Revisiting Gallager's error exponent analysis technique

Yonatan Kaspi and Neri Merhav
Department of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 32000, Israel
Email: {kaspi@tx, merhav@ee}.technion.ac.il

***Abstract*—We demonstrate a tight analysis of an expectation of a sum of exponents raised to some power, prevalent in, but not confined to, Gallager's bounding techniques. We show that the traditional analysis that uses Jensen's inequality, although tight in Gallager's random coding error exponent, might not be tight in general. Using the binary symmetric channel as an example, we show that R_c - the lowest rate at which Gallager's bound agrees with the sphere packing bound is the lowest rate for which Jensen's inequality is tight for a range of possible parameters.**

I. Introduction

We demonstrate a tight analysis of expressions of the form

$$\boldsymbol{E}\left[\sum_{m=1}^{e^{nR}} e^{n\lambda f(\hat{Q}\boldsymbol{y},\boldsymbol{x}_m)}\right]^{\rho} \tag{1}$$

where f is any continuous functional of the joint empirical distribution of $(\boldsymbol{y},\boldsymbol{x}_m)$. These expressions are frequently encountered in the analysis of random coding exponents using Gallager's techniques [1] (with f replaced by $\log P(\boldsymbol{y}|\boldsymbol{x}_m)$). Other examples where we find such expressions include list and erasure decoding [2], broadcast channels [3] and universal decoding [4]. The traditional analysis of these expressions, for $0 \leq \rho \leq 1$, uses Jensen's inequality, which allows one to insert the expectation operator into the square brackets in the above expression. The resulting expression is much easier to analyze. However, this comes at the price of uncertainty regarding the exponential tightness of the bound. Forney [2] showed that in some cases, bounds on such expressions can be tightened by introducing another parameter and using it to trade off between the tightness of Jensen's inequality and a variant of Hölder's inequality (see [2] for details). However, exponential tightness is still not guaranteed and this refinement results in *inequalities* rather than *exponential equalities*.

In this paper, we use Gallager's upper bound on the error probability of a random code to demonstrate a tight analysis of expressions having the form of (1). It is known that Gallager's bound is tight for a random code [5]. We compare the traditional analysis used by Gallager with an analysis technique that is inspired by the analysis of the Random Energy Model (REM) in statistical physics (see [6] and [7], [8] for a comprehensive survey). It was shown before that these tools improve on previous results obtained by the standard analysis technique [9], [4], [10], [11]. Here, our purpose is to gain some understanding and insights concerning the conditions under which expressions like (1) can be evaluated by the, much simpler, traditional analysis while preserving exponential tightness.

Using this relatively new technique we will show that there is a critical rate, R_c, pertaining to the number of terms in the sum of (1), above which Jensen's inequality, as a lower bound, will be tight for a range of possible $1 \geq \rho \geq \rho_c(R) \geq 0$. When the rate is lower than the critical rate, $\rho_c(R) = 1$ and Jensen's inequality is not tight except for $\rho = 1$. We show that in the binary example, our critical rate is the same as Gallager's critical rate, i.e the lowest rate for which Gallager's bound agrees with the sphere packing bound. $-\rho_c$ is the slope of Gallager's exponent. Motivated by the analysis of the REM, where there are phase transitions in the system's behavior as a function of the parameters, we draw a "phase diagram" showing regions with different behavior of the exponent of (1) in the $\rho - R$ plane. This is not directly related to the phase diagram of the REM. However, the different behavior in the different regions resembles the behavior of the REM.

The remainder of the paper is organized as follows: In Section II, we give the formal setting and notation used throughout the paper. In Section III, we apply the two analysis approaches for a general channel, and in Section IV, we specialize the results for the BSC. Finally, we conclude this work in Section V.

II. Preliminaries

We begin with notation conventions. Capital letters represent scalar random variables (RVs) and specific realizations of them are denoted by the corresponding lower case letters. Random vectors of dimension n will be denoted by boldface letters. The expectation operator will be denoted by $\boldsymbol{E}\{\cdot\}$. When we wish to emphasize the dependence of the expectation on a certain underlying probability distribution, say, Q, we subscript it by Q. i.e. $\boldsymbol{E}_Q\{\cdot\}$.

We consider a memoryless channel with a finite input alphabet $\mathcal{X}$ and finite output alphabet $\mathcal{Y}$, given by $P(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} P(y_i|x_i)$, $(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$. We are sending one of $M = e^{nR}$ messages where the codewords $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_M$ are drawn independently using a distribution $P(\boldsymbol{x}) = \prod_{i=1}^{n} P(x_i)$.

978-1-4244-8682-3/10/$26.00

The empirical distribution pertaining to a vector $\boldsymbol{x} \in \mathcal{X}^n$ will be denoted by $\hat{Q}_{\boldsymbol{x}}$ and its type class – by $T_{\boldsymbol{x}}$. In other words, $\hat{Q}_{\boldsymbol{x}} = \{\hat{q}_{\boldsymbol{x}}(a),\ a \in \mathcal{X}\}$, where $\hat{q}_{\boldsymbol{x}}(a) = n_{\boldsymbol{x}}(a)/n$, $n_{\boldsymbol{x}}(a)$ being the number of occurrences of the letter a in $\boldsymbol{x}$. Similarly, $\hat{q}_{\boldsymbol{x}|\boldsymbol{y}}(a|b) = \hat{q}_{\boldsymbol{x}\boldsymbol{y}}(a,b)/\hat{q}_{\boldsymbol{y}}(b)$ will denote the empirical conditional probability of $X = a$ given $Y = b$ (with convention that $0/0 = 0$), and $\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}}$ will denote $\{\hat{q}_{\boldsymbol{x}|\boldsymbol{y}}(a|b),\ a \in \mathcal{X},\ b \in \mathcal{Y}\}$. $T_{\boldsymbol{x}|\boldsymbol{y}}$ will denote the conditional type class of $\boldsymbol{x}$ given $\boldsymbol{y}$. The expectation w.r.t. the empirical distribution of $(\boldsymbol{x}|\boldsymbol{y})$, $\hat{Q}_{\boldsymbol{x}\boldsymbol{y}}$, will be denoted by $\boldsymbol{E}_{\hat{Q}}\{\cdot\}$, i.e., for a given function $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, we define $\boldsymbol{E}_{\hat{Q}}\{f(X,Y)\}$ as $\sum_{(a,b)\in\mathcal{X}\times\mathcal{Y}} \hat{q}_{\boldsymbol{x}\boldsymbol{y}}(a,b) f(a,b)$, where in this notation, X and Y are understood to be random variables jointly distributed according to $\hat{Q}_{\boldsymbol{x}\boldsymbol{y}}$. The entropy with respect to the empirical distribution $\hat{Q}_{\boldsymbol{x}}$ will be denoted by $H_{\hat{Q}}(X)$. The underlying empirical distributions used for calculating both $\boldsymbol{E}_{\hat{Q}}$ and $H_{\hat{Q}}$ will be understood from the context or, in case of ambiguity, will be explicitly given. Finally, the notation $a_n \doteq b_n$ means that $\frac{1}{n}\log\frac{a_n}{b_n} \to 0$ as $n \to \infty$.

III. General Memoryless Channels

Gallager's upper bound on the average probability of error [1, p. 138] is given by:

$$\begin{aligned}\overline{P_E} &\le \sum_{\boldsymbol{y}} \boldsymbol{E}\left\{P^{\frac{1}{1+\rho}}(\boldsymbol{y}|\boldsymbol{X}_m)\right\}\boldsymbol{E}\left\{\left[\sum_{m'\neq m} P^{\frac{1}{1+\rho}}(\boldsymbol{y}|\boldsymbol{X}_{m'})\right]^{\rho}\right\} \\ &= e^{-n(E_0(\rho)-\rho R)}, \qquad 0 \le \rho \le 1. \end{aligned} \tag{2}$$

where

$$E_0(\rho) = -\log\sum_y\left[\sum_x P(x)P(y|x)^{\frac{1}{1+\rho}}\right]^{1+\rho} \tag{3}$$

We assume here that $P(x)$ is the maximizing distribution of $E_0(\rho, Q)$. We know [5] that, after minimization over ρ, this bound is tight for all $0 \le R \le C$, where C is the channel capacity. While an exponentially tight analysis of the first expectation of (2) is straightforward, Gallager's analysis of the second expectation uses Jensen's inequality. Although Jensen's inequality is generally not exponentially tight, we know that at least in this case, it is (otherwise (2) would not be tight). Using a tight analysis technique, we will gain some insight into the "mechanism" of Jensen's inequality and explain why it is tight in this case, and when is it tight in expressions such as (1). We look at Gallager's expression $\boldsymbol{E}\left[\sum_{m'} P^{1/(1+\rho)}(\boldsymbol{y}|\boldsymbol{x}'_m)\right]^{\rho}$, which resembles (1) with $\lambda = \frac{1}{1+\rho}$ and $f(\hat{Q}_{\boldsymbol{x}\boldsymbol{y}}) = \log P(\boldsymbol{y}|\boldsymbol{x}'_m)$. In this section, all calculations will be made for a specific given channel output, $\boldsymbol{y}$, with empirical distribution $\hat{Q}_{\boldsymbol{y}}$. Our first step is to introduce the type–class enumerator, which counts the number of codewords that belong to the same conditional type–class, given the channel output $\boldsymbol{y}$. We note that all these codewords contribute the same likelihood to the sum and we have

$$\begin{aligned}\boldsymbol{E}\left[\sum_{m'} P^{1/(1+\rho)}(\boldsymbol{y}|\boldsymbol{x}_{m'})\right]^{\rho} &= \\ = \boldsymbol{E}\left[\sum_{\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}}} N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}}) e^{n\frac{1}{1+\rho}\boldsymbol{E}_{\hat{Q}}\log P(Y|X)}\right]^{\rho}\end{aligned} \tag{4}$$

where the inner empirical expectation is taken with respect to $\hat{Q}(\boldsymbol{x},\boldsymbol{y}) = \hat{Q}_{\boldsymbol{x}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y}) \cdot \hat{Q}_{\boldsymbol{y}}(\boldsymbol{y})$. At this point, we apply the two different analysis techniques. We start with the tight approach.

$$\begin{aligned}&\boldsymbol{E}\left[\sum_{m'} P^{1/(1+\rho)}(\boldsymbol{y}|\boldsymbol{x}_{m'})\right]^{\rho} \\ &\quad= \boldsymbol{E}\left[\sum_{\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}}} N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}}) e^{n\frac{1}{1+\rho}\boldsymbol{E}_{\hat{Q}}\log P(Y|X)}\right]^{\rho} \\ &\quad\doteq \sum_{\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}}} \boldsymbol{E}N^{\rho}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}}) e^{n\frac{\rho}{1+\rho}\boldsymbol{E}_{\hat{Q}}\log P(Y|X)}\end{aligned} \tag{5}$$

The last step is true since now we have a polynomial number of elements in the sum and therefore we can distribute ρ over the summands without loosing exponential tightness (see [10, Appendix] for a simple proof). In order to analyze the last expression, we will have to analyze the moments of the enumerators. To this effect, we note that the probability of randomly drawing, using $P(\boldsymbol{x})$, a sequence that will belong to $T_{\boldsymbol{x}|\boldsymbol{y}}$ is exponentially $e^{n(\boldsymbol{E}_{\hat{Q}}\log P(X) + H_{\hat{Q}}(\boldsymbol{x}|\boldsymbol{y}))}$. Since we draw e^{nR} codewords independently, we have

$$\boldsymbol{E}N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}}) \doteq e^{n\left(R + \boldsymbol{E}_{\hat{Q}}\log P(X) + H_{\hat{Q}}(\boldsymbol{x}|\boldsymbol{y})\right)}. \tag{6}$$

We now divide the type–classes according to the sign of the exponent of the last expression. Define:

$$\mathcal{G}_R = \left\{\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}} : R + \boldsymbol{E}_{\hat{Q}}\log P(X) + H_{\hat{Q}}(\boldsymbol{x}|\boldsymbol{y}) > 0\right\} \tag{7}$$

Using the Chernoff bound, it can be shown that for $\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}} \in \mathcal{G}_R$, the random variable $N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}})$ converges in probability double exponentially fast to its expectation. For $\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}} \in \mathcal{G}_R^c$, since the expectation vanishes with n, we are unlikely to find any codewords that belong to the conditional type–class. In this case, the dominant term (other than $\Pr\{N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}}) = 0\}$ which does not contribute to the expectation) is $\Pr(N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}}) = 1)$. This serves as intuition for the following behavior of the moments of the type–class enumerator (see [4] or [10] for details):

$$\begin{aligned}&\boldsymbol{E}N^{\rho}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}}) \\ &\quad\doteq \begin{cases} e^{n\rho\left(R + \boldsymbol{E}_{\hat{Q}}\log P(X) + H_{\hat{Q}}(\boldsymbol{x}|\boldsymbol{y})\right)} & \hat{Q}_{\boldsymbol{x}|\boldsymbol{y}} \in \mathcal{G}_R \\ e^{n\left(R + \boldsymbol{E}_{\hat{Q}}\log P(X) + H_{\hat{Q}}(\boldsymbol{x}|\boldsymbol{y})\right)} & \hat{Q}_{\boldsymbol{x}|\boldsymbol{y}} \in \mathcal{G}_R^c \end{cases}\end{aligned} \tag{8}$$

Now define

$$\begin{aligned}&A(Q) \\ &= \rho\left(R + \boldsymbol{E}_Q\log P(X) + H_Q(X|Y) + \frac{1}{1+\rho}\boldsymbol{E}_Q\log P(Y|X)\right), \\ &B(Q) \\ &= R + \boldsymbol{E}_Q\log P(X) + H_Q(X|Y) + \frac{\rho}{1+\rho}\boldsymbol{E}_Q\log P(Y|X).\end{aligned} \tag{9}$$

Substituting (8) in (5), and taking into account only the dominant exponents. we have

$$\boldsymbol{E}\left[\sum_{m'} P^{\frac{1}{1+\rho}}(\boldsymbol{y}|\boldsymbol{x}'_m)\right]^{\rho} \doteq e^{n\max_{Q\in\mathcal{G}_R} A(Q)} + e^{n\max_{Q\in\mathcal{G}_R^c} B(Q)}. \quad (10)$$

Let $s_1 = \frac{1}{1+\rho}, s_2 = \frac{\rho}{1+\rho}$ and define:

$$Q_l \stackrel{\triangle}{=} Q_l(x|y) = \frac{P(x)P^l(y|x)}{\sum_{x'} P(x')P^l(y|x')}. \quad (11)$$

Let

$$g_R(l) = R + \boldsymbol{E}_{Q_l}\log P(X) + H_{Q_l}(X|Y). \quad (12)$$

if $g_R(l) > 0$, $Q_l \in \mathcal{G}_R$. It can be shown [4] that $g_R(l)$ is decreasing with l. By a straightforward optimization, we see that the unconstrained maximizer of $A(Q)$ is Q_{s_1} and the unconstrained maximizer of $B(Q)$ is Q_{s_2}. As in [4], if the unconstrained maximizer is outside the constrained optimization domain, the maximum is obtained on the boundary of $\mathcal{G}_R$. Using this in the expressions of $A(Q), B(Q)$, we have

$$\max_{Q\in\mathcal{G}_R} A(Q) = \begin{cases} A(Q_{s_1}) & Q_{s_1}(x|y)\in\mathcal{G}_R \\ \frac{\rho}{1+\rho}\boldsymbol{E}_{Q_\delta}\log P(Y|X) & Q_{s_1}(x|y)\in\mathcal{G}_R^c \end{cases},$$

$$\max_{Q\in\mathcal{G}_R^c} B(Q) = \begin{cases} \frac{\rho}{1+\rho}\boldsymbol{E}_{Q_\delta}\log P(Y|X) & Q_{s_2}(x|y)\in\mathcal{G}_R \\ B(Q_{s_2}) & Q_{s_2}(x|y)\in\mathcal{G}_R^c \end{cases}, \quad (13)$$

where δ solves the equation $g_R(\delta) = 0$ (since $g_R(0) > 0$ and for $R < I(X;Y)$, $g_R(1) < 0$, δ exists). Taking the dominant element for each ρ and optimizing over ρ, will give us the true exponent of the variant of (1) we are analyzing here. Note that in contrast to the use of Jensen's inequality, where ρ is confined to be smaller than 1, here the parameter can take any non–negative value.

If, instead of the above tight analysis, Jensen's inequality was used we would have:

$$\begin{aligned}
&\boldsymbol{E}\left[\sum_{m'} P^{\frac{1}{1+\rho}}(\boldsymbol{y}|\boldsymbol{x}'_m)\right]^{\rho} \\
&= \boldsymbol{E}\left[\sum_{\hat{Q}\boldsymbol{x}|\boldsymbol{y}} N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}})e^{n\frac{1}{1+\rho}\boldsymbol{E}_{\hat{Q}}\log P(Y|X)}\right]^{\rho} \\
&\le \left[\sum_{\hat{Q}\boldsymbol{x}|\boldsymbol{y}} \boldsymbol{E}N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}})e^{n\frac{1}{1+\rho}\boldsymbol{E}_{\hat{Q}}\log P(Y|X)}\right]^{\rho} \qquad (\rho\le 1) \\
&\doteq \left[\sum_{\hat{Q}\boldsymbol{x}|\boldsymbol{y}} e^{n\left(R+\boldsymbol{E}_{\hat{Q}}\log P(X)+H_{\hat{Q}}(\boldsymbol{x}|\boldsymbol{y})\right)}e^{n\frac{1}{1+\rho}\boldsymbol{E}_{\hat{Q}}\log P(Y|X)}\right]^{\rho} \\
&\doteq e^{n\rho\max_{\hat{Q}\boldsymbol{x}|\boldsymbol{y}}\left(R+\boldsymbol{E}_{\hat{Q}}\log P(X)+H_{\hat{Q}}(\boldsymbol{x}|\boldsymbol{y})+n\frac{1}{1+\rho}\boldsymbol{E}_{\hat{Q}}\log P(Y|X)\right)} \\
&= e^{n\max_Q A(Q)}. \quad (14)
\end{aligned}$$

Note that here, we seek the global maximum of $A(Q)$ in contrast to the constrained maximization in (10). Also, in contrast to (10), which is exponentially tight, here we have an inequality as a result of Jensen's inequality. Comparing (10) and (14), we conclude that Jensen's inequality is exponentially tight whenever

$$\max_Q A(Q) = \max\{\max_{Q\in\mathcal{G}_R} A(Q), \max_{Q\in\mathcal{G}_R^c} B(Q)\}. \quad (15)$$

Remarks:

1. Observe that

$$\max_Q A(Q) \ge \max\{\max_{Q\in\mathcal{G}_R} A(Q), \max_{Q\in\mathcal{G}_R^c} B(Q)\} \quad (16)$$

whenever $0 \le \rho \le 1$. This of course, is not surprising since Jensen's inequality can not be tighter than our exponentially tight analysis. However, to see this analytically, observe that if the global maximizer of $A(Q)$ is in $\mathcal{G}_R$, Since $s_1 > s_2$ ($\rho < 1$) and $g_R(l)$ is decreasing with l, if $Q_{s_1} \in \mathcal{G}_R$ then $Q_{s_2} \in \mathcal{G}_R$. Since Q_{s_1} is the global optimizer and not Q_δ, $\max_Q A(Q) \ge \max_{Q\in\mathcal{G}_R^c} B(Q)$. If Q_{s_1} is not in $\mathcal{G}_R$, then there are two cases we need to verify: $Q_{s_2} \in \mathcal{G}_R$ or $Q_{s_2} \in \mathcal{G}_R^c$. When $Q_{s_2} \in \mathcal{G}_R$, since the maximizer of $A(Q)$ is Q_{s_1} and not Q_δ, $\max_Q A(Q) > \max_{Q\in\mathcal{G}_R^c} B(Q)$. When $Q_{s_2} \in \mathcal{G}_R^c$, note that $A(Q_{s_2}) \ge B(Q_{s_2})$ since $0 \le \rho \le 1$ and both expressions are negative. Since $A(Q_{s_1}) \ge A(Q_{s_2})$ we have $\max_Q A(Q) > \max_{Q\in\mathcal{G}_R^c} B(Q)$ for this case as well.

2. $\rho = 1$ is a special case where $\max_Q A(Q) = \max_Q B(Q)$. This implies the trivial fact that Jensen's inequality is tight for $\rho = 1$.

3. If there exists R, smaller than the channel capacity for which $g_R(\frac{1}{2}) > 0$, then there exists R_c, such that for all $0 \le R < R_c$, Jensen's inequality is not tight as an upper bound for all ρ except $\rho = 1$, To see this, observe that for $R = 0$, $g_R(l) \le 0$ and $\mathcal{G}_R$ is empty. Also, the global maximizer of $A(Q)$, Q_{s_1}, maximizes $g_R(l)$ for $\rho = 1$ ($s_1 = \frac{1}{2}$). Since there is a rate for which $g_R(\frac{1}{2}) > 0$ and $g_0(\frac{1}{2}) \le 0$, there is a R_c such that for all $R \le R_c$, $g_R(\frac{1}{2}) < 0$. Since for such R and all $0 \le \rho < 1$, the constrained optimizer of $A(Q)$ belongs to $\mathcal{G}_R^c$ we have

$$\max_Q A(Q) > \max\{\max_{Q\in\mathcal{G}_R} A(Q), \max_{Q\in\mathcal{G}_R^c} B(Q)\}, \quad (17)$$

meaning that Jensen's inequality is not tight for this range.

4. For large enough ρ, Jensen's inequality is tight for all $R > 0$ (possibly as a lower bound). This is true since for large enough ρ, $Q_{s_1} \in \mathcal{G}_R$. If for this ρ, $Q_{s_2} \in \mathcal{G}_R$ then $\max_Q A(Q) = \max_{Q\in\mathcal{G}_R} A(Q) > \max_{Q\in\mathcal{G}_R^c} B(Q)$ for reasons explained in the first remark. If $Q_{s_2} \in \mathcal{G}_R^c$, (meaning $\rho > 1$) then $A(Q_{s_1}) \ge B(Q_{s_2})$ and therefore Jensen's inequality is a tight lower bound (note however, that while this is true for expressions having the form of (1), in the context of Gallager's bound, this is meaningless since Gallager starts with a lower bound on the indicator function of false decoding [1]).

5. In the $R - \rho$ plane, there are three regions defined by $\max_{Q\in\mathcal{G}_R} A(Q) > \max_{Q\in\mathcal{G}_R^c} B(Q)$, $\max_{Q\in\mathcal{G}_R} A(Q) <$

$\max_{Q\in\mathcal{G}_R^c} B(Q)$, $\max_{Q\in\mathcal{G}_R} A(Q) = \max_{Q\in\mathcal{G}_R^c} B(Q)$. Denote these regions by I, II, III respectively. For $R < R_c$ (if R_c exists), we will have a transition line between regions I and II. Along this transition line, there is a discontinuity in the exponent (observe that the exponent is $B(Q)$ on the inner boundary of region II and $\rho B(Q)$ on the outer boundary). This is analogous to phase transitions found when analyzing the REM in statistical physics. For a given $R < R_c$, the best exponential bound will be given by reaching the transition line from region II (see Fig.2 in the next section for a concrete example). It is important to note that although the transition line for any $R < R_c$, is above $\rho = 1$, this does not contradict the known fact that Gallager's bound is tight. Note that here we only analyzed the second expectation of (2). When taking the first expectation into account, the optimal ρ for $R < R_c$ is luckily $\rho = 1$, both in Gallager's analysis and in the tight analysis shown here, meaning that Gallager's analysis technique is indeed tight in this case.

Note that all of the above was true for a specific channel output sequence $\boldsymbol{y}$. It is easy to show that the dependence is not on the specific output but rather on the type–class of the output. Tight bounds on expressions of the form of (1) will be given by taking $T_{\boldsymbol{y}}$ with the highest exponent for each rate.

In the next section we specialize the results of this section for the BSC. In the BSC with uniform random coding, the expressions for $A(Q)$, $B(Q)$ do not depend on the channel output and therefore are much easier to analyze.

IV. The binary symmetric channel

We will specialize the results of the previous section for the BSC. Let $p < \frac{1}{2}, \beta = \log\frac{1-p}{p}, \delta = \frac{d}{n}$.

$$\begin{aligned}
&\boldsymbol{E}\left[\sum_{m'} P^{\frac{1}{1+\rho}}(\boldsymbol{y}|\boldsymbol{x}'_m)\right]^\rho \\
&\quad = (1-p)^{n\frac{\rho}{1+\rho}}\boldsymbol{E}\left[\sum_{d=0}^{n} N(d)e^{-d\frac{1}{1+\rho}\beta}\right]^\rho \\
&\quad \doteq (1-p)^{n\frac{\rho}{1+\rho}}\sum_{d=0}^{n}\boldsymbol{E}N^\rho(d)e^{-d\beta\frac{\rho}{1+\rho}} \\
&\quad \doteq (1-p)^{n\frac{\rho}{1+\rho}}\sum_{\delta}\boldsymbol{E}N^\rho(\delta)e^{-n\delta\beta\frac{\rho}{1+\rho}}. \qquad (18)
\end{aligned}$$

In this case, the type-class enumerators of the previous section become distance enumerators, counting the number of codewords with distance d from the channel output.

The exponent of $\boldsymbol{E}N(\delta)$ is sketched in Fig.1, where we see that we have two regions. For $\delta_{GV}(R) \le \delta \le 1-\delta_{GV}(R)$, where $\delta_{GV}(R)$ solves $R + h(\delta) - \log(2) = 0$ and $h(x)$ is the binary entropy of x, we expect to find an exponential number of codewords with normalized distance δ. For rates outside this region, we are not expected to find any codewords. These two regions correspond to type-classes belonging to $\mathcal{G}_R$ and those that do not in the previous section. Let

$$\mathcal{G}_R = \{\delta : R + h(\delta) - \log(2) > 0\} \qquad (19)$$

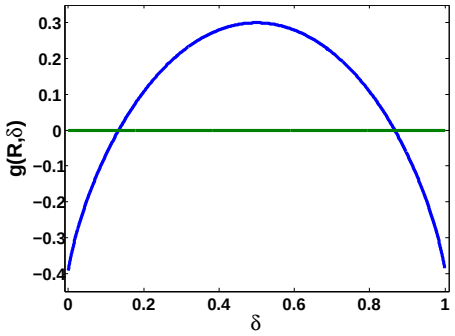

Fig. 1. The exponent of the expectation of the distance enumerator.

Now,

$$\boldsymbol{E}N^\rho(\delta) = \begin{cases} e^{n\rho(R+h(\delta)-\log(2))} & \delta\in\mathcal{G}_R \\ e^{n(R+h(\delta)-\log(2))} & \delta\in\mathcal{G}_R^c \end{cases} \qquad (20)$$

Continuing (18):

$$\begin{aligned}
&\boldsymbol{E}\left[\sum_{m'} P^{1/(1+\rho)}(\boldsymbol{y}|\boldsymbol{x}'_m)\right]^\rho = \\
&\doteq (1-p)^{n\frac{\rho}{1+\rho}}\left[\sum_{\delta\in\mathcal{G}_R} e^{n\rho(R+h(\delta)-\log(2)-\delta\beta\frac{1}{1+\rho})}\right. \\
&\qquad \left. + \sum_{\delta\in\mathcal{G}_R^c} e^{n(R+h(\delta)-\log(2)-\delta\beta\frac{\rho}{1+\rho})}\right] \\
&\doteq (1-p)^{n\frac{1}{1+\rho}}\left[e^{n\rho(R-\log(2)+\max_{\delta\in\mathcal{G}_R} h(\delta)-\delta\beta\frac{1}{1+\rho})}\right. \\
&\qquad \left. + e^{n(R-\log(2)+\max_{\delta\in\mathcal{G}_R^c} h(\delta)-\delta\beta\frac{\rho}{1+\rho})}\right] \\
&\doteq (1-p)^{n\frac{1}{1+\rho}}\left[e^{n\max_{\delta\in\mathcal{G}_R} A(\delta)} + e^{n\max_{\delta\in\mathcal{G}_R^c} B(\delta)}\right] \qquad (21)
\end{aligned}$$

Let δ_A^*, δ_B^* be the maximizers of the first and second sums in (21) respectively. We have:

$$\begin{aligned}
\delta_A^* &= \frac{p^{\frac{1}{1+\rho}}}{p^{\frac{1}{1+\rho}} + (1-p)^{\frac{1}{1+\rho}}} \\
\delta_B^* &= \frac{p^{\frac{\rho}{1+\rho}}}{p^{\frac{\rho}{1+\rho}} + (1-p)^{\frac{\rho}{1+\rho}}}. \qquad (22)
\end{aligned}$$

For $0 \le \rho \le 1$, $\delta_A^* \le \delta_B^*$ and

$$\begin{aligned}
p \le \delta_A^* &\le \frac{\sqrt{p}}{\sqrt{p}+\sqrt{1-p}} \\
\frac{\sqrt{p}}{\sqrt{p}+\sqrt{1-p}} \le \delta_B^* &\le \frac{1}{2}. \qquad (23)
\end{aligned}$$

As $\rho \to \infty$ (possible in the tight analysis technique), the boundaries of δ_A^* and δ_B^* are the same. Using this we have

for $A(\delta), B(\delta)$:

$$\begin{aligned}
&\max_{\delta\in\mathcal{G}_R} A(\delta) \\
&\begin{cases} \rho\left(R+h(\delta_A^*)-\log(2)-\beta\frac{1}{1+\rho}\delta_A^*\right) & \delta_A^* > \delta_{GV}(R) \\ -\beta\frac{\rho}{1+\rho}\delta_{GV} & \delta_A^* < \delta_{GV}(R) \end{cases}, \\
&\max_{\delta\in\mathcal{G}_R^c} B(\delta) \\
&= \begin{cases} -\beta\frac{\rho}{1+\rho}\delta_{GV} & \delta_B^* > \delta_{GV}(R) \\ R+h(\delta_B^*)-\log(2)-\beta\frac{\rho}{1+\rho}\delta_B^* & \delta_B^* < \delta_{GV}(R) \end{cases}.
\end{aligned} \tag{24}$$

The relations between $\max_\delta A(\delta)$, $\max_{\delta\in\mathcal{G}_R} A(\delta)$, $\max_{\delta\in\mathcal{G}_R^c} B(\delta)$ changes when $R > R_c$ or $R < R_c$, where $R_c = \log(2) - h(\frac{\sqrt{p}}{\sqrt{p}+\sqrt{1-p}})$ is the critical rate (the minimal R for which Gallager's bound agrees with the sphere packing bound). Since there is no dependence here on the channel output, $\boldsymbol{y}$, R_c corresponds to the critical rate of comment 3 of the previous section. When $R > R_c$, $\delta_{GV}(R) \leq \frac{\sqrt{p}}{\sqrt{p}+\sqrt{1-p}}$. Hence for some $\rho_0 > 0, \delta_A = \delta_{GV}(R)$ and the global maximizer of $A(\delta)$ is in $\mathcal{G}_R$ for all $\rho > \rho_0$. This means that Jensen's inequality is tight for all $\rho > \rho_0$. When $R < R_c$, $\delta_{GV}(R) > \frac{\sqrt{p}}{\sqrt{p}+\sqrt{1-p}}$. Hence, the global maximizer of $A(Q)$ is always in $\mathcal{G}_R^c$ and $\max_{\delta\in\mathcal{G}_R^c} B(\delta) \geq \max_{\delta\in\mathcal{G}_R} A(\delta)$. The global maximum of $A(\delta)$ for $0 < \rho < 1$ is strictly larger than $\max_{\delta\in\mathcal{G}_R^c} B(\delta)$, thus Jensen's inequality is not tight in all this range. $\max_\delta A(\delta)$ and $\max_{\delta\in\mathcal{G}_R^c} B(\delta)$ agree only in the two points in which Jensen's inequality is trivial, i.e $\rho = 0$ and $\rho = 1$.

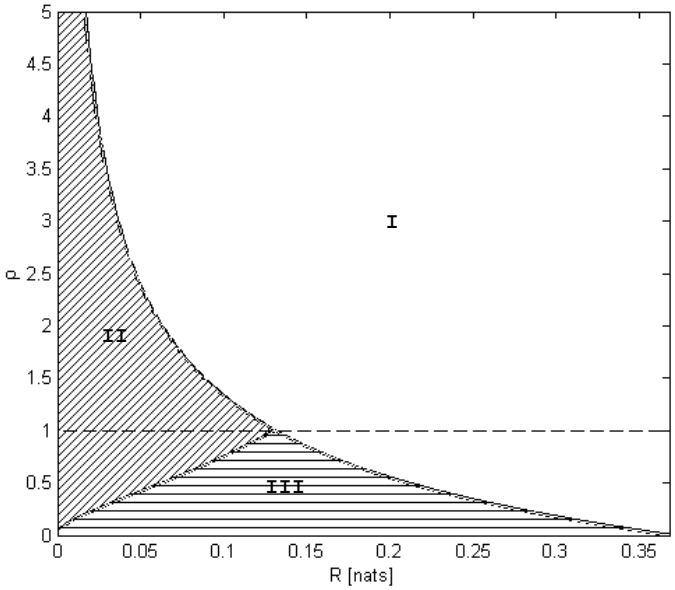

Fig. 2. Regions with different behavior of the exponent in the $R-\rho$ plane.

In Fig.2 we plot the relationship between $\max_{\delta\in\mathcal{G}_R} A(\delta)$ and $\max_{\delta\in\mathcal{G}_R^c} B(\delta)$ in the $R-\rho$ as explained in remark 5 of the previous section. in the white region (I) of Fig.2 $\max_\delta A(\delta) = \max_{\delta\in\mathcal{G}_R} A(\delta) > \max_{\delta\in\mathcal{G}_R^c} B(\delta)$ meaning that Jensen's inequality is tight (both as an upper and a lower bound, depending on whether $\rho > 1$ or $\rho < 1$). In the region marked by diagonal lines (II) we have $\max_{\delta\in\mathcal{G}_R} A(\delta) < \max_{\delta\in\mathcal{G}_R^c} B(\delta)$ and in the region marked by horizontal lines (III) we have $\max_{\delta\in\mathcal{G}_R} A(\delta) = \max_{\delta\in\mathcal{G}_R^c} B(\delta)$. In regions II and III, Jensen's inequality is not tight. As explained in Remark 5, the reason Gallager's bound is tight is that, luckily, when taking into account the exponents of $(1-p)^{n\frac{1}{1+\rho}}$ and the first expectation, the overall maximizing parameter is $\rho = 1$ which is also the maximizer of Gallager's bound and a point at which Jensen's inequality is tight.

V. Conclusion

We demonstrated a tight analysis of expressions having the form of (1), using Gallager's bound as an example. We showed that analysis through Jensen's inequality might not tight in some regions of both the number of terms in the sum of (1) (the rate) and the parameters. Although in Gallager's case the analysis is indeed tight, there are cases where it is not, as demonstrated in [4], [10]. The analysis we showed here may serve as a yardstick for examining the tightness in every given specific problem (with a specific f in (1)) of this type.

References

[1] R. G. Gallager, *Information Theory and Reliable Communication*. Wiley, 1968.

[2] G. D. Forney, "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Transactions on Information Theory*, vol. 14, no. 2, pp. 206–220, March 1968.

[3] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredachi Informatsii*, vol. 10, no. 3, pp. 3–14, 1974.

[4] N. Merhav, "Error exponents of erasure/list decoding revisited via moments of distance enumerators," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4439–4447, October 2008.

[5] R. G. Gallager, "The random coding bound is tight for the average code," *IEEE Transactions on Information Theory*, vol. 19, pp. 244–246, 1973.

[6] M. Mezard and A. Montanari, *Constraint Satisfaction Networks in Physics and Computation*. Clarendon Press.

[7] N. Merhav, *Information Theory and Statistical Physics - Lecture Notes*, http://arxiv.org/abs/1006.1565, 2010.

[8] ——, "Lecture notes on information theory and statistical physics," *submitted to Foundations and Trends in Communications and Information Theory*, June 2010.

[9] R. Etkin, N. Merhav, and E. Ordentlich, "Error exponents of optimum decoding for the interference channel," in *Proceeding of the International Symposium on Information Theory*, 2008, pp. 1523–1527.

[10] Y. Kaspi and N. Merhav, "Error exponents for broadcast channels with degraded message sets," *accepted to IEEE Transactions on Information Theory*, 2010.

[11] Y. Kaspi, "Error exponents for broadcast channels with degraded message sets," Master's thesis, Department of Electrical Engineering, Technion - Israel Institute of Technology, Haifa, Israel, April 2009.