The Entropy Power Inequality and Mrs. Gerber's Lemma for Groups of order 2^n

Varun Jog
EECS, UC Berkeley
Berkeley, CA-94720
Email: varunjog@eecs.berkeley.edu

Venkat Anantharam
EECS, UC Berkeley
Berkeley, CA-94720
Email: ananth@eecs.berkeley.edu

***Abstract*—Shannon's Entropy Power Inequality (EPI) can be viewed as characterizing the minimum differential entropy achievable by the sum of two independent random variables with fixed differential entropies. The EPI is a powerful tool and has been used to resolve a number of problems in information theory. In this paper we examine the existence of a similar entropy inequality for discrete random variables. We obtain an entropy power inequality for random variables taking values in any group of order 2^n, i.e. for such a group G we explicitly characterize the function $f_G(x, y)$ giving the minimum entropy of the group product of two independent G-valued random variables with respective entropies x and y. Random variables achieving the extremum in this inequality are thus the analogs of Gaussians, and these are also determined. It turns out that $f_G(x, y)$ is convex in x for fixed y and, by symmetry, convex in y for fixed x. This is a generalization to groups of order 2^n of the result known as Mrs. Gerber's Lemma.**

***Keywords*: Entropy, Entropy power inequality, Mrs. Gerber's Lemma, Finite groups.**

I. Introduction

The Entropy Power Inequality (EPI) relates to the so called "entropy power" of $\mathbb{R}^n$-valued random variables having densities with well defined differential entropies. It was first proposed by Shannon in 1948 [1], who also gave sufficient conditions for equality to hold. The entropy power of an $\mathbb{R}^n$-valued random variable $\mathbf{X}$ is defined as the per-coordinate variance of a circularly symmetric $\mathbb{R}^n$-valued Gaussian random variable with the same differential entropy as $\mathbf{X}$.

Theorem I.1 (Entropy Power Inequality). *For an $\mathbb{R}^n$-valued random variable $\mathbf{X}$, the entropy power of $\mathbf{X}$ is defined to be*

$$N(\mathbf{X}) = \frac{1}{2\pi e} e^{\frac{2}{n} h(\mathbf{X})}, \tag{1}$$

where $h(\mathbf{X})$ stands for the differential entropy of X. If $\mathbf{X}$ and $\mathbf{Y}$ are independent $\mathbb{R}^n$-valued random variables, the EPI states that entropy power is a super-additive function, that is

$$N(\mathbf{X}) + N(\mathbf{Y}) \leq N(\mathbf{X} + \mathbf{Y}), \tag{2}$$

with equality if and only if $\mathbf{X}$ and $\mathbf{Y}$ are Gaussian with proportional covariance matrices.

Shannon used a variational argument to show that $\mathbf{X}$ and $\mathbf{Y}$ being Gaussian with proportional covariance matrices and having the required entropies is a stationary point for $h(\mathbf{X} + \mathbf{Y})$, but this did not exclude the possibility of it being a local minimum or a saddle point. The first rigorous proof of (2) was given by Stam [2] in 1959 based on De Bruijn's identity, which couples Fisher information with differential entropy. Stam's proof was further simplified by Blachman [3].

There have been several attempts to obtain discrete versions of the EPI. For the binary symmetric channel (BSC), Wyner and Ziv [4], [5] proved a result called Mrs. Gerber's Lemma (MGL), see Theorem I.2 below, which was extended to arbitrary binary input-output channels by Witsenhausen [6]. Shamai and Wyner [7] used MGL to give a binary analog of the EPI. A version of the EPI for binomial random variables was proved in [8] and [9]. Johnson and Yu [10] obtained an EPI using the concept of Renyi thinning.

In this paper we take a different approach towards getting a discrete analog of the EPI. Even though the EPI is interpreted as an inequality in terms of the "entropy power" of random variables, it is essentially a sharp lower bound on the differential entropy of a sum of independent random variables in terms of their individual differential entropies. With discrete random variables, as long as the "sum" operation is defined we can arrive at an analogous lower bound, except with entropies instead of differential entropies. A natural case to consider is when the random variables take values in a finite group G with the binary operation '$\cdot$' and to define the function $f_G : [0, \log |G|] \times [0, \log |G|] \to [0, \log |G|]$ by

$$f_G(x, y) = \min_{H(X)=x, H(Y)=y} H(X \cdot Y). \tag{3}$$

We can then exploit the group structure and try to arrive at the explicit form of f_G.

Let us now consider a special case: $G = \mathbb{Z}_2$. We note that on $\mathbb{Z}_2$, there is a unique distribution (up to rotation) corresponding to a fixed value of entropy. We can use this to simplify $f_{\mathbb{Z}_2}$ by writing it in terms of the inverse of binary entropy, $h^{-1} : [0, \log 2] \to [0, \frac{1}{2}]$ as

$$f_{\mathbb{Z}_2}(x, y) = h(h^{-1}(x) \star h^{-1}(y)), \tag{4}$$

where $a \star b = a\bar{b} + b\bar{a}$. This is precisely the function for which Wyner and Ziv's MGL is applicable, in fact we can restate MGL in terms of $f_{\mathbb{Z}_2}$:

Theorem I.2 (Mrs. Gerber's Lemma). *$f_{\mathbb{Z}_2}(x, y)$ is convex in y for a fixed x, and by symmetry convex in x for a fixed y.*

A similar function can also be defined for $\mathbb{R}$,

$$f_{\mathbb{R}}(x,y) = \frac{1}{2}\log\left(e^{2x}+e^{2y}\right), \tag{5}$$

which satisfies the convexity property described by MGL. In fact $f_{\mathbb{R}}$ is jointly convex in (x,y). We can however easily check that $f_{\mathbb{Z}_2}$ is not jointly convex in (x,y) since $f_{\mathbb{Z}_2}(x,x) > x = \frac{x}{\log 2} f_{\mathbb{Z}_2}(\log 2, \log 2) + \left(1 - \frac{x}{\log 2}\right) f_{\mathbb{Z}_2}(0,0)$.

It seems natural to make the following conjecture:

Conjecture 1 (Generalized MGL). *If G is a finite group, then $f_G(x,y)$ is convex in x for a fixed y, and convex in y for a fixed x.*

If one is less optimistic one might make this conjecture only for abelian groups. We have carried out simulations to test Conjecture 1 for $\mathbb{Z}_3$ and $\mathbb{Z}_5$ and it appears to hold for these groups. In this paper we prove Conjecture 1 for all groups G of order 2^n. In fact we arrive at an explicit description of f_G in terms of $f_{\mathbb{Z}_2}$ for such groups. We also characterize those distributions where the minimum entropy is attained – these distributions are in this sense analogous to Gaussians in the real case. Our results support the intuition that to minimize the entropy of the group product, the random variables X and Y should be supported on the smallest possible subgroup of G (or cosets of the same) which can support them while satisfying the constraints $H(X) = x$ and $H(Y) = y$.

The structure of this paper is as follows - In section II we consider the function $f_{\mathbb{Z}_2}$ and derive certain lemmas regarding the behaviour of $f_{\mathbb{Z}_2}$ along straight lines. In section III, we use the preceding lemmas to explicitly compute $f_{\mathbb{Z}_4}$. The techniques used in this section are then recycled in section IV where we determine the form of f_G for any group G of size 2^n. In section V we discuss some other groups that seem natural to consider, and the challenges in getting a similar result for them. Since f_G is explicitly determined for all groups of order 2^n we have in effect proved an EPI for such groups. Further, the f_G we find verifies Conjecture 1 and so proves MGL for all groups of order 2^n.

II. Preliminary Inequalities for $f_{\mathbb{Z}_2}$

Consider $f : [0,\log 2] \times [0, \log 2] \to [0, \log 2]$ given by

$$f(x,y) = h(h^{-1}(x) \star h^{-1}(y)) \ .$$

Of course $f = f_{\mathbb{Z}_2}$, where $f_{\mathbb{Z}_2}$ is defined in equation (3), but it is convenient to drop the subscript in this section.

For our first lemma, we consider lines of slope $0 \le \theta \le \infty$ passing through the origin. The result we wish to prove is:

Lemma II.1. *$\frac{\partial f}{\partial x}$ (and by symmetry, $\frac{\partial f}{\partial y}$) strictly decreases along lines through the origin having slope θ, where $0 < \theta < \infty$.*

Remark II.1. When $\theta = 0$, $\frac{\partial f}{\partial x}$ is constant and is equal to 1 and when $\theta = +\infty$, $\frac{\partial f}{\partial x}$ is constant and equal to 0. The above lemma claims that for all other values $\theta \in (0,\infty)$, $\frac{\partial f}{\partial x}(x, \theta x)$ strictly decreases in x.

Lemma II.2. *For a fixed $x \neq 0$, $\frac{\partial f}{\partial x}$ strictly decreases as y increases. By symmetry, for a fixed $y \neq 0$, $\frac{\partial f}{\partial y}$ strictly decreases as x increases.*

Lemma II.3. $\left|\frac{\partial f}{\partial x}\right| \le 1$ *(and by symmetry, $\left|\frac{\partial f}{\partial y}\right| \le 1$), with strict inequality if (x,y) lies in the interior of the square $[0,\log 2] \times [0, \log 2]$.*

Proofs of Lemma II.1, II.2 and II.3: The above lemmas are all proved by first parametrizing $f(x,y)$ in (p,q) where $x = h(p)$ and $y = h(q)$ with $0 \le p,q \le \frac{1}{2}$. Using this parametrization, $\frac{\partial f}{\partial x}$ can be computed to be

$$\begin{aligned}\frac{\partial f}{\partial x} =& \frac{\partial f}{\partial p}\frac{\partial p}{\partial x}, \\ =& (1-2q)\log\left(\frac{1-p\star q}{p \star q}\right) \Big/ \log\left(\frac{1-p}{p}\right) \ .\end{aligned}$$

Lemma II.2 follows immediately from this form of f, whereas Lemma II.3 is an easy consequence of L'Hopital's rule and MGL. Lemma II.1, however, is much more harder to prove and involves taking the derivative of $\frac{\partial f}{\partial x}$ along lines through the origin, and showing that it is negative. For details, we refer the reader to [11]. ■

Lemma II.4. *$f(x,y)$ is concave along lines through the origin. More precisely, $f(x,y)$ is concave along the line $y = \theta x$ when $0 \le \theta \le \infty$, and strictly concave along this line for $0 < \theta < \infty$.*

Proof of Lemma II.4: When $\theta = 0$ or ∞, $f(x,y)$ is linear along the line $y = \theta x$, thus concave. For $0 < \theta < \infty$, by Lemma II.1, we have that $\frac{\partial f}{\partial x}$ strictly decreases along lines through the origin. By symmetry, it follows that $\frac{\partial f}{\partial y}$ also strictly decreases along lines through the origin. Since

$$\frac{df(x,\theta x)}{dx} = \frac{\partial f}{\partial x}(x,\theta x) + \theta\frac{\partial f}{\partial y}(x, \theta x) \ , \tag{6}$$

it is immediate that $\frac{df(x,\theta x)}{dx}$ also strictly decreases in x, which means that $f(x,y)$ is strictly concave along the line $y = \theta x$. ■

Lemma II.5. *If $(x_1,y_1),(x_2,y_2) \in (0,\log 2)\times(0,\log 2)$ and $\left(\frac{\partial f}{\partial x},\frac{\partial f}{\partial y}\right)\Big|_{(x_1,y_1)} = \left(\frac{\partial f}{\partial x},\frac{\partial f}{\partial y}\right)\Big|_{(x_2,y_2)}$ then $(x_1,y_1) = (x_2,y_2)$.*

Remark II.2. The above lemma says that in the interior of the unit square, the pair of partial derivatives at a point uniquely determine the point. That this fails on the boundary is seen from the fact that for any point of the form $(x,0)$ the pair of partial derivatives evaluates to $(1,0)$ and for every point of the form $(0,y)$ it is $(0,1)$.

Proof of Lemma II.5: Without loss of generality, assume $x_1 \le x_2$. We consider two cases: $y_1 \ge y_2$ or $y_1 < y_2$. Suppose $y_1 \ge y_2$, in this case we have

$$\left.\frac{\partial f}{\partial x}\right|_{(x_1,y_1)} \le \left.\frac{\partial f}{\partial x}\right|_{(x_2,y_1)} \le \left.\frac{\partial f}{\partial x}\right|_{(x_2,y_2)} \ . \tag{7}$$

The first inequality follows from MGL whereas the second inequality follows from Lemma II.2. Note also that at least one of the two inequalities is strict as $(x_1, y_1) \neq (x_2, y_2)$. Thus

$$\left.\frac{\partial f}{\partial x}\right|_{(x_1,y_1)} < \left.\frac{\partial f}{\partial x}\right|_{(x_2,y_2)} . \tag{8}$$

It remains to consider the case $y_1 < y_2$. We can also assume $x_1 < x_2$, since $x_1 = x_2$ combined with $y_1 < y_2$ gives

$$\left.\frac{\partial f}{\partial x}\right|_{(x_1,y_1)} > \left.\frac{\partial f}{\partial x}\right|_{(x_2,y_2)} .$$

Thus we assume $(x_1, y_1) < (x_2, y_2)$. Consider the line passing through the origin and (x_1, y_1). We consider two cases: either $y_2 \geq x_2 \frac{y_1}{x_1}$ or $y_2 \leq x_2 \frac{y_1}{x_1}$.
If $y_2 \geq x_2 \frac{y_1}{x_1}$,

$$\left.\frac{\partial f}{\partial x}\right|_{(x_1,y_1)} > \left.\frac{\partial f}{\partial x}\right|_{(x_2,x_2\frac{y_1}{x_1})} \geq \left.\frac{\partial f}{\partial x}\right|_{(x_2,y_2)} , \tag{9}$$

where the first inequality follows from Lemma II.1, and the second follows from Lemma II.2. If $y_2 \leq x_2 \frac{y_1}{x_1}$, we have

$$\left.\frac{\partial f}{\partial y}\right|_{(x_1,y_1)} > \left.\frac{\partial f}{\partial y}\right|_{(y_2\frac{x_1}{y_1},y_2)} \geq \left.\frac{\partial f}{\partial y}\right|_{(x_2,y_2)} , \tag{10}$$

where the first inequality follows from Lemma II.1 and the fact that $y_2 \frac{x_1}{y_1} > x_1$. The second inequality follows Lemma II.2. ■

III. An EPI and MGL for $\mathbb{Z}_4$-valued random variables

To illustrate the general structure of the proof, we consider the case of the cyclic group $\mathbb{Z}_4$. We consider two independent random variables X and Y taking values in $\mathbb{Z}_4$ and seek to determine the minimum possible entropy of the random variable $X + Y$, where $+$ stands for the group addition, and we a priori fix the entropy of X and that of Y.

Formally, we define $f_4 : [0, \log 4] \times [0, \log 4] \to [0, \log 4]$ by

$$f_4(x, y) = \min_{H(X)=x, H(Y)=y} H(X + Y) . \tag{11}$$

Thus $f_4 = f_{\mathbb{Z}_4}$, where $f_{\mathbb{Z}_4}$ is defined in equation (3). In this section we will also use the notation f_2 for $f_{\mathbb{Z}_2}$.

We will prove:

Theorem III.1.

$$f_4(x, y) = \begin{cases} x, & \textit{if } \log 2 \leq x \leq \log 4, 0 \leq y \leq \log 2 \ , \\ y, & \textit{if } 0 \leq x \leq \log 2, \log 2 \leq y \leq \log 4 \ , \\ f_2(x, y), & \textit{if } 0 \leq x, y \leq \log 2 \ , \\ f_2(x - \log 2, y - \log 2) + \log 2, & \textit{o.w.} \end{cases}$$

Corollary III.1. *$f_4(x, y)$ is convex in x for a fixed y, and by symmetry convex in y for a fixed x.*

Proof of Theorem III.1: We deal with the initial two cases first. Without loss of generality, assume $\log 2 \leq x \leq \log 4, 0 \leq y \leq \log 2$. Note that we have the trivial lower bound $f_4(x, y) \geq x$, obtained from $H(X + Y) \geq H(X)$. Since $y \leq \log 2$, let $\beta = h^{-1}(y)$ and consider the distribution of Y, $p_Y := (\beta, 0, 1 - \beta, 0)$. Also, as $\log 2 \leq x$, we can find α such that $\log 2 + H(\alpha, 1 - \alpha) = x$. Using this α, define $p_X := \left(\frac{\alpha}{2}, \frac{1-\alpha}{2}, \frac{\alpha}{2}, \frac{1-\alpha}{2}\right)$. The distribution of $X + Y$ is given by the cyclic convolution $p_X \circledast_4 p_Y$, which in this case is p_X again. Thus $H(X + Y) = H(X)$, and $f_4(x, y)$ equals the lower bound x.

Before starting on the other two cases, we derive some preliminary inequalities. We'll think of distributions on $\mathbb{Z}_4$ as a mixture of two distributions: p_E supported on $\{0, 2\}$ and p_O supported on $\{1, 3\}$. For a random variable X, we write its distribution $p_X = \alpha p_E + (1 - \alpha)p_O$, where $1 \geq \alpha \geq 0$. Similarly, we write $p_Y = \beta q_E + (1 - \beta)q_O$, where $1 \geq \beta \geq 0$. Let $X + Y = Z$. The distribution of Z is given by

$$\begin{aligned} p_Z &= p_X \circledast_4 p_Y \\ &= (\alpha p_E + (1 - \alpha)p_O) \circledast_4 (\beta q_E + (1 - \beta)q_O) \\ &= (\alpha\beta p_E \circledast_4 q_E + (1 - \alpha)(1 - \beta)p_O \circledast_4 q_O) \\ &\quad + (\alpha(1 - \beta)p_E \circledast_4 q_O + (1 - \alpha)\beta p_O \circledast_4 q_E) \ . \end{aligned}$$

The following remarkable sequence of inequalities forms the core of this paper.

$$\begin{aligned} &H(p_Z) \overset{(a)}{=} h(\alpha \star \beta) \\ &+ \overline{\alpha \star \beta} H\left(\frac{\alpha\beta}{\overline{\alpha \star \beta}} p_E \circledast_2 q_E + \frac{\bar{\alpha}\bar{\beta}}{\overline{\alpha \star \beta}} p_O \circledast_2 q_O\right) \\ &+ (\alpha \star \beta) H\left(\frac{\alpha\bar{\beta}}{\alpha \star \beta} p_E \circledast_2 q_O + \frac{\bar{\alpha}\beta}{\alpha \star \beta} p_O \circledast_2 q_E\right) \\ &\overset{(b)}{\geq} h(\alpha \star \beta) + \alpha\beta H(p_E \circledast_2 q_E) + \bar{\alpha}\bar{\beta} H(p_O \circledast_2 q_O) \\ &+ \alpha\bar{\beta} H(p_E \circledast_2 q_O) + \bar{\alpha}\beta H(p_O \circledast_2 q_E) \\ &\overset{(c)}{=} f_2(h(\alpha), h(\beta)) \\ &+ \alpha\beta f_2(H(p_E), H(q_E)) + \bar{\alpha}\bar{\beta} f_2(H(p_O), H(q_O)) \\ &+ \alpha\bar{\beta} f_2(H(p_E), H(q_O)) + \bar{\alpha}\beta f_2(H(p_O), H(q_E)) \\ &\overset{(d)}{\geq} f_2(h(\alpha), h(\beta)) + \alpha f_2(H(p_E), \beta H(q_E) + \bar{\beta} H(q_O)) \\ &+ \bar{\alpha} f_2(H(p_O), \beta H(q_E) + \bar{\beta} H(q_O)) \\ &\overset{(e)}{\geq} f_2(h(\alpha), h(\beta)) \\ &+ f_2(\alpha H(p_E) + \bar{\alpha} H(p_O), \beta H(q_E) + \bar{\beta} H(q_O)) \\ &= f_2(h(\alpha), h(\beta)) + f_2(H(X) - h(\alpha), H(Y) - h(\beta)) \ . \end{aligned}$$

Here (a) is a simple expansion of entropy, (b) is got via concavity of entropy, (c) is a restatement in terms of f_2, (d) and (e) are obtained using convexity in MGL, and the last equality follows from the chain rule of entropy. We note that concavity of entropy, and convexity in MGL are the only two ingredients needed to arrive at this lower bound.
We thus obtain the lower bound

$$f_4(x, y) \geq \min_{u,v} f_2(u, v) + f_2(x - u, y - v), \tag{12}$$

where the range of u and v is such that the right hand side makes sense.

In the third case, when $0 \leq x, y \leq \log 2$, consider the function $g : [0, x] \times [0, y] \to \mathbb{R}$ given by

$$g(u, v) := f_2(u, v) + f_2(x - u, y - v) \ .$$

As per (12), we want to minimize g over its domain. We can think of the domain as a closed rectangle with corner points $(0,0)$ and (x,y) in $\mathbb{R}^2$. Suppose the minimum is achieved in the interior of this rectangle, at a point say $(u^\star, v^\star)$, then we must have

$$\frac{\partial g}{\partial u}\bigg|_{(u^\star,v^\star)} = 0 \ , \frac{\partial g}{\partial v}\bigg|_{(u^\star,v^\star)} = 0 \ , \tag{13}$$

which implies

$$\left(\frac{\partial f_2}{\partial u}, \frac{\partial f_2}{\partial v}\right)\bigg|_{(u^\star,v^\star)} = \left(\frac{\partial f_2}{\partial u}, \frac{\partial f_2}{\partial v}\right)\bigg|_{(x-u^\star,y-v^\star)} . \tag{14}$$

By Lemma II.5, we infer that

$$(u^\star, v^\star) = (x-u^\star, y-v^\star) \ . \tag{15}$$

Thus $(u^\star, v^\star) = \left(\frac{x}{2}, \frac{y}{2}\right)$. Now let $\theta = \frac{y}{x}$, and consider the function g over the line with slope θ passing through the origin. By Lemma II.4, we know that $f_2(t, \theta t)$ is concave, and thus so is $f_2(x-t, y-\theta t)$ and so is their addition $g(t, \theta t)$. Thus, the minimum value of $g(t, \theta t)$ must be attained at the extreme points and not in the interior. Note that since $(u^\star, v^\star)$ lies on this line, it cannot be the global minimum of g on its domain. This leads us to conclude that the global minimum of g is not attained anywhere in the interior of the rectangle and therefore must be attained on the boundary.

Now consider a point $(u_0, 0)$ along the boundary. Taking the partial derivative with respect to u,

$$\frac{\partial g}{\partial u}\bigg|_{(u_0,0)} = \frac{\partial f_2}{\partial u}\bigg|_{(u_0,0)} - \frac{\partial f_2}{\partial u}\bigg|_{(x-u_0,y)} > 0 \ , \tag{16}$$

where the inequality follows from $\frac{\partial f_2}{\partial u}|_{(u_0,0)} = 1$ and $\frac{\partial f_2}{\partial u}|_{(x-u_0,y)} < 1$ by Lemma II.3. Similarly, for a boundary point of the form $(0, v_0)$ we can see that $\frac{\partial g}{\partial v}|_{(0,v_0)} > 0$. Hence, we conclude that the minimum value on the boundary is attained when $u = 0, v = 0$ and the value is $f_2(x,y)$. Thus inequality (12) reduces to

$$f_4(x,y) \geq f_2(x,y) \ . \tag{17}$$

Clearly, $f_2(x,y)$ is achieved if the random variables are supported on the $\{0,2\}$, and therefore we get

$$f_4(x,y) = f_2(x,y) \text{ for } 0 \leq x,y \leq \log 2 \ . \tag{18}$$

This completes the proof for the third case.

Moving on to the last case, define $\tilde{x} = x - \log 2$, $\tilde{y} = y - \log 2$. Let $\tilde{u} = u - \tilde{x}$ and $\tilde{v} = v - \tilde{y}$. Rewriting (12),

$$f_4(x,y) \geq \min_{\tilde{u},\tilde{v}} f_2(\log 2 - \tilde{u}, \log 2 - \tilde{v}) + f_2(\tilde{u}+\tilde{x}, \tilde{v}+\tilde{y}) \ , \tag{19}$$

where $0 \leq \tilde{u} \leq \log 2 - \tilde{x}$, $0 \leq \tilde{v} \leq \log 2 - \tilde{y}$. Just as in the previous case, define

$$g(\tilde{u}, \tilde{v}) := f_2(\log 2 - \tilde{u}, \log 2 - \tilde{v}) + f_2(\tilde{u}+\tilde{x}, \tilde{v}+\tilde{y}) \ .$$

The domain of $(\tilde{u}, \tilde{v})$ can be thought of as a closed rectangle in $\mathbb{R}^2$ with corner points $(0,0)$, $(\log 2 - \tilde{x}, \log 2 - \tilde{y})$. Suppose the minimum value is attained at $(\tilde{u}^\star, \tilde{v}^\star)$. Now using the same ideas as in the third case, we conclude that $(\tilde{u}^\star, \tilde{v}^\star)$ cannot lie in the interior of the rectangle, and on the boundary the minimum is attained when $(\tilde{u}, \tilde{v}) = (0,0)$ with it's value being $f_2(\log 2, \log 2) + f_2(\tilde{x}, \tilde{y}) = \log 2 + f_2(\tilde{x}, \tilde{y})$.

Thus inequality (12) reduces to

$$f_4(x,y) \geq \log 2 + f_2(x - \log 2, y - \log 2) \ . \tag{20}$$

Since $\log 2 \leq x, y \leq \log 4$, we can find distributions $p_X = (\frac{\alpha}{2}, \frac{1-\alpha}{2}, \frac{\alpha}{2}, \frac{1-\alpha}{2})$ and $p_Y = (\frac{\beta}{2}, \frac{1-\beta}{2}, \frac{\beta}{2}, \frac{1-\beta}{2})$ such that $H(p_X) = x$ and $H(p_Y) = y$. It's easy to check that $H(X+Y) = f_2(x - \log 2, y - \log 2) + \log 2$. Thus the bound in (20) is achieved, and we conclude that

$$f_4(x,y) = \log 2 + f_2(x - \log 2, y - \log 2) \ . \tag{21}$$

This completes the proof of Theorem III.1. ∎

Proof of Corollary III.1: Consider the function $f_x(y) = f(x,y)$. We look at two cases, $0 \leq x \leq \log 2$ and $\log 2 \leq x \leq \log 4$. In the first case,

$$f_x(y) = \begin{cases} f_2(x,y), & \text{if } 0 \leq y \leq \log 2 \ , \\ y & \text{if } \log 2 \leq y \leq \log 4 \ . \end{cases}$$

Now $f_2(x,y)$ for a fixed x and $0 \leq y \leq \log 2$ is convex by MGL, and for values of y beyond $\log 2$ the function f_x is linear with slope 1. By Claim II.3, attaching this linear part to a convex function will not affect the convexity since the slope of the linear part (= 1) is greater than or equal to the derivative of the convex part. Similarly for the second case,

$$f_x(y) = \begin{cases} x, & \text{if } 0 \leq y \leq \log 2 \ , \\ f_2(x - \log 2, y - \log 2) + \log 2 \ , & \text{o.w.} \end{cases}$$

This too, has a linear part with slope 0 attached before a convex part with slope greater equal 0 everywhere, thus the overall function continues being convex. ∎

IV. An EPI and MGL over groups of order 2^n

Let G be any group of order 2^n, with the binary operation '$\cdot$'. We look at the function $f_G : [0, n\log 2] \times [0, n\log 2] \to [0, n\log 2]$ by

$$f_G(x,y) = \min_{H(X)=x, H(Y)=y} H(X \cdot Y) \ . \tag{22}$$

Theorem IV.1. *f_G depends only on the size of G, and is denoted by f_{2^n}, where*

$$f_{2^n}(x,y) = \begin{cases} f_2(x - k\log 2, y - k\log 2) + k\log 2, \\ \qquad \textit{if } k\log 2 \leq x, y \leq (k+1)\log 2 \ , \\ \max(x,y) \quad \textit{otherwise.} \end{cases}$$

The following corollary is an immediate consequence of Theorem IV.1 and Mrs. Gerber's Lemma:

Corollary IV.1. *$f_G(x,y)$ is convex in x for a fixed y, and by symmetry is convex in y for a fixed x.*

Proof of Theorem IV.1: We deal with the second case first. Assume

$$k_1 \log 2 \leq x \leq (k_1+1)\log 2 \ , k_2 \log 2 \leq y \leq (k_2+1)\log 2 \ ,$$

where $k_1 \neq k_2$. Suppose $k_1 > k_2$. Note that we have the trivial lower bound $f_G(x,y) \geq x$, obtained from $H(X \cdot Y) \geq H(X)$.

Let H be a subgroup of G of order 2^{k_1+1}, such a group always exists by Sylow's theorem [12]. Let $\tilde{H}$ be a subgroup of H of size 2^{k_1}. Note that, regardless of whether G is abelian or not, for every $h \in H$ we have $h\tilde{H} = \tilde{H}h$, and the cosets of $\tilde{H}$ in H are therefore unambiguously defined. We fix p_X to be supported on H and constant on the two cosets of $\tilde{H}$ in H, while satisfying $H(X) = x$. Let p_Y be any distribution supported on $\tilde{H}$ such that $H(Y) = y$. It's easy to check that $X \cdot Y$ has the distribution p_X, giving $H(X \cdot Y) = H(X)$, which further coupled with the trivial lower bound implies $f_G(x, y) = x$. In case $k_1 < k_2$, we follow the exact same strategy except we choose H and $\tilde{H}$ to be of sizes 2^{k_2+1} and 2^{k_2} respectively and interchange the roles of p_X and p_Y.

For the remaining case, our proof proceeds by induction, we assume the result for groups of order 2^{n-1} and prove it for groups of order 2^n. Let H be a subgroup of G of order 2^{n-1}. Just as in the case of $\mathbb{Z}_4$, we express distributions on G as a convex combination of distributions supported on H and the coset of H in G. We can write

$$p_X = \alpha p_E + (1-\alpha) p_O \ ,$$

where $1 \geq \alpha \geq 0$, with p_E supported only on H and p_O supported on the coset of H. Similarly we write

$$p_Y = \beta q_E + (1-\beta) q_O \ ,$$

where $1 \geq \beta \geq 0$. Following the same steps (a) and (b) as in the proof of Theorem III.1, we arrive at the bound

$$\begin{aligned} H(X \cdot Y) \geq h(\alpha \star \beta) + \alpha\beta H(p_E \circledast_G q_E) + \bar{\alpha}\bar{\beta} H(p_O \circledast_G q_O) \\ + \alpha\bar{\beta} H(p_E \circledast_G q_O) + \bar{\alpha}\beta H(p_O \circledast_G q_E) \ . \end{aligned}$$

Here, for two probability distributions p and q on G, $p \circledast_G q$ denotes the probability distribution given by $p \circledast_G q(g) = \sum_{\tilde{g}} p(\tilde{g}) q(\tilde{g}^{-1} g)$. To justify the analog of step (c) in the proof of Theorem III.1, pick any $g' \notin H$, and note that p_O and q_O are supported on $g'H = Hg'$. We can 'translate' q_O to be supported on H by right multiplying the support of q_O by g'^{-1}. Call this translated distribution $\tilde{q}_O$. Now $p_E \circledast_G q_O$ is simply a translated version $p_E \circledast_G \tilde{q}_O$ obtained via right multiplying by g'. Thus

$$\begin{aligned} & H(p_E \circledast_G q_O) = H(p_E \circledast_G \tilde{q_O}) \\ & \geq f_{2^{n-1}}(H(p_E), H(\tilde{q_O})) = f_{2^{n-1}}(H(p_E), H(q_O)). \end{aligned}$$

In a similar way, for $p_O \circledast_G q_E$, we can translate p_O to $\tilde{p}_O$ via left multiplying by g'^{-1}, and conclude that $p_O \circledast_G q_E$ is a translated version of $\tilde{p_O} \circledast_G q_E$ obtained via left multiplication by g'. For $p_O \circledast_G q_O$, we translate p_O to $\tilde{p}_O$ by left multiplication and q_O to $\tilde{q}_O$ by right multiplication by g'^{-1} and conclude that $p_O \circledast_G q_O$ is a doubly translated version of $\tilde{p_O} \circledast_G \tilde{q_O}$, via left and right multiplication by g'. Thus the step analogous to (c) in the proof of Theorem III.1 is valid, and the remaining steps follow by our induction hypothesis to yield the lower bound

$$H(X \cdot Y) \geq \min_{u,v} f_2(u, v) + f_{2^{n-1}}(x-u, y-v) \ .$$

With this bound in hand, the rest of the proof, which involves an explicit computation of the lower bound and then demonstrating that it is achieved, goes through almost exactly as in the case of $\mathbb{Z}_4$ valued random variables. We refer the reader to [11] for details. For $k < n-1$, the optimal distributions turn out to be those which are supported on H and are optimal for it. For $k = n-1$, the optimal distributions are those which take constant values on the cosets of H in G. ■

V. Discussion

While this paper looks at groups of size 2^n, it is natural to ask the same questions for other groups. Note that to prove the $\mathbb{Z}_4$ case, we needed MGL for $\mathbb{Z}_2$ to arrive at the lower bound in (12). For $\mathbb{Z}_9$, proving the generalized MGL for $\mathbb{Z}_3$ is hard since the optimal distributions are not easy to parametrize. Further, even with an MGL we would still require Lemmas II.1, II.2 and II.3 for the proof technique in this paper to go through. However this need not be the only line of attack and a different strategy can possibly avoid requiring these Lemmas.

Acknowledgements

Research support from the ARO MURI grant W911NF-08-1-0233, "Tools for the Analysis and Design of Complex Multi-scale Network", from the NSF grant CNS-0910702, from the NSF Science & Technology Center grant CCF-0939370, "Science of Information", from Marvell Semiconductor Inc., and from the U.C. Discovery program is gratefully acknowledged. The authors would also like to thank the anonymous reviewers for their comments that helped improve the manuscript.

References

[1] C. Shannon, "A mathematical theory of communications, I and II," *Bell Syst. Tech. J*, 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," *Information Theory, IEEE Transactions on*, vol. 11, no. 2, pp. 267–271, 1965.

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

[5] A. Wyner, "A theorem on the entropy of certain binary sequences and applications–II," *Information Theory, IEEE Transactions on*, vol. 19, no. 6, pp. 772–777, 1973.

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

[7] S. Shamai and A. Wyner, "A binary analog to the entropy-power inequality," *Information Theory, IEEE Transactions on*, vol. 36, no. 6, pp. 1428–1430, 1990.

[8] P. Harremoes, C. Vignat, *et al.*, "An entropy power inequality for the binomial family," *JIPAM. J. Inequal. Pure Appl. Math*, vol. 4, no. 5, 2003.

[9] N. Sharma, S. Das, and S. Muthukrishnan, "Entropy power inequality for a family of discrete random variables," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pp. 1945–1949, IEEE, 2011.

[10] O. Johnson and Y. Yu, "Monotonicity, thinning, and discrete versions of the entropy power inequality," *Information Theory, IEEE Transactions on*, vol. 56, no. 11, pp. 5387–5395, 2010.

[11] V. Jog and V. Anantharam, "The entropy power inequality and Mrs. Gerber's lemma for abelian groups of order 2^n," *arXiv preprint arXiv:1207.6355*, 2012.

[12] M. Artin, "Algebra. 2nd," 2011.