Bounding Gallager's Expurgated Bound

Marco Dalai

***Abstract*—In [1], [2], a modification of the Lovász theta function, called $\vartheta(\rho)$, was introduced for the purpose of studying error exponents for codes over discrete memoryless channels. It was mentioned that this function allows one to *upper* bound the multi-letter version of Gallager's expurgated *lower* bound on the reliability function, but this property was only proved for some particular cases. In this letter, we give a general proof and we show how this can be used for the evaluation of the expurgated bound. We finally consider some algebraic properties of the function $\vartheta(\rho)$.**

I. Introduction

Let $W(y|x)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, 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. When a sequence $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ is sent through the channel, the probability of receiving a sequence $\boldsymbol{y} = (y_1, y_2, \ldots, y_n) \in \mathcal{Y}^n$ at the output is

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

The reliability function of a channel is defined as

$$E(R) = \limsup_{n\to\infty} \frac{1}{n} \log \frac{1}{\mathsf{P}_\mathrm{e}(\lceil 2^{nR} \rceil, n)},$$

where $\mathsf{P}_\mathrm{e}(M, n)$ is the smallest possible probability of error of codes with M codewords of length n. In [3], Gallager introduced a lower bound on $E(R)$ known as the expurgated bound. In its most general multi-letter form, the bound says that $E(R) \geq E_\mathrm{ex}^{(n)}(R)$ where

$$E_\mathrm{ex}^{(n)}(R) = \sup_{\rho\geq 1} \left[E_\mathrm{x}^{(n)}(\rho) - \rho R \right], \tag{2}$$

$$E_\mathrm{x}^{(n)}(\rho) = -\frac{\rho}{n} \log \min_P Q^{(n)}(\rho, P), \tag{3}$$

$$Q^{(n)}(\rho, P) = \sum_{\boldsymbol{x}_1, \boldsymbol{x}_2} P(\boldsymbol{x}_1) P(\boldsymbol{x}_2) g_n(\boldsymbol{x}_1, \boldsymbol{x}_2)^{\frac{1}{\rho}}, \tag{4}$$

$$g_n(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sum_{\boldsymbol{y}} \sqrt{\boldsymbol{W}(\boldsymbol{y}|\boldsymbol{x}_1)\boldsymbol{W}(\boldsymbol{y}|\boldsymbol{x}_2)}. \tag{5}$$

Here P runs over distributions on $\mathcal{X}^n$. By restricting P to be a product distribution, one easily sees that for any k, n, $E_\mathrm{x}^{(kn)}(\rho) \geq E_\mathrm{x}^{(n)}(\rho)$ and, in particular,

$$E_\mathrm{x}^{(n)}(\rho) \geq E_\mathrm{x}^{(1)}(\rho). \tag{6}$$

The computation of $E_\mathrm{x}^{(n)}(\rho)$ is not simple in general, due to the fact that the quadratic form $Q^{(n)}(\rho, P)$ is not necessarily convex. The case where $Q^{(n)}(\rho, P)$ is actually convex for all $\rho \geq 1$ was considered by Jelinek [4] which showed that in this case the single-letter version of the bound, i.e., with $n = 1$, is not improved by the multi-letter versions, i.e., with $n > 1$. In the other cases, the multi-letter version does often improve the single-letter one, but a quantification of this effect can be challenging. Indeed, any chosen P will give a lower bound on $E_\mathrm{ex}^{(n)}(R)$, but finding the optimal P, and hence the exact value of $E_\mathrm{ex}^{(n)}(R)$, can be a hard computational problem even for small values of n.

This purpose of this letter is to point out how a modification of the Lovász theta function [5], first introduced in [1], [2], can be used to study the function $E_\mathrm{ex}^{(n)}(R)$ and, in particular, to find upper bounds which might unexpectedly match lower bounds in some particular cases.

II. The $\vartheta(\rho)$ function

For any $x \in \mathcal{X}$, let ψ_x be the unit norm $|\mathcal{Y}|$-dimensional column vector with components $\psi_x(y) = \sqrt{W(y|x)}$. Similarly, for any $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, consider the unit norm $|\mathcal{Y}|^n$-dimensional vector $\boldsymbol{\psi}_{\boldsymbol{x}}$ with components $\boldsymbol{\psi}_{\boldsymbol{x}}(\boldsymbol{y}) = \sqrt{W^{(n)}(\boldsymbol{y}|\boldsymbol{x})}$. Since the channel is memoryless, we can write

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

where $\otimes$ is the Kronecker product. In particular, note that with this notation we can write $g_n(\boldsymbol{x}, \boldsymbol{x}') = \boldsymbol{\psi}_{\boldsymbol{x}}^\dagger \boldsymbol{\psi}_{\boldsymbol{x}'}$, so that

$$E_\mathrm{x}^{(n)}(\rho) = -\frac{\rho}{n} \log \min_P \sum_{\boldsymbol{x}, \boldsymbol{x}'} P(\boldsymbol{x}) P(\boldsymbol{x}') (\boldsymbol{\psi}_{\boldsymbol{x}}^\dagger \boldsymbol{\psi}_{\boldsymbol{x}'})^{1/\rho}. \tag{8}$$

Consider now the inner products $\psi_x^\dagger \psi_{x'} \geq 0$. For a fixed $\rho \geq 1$, consider a set of "tilted" vectors, that is, unit norm vectors $\tilde{\psi}_x$, $x \in \mathcal{X}$, in any Hilbert space such that $|\tilde{\psi}_x^\dagger \tilde{\psi}_{x'}| \leq (\psi_x^\dagger \psi_{x'})^{1/\rho}$. We call such a set of vectors $\{\tilde{\psi}_x\}$ an *orthonormal representation of degree* ρ of our channel, and call $\Gamma(\rho)$ the set of all possible such representations

$$\Gamma(\rho) = \left\{ \{\tilde{\psi}_x\} : |\tilde{\psi}_x^\dagger \tilde{\psi}_{x'}| \leq (\psi_x^\dagger \psi_{x'})^{1/\rho} \right\}, \quad \rho \geq 1. \tag{9}$$

Note that, since the original ψ_x vectors satisfy the constraints, $\Gamma(\rho)$ is non-empty. The (logarithmic[1]) *value* of an orthonormal representation is the quantity

$$V(\{\tilde{\psi}_x\}) = \min_f \max_x \log \frac{1}{|\tilde{\psi}_x^\dagger f|^2}, \tag{10}$$

where the minimum is over all unit norm vectors f. Following Lovász, we call the vector f the *handle* of the representation. Call now $\vartheta(\rho)$ the minimum value over all representations of degree ρ,

$$\vartheta(\rho) = \min_{\{\tilde{\psi}_x\} \in \Gamma(\rho)} V(\{\tilde{\psi}_x\}). \tag{11}$$

Marco Dalai is with the Department of Information Engineering, University of Brescia, Italy. Email: marco.dalai@unibs.it.

The research was supported by Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) under grant PRIN 2015 D72F1600079000.

[1] We use a logarithmic definition of the value and of the theta function for an easier comparison with rates.

Remark 1: Note that it is always possible to find an optimal representation and a handle such that $\tilde{\psi}_x^\dagger f > 0$, $\forall x$, since changing any $\tilde{\psi}_x$ with $-\tilde{\psi}_x$ gives a valid representation with the same value. In particular, there exists a representation which satisfies

$$\tilde{\psi}_x^\dagger f \geq e^{-\vartheta(\rho)/2}, \quad \forall x\,. \tag{12}$$

As already mention in [1], [2], for any fixed ρ the optimal representation $\{\tilde{\psi}_x\}$ and the value $\vartheta(\rho)$ can be efficiently obtained by solving a semidefinite optimization problem. If we consider the $(K+1)\times(K+1)$ Gram matrix

$$G = [\tilde{\psi}_1,\ldots,\tilde{\psi}_K, f]^\dagger[\tilde{\psi}_1,\ldots,\tilde{\psi}_K, f],$$

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

$$\begin{array}{rrcl} & \max & t & \\ \text{s.t.} & G(x,K+1) & \geq & t, \quad \forall x \leq K \\ & G(x,x) & = & 1, \quad \forall x \\ & G(x,x') & \leq & (\psi_x^\dagger \psi_x')^{1/\rho}, \quad x \neq x' \\ & G(x,x') & \geq & -(\psi_x^\dagger \psi_x')^{1/\rho}, \quad x \neq x' \\ & G & \text{is} & \text{positive semidefinite.} \end{array} \tag{13}$$

The solution to this problem is $t^* = e^{-\vartheta(\rho)/2}$, and both the optimal representation vectors $\tilde{\psi}_x$, $x \in \mathcal{X}$, and the handle f can be obtained by means of the spectral decomposition of the optimal G found.

III. Bounding the Expurgated Bound

Consider the $K\times K$ matrix, say $G(\rho)$, whose (x,x') element takes value $(\psi_x^\dagger\psi_{x'})^{1/\rho}$. As already mentioned, it was proved by Jelinek [4] that $E_x^{(n)}(\rho) = E_x^{(1)}(\rho)$ for all n if $G(\rho)$ is positive semidefinite. Jelinek studied a class of channels for which the matrix $G(\rho)$ is positive semidefinite for all $\rho \geq 1$ and for which, thus, the expurgated bound computed for blocks of n symbols is the same as computed for just one symbol. For a general channel W and sufficiently large ρ, however, $G(\rho)$ need not be positive semidefinite, and finding the exact value of $E_x^{(n)}(\rho)$ is a complex problem even for small n. The function $\vartheta(\rho)$ allows us to bound $E_x^{(n)}(\rho)$, as stated in the following theorem.

Theorem 1: For any channel W, any real $\rho \geq 1$ and any integer $n \geq 1$,

$$\frac{E_{\mathrm{x}}^{(n)}(\rho)}{\rho} \leq \vartheta(\rho)\,, \tag{14}$$

with equality if $G(\rho)$ is positive semidefinite.

Remark 2: It was already proved in [2] that the above inequality holds if, for the optimal representation $\{\tilde{\psi}_x\}$, all the inner products $\tilde{\psi}_x^\dagger\tilde{\psi}_{x'}$ are non-negative. Here we remove this assumption and, at the same time, give a much simpler proof. Note also that it was already observed in [2] that equality holds in (14) if $G(\rho)$ is positive semidefinite, that is, when evaluation of $E_{\mathrm{x}}^{(n)}(\rho)$ is simple. Hence, the interesting use of (14) is when $G(\rho)$ is not positive semidefinite.

Proof: Let $\{\tilde{\psi}_x\}$ be an optimal representation of degree ρ with handle f satisfying (12). 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

$$\begin{aligned} |\tilde{\boldsymbol{\psi}}_{\boldsymbol{x}}^\dagger\tilde{\boldsymbol{\psi}}_{\boldsymbol{x}'}| &= \prod_{i=1}^n |\tilde{\psi}_{x_i}^\dagger\tilde{\psi}_{x_i'}| & (15)\\ &\leq \prod_{i=1}^n (\psi_{x_i}^\dagger\psi_{x_i'})^{1/\rho} & (16)\\ &= (\boldsymbol{\psi}_{\boldsymbol{x}}^\dagger\boldsymbol{\psi}_{\boldsymbol{x}'})^{1/\rho}\,. & (17)\end{aligned}$$

Consider the argument of the log in definition of $E_x^{(n)}(\rho)$. For any fixed distribution P over $\mathcal{X}^n$, we have

$$\begin{aligned} \sum_{\boldsymbol{x},\boldsymbol{x}'} P(\boldsymbol{x})P(\boldsymbol{x}')(\boldsymbol{\psi}_{\boldsymbol{x}}^\dagger\boldsymbol{\psi}_{\boldsymbol{x}'})^{1/\rho} &\geq \sum_{\boldsymbol{x},\boldsymbol{x}'} P(\boldsymbol{x})P(\boldsymbol{x}')|\tilde{\boldsymbol{\psi}}_{\boldsymbol{x}}^\dagger\tilde{\boldsymbol{\psi}}_{\boldsymbol{x}'}| \\ &\geq \left|\sum_{\boldsymbol{x},\boldsymbol{x}'} P(\boldsymbol{x})P(\boldsymbol{x}')\tilde{\boldsymbol{\psi}}_{\boldsymbol{x}}^\dagger\tilde{\boldsymbol{\psi}}_{\boldsymbol{x}'}\right| \\ &= \left\|\sum_{\boldsymbol{x}} P(\boldsymbol{x})\tilde{\boldsymbol{\psi}}_{\boldsymbol{x}}\right\|^2. \end{aligned}$$

Set now $\boldsymbol{f} = f^{\otimes n}$. Since $\|\boldsymbol{f}\|^2 = 1$, by the Cauchy-Schwartz inequality we have

$$\left\|\sum_{\boldsymbol{x}} P(\boldsymbol{x})\tilde{\boldsymbol{\psi}}_{\boldsymbol{x}}\right\|^2 \geq \left|\sum_{\boldsymbol{x}} P(\boldsymbol{x})\tilde{\boldsymbol{\psi}}_{\boldsymbol{x}}^\dagger\boldsymbol{f}\right|^2.$$

Since for our representation we have $\tilde{\psi}_x^\dagger f \geq e^{-\vartheta(\rho)/2}$, $\forall x$, we then have, for all $\boldsymbol{x}$,

$$\begin{aligned} \tilde{\boldsymbol{\psi}}_{\boldsymbol{x}}^\dagger\boldsymbol{f} &= \prod_{i=1}^n \tilde{\psi}_{x_i}^\dagger f & (18)\\ &\geq e^{-n\vartheta(\rho)/2}, & (19)\end{aligned}$$

from which we deduce that

$$\sum_{\boldsymbol{x}} P(\boldsymbol{x})\tilde{\boldsymbol{\psi}}_{\boldsymbol{x}}^\dagger\boldsymbol{f} \geq e^{-n\vartheta(\rho)/2}\,. \tag{20}$$

Thus, for any fixed distribution P over $\mathcal{X}^n$

$$\sum_{\boldsymbol{x},\boldsymbol{x}'} P(\boldsymbol{x})P(\boldsymbol{x}')(\boldsymbol{\psi}_{\boldsymbol{x}}^\dagger\boldsymbol{\psi}_{\boldsymbol{x}'})^{1/\rho} \geq e^{-n\vartheta(\rho)}\,, \tag{21}$$

which immediately implies that

$$\frac{E_x^{(n)}(\rho)}{\rho} \leq \vartheta(\rho). \tag{22}$$

We now show that if $G(\rho)$ is positive semidefinite then equality holds for all n. First note that, as observed by Jelinek, if $G(\rho)$ is positive semidefinite then $E_x^{(n)}(\rho) = E_x(\rho)$, and the optimizing P for $n = 1$ satisfies

$$\sum_{x'} P(x')(\psi_x^\dagger\psi_{x'})^{1/\rho} \geq \sum_{i,k} P(i)P(k)(\psi_i^\dagger\psi_k)^{1/\rho}\,, \tag{23}$$

with equality for all x such that $P(x) > 0$. Since $G(\rho)$ is positive semidefinite, there exist vectors $\{\tilde{\psi}_x\}$ such that

$$\tilde{\psi}_i^\dagger\tilde{\psi}_k = (\psi_i^\dagger\psi_k)^{1/\rho}\,, \tag{24}$$

implying that $\{\tilde{\psi}_x\}$ constitute a valid representation of degree ρ. Then, the condition above for the optimality of P can be written as

$$\tilde{\psi}_x^\dagger\left(\sum_{x'} P(x')\tilde{\psi}_{x'}\right) \geq \left(\sum_i P(i)\tilde{\psi}_i\right)^\dagger \left(\sum_k P(k)\tilde{\psi}_k\right), \qquad (25)$$

with equality whenever $P(x) > 0$. Considering the unit norm vector

$$f = \frac{\sum_k P(k)\tilde{\psi}_k}{\left\|\sum_k P(k)\tilde{\psi}_k\right\|}, \qquad (26)$$

this condition can be rewritten as

$$\tilde{\psi}_x^\dagger f \geq \left\|\sum_k P(k)\tilde{\psi}_k\right\|, \qquad (27)$$

which implies

$$\log\frac{1}{|\tilde{\psi}_x^\dagger f|^2} \leq -\log\left\|\sum_k P(k)\tilde{\psi}_k\right\|^2 \qquad (28)$$

$$= -\log\left(\sum_{i,k} P(i)P(k)(\psi_i^\dagger\psi_k)^{1/\rho}\right) \qquad (29)$$

$$= \frac{E_x(\rho)}{\rho}. \qquad (30)$$

This shows that $V(\{\tilde{\psi}_x\}) \leq E_x(\rho)/\rho$ and hence $\vartheta(\rho) \leq E_x(\rho)/\rho$. Since we earlier proved that $\vartheta(\rho) \geq E_x(\rho)/\rho$ we must have equality. ■

IV. Example of application

We show in this section an example of use of $\vartheta(\rho)$ for numerically computing or bounding the expurgated bound of channels.

Consider the channel with four inputs and four outputs shown in Figure 1. For this channel the matrix $G(\rho)$ is given by

$$G(\rho) = \begin{pmatrix} 1 & .2^{1/\rho} & 0 & 0.2^{1/\rho} \\ 2^{1/\rho} & .68^{1/\rho} & .16^{1/\rho} & .04^{1/\rho} \\ 0 & .16^{1/\rho} & .68^{1/\rho} & 0 \\ .2^{1/\rho} & .04^{1/\rho} & 0 & .68^{1/\rho} \end{pmatrix}, \qquad (31)$$

which is positive semidefinite if and only if $\rho \leq \bar{\rho}$, where $\bar{\rho} \approx 2.47927$. For $\rho > \bar{\rho}$, one can try to solve the minimization in (3) numerically, but since $G(\rho)$ is not positive semidefinite it is not easy to determine whether a minimum has really been found. So, in general one determines a provable lower bound on $E_x^{(1)}(\rho)$. Figure 2(a) shows the results obtained when attempting to compute $E_x^{(1)}(\rho)$ using the Matlab function `fmincon` and the value (numerically reliable) of $\vartheta(\rho)$ determined by solving problem (13) using the library `cvx`. Since the two plots match, using (6) and (14) we deduce two important facts: that our evaluation of $E_x^{(1)}(\rho)$ is correct and that $E_x^{(n)}(\rho) = E_x^{(1)}(\rho)$ for the range of values of ρ tested. Figure 2(b) shows the resulting plot of the expurgated bound, which tells us another interesting fact, that is that the expurgated bound only improves the random coding bound at rates[2] $R < 0.6951$, which is only very slightly larger than $\log(2) \approx 0.9631$, the zero-error capacity. This example shows that $\vartheta(\rho)$ can be used as an effective tool for studying $E_{ex}^{(n)}(R)$ in those situations where Jelinek's study does not apply.

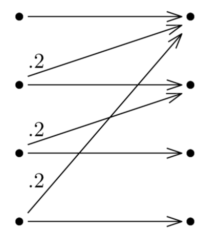


Fig. 1. Channel for which we compare $E_x(\rho)/\rho$ and ϑ in Figure 2.

V. On the algebraic properties of $\vartheta(\rho)$

It seems legitimate to ask whether the function $\vartheta(\rho)$ preserves some of the very interesting properties of Lovász's ϑ. Here we show how one of the representations derived by Lovász changes when we force the representations to be in $\Gamma(\rho)$. The theorem presented in this section generalizes Lovász's [5, Th. 3].

For any $\gamma > 0$, let $\mathcal{A}(\gamma)$ be the set of symmetric $n\times n$ matrices $A = (a_{i,j})$ such that

$$a_{i,i} = 1$$
$$|a_{i,j} - 1| \leq \gamma\,(\psi_i^\dagger\psi_j)^{1/\rho}, \quad i \neq j.$$

Define then the function

$$r(\gamma) = \inf_{A\in\mathcal{A}(\gamma)} \lambda_{\max}(A), \qquad (32)$$

where $\lambda_{\max}(A)$ represents the largest eigenvalue of the matrix A.

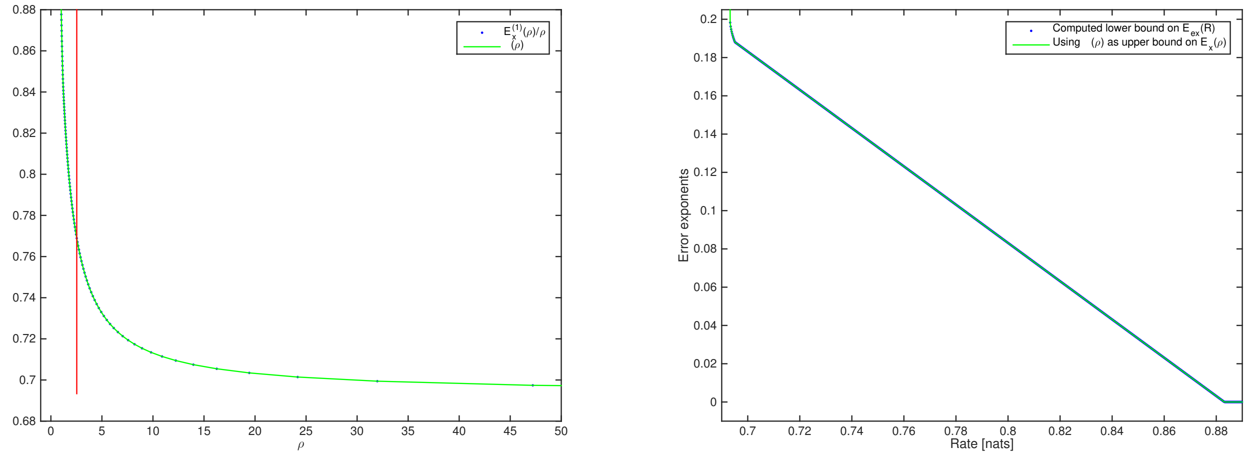
Theorem 2: $\vartheta(\rho) = \log\gamma^*$, where γ^* is the only solution of the equation $\gamma = r(\gamma)$.

Proof: We first prove that the equation $\gamma = r(\gamma)$ has one unique solution. Note that the $\mathcal{A}(\gamma) \subseteq \mathcal{A}(\gamma')$ if $\gamma < \gamma'$, which implies that $r(\gamma)$ is a non-increasing function. Furthermore, if $A_1 \in \mathcal{A}(\gamma_1)$ and $A_2 \in \mathcal{A}(\gamma_2)$ then, for any $\alpha \in [0,1]$, the matrix $A = \alpha A_1 + (1-\alpha)A_2$ is in $\mathcal{A}(\alpha\gamma_1 + (1-\alpha)\gamma_2)$ and $\lambda_{\max}(A) \leq \alpha\lambda_{\max}(A_1) + (1-\alpha)\lambda_{\max}(A_2)$. This implies that $r(\gamma)$ is a convex function of γ and, hence, it is continuous. Since $r(\gamma)$ is non-increasing and continuous, there is only one solution γ^* to the equation $r(\gamma) = \gamma$. We now proceed to the proof that the solution is $\gamma^* = e^{\vartheta(\rho)}$.

We first prove that $e^{\vartheta(\rho)} \geq \gamma^*$. Let $\{\tilde{\psi}_x\}$ be a representation of degree ρ with handle f and consider the matrix A with elements

$$a_{i,i} = 1$$
$$a_{i,j} = 1 - \frac{\tilde{\psi}_i^\dagger\tilde{\psi}_j}{(\tilde{\psi}_i^\dagger f)(\tilde{\psi}_j^\dagger f)}, \quad i \neq j.$$

[2]The exact computation of $E_x^{(n)}(\rho)$ implies that the computation of $E_{ex}^{(n)}(R)$ is guaranteed to be correct for $R > \vartheta(\rho_{\max})$, where $\rho_{\max}$ is the largest ρ tested, since larger values of ρ might improve $E_{ex}^{(n)}(R)$ only at lower rates. In our computation, we had $\rho_{\max} = 1000$ and $\vartheta(\rho_{\max}) \approx 0.6933$.



(a) Comparison of $E_{\rm x}(\rho)/\rho$ and $\vartheta(\rho)$.

(b) Plot of the expurgated bound.

Fig. 2. Numerical results for the channel defined in Figure 1. The red line in Figure 2(a) represents the largest ρ for which $G(\rho)$ is positive semidefinite.

Note that $A \in \mathcal{A}(e^{\vartheta(\rho)})$, since $|\tilde{\psi}_i^\dagger \tilde{\psi}_j| \leq (\psi_i^\dagger \psi_j)^{1/\rho}$ and $|\tilde{\psi}_x^\dagger f| \geq e^{-\vartheta(\rho)/2}, \forall x$. Then, as in [5, Th. 3], we have

$$-a_{i,j} = \left(f - \frac{\tilde{\psi}_i}{f^\dagger \tilde{\psi}_i}\right)^\dagger \left(f - \frac{\tilde{\psi}_j}{f^\dagger \tilde{\psi}_j}\right), \quad i \neq j, \tag{33}$$

and

$$e^{\vartheta(\rho)} - a_{i,i} = \left\| f - \frac{\tilde{\psi}_i}{f^\dagger \tilde{\psi}_i} \right\|^2 + \left(e^{\vartheta(\rho)} - \frac{1}{\|f^\dagger \tilde{\psi}_i\|^2} \right). \tag{34}$$

This implies that $e^{\vartheta(\rho)} I - A$ is positive semidefinite, and hence the largest eigenvalue of A is at most $e^{\vartheta(\rho)}$, implying that $r(e^{\vartheta(\rho)}) \leq e^{\vartheta(\rho)}$ or, in other words, that $e^{\vartheta(\rho)} \geq \gamma^*$.

Conversely, we prove that $e^{\vartheta(\rho)} \leq \gamma^*$. To do this, we prove that if $\gamma_1 > \gamma^*$, then $e^{\vartheta(\rho)} < \gamma_1$. Let then $\gamma_1 > \gamma^*$, so that $r(\gamma_1) < \gamma_1$. Since $r(\gamma)$ is continuous, there exists a $\gamma < \gamma_1$ such that $r(\gamma) < \gamma$ as well. For this γ, let A be a matrix in $\mathcal{A}(\gamma)$ achieving $r(\gamma)$, that is $\lambda_{\max}(A) = r(\gamma)$. Then, since $\gamma > r(\gamma) = \lambda_{\max}(A)$, the matrix $\gamma I - A$ is positive semi-definite and there exists a set of n vectors $\{v_i\}$ such that

$$\gamma \delta_{ij} - a_{i,j} = v_i^\dagger v_j, \tag{35}$$

where δ_{ij} is the Kronecker delta. Let f be a unit norm vector orthogonal to all the v_i (expand the space dimension if necessary) and set

$$\tilde{\psi}_i = \frac{1}{\sqrt{\gamma}}(f + v_i), \tag{36}$$

so that

$$\tilde{\psi}_i^\dagger \tilde{\psi}_j = \frac{1}{\gamma}(1 + v_i^\dagger v_j) \tag{37}$$

$$= \frac{1}{\gamma}(1 + \gamma\delta_{ij} - a_{i,j}). \tag{38}$$

Since $A \in \mathcal{A}(\gamma)$, we then easily check that

$$\tilde{\psi}_i^\dagger \tilde{\psi}_i = 1$$

and

$$|\tilde{\psi}_i^\dagger \tilde{\psi}_j| = \frac{|1 - a_{i,j}|}{\gamma} \leq (\psi_i^\dagger \psi_j)^{1/\rho},$$

so that the vectors $\{\tilde{\psi}_i\}$ form an orthogonal representation of degree ρ for the channel. Furthermore, we find that for all i,

$$\frac{1}{\tilde{\psi}_i^\dagger f} = \sqrt{\gamma}, \tag{39}$$

which means that the value of this representation is not larger than $\log \gamma$. Hence, $\vartheta(\rho) \leq \log \gamma < \log \gamma_1$ or, in other words, $e^{\vartheta(\rho)} < \gamma_1$ as was to be proven. ■

As in [5], the last part of the proof above, used with $\gamma = \gamma^* = e^{\vartheta(\rho)}$, also implies the following result.

Theorem 3: There always exists a representation $\{\tilde{\psi}_x\}$ of degree ρ with handle f such that

$$\tilde{\psi}_i^\dagger f = e^{-\vartheta(\rho)/2}, \quad \forall i. \tag{40}$$

VI. Conclusions

We presented a technique for bounding, and in certain cases numerically compute the multi-letter version of Gallager's expurgated bound $E_{\rm ex}^{(n)}(R)$. The procedure is based on an extension of Lovász's method for bounding the capacity of graphs.

References

[1] M. Dalai, "An "Umbrella" Bound of the Lovász-Gallager Type," in *Proc. IEEE Intern. Symp. Inform. Theory*, 2013, pp. 3025–3029.

[2] ——, "Lower Bounds on the Probability of Error for Classical and Classical-Quantum Channels," *IEEE Trans. Inform. Theory*, vol. 59, no. 12, pp. 8027 – 8056, 2013.

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

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

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