ISIT'2013 1569717055

A New Entropy Power Inequality for Integer-valued Random Variables

Saeid Haghighatshoar*, Emmanuel Abbe†, Emre Telatar*
*EPFL, Lausanne, Switzerland, {saeid.haghighatshoar, emre.telatar}@epfl.ch
†Princeton University, Princeton, NJ, USA, eabbe@princeton.edu

***Abstract*—The entropy power inequality (EPI) provides lower bounds on the differential entropy of the sum of two independent real-valued random variables in terms of the individual entropies. Versions of the EPI for discrete random variables have been obtained for special families of distributions with the differential entropy replaced by the discrete entropy, but no universal inequality is known (beyond trivial ones). More recently, the sumset theory for the entropy function yields a sharp inequality $H(X+X')-H(X)\geq \frac{1}{2}-o(1)$ when X, X' are i.i.d. with high entropy. This paper provides the inequality $H(X+X')-H(X)\geq g(H(X))$, where X, X' are arbitrary i.i.d. integer-valued random variables and where g is a universal strictly positive function on $\mathbb{R}_+$ satisfying $g(0)=0$. Extensions to non identically distributed random variables and to conditional entropies are also obtained.**

***Index Terms*—Entropic inequalities, Entropy power inequality, Mrs. Gerber's lemma, Shannon sumset theory.**

I. Introduction

For a continuous random variable[1] $\boldsymbol{X}$ on $\mathbb{R}^n$, let $h(\boldsymbol{X})$ be the differential entropy of $\boldsymbol{X}$ and let $N(\boldsymbol{X})=2^{\frac{2}{n}h(\boldsymbol{X})}$ denote the *entropy power* of $\boldsymbol{X}$. If $\boldsymbol{Y}$ is another continuous $\mathbb{R}^n$-valued random variable independent of $\boldsymbol{X}$, the EPI states that

$$N(\boldsymbol{X}+\boldsymbol{Y})\geq N(\boldsymbol{X})+N(\boldsymbol{Y}), \tag{1}$$

with equality if and only if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are Gaussian with proportional covariance matrices. A weaker statement of the EPI, yet of key use in applications, is the following inequality stated here for $n=1$,

$$h(X+X')-h(X)\geq \frac{1}{2}, \tag{2}$$

where X, X' are i.i.d., and where equality holds if and only if X is Gaussian.

The EPI was first proposed by Shannon [1] who used a variational argument to show that Gaussian random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ with proportional covariance matrices and specified differential entropies constitute a stationary point for $h(\boldsymbol{X}+\boldsymbol{Y})$. The first rigorous proof of the EPI was given by Stam [2] in 1959, using the De Bruijin's identity. This proof was further simplified by Blachman [3]. Another proof was given by Lieb [4] using an extension of Young's inequality.

While there is a wide variety of entropic inequalities, the EPI is the only general inequality in information theory estimating the entropy of a sum of independent random variables by means of the individual entropies. It is used as a key ingredient to prove converse results for some of the coding theorems [5]–[9].

There have been numerous extensions and simplifications of the EPI over the reals, refer to the references in [16]. There have also been several attempts to obtain discrete versions of the EPI, using Shannon entropy. Of course, it is not clear what is meant by a discrete version of the EPI, since (1), (2) clearly do not hold verbatim for Shannon entropy.

Several extensions have yet been developed. First, there have been some extensions using finite field additions, for example, the so-called Mrs. Gerber's Lemma (MGL) proved in [11] by Wyner and Ziv for binary alphabets. The MGL was further extended by Witsenhausen [12] to non binary alphabets, who also provided counter-examples for the case of general alphabets. More recently, [16] obtained EPI and MGL results for abelian groups of order 2^n. Second, concerning discrete random variables and addition over the reals, Harremoes and Vignat [13] proved that the discrete EPI holds for binomial random variables with parameter $\frac{1}{2}$, namely if $X_n\sim B(n,\frac{1}{2})$ then for every $m,n\geq 1$, $2^{2H(X_m+X_n)}\geq 2^{2H(X_n)}+2^{2H(X_m)}$. This result was later on generalized by Sharma, Das and Muthukrishnan [14]. Yu and Johnson [15] obtained a version of the EPI for discrete random variables using the notion of thinning.

More recently, Tao established in [17] a sumset theory for Shannon entropy, obtaining in particular the sharp inequality

$$H(X+X')-H(X)\geq \frac{1}{2}-o(1),$$

where $o(1)$ vanishes when $H(X)$ tends to infinity. Further results were obtained for the differential entropy in [18].

In this paper, we are interested in integer-valued random variables with arithmetic over the reals. We show that there exists an increasing function $g:\mathbb{R}_+\to\mathbb{R}_+$, such that $g(c)=0$ if and only if $c=0$, and

$$H(X+X')-H(X)\geq g(H(X)),$$

for any i.i.d. integer-valued random variables X, X'. We further generalize the result to non identically distributed random variables and to conditional entropies.

The main motivation for this paper came from [10], where the source polarization phenomenon introduced in [19] was extended to integer-valued random variables under linear measurements over the reals taken by Hadamard matrices.

[1] All continuous random variables have well-defined differential entropies.

II. Results

In this section, we will give an overview of the results proved in the paper. The first theorem gives an EPI for i.i.d. integer-valued random variables.

Theorem 1. *There is a function $g : \mathbb{R}_+ \to \mathbb{R}_+$ such that for any two i.i.d. $\mathbb{Z}$-valued random variables X, X',*

$$H(X + X') - H(X) \geq g(H(X)).$$

Moreover, g is an increasing function, $\lim_{c\to\infty} g(c) = \frac{1}{8}\log_2(e)$ and $g(c) = 0$ if and only if $c = 0$.

Remark 1. The function g in Theorem 1 is given by

$$g(c) = \min_{x\in[0,1]}\{(cx - h_2(x)) \vee \frac{(1-x)^2((1-x)\vee(4x-2)^+)^2\log_2(e)}{8}\},$$

where h_2 is the binary entropy function and for $a, b \in \mathbb{R}$, $a \vee b$ denotes the maximum of a and b.

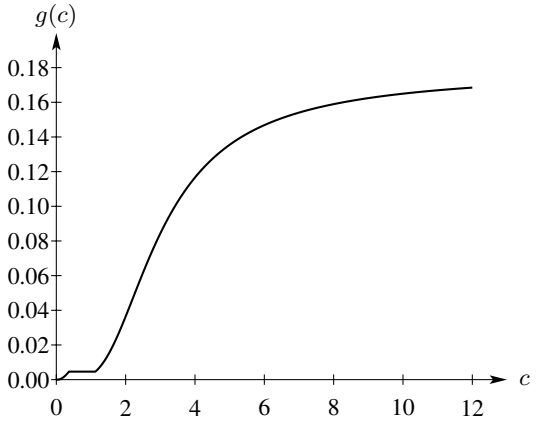

Fig. 1: The EPI gap for i.i.d. integer-valued random variables

Remark 2. As mentioned in the introduction, a recent result by Tao [17] implies that for two i.i.d. $\mathbb{Z}$-valued random variables X, X' of very large entropy, $H(X + X') - H(X) \approx \frac{1}{2}$. In comparison with this result, we get the asymptotic lower bound $\frac{1}{8}\log_2(e) \approx 0.18$, which also holds for the general independent case provided that the entropy of both random variables approaches infinity.

The next theorem extends the i.i.d. result to the general independent case.

Theorem 2. *There is a function $g : \mathbb{R}_+^2 \to \mathbb{R}_+$ such that for any two independent $\mathbb{Z}$-valued random variables X, X',*

$$H(X + X') - \frac{H(X) + H(X')}{2} \geq g(H(X), H(X')).$$

Moreover, g is a positive and doubly-increasing[2] function, $\lim_{(c,d)\to(\infty,\infty)} g(c, d) = \frac{1}{8}\log_2(e)$ and $g(c, d) = 0$ if and only if $c = d = 0$.

[2]A function $g : \mathbb{R}_+^2 \to \mathbb{R}_+$ is doubly-increasing if for any value of one of the arguments, it is an increasing function of the other argument.

Remark 3. One might be tempted to prove the stronger bound

$$H(X + X') - (H(X) \vee H(X')) \geq g(H(X), H(X')), \quad (3)$$

for some doubly-increasing function g. However, this fails because, for example, assume that X, X' are random variables uniformly distributed over $\{1, 2, \ldots, M\}$ and $\{1, 2, \ldots, NM\}$, for some number $N \geq 2$. It is not difficult to show that

$$H(X + X') - (H(X) \vee H(X')) \leq \log_2(\frac{N+1}{N}),$$

which decreases to 0 with increasing N. Hence, the strong inequality (3) does not hold universally over all integer-valued random variables.

The next theorem extends the results in Theorem 1 to the conditional case.

Theorem 3. *There is a function $\tilde{g} : \mathbb{R}_+ \to \mathbb{R}_+$ such that for any two i.i.d. $\mathbb{Z}$-valued pairs of random variables (X, Y) and (X', Y'),*

$$H(X + X'|Y, Y') - H(X|Y) \geq \tilde{g}(H(X|Y)).$$

Moreover, $\tilde{g} : \mathbb{R}_+ \to \mathbb{R}_+$ is an increasing function and $\tilde{g}(c) = 0$ if and only if $c = 0$.

Remark 4. The function $\tilde{g}$ is given by

$$\tilde{g}(c) = \min_{\delta\in[0,\frac{1}{2}]}\{(g(c, c) - h_2(\delta)) \ \vee \ \delta^2 g(c, c)\}, \quad (4)$$

where g is as in Theorem 2.

III. Proof techniques

In this part, we will give an overview and also some intuition about the techniques used for proving the theorems.

A. EPI for i.i.d. random variables

We will start from the EPI for i.i.d. random variables. The main idea of the proof is to find suitable bounds for $H(p \star p) - H(p)$ in two different cases: one case in which p is close to a spiky distribution (a unit mass at a single point) and the other case where p is close to a uniform distribution over a subset of $\mathbb{Z}$.

Lemma 1. *Assume that p is a probability distribution over $\mathbb{Z}$ with $H(p) = c$ and let $x = \|p\|_\infty$. Then*

$$H(p \star p) - c \geq cx - h_2(x).$$

Remark 5. If p is a spiky distribution then $\|p\|_\infty \approx 1$ and Lemma 1 gives the bound $H(p \star p) - c \gtrsim c$, which is the tightest lower bound we can hope.

Lemma 2. *Assume that p_1, p_2 and p are arbitrary probability distributions over $\mathbb{Z}$ such that p_1 and p_2 have non-overlapping supports and $\|p\|_\infty = x$. Then*

$$\|p \star p_1 - p \star p_2\|_1 \geq 2(2x - 1)^+.$$

Lemma 3. *Let $c > 0$, $0 < \alpha \leq \frac{1}{2}$ and $n \in \mathbb{Z}$. Assume that p is a probability distribution over $\mathbb{Z}$ such that $\alpha \leq p((-\infty, n]) \leq 1 - \alpha$ and $H(p) = c$. Then,*

$$\|p \star p_1 - p \star p_2\|_1 \geq 2\alpha,$$

where $p_1 = \frac{1}{p((-\infty,n])} p|_{(-\infty,n]}$ and $p_2 = \frac{1}{p([n+1,\infty))} p|_{[n+1,\infty)}$ are scaled restrictions of p to $(-\infty, n]$ and $[n+1, \infty)$ respectively.

Lemma 4. *Assuming the hypotheses of Lemma 3,*

$$H(p \star p) - c \geq \frac{\alpha^2 \|p \star p_1 - p \star p_2\|_1^2}{2} \log_2(e).$$

Lemma 5. *Assume that p is a probability distribution over $\mathbb{Z}$ with $H(p) = c$ and $\|p\|_\infty = x$. Then*

$$H(p \star p) - c \geq \frac{(1-x)^2((1-x) \vee (4x-2)^+)^2}{8} \log_2(e).$$

Proof: We only give a sketch of the proof. Let $\alpha = \frac{1-x}{2}$. It can be checked that α satisfies the conditions of Lemma 3, which along with Lemma 2, implies that $\|p \star p_1 - p \star p_2\|_1 \geq (1-x) \vee (4x-2)^+$. Therefore, using Lemma 4, we get the result. ∎

Now that we have the required bounds in the spiky and non-spiky cases, we can combine them to prove Theorem 1.

Proof of Theorem 1: Assume that X, X' are arbitrary i.i.d. random variables with a probability distribution p over $\mathbb{Z}$ with $H(p) = c$ and $\|p\|_\infty = x$. It is easy to see that $x \geq 2^{-c}$. Using Lemma 1 and Lemma 5, it results that $H(p \star p) - c \geq l(c)$, where

$$l(c) = \min_{x \in [2^{-c}, 1]} \{(cx - h_2(x)) \vee \frac{(1-x)^2((1-x) \vee (4x-2)^+)^2}{8} \log_2(e)\}.$$

We will use a simpler lower bound given by

$$g(c) = \min_{x \in [0,1]} \{(cx - h_2(x)) \vee \frac{(1-x)^2((1-x) \vee (4x-2)^+)^2}{8} \log_2(e)\},$$

where obviously $l(c) \geq g(c)$. It is easy to check that $g(c)$ is a continuous function of c. The monotonicity of g follows from monotonicity of $cx - h_2(x)$ with respect to c, for every $x \in [0,1]$. For strict positivity, note that $(1-x)^2((1-x) \vee (4x-2)^+)^2$ is strictly positive for $x \in [0,1)$ and it is 0 when $x = 1$, but $\lim_{x \to 1} cx - h_2(x) = c$. Hence, for $c > 0$, $g(c) > 0$. It can also be shown that $\lim_{c \to \infty} g(c) = \frac{1}{8} \log_2(e)$. ∎

Figure 1 shows the EPI gap with the asymptotic value 0.18.

B. EPI for non-i.i.d. random variables

Theorem 2 is an extension of Theorem 1 to independent but non identically distributed random variables. Similar to the i.i.d. case, the idea is to distinguish between the spiky and non-spiky distributions.

Lemma 6. *Assume that p and q are two probability distributions over $\mathbb{Z}$ with $H(p) = c$ and $H(q) = d$. Suppose that $x = \|p\|_\infty$ and $y = \|q\|_\infty$. Then,*

$$2H(p \star q) - c - d \geq dx - h_2(x) + cy - h_2(y). \quad (5)$$

When at least one of the distributions is spiky, Lemma 6 gives a relatively tight bound. Hence, we should try to find a good bound for the non-spiky case.

Lemma 7. *Let p, q be two probability distributions over $\mathbb{Z}$. Assume that there are $0 < \alpha, \beta < \frac{1}{2}$ and $m, n \in \mathbb{Z}$ such that $\alpha \leq p((-\infty, m]) \leq 1 - \alpha$ and $\beta \leq q((-\infty, n]) \leq 1 - \beta$. Then*

$$\|q \star p_1 - q \star p_2\|_1 + \|p \star q_1 - p \star q_2\|_1 \geq 2(\alpha + \beta),$$

where $p_1 = \frac{1}{p((-\infty,m])} p|_{(-\infty,m]}$, $p_2 = \frac{1}{p([m+1,\infty))} p|_{[m+1,\infty)}$, $q_1 = \frac{1}{q((-\infty,n])} q|_{(-\infty,n]}$, and $q_2 = \frac{1}{q([n+1,\infty))} q|_{[n+1,\infty)}$.

Lemma 8. *Assume that the hypotheses of Lemma 7 hold and let $H(p) = c$ and $H(q) = d$. Then*

$$H(p \star q) - d \geq \frac{\alpha^2 \|q \star p_1 - q \star p_2\|_1^2}{2} \log_2(e),$$
$$H(p \star q) - c \geq \frac{\beta^2 \|p \star q_1 - p \star q_2\|_1^2}{2} \log_2(e).$$

Lemma 9. *Let p and q be probability distributions over $\mathbb{Z}$ with $H(p) = c$, $H(q) = d$, $\|p\|_\infty = x$ and $\|q\|_\infty = y$. Then $2H(p \star q) - c - d \geq l(x, y)$, where*

$$l(x,y) = \min_{(a,b) \in T(x,y)} \frac{(1-x)^2 a^2 + (1-y)^2 b^2}{8} \log_2(e),$$

and $T(x, y)$, for $(x, y) \in [0,1] \times [0,1]$, is the set of all $(a, b) \in \mathbb{R}_+^2$ given by the following inequalities:

$$a \geq (4y-2)^+, b \geq (4x-2)^+, a + b \geq 2 - x - y.$$

Moreover, $l(x, y)$ is a continuous function, $l(x, y) \geq 0$ and $l(x, y) = 0$ if and only if $(x, y) = (1, 1)$.

Proof: Let $\alpha = \frac{1-x}{2}$ and $\beta = \frac{1-y}{2}$. It can be checked that α, β satisfy the conditions of Lemma 7, thus using Lemma 8, we obtain

$$2H(p \star q) - c - d \geq \frac{\alpha^2 a^2 + \beta^2 b^2}{2} \log_2(e) = \frac{(1-x)^2 a^2 + (1-y)^2 b^2}{8} \log_2(e),$$

where $a = \|q \star p_1 - q \star p_2\|_1$ and $b = \|p \star q_1 - p \star q_2\|_1$. Also, from Lemma 7, we have

$$a + b \geq 2(\alpha + \beta) = 2 - x - y. \quad (6)$$

Moreover, applying Lemma 2 to the distribution p with $\|p\|_\infty = x$ and q_1, q_2 with disjoint supports, and similarly to q with $\|q\|_\infty = y$ and p_1, p_2 with disjoint supports, we get

$$b \geq (4x-2)^+, a \geq (4y-2)^+. \quad (7)$$

Therefore, $2H(p \star q) - c - d \geq l(x, y)$, where

$$l(x,y) = \min_{(a,b) \in T(x,y)} \frac{(1-x)^2 a^2 + (1-y)^2 b^2}{8} \log_2(e),$$

and $T(x,y)$ is defined by the three inequalities derived in (6) and (7). The continuity of $l(x,y)$ can be easily checked. For the last part of the lemma, notice that if $M := x \vee y < 1$ then

$$\begin{aligned} l(x,y) &\geq \min_{a+b \geq 2-2M} \frac{(1-M)^2(a^2+b^2)}{8} \log_2(e) \\ &\geq \frac{(1-M)^4}{4} \log_2(e) > 0. \end{aligned}$$

Moreover, if $x \vee y = 1$ but $(x,y) \neq (1,1)$ then, for example, $y \in [0,1), x = 1$, which implies that $b \geq 2$. Therefore, we get $l(x,y) \geq \frac{(1-y)^2}{2} \log_2(e)$, which is strictly positive. A similar argument applies to $x \in [0,1), y = 1$. Therefore, over $(x,y) \in [0,1] \times [0,1]$, $l(x,y) \geq 0$ and $l(x,y) = 0$ if and only if $(x,y) = (1,1)$. ∎

Proof of Theorem 2: Let X, X' be independent integer-valued random variables with probability distributions p, q, where $H(p) = c$ and $H(q) = d$. Let $x = \|p\|_\infty$ and $y = \|q\|_\infty$. It is easy to check that $x \geq 2^{-c}, y \geq 2^{-d}$. Using Lemma 6 and Lemma 9, we obtain that $H(p \star q) - \frac{c+d}{2} \geq s(c,d)$, where $s(c,d)$ is given by

$$\frac{1}{2} \min_{(x,y) \in R(c,d)} \big\{ (dx - h_2(x) + cy - h_2(y)) \ \vee \ l(x,y) \big\},$$

for $R(c,d) = [2^{-c}, 1] \times [2^{-d}, 1]$. We will use a simpler lower bound given by

$$g(c,d) = \frac{1}{2} \min_{(x,y) \in R} \big\{ (dx - h_2(x) + cy - h_2(y)) \vee l(x,y) \big\},$$

where $R = [0,1] \times [0,1]$. It is easy to see that $g(c,d)$ is a continuous function. Also, notice that $l(x,y)$ in the definition of g is strictly positive except for $(x^\star, y^\star) = (1,1)$. But $\lim_{(x,y) \to (1,1)} dx - h_2(x) + cy - h_2(y) = c + d$, which is strictly positive unless $c = d = 0$. Therefore, for $(c,d) \neq (0,0)$, $g(c,d) > 0$.

The function $dx - h_2(x) + cy - h_2(y)$ is a doubly-increasing function of (c,d) over R, which implies that $g(c,d)$ must be a doubly-increasing function of (c,d). Also, it can be checked that for high values of c and d, the outer minimum in the definition of g is achieved in a small neighborhood of $(0,0)$, namely, $[0,\epsilon] \times [0,\epsilon]$, where ϵ approaches zero as c, d get large. From the continuity of $l(x,y)$, it can be shown that in this range, the value of $l(x,y)$ is close to

$$\min_{(a,b): a,b \geq 0, a+b \geq 2} \frac{a^2+b^2}{8} \log_2(e) = \frac{1}{4} \log_2(e).$$

This implies that $\lim_{(c,d) \to (\infty,\infty)} g(c,d) = \frac{1}{8} \log_2(e)$. ∎

C. Conditional EPI

In this part, we will prove the EPI result for the conditional case, where we will find a lower bound for the conditional entropy gap, $H(X + X'|Y,Y') - H(X|Y)$, for i.i.d. $\mathbb{Z}$-valued pairs (X,Y) and (X',Y') assuming that $H(X|Y) = c$, for some positive number c. To simplify the problem, we drop the i.i.d. condition and assume that (X,Y) and (X',Y') are independent pairs with $H(X|Y) = H(X'|Y') = c$. Let $t_n(c)$ be the infimum of $H(X + X'|Y,Y') - c$ over all independent pairs $(X,Y), (X',Y')$ having a conditional entropy equal to c and with Y and Y' having a support size at most n. Also, assume that $t_\infty(c)$ is the corresponding infimum when there is no constraint on the support size of Y, Y'. We prove the following lemma.

Lemma 10. *For every $n \geq 2$, $t_\infty(c) = t_n(c)$.*

Proof: Obviously, $t_n(c) \geq t_\infty(c)$. Moreover, given any $\epsilon > 0$ there is an ϵ-optimal independent pair (X,Y) and (X',Y') such that $H(X+X'|Y,Y') - c \leq t_\infty(c) + \epsilon$. Let q, q' denote the distribution of Y, Y' and let p_i, p'_j be the conditional distribution of X, X' given $Y = i, Y' = j$. Let

$$V = \{ \boldsymbol{v}_{ij} \in \mathbb{R}^3 : \boldsymbol{v}_{ij} = (H(p_i \star p'_j), H(p_i), H(p'_j)),\ i,j \in \mathbb{Z} \}.$$

It is easy to see that

$$\sum_{i,j \in \mathbb{Z}} q_i q'_j \boldsymbol{v}_{ij} = (H(X + X'|Y,Y'), c, c) := \boldsymbol{h},$$

which implies that the three dimensional vector $\boldsymbol{h} := (H(X + X'|Y,Y'), c, c)$ can be written as a convex combination of the vectors $\boldsymbol{v}_{ij} \in V$ with weights $q_i q'_j$. Let $\boldsymbol{v}_i = \sum_j q'_j \boldsymbol{v}_{ij}$. Then, we have $\sum_i q_i \boldsymbol{v}_i = \boldsymbol{h}$. Notice that the second component of $\boldsymbol{v}_i$ is equal to $H(p_i)$. Also, the third component is equal to c independent of i, which implies that there are only two components depending on i in $\boldsymbol{v}_i$. Therefore, by Carathèodory theorem, it is possible to write $\boldsymbol{h}$ as a convex combination of at most three $\boldsymbol{v}_i, i \in \mathbb{Z}$, which without loss of generality, we can assume to be $\{\boldsymbol{v}_0, \boldsymbol{v}_1, \boldsymbol{v}_2\}$. In other words, there are positive $\gamma_i, i = 0,1,2$, $\sum_{i=0}^2 \gamma_i = 1$ and $\boldsymbol{h} = \sum_{i=0}^2 \gamma_i \boldsymbol{v}_i$. Also, note that if we change the distribution of Y from q to γ, the resulting $(X,Y), (X',Y')$ is again an ϵ-optimal solution. Now, we claim that we can simplify the problem further and find a probability triple $\psi = (\psi_0, \psi_1, \psi_2)$ for Y with at most 2 non-zero elements such that $\sum_{i=0}^2 \psi_i H(p_i) = c$ and

$$\sum_{i=0}^2 \psi_i \boldsymbol{v}_i^{(1)} \leq \sum_{i=0}^2 \gamma_i \boldsymbol{v}_i^{(1)} = \sum_{i=0}^2 q_i \boldsymbol{v}_i^{(1)} = H(X + X'|Y,Y'),$$

where $\boldsymbol{v}_i^{(1)}$ denotes the first coordinate of the vector $\boldsymbol{v}_i$. This implies that if we replace the distribution γ for Y by ψ, which has a support size at most 2, we get a lower $H(X+X'|Y,Y')$.

To prove the claim, let us consider the following problem

$$\text{minimize } \sum_{i=0}^2 \psi_i \boldsymbol{v}_i^{(1)} \text{ s.t. } \begin{cases} \sum_{i=0}^2 \psi_i = 1, \\ \sum_{i=0}^2 \psi_i H(p_i) = c, \\ \psi_i \geq 0. \end{cases}$$

First of all, notice that as $\sum_{i=0}^2 \gamma_i H(p_i) = c$, γ is in the feasible set. Therefore, the feasible set is a non-empty subset of the three dimensional probability simplex. Also, as the objective function is linear in ψ, the optimal point must be at the boundary of the feasible set, which implies that there is an optimal solution with at most two non-zero components and this proves the claim. By a similar argument, we can assume that q' has also support size at most 2 and we obtain a new ϵ-optimal solution in which the support of both q and q' has

at most size 2. This implies that for any $\epsilon > 0$ and any $n \geq 2$, $t_n(c) \leq t_2(c) \leq t_\infty(c) + \epsilon$. Therefore, $t_n(c) = t_\infty(c)$. ∎

Using Lemma 10, without loss of generality, we can assume that Y, Y' are binary random variables. We will use the following lemmas to prove Theorem 3.

Lemma 11. *Let $(X, Y), (X', Y')$ be an independent pair of random variables, where Y and Y' are binary valued with $\mathbb{P}(Y = 0) = \alpha$, $\mathbb{P}(Y' = 0) = \beta$ and $H(X|Y) = H(X'|Y') = c$. Then,*

$$H(X + X'|Y, Y') - c \geq g(c, c) - (h_2(\alpha) \wedge h_2(\beta)),$$

where g is as in Theorem 2.

Proof: We only give a sketch of the proof. Let p_0, p_1 be the conditional distribution of X given $Y = 0$ and $Y = 1$, thus the distribution of X is $p_X = \alpha p_0 + (1 - \alpha)p_1$. Let $p_1^{(n)}$ be the distribution p_1 shifted to the right by n steps, namely, $p_1^{(n)}(i) = p_1(i - n)$. It is easy to check that if we replace p_1 by $p_1^{(n)}$, then $H(X|Y)$ and $H(X + X'|Y, Y')$ will remain unchanged. However, the new distribution of X will be $p_X = \alpha p_0 + (1-\alpha)p_1^{(n)}$ and the idea is that for n high enough, Y can be predicted from X, thus we can make $H(Y|X)$ as small as desired. A similar argument holds for $H(Y'|X')$. Assume that for a given $\epsilon > 0$, by a suitable shift of the distributions, we have $H(Y|X), H(Y'|X') < \epsilon$. Then, we obtain the following:

$$\begin{aligned}
&H(X + X'|Y, Y') - c \\
&\quad = H(X + X') - H(X) - I(X + X'; Y, Y') + I(X; Y) \\
&\quad \geq H(X + X') - H(X) - H(Y, Y') + H(Y) - H(Y|X) \\
&\quad \geq H(X + X') - H(X) - H(Y') - \epsilon \\
&\quad \geq g(H(X), H(X')) - h_2(\beta) - \epsilon \\
&\quad \geq g(c, c) - h_2(\beta) - \epsilon,
\end{aligned}$$

where we used $H(X), H(X') \geq c$ and the doubly-increasing property of g. Similarly, one can show that $H(X+X'|Y, Y') - c \geq g(c, c) - h_2(\alpha) - \epsilon$. As $\epsilon > 0$ is arbitrary, we get the desired result $H(X + X'|Y, Y') - c \geq g(c, c) - (h_2(\alpha) \wedge h_2(\beta))$. ∎

Lemma 12. *Assume that all of the conditions of Lemma 11 hold. Suppose there is a $0 \leq \delta \leq \frac{1}{2}$ such that $\delta < \alpha, \beta < 1 - \delta$. Then $H(X + X'|Y, Y') - c \geq \delta^2 g(c, c)$.*

Proof: Assume that the distribution of Y, Y' is q, q'. Also, for $k, l \in \{0, 1\}$, let p_k, p'_l be the conditional distribution of X, X' given $Y = k, Y' = l$. By the assumption, $\delta < \alpha, \beta < 1 - \delta$, we have $q_r, q'_s > \delta$ for any $r, s \in \{0, 1\}$. There must be $i, j \in \{0, 1\}$ such that $H(p_i), H(p'_j) \geq c$. Therefore, we have

$$\begin{aligned}
&H(X + X'|Y, Y') - c \\
&\quad = \sum_{k,l=0}^{1} q_k q'_l (H(p_k \star p'_l) - \frac{H(p_k) + H(p'_l)}{2}) \\
&\quad \geq q_i q'_j (H(p_i \star p'_j) - \frac{H(p_i) + H(p'_j)}{2}) \\
&\quad \geq \delta^2 g(H(p_i), H(p'_j)) \geq \delta^2 g(c, c),
\end{aligned}$$

where we used the doubly-increasing property of g. ∎

Proof of Theorem 3: The proof follows by combining the results obtained in Lemma 11 and 12. Let $\delta = \min\{\alpha, 1 - \alpha, \beta, 1 - \beta\}$. Then $0 \leq \delta \leq \frac{1}{2}$ and using Lemma 12, we get the lower bound $\delta^2 g(c, c)$. Similarly, from Lemma 11 and using the fact that $\min\{h_2(\alpha), h_2(\beta)\} = h_2(\delta)$, we get the lower bound $g(c, c) - h_2(\delta)$. Combining the two, we obtain the desired lower bound

$$\tilde{g}(c) = \min_{\delta \in [0, \frac{1}{2}]} \{(g(c, c) - h_2(\delta)) \ \vee \ \delta^2 g(c, c)\}.$$

The monotonicity of $\tilde{g}$ follows from the monotonicity of $g(c, c)$ with respect to c. Also, notice that for $c > 0$, $\delta^2 g(c, c)$ is strictly positive unless $\delta = 0$ but $\lim_{\delta \to 0} g(c, c) - h_2(\delta) = g(c, c)$, which is strictly positive if $c > 0$. Therefore, for $c > 0$, we have $\tilde{g}(c) > 0$. ∎

REFERENCES

[1] C. Shannon,"A mathematical theory of communications, I and II," Bell Systems Technical Journal, vol. 27, pp. 379-423, 1948.

[2] A. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," Information and Control, vol. 2, no. 2, pp. 101-112, 1959.

[3] N. Blachman, "The convolution inequality for entropy powers," IEEE Transactions on Information Theory, vol. 11, no. 2, pp. 267–271, 1965.

[4] E. Lieb, "Proof of an entropy conjecture of Wehrl," Communications in Mathematical Physics, vol. 62, no. 1, pp. 35-41, 1978.

[5] P. Bergmans, "Random coding theorem for broadcast channels with degraded components," IEEE Transactions on Information Theory, vol. 19, no. 2, pp. 197–207, 1973.

[6] S. Leung-Yan-Cheong and M. Hellman, "The Gaussian wire-tap channel," IEEE Transactions on Information Theory, vol. 24, no. 4, pp. 451-456, 1978.

[7] L. Ozarow, "On a source-coding problem with two channels and three receivers," Bell Syst. Tech. J, vol. 59, no. 10, pp. 1909–1921, 1980.

[8] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," IEEE Transactions on Information Theory, vol. 44, no. 3, pp. 1057-1070, 1998.

[9] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," IEEE Transactions on Information Theory, vol. 52, no. 9, pp. 3936–3964, 2006.

[10] S. Haghighatshoar, E. Abbe, E. Telatar, "Adaptive sensing using deterministic partial Hadamard matrices," In Proc. International Symposium on Information Theory, pp. 1842–1846, 2012.

[11] A. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications I," IEEE Transactions on Information Theory, vol. 19, no. 6, pp. 769-772, 1973.

[12] H. Witsenhausen, "Entropy inequalities for discrete channels," IEEE Transactions on Information Theory, vol. 20, no. 5, pp. 610-616, 1974.

[13] P. Harremoes, C. Vignat, "An entropy power inequality for the binomial family," Journal of Inequalities in Pure and Applied Mathematics, vol. 4, no. 5, 2003.

[14] N. Sharma, S. Das, and S. Muthukrishnan, "Entropy power inequality for a family of discrete random variables," In Proc. of International Symposium on Information Theory, pp. 1945–1949, 2011.

[15] O. Johnson, Y. Yu, "Monotonicity, thinning, and discrete versions of the entropy power inequality," IEEE Transaction on Information Theory, vol. 56, pp. 5387– 5395, 2010.

[16] A. Jog, V. Anantharam, "The Entropy Power Inequality and Mrs. Gerber's Lemma for Abelian Groups of Order 2^n," arXiv:1207.6355, 2012.

[17] T. Tao,"Sumset and inverse sumset theory for Shannon entropy," Combinatorics, Probability & Computing, vol. 19, no. 4, pp. 603-639, 2010.

[18] I. Kontoyiannis and M. Madiman, "Sumset and inverse sumset inequalities for differential entropy and mutual information," arXiv:1206.0489, 2012.

[19] E. Arıkan, "Source polarization," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, 2010.