Converse bounds for interference channels via coupling and proof of Costa's conjecture

Yury Polyanskiy and Yihong Wu

***Abstract*—It is shown that under suitable regularity conditions, differential entropy is $O(\sqrt{n})$-Lipschitz as a function of probability distributions on $\mathbb{R}^n$ with respect to the quadratic Wasserstein distance. Under similar conditions, (discrete) Shannon entropy is shown to be $O(n)$-Lipschitz in distributions over the product space with respect to Ornstein's $\bar{d}$-distance (Wasserstein distance corresponding to the Hamming distance). These results together with Talagrand's and Marton's transportation-information inequalities allow one to replace the unknown multi-user interference with its i.i.d. approximations. As an application, a new outer bound for the two-user Gaussian interference channel is proved, which, in particular, settles the "missing corner point" problem of Costa (1985).**

I. Introduction

Arguably, a key novel effect in multi-user information theory is multi-user interference, where one user's codebook creates complicated non-i.i.d. disturbance for other users. A convenient workaround would be to have rigorous approximation results allowing replacing a complicated non-i.i.d. interference with a simpler i.i.d. one. Such approximation is the key contribution of this paper.

As a concrete example, we consider the so-called "missing corner point" problem in the capacity region of the two-user Gaussian interference channels (GIC) [1], which has attracted renewed attention recently as witnessed by [2]–[5] and Sason's comprehensive treatment in [6].

Mathematically, the key question for settling "missing corner point" is the following: given independent n-dimensional random vectors X_1, X_2, G_2, Z with the latter two being Gaussian, is it true that

$$D(P_{X_2+Z} \| P_{G_2+Z}) = o(n) \tag{1}$$
$$\stackrel{?}{\Rightarrow} |h(X_1 + X_2 + Z) - h(X_1 + G_2 + Z)| = o(n).$$

This paper proves that indeed under suitable regularity conditions, the difference in entropy (in both continuous and discrete cases) can be bounded by the *Wasserstein distance*, a notion originating from optimal transportation theory which turns out to be the main tool of this paper. The Wasserstein distance, in turn, can be bounded by Kullback-Leibler divergence by virtue of Marton's and Talagrand's information-transportation inequalities [7], [8].

We start with the definition of the Wasserstein distance on the Euclidean space. Given probability measures P, Q on $\mathbb{R}^n$, define their p-Wasserstein distance ($p \geq 1$) as

$$W_p(P, Q) \triangleq \inf(\mathbb{E}[\|X - Y\|^p])^{1/p}, \tag{2}$$

where $\|\cdot\|$ denotes the Euclidean distance and the infimum is taken over all couplings of P and Q, i.e., joint distributions P_{XY} whose marginals satisfy $P_X = P$ and $P_Y = Q$. The following dual representation of the W_1 distance is useful:

$$W_1(P, Q) = \sup_{\mathrm{Lip}(f) \leq 1} \int f dP - \int f dQ. \tag{3}$$

It is easy to see that in order to control $|h(X) - h(\tilde{X})|$ by means of $W_2(P_X, P_{\tilde{X}})$, one necessarily needs to assume some regularity properties of P_X and $P_{\tilde{X}}$; otherwise, choosing one to be a fine quantization of the other creates an infinite gap between differential entropies, while keeping the W_2-distance arbitrarily small. Our main result in Section II shows that under moment constraints and certain conditions on the densities (which are in particular satisfied by convolutions with Gaussians), various information measures such as differential entropy and mutual information on $\mathbb{R}^n$ are in fact $\sqrt{n}$-Lipschitz continuous with respect to the W_2-distance. These results have natural counterparts in the discrete case, where the Euclidean distance is replaced by the Hamming distance (Section IV).

Furthermore, *transportation-information inequalities*, such as those due to Marton [7] and Talagrand [8], allow us to bound the Wasserstein distance by the KL divergence. For example, Talagrand's inequality states that if $Q = \mathcal{N}(0, \Sigma)$, then

$$W_2^2(P, Q) \leq \frac{2\sigma_{\max}(\Sigma)}{\log e} D(P\|Q)\,, \tag{4}$$

where $\sigma_{\max}(\Sigma)$ denotes the maximal singular value of the covariance matrix Σ. Invoking (4) in conjunction with the Wasserstein continuity of the differential entropy, we establish (1) and prove a new outer bound for the capacity region of the two-user GIC, finally settling the missing corner point in [1]. See Section III for details.

Y.P. is with the Department of EECS, MIT, Cambridge, MA, email: yp@mit.edu. Y.W. is with the Department of ECE and the Coordinated Science Lab, University of Illinois at Urbana-Champaign, Urbana, IL, email: yihongwu@illinois.edu. This research has been supported in part by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-09-39370, by the NSF CAREER award CCF-12-53205, by NSF grants IIS-14-47879, CCF-14-23088 and CCF-15-27105 and by the Simons Institute for the Theory of Computing. We are also indebted to Prof. Chandra Nair who brought this question to our attention.

978-1-5090-1805-5/16/$31.00 ©2016 IEEE

Notations: Throughout this paper $\log$ is with respect to an arbitrary base, which also specifies the units of the differential entropy $h(\cdot)$, Shannon entropy $H(\cdot)$, mutual information $I(\cdot;\cdot)$, and divergence $D(\cdot\|\cdot)$. The natural logarithm is denoted by ln. The norm of $x \in \mathbb{R}^n$ is denoted by $\|x\| \triangleq (\sum_{j=1}^n x_j^2)^{1/2}$. For random variables X and Y, let $X \perp\!\!\!\perp Y$ denote their independence.

Proofs: A full version of this work containing all proofs and extensions is available in [9].

II. Wasserstein-continuity of entropy

Proposition 1. *Let B satisfy $\|B\| \le \sqrt{nP}$ (a.s.), $G \sim \mathcal{N}(0, \sigma_G^2 I_n)$ be independent of B and $V = B + G$. For any U, we have*

$$h(U) - h(V) \le \frac{\log e}{2\sigma_G^2}\left(\mathbb{E}[\|U\|^2] - \mathbb{E}[\|V\|^2]\right) + \frac{\log e}{\sigma_G^2}\sqrt{nP}W_1(P_U, P_V)\,. \quad (5)$$

Proof. First notice that the density p_V of V satisfies

$$\nabla \log p_V(v) = \frac{\log e}{\sigma_G^2}(\hat{B}(v) - v)\,, \quad (6)$$

where $\hat{B}(v) \triangleq \mathbb{E}[B|V = v] = \frac{\mathbb{E}[Bp_G(v-B)]}{\mathbb{E}[p_G(v-B)]}$ satisfies $\|\hat{B}(v)\| \le \sqrt{nP}$, since $\|B\| \le \sqrt{nP}$ almost surely (as in [10, Proof of Theorem 8]). Denoting κ the appropriate constant and $\bar{t} \triangleq 1 - t$, we get

$$\begin{aligned}
\log \frac{p_V(v)}{p_V(u)} &= \\
&= \int_0^1 dt\,\langle \nabla \log p_V(tv + \bar{t}u), v - u\rangle \quad (7)\\
&= \kappa \int_0^1 dt\langle \hat{B}(tv + \bar{t}u), v - u\rangle - \frac{\kappa}{2}(\|v\|^2 - \|u\|^2) \quad (8)\\
&\le \kappa\sqrt{nP}\|v - u\| - \frac{\kappa}{2}(\|v\|^2 - \|u\|^2)\,. \quad (9)
\end{aligned}$$

Taking the expectation of the last equation under the W_1-optimal coupling, we obtain (5) after noticing

$$h(U) - h(V) + D(P_U\|P_V) = \mathbb{E}\left[\log \frac{p_V(V)}{p_V(U)}\right]\,.$$

□

Corollary 2. *Let A, B, G, Z be mutually independent, with $G \sim \mathcal{N}(0, \sigma_G^2 I_n)$, $Z \sim \mathcal{N}(0, \sigma_Z^2 I_n)$, and B satisfying $\|B\| \le \sqrt{nP}$ (a.s.). Furthermore, assume $\mathbb{E}[A] = \mathbb{E}[B] = 0$ and $\mathbb{E}[\|A\|^2] = \mathbb{E}[\|G\|^2]$. Then, for every $c \in [0, 1]$, we have*

$$\begin{aligned}
&h(B + A + Z) - h(B + G + Z)\\
&\le \frac{\sqrt{2nP(\sigma_G^2 + c^2\sigma_Z^2)\log e}}{\sigma_G^2 + \sigma_Z^2}\sqrt{D(P_{A+cZ}\|P_{G+cZ})}\,.
\end{aligned}$$

Proof. First, notice that by definition the Wasserstein distance is non-increasing under convolutions, i.e., $W_2(P_1 * Q, P_2 * Q) \le W_2(P_1, P_2)$. Since $c \le 1$ and the Gaussian distribution is stable, we have

$$W_2(P_{B+A+Z}, P_{B+G+Z}) \le W_2(P_{A+cZ}, P_{G+cZ}),$$

and via Talagrand's inequality (4) for some $\kappa > 0$

$$W_2(P_{A+cZ}, P_{G+cZ}) \le \sqrt{\kappa D(P_{A+cZ}\|P_{G+cZ})}\,.$$

From here we apply Proposition 1 with G replaced by $G + Z$. □

III. Gaussian interference channels

A. New outer bound

Consider the two-user Gaussian interference channel (GIC)

$$\begin{aligned}
Y_1 &= X_1 + bX_2 + Z_1\\
Y_2 &= aX_1 + X_2 + Z_2\,,
\end{aligned} \quad (10)$$

with $a, b \ge 0$, $Z_i \sim \mathcal{N}(0, I_n)$, and a power constraint on the n-letter codebooks:

$$\|X_1\| \le \sqrt{nP_1}, \quad \|X_2\| \le \sqrt{nP_2} \quad \text{a.s.} \quad (11)$$

Denote by $\mathcal{R}(a, b)$ the capacity region of the GIC in (10). As an application of the results developed in Section II, we prove an outer bound for the capacity region.

Theorem 3. *Let $0 < a \le 1$, $C_2 = \frac{1}{2}\log(1 + P_2)$, and $\tilde{C}_2 = \frac{1}{2}\log(1 + \frac{P_2}{1+a^2P_1})$. Then, for any $b \ge 0$ and $\tilde{C}_2 \le R_2 \le C_2$, any rate pair $(R_1, R_2) \in \mathcal{R}(a, b)$ satisfies*

$$R_1 \le \frac{1}{2}\log\min\left\{A - \frac{1}{a^2} + 1, R_c\right\}\,, \quad (12)$$

where

$$R_c = A\frac{(1 + P_2)(1 - (1 - a^2)\exp(-2\delta)) - a^2}{P_2}\,, \quad (13)$$

$$A = (P_1 + a^{-2}(1 + P_2))\exp(-2R_2), \quad (14)$$

$$\delta = C_2 - R_2 + a\sqrt{\frac{2P_1(C_2 - R_2)\log e}{1 + P_2}}. \quad (15)$$

Consequently, $R_2 \ge C_2 - \epsilon$ implies that $R_1 \le \frac{1}{2}\log(1 + \frac{a^2P_1}{1+P_2}) - \epsilon'$, where $\epsilon' = O(\sqrt{\epsilon})$ as $\epsilon \to 0$.

Remark 1. The first part of the bound (12) coincides with Sato's outer bound [11] and [12, Theorem 2] by Kramer, with the latter having been obtained by reducing the Z-interference channel to the degraded broadcast channel; the second part of (12) is new and it settles the missing corner point of the capacity region. The location of this corner point was first proposed by Costa [13], but with a flawed proof, as pointed out in [14]. The high-level difference between our proof and that of [13] is the replacement of Pinsker's inequality by Talagrand's and the use of a coupling argument.[1]

Proof. Without loss of generality, assume that all random variables have zero mean. First of all, setting $b = 0$ (which is equivalent to granting the first user access to X_2) will not shrink the capacity region of the interference channel in (10). Therefore, to prove the desired outer bound, it suffices to focus on the following Z-interference channel henceforth:

$$\begin{aligned} Y_1 &= X_1 + Z_1 \\ Y_2 &= aX_1 + X_2 + Z_2 \,. \end{aligned} \tag{16}$$

Let (X_1, X_2) be n-dimensional random variables corresponding to the encoder output of the first and second user, which are uniformly distributed on the respective codebook. For $i = 1, 2$ we define $R_i \triangleq \frac{1}{n} I(X_i; Y_i)$. By Fano's inequality there is no difference asymptotically between this definition of rate and the operational one. Define the entropy-power function of the X_1-codebook:

$$N_1(t) \triangleq \exp\left\{\frac{2}{n} h(X_1 + \sqrt{t} Z)\right\}, \qquad Z \sim \mathcal{N}(0, I_n)\,.$$

We know the following general properties of $N_1(t)$:

- N_1 is monotonically increasing.
- $N_1(0) = 0$ (since X_1 is uniform over the codebook).
- $N_1'(t) \geq 2\pi e$ (since $N_1(t+\delta) \geq N_1(t) + 2\pi e \delta$ by entropy power inequality).
- $N_1(t)$ is concave (Costa's entropy power inequality [1]).
- $N_1(t) \leq 2\pi e(P_1 + t)$ (Gaussian distribution maximizes differential entropy).

We can then express R_1 in terms of the entropy power function as

$$R_1 = \frac{1}{2} \log \frac{N_1(1)}{2\pi e}\,. \tag{17}$$

It remains to upper bound $N_1(1)$. We only show the second part of the bound. Note that

$$\begin{aligned} nR_2 &= h(X_2 + aX_1 + Z) - h(aX_1 + Z) \leq \\ &\frac{n}{2} \log 2\pi e(1 + P_2 + a^2 P_1) - h(aX_1 + Z)\,, \end{aligned} \tag{18}$$

and therefore

$$N_1\left(\frac{1}{a^2}\right) \leq 2\pi e A\,. \tag{19}$$

Let $G_2 \sim \mathcal{N}(0, P_2 I_n)$. Using $\mathbb{E}[\|X_2\|^2] \leq nP_2$ and $X_1 \perp\!\!\!\perp X_2$, we obtain

$$\begin{aligned} nR_2 = I(X_2; Y_2) &\leq I(X_2; Y_2 | X_1) = I(X_2; X_2 + Z_2) \\ &\leq nC_2 - D(P_{X_2+Z_2} \| P_{G_2+Z_2}), \end{aligned}$$

[1]After circulating our initial draft, we were informed that the authors of [3] posted an updated manuscript [15] that also proves Costa's conjecture. Their method is based on the analysis of the minimum mean-square error (MMSE) properties of good channel codes, but we were not able to verify all the details. A further update is in [16].

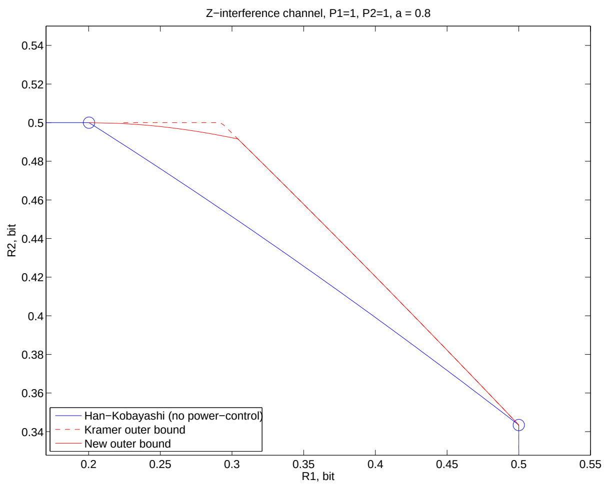

Fig. 1. Illustration of the "missing corner point": The bound in Theorem 3 establishes the location of the upper corner point, as conjectured by Costa [13]. The bottom corner point has been established by Sato [11].

that is,

$$D(P_{X_2+Z_2} \| P_{G_2+Z_2}) \leq n(C_2 - R_2)\,. \tag{20}$$

Furthermore,

$$nR_2 = h(aX_1 + X_2 + Z_2) - h(aX_1 + G_2 + Z_2) \tag{21}$$

$$+ h(aX_1 + G_2 + Z_2) - h(aX_1 + Z_2)\,. \tag{22}$$

Note that the second term (22) is precisely $\frac{n}{2} \log \frac{N_1(\frac{1}{a^2})}{N_1(\frac{1+P_2}{a^2})}$. The first term (21) can be bounded by applying Corollary 2 and (20) with $B = aX_1$, $A = X_2$, $G = G_2$, and $c = 1$ to get an estimate

$$n\sqrt{\frac{2a^2 P_1 (C_2 - R_2) \log e}{1 + P_2}}\,. \tag{23}$$

Combining (21)–(23), yields

$$N_1\left(\frac{1}{a^2}\right) \leq \frac{\exp(2\delta)}{1 + P_2} N_1\left(\frac{1 + P_2}{a^2}\right)\,. \tag{24}$$

where δ is defined in (15). From the concavity of $N_1(t)$ and (24)

$$N_1(1) \leq \gamma N_1\left(\frac{1}{a^2}\right) - (\gamma - 1) N_1\left(\frac{1 + P_2}{a^2}\right) \tag{25}$$

$$\leq N_1\left(\frac{1}{a^2}\right)\left(\gamma - (\gamma - 1)\frac{1 + P_2}{\exp(2\delta)}\right), \tag{26}$$

where $\gamma = 1 + \frac{1 - a^2}{P_2} > 1$. In view of (17), by upper bounding $N_1\left(1/a^2\right)$ in (26) via (19), we get after some simplifications the second part of (12). $\square$

The outer bound in (12) is evaluated in Fig. 1 for the case of $b = 0$ (Z-interference), where we also plot (just for reference) the simple Han-Kobayashi inner bound for the Z-GIC in (16), attained by choosing $X_1 = U + V$ with $U \perp\!\!\!\perp V$ jointly Gaussian.

B. Corner points of the capacity region

The two corner points of the capacity region are defined as follows:

$$C_1'(a,b) \triangleq \max\{R_1 : (R_1, C_2) \in \mathcal{R}(a,b)\}\,, \quad (27)$$
$$C_2'(a,b) \triangleq \max\{R_2 : (C_1, R_2) \in \mathcal{R}(a,b)\}\,, \quad (28)$$

where $C_i = \frac{1}{2}\log(1+P_i)$. As a corollary, Theorem 3 completes the picture of the corner points for the capacity region of GIC for all values of $a, b \in \mathbb{R}_+$. We refer to [9] for complete details.

For $a > 1$ and $b > 1$ (strong interference) the capacity region is well known [17], [18], and so we assume $a \le 1$ henceforth. For the top corner, we have that $C_1'(a,b)$ equals

$$\begin{cases} \frac{1}{2}\log\left(1+\frac{a^2P_1}{1+P_2}\right), & 0<a\le 1, b\ge 0 \\ C_1, & a=0, b=0 \\ C_1, & a=0, b\ge\sqrt{1+P_1} \\ \frac{1}{2}\log\left(1+\frac{P_1+(b^2-1)P_2}{1+P_2}\right), & a=0, 1<b<\sqrt{1+P_1} \\ \frac{1}{2}\log\left(1+\frac{P_1}{1+b^2P_2}\right), & a=0, 0<b\le 1. \end{cases} \quad (29)$$

Note that, for any $b \ge 0$, $a \mapsto C_1'(a,b)$ is discontinuous as $a \downarrow 0$. The bottom corner point $C_2'(a,b)$ equals

$$\begin{cases} \frac{1}{2}\log\left(1+\frac{P_2}{1+a^2P_1}\right), & b=0 \\ \frac{1}{2}\log\left(1+\frac{P_2}{1+a^2P_1}\right), & b\ge\sqrt{\frac{1+P_1}{1+a^2P_1}} \\ \frac{1}{2}\log\left(1+\frac{b^2P_2}{1+P_1}\right), & 1<b<\sqrt{\frac{1+P_1}{1+a^2P_1}} \\ \frac{1}{2}\log\left(1+\frac{b^2P_2}{1+P_1}\right), & 0<b\le 1\,, \end{cases} \quad (30)$$

which is discontinuous as $b \downarrow 0$ for any fixed $a \in [0,1]$.

IV. Discrete version

Fix a finite alphabet $\mathcal{X}$ and an integer n. On the product space $\mathcal{X}^n$ we define the Hamming distance

$$d_H(x,y) = \sum_{j=1}^n \mathbf{1}_{\{x_j \ne y_j\}}\,,$$

and consider the corresponding Wasserstein distance W_1. In fact, $\frac{1}{n}W_1(P,Q)$ is known as Ornstein's $\bar{d}$-distance [7], [19], namely,

$$\bar{d}(P,Q) = \frac{1}{n}\inf \mathbb{E}[d_H(X,Y)], \quad (31)$$

where the infimum is taken over all couplings P_{XY} of P and Q. We next formulate the analog of Proposition 1 for the discrete setting.

Proposition 4. *Let $P_{Y|X,A}$ be a two-input blocklength-n memoryless channel, namely, $P_{Y|X,A}(y|x,a) = \prod_{j=1}^n W(y_j|x_j,a_j)$, where $W(\cdot|\cdot)$ is a stochastic matrix and $y \in \mathcal{Y}^n, x \in \mathcal{X}^n, a \in \mathcal{A}^n$. Let $X, A, \tilde{A}$ be independent n-dimensional discrete random vectors. Let Y and $\tilde{Y}$ be the outputs generated by (X,A) and $(X,\tilde{A})$, respectively. Then*

$$|H(Y) - H(\tilde{Y})| \le cn\bar{d}(P_Y, P_{\tilde{Y}}) \quad (32)$$
$$D(P_Y\|P_{\tilde{Y}}) + D(P_{\tilde{Y}}\|P_Y) \le 2cn\bar{d}(P_Y, P_{\tilde{Y}}) \quad (33)$$
$$|I(X;Y) - I(X;\tilde{Y})| \le 2cn\mathbb{E}[\bar{d}(P_{Y|X}, P_{\tilde{Y}|X})] \quad (34)$$

where

$$c \triangleq \max_{x,a,y,y'} \log\frac{W(y|x,a)}{W(y'|x,a)}, \quad (35)$$
$$\mathbb{E}[\bar{d}(P_{Y|X}, P_{\tilde{Y}|X})] \triangleq \sum_{x\in\mathcal{X}^n} P_X(x)\bar{d}(P_{Y|X=x}, P_{\tilde{Y}|X=x}). \quad (36)$$

Proof. The function $y \mapsto \log P_Y(y)$ is c-Lipschitz with respect to the Hamming distance (cf. [10, Eqn. (58)]). From Lipschitz continuity we conclude the existence of a coupling $P_{Y,\tilde{Y}}$, such that

$$\mathbb{E}\left[\left|\log\frac{P_Y(Y)}{P_Y(\tilde{Y})}\right|\right] \le cn\bar{d}(P_Y, P_{\tilde{Y}})\,.$$

The rest of the proof of (32) and (33) is straightforward [9]. To get the inequality for mutual informations, apply (32) to estimate $|H(Y|X=x) - H(\tilde{Y}|X=x)|$ in terms of $\bar{d}(P_{Y|X=x}, P_{\tilde{Y}|X=x})$ and average it over X. □

We next show how to determine corner points of capacity regions of discrete memoryless interference channels (DMIC). We will need two extra results. First is Marton's transportation inequality that will help convert Proposition 4 to bounds in terms of KL divergence as follows. When Q is a product distribution [7, Lemma 1] states:

$$\bar{d}(P,Q) \le \sqrt{\frac{D(P\|Q)}{2n\log e}}. \quad (37)$$

Second is an auxiliary tensorization result which appears to be a standard exercise for degraded channels.[2]

Proposition 5. *Let $X^n \to A^n \to B^n$, where the memoryless channels $P_{A|X}$ and $P_{B|A}$ of blocklength n satisfy $P_{B|A=a} \not\perp P_{B|A=a'}$, $\forall a \ne a'$, and $P_{A|X=x} \ne P_{A|X=x'}$, $\forall x \ne x'$. Then, there exists a continuous function $g : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $g(0) = 0$, such that, for all n, $I(X^n; A^n) \le I(X^n; B^n) + \epsilon n$ implies*

$$H(X^n) \le I(X^n; B^n) + g(\epsilon)n\,. \quad (38)$$

We are now ready to state a non-trivial example of corner points for the capacity region of DMIC.

Theorem 6. *Consider the two-user DMIC*

$$Y_1 = X_1\,, \quad (39)$$
$$Y_2 = X_2 + X_1 + Z_2 \mod 3\,, \quad (40)$$

[2]This is the analog of the following property of Gaussian channels: For i.i.d. Gaussian Z and $t_1 < t_2 < t_3$ we have $I(X; X+t_2Z) = I(X; X+t_3Z)+o(n)$ implies $I(X; X+t_1Z) = I(X; X+t_3Z)+o(n)$. This follows from Costa's EPI.

where $X_1 \in \{0,1,2\}^n$, $X_2 \in \{0,1\}^n$, $Z_2 \in \{0,1,2\}^n$ are mutually independent, and $Z_2 \sim P_2^{\otimes n}$ is i.i.d. for some non-uniform P_2 containing no zeros. The maximal rate achievable by user 2 is

$$C_2 = \max_{\mathrm{supp}(Q)\subset\{0,1\}} H(Q * P_2) - H(P_2). \quad (41)$$

At this rate, the maximal rate of user 1 is

$$C_1' = \log 3 - \max_{\mathrm{supp}(Q)\subset\{0,1\}} H(Q * P_2). \quad (42)$$

Proof. Given a sequence of codes with vanishing probability of error and rate pairs (R_1, R_2), where $R_2 = C_2 - \epsilon$, we show that $R_1 \le C_1' - \epsilon'$, where $\epsilon' \to 0$ as $\epsilon \to 0$. Let Q_2 be the maximizer of (41), i.e., the capacity-achieving distribution of the channel $X_2 \mapsto X_2 + Z_2$. Let $\tilde{X}_2 \in \{0,1\}^n$ be distributed according to Q_2^n. Then, $\tilde{X}_2 + Z_2 \sim P_3^{\otimes n}$, where $P_3 = Q_2 * P_2$. By Fano's inequality,

$$D(P_{X_2+Z_2} \| P_{\tilde{X}_2+Z_2}) \le n\epsilon + o(n).$$

Since $P_{\tilde{X}_2+Z_2} = P_3^{\otimes n}$ is a product distribution, Marton's inequality (37) yields

$$\bar{d}(P_{X_2+Z_2}, P_{\tilde{X}_2+Z_2}) \le \sqrt{\frac{\epsilon}{2\log e}} + o(1).$$

Applying (34) in Proposition 4 and in view of the translation invariance of the $\bar{d}$-distance, we obtain

$$|I(X_1;Y_2) - I(X_1;X_1+\tilde{X}_2+Z_2)| \le (\alpha\sqrt{\epsilon}+o(1))n,$$

for a finite constant α. On the other hand,

$$I(X_1;X_1+Z_2) = I(X_1;Y_2) + I(X_1;X_2|Y_2) = I(X_1;Y_2) + o(n), \quad (43)$$

where $I(X_1;X_2|Y_2) \le H(X_2|Y_2) = o(n)$ by Fano's inequality. Combining the last two displays, we have

$$I(X_1;X_1+\tilde{X}_2+Z_2) \le I(X_1;X_1+Z_2) + (\alpha\sqrt{\epsilon}+o(1))n.$$

Next we apply Proposition 5, with $X = X_1 \to A = X_1 + Z_2 \to B = A + \tilde{X}_2$, to get

$$H(X_1) \le I(X_1;X_1+\tilde{X}_2+Z_2) + g(\alpha\sqrt{\epsilon})n \le nC_1' + o(n),$$

where the last inequality follows from the fact that $\max_{X_1} I(X_1;X_1+\tilde{X}_2+Z_2) = nC_1'$, attained by X_1 uniform on $\{0,1,2\}^n$.

Finally, note that the rate pair (C_1', C_2) is achievable by a random MAC-code for $(X_1, X_2) \to Y_2$, with X_1 uniform on $\{0,1,2\}^n$ and $X_2 \sim Q_2^{\otimes n}$. □

References

[1] M. H. Costa, "A new entropy power inequality," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 751–760, 1985.

[2] M. H. M. Costa, "Noisebergs in Z-Gaussian interference channels," in *Proc. Information Theory and Applications Workshop (ITA)*, San Diego, CA, Feb. 2011.

[3] R. Bustin, H. V. Poor, and S. Shamai, "The effect of maximal rate codes on the interfering message rate," in *Proc. 2014 IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, USA, Jul. 2014, pp. 91–95.

[4] M. H. Costa and O. Rioul, "From almost Gaussian to Gaussian: Bounding differences of differential entropies," in *Proc. Information Theory and Applications Workshop (ITA)*, San Diego, CA, Feb. 2015.

[5] O. Rioul and M. H. Costa, "Almost there – corner points of Gaussian interference channels," in *Proc. Information Theory and Applications Workshop (ITA)*, San Diego, CA, Feb. 2015.

[6] I. Sason, "On the corner points of the capacity region of a two-user Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 61, no. 7, pp. 3682–3697, July 2015.

[7] K. Marton, "A simple proof of the blowing-up lemma (corresp.)," *IEEE Trans. Inf. Theory*, vol. 32, no. 3, pp. 445–446, 1986.

[8] M. Talagrand, "Transportation cost for Gaussian and other product measures," *Geometric and Functional Analysis*, vol. 6, no. 3, pp. 587–600, 1996.

[9] Y. Polyanskiy and Y. Wu, "Wasserstein continuity of entropy and outer bounds for interference channels," *arXiv preprint arXiv:1504.04419*, Apr. 2015.

[10] Y. Polyanskiy and S. Verdú, "Empirical distribution of good channel codes with non-vanishing error probability," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 5–21, Jan. 2014.

[11] H. Sato, "On degraded Gaussian two-user channels (corresp.)," *IEEE Trans. Inf. Theory*, vol. 24, no. 5, pp. 637–640, 1978.

[12] G. Kramer, "Outer bounds on the capacity of Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 3, pp. 581–586, 2004.

[13] M. H. Costa, "On the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 31, no. 5, pp. 607–615, 1985.

[14] I. Sason, "On achievable rate regions for the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1345–1356, 2004.

[15] R. Bustin, H. V. Poor, and S. Shamai, "The effect of maximal rate codes on the interfering message rate," *arXiv preprint arXiv:1404.6690v4*, Apr 2015.

[16] ——, "Optimal point-to-point codes in interference channels: An incremental I-MMSE approach," *arXiv preprint arXiv:1510.08213*, Oct 2015.

[17] A. Carleial, "A case where interference does not reduce capacity (corresp.)," *IEEE Trans. Inf. Theory*, vol. 21, no. 5, pp. 569–570, Sep 1975.

[18] H. Sato, "The capacity of the Gaussian interference channel under strong interference (corresp.)," *IEEE Trans. Inf. Theory*, vol. 27, no. 6, pp. 786–788, 1981.

[19] R. M. Gray, D. L. Neuhoff, and P. C. Shields, "A generalization of Ornstein's $\bar{d}$ distance with applications to information theory," *The Annals of Probability*, pp. 315–328, 1975.