A New Quantum Data Processing Inequality

Salman Beigi

School of Mathematics, Institute for Research in Fundamental Sciences (IPM), P.O. Box 19395-5746, Tehran, Iran

(Dated: March 8, 2023)

Quantum data processing inequality bounds the set of bipartite states that can be generated by two far apart parties under local operations; having access to a bipartite state as a resource, two parties cannot locally transform it to another bipartite state with a mutual information greater than that of the resource state. Nevertheless, due to the additivity of quantum mutual information under tensor product, the data processing inequality gives no bound when the parties are provided with an arbitrary number of copies of the resource state. In this paper we introduce a measure of correlation on bipartite quantum states, called maximal correlation, that is not additive and gives the same value when computed for multiple copies. Then, by proving a data processing inequality for this measure, we find a bound on the set of states that can be generated under local operations even when an arbitrary number of copies of the resource state is available.

I. INTRODUCTION

Let $\rho_{AB}$ be a bipartite quantum state on registers $A$ and $B$, and assume that an arbitrary number of copies of $\rho_{AB}$ are shared between two parties Alice and Bob. The goal of Alice and Bob is to generate some bipartite state $\sigma_{EF}$ under local operations but without communication. That is for some $n$, they want to apply local super-operators $\Phi_{A^n \to E}$ and $\Psi_{B^n \to F}$ such that

$$\Phi \otimes \Psi(\rho_{AB}^n) = \sigma_{EF}.$$  

Typical examples of this problem are entanglement distillation and common randomness distillation under local operations, in which case $\sigma_{EF}$ is an ebit or one bit of shared randomness.

To answer this question one cannot look for such local operators by brute-force search since we assume $n$, the number of copies of the resource state $\rho_{AB}$ is arbitrarily large. On the other hand to obtain some necessary conditions on the existence of $\Phi$ and $\Psi$ one may compare the strength of correlations of $\rho_{AB}^n$ and $\sigma_{EF}$. If $\sigma_{EF}$ is more correlated than $\rho_{AB}^n$, then such operations do not exists since local transformations do not generate correlation. Nevertheless again since $n$ can be arbitrarily large, the standard measures of correlation provide us with no bound. For instance the data processing inequality of mutual information states that if $\rho_{AB}^n$ can be locally transformed to $\sigma_{EF}$ then

$$n I(A;B)_{\rho} = I(A^n;B^n)_{\rho^n} \geq I(E;F)_{\sigma},$$  

where $I(\cdot,\cdot)$ denotes the quantum mutual information. This inequality is loose for sufficiently large $n$ and gives us no bound if $I(A;B)_{\rho} \neq 0$.

In the classical case (where $\rho_{AB}$ and $\sigma_{EF}$ are bipartite random variables) there is a measure of correlation called Hirschfeld-Gebelein-Rényi maximal correlation or simply the maximal correlation $[1-4]$. Maximal correlation has two main properties that are useful for the problem of local state transformation described above. First, it is not additive on independent copies of a bipartite distribution, and indeed gives the same number when computed on independent copies. Second, it satisfies a data processing inequality. Using these two properties maximal correlation gives a non-trivial bound on our problem in the classical case.

The main contribution of this paper is to generalize maximal correlation to the quantum case.
II. A NEW MEASURE OF CORRELATION

Let $\mathcal{H}_A$ be the Hilbert space corresponding to a quantum register $A$, which in this paper is assumed to be finite dimensional. Denote the space of linear operators acting on $\mathcal{H}_A$ by $L(\mathcal{H}_A)$. Similarly define $L(\mathcal{H}_B)$ and equip these two spaces with the Hilbert-Schmidt inner product, i.e., $\langle M, N \rangle = \text{tr}(M^\dagger N)$. This inner product induces a norm on the space of linear operators which we denote by $\| \cdot \|_2$.

For a bipartite quantum state $\rho_{AB}$ we define its maximal correlation by

$$\mu(\rho_{AB}) = \max |\text{tr}(\rho_{AB}X_A \otimes Y_B^\dagger)|,$$

where $\rho_A$ and $\rho_B$ are the reduced density matrices on subsystems $A$ and $B$ respectively, and $X_A \in L(\mathcal{H}_A)$ and $Y_B \in L(\mathcal{H}_B)$. Indeed, $\mu(\rho_{AB})$ is the maximum of the (absolute value of the) expectation of the tensor product of two local operators that have zero expectation and variance 1.

Maximal correlation in the classical case, where $\rho_{AB}$ is a joint distribution, is first introduced by Hirschfeld [1] and Gebelein [2] and then studied by Rényi [3, 4]. Witsenhausen in [5] proved that maximal correlation of several independent copies of a joint distribution equals to that of a single copy. This parameter has recently been revisited by several authors; see e.g. [6–9].

To study properties of $\mu(\rho_{AB})$, let us define

$$\tilde{\rho}_{AB} = (I_A \otimes \rho_B^{-1/2})\rho_{AB}(\rho_A^{-1/2} \otimes I_B),$$

where inverses of $\rho_A$ and $\rho_B$ are defined on their supports. Note that $\tilde{\rho}_{AB}$ is not even hermitian, so is not a density matrix.

**Theorem 1** $\mu(\rho_{AB})$ is equal to the second Schmidt coefficient of $\tilde{\rho}_{AB}$ as a vector in the bipartite Hilbert space $L(\mathcal{H}_A) \otimes L(\mathcal{H}_B)$.

**Proof:** Let $R_A = \rho_A^{1/2}X_A$ and $S_B = Y_B^\dagger \rho_B^{1/2}$. With this change of variables $\mu(\rho_{AB})$ is equivalently equal to

$$\mu(\rho_{AB}) = \max |\text{tr}(\tilde{\rho}_{AB}R_A \otimes S_B)|,$$

$$\langle \rho_A^{1/2}, R_A \rangle = \langle \rho_B^{1/2}, S_B \rangle = 0,$$

$$\|R_A\|_2 = \|S_B\|_2 = 1.$$  

Let

$$\tilde{\rho}_{AB} = \sum_i \lambda_i M_i \otimes N_i,$$

be the Schmidt decomposition of $\tilde{\rho}_{AB} \in L(\mathcal{H}_A) \otimes L(\mathcal{H}_B)$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ are the Schmidt coefficients and $\{M_i\}$ and $\{N_i\}$ are orthonormal bases for $L(\mathcal{H}_A)$ and $L(\mathcal{H}_B)$, respectively. Note that

$$\lambda_1 = \max_{\|V\|_2 = \|W\|_2 = 1} \text{tr}(\tilde{\rho}_{AB}V_A \otimes W_B),$$

be the Schmidt decomposition of $\tilde{\rho}_{AB} \in L(\mathcal{H}_A) \otimes L(\mathcal{H}_B)$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ are the Schmidt coefficients and $\{M_i\}$ and $\{N_i\}$ are orthonormal bases for $L(\mathcal{H}_A)$ and $L(\mathcal{H}_B)$, respectively. Note that
and using Cauchy-Schwarz inequality we have

\[
\lambda_1 = \text{tr} \left( M_1^\dagger \otimes N_1^\dagger \tilde{\rho}_{AB} \right)
\]

\[
= \text{tr} \left[ \left( \rho_A^{-1/2} M_1^\dagger \right) \otimes \left( N_1^\dagger \rho_B^{-1/2} \right) \rho_{AB} \right]
\]

\[
= \text{tr} \left[ \left( \rho_A^{-1/2} M_1^\dagger \otimes I_B \right) \left( I_A \otimes N_1^\dagger \rho_B^{-1/2} \right) \rho_{AB} \right]
\]

\[
\leq \left[ \text{tr} \left( \rho_A^{-1/2} M_1^\dagger M_1 \rho_A^{-1/2} \otimes I_B \right) \right]^{1/2} \cdot \left[ \text{tr} \left( \rho_A^{-1/2} N_1^\dagger N_1 \rho_B^{-1/2} \right) \right]^{1/2}
\]

\[
= \left[ \text{tr} \left( \rho_A^{-1/2} M_1^\dagger M_1 \rho_A^{-1/2} \right) \right]^{1/2} \cdot \left[ \text{tr} \left( \rho_B^{-1/2} N_1^\dagger N_1 \rho_B^{-1/2} \right) \right]^{1/2}
\]

\[
= \left[ \text{tr} (M_1^\dagger M_1) \right]^{1/2} \cdot \left[ \text{tr} (N_1^\dagger N_1) \right]^{1/2}
\]

\[
= 1.
\]

On the other hand, observe that \( \|\rho_A^{1/2}\|_2 = \|\rho_B^{1/2}\|_2 = 1 \) and \( |\text{tr}(\tilde{\rho}_{AB}\rho_A^{1/2} \otimes \rho_B^{1/2})| = |\text{tr}(\rho_{AB})| = 1 \).

As a result,

\[
\lambda_1 = 1,
\]

and we can take \( M_1 = \rho_A^{1/2} \) and \( N_1 = \rho_B^{1/2} \).

Now for \( R_A \) and \( S_B \) satisfying [11] and [13] we have

\[
|\text{tr}(\tilde{\rho}_{AB} R_A \otimes S_B)| = \left| \sum_{i \geq 2} \lambda_i \langle R^\dagger, M_i \rangle \langle S^\dagger, N_i \rangle \right|
\]

\[
= \left| \sum_{i \geq 2} \lambda_i \langle R^\dagger, M_i \rangle \langle S^\dagger, N_i \rangle \right|
\]

\[
\leq \left( \sum_{i \geq 2} \lambda_i |\langle R^\dagger, M_i \rangle|^2 \right)^{1/2} \cdot \left( \sum_{i \geq 2} \lambda_i |\langle S^\dagger, N_i \rangle|^2 \right)^{1/2}
\]

\[
\leq \lambda_2,
\]

where in the last line we use \( 1 = \|R\|_2^2 = \sum_{i \geq 2} |\langle R^\dagger, M_i \rangle|^2 \) and similarly \( 1 = \sum_{i \geq 2} |\langle S^\dagger, N_i \rangle|^2 \).

These inequalities are tight for \( R = M_2^\dagger \) and \( S = N_2^\dagger \). We conclude that \( \mu(\rho_{AB}) = \lambda_2 \).

\[\square\]

Let us consider the special case where \( A \) and \( B \) are classical registers. If \( \{i\} : 1 \leq i \leq d_A \) and \( \{k\} : 1 \leq k \leq d_B \) are computational bases of \( \mathcal{H}_A \) and \( \mathcal{H}_B \) respectively, then \( \rho_{AB} \) is diagonal in the basis \( \{i\}\{k\} : 1 \leq i \leq d_A, 1 \leq k \leq d_B \). Let us denote \( p_{ik} = \langle i|\rho_{AB}|i\rangle \), so we can think of a joint distribution \( P_{AB} \) with marginals \( P_A \) and \( P_B \). Then it is easy to see that

\[
\mu(P_{AB}) = \max \mathbb{E}(f(i)g(k))
\]

\[
\mathbb{E}(f(i)) = \mathbb{E}(g(k)) = 0,
\]

\[
\mathbb{E}(f(i)^2) = \mathbb{E}(g(k)^2) = 1,
\]

where the maximum is taken over all real functions \( f \) and \( g \) defined on \( \{1, \ldots, d_A\} \) and \( \{1, \ldots, d_B\} \).
respectively, and $E$ denotes the expectation value with respect to $P_{AB}$.

Maximal correlation can be reformulated using Theorem 1. Define
\[
\tilde{p}_{ik} = p_i^{-1/2} p_k^{-1/2} p_{ik},
\]
and let $\tilde{P}_{AB}$ be a $d_A \times d_B$ matrix whose $ik$-th entry is $\tilde{p}_{ik}$. It is easy to see that Schmidt coefficients of $\tilde{\rho}_{AB}$ are in one-to-one correspondence with singular values of $\tilde{P}_{AB}$. So in the classical case $\mu(P_{AB})$ is equal to the second singular value of $\tilde{P}_{AB}$. For example if $P_{AB}$ denotes two perfectly correlated random variables, then $P_{AB}$ is the identity matrix and $\mu(P_{AB}) = 1$. This latter formulation of maximal correlation in the classical case is found by Kang and Ulukus [6] and Kumar [7].

**Theorem 2** $\mu(\cdot)$ satisfies the following properties:

(a) $\mu(\rho_{AB} \otimes \sigma_{A'B'}) = \max\{\mu(\rho_{AB}), \mu(\sigma_{A'B'})\}$.

(b) Let $\Phi : L(\mathcal{H}_B) \to L(\mathcal{H}_{B'})$ be a completely positive trace-preserving super-operator. Let $\sigma_{AB} = \mathcal{I}_A \otimes \Phi_B(\rho_{AB})$. Then $\mu(\sigma_{AB}) \leq \mu(\rho_{AB})$.

**Proof:** (a) Let $\lambda_1 = 1 \geq \lambda_2 \geq \cdots$ and $\xi_1 = 1 \geq \xi_2 \geq \cdots$ be the Schmidt coefficients of $\tilde{\rho}_{AB}$ and $\tilde{\sigma}_{A'B'}$ respectively. By Theorem 1 $\mu(\rho_{AB} \otimes \sigma_{A'B'})$ is equal to the second Schmidt coefficient of $\tilde{\rho}_{AB} \otimes \tilde{\sigma}_{A'B'}$ which is equal to
\[
\max\{\lambda_1, \xi_2, \xi_1, \lambda_2\} = \max\{\lambda_2, \xi_2\} = \max\{\mu(\rho_{AB}), \mu(\sigma_{A'B'})\}.
\]

(b) Any completely positive trace-preserving map is a composition of an isometry and a partial trace. Local isometries obviously do not change $\mu(\rho_{AB})$. Moreover $\mu(\sigma_{AB}) \leq \mu(\sigma_{A'A'})$ is easy to prove. Here we present a proof for the case where $\Phi_B$ is only 2-positive and not necessarily completely positive.

Let $X_A$ and $Y_B$ be the optimizers for $\sigma_{AB}$ satisfying (2) and (3). Let $\Phi^*$ be the adjoint of $\Phi$, i.e., $\text{tr}(\Phi(M)N) = \text{tr}(M\Phi^*(N))$. Note that $\Phi^*$ is 2-positive since $\Phi$ is 2-positive, and $\Phi^*(I) = I$ because $\Phi$ is trace-preserving. Define $Z = \Phi^*(Y)$. Observe that $\sigma_A = \rho_A$ and $\sigma_{B'} = \Phi(\rho_B)$. Thus $\text{tr}(\rho_A X) = \text{tr}(\sigma_A X) = 0$ and $\text{tr}(\rho_A XX^\dagger) = \text{tr}(\sigma_A XX^\dagger) = 1$. Moreover,
\[
\text{tr}(\rho_B Z) = \text{tr}(\rho_B \Phi^*(Y)) = \text{tr}(\Phi(\rho_B)Y) = \text{tr}(\sigma_B Y) = 0,
\]
and
\[
\text{tr}(\rho_{AB} X \otimes Z^\dagger) = \text{tr}(\rho_{AB} X \otimes \Phi^*(Y^\dagger)) = \text{tr}(\mathcal{I}_A \otimes \Phi_B(\rho_{AB})X \otimes Y^\dagger) = \text{tr}(\sigma_{AB'} X \otimes Y^\dagger) = \mu(\sigma_{AB'}).\]

where we use the fact that both $\Phi$ and $\Phi^*$ are hermitian-preserving. Therefore, we conclude that $\mu(\rho_{AB}) \geq \mu(\sigma_{AB'})$ if $\text{tr}(\rho_B ZZ^\dagger) \leq 1$. To prove the latter, observe that
\[
\begin{pmatrix}
YY^\dagger & Y \\
Y^\dagger & I
\end{pmatrix} = 
\begin{pmatrix}
Y \\
I
\end{pmatrix} \begin{pmatrix}
Y^\dagger & I
\end{pmatrix}
\]
is positive semidefinite. Since $\Phi^*$ is 2-positive,
\[
\begin{pmatrix}
\Phi^*(YY^\dagger) & \Phi^*(Y) \\
\Phi^*(Y^\dagger) & \Phi^*(I)
\end{pmatrix} = 
\begin{pmatrix}
\Phi^*(YY^\dagger) & \Phi^*(Y) \\
\Phi^*(Y^\dagger) & I
\end{pmatrix},
\]
is positive semidefinite. This means that 
\[ \Phi^*(YY^\dagger) \geq \Phi^*(Y)\Phi^*(Y^\dagger). \]
Now using \( \rho_B \geq 0 \) we have
\[
\begin{align*}
\text{tr}(\rho_B ZZ^\dagger) &= \text{tr}(\rho_B \Phi^*(Y)\Phi^*(Y^\dagger)) \\
&\leq \text{tr}(\rho_B \Phi^*(YY^\dagger)) \\
&= \text{tr}(\Phi(\rho_B)YY^\dagger) \\
&= \text{tr}(\sigma_B YY^\dagger) \\
&= 1.
\end{align*}
\]
We are done. \( \square \)

The following corollary is the main result of this paper and is a direct consequence of the above theorem.

**Corollary 3** Suppose that \( \rho_{AB}^{\otimes n} \), for some \( n \), can be locally transformed to \( \sigma_{EF} \) (under completely positive trace-preserving super-operators). Then
\[
\mu(\rho_{AB}) \geq \mu(\sigma_{EF}).
\]

The following example reveals the strength of this corollary. Let \( |\psi\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \) be the Bell state on two qubits. Define
\[
\rho_{AB}^{(p)} = (1 - p) \frac{I_{AB}}{4} + p |\psi\rangle\langle\psi|_{AB},
\]
where \( 0 \leq p \leq 1 \) and \( I_{AB}/4 \) is the maximally mixed state. Note that \( \rho_{A}^{(p)} = I_A/2 \) and \( \rho_{B}^{(p)} = I_B/2 \) for every \( p \). Therefore,
\[
\begin{align*}
\mu(\rho_{AB}^{(p)}) &= \max \text{tr}(\rho_{AB}^{(p)} X \otimes Y^\dagger) \\
&= \text{tr}(X) = \text{tr}(Y) = 0, \\
&= \text{tr}(XX^\dagger) = \text{tr}(YY^\dagger) = 2.
\end{align*}
\]
For \( X \) and \( Y \) satisfying the above equations we have
\[
|\text{tr}(\rho_{AB}^{(p)} X \otimes Y^\dagger)| = \left| \frac{1-p}{4} \text{tr}(X \otimes Y^\dagger) + p|\psi\rangle\langle\psi|_{AB}\right| \\
= \frac{p}{2} |\text{tr}(X^T Y^\dagger)| \\
\leq \frac{p}{2} \|X\|_2 \cdot \|Y\|_2 \\
= p.
\]
This upper bound is achievable at \( X = Y = |0\rangle\langle0| - |1\rangle\langle1| \). Therefore,
\[
\mu(\rho_{AB}^{(p)}) = p.
\]
We conclude that entanglement cannot be distilled from \( \rho_{AB}^{(p)} \) for \( p < 1 \) under local operations (but no communication) because for the maximally entangled state we have \( \mu(|\psi\rangle\langle\psi|_{AB}) = \mu(\rho_{AB}^{(1)}) = 1 \). In fact even common randomness cannot be extracted from these states (under local operations).
since for two perfectly correlated bits $U, V$, we have $\mu(P_{UV}) = 1$. Moreover, having infinitely many copies of $\rho_{AB}^{(p)}$ one cannot locally generate a single copy of $\rho_{AB}^{(q)}$ if $q > p$.

**III. OTHER SCHMIDT COEFFICIENTS**

In this section we generalize the data processing inequality of the previous section for $\mu(\cdot)$ to other Schmidt coefficients of $\tilde{\rho}_{AB}$. These new inequalities, however, do not hold in the $n$-letter case in the sense of Corollary 3.

**Theorem 4**  Let $1 = \mu_1(\rho_{AB}) \geq \mu_2(\rho_{AB}) \geq \mu_3(\rho_{AB}) \geq \cdots$ be Schmidt coefficients of $\tilde{\rho}_{AB}$. Let $\Phi_B : \mathcal{L}(\mathcal{H}_B) \rightarrow \mathcal{L}(\mathcal{H}_{B'})$ be a completely positive trace-preserving super-operator and let $\sigma_{AB'} = \mathcal{I}_A \otimes \Phi_B(\rho_{AB})$. Then for every $i$ we have

$$\mu_i(\rho_{AB}) \geq \mu_i(\sigma_{AB'}).$$

To prove this theorem it is more convenient to use the isomorphism between the Hilbert spaces $\mathcal{V} \otimes \mathcal{W}$ and $\mathcal{L}(\mathcal{V}, \mathcal{W})$, and the fact that Schmidt coefficients are mapped to singular values under this isomorphism. To be more precise, let us fix an orthonormal basis $\{|i\rangle, \ldots, |d - 1\rangle\}$ for $\mathcal{H}_A$. Then for every $Z_{AB} \in \mathcal{L}(\mathcal{H}_A) \otimes \mathcal{L}(\mathcal{H}_B)$ there exists a super-operator $\Omega_Z : \mathcal{L}(\mathcal{H}_A) \rightarrow \mathcal{L}(\mathcal{H}_B)$ such that

$$Z_{AB} = \sum_{i,j=0}^{d-1} |i\rangle\langle j| \otimes \Omega_Z(|j\rangle\langle i|).$$

Using the fact that $\{|i\rangle\langle j| : i, j = 0, \ldots, d - 1\}$ is an orthonormal basis for $\mathcal{L}(\mathcal{H}_A)$ it is easy to see that Schmidt coefficients of $Z_{AB}$ are equal to singular values of $\Omega_Z$.

By the above notation we may consider super-operators $\Omega_\rho$ and $\Omega_{\tilde{\rho}}$. Observe that

$$\sum_{i,j} |i\rangle\langle j| \otimes \rho_B^{-1/2} \Omega_\rho(\rho_A^{-1/2}) |j\rangle\langle i| = \sum_{i,j,k} |i\rangle\langle j| \otimes \rho_B^{-1/2} \Omega_\rho(\rho_A^{-1/2}) |k\rangle\langle i|$$

$$= \sum_{i,j,k} |i\rangle\langle k| \rho_A^{-1/2} |j\rangle\langle j| \otimes \rho_B^{-1/2} \Omega_\rho(\rho_A^{-1/2}) |i\rangle,$$

$$= \sum_{i,k} |i\rangle\langle k| \rho_A^{-1/2} \otimes \rho_B^{-1/2} \Omega_\rho(\rho_A^{-1/2}) |i\rangle$$

$$= (I_A \otimes \rho_B^{-1/2}) \left( \sum_{i,k} |i\rangle\langle k| \otimes \Omega_\rho(|k\rangle\langle i|) \right) (\rho_A^{-1/2} \otimes I_A)$$

$$= (I_A \otimes \rho_B^{-1/2}) \rho_{AB} (\rho_A^{-1/2} \otimes I_A)$$

$$= \tilde{\rho}_{AB}.$$

Therefore, by definition we have

$$\Omega_{\tilde{\rho}}(X) = \rho_B^{-1/2} \Omega_\rho(\rho_A^{-1/2} X). \quad (6)$$

We are now ready to prove Theorem 4.
Proof: From the definitions it is clear that $\Omega_{\sigma} = \Phi \circ \Omega_{\rho}$, and then from (10) and $\sigma_A = \rho_A$ we have

$$
\Omega_{\tilde{\sigma}}(X) = \sigma_B^{-1/2} \Omega_{\sigma}(\sigma_A^{-1/2} X) \\
= \sigma_B^{-1/2} \Phi \left( \Omega_{\rho} (\rho_A^{-1/2} X) \right) \\
= \sigma_B^{-1/2} \Phi \left( \rho_B^{-1/2} \Omega_{\rho} (X) \right).
$$

This means that if we define $\Psi : \mathcal{L}(\mathcal{H}_B) \to \mathcal{L}(\mathcal{H}_B')$ by $\Psi(Y) = \sigma_B^{-1/2} \Phi(\rho_B^{1/2} Y)$ then

$$
\Omega_{\tilde{\sigma}} = \Psi \circ \Omega_{\tilde{\rho}}.
$$

Thus given the correspondence between singular values and Schmidt coefficients we conclude that

$$
\mu_i(\sigma_{AB'}) \leq \|\Psi\| \cdot \mu_i(\rho_{AB}),
$$

where $\|\Psi\| = \|\Psi\|_{\infty}$ denotes the operator norm of $\Psi$, and we use (for example) Problem III.6.2 of [10]. Thus it suffices to show that $\|\Psi\| \leq 1$.

Fix an orthonormal basis $\{|0\rangle, |1\rangle, \ldots, |d'-1\rangle\}$ for $\mathcal{H}_B$ and define

$$
\tau_{BB'} = \sum_{k,l=0}^{d-1} |k\rangle\langle l| \otimes \Phi \left( \rho_B^{1/2} |k\rangle\langle l| \rho_B^{1/2} \right).
$$

It is easy to verify that $\tau_{BB'}$ is a density matrix with marginals $\tau_B' = \Phi(\rho_B) = \sigma_B$ and $\tau_B = \rho_B^*$ where by $\rho_B^*$ we mean the entry-wise complex conjugate of $\rho_B$ (with respect to the chosen basis). Moreover, we have $\Omega_{\tau}(X) = \sigma_B^{-1/2} \Phi(\rho_B^{1/2} X^T)$ and then $\Omega_{\tau}(X) = \Psi(X^T)$. As a result, $\|\Psi\| = \|\Omega_{\tau}\|$ which we know is equal to the maximum Schmidt coefficient of $\tau_{BB'}$ which is 1.

\[ \Box \]

IV. EXTREME VALUES

In this section we study the extreme values of maximal correlation.

Theorem 5 0 ≤ $\mu(\rho_{AB})$ ≤ 1 and the followings hold:

(a) $\mu(\rho_{AB}) = 0$ if and only if $\rho_{AB} = \rho_A \otimes \rho_B$, i.e., $\rho_{AB}$ contains neither classical nor quantum correlation.

(b) $\mu(\rho_{AB}) = 1$ if and only if there exist nontrivial local operators $X_A, Y_B$ such that $\rho_{AB}(X_A \otimes I_B) = \rho_{AB}(I_A \otimes Y_B)$. Furthermore, if $X_A, Y_B$ are hermitian, then there are local measurements $\{M_A, I_A - M_A\}$ and $\{N_B, I_B - N_B\}$ such that $\text{tr} (\rho_{AB} M_A \otimes N_B) \neq 0,1$, and

$$
\text{tr} (\rho_{AB} (M_A \otimes (I_B - N_B))) = \text{tr} (\rho_{AB} ((I_A - M_A) \otimes N_B)) = 0.
$$

Proof: $\mu(\rho_{AB}) \geq 0$ is clear from the definition and $\mu(\rho_{AB}) = \lambda_2 \leq \lambda_1 = 1$ follows from the proof of Theorem II.

(a) $\mu(\rho_{AB}) = 0$ if and only if all Schmidt coefficients of $\bar{\rho}_{AB}$ except the first one ($\lambda_1 = 1$) are zero, which means that

$$
\bar{\rho}_{AB} = M_1 \otimes N_1 = \rho_A^{1/2} \otimes \rho_B^{1/2},
$$
or equivalently \( \rho_{AB} = \rho_A \otimes \rho_B \).

(b) Suppose that \( \mu(\rho_{AB}) = 1 \), so there exist \( X_A \) and \( Y_B \) satisfying (2) and (3), and \( \text{tr}(\rho_{AB} X_A \otimes Y_B^\dagger) = 1 \). We assume without loss of generality that \( \rho_A \) and \( \rho_B \) are invertible. Define \( Z_{AB} = X_A \otimes I_B - I_A \otimes Y_B \). Observe that

\[
\text{tr}(\rho_{AB} Z Z^\dagger) = \text{tr}(\rho_{AB} XX^\dagger \otimes I_B) + \text{tr}(\rho_{AB} I_A \otimes YY^\dagger) - \text{tr}(\rho_{AB} X \otimes Y^\dagger) - \text{tr}(\rho_{AB} X^\dagger \otimes Y) = \text{tr}(\rho_A X X^\dagger) + \text{tr}(\rho_B Y Y^\dagger) - 2 = 0.
\]

Since both \( \rho_{AB} \) and \( ZZ^\dagger \) are positive semidefinite we conclude that \( \rho_{AB} Z Z^\dagger = 0 \) and in fact \( \rho_{AB} Z = 0 \). Equivalently, we obtain

\[
\rho_{AB}(X_A \otimes I_B) = \rho_{AB}(I_A \otimes Y_B). \tag{7}
\]

Conversely, if nontrivial operators \( X_A, Y_B \) satisfying the above equation exists, then shifting them with identity we can assume that \( \text{tr}(\rho_A X_A) = \text{tr}(\rho_{AB}(X_A \otimes I_B)) = \text{tr}(\rho_{AB}(I_A \otimes Y_B)) = \text{tr}(\rho_B Y_B) = 0 \). Moreover, by scaling them we can assume that \( \text{tr}(\rho_A X_A X_A^\dagger) = 1 \) which implies

\[
\text{tr}(\rho_B Y_B Y_B^\dagger) = \text{tr}(\rho_{AB}(I_A \otimes Y_B)(I_A \otimes Y_B^\dagger)) = \text{tr}(\rho_A X_A X_A^\dagger) = 1,
\]

where here we use the fact that (7) implies \((X_A^\dagger \otimes I_B) \rho_{AB} = (I_A \otimes Y_B^\dagger) \rho_{AB}\). Thus, \( X_A, Y_B \) satisfy (2) and (3), and we have

\[
\text{tr}(\rho_{AB} X_A \otimes Y_B^\dagger) = \text{tr}(\rho_{AB}(X_A \otimes I_B)(I_A \otimes Y_B^\dagger)) = \text{tr}(\rho_{AB}(I_A \otimes Y_B Y_B^\dagger)) = \text{tr}(\rho_B Y_B Y_B^\dagger) = 1.
\]

Suppose further that \( X_A, Y_B \) are hermitian. Observe that

\[
\rho_{AB}(X_A^2 \otimes I_B) = \rho_{AB}(X_A \otimes I_B)(X_A \otimes I_B)
\]

\[
= \rho_{AB}(I_A \otimes Y_B)(X_A \otimes I_B)
\]

\[
= \rho_{AB}(X_A \otimes I_B)(I_A \otimes Y_B)
\]

\[
= \rho_{AB}(I_A \otimes Y_B)(I_A \otimes Y_B)
\]

\[
= \rho_{AB}(I_A \otimes Y_B^2).
\]

More generally, for every polynomial \( q(t) \) we have \( \rho_{AB}(q(X_A) \otimes I_B) = \rho_{AB}(I_A \otimes q(Y_B)) \). Using (2), \( X_A \) is not a multiple of identity and has a non-trivial eigenspace. On the other hand orthogonal projections on eigenspaces of a hermitian operator can be written as polynomials in terms of that operator with real coefficients. Therefore, there exists a non-zero orthogonal projection \( q(X_A) = M_A \neq I_A \) and a hermitian operator \( q(Y_B) = N_B \) such that

\[
\rho_{AB}(M_A \otimes I_B) = \rho_{AB}(I_B \otimes N_B).
\]
Replacing $N_A$ with $N_A^2$, we may assume that $N_B$ is positive semidefinite because
\[
\rho_{AB}(I_B \otimes N_A^2) = \rho_{AB}(M_A^2 \otimes I_B) = \rho_{AB}(M_A \otimes I_B).
\]

Note that
\[
\text{tr}(\rho_B N_B^n) = \text{tr}(\rho_{AB}(I_A \otimes N_B^n)) = \text{tr}(\rho_{AB}(M_A \otimes I_B)) = \text{tr}(\rho_{AM_A}).
\]

Moreover, since $\rho_A$ is full-rank and $M_A$ is a non-trivial projection, $0 < \text{tr}(\rho_{AM_A}) < 1$. Now if $N_A$ has an eigenvalue greater than 1, since $\rho_B$ is full-rank, $\text{tr}(\rho_B N_B^n)$ would tend to infinity as $n$ goes to infinity. We conclude that all eigenvalues of $N_B$ are less than or equal to 1 and $N_B \leq I_B$.

Consider the local measurements $\{M_A, I_A - M_A\}$ and $\{N_B, I_B - N_B\}$ to be applied on $\rho_{AB}$. The probability of obtaining $M_A$ and $I_B - N_B$ is equal to
\[
\text{tr}(\rho_{AB}(M_A \otimes (I_B - N_B))) = \text{tr}(\rho_{AB}(M_A \otimes I_B)) - \text{tr}(\rho_{AB}(M_A \otimes N_B))
\]
\[
= \text{tr}(\rho_{AB}(M_A \otimes I_B)) - \text{tr}(\rho_{AB}(I_A \otimes N_B)(M_A \otimes I_B))
\]
\[
= \text{tr}(\rho_{AB}(M_A \otimes I_B)) - \text{tr}(\rho_{AB}(M_A \otimes I_B))(M_A \otimes I_B))
\]
\[
= \text{tr}(\rho_{AB}(M_A \otimes I_B)) - \text{tr}(\rho_{AB}(M_A \otimes I_B)) = 0.
\]

Similarly we have $\text{tr}(\rho_{AB}((I_A - M_A) \otimes N_B)) = 0$. We have $\text{tr}(\rho_{AB}M_A \otimes N_B) \neq 0,1$ because
\[
\text{tr}(\rho_{AB}M_A \otimes N_B) = \text{tr}(\rho_{AB}M_A^2 \otimes I_B) = \text{tr}(\rho_{AM_A})
\]
is strictly between 0 and 1.

\[\square\]

**Remark 6** When the marginal states $\rho_A, \rho_B$ are maximally mixed, the optimal operators $X_A, Y_B$ can be chosen to be symmetric. The point is that in this case $\tilde{\rho}_{AB}$ is proportional to $\rho_{AB}$ that is hermitian and belongs to the tensor product of spaces of hermitian operators acting on subsystems $A$ and $B$ as real Hilbert spaces. Therefore, in this case the Schmidt decomposition of $\tilde{\rho}_{AB}$ consists of Hermitian operators and following the proof of Theorem 7 it can be verified that the optimal operators $X_A, Y_B$ in the definition of quantum maximal correlation can be chosen to be hermitian.

**Example 7** In this example we show that optimal $X_A, Y_B$ are not hermitian in general. Let
\[
\rho_{AB} = \frac{1}{\sqrt{2}}(|00\rangle\langle00| + |++\rangle\langle++|),
\]
where $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. Let $X = Y = |0\rangle\langle0| - 2|0\rangle\langle1| - |1\rangle\langle1|$. Then, it is not hard to verify that $\text{tr}(\rho_{AXA}) = \text{tr}(\rho_{BY}) = 0$, $\text{tr}(\rho_{AXA}\rho_{AXA}) = \text{tr}(\rho_{BY}\rho_{BY}) = 1$ and $\text{tr}(\rho_{ABX_A}Y_B) = 1$. Therefore, $\mu(A, B) = 1$. Note that $X_A, Y_B$ are not hermitian and no other such hermitian operators exist that satisfy the above equations. To prove the latter claim, if such hermitian operators exists, then there are local measurements $\{M_A, I_A - M_A\}$ and $\{N_B, I_B - N_B\}$ that satisfy part (b) of Theorem 5. Nevertheless, it is not hard to verify that such local measurements do not exist.

We say that $\rho_{AB}$ has a common data (in the asymptotic sense) if for every $\epsilon > 0$ there exists $n$ and local measurements $\{M_0, M_1 = I - M_0\}$ and $\{N_0, N_1 = I - N_0\}$ with outcomes $U, V$ such that
\[
\text{Pr}(U \neq V) = \text{tr}(\rho_{AB}^{\otimes n}M_0 \otimes N_1) + \text{tr}(\rho_{AB}^{\otimes n}M_1 \otimes N_0) \leq \epsilon,
\]
and \( \Pr(U = 0, V = 0), \Pr(U = 1, V = 1) \geq c \) where \( c > 0 \) is a constant independent of \( n \) and \( \epsilon \).

In the classical case a joint distribution \( P_{UV} \) has a common data if it is decomposable. Decomposability means that \( U \) and \( V \), the ranges (set of alphabets) of \( U, V \), can be decomposed as disjoint unions \( U = U_0 \cup U_1 \) and \( V = V_0 \cup V_1 \) such that \( \Pr[U_i \times V_j] \) is equal to zero if \( i \neq j \) and is positive otherwise. It is shown in [5] that decomposability is equivalent to having a common data even in the asymptotic sense, and that both of these are equivalent to \( \mu(P_{UV}) = 1 \). Here we attempt to generalize this to the quantum case.

**Lemma 8** Let \( U, V \) be two binary random variables such that \( p_{01}, p_{10} \leq \epsilon \). Then

\[
\mu(P_{UV}) \geq 1 - \frac{\epsilon}{p_{00}p_{11}} - \frac{2\epsilon^2}{p_{00}p_{11}}.
\]

**Proof:** \( \mu(P_{UV}) \) is equal to the second singular value of

\[
\tilde{P}_{UV} = \begin{pmatrix}
p_{00} & p_{01} \\
p_{10} & p_{11}
\end{pmatrix}.
\]

We know that the first singular value of \( P_{UV} \) is 1, so the second singular value is equal to

\[
\mu(P_{UV}) = |\det \tilde{P}_{UV}| = \frac{|p_{00}p_{11} - p_{01}p_{10}|}{\sqrt{(p_{00} + p_{01})(p_{00} + p_{10})(p_{10} + p_{11})(p_{01} + p_{11})}}
\]

\[
\geq \frac{p_{00}p_{11} - p_{01}p_{10}}{(p_{00} + \epsilon)(p_{11} + \epsilon)} - \frac{\epsilon^2}{p_{00}p_{11}}
\]

\[
\geq 1 - \frac{\epsilon(p_{00} + p_{11}) + \epsilon^2}{p_{00}p_{11}} - \frac{\epsilon^2}{p_{00}p_{11}}
\]

\[
\geq 1 - \frac{\epsilon}{p_{00}p_{11}} - \frac{2\epsilon^2}{p_{00}p_{11}}.
\]

\( \square \)

**Theorem 9** \( \rho_{AB} \) has a common data (in the asymptotic sense) only if \( \mu(\rho_{AB}) = 1 \).

**Proof:** If \( \rho_{AB} \) has a common data in the asymptotic sense, then by definition for every \( \epsilon > 0 \) and sufficiently large \( n \), \( \rho_{AB}^{\otimes n} \) under local measurements can be transformed to random variables \( U \) and \( V \) such that \( \Pr(U \neq V) \leq \epsilon \) and \( \Pr(U = 0, V = 0), \Pr(U = 1, V = 1) \geq c \) for some constant \( c > 0 \) that is independent of \( n \) and \( \epsilon \). Thus, using Lemma 8 and Corollary 3 we have

\[
\mu(\rho_{AB}) \geq 1 - \frac{\epsilon}{c^2} - \frac{2\epsilon^2}{c^2}.
\]

The claim follows since this inequality holds for all \( \epsilon > 0 \).

\( \square \)

In the example at the end of Section II we see that maximal correlation can indeed take any value between 0 and 1. On pure states however it takes only the extreme values.

**Proposition 10** (i) Suppose \( \rho_{AB} \) is pure. Then \( \mu(\rho_{AB}) = 0 \) if \( \rho_{AB} \) is separable and \( \mu(\rho_{AB}) = 1 \) if \( \rho_{AB} \) is entangled.
(ii) If $\|\rho_{AB} - \tau_{AB}\|_\text{tr} \leq \epsilon$ where $\tau_{AB}$ is a maximally entangled state and $\epsilon \leq 1/10$, then $\mu(\rho_{AB}) \geq 1 - 9\epsilon$.

**Proof:** (i) If $\rho_{AB}$ is separable, $\mu(\rho_{AB}) = 0$ follows from part (a) of Theorem 5. Thus suppose $\rho_{AB} = |\psi\rangle\langle\psi|_{AB}$ where $|\psi\rangle_{AB}$ is entangled with Schmidt decomposition

$$|\psi\rangle_{AB} = \sum_i \alpha_i |v_i\rangle_A \otimes |w_i\rangle_B.$$  

Since $|\psi\rangle_{AB}$ is entangled at least two of the Schmidt coefficients (say) $\alpha_1$ and $\alpha_2$ are non-zero. Define

$$X_A = c\alpha_2^2 |v_1\rangle\langle v_1| - c\alpha_1^2 |v_2\rangle\langle v_2|,$$

and

$$Y_A = c\alpha_2^2 |w_1\rangle\langle w_1| - c\alpha_1^2 |w_2\rangle\langle w_2|,$$

where $c^{-1} = \alpha_1 \alpha_2 (\alpha_1^2 + \alpha_2^2)^{1/2}$. Then $X_A$ and $Y_B$ satisfy (2), (3) and $\text{tr}(\rho_{AB} X_A \otimes Y_B) = 1$. As a result $\mu(\rho_{AB}) = 1$.

(ii) Since $\tau_{AB}$ is a maximally entangled state, there are local measurements $\{M_A^0, M_A^1\}$ and $\{N_B^0, N_B^1\}$ such that $\text{tr}(\tau_{AB} M_A^0 \otimes N_B^0) = \text{tr}(\tau_{AB} M_A^1 \otimes N_B^0) = 0$ and

$$\text{tr}(\tau_{AB} M_A^0 \otimes N_B^0) = \frac{1}{d} \left\lfloor \frac{d}{2} \right\rfloor, \quad \text{tr}(\tau_{AB} M_A^1 \otimes N_B^0) = \frac{1}{d} \left\lceil \frac{d}{2} \right\rceil,$$

where $d$ is the minimum of the dimensions of registers $A$ and $B$. Let

$$p_{uv} = \text{tr}(\rho_{AB} M_A^u \otimes N_B^v).$$

Then, using $\|\rho_{AB} - \tau_{AB}\|_\text{tr} \leq \epsilon$ we find that

$$|p_{00} - \frac{1}{d} \left\lfloor \frac{d}{2} \right\rfloor| + |p_{11} - \frac{1}{d} \left\lceil \frac{d}{2} \right\rceil| + p_{01} + p_{10} \leq \epsilon,$$

which using $\epsilon \leq 1/10$ implies

$$p_{00} p_{11} \geq \left( \frac{1}{d} \left\lfloor \frac{d}{2} \right\rfloor - \epsilon \right) \left( \frac{1}{d} \left\lceil \frac{d}{2} \right\rceil - \epsilon \right) \geq \left( \frac{1}{3} - \epsilon \right) \left( \frac{2}{3} - \epsilon \right) \geq \frac{7 \times 17}{30^2}, \tag{8}$$

and $p_{01} + p_{10} < \epsilon$. By Corollary 3 we have $\mu(\rho_{AB}) \geq \mu(P_{UV})$. So it suffices to show that $\mu(P_{UV}) \geq 1 - 9\epsilon$ which is a simple consequence of Lemma 8.

Although part (ii) of this proposition states the continuity of $\mu(\cdot)$ at maximally entangled states, by part (i) it takes values 0 or 1 on pure states and is not continuous at separable states. We note that such a seemingly undesirable property is unavoidable. The main point is that for every entangled pure state $|\psi\rangle_{AB}$, its $n$-fold tensor product $|\psi\rangle_{AB}^\otimes n$ is close to a maximally entangled state. Then, considering the continuity of $\mu(\cdot)$ at maximally entangled states, we conclude that $\mu(|\psi\rangle\langle\psi|_{AB}) = 1$. \qed
V. CONCLUDING REMARKS

In this paper we generalized a measure of bipartite correlation called maximal correlation to quantum states. We showed that this measure satisfies $\mu(\rho_{AB}^n) = \mu(\rho_{AB})$, and proved a data processing type inequality for it. We note that, being a Schmidt coefficient, the quantum maximal correlation can be computed efficiently.

$\mu(\cdot)$ is a measure of total classical and quantum correlations and takes its maximum value on two perfectly correlated bits. This implies that $\mu(\cdot)$ is not well-behaved under (even one bit of) classical communication. It is tempting to look for a measure of quantum correlation with the above properties that vanishes on classically correlated states. In that case we could use such a measure to study the problem of entanglement distillation under LOCC maps. Even proving the nonexistence of such a measure would be interesting.

Acknowledgements. The author is thankful to Amin Gohari for introducing [6], to Yury Polyanskiy for sending a copy of [8], and to Saleh Rahimi-Keshari for sharing Example [7]. The author is also grateful to the unknown referee who pointed an error in the statement of Theorem [9]. This research was in part supported by National Elites Foundation and by a grant from IPM (No. 91810409).

[1] H. O. Hirschfeld, A connection between correlation and contingency, Proc. Cambridge Philosophical Soc. 31, 520-524 (1935).
[2] H. Gebelein, Das statistische problem der Korrelation als variations-und Eigenwertproblem und sein Zusammenhang mit der Ausgleichungsrechnung, Z. für angewandte Math. und Mech. 21, 364–379 (1941).
[3] A. Rényi, New version of the probabilistic generalization of the large sieve, Acta Math. Hung. 10, 217-226 (1959).
[4] A. Rényi, On measures of dependence, Acta Math. Hung. 10, 441-451 (1959).
[5] H. S. Witsenhausen, On sequences of pairs of dependent random variables, SIAM Journal on Applied Mathematics, 28: 100-113 (1975).
[6] W. Kang and S. Ulukus, A New Data Processing Inequality and Its Applications in Distributed Source and Channel Coding, IEEE Transactions on Information Theory 57, 56–69 (2011).
[7] G. Kumar, Binary Rényi Correlation, http://www.stanford.edu/~gowthamr/research/binary_renyi_correlation.pdf (December 2012).
[8] Y. Polyanskiy, Hypothesis testing via a comparator, Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on, 2206-2210 (2012).
[9] S. Kamath and V. Anantharam, Non-interactive Simulation of Joint Distributions: The Hirschfeld-Gebelein-Rényi Maximal Correlation and the Hypercontractivity Ribbon, Proceedings of the 50th Annual Allerton Conference on Communications, Control and Computing (2012).
[10] R. Bhatia, Matrix Analysis, Springer (2010).
[11] This example is shared with us by Saleh Rahimi-Keshari.