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

Wojciech Roga,[1] Mark Fannes,[2] and Karol Życzkowski[1,3]
[1]Smoluchowski Institute of Physics, Jagiellonian University, ul. Reymonta 4, 30-059 Kraków, Poland
[2]Instituut voor Theoretische Fysica, Universiteit Leuven, B-3001 Leuven, Belgium
[3]Centrum Fizyki Teoretycznej, Polska Akademia Nauk, Al. Lotników 32/44, 02-668 Warszawa, Poland
(Received 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 ρ_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 $\{\rho_i\}$:

$$H(X{:}Y) \leq S\Big(\sum_i q_i\rho_i\Big) - \sum_i q_i S(\rho_i) =: \chi(\{q_i, \rho_i\}). \quad (1)$$

Here, $S(\rho)$ is the von Neumann entropy of ρ. The right-hand side of (1) is the Holevo χ quantity of the ensemble $\{q_i, \rho_i\}$. Note that χ 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 χ:

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

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 Φ 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 $V{:}\mathcal{H}_A \to \mathcal{H}_{AB}$:

$$\Phi{:}\rho \mapsto \rho' = \mathrm{Tr}_B V\rho V^\dagger. \quad (2)$$

Note that the density matrix $V\rho V^\dagger$ has, up to multiplicities of zero, the same eigenvalues as ρ. 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

$$V|\psi\rangle = \sum_i K_i|\psi\rangle \otimes |i\rangle \quad \text{and}$$
$$V\rho V^\dagger = \sum_{ij} K_i\rho K_j^\dagger \otimes |i\rangle\langle j|, \quad (3)$$

which yields the Kraus form

$$\Phi(\rho) = \sum_i K_i\rho K_i^\dagger. \quad (4)$$

As V is an isometry, we have a resolution of the identity:

$$\sum_i K_i^\dagger K_i = \mathbb{1}. \tag{5}$$

Complementary to Φ is a quantum map $\tilde{\Phi}: A \to B$ defined by tracing over the principal system [4–6]:

$$\tilde{\Phi}: \rho \mapsto \sigma = \mathrm{Tr}_A V\rho V^\dagger = \sum_{ij} (\mathrm{Tr}\rho K_j^\dagger K_i)|i\rangle\langle j|. \tag{6}$$

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 ρ 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 ρ into one of the output states

$$\rho_i' := \frac{1}{\mathrm{Tr}\rho K_i^\dagger K_i} K_i \rho K_i^\dagger \tag{7}$$

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

Our main result is contained in the following bounds.

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

$$\chi(\{q_i, \rho_i'\}) \le S(\sigma) \le H(\{q_i\}) \tag{8}$$

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

$$\sum_i q_i S(\rho_i') \le S(\rho). \tag{9}$$

Proof.—(a) The rightmost inequality $S(\sigma) \le H(\{q_i\})$ is a direct consequence of the majorization theorem, which says that for any state $S(\sigma) \le S[\mathrm{diag}(\sigma)]$; see, e.g., [7]. To

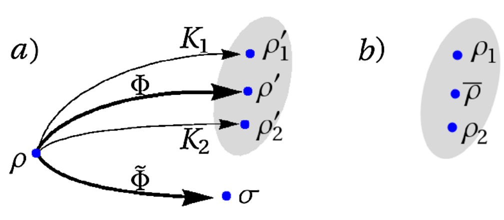

FIG. 1 (color online). (a) A dynamical picture: An initial quantum state ρ is sent by a map Φ into ρ', while the complementary map $\tilde{\Phi}$ sends it into σ. A Kraus operator K_i maps ρ into ρ_i' with probability q_i so that ρ' is the barycenter of the ensemble $\{q_i, \rho_i'\}$. (b) A static picture: The ensemble $\{q_i, \rho_i\}$ determines the barycenter $\bar{\rho}$.

prove the left inequality, consider the isometry W and the three-partite quantum state ω_{123}:

$$W|\psi\rangle := \sum_i K_i|\psi\rangle \otimes |i\rangle \otimes |i\rangle \omega_{123} := W\rho W^\dagger$$
$$= \sum_{ij} K_i \rho K_j^\dagger \otimes |i\rangle\langle j| \otimes |i\rangle\langle j|. \tag{10}$$

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

$$S(\omega_{23}) = S(\sigma), \qquad S(\omega_1) = S\Big(\sum_i q_i \rho_i'\Big),$$

and $$\sum_i q_i S(\rho_i') = \sum_i \mathrm{Tr}A_{ii} \ln \mathrm{Tr}A_{ii} - \sum_i \mathrm{Tr}A_{ii} \ln A_{ii}$$
$$= S(\omega_{13}) - S(\omega_2). \tag{11}$$

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

$$S(\omega_1) + S(\omega_2) \le S(\omega_{13}) + S(\omega_{23}) \tag{12}$$

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

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

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

The Jamiołkowski isomorphism represents a quantum map Φ acting on an N-level system by a density matrix σ_Φ on an extended space: $\sigma_\Phi = \Phi \otimes \mathrm{id}(|\phi^+\rangle\langle\phi^+|)$, 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 Φ can be quantified by its entropy [9], defined as the entropy of the corresponding state $S(\Phi) := S(\sigma_\Phi)$. If the initial state ρ is maximally mixed, then the exchange entropy $S(\sigma)$ is equal to the entropy of the map [10]. Theorem 1 yields now a simple interpretation of the entropy of a map: It is an upper bound for the Holevo quantity (1) for a transformation of the maximally mixed state $\rho_* = \mathbb{1}/N$. Furthermore, the entropy of a map is an upper bound for the Holevo quantity associated with ensembles of Kraus maps acting on the mixed state ρ_*:

$$\max_{\{K_i\}} \chi\left(\left\{\mathrm{Tr}K_iK_i^\dagger/N, \frac{K_iK_i^\dagger}{\mathrm{Tr}K_iK_i^\dagger}\right\}\right) \leq S(\Phi), \tag{13}$$

where the maximum is taken over sets of Kraus operators that realize the same quantum operation: $\Phi(\rho) = \sum_i K_i\rho K_i^\dagger$.

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 $I_{\mathrm{coh}}^{\rho} := S[\Phi(\rho)] - S(\sigma)$ is the coherent information [3], so Eq. (14) implies that $I_{\mathrm{coh}}^{\rho} \leq S(\rho)$. We are now in position to refine this bound.

Proposition 2.—Consider a state ρ and quantum operations Φ_1 and Φ_2, with $\Phi_1(\rho) = \sum_i K_i\rho K_i^\dagger$ and a quantum ensemble $\{q_i, \rho_i'\}$, where $q_i := \mathrm{Tr}\rho K_i^\dagger K_i$ and $\rho_i' := K_i\rho K_i^\dagger/q_i$. Then (a) the coherent information for the quantum operation Φ_1 is bounded by

$$I_{\mathrm{coh}}^{\rho}(\Phi_1) \leq \sum_i q_i S(\rho_i') \leq S(\rho). \tag{15}$$

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

$$I_{\mathrm{coh}}^{\rho}(\Phi_2 \circ \Phi_1) \leq \sum_i p_i S[\Phi_2(\rho_i')]. \tag{16}$$

Proof.—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

$$\omega_{1234}' := \sum_{ijk\ell} L_iK_k\rho K_\ell^\dagger L_j^\dagger \otimes |i\rangle\langle j| \otimes |k\rangle\langle \ell| \otimes |k\rangle\langle \ell|, \tag{17}$$

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

$$S(\omega_1') + S(\omega_{24}') \leq S(\omega_{234}') + S(\omega_{13}').$$

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

$$S(\omega_1') + S(\omega_4') \leq S(\omega_{234}') + S(\omega_{13}').$$

Therefore, it is sufficient to prove that $S(\omega_4') \leq S(\omega_{24}')$. As the matrix ω_4' is diagonal and consists of the traces of the blocks of the block diagonal matrix ω_{24}', ω_{24}' is more mixed than ω_4' 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 $\bar{\rho} := \sum_i q_i\rho_i$. For an ensemble of classical measures $\{q_i, \mu_i\}$, the generalized Jensen-Shannon divergence (JSD) quantifies the concavity of the Shannon entropy:

$$\mathrm{JSD}(\{q_i, \mu_i\}) := H\left(\sum_i q_i\mu_i\right) - \sum_i q_i H(\mu_i). \tag{18}$$

The Holevo quantity χ, 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 Φ in Theorem 1 or its Kraus form. For an initial state ρ and a Kraus operator K_i, we consider the polar decomposition $K_i\rho^{1/2} = X_iU_i$. 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:

$$\sigma_{ij} = \mathrm{Tr}K_i\rho K_j^\dagger = \sqrt{q_iq_j}\mathrm{Tr}\sqrt{\rho_i}U_iU_j^\dagger\sqrt{\rho_j}. \tag{19}$$

In this way we arrive at Theorem 3.

Theorem 3.—Consider a quantum ensemble $\{q_i, \rho_i\}$ and a collection of unitary matrices $\{U_i\}$ and construct the correlation matrix σ 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 χ, we need to minimize the entropy of the correlation matrix (19) over sets of 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 ρ_1 and ρ_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

$$\sigma_\lambda = \begin{pmatrix} \lambda & \sqrt{\lambda(1-\lambda)}\sqrt{F} \\ \sqrt{\lambda(1-\lambda)}\sqrt{F} & 1-\lambda \end{pmatrix}, \tag{20}$$

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

Proof.—Given λ, ρ_1, and ρ_2, the entropy $S(\sigma)$ is minimal, if the absolute value of the off-diagonal element σ_{12} is maximal. As $|\mathrm{Tr}AB| \leq \|A\|\mathrm{Tr}|B|$, we have the upper bound

$$|\mathrm{Tr}\rho_2^{1/2}\rho_1^{1/2}U_1U_2^\dagger| \leq \mathrm{Tr}|\rho_2^{1/2}\rho_1^{1/2}| = \sqrt{F}.$$

Moreover, the inequality is saturated by choosing for $U_1U_2^\dagger$ the adjoint of the unitary of the polar decomposition of $\rho_2^{1/2}\rho_1^{1/2}$.

Let us now set $\lambda = 1-\lambda = \frac{1}{2}$. In this case $\chi = \frac{1}{2}[S(\rho_1\|\bar{\rho}) + S(\rho_2\|\bar{\rho})]$, where $\bar{\rho} = \frac{1}{2}(\rho_1 + \rho_2)$ and $S(\rho_1\|\rho_2) := \mathrm{Tr}\rho_1(\ln\rho_1 - \ln\rho_2)$ is the quantum relative entropy [3]. The spectrum of the correlation matrix (20) is $(\mu, 1-\mu)$ with $\mu = \frac{1}{2}(1-\sqrt{F})$. Hence we get an explicit formula for the exchange entropy $S(\sigma)$ and the universal bound

$$\chi \leq H_2\{\tfrac{1}{2}[1 - \sqrt{F(\rho_1, \rho_2)}]\}, \tag{21}$$

where $H_2(x) := -x\ln x - (1-x)\ln(1-x)$ is the Shannon entropy of a probability vector of size 2.

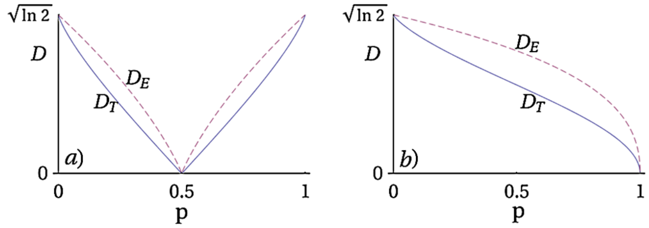

FIG. 2 (color online). The transmission distance D_T (solid line) is bounded from above by the entropic distance D_E (dashed line) (a) for pairs of classical states $P = (p, 1-p)$ and $Q = (1-p, p)$ and (b) for pairs of classical states $P = (p, 1-p)$ and $Q = (1, 0)$.

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:

$$D_E(\rho_1, \rho_2) := \sqrt{H_2\{\tfrac{1}{2}[1 - \sqrt{F(\rho_1, \rho_2)}]\}}. \tag{22}$$

To show that D_E is a distance, one may use the Bures distance $D_B(\rho_1, \rho_2) = \sqrt{2 - 2\sqrt{F(\rho_1, \rho_2)}}$. Both quantities are functions of fidelity, and one can therefore write the entropic distance as a function of Bures's distance $D_E(D_B) = \sqrt{H_2(D_B^2/4)}$. 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] $D_H(P, Q) = \sqrt{\sum_i(\sqrt{p_i} - \sqrt{q_i})^2}$. The entropic distance D_E between two classical states is then a concave function of their Hellinger distance: $D_E(P, Q) = \sqrt{H_2[D_H^2(P, Q)/4]}$. Although JSD does not satisfy the triangle inequality, its square root, the transmission distance [15], does [18]. Hence, inequality (21) implies

$$D_T(P, Q) := \sqrt{\mathrm{JSD}(P, Q)} \le D_E(P, Q), \tag{23}$$

which is illustrated in Fig. 2 and used in Ref. [17]. The square root of χ 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 χ 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.

[1] A. S. Holevo, Probl. Inf. Transm. **9**, 177 (1973).
[2] F. Benatti, J. Math. Phys. (N.Y.) **37**, 5244 (1996).
[3] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England, 2000).
[4] R. Alicki and M. Fannes, *Quantum Dynamical Systems* (Oxford University Press, Oxford, 2001).
[5] A. S. Holevo, Probab. Theory Appl. **51**, 133 (2005).
[6] C. King, K. Matsumoto, M. Nathanson, and M. B. Ruskai, Markov Processes Relat. Fields **13**, 391 (2007).
[7] I. Bengtsson and K. Życzkowski, *Geometry of Quantum States* (Cambridge University Press, Cambridge, England, 2006).
[8] M. B. Ruskai, J. Math. Phys. (N.Y.) **43**, 4358 (2002).
[9] K. Życzkowski and I. Bengtsson, Open Syst. Inf. Dyn. **11**, 3 (2004).
[10] W. Roga, M. Fannes, and K. Życzkowski, J. Phys. A **41**, 035305 (2008).
[11] G. Lindblad, in *Quantum Aspects of Optical Communication*, edited by C. Bendjaballah *et al.*, Lect. Notes Phys. Vol. 378 (Springer-Verlag, Berlin, 1991), pp. 79–80.
[12] A. S. Holevo and M. E. Shirokov, arXiv:1004.2495v1.
[13] B. Fuglede and F. Topsøe, in *Proceedings of the 2004 IEEE International Symposium on Information Theory* (IEEE, New York, 2004), p. 31.
[14] A. P. Majtey, P. W. Lamberti, and D. P. Prato, Phys. Rev. A **72**, 052310 (2005).
[15] J. Briët and P. Herremoës, Phys. Rev. A **79**, 052311 (2009).
[16] R. Jozsa, J. Mod. Opt. **41**, 2315 (1994).
[17] P. W. Lamberti, M. Portesi, and J. Sparacino, Int. J. Quantum. Inform. **7**, 1009 (2009).
[18] D. M. Endres and J. E. Schindelin, IEEE Trans. Inf. Theory **49**, 1858 (2003).
[19] P. W. Lamberti, A. P. Majtey, A. Borras, M. Casas, and A. Plastino, Phys. Rev. A **77**, 052311 (2008).