Random Coding Bounds that Attain the Joint Source-Channel Exponent

Adrià Tauste Campo, Gonzalo Vazquez-Vilar, Albert Guillén i Fàbregas
tobias.koch@eng.cam.ac.uk, tobi.koch@eng.cam.ac.uk
1University of Cambridge, 2Universitat Pompeu Fabra, 3Institució Catalana de Recerca i Estudis Avançats (ICREA)
Email: {atauste,gvazquez,guillen,alfonso.martinez}@ieee.org, tobi.koch@eng.cam.ac.uk

Abstract—This paper presents a random-coding upper bound on the average error probability of joint source-channel coding that attains Csiszár’s error exponent. The bound is based on a code construction for which source messages are assigned to disjoint subsets (classes), and codewords generated according to a distribution that depends on the class of the source message. For a single class, the bound recovers Gallager’s exponent; identifying the classes with source type classes, it recovers Csiszár’s exponent. Moreover, it is shown that as a two appropriately designed classes are sufficient to attain Csiszár’s exponent.

I. INTRODUCTION

We study the problem of transmitting a length-$k$ discrete memoryless source over a discrete memoryless channel using length-$n$ block codes. The source is characterized by a distribution $P_Y(y) = \prod_{i=1}^{k} P_Y(v_i)$, $v = (v_1, \ldots, v_k) \in \mathcal{V}^k$, where $\mathcal{V}$ is a discrete alphabet with cardinality $|\mathcal{V}|$. The channel law is given by a conditional probability distribution $P_{Y|X}(y|x) = \prod_{i=1}^{n} P_{Y|X}(y_i|x_i)$, $x = (x_1, \ldots, x_n) \in \mathcal{X}^n$, where $\mathcal{X}$ and $\mathcal{Y}$ are discrete alphabets with cardinalities $|\mathcal{X}|$ and $|\mathcal{Y}|$, respectively.

In this joint source-channel coding (JSCC) setup, the encoder maps the source message $v$ to a length-$n$ codeword $x(v)$, which is then transmitted over the channel. We refer to the ratio $t = k/n$ as the transmission rate. Based on the channel output $y$, the decoder guesses a source message $\hat{v}$ according to the maximum a posteriori (MAP) criterion, i.e.,

$$\hat{v} = \arg \max_v P_Y(v)P_{Y|X}(y|x(v)).$$

(1)

When clear from the context, we shall simplify notation by writing $x$ instead of $x(v)$, making the message $v$ implicit.

We study the random-coding average error probability $\bar{\epsilon}$ by means of the random-coding union (RCU) bound [1, 2]:

$$\bar{\epsilon} \leq E \left[ \min \left\{ 1, \sum_{v \neq \mathcal{V}} \Pr \left( \frac{P_Y(v)P_{Y|X}(Y|X) \geq 1}{P_Y(V)P_{Y|X}(Y|X)} \right) \right\} \right],$$

(2)

where the expectation is taken according to the joint distribution $P_Y P_{Y|X} P_{Y|X}$ and the probability computed with respect to the distribution $P_{Y|X}$ for each $v'$ in the summation.

Optimization over the conditional distributions $P_{Y|X}$, necessary to obtain the tightest possible bound in (2), quickly becomes computationally unfeasible as the block length grows large. In this paper, we focus instead on the exponential decay of (2) with respect to $n$ and slightly loosen the RCU bound into a convenient form that is proven to attain the JSCE exponent found by Csiszár [3].

II. PREVIOUS WORK

In [4, Prob. 5.16] Gallager provided an upper bound on $\bar{\epsilon}$ when the codewords corresponding to different source messages are drawn independently according to a distribution $P_X$:

$$\bar{\epsilon} \leq e^{-E_0(\rho, P_Y|X, P_X) + E_0(\rho, P_Y)},$$

(3)

where $E_0(\rho, P_Y|X, P_X)$ denotes Gallager’s channel function,

$$E_0(\rho, P_Y|X, P_X) \triangleq -\log \left( \sum_{x} P_X(x) P_{Y|X}(y|x)^{1+\rho} \right)$$

(4)

and where $E_0(\rho, P_Y)$ denotes Gallager’s source function,

$$E_0(\rho, P_Y) \triangleq \log \left( \sum_{v} P_Y(v)^{1+\rho} \right).$$

(5)

While the derivation of (3) assumes that $P_X$ is independent of the source message, it may be proved that considering an arbitrary distribution $P_X|Y$ does not improve the bound.

When $P_X$ is a product distribution, i.e., $P_X(x) = \prod_{i=1}^{n} P_X(x_i)$, the bound in (3) becomes

$$\bar{\epsilon} \leq e^{-n \left( E_0(\rho, P_Y|X, P_X) - t E_0(\rho, P_Y) \right)},$$

(6)

thus proving that $\bar{\epsilon}$ exponentially vanishes with respect to $n$. By maximizing over $P_X$ and $\rho$, this bound provides a lower bound on the JSCE exponent $E_1$

$$E_1 \geq E_1^0 \triangleq \max_{\rho \in [0,1]} \left\{ E_0(\rho, P_Y|X) - t E_0(\rho, P_Y) \right\},$$

(7)

where we define $E_0(\rho, P_Y|X) \triangleq \max_{P_X} E_0(\rho, P_Y|X, P_X)$.

Csiszár refined Gallager’s result using a code construction based on fixed composition codes [3]. Specifically, he showed
that for all $\delta > 0$, there exists an $n_0 \in \mathbb{N}$ such that, for $n \geq n_0$
\[ \varepsilon \leq \frac{M_k}{n} e^{-n(\varepsilon(R_i, P_{Y^j}) + E_r(R_i, P_{Y^j}|X) - 2\delta)}, \]
(8)
where $M_k$ is the number of source-type classes in $\mathcal{V}_k$ [5], $e(R, P_{Y^j})$ is the source reliability function [5]–[7]
\[ e(R, P_{Y^j}) \triangleq \sup_{\rho \geq 0} \{ \rho R - E_o(\rho, P_{Y^j}) \}, \]
(9)
and where $E_i(R, P_{Y^j}|X)$ is the channel random-coding exponent, given by [4]
\[ E_i(R, P_{Y^j}|X) \triangleq \max_{\rho \in [0, 1]} \{ E_0(\rho, P_{Y^j}|X) - \rho R \}. \]
(10)
Eq. (8) leads to another lower bound on the exponent $E_j$,
\[ E_j \geq E_j^{Cs}, \]
(11)
\[ \triangleq \min_{h(\mathcal{V}_j) \leq R \leq R_c} \left\{ t e \left( \frac{R}{t}, P_{Y^j} \right) + E_i(R, P_{Y^j}|X) \right\}, \]
(12)
where $R_c \triangleq t \log |\mathcal{V}_j|$, and $E_i(R, P_{Y^j}|X)$ is the concave hull of $E_0(\rho, P_{Y^j}|X)$, $E_j^{Cs}$ gives the actual ISCC error exponent $E_j$ when the minimum in (12) is attained at a rate $R \geq R_c$, where $R_c$ is the critical rate of the channel [3].
Zhong et al. [9] quantified the improvement of Csiszár’s exponent (12) over Gallager’s (7) via Fenchel’s duality theorem [10, Thm. 31.1], which allows one to rewrite (12) as
\[ E_j^{Cs} = \max_{\rho \in [0, 1]} \{ E_0(\rho, P_{Y^j}|X) - t E_o(\rho, P_{Y^j}) \}, \]
(13)
where $E_0(\rho, P_{Y^j}|X)$ denotes the concave hull of $E_0(\rho, P_{Y^j}|X)$, defined as the pointwise inifimum over the family of affine functions that upper-bound $E_0(\rho, P_{Y^j}|X)$, as a function of $\rho$ in $\mathbb{R}$. [10, Cor. 12.1.1]. It follows that $E_j^{Cs} \geq E_j^{G}$, with the inequality possibly strict, as shown in an example in [9].

III. RANDOM CODING BOUND

We have recently proposed a random-coding upper bound which attains Csiszár’s exponent [11]. The derivation of this bound involves the following steps:

1) Define a partition $\mathcal{P}_k$ of the message set $\mathcal{V}_k$ into $N_k$ disjoint subsets $A_1, \ldots, A_{N_k}$ satisfying $\bigcup_{i=1}^{N_k} A_i = \mathcal{V}_k$.
   We shall refer to these subsets as classes.
2) Assign a channel input distribution $P_{X(i)}$ to each class $A_i$. Then, for each source message $v \in A_i$ randomly and independently generate codewords $x(v) \in X^n$ according to $P_{X(i)}$.
3) Upper-bound the probability of error using Gallager’s bounding techniques [4].

In the following we define
\[ E_{X(i)}^{(i)}(\rho, P_{V}) \triangleq \log \left( \sum_{x \in A_i} P_{V}(v) \right)^{1+\rho}, \]
(14)
for $i = 1, \ldots, N_k$.

Theorem 1: For every partition $\mathcal{P}_k$, for every set of product channel input distributions $P_{X(i)}^{(i)}(x) = \prod_{i=1}^{N_k} P_{X(i)}^{(i)}(x_i)$, $i = 1, \ldots, N_k$, and for every set of parameters $\rho_1, \ldots, \rho_{N_k} \in [0, 1]$, the average probability of error is upper-bounded by
\[ \tilde{\varepsilon} \leq \tilde{\varepsilon}_B(\mathcal{P}_k) \]
\[ \triangleq h(k) \sum_{i=1}^{N_k} \max_{\rho \in [0, 1]} \{ n E_0(\rho_i, P_{Y^j}|X) - E_{X(i)}^{(i)}(\rho_i, P_{V}) \}, \]
(16)
where $h(k) \triangleq 2N_k(k+1)^{\nu}(k+1)^{\nu}$. Proof: See [11].

If we choose the partition $\mathcal{P}_k$ such that $N_k = 1$ and $A_1 = \mathcal{V}_k$ for $k = 1, 2, \ldots$, then $E_{X(i)}^{(i)}(\rho_i, P_{V}) = E_{X(i)}^{G}(\rho_i, P_{V})$ and $\log h(k)/k \to 0$ as $n \to \infty$. Hence, (15) recovers Gallager’s bound on the error exponent (7)
\[ \lim_{n \to \infty} - \frac{1}{n} \log \tilde{\varepsilon}_B(\mathcal{P}_k) = E_j^G. \]
(17)

With a more judicious choice of $\mathcal{P}_k$ the upper bound (15) also recovers Csiszár’s lower bound on the error exponent (12). Specifically, (12) can be achieved by identifying the classes $A_1, \ldots, A_{N_k}$ with the source-type classes $T_1, \ldots, T_{N_k}$.

A. Attaining Csiszár’s exponent with two classes

Let $\mathcal{T}(v)$ denote the source-type class associated to message $v$. We define the partition $\mathcal{P}_k(\lambda_0)$ as follows. For some $\lambda_0 \geq 0$ and every $k \geq 1$, we assign the source messages into two sets, $A_1$ and $A_2$, respectively defined as
\[ A_1 \triangleq \{ v : |\mathcal{T}(v)| \geq e^{tk_0} \lambda_0 \}, \]
(20)
\[ A_2 \triangleq \{ v : |\mathcal{T}(v)| < e^{tk_0} \lambda_0 \}. \]
(21)

Then, we have the following theorem.

Theorem 2: Consider the family of partitions $\{ \mathcal{P}_k = \mathcal{P}_k(\lambda_0), \lambda_0 \in [0, R_c] \}$ for every $k \geq 1$. Then,
\[ \lim_{\lambda_0 \to 0} \sup_{0 \leq v \leq R_c} \lim_{n \to \infty} - \frac{1}{n} \log \tilde{\varepsilon}_B(\mathcal{P}_k(\lambda_0)) \geq E_j^{Cs}. \]
(22)

Proof: See Section V.

The division of $\mathcal{V}_k$ into the classes $A_1$ and $A_2$ for the best threshold $\lambda_0^*$, together with the optimal distributions $P_{X(i)}^{(i)}$ and $P_{X}^{(2)}$, induce a conditional distribution $P_{X|V}(x|v)$ given by
\[ P_{X|V}(x|v) = \begin{cases} \prod_{i=1}^{n} P_{X(i)}^{(i)}(x_i), & v \in A_1, \\ \prod_{i=1}^{n} P_{X}^{(2)}(x_i), & v \in A_2, \end{cases} \]
(23)
By construction, the new random coding bound (16) is looser with respect to the RCU bound (2) for all \( n \) and \( P_X(1) = 0.041 \). Therefore, the source entropy is \( H(V) = 0.2469 \) bits/source symbol, the channel capacity is \( C = 0.9468 \) bits/channel use, and the critical rate is \( R_c = 0.4564 \) bits/channel use.

As Gallager observed, optimizing the \( E_0(\rho, P_{Y|X}, P_X) \) function over the input distribution may lead to a discontinuity of the derivative of the function \( E_0(\rho, P_{Y|X}) \) with respect to \( \rho \). In this example, the optimal distribution abruptly changes from \( P_X^{(1)} = \left( \frac{1}{4} \frac{1}{4} \frac{1}{4} \frac{1}{4} \right) \) to \( P_X^{(2)} = \left( \frac{1}{2} \frac{1}{2} \frac{1}{2} \frac{1}{2} \right) \) for some \( \rho \in [0, 1] \). In turn, this implies that \( E_0(\rho, P_{Y|X}) \) is not concave. Therefore, for some \( \rho \), \( E_0(\rho, P_{Y|X}) > E_0(\rho, P_{Y|X}) \).

In Fig. 1 we plot several bounds on the JSCC exponent based on the aforementioned random-coding bounds. For Gallager and Csiszár exponents we use the arguments in (7) and (13) as a function of \( \rho \) respectively. The figure shows how the non-concavity of Gallager’s exponent function around the optimal \( \rho \) of Csiszár’s exponent translates into a loss in exponent. For reference purposes, Fig. 1 shows the value of Csiszár’s exponent with a horizontal dashed line. For the random-coding bound (16) with two classes, we apply Theorem 2 and plot the exponent of each individual class as a function of \( \rho \). The optimum value of \( \lambda_0 \) that determines the partition is given by \( \lambda_0 = 0.732 \). The resulting exponent is obtained by first individually maximizing the exponent of each class over \( \rho \), and then choosing the minimum. The figure illustrates that for the optimal partition, the exponent of both classes coincides with Csiszár’s.

In Fig. 2 we compare Gallager’s upper bound (6), the RCU bounds with a single input distribution (either \( P_X^{(1)} \) or \( P_X^{(2)} \)) and the RCU bound when the two-class construction is employed and each class distribution independently optimized. While the single-class RCU bound is tighter than Gallager’s bound, it attains the same asymptotic slope. The two-class construction achieves a tighter upper bound on the random coding error probability. In this example, Csiszár’s exponent (12) is attained at \( R^* = 0.7273 > R_c \) and thus, \( E_1^{\text{sc}} \) is tight and gives the JSCC exponent.

V. PROOF OF THEOREM 2

This section provides proof of Theorem 2. To this end we first introduce two lemmas which are then used in the derivation of the main result.

Consider the partition \( \bar{P}_k(\lambda_0) \) defined by (20)-(21). By noting that \( E_k^{(i)}(\rho) \triangleq E_k^{(i)}(\rho, P_Y) \) is a continuous first-order differentiable function of \( \rho \) we define, for \( i = 1, 2 \),

\[
\lambda_i(\rho) \triangleq \frac{1}{n} \frac{\partial E_k^{(i)}(\rho)}{\partial \rho}.
\]

For future reference, we also define

\[
F_1(\lambda, \rho) \triangleq \max_{\rho_1 \in [0, \rho]} \left\{ E_0(\rho_1, P_{Y|X}) + \lambda(\rho - \rho_1) \right\}, \quad \text{and} \quad F_2(\lambda, \rho) \triangleq \max_{\rho_2 \in [\rho, 1]} \left\{ E_0(\rho_2, P_{Y|X}) + \lambda(\rho - \rho_2) \right\}.
\]

---

Figure 1. Error exponent bounds. Csiszár’s and Gallager’s curves correspond to \( E_0(\rho, P_{Y|X}) - tE_1(\rho, P_Y) \) and \( E_0(\rho, P_{Y|X}) - tE_2(\rho, P_Y) \), respectively. Single class correspond to \( E_0(\rho, P_{Y|X}) - \lim_{n \to \infty} \frac{1}{n} E_0^{(1)}(\rho, P_Y) \), for \( i = 1, 2 \).

Figure 2. Random coding upper bounds to the error probability \( \varepsilon \).
Lemma 1: For every $\rho \in [0,1]$, the limits $\lim_{n \to \infty} \lambda_i(\rho)$, $i = 1, 2$ exist. Furthermore, for every $0 \leq \lambda_0 \leq R_V$ we have

$$\lim_{n \to \infty} \lambda_2(1) \leq \lambda_0 \leq \lim_{n \to \infty} \lambda_1(0).$$

(28)

Lemma 2: For $\rho \in [\hat{\rho}, 1]$ where $\hat{\rho}$ is the smallest value $\hat{\rho}$ that satisfies $\hat{\rho}_c = \arg \max_{\rho' \in [0,1]} \{ E_0(\rho', P_Y|X) - \rho'R_V \}$, it holds that

$$\min_{\lambda \in [0,R_V]} \{ F_i(\lambda, \rho) \} = \min_{\lambda \in [0,R_V]} \{ F_i(\lambda, \rho) \}. $$

(29)

Armed with the above two lemmas, we proceed to prove the lower bound (22). Since $E_i(\rho) \geq \rho$, we can bound $E_i(\rho)$, for $i = 1, 2$ and $\rho \in [0,1]$, as

$$E_i(\rho) \geq E_i(1) + n\lambda_i(1) - \rho - \rho_1$$

(30)

$$E_i(\rho) \geq E_i(1) + n\lambda_i(0) - \rho - \rho_1,$$  

(31)

$$E_i(\rho) \geq E_i(2) + n\lambda_i(2) - \rho - \rho_2,$$  

(32)

$$E_i(\rho) \geq E_i(2) + n\lambda_i(1) - \rho - \rho_2,$$  

(33)

Since the limit $\lim_{n \to \infty} \frac{1}{n} E_i(\rho)$ exists for $\rho \in [0,1], Lemma 1, together with (31) and (33), shows that

$$\lim_{n \to \infty} \frac{1}{n} E_i(\rho) \geq \lim_{n \to \infty} \frac{1}{n} E_i(1) + \lambda_0(\rho_0 - \rho),$$  

(34)

$$\lim_{n \to \infty} \frac{1}{n} E_i(\rho) \geq \lim_{n \to \infty} \frac{1}{n} E_i(2) + \lambda_0(\rho_0 - \rho_2).$$  

(35)

for $\rho_1 \in [0, \rho_0]$ and $\rho_2 \in [\rho_0, 1]$. Then, we have that, for an arbitrary dummy variable $\rho_0 \in [0,1],$

$$\lim_{n \to \infty} \frac{1}{n} \log \epsilon_B(\hat{\rho}_k(\lambda_0))\|$$

(36)

where (37) follows by noting that $h(tn)$ is subexponential in $n$; in (38) we used that the limit exists for each set of the parameters over which the optimization is performed; (39) follows from restricting the intervals over which $\rho_1$ and $\rho_2$ are maximized; in (40) we applied (34)-(35) and used the definition of $F_i(\lambda, \rho), i = 1, 2$, cf. (26)-(27); and in (41) we used that, by definition, $E_i(\rho) \leq E_i(\rho, P_Y) = kE_i(\rho, P_Y)$ since the sum is taken over a smaller number of terms.

As (36)-(41) hold for arbitrary $\rho_0 \in [\hat{\rho}, 1]$ (with $\hat{\rho}$ defined in Lemma 2), we obtain upon maximizing over $\lambda \in [0, R_V]$ and $\rho_0 \in [\hat{\rho}, 1]$

$$\sup_{\lambda \in [0,R_V]} \lim_{n \to \infty} \frac{1}{n} \log \epsilon_B(\hat{\rho}_k(\lambda_0))\|$$

(42)

$$\geq \max_{\rho_0 \in [\hat{\rho}, 1]} \left\{ \max_{\lambda \in [0,R_V]} \left\{ \min_{i=1,2} \{ F_i(\lambda, \rho_0) - tE_i(\rho_0, P_Y) \} \right\} \right\}$$

(43)

$$= \max_{\rho_0 \in [\hat{\rho}, 1]} \left\{ \min_{\lambda \in [0,R_V]} \left\{ \max_{i=1,2} \{ F_i(\lambda, \rho_0) - tE_i(\rho_0, P_Y) \} \right\} \right\}$$

(44)

$$\geq \max_{\rho_0 \in [\hat{\rho}, 1]} \left\{ \max_{\lambda \geq 0 \in [0,1]} \left\{ \max_{\rho \in [0,1]} \{ E_0(\rho, P_Y|X) + \lambda_0(\rho_0 - \rho) \} \right\} - tE_i(\rho_0, P_Y) \right\}$$

(45)

$$= \max_{\rho_0 \in [\hat{\rho}, 1]} \left\{ E_0(\rho_0, P_Y|X) - tE_i(\rho_0, P_Y) \right\}$$

(46)

where (42) we used that $F_i(\lambda, \rho), i = 1, 2$ are continuous functions of $\lambda$ to write a maximum instead of a supremum; in (43) we applied Lemma 2; (44) follows from the fact that $\max_{i=1,2} \max_{\rho \in [0,1]} f(x) = \max_{\rho \in [0,1], f(x)} f(x); \text{in (45) we relaxed the range over which } \lambda_0 \text{ is optimized}; \text{and finally, (46) follows from the fact that the convex hull of a function can be written as the double conjugate of the original function [10, Thm 12.2]. The concave hull is then the negative of the double conjugate of the negative of the original function.}

In order to conclude the proof it remains to show that the range of $\rho_0$ over which the argument of (46) is optimized can be extended to $\rho_0 \in [0,1]$ without violating the inequality chain (42)-(46). We prove it by contradiction. To this end, let us consider a $\rho_0^*$ that satisfies

$$\rho_0^* = \arg \max_{\rho_0 \in [0,1]} \{ E_0(\rho_0, P_Y|X) - tE_i(\rho_0, P_Y) \},$$

(47)

and assume that $\rho_0^* < \hat{\rho}.$

From the definition of the concave hull, it follows that the smallest value of $\rho^*$ that satisfies $\rho^* = \arg \max_{\rho \in [0,1]} \{ E_0(\rho, P_Y|X) - \rho R_V \}$ equals $\hat{\rho}.$ Then we have
\[ E_0(\rho_0^*, P_{Y|X}) - tE_s(\rho_0, P_V) \]
\[ = \tilde{E}_0(\rho_0^*, P_{Y|X}) - \rho_0^* R_V + \rho_0^* R_V - tE_s(\rho_0^*, P_V) \tag{48} \]
\[ < \tilde{E}_0(\hat{\rho}, P_{Y|X}) - \hat{\rho} R_V + \rho_0^* R_V - tE_s(\rho_0^*, P_V) \tag{49} \]
\[ \leq \tilde{E}_0(\hat{\rho}, P_{Y|X}) + R_V(\rho_0^* - \hat{\rho}) - t \left( E_s(\hat{\rho}, P_V) + \frac{\partial E_s(\rho, P_V)}{\partial \rho}(\rho_0^* - \hat{\rho}) \right) \tag{50} \]
\[ \leq \tilde{E}_0(\hat{\rho}, P_{Y|X}) - tE_s(\hat{\rho}, P_V) \tag{51} \]

where in (49) we used the definition of \( \hat{\rho} \); (50) follows from the convexity of \( E_s(\rho, P_V) \); and in (51) we used that

\[
\frac{\partial E_s(\rho, P_V)}{\partial \rho} \leq \lim_{\rho' \to \infty} \frac{\partial E_s(\rho, P_V)}{\partial \rho} \bigg|_{\rho = \rho'} = \frac{R_V}{t}. \tag{52}
\]

From (48)-(51) it follows that by choosing \( \rho_0 = \hat{\rho} \) we would achieve an objective strictly larger than by choosing \( \rho_0 = \rho_0^* \), hence contradicting the initial assumption.

It thus follows that

\[
\max_{\rho_0 \in [\hat{\rho}, 1]} \left\{ \tilde{E}_0(\rho_0, P_{Y|X}) - tE_s(\rho_0, P_V) \right\} = \max_{\rho_0 \in [0, 1]} \left\{ \tilde{E}_0(\rho_0, P_{Y|X}) - tE_s(\rho_0, P_V) \right\}. \tag{53}
\]

Since (53) is equal to (13), this concludes the proof.