Lovász's Theta Function, Rényi's Divergence and the Sphere-Packing Bound

Marco Dalai

Department of Information Engineering
University of Brescia - Italy
Email: marco.dalai@ing.unibs.it

Abstract—**Lovász's bound to the capacity of a graph and the the sphere-packing bound to the probability of error in channel coding are given a unified presentation as information radii of the Csiszár type using the Rényi divergence in the classical-quantum setting. This brings together two results in coding theory that are usually considered as being of a very different nature, one being a "combinatorial" result and the other being "probabilistic". In the context of quantum information theory, this difference disappears.**

I. INTRODUCTION

One of the central topics in coding theory is the problem of bounding the probability of error of optimal codes for communication over a given channel. Shannon [1] introduced the notion of channel capacity C, which represents the largest rate at which information can be sent through the channel with probability of error that vanishes with increasing block-length. He then also introduced [2] the notion of zero-error capacity C_0 as the largest rate at which information can be sent with probability of error precisely equal to zero. For rates in the range $C_0 < R < C$, the probability of error is known to decrease exponentially in the block-length n as

$$P_e \approx e^{-nE(R)}, \tag{1}$$

where $E(R)$ is the so called reliability function of the channel. While in the the region of high rates the function $E(R)$ is known exactly, in the low rate region little is known about P_e; determining both $E(R)$ and C_0 is an unsolved problem and only upper and lower bounds for these quantities are known.

Two of the most important contributions to the study of $E(R)$ and of C_0, which came respectively in the '60s and in the '70s, are the sphere-packing bound $E(R) \leq E_{sp}(R)$ [3] and Lovász's bound $C_0 \leq \vartheta$ [4]. These two bounds are usually considered as being the result of totally unrelated methods. In this paper, we show that this is not the case, and that Lovász's result comes as a special case of the sphere-packing bound once we move to the more general context of classical-quantum channels. In order to do that, we extend to the classical-quantum case a result of Csiszár that allows us to express the sphere-packing exponent [5] in terms of an information radius using the Rényi divergence. Lovász's result then emerges naturally as a special case. This leads to a unified view of two of the most important bounds to $E(R)$ and to C_0, showing that quantum information theory is a useful tool to attack problems at the intersection of probability and combinatorics in classical information theory.

II. CLASSICAL CHANNELS

A. Basic notations and definitions

Let $W(x|y)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, be the transition probabilities of a discrete memoryless channel $W : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ are finite sets. For a sequence $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ and a sequence $\mathbf{y} = (y_1, y_2, \ldots, y_n) \in \mathcal{Y}^n$, the probability of observing $\mathbf{y}$ at the output of the channel given $\mathbf{x}$ at the input is

$$W^{(n)}(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} W(y_i|x_i). \tag{2}$$

A block code with M messages and block-length n is a mapping from a set $\{1, 2, \ldots, M\}$ of M messages onto a set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\}$ of M sequences in $\mathcal{X}^n$. The rate R of the code is defined as $R = \log M / n$. A decoder is a mapping from $\mathcal{Y}^n$ into the set of possible messages $\{1, 2, \ldots, M\}$. If message m is to be sent, the encoder transmits the codeword $\mathbf{x}_m$ through the channel. An output sequence $\mathbf{y}$ is received by the decoder, which maps it to a message $\hat{m}$. An error occurs if $\hat{m} \neq m$.

Let $Y_m \subseteq \mathcal{Y}^n$ be the set of output sequences that are mapped into message m. When message m is sent, the probability of error is

$$P_{e|m} = \sum_{\mathbf{y} \notin Y_m} W^{(n)}(\mathbf{y}|\mathbf{x}_m). \tag{3}$$

The maximum error probability of the code is defined as the largest $P_{e|m}$, that is,

$$P_{e,\max} = \max_m P_{e|m}. \tag{4}$$

Let $P_{e,\max}^{(n)}(R)$ be the smallest maximum error probability among all codes of length n and rate at least R. Shannon's theorem [1] states that sequences of codes exists such that $P_{e,\max}^{(n)}(R) \to 0$ as $n \to \infty$ for all rates smaller than a constant C, called *channel capacity*, which is given by the expression

$$C = \max_P \sum_{x,y} P(x)W(y|x) \log \frac{W(y|x)}{\sum_{x'} P(x')W(y|x')}, \tag{5}$$

where the maximum is over all probability distributions on the input alphabet.

For $R < C$, Shannon's theorem only asserts that $P_{e,\max}^{(n)}(R) \to 0$ as $n \to \infty$. For a range of rates $C_0 \leq R \leq C$, the optimal probability of error $P_{e,\max}^{(n)}(R)$ is known to have

an exponential decrease in n, and it is thus useful to define the *reliability function* of the channel as

$$E(R) = \limsup_{n \to \infty} -\frac{1}{n} \log P_{e,\max}^{(n)}(R). \qquad (6)$$

The value C_0 is the so called *zero-error capacity*, also introduced by Shannon [2], which is defined as the highest rate at which communication is possible with probability of error precisely equal to zero. More formally,

$$C_0 = \sup\{R \,:\, P_{e,\max}^{(n)}(R) = 0 \text{ for some } n\}. \qquad (7)$$

For $R < C_0$, we may define the reliability function $E(R)$ as being infinite. Determining the reliability function $E(R)$ (at low positive rates) and the zero-error capacity C_0 of a general channel is still an unsolved problem.

B. Reliability and zero-error capacity

In order to study the zero-error capacity of a channel, it is important to consider when two input symbols or two input sequences are confusable and when they are not. Note that two input symbols x and x' cannot be confused at the output if and only if the associated conditional distribution $W(\cdot|x)$ and $W(\cdot|x')$ have disjoint supports. Furthermore, two sequences $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{x}' = (x'_1, \ldots, x'_n)$ cannot be confused if and only if there exists at least one index i such that symbols x_i and x'_i are not confusable. For a given channel W, it is then useful to define a *confusability graph* $G(W)$ whose vertices are the elements of $\mathcal{X}$ and whose edges are the elements $(x, x') \in \mathcal{X}^2$ such that x and x' are confusable. It is then easily seen that C_0 only depends on $G(W)$. Furthermore, for any G, we can always find a channel W such that $G(W) = G$. Thus, we may equivalently speak of the zero-error capacity of a channel W or of the *capacity $C(G)$ of the graph* G if $G = G(W)$, and we will use those two notions interchangeably through the paper.

A first upper bound to C_0 was obtained by Shannon [2], who upper bounded C_0 with the zero-error capacity C_{FB} when perfect feedback is available. He could prove by means of a combinatorial argument that, if $C_0 > 0$, then

$$C_{FB} = \max_P -\log \max_y \sum_{x:W(y|x)>0} P(x). \qquad (8)$$

Given a graph G, then, the best bound to $C(G)$ is obtained by using the channel W' with $G(W') = G$ which minimizes C_{FB}. Interestingly enough, this bound can also be obtained by a rather different method that relies on bounding the reliability function $E(R)$. In particular, the so called sphere-packing bound, first derived in [6] and later rigorously proved in [3], states that $E(R) \le E_{sp}(R)$, where $E_{sp}(R)$ is defined by

$$E_{sp}(R) \ge \sup_{\rho \ge 0} [E_0(\rho) - \rho R]$$

$$E_0(\rho) = \max_P E_0(\rho, P)$$

$$E_0(\rho, P) = -\log \sum_y \left(\sum_x P(x) W(y|x)^{1/(1+\rho)} \right)^{1+\rho}.$$

The function $E_{sp}(R)$ is finite for all rates R larger than the quantity

$$R_\infty = \max_P -\log \max_y \sum_{x:W(y|x)>0} P(x), \qquad (9)$$

which implies that $E(R)$ is finite for $R > R_\infty$ and thus that $C_0 \le R_\infty$. Interestingly enough, we see that if $C_0 > 0$ then $R_\infty = C_{FB}$. This implies that in all cases of practical interest, Shannon's bound to C_0, which was first derived by means of a combinatorial method, can also be deduced from the sphere-packing bound, which is instead derived in a probabilistic setting.

A major breakthrough came with Lovász's 1979 work [4]. Given a confusability graph G, Lovász calls an *orthonormal representation* of G any set $\{u_x\}_{x \in \mathcal{X}}$ of unit norm vectors in any Hilbert space such that u_x and $u_{x'}$ are orthogonal if symbols x and x' are not confusable. We will use here the *bra-ket* notation $\langle a|b \rangle$ for the scalar product between two vectors a and b. He then defines the *value* of a representation $\{u_x\}$ as[1]

$$V(\{u_x\}) = \min_c \max_x \log \frac{1}{|\langle u_x|c \rangle|^2}, \qquad (10)$$

where the minimum is over all unit norm vectors c. The vector c that achieves the minimum above is called the *handle* of the representation. Lovász shows that any orthonormal representation satisfies $V(\{u_x\}) \ge C_0$. Optimizing over all representations, he thus gives a bound for C_0 in the form $C_0 \le \vartheta$, where

$$\vartheta = \min_{\{u_x\}} \min_c \max_x \log \frac{1}{|\langle u_x|c \rangle|^2}$$

is the so called Lovász theta function. This result is usually considered to be of a purely combinatorial nature and no probabilistic interpretation seems to have emerged up to now. It is interesting to note, however, that a possible representation for the confusability graph of a channel W can simply be constructed by taking the set of $|\mathcal{Y}|$-dimensional real valued vectors $\{\varphi_x\}$ with components $\varphi_x(y) = \sqrt{W(y|x)}$. As we will show later, the value of this representation $V(\{\varphi_x\})$ is precisely the cut-off rate of the channel, which is never smaller than C_0. Clearly, using different channels W' (with $G(W') = G(W)$), we may upper bound C_0 with the lowest of their cut-off rates. Nicely enough, it turns out that this would lead precisely to the same upper bound obtained by means of C_{FB} (or R_∞). Lovász's theta function achieves a smaller upper bound to C_0 due to the fact that it allows the components of the vectors of a representation to take on negative values. Lovász's approach seems thus to suggest bounding the zero-error capacity by considering the use of quantum-theoretic wave functions in place of classical probability distributions.

---

[1]We use a logarithmic version of the theta function so as to make its comparison with rates more straightforward.

C. Rényi's Information Radii

It is known [7] that the capacity of a classical channel can be written as an information radius according to the expression

$$C = \min_{Q} \max_{x} D(W(\cdot|x)||Q), \tag{11}$$

where $D(\cdot||\cdot)$ is the Kullback-Leibler divergence. This min-max formulation was extended by Csiszár [8] to describe the reliability function in the high rate region. Here, since we are only interested in upper bounds to $E(R)$, it is useful to consider the sphere-packing exponent $E_{sp}(R)$, for which Csiszár's min-max expression holds with full generality. The function $E_{sp}(R)$ equals the upper envelope of all the lines $E_0(\rho) - \rho R$, and an important quantity is the value $R_\rho = E_0(\rho)/\rho$ at which each of these lines meets the R axis.[2] Given two distributions Q_1 and Q_2 on the channel output $\mathcal{Y}$, define the *Rényi divergence* of order $\alpha \in (0,1)$ of Q_1 from Q_2 as

$$D_\alpha(Q_1||Q_2) = \frac{1}{\alpha - 1} \log \sum_y Q_1(y)^\alpha Q_2(y)^{1-\alpha}. \tag{12}$$

It is then shown in [8, Prop. 1] that

$$R_\rho = \min_{Q} \max_{x} D_\alpha(W(\cdot|x)||Q), \quad \alpha = 1/(1+\rho). \tag{13}$$

Using the known properties of the Rényi divergence (see [8]), we find that when $\rho \to 0$ the above expression (with $\alpha \to 1$) gives the already mentioned expression for the capacity (11), while for $\rho \to \infty$ we obtain

$$R_\infty = \min_{Q} \max_{x} -\log \sum_{y:W(y|x)>0} Q(y), \tag{14}$$

which is the dual formulation of (9).

It is evident that there is an interesting similarity between the min-max expression for R_ρ of a channel W and the value of a representation in Lovász' sense. In the next sections, we will show that this similarity is not a simple coincidence. Lovász' bound to C_0 and the sphere-packing bound to $E(R)$ are based on the very same idea and can be described in a unified way in probabilistic terms in the context of quantum information theory. By considering the extension of the sphere-packing bound to classical-quantum channels, we will show that Lovász' bound emerges naturally, in that case, as a consequence of the bound $C_0 \le R_\infty$.

Remark 1: A very nice fact, apparently not reported in the literature, is that the usual cut-off rate of a classical channel W, evaluated according to equation (13) with $\alpha = 1/2$, is precisely the value $V(\{\varphi_x\})$ of the representation $\{\varphi_x\}$ with $\varphi_x = \sqrt{W(\cdot|x)}$. In this paper, however, we will interpret Lovász's value of a representation $\{u_x\}$ in relation to the rate R_∞ of a pure-state classical-quantum channel with state vectors $|u_x\rangle$. It turns out [9] that the cut-off rate of a classical channel W precisely equals the rate R_∞ of a pure-state classical-quantum channel with state vectors $|\varphi_x\rangle$ as defined above, but the true reason for this equivalence is not yet clear.

---

[2] Here, since we also consider the true zero-error capacity C_0, we do not adopt Csiszár's notation of *channel capacity of order α*.

III. CLASSICAL-QUANTUM CHANNELS

A. Basic notions and the sphere-packing bound

We introduce here the minimal notions and results on classical-quantum channels so as to make this paper as self-contained as possible. The interested reader may refer to [10] [11] for more details.

Following [12], consider a classical-quantum channel with a finite input alphabet $\mathcal{X}$ with associated density operators S_x, $x \in \mathcal{X}$ in a finite dimensional Hilbert space[3] $\mathcal{H}$. The n-fold product channel acts in the tensor product space $\mathcal{H}^{\otimes n}$ of n copies of $\mathcal{H}$. To a codeword $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is associated the signal state $\mathbf{S_x} = S_{x_1} \otimes S_{x_2} \cdots \otimes S_{x_n}$. A block code with M codewords is a mapping from a set of M messages $\{1, \ldots, M\}$ into a set of M codewords $\mathbf{x}_1, \ldots, \mathbf{x}_M$. The rate of the code is defined as $R = \frac{\log M}{n}$.

A quantum decision scheme for such a code is a so-called POVM (see for example [11]), that is, a collection of M positive operators[4] $\{\Pi_1, \Pi_2, \ldots, \Pi_M\}$ such that $\sum \Pi_m \le \mathbb{1}$, where $\mathbb{1}$ is the identity operator. The probability that message m' is decoded when message m is transmitted is $P(m'|m) = \mathrm{Tr}\,\Pi_{m'} \mathbf{S}_{\mathbf{x}_m}$. The probability of error after sending message m is

$$P_{e|m} = 1 - \mathrm{Tr}\,(\Pi_m \mathbf{S}_{\mathbf{x}_m}). \tag{15}$$

We then define $P_{e,\max}$, $P_{e,\max}^{(n)}(R)$, C, C_0 and $E(R)$ precisely as in the classical case.

As in the classical case, we can still express C_0 as the capacity of a confusability graph (see [13] for more general results) where, in this case, two input symbols are confusable if and only if $\mathrm{Tr}(S_x S_{x'}) > 0$. In fact, if a code with M codewords satisfies $P_{e,\max} = 0$, then for each $m \ne m'$ we must have $\mathrm{Tr}(\Pi_m \mathbf{S}_{\mathbf{x}_m}) = 1$ and $\mathrm{Tr}(\Pi_m \mathbf{S}_{\mathbf{x}_{m'}}) = 0$. This is possible if and only if the signals $\mathbf{S}_{\mathbf{x}_m}$ and $\mathbf{S}_{\mathbf{x}_{m'}}$ are orthogonal, that is $\mathrm{Tr}(\mathbf{S}_{\mathbf{x}_m} \mathbf{S}_{\mathbf{x}_{m'}}) = 0$. But, using the property that $\mathrm{Tr}((A \otimes B)(C \otimes D)) = \mathrm{Tr}(AC)\,\mathrm{Tr}(BD)$, we have

$$\mathrm{Tr}(\mathbf{S}_{\mathbf{x}_m} \mathbf{S}_{\mathbf{x}_{m'}}) = \prod_{i=1}^{n} \mathrm{Tr}(S_{x_{m,i}} S_{x_{m',i}}). \tag{16}$$

This implies that $\mathrm{Tr}(S_{x_{m,i}} S_{x_{m',i}}) = 0$ for at least one value of i. Thus, evaluating the zero-error capacity in the classical-quantum setting amounts to evaluating the capacity of a graph as defined in the previous section. In this sense, there is no difference between classical and classical-quantum channels and, given a graph G, we can interpret the capacity $C(G)$ as either the zero error capacity C_0 of a classical or of a classical-quantum channel with that confusability graph. (For recent results on the zero-error communication via general quantum channels see [14] and references therein).

For classical-quantum channels, bounds to the reliability function $E(R)$ have been developed which partially match

---

[3] The S_x can thus be represented as positive semi-definite Hermitian matrices with unit trace.

[4] The operators Π_m can thus be represented as positive semi-definite matrices. The notation $\sum_m \Pi_m \le \mathbb{1}$ simply means that $\mathbb{1} - \sum_m \Pi_m$ is positive semidefinite. Note that, by construction, all the eigenvalues of each operator Π_m must be in the interval $[0, 1]$.

those of the classical case. Lower bounds to the reliability function were obtained in [15] and [12], while upper bounds have remained relatively unexplored until recently. For general $R > 0$, the first upper bound to $E(R)$ was obtained in [5] as an extension of the classical sphere-packing bound of [3]. The bound can be stated as follows.

Theorem 1 (Sphere Packing Bound [5][9]): For all positive rates R and all positive $\varepsilon < R$,

$$E(R) \leq E_{sp}(R - \varepsilon), \tag{17}$$

where $E_{sp}(R)$ is defined by the relations

$$E_{sp}(R) = \sup_{\rho \geq 0} [E_0(\rho) - \rho R] \tag{18}$$

$$E_0(\rho) = \max_P E_0(\rho, P) \tag{19}$$

$$E_0(\rho, P) = -\log \operatorname{Tr} \left(\sum_x P(x) S_x^{1/(1+\rho)} \right)^{1+\rho}. \tag{20}$$

B. Quantum Rényi's Information Radii

We now extend Csiszár's result to give a characterization of the sphere packing bound for classical-quantum channels in terms of Rényi's information measures. Given two density operators F_1 and F_2 in $\mathcal{H}$, and $\alpha \in (0, 1)$, define the Rényi divergence of order α of F_1 from F_2 as

$$D_\alpha(F_1 \| F_2) = \frac{1}{\alpha - 1} \log \operatorname{Tr} F_1^\alpha F_2^{1-\alpha}. \tag{21}$$

As in the classical case, for $\rho > 0$, let then

$$R_\rho = E_0(\rho)/\rho. \tag{22}$$

Then we have the following result.

Theorem 2: For a classical-quantum channel with states S_x, $x \in \mathcal{X}$ and $\rho > 0$, the rate R_ρ defined above satisfies

$$R_\rho = \min_F \max_x D_\alpha(S_x \| F), \quad \alpha = 1/(1 + \rho). \tag{23}$$

Proof: Setting $\alpha = 1/(1 + \rho)$, we can write

$$R_\rho = \max_P \frac{1}{\alpha - 1} \log \left[\operatorname{Tr} \left(\sum_x P(x) S_x^\alpha \right)^{1/\alpha} \right]^\alpha \tag{24}$$

and, defining $A(\alpha, P) = \sum_x P(x) S_x^\alpha$, we can write

$$R_\rho = \max_P \frac{1}{\alpha - 1} \log \| A(\alpha, P) \|_{1/\alpha}, \tag{25}$$

where $\| \cdot \|_r$ is the Schatten r-norm. From the Hölder inequality we know that, for any positive operators A and B, we have

$$\| A \|_{1/\alpha} \| B \|_{1/(1-\alpha)} \geq \operatorname{Tr}(AB) \tag{26}$$

with equality if an only if $B = \gamma A^{1-1/\alpha}$ for some scalar coefficient γ. Thus we can write

$$\| A \|_{1/\alpha} = \max_{\| B \|_{1/(1-\alpha)} \leq 1} \operatorname{Tr}(AB), \tag{27}$$

where B runs over positive operators in the unit ball in the $(1/(1-\alpha))$-norm. Using this expression for the Schatten norm we obtain

$$R_\rho = \max_P \frac{1}{\alpha - 1} \log \max_{\| B \|_{1/(1-\alpha)} \leq 1} \operatorname{Tr}(A(\alpha, P)B) \tag{28}$$

$$= \frac{1}{\alpha - 1} \log \min_P \max_{\| B \|_{1/(1-\alpha)} \leq 1} \operatorname{Tr} \left(\sum_x P(x) S_x^\alpha B \right). \tag{29}$$

In the last expression, the minimum and the maximum are both taken over convex sets and the objective function is linear both in P and B. Thus, we can interchange the order of maximization and minimization to get

$$R_\rho = \frac{1}{\alpha - 1} \log \max_{\| B \|_{1/(1-\alpha)} \leq 1} \min_P \sum_x P(x) \operatorname{Tr}(S_x^\alpha B) \tag{30}$$

$$= \frac{1}{\alpha - 1} \log \max_{\| B \|_{1/(1-\alpha)} \leq 1} \min_x \operatorname{Tr}(S_x^\alpha B). \tag{31}$$

Now, we note that the maximum over B can always be achieved by a positive operator, since all the S_x^α are positive operators. Thus, we can change the dummy variable B with $F = B^{1/(1-\alpha)}$, where F is now a positive operator constrained to satisfy $\| F \|_1 \leq 1$, that is, it is a density operator. Using F, we get

$$R_\rho = \frac{1}{\alpha - 1} \log \max_F \min_x \operatorname{Tr}(S_x^\alpha F^{1-\alpha}) \tag{32}$$

$$= \min_F \max_x \frac{1}{\alpha - 1} \log \operatorname{Tr}(S_x^\alpha F^{1-\alpha}) \tag{33}$$

$$= \min_F \max_x D_\alpha(S_x \| F). \tag{34}$$

where F now runs over all density operators. ∎

It is obvious that, if all operators S_x commute, which means that the channel is classical, than the optimal F is diagonal in the same basis where the S_x are, and we thus recover Csiszár's expression for the classical case. Furthermore, for $\rho \to 0$ (that is, $\alpha \to 1$) we obtain the expression of the capacity as an information radius already established for classical-quantum channels [16]. When $\rho = 1$ (that is, $\alpha = 1/2$) then, we obtain an alternative expression for the so called quantum cut-off rate [17]. The most important case in our context, however, is the case when $\rho \to \infty$ (that is, $\alpha \to 0$). Taking the limit in Theorem 2, letting S_x^0 be the projector in the subspace of S_x, we obtain

$$R_\infty = \min_F \max_x \log \frac{1}{\operatorname{Tr}(S_x^0 F)}, \tag{35}$$

where the minimum is again over all density operators F. Note that the argument of the min-max in (35) coincides with $D_{\min}(S_x \| F)$ according to the definition of $D_{\min}$ introduced in [18].

The analogy with the Lovász theta function becomes evident if we consider a special case of (35). Assume that the states S_x are pure and set $S_x = |u_x\rangle\langle u_x|$. Consider for a moment the search for the optimum F when restricted to rank-one operators, that is $F = |f\rangle\langle f|$. We see that in this case we can

write $\mathrm{Tr}(S_x^0 F) = |\langle u_x | f \rangle|^2$. When searching over all possible F, we thus find that for this channel we have

$$R_\infty \leq V(\{u_x\}). \tag{36}$$

Hence, we see that Lovász's bound $C_0 \leq V(\{u_x\})$ can be deduced as a consequence of $C_0 \leq R_\infty$. For a given graph G, one may want to bound $C(G)$ with the smallest R_∞ over all channels with confusability graph G. This is discussed in the next section section.

IV. Sphere Packing and the Lovász Theta Function

For a given confusability graph G, inspired by (35), we define a *representation* of G any set of projectors $\{U_x\}$ such that $U_x U_{x'} = 0$ if symbols x and x' cannot be confused. Furthermore, we introduce an alternative definition of *value*

$$V_{sp}(\{U_x\}) = \min_F \max_x \log \frac{1}{\mathrm{Tr}\,(U_x F)}, \tag{37}$$

where the minimum is over all density operators F. The optimal F will be called again the handle of the representation. We can then finally define the quantity.

$$\vartheta_{sp} = \min_{\{U_x\}} \min_F \max_x \log \frac{1}{\mathrm{Tr}\,(U_x F)}, \tag{38}$$

where $\{U_x\}$ runs over all representations of the graph G. We then have the following result.

Theorem 3: For any graph, we have

$$C(G) \leq \vartheta_{sp} \leq \vartheta. \tag{39}$$

Proof: The fact that $\vartheta_{sp} \leq \vartheta$ is obvious, since Lovász's ϑ is obtained by restricting the minimization in the definition of ϑ_{sp} to rank-one projectors $U_x = |u_x\rangle\langle u_x|$ and handle $F = |f\rangle\langle f|$. That $C_0 \leq \vartheta_{sp}$ should be clear in light of the above discussion on the bound $E(R) \leq E_{sp}(R)$. It is instructive, however, to present a self-contained proof along the same argument used by Lovász.

Consider an optimal representation $\{U_x\}$ and, to a sequence of symbols $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, associate the operator (projector) $\mathbf{U_x} = U_{x_1} \otimes U_{x_2} \cdots \otimes U_{x_n}$. Consider then a zero-error code with M codewords of length n, $\mathbf{x}_1, \ldots, \mathbf{x}_M$, and their associated projectors $\mathbf{U_{x_1}}, \ldots, \mathbf{U_{x_M}}$. Then, as proved before, for $m \neq m'$ we have $\mathrm{Tr}(\mathbf{U_{x_m}} \mathbf{U_{x_{m'}}}) = 0$. Hence, since the states $\{\mathbf{U_{x_m}}\}$ are orthogonal projectors, we clearly have

$$\sum_{m=1}^M \mathbf{U_{x_m}} \leq \mathbb{1}, \tag{40}$$

where $\mathbb{1}$ is the identity operator. Consider now the state $\mathbf{F} = F^{\otimes n}$ where F is the handle of the representation $\{U_x\}$. Note that, for each m, we have

$$\mathrm{Tr}(\mathbf{U_{x_m}} \mathbf{F}) = \prod_{i=1}^n \mathrm{Tr}(U_{x_{m,i}} F)$$
$$\geq e^{-n\vartheta_{sp}}.$$

So, using (40), we deduce that

$$1 \quad \geq \quad \sum_{m=1}^M \mathrm{Tr}(\mathbf{U_{x_m}} \mathbf{F}) \tag{41}$$
$$\geq \quad M e^{-n\vartheta_{sp}}. \tag{42}$$

and hence that $M \leq e^{n\vartheta_{sp}}$. $\blacksquare$

Note added in the final version: Schrijver [19] has observed that Lemma 4 and Corollary 1 in [4] apply *mutatis mutandis* with our definitions of representation and of ϑ_{sp}. Then, Theorem 5 in [4] implies $\vartheta \leq \vartheta_{sp}$, proving that $\vartheta_{sp} = \vartheta$. This conclusively shows that the sphere-packing bound, when applied to classical-quantum channels, gives precisely Lovász' bound to C_0 and that pure state channels suffice for this purpose. This also implies that for Lovász's optimal representations there is always a rank-one minimizing F in (35). It is worth pointing out that this is not true in general and that strict inequality holds in (36) for some channels.

References

[1] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Sys. Tech. Journal*, vol. 27, pp. 379–423,623–656, 1948.

[2] ——, "The Zero-Error Capacity of a Noisy Channel," *IRE Trans. Inform. Theory*, vol. IT-2, pp. 8–19, 1956.

[3] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower Bounds to Error Probability for Coding in Discrete Memoryless Channels. I," *Information and Control*, vol. 10, pp. 65–103, 1967.

[4] L. Lovasz, "On the Shannon Capacity of a Graph," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 1–7, 1979.

[5] M. Dalai, "Sphere Packing Bound for Quantum Channels," in *Proc. IEEE Intern. Symp. Inform. Theory*, 2012, pp. 160 – 164.

[6] R. M. Fano, *Transmission of Information: A Statistical Theory of Communication.* Wiley, New York, 1961.

[7] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* Academic Press, 1981.

[8] I. Csiszár, "Generalized Cutoff Rates and Rényi's Information Measures," *IEEE Trans. Inform. Theory*, vol. 41, no. 1, pp. 26 –34, Jan. 1995.

[9] M. Dalai, "Lower Bounds to the Probability of Error for Classical and Classical-Quantum Channels," *arXiv:1201.5411v3 [cs.IT]*.

[10] M. Hayashi, *Quantum Information: An Introduction.* Springer, 2006.

[11] M. M. Wilde, *From Classical to Quantum Shannon Theory.* arXiv:1106.1445 [quant-ph], 2012.

[12] A. S. Holevo, "Reliability Function of General Classical-Quantum Channel," *Information Theory, IEEE Transactions on*, vol. 46, no. 6, pp. 2256 –2261, Sep. 2000.

[13] R. Medeiros, R. Alleaume, G. Cohen, and F. De Assis, "Zero-Error Capacity of Quantum Channels and Noiseless Subsystems," in *Proc. Int. Telecomm. Symposium.* Fortaleza, Brazil, 2006, pp. 900–905.

[14] R. Duan, S. Severini, and A. Winter, "Zero-error Communication via Quantum Channels, Noncommutative Graphs, and a Quantum Lovász Theta Function," *IEEE Trans. Inform. Theory*, vol. 59, no. 2, pp. 1164–1174, 2013.

[15] M. V. Burnashev and A. S. Holevo, "On Reliability Function of Quantum Communication Channel," *Probl. Peredachi Inform.*, vol. 34, no. 2, pp. 1–13, 1998.

[16] M. Hayashi and H. Nagaoka, "General Formulas for Capacity of Classical-Quantum Channels," *Information Theory, IEEE Transactions on*, vol. 49, no. 7, pp. 1753 – 1768, july 2003.

[17] M. Ban, K. Kurokawa, and O. Hirota, "Cut-Off Rate for Quantum Communication Channels with Entangled Measurement," *Quantum and Semiclassical Optics: Journal of the European Optical Society Part B*, vol. 10, no. 2, p. L7, 1998.

[18] N. Datta, "Min- and Max-Relative Entropies and a New Entanglement Monotone," *Information Theory, IEEE Transactions on*, vol. 55, no. 6, pp. 2816–2826, 2009.

[19] A. Schrijver, Personal communication, May 2013.