An "Umbrella" Bound of the Lovász-Gallager Type

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

Abstract—We propose a novel approach for bounding the probability of error of discrete memoryless channels with a zero-error capacity based on a combination of Lovász' and Gallager's ideas. The obtained bounds are expressed in terms of a function $\vartheta(\rho)$, introduced here, that varies from the cut-off rate of the channel to the Lovázs theta function as ρ varies from 1 to ∞ and which is intimately related to Gallager's expurgated coefficient. The obtained bound to the reliability function, though loose in its present form, is finite for all rates larger than the Lovász theta function.

I. INTRODUCTION

One of the most intriguing topic in coding theory is the problem of bounding the probability of error of optimal codes at low rates. While at high rates the asymptotic behaviour of the probability of error for optimal codes is now very well understood, very little is known in the low rate region. 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. Both determining $E(R)$ for small R and even determining C_0 is an unsolved problem and only upper and lower bounds for these quantities are known. Lovász gave an important improvement in upper bounding C_0 by means of his ϑ function, thus enlarging the range of values over which $E(R)$ is known to be finite. However, Lovász's result was never exploited to find actual bounds to $E(R)$ for rates R immediately above ϑ.

In this paper we propose a first attempt to bound the probability of error at rates $R > \vartheta$ by combining Lovász's method with the idea used by Gallager in the development of his expurgated bound. Even if the proposed approach does not lead yet to good bounds to the reliability function, we believe it sheds some light on this relatively unexplored topic.

II. BASIC NOTIONS

A. Reliability of DMCs

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 $\mathcal{X} = \{1, 2, \ldots, K\}$ and $\mathcal{Y} = \{1, 2, \ldots, J\}$ are finite sets. For a sequence $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ and a sequence $\boldsymbol{y} = (y_1, y_2, \ldots, y_n) \in \mathcal{Y}^n$, the probability of observing $\boldsymbol{y}$ at the output of the channel given $\boldsymbol{x}$ at the input is

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

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 $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\}$ 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 $\boldsymbol{x}_m$ through the channel. An output sequence $\boldsymbol{y}$ 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

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

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 exist such that $P_{e,\max}^{(n)}(R) \to 0$ as $n \to \infty$ for all rates smaller than a constant C, called *channel capacity*. 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 $C_0 < R < C$, 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

$$E(R) = \limsup_{n \to \infty} -\frac{1}{n} \log P_{e,\max}^{(n)}(R). \tag{5}$$

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,

$$C_0 = \sup\{R : P_{e,\max}^{(n)}(R) = 0 \text{ for some } n\}. \tag{6}$$

For $R < C_0$, we may define the reliability function $E(R)$ as being infinite. Note that $C_0 > 0$ if and only if there are at least

two input symbols x and x' which are not confusable at the output, meaning that $W(y|x)W(y|x')$ is zero for all values of y. 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.

One of the most famous results in this direction is Lovász's upper bound to C_0. Lovász proves that C_0 is upper bounded by a quantity ϑ defined as

$$\vartheta = \min_{\{\boldsymbol{u}_x\}} \min_{\boldsymbol{c}} \max_x \log \frac{1}{|\boldsymbol{u}_x^\dagger \boldsymbol{c}|^2}$$

where $\{\boldsymbol{u}_x\}_{x \in \mathcal{X}}$ runs over all sets of unit norm vectors in any Hilbert space such that $\boldsymbol{u}_x$ and $\boldsymbol{u}_{x'}$ are orthogonal if symbols x and x' are not confusable and $\boldsymbol{c}$ runs over all unit norm vectors. Here, $\cdot^\dagger$ denotes conjugate transpose and $\boldsymbol{u}_x^\dagger \boldsymbol{c}$ is the scalar product between $\boldsymbol{u}_x$ and $\boldsymbol{c}$.

B. Bhattacharyya distances and scalar products

Here, we briefly recall some important connections between the reliability function $E(R)$ and the Bhattacharyya distance between codewords. This connection is of great importance since the Bhattacharyya distance between distributions is related to a scalar product between unit norm vectors in a Hilbert space. It is this property that creates an underlying common substrate for Lovász's approach and for bounding the reliability function.

For a generic input symbol x, consider the unit norm $|\mathcal{Y}|$-dimensional column vector $\boldsymbol{\psi}_x$ with components $\psi_x(y) = \sqrt{W(y|x)}$. We call this the *state vector* of input symbol x, in analogy with the input signals of pure-state classical-quantum channels (see comment at the end of Section IV). In the same way, for an input sequence $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, consider the unit norm $|\mathcal{Y}|^n$-dimensional column vector $\boldsymbol{\Psi}_{\boldsymbol{x}}$ whose components are the values $\sqrt{W^{(n)}(\boldsymbol{y}|\boldsymbol{x})}$, that is, $\boldsymbol{\Psi}_{\boldsymbol{x}}$ is simply the element-wise square root of the conditional output distribution given the input sequence $\boldsymbol{x}$. Then, since the channel is memoryless, we can write

$$\boldsymbol{\Psi}_{\boldsymbol{x}} = \boldsymbol{\psi}_{x_1} \otimes \boldsymbol{\psi}_{x_2} \otimes \cdots \boldsymbol{\psi}_{x_n} \tag{7}$$

where $\otimes$ is the Kronecker product. Let for ease of notation $\boldsymbol{\Psi}_m$ be the state vector of the codeword $\boldsymbol{x}_m$; then we can represent our code $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\}$ by means of their associated state vectors $\{\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \ldots, \boldsymbol{\Psi}_M\}$. Since all square roots are taken positive, note that our channel has a positive zero-error capacity if and only if there are at least two state vectors $\boldsymbol{\psi}_x$, $\boldsymbol{\psi}_{x'}$ such that $\boldsymbol{\psi}_x^\dagger \boldsymbol{\psi}_{x'} = 0$. This implies that codes can be built such that $\boldsymbol{\Psi}_m^\dagger \boldsymbol{\Psi}_{m'} = 0$ for some m, m', that is, the two codewords m and m' cannot be confused at the output. However, the scalar product $\boldsymbol{\Psi}_m^\dagger \boldsymbol{\Psi}_{m'}$ plays a more general role since it is related to the so called Bhattacharyya distance between the two codewords m and m'. In particular, in a binary hypothesis testing between codeword m and m', an extension of the Chernoff Bound allows to assert that the minimum error probability asymptotically satisfies [3]

$$P_e \doteq \min_{0 \le s \le 1} \sum_{\boldsymbol{y}} W^{(n)}(\boldsymbol{y}|\boldsymbol{x}_m)^{1-s} W^{(n)}(\boldsymbol{y}|\boldsymbol{x}_{m'})^s \tag{8}$$

where $\doteq$ means equivalence to the first order in the exponent. For $s = 1/2$, the sum above obviously equals $\boldsymbol{\Psi}_m^\dagger \boldsymbol{\Psi}_{m'}$. It is easily shown that the minimum above is always between $(\boldsymbol{\Psi}_m^\dagger \boldsymbol{\Psi}_{m'})^2$ and $\boldsymbol{\Psi}_m^\dagger \boldsymbol{\Psi}_{m'}$, and it equals the latter for a class of channels, called pairwise reversible channels, that have some symmetry with respect to the input symbols[1] [3]. Obviously, for a given code, the probability of error $P_{e,\max}$ is lower bounded by the probability of error in each binary hypothesis test between two codewords. Hence, we find that $P_{e,\max}$ asymptotically satisfies

$$-\frac{1}{n} \log P_{e,\max} \le -\frac{2}{n} \log \max_{m \ne m'} (\boldsymbol{\Psi}_m^\dagger \boldsymbol{\Psi}_{m'}) + o(n), \tag{9}$$

where the coefficient 2 can be removed if the channel is pairwise reversible. It is thus obvious that it is possible to upper bound $E(R)$ by lower bounding the quantity

$$\gamma = \max_{m \ne m'} \boldsymbol{\Psi}_m^\dagger \boldsymbol{\Psi}_{m'} \tag{10}$$

Lovász's work aims at finding a value ϑ as small as possible that allows to conclude that, for a set of $M = e^{nR} > e^{n\vartheta}$ codewords, γ cannot be zero, and thus at least two codewords are confusable. Here, instead, we want something more, that is, finding a lower bound on γ for each code with rate $R > \vartheta$ so as to deduce an upper bound to $E(R)$ for all $R > \vartheta$.

III. AN "UMBRELLA" BOUND

Consider the scalar products between the channel state vectors $\boldsymbol{\psi}_x^\dagger \boldsymbol{\psi}_{x'} \ge 0$. For a fixed $\rho \ge 1$, consider then a set of "tilted" state vectors, that is, unit norm vectors $\{\tilde{\boldsymbol{\psi}}_x\}$ in any Hilbert space such that $|\tilde{\boldsymbol{\psi}}_x^\dagger \tilde{\boldsymbol{\psi}}_{x'}| \le (\boldsymbol{\psi}_x^\dagger \boldsymbol{\psi}_{x'})^{1/\rho}$. We call such a set of vectors $\{\tilde{\boldsymbol{\psi}}_x\}$ an *orthonormal representation of degree ρ* of our channel, and call $\Gamma(\rho)$ the set of all possible such representations

$$\Gamma(\rho) = \left\{ \{\tilde{\boldsymbol{\psi}}_x\} : |\tilde{\boldsymbol{\psi}}_x^\dagger \tilde{\boldsymbol{\psi}}_{x'}| \le (\boldsymbol{\psi}_x^\dagger \boldsymbol{\psi}_{x'})^{1/\rho} \right\}, \quad \rho \ge 1. \tag{11}$$

Observe that $\Gamma(\rho)$ is non-empty since the original $\boldsymbol{\psi}_k$ vectors satisfy the constraints. The *value* of an orthonormal representation is the quantity

$$V(\{\tilde{\boldsymbol{\psi}}_x\}) = \min_{\boldsymbol{f}} \max_x \log \frac{1}{|\tilde{\boldsymbol{\psi}}_x^\dagger \boldsymbol{f}|^2}, \tag{12}$$

where the minimum is over all unit norm vectors $\boldsymbol{f}$. The optimal choice of the vector $\boldsymbol{f}$ is called, with Lovász, the handle of the representation. We call it $\boldsymbol{f}$ to point out that this vector plays essentially the same role as the auxiliary output distribution f used in the sphere-packing bound of [4]. Due to space limitation, we cannot discuss this detail here; see the comment at the end of Section IV.

Call now $\vartheta(\rho)$ the minimum value over all representations of degree ρ,

$$\vartheta(\rho) = \min_{\{\tilde{\boldsymbol{\psi}}_x\} \in \Gamma(\rho)} V(\{\tilde{\boldsymbol{\psi}}_x\}). \tag{13}$$

---

[1]Somehow tautologically, pairwise reversible channels are those for which the minimum is achieved for $s = 1/2$.

We have the following result.

Theorem 1: For any code of block-length n with M codewords and any $\rho \geq 1$ we have

$$\max_m \sum_{m' \neq m} (\boldsymbol{\Psi}_m^\dagger \boldsymbol{\Psi}_{m'}) \geq \frac{\left(Me^{-n\vartheta(\rho)} - 1\right)^\rho}{(M-1)^{\rho-1}} \qquad (14)$$

Corollary 1: For the reliability function of a general DMC we have the bound

$$E(R) \leq 2\rho\,\vartheta(\rho), \qquad R > \vartheta(\rho), \qquad (15)$$

where the coefficient 2 can be removed if the channel is pairwise reversible.

Proof: For an input sequence $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ call, in analogy with (7), $\tilde{\boldsymbol{\Psi}}_{\boldsymbol{x}} = \tilde{\psi}_{x_1} \otimes \tilde{\psi}_{x_2} \otimes \cdots \tilde{\psi}_{x_n}$. Observe first that, for any two input sequences $\boldsymbol{x}$ and $\boldsymbol{x}'$, we have

$$|\tilde{\boldsymbol{\Psi}}_{\boldsymbol{x}}^\dagger \tilde{\boldsymbol{\Psi}}_{\boldsymbol{x}'}| = \prod_{i=1}^n |\tilde{\psi}_{x_i}^\dagger \tilde{\psi}_{x_i'}| \qquad (16)$$

$$\leq \prod_{i=1}^n (\psi_{x_i}^\dagger \psi_{x_i'})^{1/\rho} \qquad (17)$$

$$= (\tilde{\boldsymbol{\Psi}}_{\boldsymbol{x}}^\dagger \tilde{\boldsymbol{\Psi}}_{\boldsymbol{x}'})^{1/\rho} \qquad (18)$$

Furthermore, note that, for an optimal representation of degree ρ with handle $\boldsymbol{f}$, we have $|\tilde{\psi}_x^\dagger \boldsymbol{f}|^2 \geq e^{-\vartheta(\rho)}$, $\forall x$. Set now $\boldsymbol{F} = \boldsymbol{f}^{\otimes N}$. We then have

$$|\tilde{\boldsymbol{\Psi}}_{\boldsymbol{x}}^\dagger \boldsymbol{F}|^2 = \prod_{i=1}^n |\tilde{\psi}_{x_i}^\dagger \boldsymbol{f}|^2 \qquad (19)$$

$$\geq e^{-n\vartheta(\rho)}. \qquad (20)$$

Let us first check how Lovász's bound is obtained. Lovász's approach is to bound the number M of codewords with orthogonal state vectors, using the property that if $\tilde{\boldsymbol{\Psi}}_1, \tilde{\boldsymbol{\Psi}}_2, \ldots \tilde{\boldsymbol{\Psi}}_M$ form a set of orthonormal vectors, then

$$1 = \|\boldsymbol{F}\|_2^2 \qquad (21)$$

$$\geq \sum_m |\tilde{\boldsymbol{\Psi}}_m^\dagger \boldsymbol{F}|^2 \qquad (22)$$

$$\geq Me^{-n\vartheta(\rho)}. \qquad (23)$$

Hence, if $M > e^{n\vartheta(\rho)}$, there are at least two non-orthogonal vectors in the set, say $|\tilde{\boldsymbol{\Psi}}_m^\dagger \tilde{\boldsymbol{\Psi}}_{m'}|^2 > 0$. But this implies that $(\boldsymbol{\Psi}_m^\dagger \boldsymbol{\Psi}_{m'})^2 \geq |\tilde{\boldsymbol{\Psi}}_m^\dagger \tilde{\boldsymbol{\Psi}}_{m'}|^{2\rho} > 0$. Hence, if $R > \vartheta(\rho)$, no zero-error code can exist. We still have the freedom in the choice of ρ and it is obvious that larger values of ρ can only give better results. Hence, it is preferable to simply work in the limit of $\rho \to \infty$ and thus build the representation $\tilde{\psi}_1, \tilde{\psi}_2, \ldots, \tilde{\psi}_K$ under the only constraint that $|\tilde{\psi}_x^\dagger \tilde{\psi}_{x'}| = 0$ whenever $|\psi_x^\dagger \psi_{x'}| = 0$. This gives precisely Lovász' result.

Now, instead of bounding R under the hypothesis of zero-error communication, we want to bound the probability of error for a given $R > \vartheta(\rho)$. Considering the tilted state vectors of the code, we can rewrite equation (20) as

$$|\tilde{\boldsymbol{\Psi}}_m^\dagger \boldsymbol{F}|^2 = \boldsymbol{F}^\dagger \left(\tilde{\boldsymbol{\Psi}}_m \tilde{\boldsymbol{\Psi}}_m^\dagger\right) \boldsymbol{F} \qquad (24)$$

$$\geq e^{-n\vartheta(\rho)}. \qquad (25)$$

The second expression above has the benefit of easily allowing averaging this expression over different codewords. So, we can average this expression over all m and, defining the matrix $\boldsymbol{\Phi} = \left(\tilde{\boldsymbol{\Psi}}_1, \ldots, \tilde{\boldsymbol{\Psi}}_M\right)/\sqrt{M}$, we get

$$\boldsymbol{F}^\dagger \boldsymbol{\Phi}\boldsymbol{\Phi}^\dagger \boldsymbol{F} \geq e^{-n\vartheta(\rho)}. \qquad (26)$$

Since $\boldsymbol{F}$ is a unit norm vector, this implies that the matrix $\boldsymbol{\Phi}\boldsymbol{\Phi}^\dagger$ has at least one eigenvalue larger than or equal to $e^{-n\vartheta(\rho)}$. This in turn implies that also the matrix $\boldsymbol{\Phi}^\dagger \boldsymbol{\Phi}$ has itself an eigenvalue larger than or equal to $e^{-n\vartheta(\rho)}$, that is

$$\lambda_{\max}\left(\boldsymbol{\Phi}^\dagger \boldsymbol{\Phi}\right) \geq e^{-n\vartheta(\rho)}. \qquad (27)$$

It is known that for a given matrix A with elements $\{A_{i,j}\}$, the following inequality holds

$$\lambda_{\max}(A) \leq \max_i \sum_j |A_{i,j}|. \qquad (28)$$

Using this inequality with $A = \boldsymbol{\Phi}^\dagger \boldsymbol{\Phi}$, since $A_{i,j} = \tilde{\boldsymbol{\Psi}}_i^\dagger \tilde{\boldsymbol{\Psi}}_j/M$, we get

$$e^{-n\vartheta(\rho)} \leq \max_i \sum_j \frac{|\tilde{\boldsymbol{\Psi}}_i^\dagger \tilde{\boldsymbol{\Psi}}_j|}{M} \qquad (29)$$

$$= \frac{1}{M}\left(1 + \max_i \sum_{j \neq i} |\tilde{\boldsymbol{\Psi}}_i^\dagger \tilde{\boldsymbol{\Psi}}_j|\right). \qquad (30)$$

We then deduce

$$\frac{Me^{-n\vartheta(\rho)} - 1}{M - 1} \leq \max_i \frac{1}{M-1} \sum_{j \neq i} |\tilde{\boldsymbol{\Psi}}_i^\dagger \tilde{\boldsymbol{\Psi}}_j| \qquad (31)$$

$$\leq \max_i \frac{1}{M-1} \sum_{j \neq i} \left(\boldsymbol{\Psi}_i^\dagger \boldsymbol{\Psi}_j\right)^{1/\rho} \qquad (32)$$

$$\leq \max_i \left(\frac{1}{M-1} \sum_{j \neq i} \boldsymbol{\Psi}_i^\dagger \boldsymbol{\Psi}_j\right)^{1/\rho}, \qquad (33)$$

where the last step is due to the Jensen inequality, since $\rho \geq 1$. Extracting the sum from this inequality we obtain the inequality stated in the theorem.

To prove the corollary, simply note that

$$\max_{i \neq j} \boldsymbol{\Psi}_i^\dagger \boldsymbol{\Psi}_j \geq \max_i \frac{1}{M-1} \sum_{j \neq i} |\boldsymbol{\Psi}_i^\dagger \boldsymbol{\Psi}_j| \qquad (34)$$

$$\geq \left(\frac{Me^{-n\vartheta(\rho)} - 1}{M - 1}\right)^\rho \qquad (35)$$

$$\geq \left(e^{-n\vartheta(\rho)} - e^{-nR}\right)^\rho. \qquad (36)$$

The bound is trivial if $R \leq \vartheta(\rho)$. For $R > \vartheta(\rho)$, instead, the second term in the parenthesis decreases exponentially faster than the first, which leads us to the conclusion that

$$-\frac{1}{n} \log \max_{m \neq m'} \boldsymbol{\Psi}_m^\dagger \boldsymbol{\Psi}_{m'} \leq \rho\vartheta(\rho) + o(1). \qquad (37)$$

The bounds in terms of $E(R)$ are then obtained by simply taking the limit $n \to \infty$ and using the bound (9). ∎

We close this section with a comment on the computation of the function $\vartheta(\rho)$. There is no essential difference with respect to the evaluation of the Lovász theta function. The optimal representation $\{\tilde{\boldsymbol{\psi}}_x\}$ for any fixed ρ, can be obtained by solving a semidefinite optimization problem. If we consider the $(K+1) \times (K+1)$ Gram matrix

$$G = [\tilde{\boldsymbol{\psi}}_1, \ldots, \tilde{\boldsymbol{\psi}}_K, \boldsymbol{f}]^\dagger [\tilde{\boldsymbol{\psi}}_1, \ldots, \tilde{\boldsymbol{\psi}}_K, \boldsymbol{f}] \qquad (38)$$

we note that finding the optimal representation amounts to solving the problem

$$\begin{aligned}
\max \quad & V \\
\text{s.t.} \quad G(k, K+1) \geq & V, \quad \forall k \leq K \\
G(k, k) = & 1, \quad \forall k \\
G(k, i) \leq & \boldsymbol{\psi}_k^\dagger \boldsymbol{\psi}_i^{1/\rho} \\
& 1 \leq k < K,\ k < i \leq K \\
G \text{ is } & \text{positive semidefinite}
\end{aligned} \qquad (39)$$

The solution to this problem is $V = \vartheta(\rho)$ and both the representation vectors $\{\tilde{\boldsymbol{\psi}}_x\}$ and the handle $\boldsymbol{f}$ can be obtained by means of the spectral decomposition of the optimal G found.

IV. CONNECTIONS WITH OTHER RESULTS IN CHANNEL THEORY

A first important comment about $\vartheta(\rho)$ concerns the result obtained for $\rho = 1$; the value $\vartheta(1)$ is in fact simply the cut-off rate of the channel. Indeed, for $\rho = 1$, we can without loss of generality use the obvious representation $\tilde{\boldsymbol{\psi}}_x = \boldsymbol{\psi}_x, \forall x$, since any different optimal representation will simply be a rotation of this (or an equivalent description in a space with a different dimension). In this case, all the components of all the vectors $\{\tilde{\boldsymbol{\psi}}_x\}$ are non-negative and this easily implies that the optimal f can as well be chosen with non-negative components, since changing a supposedly negative component of $\boldsymbol{f}$ to its absolute value can only improve the result. Thus, $\boldsymbol{f}$ can be written as the square root of a probability distribution Q on $\mathcal{Y}$, and we have

$$\begin{aligned}
\vartheta(1) &= \min_{\boldsymbol{f}} \max_x \log \frac{1}{|\boldsymbol{\psi}_x^\dagger \boldsymbol{f}|^2} \qquad &(40) \\
&= \min_Q \max_x \left(-2 \log \sum_y \sqrt{W(y|x)Q(y)} \right) &(41)
\end{aligned}$$

where the minimum is now over all probability distributions Q. As observed by Csiszár [6, Proposition 1, with $\alpha = 1/2$], this expression equals the cut-off rate R_1 of the channel defined as

$$\begin{aligned}
R_1 &= \max_P -\log \sum_{x,x'} P(x)P(x') \left(\sum_y \sqrt{W(y|x)W(y|x')} \right) \\
&= \max_P -\log \sum_{x,x'} P(x)P(x') \boldsymbol{\psi}_x^\dagger \boldsymbol{\psi}_{x'}
\end{aligned}$$

Another important characteristic of the function $\vartheta(\rho)$ is observed in the limit $\rho \to \infty$. In the limit, the only constraint on the representations is that $|\tilde{\boldsymbol{\psi}}_x^\dagger \tilde{\boldsymbol{\psi}}_{x'}| = 0$ whenever

$|\boldsymbol{\psi}_x^\dagger \boldsymbol{\psi}_{x'}| = 0$. Hence, when $\rho \to \infty$, the set of possible representations is precisely the same considered by Lovász [7], and we thus have $\vartheta(\rho) \to \vartheta$ as $\rho \to \infty$. So, the value of $\vartheta(\rho)$ moves from the cut-off rate R_1 to the Lovász bound ϑ when ρ varies from 1 to ∞. This clearly implies that the bound of Corollary 1 is finite for all $R > \vartheta$ and thus it allows to bound the zero-error capacity of the channel as

$$\begin{aligned}
C_0 &\leq \lim_{\rho \to \infty} \vartheta(\rho) \qquad &(42) \\
&= \vartheta. &(43)
\end{aligned}$$

In general, the function $\vartheta(\rho)$ turns out to be strongly related to the coefficient $E_x(\rho)$ used in the expurgated bound of Gallager [8] and defined, using our definition of $\boldsymbol{\psi}_x$, as

$$E_x(\rho) = \max_P \left[-\rho \log \sum_{x,x'} P(x)P(x')(\boldsymbol{\psi}_x^\dagger \boldsymbol{\psi}_{x'})^{1/\rho} \right]. \quad (44)$$

In order to present this relation, it is instructive to consider first the so called *non-negative definite* channels as defined by Jelinek [9]. These are channels for which the matrix $\tilde{C}$ with (i,j) element $\tilde{c}_{i,j} = (\boldsymbol{\psi}_i^\dagger \boldsymbol{\psi}_j)^{1/\rho}$ is positive semidefinite for all $\rho \geq 1$. For example, the binary symmetric channel (BSC) is non-negative definite. It was proved by Jelinek that, for these channels, the expurgated coefficient $E_x^{(n)}(\rho)$ computed over the n-fold extensions of the channel (and normalized to n) has the same value as $E_x(\rho)$. It is also known that for these channels, the inputs can be partitioned in subsets such that all pairs of symbols from the same subset are confusable and no pair of symbols from different subsets are confusable. The zero error capacity in this case is simply the logarithm of the number of such subsets. For these channels, since the matrix $\tilde{C}$ is positive semidefinite, there exists a set of vectors $\tilde{\boldsymbol{\psi}}_1, \tilde{\boldsymbol{\psi}}_2, \ldots, \tilde{\boldsymbol{\psi}}_K$ such that $\tilde{\boldsymbol{\psi}}_i^\dagger \tilde{\boldsymbol{\psi}}_j = \tilde{c}_{i,j}$, that is, for all $\rho \geq 1$, representations of degree ρ exist that satisfy all the constraints with equality. In this case, the equivalence with the cut-off rate that we have seen for $\rho = 1$ can be in a sense extended to other ρ values. It can be proved [5, Th. 9] that we can write

$$\begin{aligned}
\vartheta(\rho) &= \min_{\boldsymbol{f}} \max_x \log \frac{1}{|\boldsymbol{f}^\dagger \tilde{\boldsymbol{\psi}}_x|^2} \qquad &(45) \\
&= \max_P \left[-\log \sum_{x,x'} P(x)P(x') \tilde{\boldsymbol{\psi}}_x^\dagger \tilde{\boldsymbol{\psi}}_{x'} \right] &(46) \\
&= \max_P \left[-\log \sum_{x,x'} P(x)P(x')(\boldsymbol{\psi}_x^\dagger \boldsymbol{\psi}_{x'})^{1/\rho} \right] &(47)
\end{aligned}$$

Hence, under such circumstances, we find that $\vartheta(\rho) = E_x(\rho)/\rho$. For example, for the BSC with transition probability ε, we have

$$\vartheta(\rho) = -\log \left(\frac{1}{2} + \frac{1}{2}[4\varepsilon(1-\varepsilon)]^{1/2\rho} \right). \qquad (48)$$

In general, for non-negative definite channels, the bound of Corollary 1 is obtained by drawing the curve parameterized as $(E_x(\rho)/\rho, 2E_x(\rho))$ in the (R, E) plane. Thus, it is seen

that that bound is loose in general. It is however somehow tight in the sense that, in this particular case,

$$C_0 = \lim_{\rho \to \infty} \vartheta(\rho), \tag{49}$$

(which is however trivial) and, if $C_0 = 0$, the bound gives

$$E(0) \leq \lim_{\rho \to \infty} 2\rho\, \vartheta(\rho) \tag{50}$$
$$= \lim_{\rho \to \infty} 2E_x(\rho) \tag{51}$$
$$= 2E_{ex}(0), \tag{52}$$

where $E_{er}(R)$ is Gallager's expurgated lower bound to $E(R)$. If the channel is pairwise reversible, this can then be improved to $E(R) \leq E_{ex}(0)$, which is obviously tight.

For general channels with a non-trivial zero-error capacity, like for example any channel whose confusability graph is a pentagon, what happens is that the matrix $\tilde{C}$ is in general positive semidefinite only for values of ρ in a range $[1, \bar{\rho}]$ and then it becomes not positive semidefinite for some $\rho > \bar{\rho}$. This implies that for $\rho > \bar{\rho}$, representations that satisfy all the constraints with equality do not exist in general. In this case, the two expressions in equations (46) and (47) are no more equal and in general they could both differ from $\vartheta(\rho)$. If all the values $\tilde{\psi}_x^\dagger \tilde{\psi}_{x'}$ are nonnegative[2], however, then it can be proved that the expression in (46) equals $\vartheta(\rho)$ [5, Th. 9]. In this case, we see the interesting difference between $\vartheta(\rho)$ and $E_x(\rho)/\rho$. The two quantities follow respectively (46) and (47). When $\rho \to \infty$, $\vartheta(\rho)$ tends to ϑ, an upper bound to C_0. The value $E_x(\rho)/\rho$ instead is known to converge to the independence number of the confusability graph of the channel [10], a lower bound to C_0.

More generally, if $\tilde{\psi}_x^\dagger \tilde{\psi}_{x'} \geq 0$, $\forall x, x'$, since $\vartheta(\rho)$ is given by equation (46), it is an upper bound to (47) and thus to $E_x(\rho)/\rho$. It can then also be proved that (46) is multiplicative, in this case, over the n-fold tensor power of the representation $\{\tilde{\psi}_k\}$. This implies that, for all n, $\vartheta(\rho)$ is an upper bound to the (normalized) expurgated bound $E_x^{(n)}(\rho)/\rho$ computed for the n-fold memoryless extension of the channel. That is, $\vartheta(\rho)$ generalizes ϑ in the sense that, in the same way as

$$\vartheta \geq \sup_n \lim_{\rho \to \infty} \frac{E_x^{(n)}(\rho)}{\rho} = C_0, \tag{53}$$

also

$$\vartheta(\rho) \geq \sup_n \frac{E_x^{(n)}(\rho)}{\rho}. \tag{54}$$

The discussion of this point with generality requires some technicalities and will hopefully be given in a future work (see footnote 2).

Is is worth pointing out that, for some channels, the optimal representation may even stay fixed for ρ larger than some given finite value $\rho_{\max}$ and $\vartheta(\rho)$ is thus constant for $\rho \geq \rho_{\max}$ (in this case, the bounds are useless for $\rho > \rho_{\max}$). This happens for the famous example for the noisy typewriter channel with

---

[2]We conjecture that the optimal representation, in terms of Lovász's definition of *value*, always satisfies this condition. We have not yet investigated this aspect, but have never found a counterexample.
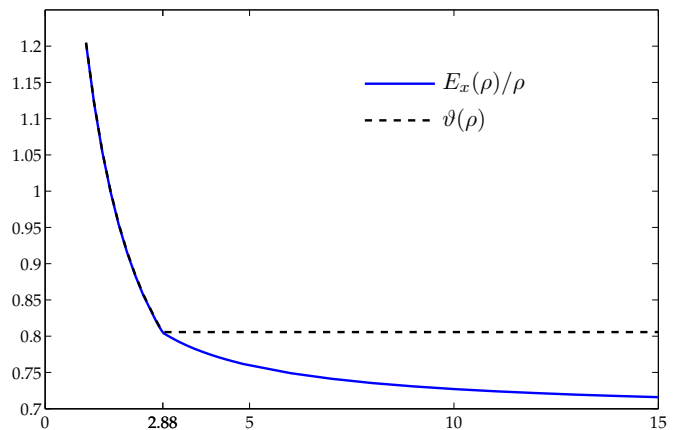

Fig. 1. Plot of $\vartheta(\rho)$ and $E_x(\rho)/\rho$ for the noisy typewriter channel with five inputs and crossover probability $1/2$.

five inputs and crossover probability $1/2$. In this case $\bar{\rho} = \rho_{\max} \approx 2.88$; as shown in Fig. 1, for $\rho < \rho_{\max}$ we have $\vartheta(\rho) = E_x(\rho)/\rho$ while, for $\rho \geq \rho_{\max}$, $\vartheta(\rho) = C_0 = \log\sqrt{5}$.

We close the paper with a comment on the relation between the results presented here and some recent results in quantum information theory. In that context, it is revealed [11] that Lovász's idea is intimately related to the sphere-packing bound of [4]. The bound to $E(R)$ derived here is a special case of a more general bound that can be derived by properly applying the sphere-packing bound for classical-quantum channels [12], [5]. In particular, while the construction of the representation $\{\tilde{\psi}_x\}$ was introduced here as a purely mathematical trick to bound $E(R)$, this procedure can be interpreted in the context of classical-quantum channels as a natural way to bound $E(R)$ by comparing the original channel with an auxiliary one.

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. II," *Information and Control*, vol. 10, pp. 522–552, 1967.

[4] ——, "Lower Bounds to Error Probability for Coding in Discrete Memoryless Channels. I," *Information and Control*, vol. 10, pp. 65–103, 1967.

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

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

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

[8] R. G. Gallager, "A Simple Derivation of the Coding Theorem and Some Applications," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 3–18, 1965.

[9] F. Jelinek, "Evaluation of Expurgated Error Bounds," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 501–505, 1968.

[10] I. Korn, "On the Lower Bound of Zero-Error Capacity," *IEEE Trans. on Inform. Theory*, vol. 14, no. 3, pp. 509 – 510, may 1968.

[11] M. Dalai, "Lovász's Theta Function, Rényi's Divergence and the Sphere-Packing Bound," in *Proc. IEEE Intern. Symp. Inform. Theory*, 2013.

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