omega_{234}) + S(\omega_{13}').
\]

The inequality (16) which we want to prove can be rewritten as

\[
S(\omega_1') + S(\omega_2') \leq S(\omega_{234}) + S(\omega_{13}').
\]

Therefore, it is sufficient to prove that \( S(\omega_2') \leq S(\omega_{24}) \). As the matrix \( \omega_2' \) is diagonal and consists of the traces of the blocks of the block diagonal matrix \( \omega_{24} \), \( \omega_{24} \) is more mixed than \( \omega_2' \) and has therefore larger entropy.

Let us now consider the static case in Fig. 1(b): a quantum ensemble \( \{q_i, \rho_i\} \) which determines the average state \( \bar{\rho} := \sum q_i \rho_i \). For an ensemble of classical measures \( \{q_i, \mu_i\} \), the generalized Jensen-Shannon divergence (JSD) quantifies the concavity of the Shannon entropy:

\[
\text{JSD} \left( \{q_i, \mu_i\} \right) := H\left( \sum q_i \mu_i \right) - \sum q_i H(\mu_i). \tag{18}
\]

The Holevo quantity \( \chi \), which is its quantum analogue, is also called the quantum Jensen-Shannon divergence [13–15].

It is intuitively clear that Theorem 1 can be reformulated in terms of concavity of entropy without referring to the map \( \Phi \) in Theorem 1 or its Kraus form. For an initial state \( \rho \) and a Kraus operator \( K_i \), we consider the polar decomposition \( K_i \rho^{1/2} = X_i U_i \). Here \( X_i \) is a Hermitian matrix and \( U_i \) is unitary. Note that \( X_i^2 = K_i \rho K_i^\dagger \), and this is equal to \( q_i \rho_i' \). Therefore, \( K_i \sqrt{\rho} = \sqrt{q_i \rho_i'} U_i \), so the elements of the correlation matrix (6) read:

\[
\sigma_{ij} = \text{Tr} K_i \rho K_j^\dagger = \sqrt{q_i q_j} \text{Tr} \sqrt{\rho} U_i U_j^\dagger \sqrt{\rho}.
\]

In this way we arrive at Theorem 3.

Theorem 3.—Consider a quantum ensemble \( \{q_i, \rho_i\} \) and a collection of unitaries matrices \( \{U_i\} \) and construct the correlation matrix \( \sigma \) as in (19). Then \( \chi(\{q_i, \rho_i\}) \leq S(\sigma) \).

As a simple application we consider the case of an ensemble with 2 elements. To obtain the lowest upper bound for \( \chi \), we need to minimize the entropy of the correlation matrix (19) over unitaries \( U_1, U_2 \). This is equivalent to finding the POVM, which minimizes \( S(\sigma) \) among all measurements which result in the same ensemble of output states.

Proposition 4.—Consider two density matrices \( \rho_1 \) and \( \rho_2 \) occurring with probabilities \( \lambda, 1 - \lambda \). The smallest entropy of the correlation matrix (19) over unitaries \( U_1 \) and \( U_2 \) is achieved for the matrix

\[
\sigma_{ij} = \left( \frac{\lambda}{\sqrt{\lambda(1 - \lambda)\sqrt{F}}} \right) \left( \sqrt{\lambda(1 - \lambda)\sqrt{F}} ight), \tag{20}
\]

where \( \sqrt{F} \) is the root fidelity [16]: \( \sqrt{F} = \text{Tr} \sqrt{\rho_1^{1/2} \rho_2 \rho_1^{1/2}} \).

\[\text{Proof.}—\text{Given } \lambda, \rho_1, \text{ and } \rho_2, \text{ the entropy } S(\sigma) \text{ is minimal, if the absolute value of the off-diagonal element } \sigma_{ij} \text{ is maximal. As } |\text{Tr}AB| \leq ||A||\text{Tr}|B|, \text{ we have the upper bound}\]

\[
|\text{Tr} \rho_2^{1/2} \rho_1^{1/2} U_1 U_2^\dagger| \leq |\text{Tr}|\rho_2^{1/2} \rho_1^{1/2}| = \sqrt{F}.
\]

Moreover, the inequality is saturated by choosing \( U_1 U_2^\dagger \) the adjoint of the unitary of the polar decomposition of \( \rho_2^{1/2} \rho_1^{1/2} \).

Let us now set \( \lambda = 1 - \mu = \frac{1}{2} \). In this case \( \chi = \frac{1}{2} S(\rho_1) + S(\rho_2) \). Let us now set \( \lambda = 1 - \mu = \frac{1}{2} \). In this case \( \chi = \frac{1}{2} S(\rho_1) + S(\rho_2) \).

\[
\chi \leq H_2\left[ 1 - \sqrt{F(\rho_1, \rho_2)} \right]. \tag{21}
\]

where \( H_2(x) := -x \ln x - (1 - x) \ln(1 - x) \) is the Shannon entropy of a probability vector of size 2.
The right-hand side of (21) can be used to characterize closeness between quantum states. Although the entropy $H_2$ does not obey the triangle inequality, its square root does. Such an entropic distance was advocated by Lamberti, Portesi, and Sparacino [17] as a natural metric in the space of quantum states:

$$D_E(p_1, p_2) := \sqrt{H_2^2[1 - F(p_1, p_2)]},$$

(22)

To show that $D_E$ is a distance, one may use the Bures distance $D_B(p_1, p_2) = \sqrt{2 - 2F(p_1, p_2)}$. Both quantities are functions of fidelity, and one can therefore write the entropic distance as a function of Bures’s distance $D_E(D_B) = \sqrt{H_2^2(D_B^2/4)}$. As this function is concave, $D_E$ satisfies the axioms of a distance.

Turning now to the classical case, i.e., diagonal density matrices, the root fidelity reduces to the Bhattacharyya coefficient $B(P, Q) = \sum_i \sqrt{p_i q_i}$, while the Bures distance is equivalent to the Hellinger distance [7] $D_B(P, Q) = \sqrt{\sum_i (\sqrt{p_i} - \sqrt{q_i})^2}$. The entropic distance $D_E$ between two classical states is then a concave function of their Hellinger distance: $D_E(P, Q) = \sqrt{H_2^2[D_B^2(P, Q)/4]}$. Although JSD does not satisfy the triangle inequality, its square root, the transmission distance [15], does [18]. Hence, inequality (21) implies

$$D_T(P, Q) := \sqrt{\text{JSD}(P, Q)} \leq D_E(P, Q),$$

(23)

which is illustrated in Fig. 2 and used in Ref. [17]. The square root of $\chi$ is conjectured to be a distance in the space of all quantum states [15,19], which is proved for pure quantum states.

We showed that the Holevo quantity $\chi$ is bounded by the exchange entropy $S(\sigma)$ and analyzed several consequences of this universal inequality for classical and quantum information theory. In particular, we bounded the transmission distance between classical distributions by their entropic distance.

It is a pleasure to thank R. Alicki and M., P., and R. Horodecki for fruitful discussions. Helpful correspondence and suggestions by N. Datta, P. Herremoës, A. S. Holevo, P. Lamberti, M. B. Ruskai, and F. Topsøe are gratefully acknowledged. This work was supported by Grant No. DFG-SFB/38/2007 of the Polish Ministry of Science and Higher Education and by the Belgian Interuniversity Attraction Poles Programme P6/02.