This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2874989, IEEE Transactions on Information Theory

Tradeoff relations between accessible information, informational power, and purity

Michele Dall'Arno and Francesco Buscemi

Abstract—**The accessible information and the informational power quantify the maximum amount of information that can be extracted from a quantum ensemble and by a quantum measurement, respectively. Here, we investigate the tradeoff between the accessible information (informational power, respectively) and the purity of the states of the ensemble (the elements of the measurement, respectively). Under any given *lower* bound on the purity, i) we compute the minimum informational power and show that it is attained by the depolarized uniformly-distributed measurement; ii) we give a lower bound on the accessible information. Under any given *upper* bound on the purity, i) we compute the maximum accessible information and show that it is attained by an ensemble of pairwise commuting states with at most two distinct non-null eigenvalues; ii) we give a lower bound on the maximum informational power. The present results provide, as a corollary, novel sufficient conditions for the tightness of the Jozsa-Robb-Wootters lower bound to the accessible information.**

Index Terms—**Accessible information, informational power, purity.**

We address the problem of communicating classical information over the most general physical channel, that is a quantum channel (classical channels being a particular instance of the quantum case). In particular, we consider the case in which the sender is allowed to encode a classical random variable X on a quantum system, which is then transmitted to a receiver and measured, thus producing an output classical random variable Y. The encoding here produces an ensemble of quantum states, one for each letter in the input alphabet $\mathcal{X} = \{x\}$, whereas the measurement returns a letter in the output alphabet $\mathcal{Y} = \{y\}$.

When the input ensemble is fixed, the final measurement can be optimized to achieve the maximum amount of mutual input-output information $I(X;Y)$. This quantity is defined as the *accessible information* of the ensemble [1]. By direct analogy, the maximum amount of input-output information that can be established for a fixed measurement, by optimizing over all possible input ensembles, is defined as the *informational power* of the measurement [2], [3], [4], [5], [6], [7], [8], [9], [10]. A duality relation between these two information-theoretic measures was established in Ref. [2]. Within this context, one is generally interested in bounding the accessible information and the informational power that can be achieved given some resources, for example for fixed Hilbert space dimension.

A family of quantum states or measurement operators is called "pure" if all its elements are represented by rank-one operators. Mathematically speaking, the *purity* of a positive semi-definite operator X is given by $P(X) = \mathrm{Tr}[X^2]/(\mathrm{Tr}[X])^2$. Intuitively, this number is usually considered as a good proxy for the "classical uncertainty" contained in a state or in a measurement: the higher the purity, the less the classical uncertainty. As it can be readily shown, the purity reaches its maximum ($P = 1$) on rank-one operators. Our main result is to derive analytical bounds on the accessible information and the informational power that consider the purity as a free variable in the problem. In this sense, purity can be considered as a resource, only available in limited amounts.

As an example, let us consider lower bounds on the accessible information [11], [12] and the informational power [5]. These are typically expressed in terms of a quantity called *subentropy* [11] (see Ref. [13] for a study of its properties). In this sense, the subentropy of a given state ρ quantifies the minimum accessible information of any ensemble of *pure* states averaging to ρ. Hence, known subentropy-like lower bounds on informational measures hold only if the optimization is restricted to pure states and measurement elements. In this paper we generalize similar lower and upper bounds by investigating tradeoff relations between accessible information, informational power, and purity, which can now be bounded by any given value $0 \leq P \leq 1$.

More concretely, our contribution is two-fold. First, we consider the case in which an arbitrary *lower* bound on the purity is given. In this case, we derive the minimum informational power of any measurement when its elements are subject to such a purity constraint. We show that it is attained by the "depolarized Scrooge measurement," that is, the uniformly depolarized, uniformly distributed measurement. We also derive a lower bound on the accessible information. This result has important connections with previous literature. It proves a conjecture formulated in Ref. [10], where the accessible information and the informational power of mixed t-design ensembles and measurements – including depolarized Scrooge structures – were studied. In the process, our result corrects Eq. A24, Property 11, of Ref. [14].

The second set of results concerns the case in which an arbitrary *upper* bound on the purity is enforced. Under this assumption, we derive the maximum accessible information of any ensemble when its states are subject to such a purity constraint. We prove that it is attained by a particular class of ensembles with commuting states, each with at most two different non-null eigenvalues. Additionally, we derive a lower bound on the maximum informational power. This result too

M. Dall'Arno is with the Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, 117543, Singapore (e–mail: cqtmda@nus.edu.sg).

F. Buscemi is with the Graduate School of Informatics, Nagoya University, Chikusa-ku, 464-8601, Nagoya, Japan (e–mail: buscemi@i.nagoya-u.ac.jp).

has important connections with previous literature. It allows us to simplify a proof, given in Ref. [15] adopting a topological approach, of the tradeoff between purity (therein referred to as the index of coincidence of a classical probability distribution) and Shannon entropy. Moreover, our formulation can be extended to encompass the case of arbitrary Rényi entropy, not only Shannon's.

Our findings have implications for the problem of the tightness of the Jozsa-Robb-Wootters lower bound on the accessible information, given in Eq.(33) of Ref. [11]. Prior to this work, not much was known about this problem, except for the cases of uniformly distributed pure states (Scrooge ensemble). This contrasts with the case of the Holevo upper bound on accessible information, for which general necessary and sufficient conditions for tightness are known [16], [17]. As a consequence of our results, it follows that the Jozsa-Robb-Wootters lower bound is also tight for uniformly *depolarized* (thus, not pure) Scrooge ensembles.

I. Main results

We consider a quantum system associated with a (finite) n-dimensional Hilbert space $\mathcal{H}$, and we denote with $\mathrm{Lin}(\mathcal{H})$ the space of linear operators on $\mathcal{H}$. Quantum states of such a system are represented by density matrices $\rho \in \mathrm{Lin}(\mathcal{H})$, that is, positive-semidefinite ($\rho \geq 0$) unit-trace ($\mathrm{Tr}[\rho] = 1$) operators. Any discrete quantum ensemble of such a system is represented by a family of sub-normalized states $\{\rho_x \in \mathrm{Lin}(\mathcal{H})\}$, that is, $\rho_x \geq 0$ for any x and $\mathrm{Tr} \sum_x \rho_x = 1$. Equivalently, $\rho := \sum_x \rho_x$ is a quantum state, and we say that the states composing the ensemble average to ρ. Any discrete quantum measurement on such a system is represented by a POVM, that is a family $\{\pi_y \in \mathrm{Lin}(\mathcal{H})\}$ of positive semi-definite operators, such that $\sum_y \pi_y = \mathbb{1}$, where $\mathbb{1}$ represents the unit element, that is the element with probability 1 over any state. The joint probability distribution of outcome y given input x is given by the Born rule, that is $p_{x,y} = \mathrm{Tr}[\rho_x \pi_y]$. In the continuous case, summations must be replaced by integrals. In the following, we will consider both discrete and continuous ensembles and POVMs, and for simplicity we will adopt the discrete notation wherever it suffices.

The accessible information [1] $A(\{\rho_x\})$ and the informational power [2] $W(\{\pi_y\})$ are operationally defined as the maximum amount of information that can be extracted from ensemble $\{\rho_x\}$ and by POVM $\{\pi_y\}$, respectively:

$$A(\{\rho_x\}) = \max_{\{\pi_y\}} I(\{\mathrm{Tr}[\rho_x \pi_y]\}),$$
$$W(\{\pi_y\}) = \max_{\{\rho_x\}} I(\{\mathrm{Tr}[\rho_x \pi_y]\}),$$

where the maxima are over any POVM $\{\pi_y\}$ and ensemble $\{\rho_x\}$, respectively, and $I(\{p_{x,y}\})$ denotes the mutual information of the joint probability distribution $\{p_{x,y}\}$, that is

$$I(\{p_{x,y}\}) := \sum_{x,y} p_{x,y} \ln \frac{p_{x,y}}{p_x p_y},$$

where $\{p_x := \sum_y p_{x,y}\}$ and $\{p_y := \sum_x p_{x,y}\}$ are the marginals of $\{p_{x,y}\}$.

The accessible information and the informational power are related by the following duality formula [2], which holds for any POVM $\{\pi_y\}$:

$$W(\{\pi_y\}) = \max_{\rho} A\left(\{\sqrt{\rho}\pi_y\sqrt{\rho}\}\right), \tag{1}$$

where the maximum is over any state ρ. The Jozsa-Robb-Wootters lower bound on the accessible information [11] of any ensemble $\{\rho_x\}$ is given by

$$A(\{\rho_x\}) \geq Q(\rho) - \sum_x \mathrm{Tr}[\rho_x] Q\left(\frac{\rho_x}{\mathrm{Tr}[\rho_x]}\right), \tag{2}$$

where $\rho := \sum_x \rho_x$ and $Q(\rho)$ denotes the subentropy [11] of ρ (usually defined by Eq. (6), although here we regard $Q(\rho)$ as a particular case of the quantity $Q_A(\rho, P)$ defined by Eq. (4)). The Holevo upper bound [16], [18] on accessible information is given by

$$A(\{\rho_x\}) \leq S(\rho) - \sum_x \mathrm{Tr}[\rho_x] S\left(\frac{\rho_x}{\mathrm{Tr}[\rho_x]}\right), \tag{3}$$

where $S(\rho)$ denotes the Von Neumann entropy [16] of ρ. It is well-known [16], [17] that the bound in Eq. (3) is tight if and only if ρ_x's are pairwise commuting.

The aim of this work is to study lower and upper bounds on the accessible information $A(\{\rho_x\})$ and the informational power $W(\{\pi_y\})$ under constraints on the purity P of states $\{\rho_x\}$ and POVM elements $\{\pi_y\}$, where $P(X) := \mathrm{Tr}[X^2]/\mathrm{Tr}[X]^2$ for any self-adjoint operator X.

A. Minimum information under purity constraint

Our first result is a lower bound on the accessible information and informational power. For fixed Hilbert space dimension n, denote with $Q_A(\rho, P)$ the minimum of the accessible information $A(\{\rho_x\})$ of any ensemble $\{\rho_x\}$ averaging to state ρ such that $P(\rho_x) \geq P$ for any x, that is

$$Q_A(\rho, P) := \min_{\substack{\{\rho_x\} \\ P(\rho_x) \geq P \\ \sum_x \rho_x = \rho}} A(\{\rho_x\}). \tag{4}$$

Analogously, denote with $Q_W(P)$ the minimum of the informational power $W(\{\pi_y\})$ of any POVM $\{\pi_y\}$ such that $P(\{\pi_y\}) \geq P$ for any y. That is,

$$Q_W(P) := \min_{\substack{\{\pi_y\} \\ P(\pi_y) \geq P}} W(\{\pi_y\}). \tag{5}$$

If $P = 1$, the quantity $Q_A(\rho, 1) =: Q(\rho)$ reduces to the well-known subentropy [11]. Notice that, by definition, the subentropy $Q(\phi)$ of any pure state ϕ is zero. Ref. [11] shows that $Q(\rho)$ is attained by the ρ-distorted Scrooge ensemble, that is, the ensemble of pure states $\{n\sqrt{\rho}\phi_x^*\sqrt{\rho}\}$, where $\{\phi_x^*\}$ denotes the uniformly (Haar) distributed ensemble. If $\rho = \sum_k \lambda_k |\lambda_k\rangle\langle\lambda_k|$ is a spectral decomposition of ρ, in the absence of null eigenvalues and degeneracies one explicitly obtains the formula

$$Q(\rho) = -\sum_k \frac{\lambda_k^n \ln \lambda_k}{\prod_{j \neq k}(\lambda_k - \lambda_j)}. \tag{6}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2874989, IEEE Transactions on Information Theory

Limits must be considered in case of null eigenvalues and degeneracies. The formula (6) is often used to define the subentropy. The following expressions for $Q_A(\rho, 1)$ [11] and $Q_W(1)$ [5] follow

$$\max_\rho Q_A(\rho, 1) = Q_W(1) = \ln n - \Sigma_n. \tag{7}$$

Here and in the following we set $\Sigma_k := \sum_{j=2}^k 1/j$.

Our first main result consists of generalizing Eq. (7) to the case of arbitrary purity $P \in [1/n, 1]$.

Theorem 1 (Lower bound under purity constraint)**.** *One has*

$$\begin{aligned}&\max_\rho Q_A(\rho, P)\\ \geq &Q_W(P)\\ =&\ln n - \sum_{k=2}^{n} \binom{n}{k} \frac{a^k (\ln a - \Sigma_k)}{(b-a)^{k-1}} + \frac{b^n (\ln b - \Sigma_n)}{(b-a)^{n-1}},\end{aligned}$$

where $a := (1-\epsilon)/n$ *and* $b := \epsilon + (1-\epsilon)/n$*, with* $\epsilon := \sqrt{(nP-1)/(n-1)}$*. The quantity* $Q_W(P)$ *is attained by the* ϵ*-depolarized Scrooge POVM* $\{n\mathcal{D}_\epsilon(\phi_y^*)\}$*.*

Here and in the following we denote with $\mathcal{D}_\epsilon$, and we call "depolarizing map," the positive (but not completely–positive) linear map given by

$$\mathcal{D}_\epsilon(\rho) := \epsilon\rho + (1-\epsilon)\operatorname{Tr}[\rho]\frac{\mathbb{1}}{n}, \qquad -\frac{1}{n-1} \leq \epsilon \leq 1.$$

Notice that the above map is self-dual with respect to the trace: in other words, it acts on states and measurements in the same way. Also, the map $\mathcal{D}_\epsilon$ is completely positive for $-(n^2-1)^{-1} \leq \epsilon \leq 1$, as shown in Ref. [19], and coincides with the *depolarizing channel* for $0 \leq \epsilon \leq 1$.

B. Maximum information under purity constraint

Our second result is an upper bound on the accessible information and informational power. For fixed Hilbert space dimension n, denote with $S_A(\rho, P)$ the maximum of the accessible information $A(\{\rho_x\})$ of any ensemble $\{\rho_x\}$ averaging to state ρ such that $P(\rho_x) \leq P$ for any x, that is

$$S_A(\rho, P) := \max_{\substack{\{\rho_x\}\\ P(\rho_x)\leq P\\ \sum_x \rho_x = \rho}} A(\{\rho_x\}). \tag{8}$$

Analogously, denote with $S_W(P)$ the maximum of the informational power $W(\{\pi_y\})$ of any POVM $\{\pi_y\}$ such that $P(\pi_y) \leq P$ for any y, that is

$$S_W(P) := \max_{\substack{\{\pi_y\}\\ P(\pi_y)\leq P}} W(\{\pi_y\}).$$

If $P = 1$, the quantity $S_A(\rho, P) =: S(\rho)$ reduces to the well-known Von Neumann entropy. Notice that, by definition, the entropy $S(\phi)$ of any pure state ϕ is zero. It is well-known that $S(\rho)$ is attained by the ensemble given by the spectral decomposition of ρ, and is given by

$$S(\rho) = -\operatorname{Tr}[\rho \log \rho]. \tag{9}$$

The formula (9) is often used to define the entropy. The following expressions for $S_A(\rho, 1)$ and $S_W(1)$ [5] follow

$$S_W(1) = \max_\rho S_A(\rho, 1) = \ln n. \tag{10}$$

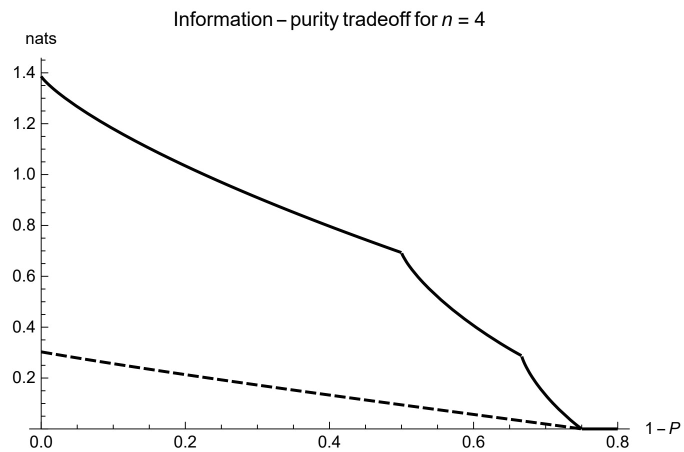
Our second main result consists of generalizing Eq. (10) to the case of arbitrary purity $P \in [1/n, 1]$.

Theorem 2 (Upper bound under purity constraint)**.** *One has*

$$S_W(P) \geq \max_\rho S_A(\rho, P) = \ln n + \lfloor P^{-1} \rfloor a \ln a + b \ln b,$$

where $a := (1 + \sqrt{(P\alpha - 1)/\lfloor P^{-1}\rfloor})/\alpha$ *and* $b := (1 - \sqrt{\lfloor P^{-1}\rfloor(P\alpha-1)})/\alpha$*, with* $\alpha := \lfloor P^{-1} \rfloor + 1$*. The quantity* $\max_\rho S_A(\rho, P)$ *is attained by any ensemble* $\{\rho_x\}$ *of* n *states such that* $\rho_x = a|x\rangle\langle x| + b\sum_{k\neq x}|k\rangle\langle k|$ *for any* x*, for any orthonormal basis* $\{|k\rangle\}$*.*

The results of Theorem 1 and Theorem 2 are depicted in Fig. 1.



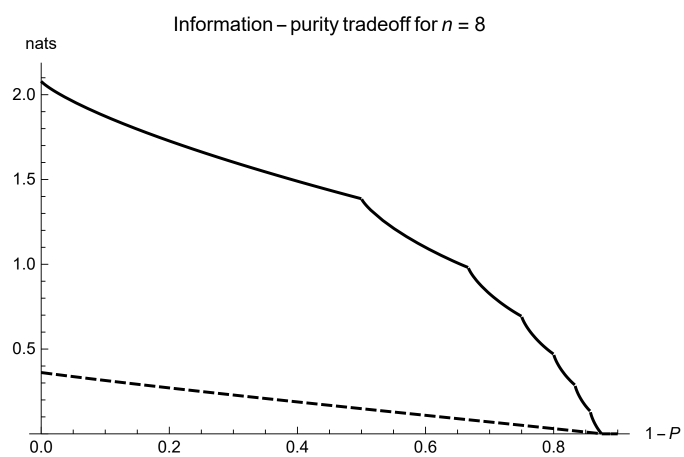


Figure 1. Tradeoff between information and impurity $1 - P$, with $P \in [1/n, 1]$, for quantum ensembles and quantum measurements, for different values of the dimension n. The quantity $Q_W(P)$, as given by Theorem 1, is represented by the lower dashed curve. The maximum value attained at $P = 1$ is $\ln n - \Sigma_n$. The quantity $\max_\rho S_A(\rho, P)$, as given by Theorem 2, is represented by the upper solid curve. Notice that, as a consequence of the dependence on $\lfloor P \rfloor^{-1}$ in Theorem (2), the quantity $\max_\rho Q_A(\rho, P)$ has n singularities, one for each $P = 1/k$, for any integer $1 \leq k \leq n$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2874989, IEEE Transactions on Information Theory

II. Proofs

A. Minimum information under purity constraint

The aim of this section is to prove Theorem 1. Notice that by replacing the maximum over ρ with $\rho = \mathbb{1}/n$ in $\max_\rho Q_A(\rho, P)$ one immediately has

$$\max_\rho Q_A(\rho, P) \geq Q_A\left(\frac{\mathbb{1}}{n}, P\right).$$

Since Eq. (1), with the choice $\rho = \mathbb{1}/n$, implies that

$$W(\{\pi_y\}) \geq A\left(\left\{\frac{\pi_y}{n}\right\}\right),$$

one immediately has

$$Q_W(P) \geq Q_A\left(\frac{\mathbb{1}}{n}, P\right).$$

In other words, both quantities $\max_\rho Q_A(\rho, P)$ and $Q_W(P)$ are lower bounded by the same quantity $Q_A(\mathbb{1}/n, P)$.

In turn, this common lower bound can be lower bounded by Eq. (2). Recalling [11] that $Q(\mathbb{1}/n) = \ln n - \Sigma_n$, one has

$$Q_A\left(\frac{\mathbb{1}}{n}, P\right) \geq \ln n - \Sigma_n - \max_{\substack{\{\rho_x\} \\ P(\rho_x)\geq P \\ \sum_x \rho_x = \mathbb{1}/n}} \sum_x \mathrm{Tr}[\rho_x] Q\left(\frac{\rho_x}{\mathrm{Tr}[\rho_x]}\right). \quad (11)$$

Let us consider the last term in the r.h.s. of Eq. (11). Since relaxing the constraint $\sum_x \rho_x = \mathbb{1}/n$ can only increase the maximum and the maximum of the average over x is not larger than the largest element, one has

$$\max_{\substack{\{\rho_x\} \\ P(\rho_x)\geq P \\ \sum_x \rho_x = \mathbb{1}/n}} \sum_x \mathrm{Tr}[\rho_x] Q\left(\frac{\rho_x}{\mathrm{Tr}[\rho_x]}\right) \leq \max_{\substack{\rho \\ P(\rho)\geq P}} Q(\rho).$$

By replacing this result in Eq. (11) one obtains

$$Q_A\left(\frac{\mathbb{1}}{n}, P\right) \geq \ln n - \Sigma_n - \max_{\substack{\rho \\ P(\rho)\geq P}} Q(\rho). \quad (12)$$

Hence, it suffices to compute the maximum of the subentropy $Q(\rho)$, under the constraint $P(\rho) \geq P$. First, notice that, without loss of generality, the constraint $P(\rho) \geq P$ can be replaced with $P(\rho) = P$, that is

$$\max_{\substack{\rho \\ P(\rho)\geq P}} Q(\rho) = \max_{\substack{\rho \\ P(\rho)= P}} Q(\rho).$$

Indeed, for any state ρ such that $P(\rho) > P$, there exists a value of ϵ such that the depolarized state $\mathcal{D}_\epsilon(\rho)$ is such that $P(\rho) = P$, and $Q(\mathcal{D}_\epsilon(\rho)) > Q(\rho)$. This follows from the fact [13] that $Q(\rho)$ is concave in ρ and maximized by $\rho = \mathbb{1}/n$, and from the fact that $P(\rho)$ is convex in ρ and minimized by $\rho = \mathbb{1}/n$, and hence $Q(\mathcal{D}_\epsilon(\rho))$ and $P(\mathcal{D}_\epsilon(\rho))$ are monotonically increasing and decreasing in ϵ, respectively.

Hence, in the following Lemma we compute the maximum of the subentropy $Q(\rho)$ under constraint $P(\rho) = P$.

Lemma 1. *The maximum of the subentropy $Q(\rho)$ over any state ρ with purity $P(\rho) = P$, for any P, is attained by any ϵ-depolarized pure state $\mathcal{D}_\epsilon(\phi)$, with*

$$\epsilon = \sqrt{\frac{nP-1}{n-1}}.$$

Explicitly one has

$$\max_{\substack{\rho \\ P(\rho)=P}} Q(\rho) = \sum_{k=2}^{n} \binom{n}{k} \frac{a^k (\ln a - \Sigma_k)}{(b-a)^{k-1}} - \frac{b^n (\ln b - \Sigma_n)}{(b-a)^{n-1}} - \Sigma_n,$$

where a and b are the eigenvalues of $\mathcal{D}_\epsilon(\phi)$ with multiplicity $n-1$ and 1, respectively, that is

$$\begin{cases} a := \frac{1-\epsilon}{n}, \\ b := \epsilon + \frac{1-\epsilon}{n}. \end{cases}$$

Proof. We discuss here a sketch of our proof, which is formally provided in the Appendix. Our proof is based on a result of Ref. [14], where the maximization of the subentropy was considered under a constraint on the symmetric polynomial of degree two. We first show that such a constraint is equivalent to a purity constraint, and hence the same state ρ is optimal for the optimization problem considered here. Then, we compute the accessible information $Q(\rho)$ of such an optimal state, a non-trivial task given the $n-1$ degeneracy of its spectrum and hence the impossibility to directly apply Eq. (6). The explicit calculation is carried out in two equivalent ways: by means of an integral representation [11] of the subentropy, and by means of a formula for divided differences [20]. □

Applying Lemma 1 to the bound (12), we can now lower bound the two quantities of interest as follows:

$$\max_\rho Q_A(\rho, P) \geq \ln n - \sum_{k=2}^{n} \binom{n}{k} \frac{a^k (\ln a - \Sigma_k)}{(b-a)^{k-1}} + \frac{b^n (\ln b - \Sigma_n)}{(b-a)^{n-1}}, \quad (13)$$

$$Q_W(P) \geq \ln n - \sum_{k=2}^{n} \binom{n}{k} \frac{a^k (\ln a - \Sigma_k)}{(b-a)^{k-1}} + \frac{b^n (\ln b - \Sigma_n)}{(b-a)^{n-1}}. \quad (14)$$

We prove now the tightness of the lower bound in Eq. (14). In Ref. [10], in the context of mixed t-designs, the accessible information of the ϵ-depolarized Scrooge ensemble $\{\mathcal{D}_\epsilon(\phi^*_x)\}$ and the informational power of the ϵ-depolarized Scrooge POVM $\{n\mathcal{D}_\epsilon(\phi^*_y)\}$ were derived for $0 \leq \epsilon \leq 1$. We generalize that result to the case $-(n-1)^{-1} \leq \epsilon \leq 1$.

To this aim, we generalize an upper bound to the informational power derived in Ref. [10] to the case of accessible information. We start by noticing that by definition

$$I(\{\mathrm{Tr}[\rho_x \pi_y]\}) = \ln n + \sum_{x,y} \mathrm{Tr}[\rho_x \pi_y] \ln\left(\frac{\mathrm{Tr}[\rho_x \pi_y]}{\mathrm{Tr}[\rho_x]\,\mathrm{Tr}[\pi_y]}\right) - \sum_y \mathrm{Tr}[\rho \pi_y] \ln\left(n\frac{\mathrm{Tr}[\rho\pi_y]}{\mathrm{Tr}[\pi_y]}\right).$$

Since both $\{\mathrm{Tr}[\pi_y]/n\}$ and $\{\mathrm{Tr}[\rho\pi_y]\}$ are probability distributions, the last term in the r.h.s. is the relative entropy $D(\{\mathrm{Tr}[\rho\pi_y]\}||\{\mathrm{Tr}[\pi_y]/n\})$, which is non-negative, and null if $\rho = \mathbb{1}/n$. Hence, disregarding the last term in the r.h.s. and setting $\eta(x) := -x\ln x$, one has

$$\begin{aligned} & I(\{\mathrm{Tr}[\rho_x\pi_y]\}) \\ \leq & \ln n + \sum_{x,y} \mathrm{Tr}[\rho_x\pi_y] \ln\left(\frac{\mathrm{Tr}[\rho_x\pi_y]}{\mathrm{Tr}[\rho_x]\,\mathrm{Tr}[\pi_y]}\right) \\ = & \ln n + \sum_{x,y} \mathrm{Tr}[\rho_x]\,\mathrm{Tr}[\pi_y] \frac{\mathrm{Tr}[\rho_x\pi_y]}{\mathrm{Tr}[\rho_x]\,\mathrm{Tr}[\pi_y]} \ln\left(\frac{\mathrm{Tr}[\rho_x\pi_y]}{\mathrm{Tr}[\rho_x]\,\mathrm{Tr}[\pi_y]}\right) \\ = & \ln n - \sum_{x,y} \mathrm{Tr}[\rho_x]\,\mathrm{Tr}[\pi_y]\ \eta\left(\frac{\mathrm{Tr}[\rho_x\pi_y]}{\mathrm{Tr}[\rho_x]\,\mathrm{Tr}[\pi_y]}\right), \end{aligned}$$

which is nicely symmetric in the ensemble and the POVM (notice that in the denominator we have $\mathrm{Tr}[\pi_y]$ rather than $\mathrm{Tr}[\rho\pi_y]$). Hence, we can use it to upper bound both the accessible information and the informational power in the same way. Notice also that the argument of η does not depend on the traces of ρ_x and π_y: these can be rescaled at will without changing the value of the ratio within parentheses. Thus we can recast the problem as an optimization over a single normalized state, as follows.

By definition, the accessible information is the maximum of the mutual information over all POVMs, hence

$$\begin{aligned} & A(\{\rho_x\}) \\ \leq & \ln n - n \min_{\{\pi_y\}} \sum_{x,y} \mathrm{Tr}[\rho_x] \frac{\mathrm{Tr}[\pi_y]}{n}\ \eta\left(\frac{\mathrm{Tr}[\rho_x\pi_y]}{\mathrm{Tr}[\rho_x]\,\mathrm{Tr}[\pi_y]}\right). \end{aligned}$$

In the above equation, we introduced a factor n, so that the coefficient $\mathrm{Tr}[\pi_y]/n$ is a probability distribution. Hence, the minimum over $\{\pi_y\}$ of the average over y is not less than the global minimum, i.e., it can be bounded as follows:

$$A(\{\rho_x\}) \leq \ln n - n \min_{\phi} \sum_x \mathrm{Tr}[\rho_x]\ \eta\left(\frac{\mathrm{Tr}[\rho_x\phi]}{\mathrm{Tr}[\rho_x]}\right), \quad (15)$$

where now the minimum is taken over a single normalized state ϕ (which can be chosen pure, although this does not matter at this point). Notice that equality holds if $\sum_x \rho_x = \mathbb{1}/n$ and $\mathbb{1}/n$ belongs to the convex hull of the set of states attaining the minima over ϕ. The former condition is sufficient for the relative entropy $D(\{\mathrm{Tr}[\rho\pi_y]\}||\{\mathrm{Tr}[\pi_y]/n\})$ to be zero, as discussed before. The latter condition, instead, is necessary and sufficient for the r.h.s. of Eq. (15) to be equivalent to the r.h.s. of the previous equation.

Along exactly the same lines, by definition of informational power, one has

$$\begin{aligned} & W(\{\pi_y\}) \\ \leq & \ln n - n \min_{\{\rho_x\}} \sum_{x,y} \mathrm{Tr}[\rho_x] \frac{\mathrm{Tr}[\pi_y]}{n}\ \eta\left(\frac{\mathrm{Tr}[\rho_x\pi_y]}{\mathrm{Tr}[\rho_x]\,\mathrm{Tr}[\pi_y]}\right). \end{aligned}$$

Again, since $\mathrm{Tr}[\rho_x]$ is a probability distribution over x, the minimum over $\{\rho_x\}$ of the average over x is lower bounded by the minimum over a single normalized state ϕ as follows

$$W(\{\pi_y\}) \leq \ln n - n \min_{\phi} \sum_y \frac{\mathrm{Tr}[\pi_y]}{n}\ \eta\left(\frac{\mathrm{Tr}[\pi_y\phi]}{\mathrm{Tr}[\pi_y]}\right), \quad (16)$$

with equality if $\mathbb{1}/n$ belongs to the convex hull of the set of states attaining the minima over ϕ.

We compute the bounds in Eq. (15) and Eq. (16) for the depolarized version of the uniformly distributed pure ensemble, that is $\{\rho_x = \mathcal{D}_\epsilon(\phi_x^*)\}$, and for the depolarized version of the uniformly distributed rank-one POVM, that is $\{n\mathcal{D}_\epsilon(\phi_y^*)\}$, respectively. We also show that, in these cases, the bounds are tight. To these aims, first notice that the summation in the r.h.s. of Eq. (15) and Eq. (16), which are identical in form, must be replaced in these cases by an integral over uniform measure $\mathrm{d}\mu_x$. Since, by direct calculation,

$$\mathrm{Tr}[\mathcal{D}_\epsilon(\phi_g^*)\phi] = (b-a)\left|\langle\phi|\phi_g^*\rangle\right|^2 + a\ ,$$

one has, setting $g(x) := (b-a)x + a$,

$$\begin{aligned} & \min_\phi \int \mathrm{d}\mu_x\ \langle\phi_x^*|\phi_x^*\rangle\, \eta\left(\frac{\mathrm{Tr}[\mathcal{D}_\epsilon(\phi_x^*)\phi]}{\langle\phi_x^*|\phi_x^*\rangle}\right) \\ = & \min_\phi \int \mathrm{d}\mu_x\ \langle\phi_x^*|\phi_x^*\rangle\, \eta\circ g\left(\frac{\left|\langle\phi|\phi_x^*\rangle\right|^2}{\langle\phi_x^*|\phi_x^*\rangle}\right). \quad (17) \end{aligned}$$

Due to unitary invariance, the minimum over ϕ is independent of ϕ, so in the following ϕ will denote an arbitrarily chosen pure state. Hence, the bounds in Eq. (15) and Eq. (16) are tight.

To compute the integral in the r.h.s. of Eq. (17), we resort to the following result, proved in Refs. [21] and [22]. For any integrable function f one has

$$\begin{aligned} & \int \mathrm{d}\mu_x\, \langle\phi_x^*|\phi_x^*\rangle\, f\left(\frac{\left|\langle\phi|\phi_x^*\rangle\right|^2}{\langle\phi_x^*|\phi_x^*\rangle}\right) \\ = & (n-1)!\left[[f]^{n-1}(1) - \sum_{k=2}^n \frac{[f]^{k-1}(0)}{(n-k)!}\right], \quad (18) \end{aligned}$$

where $\{[f]^m\}_{m=1}^{n-1}$ represents a choice of m-degree antiderivatives of f, namely $[f] := \int \mathrm{d}x f(x)$ and $[f]^m := [[f]^{m-1}]$. Of course, for any choice of $[f]^{m-1}$, one has that $[f]^m$ is uniquely defined up to a constant, but Eq. (18) is independent of such a choice (see Refs. [21], [22]).

It was also shown in Refs. [21] and [22] that

$$[\eta]^m = -\frac{x^{m+1}}{(m+1)!}\left(\log x - \Sigma_{m+1}\right)\ ,$$

and, given that g is an affine function, by direct computation one immediately has

$$[\eta\circ g]^m = \frac{1}{(b-a)^m}[\eta]^m \circ g\ . \quad (19)$$

Since in our case one has $f = \eta\circ g$, by replacing Eq. (19) into Eq. (18) we obtain the accessible information $A(\{\mathcal{D}_\epsilon(\phi_x^*)\})$ of the depolarized version of the uniformly distributed ensemble

$\{\mathcal{D}_\epsilon(\phi_x^*)\}$, and the informational power $W(\{n\mathcal{D}_\epsilon(\phi_y^*)\})$ of the depolarized version of the uniformly distributed rank-one POVM $\{n\mathcal{D}_\epsilon(\phi_y)\}$, as follows

$$\begin{aligned}&A(\{\mathcal{D}_\epsilon(\phi_x^*)\})\\ =&W(\{n\mathcal{D}_\epsilon(\phi_y^*)\})\\ =&\ln n - \sum_{k=2}^{n}\binom{n}{k}\frac{a^k(\ln a - \Sigma_k)}{(b-a)^{k-1}} + \frac{b^n(\ln b - \Sigma_n)}{(b-a)^{n-1}} ,\end{aligned}$$

which proves the tightness of the bound on the informational power in Eq.(14) (but not the tightness of the bound on the accessible information in Eq.(13), given that it is a maximin problem).

Summarizing, we have the following first main result.

Theorem 1 (Lower bound under purity constraint)**.** *One has*

$$\begin{aligned}&\max_\rho Q_A(\rho, P)\\ \geq& Q_W(P)\\ =&\ln n - \sum_{k=2}^{n}\binom{n}{k}\frac{a^k(\ln a - \Sigma_k)}{(b-a)^{k-1}} + \frac{b^n(\ln b - \Sigma_n)}{(b-a)^{n-1}},\end{aligned}$$

where $a := (1-\epsilon)/n$ and $b := \epsilon + (1-\epsilon)/n$, with $\epsilon := \sqrt{(nP-1)/(n-1)}$. The quantity $Q_W(P)$ is attained by the ϵ-depolarized Scrooge POVM $\{n\mathcal{D}_\epsilon(\phi_y^*)\}$.

Theorem 1 sheds new light on the problem of the tightness of the Jozsa-Robb-Wootters lower bound on the accessible information in Eq. (2). Indeed, from Eq. (11) it follows that a sufficient condition for tightness is that the ensemble $\{\rho_x\}$ is the ϵ-depolarized Scrooge ensemble $\{\mathcal{D}_\epsilon(\phi_x^*)\}$, for any $0 \leq \epsilon \leq 1$. This generalizes the previously known fact that the bound in Eq. (2) is tight for $\epsilon = 1$.

B. Maximum information under purity constraint

The aim of this section is to prove Theorem 2. By applying Eq. (3) and using the bound $S(\rho) \leq \ln n$ we have

$$\max_\rho S_A(\rho, P) \leq \ln n - \min_{\substack{\{\rho_x\}\\ P(\rho_x)\leq P}} \sum_x \mathrm{Tr}[\rho_x] S\left(\frac{\rho_x}{\mathrm{Tr}[\rho_x]}\right),$$

which is tight if and only if the minimum over $\{\rho_x\}$ is attained by an ensemble of commuting states averaging to the maximally mixed state. Since the minimum of the average of $S(\rho_x/\mathrm{Tr}[\rho_x])$ is not smaller that the minimum of $S(\rho)$, one has

$$\max_\rho S_A(\rho, P) \leq \ln n - \min_{\substack{\rho\\ P(\rho)\leq P}} S(\rho),$$

which is tight if and only if the maximally mixed state belongs to the convex hull of the set of states attaining the minimum over ρ.

Hence, in the following we address the problem of minimizing the Von Neumann entropy under an upper bound on the purity. First, notice that since $P(\rho) \leq P$ defines a convex set and $S(\rho)$ is concave, the minimum is attained on the boundary, that is

$$\min_{\substack{\rho\\ P(\rho)\leq P}} S(\rho) = \min_{\substack{\rho\\ P(\rho)=P}} S(\rho).$$

In Ref. [15], the maximum and minimum of the Von Neumann entropy under an equality constraint on the purity (therein referred to as the index of coincidence of a classical probability distribution) were derived with a topological approach. We also notice that analogous results were discussed in Ref. [23] to characterize maximally entangled states for given purity of the marginals. In the following Lemma, we provide a simple proof of a partial result of Ref. [15], that we generalize to the case of arbitrary Rényi entropy $H_\alpha(\vec\lambda) := (1-\alpha)^{-1}\ln\sum_k \lambda_k^\alpha$. The case of Von Neumann entropy is recovered since $S(\rho) = \lim_{\alpha\to 1} H_\alpha(\vec\lambda)$, where $\rho := \sum_k \lambda_k |\lambda_k\rangle\langle\lambda_k|$, and the purity constraint $P(\rho) = P$ becomes $|\vec\lambda|_2^2 = P$.

Lemma 2. *Under constraints $\vec\lambda \geq 0$, $|\vec\lambda|_1 = 1$, and $|\vec\lambda|_2^2 = P$, the extrema of $H_\alpha(\vec\lambda)$ are attained by a $\vec\lambda$ with at most two different non-null eigenvalues, that is*

$$\vec\lambda = (a_\pm, \dots, a_\pm, b_\pm, \dots b_\pm, 0, \dots 0),$$

where (a_+, b_+) and (a_-, b_-) are the only two assignments that satisfy the constraints, and are explicitly given by

$$a_\pm := \frac{1 \pm \sqrt{\frac{n_b}{n_a}\left(P(n_a+n_b)-1\right)}}{n_a+n_b}, \tag{20}$$

$$b_\pm := \frac{1 \mp \sqrt{\frac{n_a}{n_b}\left(P(n_a+n_b)-1\right)}}{n_a+n_b}, \tag{21}$$

where n_a and n_b denote the multiplicity of $a_\pm$ and $b_\pm$, respectively. Explicitly one has

$$\min_{\substack{\vec\lambda\geq 0\\ |\vec\lambda|_1=1\\ |\vec\lambda|_2^2=P}} H_\alpha\left(\vec\lambda\right) = \min_{n_a,n_b,\pm}\left[\frac{1}{1-\alpha}\ln\left(n_a a_\pm^\alpha + n_b b_\pm^\alpha\right)\right],$$

$$\max_{\substack{\vec\lambda\geq 0\\ |\vec\lambda|_1=1\\ |\vec\lambda|_2^2=P}} H_\alpha\left(\vec\lambda\right) = \max_{n_a,n_b,\pm}\left[\frac{1}{1-\alpha}\ln\left(n_a a_\pm^\alpha + n_b b_\pm^\alpha\right)\right].$$

Proof. We discuss here a sketch of our proof, which is formally provided in the Appendix. First, we notice that the equality and inequality constrained optimizations of the Rényi entropy are equivalent to a set of equality-only constrained optimizations in smaller dimensions. This allows us to successfully apply the method of Lagrange multipliers to solve such a set of optimization problems. □

We remark that Lemma 2 is in closed form, because it involves a minimization over n_a and n_b, non-negative integers such that $n_a + n_b \leq n$. However, Ref. [15] provides additional insight (for the case of Von Neumann entropy) since such a minimization is solved therein. It was shown in Ref. [15] that for the Von Neumann entropy $S(\rho)$ the minimum over n_a and n_b is attained by $n_a = \lfloor P^{-1}\rfloor$ and $n_b = 1$, and by a_+, b_+. So we have the following upper bound on the accessible information

$$\max_\rho Q_A(\rho, P) \leq \ln n + n_a a_+ \ln a_+ + b_+ \ln b_+,$$

where $n_a = \lfloor P^{-1}\rfloor$ and a_+, b_+ are as given by Lemma 2. Moreover, this bound is tight, since the maximally mixed state

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2874989, IEEE Transactions on Information Theory

belongs to the convex hull of the set of states obtained by considering all the permutations of the eigenvalues $\vec{\lambda}$ as given by Lemma 2, for some fixed basis $\{|\lambda_k\rangle\}$. By taking the same structure as a POVM $\{\pi_y\}$ one also has

$$W(\{\pi_y\}) = \ln n + n_a a_+ \ln a_+ + b_+ \ln b_+.$$

Hence we have our second main result

Theorem 2 (Upper bound under purity constraint)**.** *One has*

$$S_W(P) \geq \max_\rho S_A(\rho, P) = \ln n + \lfloor P^{-1} \rfloor a \ln a + b \ln b,$$

where $a := (1 + \sqrt{(P\alpha - 1)/\lfloor P^{-1}\rfloor})/\alpha$ *and* $b := (1 - \sqrt{\lfloor P^{-1}\rfloor (P\alpha - 1)})/\alpha$, *with* $\alpha := \lfloor P^{-1}\rfloor + 1$. *The quantity* $\max_\rho S_A(\rho, P)$ *is attained by any ensemble* $\{\rho_x\}$ *of* n *states such that* $\rho_x = a\,|x\rangle\langle x| + b \sum_{k\neq x} |k\rangle\langle k|$ *for any* x, *for any orthonormal basis* $\{|k\rangle\}$.

III. Conclusion

Known subentropy-like lower bounds on informational measures (accessible information and informational power) all assume the optimization to be restricted to pure states and POVM elements. In this work, we relaxed this assumption, by regarding purity as a resource, thus recasting the problem as an information-purity tradeoff. In particular, we computed the minimum informational power when the purity is lower bounded, and the maximum accessible information when the purity is upper bounded. We provided bounds for the other cases. We also discussed the problem of the tightness of the Jozsa-Robb-Wootters lower bound on accessible information, giving new cases in which it is tight.

We conclude by discussing some relevant open problems:

- It is still an open problem whether our bounds in Theorem 1 and 2 are tight.
- In Lemma 1 we derived the maximum of the subentropy under a purity constraint; analogously, in Lemma 2 we derived the maximum and minimum of the Von Neumann entropy under a purity constraint. It is still open the problem of deriving the minimum of the subentropy under a purity constraint. One approach to this problem would involve extending the proof technique of Lemma 5 of Ref. [14].
- Here we showed that the Jozsa-Robb-Wootters bound in Eq. (2) is tight not only for the Scrooge pure ensembles and measurements, but also when these are ϵ-depolarized. It is still open the problem of deriving necessary and sufficient conditions for the tightness of the subentropy lower bounds in general.
- Closed expressions for the quantities $Q_A(\rho, P)$ and $S_A(\rho, P)$ are well-known for the case $P = 1$, for any ρ. We introduced closed expressions for any P, in the case $\rho = \mathbb{1}/n$. Deriving closed expressions for $Q_A(\rho, P)$ and $S_A(\rho, P)$ for any ρ and P is still an open problem. In this sense, there is still a trade-off in our current understanding of the relation between information and purity.

Appendix: Proofs of the Lemmas

Here we prove Lemmas 1 and 2, that we recall for convenience.

A. Maximum subentropy under purity constraint

Lemma 1. *The maximum of the subentropy* $Q(\rho)$ *over any state* ρ *with purity* $P(\rho) = P$, *for any* P, *is attained by any* ϵ*-depolarized pure state* $\mathcal{D}_\epsilon(\phi)$, *with*

$$\epsilon = \sqrt{\frac{nP - 1}{n - 1}}. \tag{22}$$

Explicitly one has

$$\max_{\substack{\rho \\ P(\rho)=P}} Q(\rho) = \sum_{k=2}^{n} \binom{n}{k} \frac{a^k (\ln a - \Sigma_k)}{(b-a)^{k-1}} - \frac{b^n (\ln b - \Sigma_n)}{(b-a)^{n-1}} - \Sigma_n,$$

where a *and* b *are the eigenvalues of* $\mathcal{D}_\epsilon(\phi)$ *with multiplicity* $n-1$ *and* 1, *respectively, that is*

$$\begin{cases} a := \frac{1-\epsilon}{n}, \\ b := \epsilon + \frac{1-\epsilon}{n}. \end{cases}$$

Proof. Upon setting $\rho =: \sum_k \lambda_k |\lambda_k\rangle\langle\lambda_k|$ and $e_2(\rho) := \sum_{k<j} \lambda_k \lambda_j$, it has been proven [14] (see Property 7 and Lemma 5 therein) that

$$\max_{\substack{\rho \\ e_2(\rho)=E}} Q(\rho)$$

is attained by $\rho = \mathcal{D}_\epsilon(\phi)$, where ϕ is any pure state and ϵ is the non-negative parameter such that the constraint $e_2(\rho) = E$ is satisfied. Since by explicit computation one has $P(\rho) = 1 - 2e_2(\rho)$, the maximum under purity constraint is also attained by $\rho = \mathcal{D}_\epsilon(\phi)$ and ϵ is the non-negative parameter such that the constraint $P(\rho) = P$ is satisfied. By explicit computation one has

$$P(\rho) = \frac{(n-1)\epsilon^2 + 1}{n},$$

hence Eq. (22) immediately follows.

In order to compute $Q(\mathcal{D}_\epsilon(\phi))$, Eq. (6) is unpractical as the spectrum of $\mathcal{D}_\epsilon(\phi)$ is degenerate. Here we compute $Q(\mathcal{D}_\epsilon(\phi))$ using the integral representation of $Q(\rho)$ derived in Ref. [11]. One has $Q(\rho) = G(\rho) - \Sigma_n$, where

$$G(\rho) := -n \int \mathrm{d}x \left(\sum_{k=1}^{n} \lambda_k x_k \right) \ln \left(\sum_{k=1}^{n} \lambda_k x_k \right),$$

and $\rho = \sum_{k=1}^n \lambda_k |\lambda_k\rangle\langle\lambda_k|$. The integral is over the simplex of probabilities given by $x_k \geq 0$ for any k and $\sum_k x_k = 1$, that is

$$\int \mathrm{d}x := N \int_0^1 \mathrm{d}x_1 \cdots \int_0^{1-x_1 \cdots - x_{n-2}} \mathrm{d}x_{n-1},$$

where N denotes a normalization factor that was derived e.g. in Eq. (A1), Appendix 1, of Ref. [11].

For the sake of completeness, let us here compute N again by iteratively applying the integration formula

$$\int_0^\beta \mathrm{d}x\, (\beta - x)^m = \frac{\beta^{m+1}}{m+1},$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2874989, IEEE Transactions on Information Theory

easily obtained by replacing $t := \beta - x$, thus eventually obtaining

$$\int_0^1 \mathrm{d}x_1 \cdots \int_0^{1-x_1 \cdots - x_{k-1}} \mathrm{d}x_k = \frac{1}{k!}, \qquad \forall k,$$

Hence the condition $\int \mathrm{d}x = 1$ requires $N = (n-1)!$.

To compute $G(\mathcal{D}_\epsilon(\phi))$, notice that for $\rho = \mathcal{D}_\epsilon(\phi)$ one has $\lambda_k = a$ for $1 \le k \le n-1$ and $\lambda_n = b$, with $a = (1-\epsilon)/n$ and $b = \epsilon + (1-\epsilon)/n$. We set $c := b - a$.

We compute now $G(\mathcal{D}_\epsilon(\phi))$ by iteratively applying the integration formulas

$$\int_0^{1-\alpha} \mathrm{d}x (b - c(\alpha + x))^{m-1} \ln(b - c(\alpha + x)) \\ = \frac{(b - c\alpha)^m \left(\ln(b - c\alpha) - \frac{1}{m} \right) - a^m \left(\ln a - \frac{1}{m} \right)}{mc},$$

easily derived by substituting $t := b - c(\alpha - x)$ and by partial integration, and

$$\int_0^{1-\alpha} \mathrm{d}x (b - c(\alpha + x))^{m-1} = \frac{(b - c\alpha)^m - a^m}{mc},$$

easily derived by substituting $t := b - c(\alpha - x)$, thus eventually obtaining

$$G(\mathcal{D}_\epsilon(\phi)) = \sum_{k=2}^n \binom{n}{k} \frac{a^k (\ln a - \Sigma_k)}{(b-a)^{k-1}} - \frac{b^n (\ln b - \Sigma_n)}{(b-a)^{n-1}},$$

or equivalently

$$Q(\mathcal{D}_\epsilon(\phi)) \\ = \sum_{k=2}^n \binom{n}{k} \frac{a^k (\ln a - \Sigma_k)}{(b-a)^{k-1}} - \frac{b^n (\ln b - \Sigma_n)}{(b-a)^{n-1}} - \Sigma_n. \quad (23)$$

An alternative is to compute $Q(\mathcal{D}_\epsilon(\phi))$ by divided differences (see Eq. 11 of Ref. [20]), in which case one has

$$Q(\mathcal{D}_\epsilon(\phi)) = \frac{1}{(n-2)!} \frac{\partial^{n-2}}{\partial a^{n-2}} \left(\frac{a^n \log a}{b-a} - \frac{b^n \log b}{b-a} \right). \quad (24)$$

One immediately has

$$\frac{\partial^{n-2}}{\partial a^{n-2}} \frac{b^n \log b}{b-a} = (n-2)! \frac{b^n \log b}{(b-a)^{n-1}}. \quad (25)$$

By applying the multinomial theorem, one also has

$$\frac{\partial^{n-2}}{\partial a^{n-2}} \frac{a^n \log a}{b-a} \\ = \sum_{k_1+k_2+k_3=n-2} \frac{(n-2)!}{k_1! k_2! k_3!} \times \\ \times \left(\frac{\partial^{k_1}}{\partial a^{k_1}} a^n \right) \left(\frac{\partial^{k_2}}{\partial a^{k_2}} \log a \right) \left(\frac{\partial^{k_3}}{\partial a^{k_3}} (b-a)^{-1} \right).$$

Since of course

$$\frac{\partial^k}{\partial a^k} a^n = \frac{n!}{(n-k)!} a^{n-k},$$

$$\frac{\partial^k}{\partial a^k} \log a = \begin{cases} \log a & \text{if } k = 0, \\ (-1)^{k-1} (k-1)! a^{-k} & \text{if } k > 0, \end{cases}$$

$$\frac{\partial^k}{\partial a^k} (b-a)^{-1} = k! (b-a)^{-(k+1)},$$

one has

$$\frac{\partial^{n-2}}{\partial a^{n-2}} \frac{a^n \log a}{b-a} \\ = (n-2)! \sum_{k=2}^n \frac{a^k}{(b-a)^{k-1}} \times \\ \times \left[\binom{n}{k} \log a - \sum_{j=1}^{n-2} \binom{n}{k+j} \frac{(-1)^j}{j} \right]. \quad (26)$$

Combining Eq. (24), Eq. (25), and Eq. (26) one finally has

$$Q(\mathcal{D}_\epsilon(\phi)) \\ = \sum_{k=2}^n \binom{n}{k} \frac{a^k (\ln a - \Sigma_k)}{(b-a)^{k-1}} - \frac{b^n (\ln b - \Sigma_n)}{(b-a)^{n-1}} - \Sigma_n + \\ + \Sigma_n \left[(n-1)a + b - 1 \right]. \quad (27)$$

which differs from Eq. (23) only by a term that vanishes due to unit-trace. The r.h.s. of Eq. (27) should replace the r.h.s. of Eq. A24, Property 11, of Ref. [14]. □

B. Extremal Rényi entropies under purity constraint

Lemma 2. *Under constraints $\vec{\lambda} \ge 0$, $|\vec{\lambda}|_1 = 1$, and $|\vec{\lambda}|_2^2 = P$, the extrema of $H_\alpha(\vec{\lambda})$ are attained by a $\vec{\lambda}$ with at most two different non-null eigenvalues, that is*

$$\vec{\lambda} = (a_\pm, \dots, a_\pm, b_\pm, \dots b_\pm, 0, \dots 0),$$

where (a_+, b_+) and (a_-, b_-) are the only two assignments that satisfy the constraints, and are explicitly given by

$$a_\pm := \frac{1 \pm \sqrt{\frac{n_b}{n_a} \left(P(n_a + n_b) - 1 \right)}}{n_a + n_b}, \quad (28)$$

$$b_\pm := \frac{1 \mp \sqrt{\frac{n_a}{n_b} \left(P(n_a + n_b) - 1 \right)}}{n_a + n_b}, \quad (29)$$

where n_a and n_b denote the multiplicity of $a_\pm$ and $b_\pm$, respectively. Explicitly one has

$$\min_{\substack{\vec{\lambda} \ge 0 \\ |\vec{\lambda}|_1 = 1 \\ |\vec{\lambda}|_2^2 = P}} H_\alpha\left(\vec{\lambda}\right) = \min_{n_a, n_b, \pm} \left[\frac{1}{1-\alpha} \ln \left(n_a a_\pm^\alpha + n_b b_\pm^\alpha \right) \right],$$

$$\max_{\substack{\vec{\lambda} \ge 0 \\ |\vec{\lambda}|_1 = 1 \\ |\vec{\lambda}|_2^2 = P}} H_\alpha\left(\vec{\lambda}\right) = \max_{n_a, n_b, \pm} \left[\frac{1}{1-\alpha} \ln \left(n_a a_\pm^\alpha + n_b b_\pm^\alpha \right) \right].$$

Proof. We consider the following optimization problems

$$\min_{\substack{\{\lambda_k\}_{k=1}^n \ge 0 \\ |\vec{\lambda}|_1 = 1 \\ |\vec{\lambda}|_2^2 = P}} H_\alpha(\vec{\lambda}), \qquad \text{and} \qquad \max_{\substack{\{\lambda_k\}_{k=1}^n \ge 0 \\ |\vec{\lambda}|_1 = 1 \\ |\vec{\lambda}|_2^2 = P}} H_\alpha(\vec{\lambda}). \quad (30)$$

We iteratively recast these equality- and inequality-constrained programs in dimension n into a set of equality-constrained programs in smaller dimensions. Indeed, the extrema in Eq.(30) are obtained by

$$\min_{\substack{\{\lambda_k\}_{k=1}^n \\ |\vec{\lambda}|_1 = 1 \\ |\vec{\lambda}|_2^2 = P}} H_\alpha(\vec{\lambda}), \qquad \text{and} \qquad \max_{\substack{\{\lambda_k\}_{k=1}^n \\ |\vec{\lambda}|_1 = 1 \\ |\vec{\lambda}|_2^2 = P}} H_\alpha(\vec{\lambda}),$$

and on the positivity faces, characterized by at least one entry equal to zero or equivalently by dimension $n-1$. Then, the problem on the positivity faces is

$$\min_{\substack{\{\lambda_k\}_{k=1}^{n-1}\geq 0\\ |\vec{\lambda}|_1=1\\ |\vec{\lambda}|_2^2=P}} H_\alpha(\vec{\lambda}), \qquad \text{and} \qquad \max_{\substack{\{\lambda_k\}_{k=1}^{n-1}\geq 0\\ |\vec{\lambda}|_1=1\\ |\vec{\lambda}|_2^2=P}} H_\alpha(\vec{\lambda}).$$

By iterating, the solutions of the programs in Eq. (30) are the solutions of this set of programs

$$\min_{\substack{\{\lambda_k\}_{k=1}^{m}\\ |\vec{\lambda}|_1=1\\ |\vec{\lambda}|_2^2=P}} H_\alpha(\vec{\lambda}), \qquad \text{and} \qquad \max_{\substack{\{\lambda_k\}_{k=1}^{m}\\ |\vec{\lambda}|_1=1\\ |\vec{\lambda}|_2^2=P}} H_\alpha(\vec{\lambda}), \tag{31}$$

for $1 \leq m \leq n$.

We now proceed solving the programs in Eq. (31). Notice first that the extrema of H_α are attained in the same points as the extrema of $\sum_k \lambda_k^\alpha$ since $\ln$ is monotonic increasing. By introducing Lagrange multipliers μ and ν one has

$$F_\alpha := \sum_k \lambda_k^\alpha - \mu\left(\sum_k \lambda_k - 1\right) - \nu\left(\sum_k \lambda_k^2 - P\right),$$

which for $\alpha = 1$ becomes

$$F_1 := -\sum_k \lambda_k \ln \lambda_k - \mu\left(\sum_k \lambda_k - 1\right) - \nu\left(\sum_k \lambda_k^2 - P\right).$$

Thus one has

$$\frac{\partial F_\alpha}{\partial \lambda_k} = \alpha\lambda_k^{\alpha-1} - \mu - 2\nu\lambda_k, \tag{32}$$

which for $\alpha = 1$ becomes

$$\frac{\partial F_1}{\partial \lambda_k} = -\ln\lambda_k - 1 - \mu - 2\nu\lambda_k. \tag{33}$$

Since Eq. (32) and Eq. (33) depend on λ_k only (that would not be the case if we had not removed the positivity constraint) and have well-defined concavity, the optimal $\vec{\lambda}$ has at most two different non-null entries, that we call a and b, and that do not depend on k. Then, the constraints $|\vec{\lambda}|_1 = 1$ and $|\vec{\lambda}|_2^2 = P$ give the following conditions on a and b:

$$\begin{cases} n_a a + n_b b = 1, \\ n_a a^2 + n_b b^2 = P. \end{cases}$$

This systems admits two solutions, (a_+, b_+) and (a_-, b_-), as given by Eq. (28) and Eq. (29). Due to the constraint $n_a + n_b \leq n$, the number of such $\vec{\lambda}$'s is finite, which provides a closed form solution of the optimization. $\square$

Acknowledgments

The authors are grateful to Anna Szymusiak for pointing out Refs. [21], [22] and to Richard Jozsa for insightful comments based on an early version of this work. M. D. acknowledges support from the Ministry of Education and the Ministry of Manpower (Singapore). F. B acknowledges support from the Japan Society for the Promotion of Science (JSPS) KAKENHI, Grant No. 17K17796. This work was partly supported by the program for FRIAS-Nagoya IAR Joint Project Group. This work is dedicated to the memory of Graeme Mitchison.

References

[1] M. M. Wilde, *Quantum Information Theory* (Second Edition, Cambridge University Press, 2017).
[2] M. Dall'Arno, G. M. D'Ariano, and M. F. Sacchi, *Informational power of quantum measurements*, Phys. Rev. A **83**, 062304 (2011).
[3] A. S. Holevo, *Information capacity of a quantum observable*, Problems Inform. Transmission **48**, 1 (2012).
[4] W. Słomczyński and A. Szymusiak, *Highly symmetric POVMs and their informational power*, Quant. Inf. Proc. **15**, 565 (2014).
[5] M. Dall'Arno, F. Buscemi, and M. Ozawa, *Tight bounds on accessible information and informational power*, J. Phys. A: Math. Theor. **47**, 235302 (2014).
[6] A. Szymusiak, *Maximally informative ensembles for SIC-POVMs in dimension 3*, J. Phys. A: Math. Theor. **47**, 445301 (2014).
[7] M. Dall'Arno, *Accessible Information and Informational Power of Quantum 2-designs*, Phys. Rev. A **90**, 052311 (2014).
[8] M. Dall'Arno, *Hierarchy of Bounds on Accessible Information and Informational Power*, Phys. Rev. A **92**, 012328 (2015).
[9] A. Szymusiak and W. Słomczyński, *Informational power of the Hoggar SIC-POVM*, Phys. Rev. A **94**, 012122 (2016).
[10] S. Brandsen, M. Dall'Arno, and A. Szymusiak, *Communication capacity of mixed quantum t designs*, Phys. Rev. A **94**, 022335 (2016).
[11] R. Jozsa, D. Robb, and W. K Wootters, *A Lower Bound for Accessible Information in Quantum Mechanics* Phys. Rev. A **49**, 668 (1994).
[12] C.A. Fuchs and C.M. Caves, *Ensemble-Dependent Bounds for Accessible Information in Quantum Mechanics*. Phys. Rev. Lett. **73**, 3047 (1994).
[13] N. Datta, T. Dorlas, R. Jozsa, and F. Benatti, *Properties of subentropy*. Journal of Mathematical Physics **55**, 062203 (2014).
[14] R. Jozsa and G. Mitchison, *Symmetric polynomials in information theory: Entropy and subentropy*, Journal of Mathematical Physics **56**, 062201 (2015).
[15] P. Harremoës and F. Topsøe , *Inequalities Between Entropy and Index of Coincidence Derived From Information Diagrams*, IEEE Trans. Inf. Theory **47**, 2944 (2001)
[16] A. S. Holevo, *Bounds for the quantity of information transmitted by a quantum communication channel*, Problems Inform. Transmission **9**, 177 (1973).
[17] M. B. Ruskai, *Inequalities for Quantum Entropy: A Review with Conditions for Equality*, J. Math. Phys. **43**, 4358 (2002).
[18] A. S. Holevo, *Statistical structure of quantum theory*, Lecture Notes in Physics. Monographs **67** (Springer-Verlag, Berlin, 2001).
[19] C. King, "The capacity of the quantum depolarizing channel", IEEE Transactions on Information Theory **49**, 221 (2003).
[20] I. S. B. Sardharwalla, S. Strelchuk, and R. Jozsa, *Quantum conditional query complexity*, Quantum Information & Computation **17**, 0541 (2017).
[21] K. R. W. Jones, *Principles of quantum inference*, Annals of Physics **207**, 140 (1991).
[22] K. R. W. Jones, *Quantum limits to information about states for finite dimensional Hilbert space*, Journal of Physics A: Mathematical and General **24**, 121 (1991).
[23] T.-C. Wei, K. Nemoto, P. M. Goldbart, P. G. Kwiat, W. J. Munro, and F. Verstraete, *Maximal entanglement versus entropy for mixed quantum states*. Phys. Rev. A **67**, 022110 (2003).

Michele Dall'Arno received his PhD in theoretical physics from the University of Pavia, Italy, in 2011. After post-doc positions in ICFO, Barcelona, and Nagoya, Japan, he joined the Centre for Quantum Technologies, National University of Singapore, where he is a post-doctoral researcher since 2014.

Francesco Buscemi received his PhD in theoretical physics from the University of Pavia, Italy, in 2006. After post-doc positions in Tokyo, Japan, and Cambridge, UK, he joined Nagoya University in 2009, where he is a tenured associated professor in mathematical informatics since 2014. In 2018 he was awarded the Birkhoff-von Neumann Prize by the International Quantum Structures Association.