On Hypercontractivity and the Mutual Information between Boolean Functions

Venkat Anantharam*, Amin Aminzadeh Gohari†, Sudeep Kamath*, Chandra Nair‡
*EECS Department, University of California, Berkeley,
{ananth, sudeep}@eecs.berkeley.edu
†EE Department, Sharif University of Technology, Tehran, Iran
aminzadeh@sharif.edu
‡IE Department, The Chinese University of Hong Kong
chandra@ie.cuhk.edu.hk

***Abstract*—Hypercontractivity has had many successful applications in mathematics, physics, and theoretical computer science. In this work we use recently established properties of the hypercontractivity ribbon of a pair of random variables to study a recent conjecture regarding the mutual information between binary functions of the individual marginal sequences of a sequence of pairs of random variables drawn from a doubly symmetric binary source.**

I. Introduction

Let (X,Y) be a pair of $\{0,1\}$-valued random variables such that X and Y are uniformly distributed and $\mathsf{Pr}(X=0,Y=1) = \mathsf{Pr}(X=1,Y=0) = \frac{1}{2}\alpha$. This joint distribution is sometimes referred to as the doubly symmetric binary source, $\mathrm{DSBS}(\alpha)$.

Define for $x \in [0,1]$ the binary entropy function $h(x) := x\log_2\frac{1}{x} + (1-x)\log_2\frac{1}{1-x}$, with the convention that $0\log_2 0 = 0$.

The following isoperimetric information inequality was conjectured by Kumar and Courtade in [**?**]. They also provided some evidence for its validity.

Conjecture 1: (Kumar-Courtade [**?**]) If $\{(X_i,Y_i)\}_{i=1}^n$ are drawn i.i.d. from $\mathrm{DSBS}(\alpha)$, and $b : \{0,1\}^n \to \{0,1\}$ is any Boolean function, then $I(b(X^n);Y^n) \leq I(X_1;Y_1) = 1-h(\alpha)$.

Using perturbation based arguments it can be shown that Conjecture **??** is equivalent to Conjecture **??** below.

Conjecture 2: If $\{(X_i,Y_i)\}_{i=1}^n$ are drawn i.i.d. from $\mathrm{DSBS}(\alpha)$, and the Markov chain $W - X^n - Y^n - Z$ holds with W binary-valued, then $I(W;Z) \leq I(X_1;Y_1) = 1-h(\alpha)$.

In this document we study a weaker form of the above conjecture, as stated below.

Conjecture 3: If $\{(X_i,Y_i)\}_{i=1}^n$ are drawn i.i.d. from $\mathrm{DSBS}(\alpha)$, and $b, b' : \{0,1\}^n \to \{0,1\}$ are any Boolean functions, then $I(b(X^n);b'(Y^n)) \leq I(X_1;Y_1) = 1-h(\alpha)$.

Remark: In the statement of Conjecture **??**, if one additionally assumes $b = b'$, then the statement is known to be true [**?**].

Since n is arbitrary in the statement of the conjecture, it is not in a form that is amenable to brute-force numerical verification. In this paper we present a stronger conjecture (Conjecture **??**) relating to an arbitrary pair of binary random variables that would imply Conjecture **??**.

Conjecture **??** relates the chordal slope of the hypercontractivity ribbon of a pair of binary random variables (X,Y) at infinity, denoted $\mathsf{s}^*(X;Y)$, to their mutual information, $I(X;Y)$. This motivates the study of $\mathsf{s}^*(X;Y)$ for binary pairs of random variables (X,Y). We provide some results about this quantity, including a certain form of duality.

A. A remark on Conjecture **??**

A natural question to ask is whether Conjectures **??** and **??** are more general, i.e. if $\{(X_i,Y_i)\}_{i=1}^\infty$ are generated i.i.d. from an *arbitrary* binary-valued pair source $\mu_{X,Y}(x,y)$ and if $b, b' : \{0,1\}^n \mapsto \{0,1\}$, then do we have $I(b(X^n);b'(Y^n)) \leq I(X_1;Y_1)$?. This can be shown to be false. For example, consider (X,Y) to have the joint distribution of a successive pair of random variables from a stationary ergodic Markov chain with state space $\{0,1\}$ with transition probabilities $P(Y=1|X=0)=\alpha, P(Y=0|X=1)=\beta$ (see Fig. **??**).

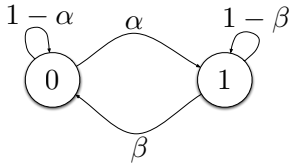

Fig. 1. A simple two-state Markov chain

Then, (X,Y) have joint distribution given by the matrix $\begin{bmatrix} \frac{\beta(1-\alpha)}{\alpha+\beta} & \frac{\alpha\beta}{\alpha+\beta} \\ \frac{\alpha\beta}{\alpha+\beta} & \frac{\alpha(1-\beta)}{\alpha+\beta} \end{bmatrix}$. For $(X_1,Y_1),(X_2,Y_2)$ drawn

i.i.d. from this joint distribution with $\alpha = 0.01, \beta = 0.04$, we can compute $I(X_1;Y_1) = 0.6088\ldots < I(X_1 \oplus X_2; Y_1 \oplus Y_2) = 0.70\ldots$. Thus, Conjectures **??** and **??** seem somewhat special, for DSBS sources only.

II. Preliminaries

Definition 1: For a pair of random variables $(X,Y) \sim \mu_{X,Y}(x,y)$ on $\mathcal{X} \times \mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ are finite sets, we define the *hypercontractivity ribbon*

$$\mathcal{R}(X;Y) \subseteq \{(\mathsf{p},\mathsf{q}) : 1 \le \mathsf{q} \le \mathsf{p}\}$$

as follows: for $1 \le \mathsf{q} \le \mathsf{p}$, we have $(\mathsf{p},\mathsf{q}) \in \mathbb{R}(X;Y)$ if

$$||\mathbb{E}[g(Y)|X]||_{\mathsf{p}} \le ||g(Y)||_{\mathsf{q}} \quad \forall g : \mathcal{Y} \mapsto \mathbb{R}. \tag{1}$$

For a given $\mathsf{p} \ge 1$, define $s^{(\mathsf{p})}(X,Y)$ as

$$s^{(\mathsf{p})}(X;Y) := \inf\{r : (\mathsf{p}, \mathsf{p}r) \in \mathcal{R}(X;Y)\}.$$

It is easy to see that $s^{(\mathsf{p})}(X;Y)$ is decreasing in p. Let

$$\mathsf{s}^*(X;Y) := \lim_{\mathsf{p}\to\infty} s^{(\mathsf{p})}(X;Y).$$

In this paper we study $\mathsf{s}^*(X;Y)$ for pairs of binary random variables, in our attempts to establish Conjecture **??**. Below, we provide some known results regarding the quantity $\mathsf{s}^*(X;Y)$. These results apply to general pairs of finite random variables.

A. Alternate characterizations of $\mathsf{s}^*(X;Y)$

Let $(X,Y) \sim p_{X,Y}(x,y)$ be finite valued random variables such that $p_X(x) > 0$ and $p_Y(y) > 0$ for every $x \in \mathcal{X}, y \in \mathcal{Y}$.

In [**?**] it was shown that

$$\mathsf{s}^*(X;Y) := \sup_{r_X \ll p_X} \frac{D(r_Y \| p_Y)}{D(r_X \| p_X)},$$

where the supremum is taken over r_X running over the set of distributions on $\mathcal{X}$ (hence is absolutely continuous with respect to p_X due to our positivity assumption of $p_X(x)$) and r_Y is the marginal distribution induced on $\mathcal{Y}$ by $r_{X,Y}(x,y) = r_X(x)p_{Y|X}(y|x)$. They also showed that $\mathsf{s}^*(X;Y)$ satisfies the following two properties:

- (T) *Tensorization*: If $\{(X_i,Y_i)\}_{i=1}^n$ are drawn i.i.d., then $\mathsf{s}^*(X^n;Y^n) = \mathsf{s}^*(X_1;Y_1)$.
- (D) *Data Processing Inequality*: If $W - X - Y - Z$ is a Markov chain, then we have $\mathsf{s}^*(X;Y) \ge \mathsf{s}^*(W;Z)$.

For $(X,Y) \sim \mathrm{DSBS}(\alpha)$, $\mathsf{s}^*(X;Y) = (1-2\alpha)^2$. This result dates back to Bonami [**?**] and Beckner [**?**], and is also independently derived in [**?**].

Recently it was shown [**?**] that

$$\mathsf{s}^*(X;Y) = \sup_{U : U-X-Y, I(U;X)>0} \frac{I(U;Y)}{I(U;X)}.$$

Given a joint distribution $p(x,y)$, consider the conditional distribution $p_{Y|X}(y|x)$ as defining a channel, $\mathcal{C}$, from X to Y. Fix this transition probability and consider the following function as we vary the input distribution:

$$t_\lambda^{\mathcal{C}}(q(x)) := H_q(Y) - \lambda H_q(X).$$

Let $\mathcal{K}(t_\lambda^{\mathcal{C}})_{q_0(x)}$ denote the lower convex envelope of the function $t_\lambda^{\mathcal{C}}(q(x))$ evaluated at the input distribution $q_0(x)$.

*Theorem 1 ([**?**]):* For $(X,Y) \sim p(x,y)$, we have

$$\mathsf{s}^*(X;Y) := \inf\{\lambda : \mathcal{K}(t_\lambda^{\mathcal{C}})_{p(x)} = t_\lambda^{\mathcal{C}}(p(x))\}. \tag{2}$$

By Theorem **??**, we know that the point $\left(p(x), t^{\mathcal{C}}_{\mathsf{s}^*_p(X;Y)}(p(x))\right)$ lies on the lower convex envelope of the curve $q(x) \mapsto t^{\mathcal{C}}_{\mathsf{s}^*_p(X;Y)}(q(x))$, where $\mathsf{s}^*_p(X;Y)$ is $\mathsf{s}^*(X;Y)$ evaluated at $p(x)p(y|x)$.

B. Lower bound on $\mathsf{s}^*(X;Y)$

$\mathsf{s}^*(X;Y)$ is bounded from below by $\rho_m(X;Y)^2$ defined as follows:

Definition 2: For jointly distributed random variables (X,Y), define their Hirschfeld-Gebelein-Rényi *maximal correlation* $\rho_m(X;Y) := \sup \mathbb{E} f(X)g(Y)$ where the supremum is over $f : \mathcal{X} \mapsto \mathbb{R}, g : \mathcal{Y} \mapsto \mathbb{R}$ such that $\mathbb{E}f(X) = \mathbb{E}g(Y) = 0$ and $\mathbb{E}f(X)^2, \mathbb{E}g(Y)^2 \le 1$.

For $(X,Y) \sim \mathrm{DSBS}(\epsilon)$, the inequality $\mathsf{s}^*(X;Y) \ge \rho_m(X;Y)^2$ holds with equality. It is easy to show that $\rho_m(X;Y) = |1-2\epsilon|$ and $\mathsf{s}^*(X;Y) = (1-2\epsilon)^2$ [**?**].

C. Main Conjecture

We will make progress towards Conjecture **??** by stating our main conjecture.

Conjecture 4: For any binary-valued random variable pair (W,Z), we have

$$h\left(\frac{1-\sqrt{\mathsf{s}^*(W;Z)}}{2}\right) + I(W;Z) \le 1, \tag{3}$$

with equality if and only if $(W,Z) \sim \mathrm{DSBS}(\alpha)$ for some $0 \le \alpha \le 1$, or if W and Z are independent.

Note that Conjecture **??** implies Conjecture **??**. Indeed, when $(X^n,Y^n) \sim \prod_i p(x_i,y_i)$ where $p(x,y)$ corresponds to $DSBS(\alpha)$ then

$$\begin{aligned}
&I(b(X^n); b'(Y^n)) \\
&\quad \le 1 - h\left(\frac{1-\sqrt{\mathsf{s}^*(b(X^n);b'(Y^n))}}{2}\right) && (4)\\
&\quad \le 1 - h\left(\frac{1-\sqrt{\mathsf{s}^*(X^n;Y^n)}}{2}\right) && (5)\\
&\quad \le 1 - h\left(\frac{1-\sqrt{\mathsf{s}^*(X;Y)}}{2}\right) && (6)\\
&\quad = 1 - h\left(\frac{1-\sqrt{(1-2\alpha)^2}}{2}\right) && (7)\\
&\quad = 1 - h(\alpha),
\end{aligned}$$

where (**??**) is from Conjecture **??**, (**??**) follows from data processing property (using $b(X^n) \to X^n \to Y^n \to b'(Y^n)$ is Markov), (**??**) follows from tensorization property of s^*, and (**??**) uses the result that $\mathsf{s}^*(X;Y) = (1-2\alpha)^2$ when $(X,Y) \sim DSBS(\alpha)$.

One advantage of Conjecture **??** over Conjecture **??** is that Conjecture **??** can be subject to numerical verification (because of the cardinality of two on W and Z). Extensive numerical simulations seems to validate Conjecture **??**. Indeed, it may be possible to obtain a computer assisted proof. However, our focus is to get an analytical proof.

Remark: It can be shown that ρ_m too satisfies the tensorization and data processing inequality properties [**?**]. Thus, if

$$h\left(\frac{1-\rho_m(W;Z)}{2}\right) + I(W;Z) \le 1, \tag{8}$$

held whenever W, Z are binary, this would have implied Conjecture **??**. However, (**??**) fails for some distributions $p_{W,Z}$ with W, Z binary-valued.

Remark: It can be shown in a similar way that if

$$h\left(\frac{1-\sqrt{\mathsf{s}^*(W;Z)}}{2}\right) + I(W;Z) \le 1, \tag{9}$$

held whenever W is binary and Z is finite-valued, then it would have implied Conjecture **??**. However, (**??**) fails for some distributions $p_{W,Z}$ when W is binary-valued and Z is ternary-valued.

III. Properties of s^*

One of the difficulties for proving Conjecture **??** analytically is that we do not have any explicit expression for s^*, except in certain special cases. This motivates studying s^* for pairs of binary valued random variables. Further Conjecture **??** provides some insights on s^* for binary valued random variables. Thus, we might ask if there are simple characterization of s^* (and more generally the hypercontractivity ribbon) particularly for binary valued random variables.

A. A Duality property of $\mathsf{s}^*(W;Z)$

Theorem 2: Given a pair of binary-valued random variables $(W,Z) \sim p(w,z)$ (notation in Fig. **??**) with their joint distribution satisfying $0 < c, d < 1$, let $r_W^* \neq p_W$ be a maximizer of

$$\mathsf{s}_p^*(W;Z) = \sup_{r_W \neq p_W} \frac{D(r_Z \| p_Z)}{D(r_W \| p_W)}.$$

Let $r_{WZ}^* := r_W^* p_{Z|W}$. Then p_W is a maximizer of

$$\mathsf{s}_r^*(W;Z) = \sup_{q_W \neq r_W^*} \frac{D(q_Z \| r_Z^*)}{D(q_W \| r_W^*)} \tag{10}$$

and $\mathsf{s}_p^*(W;Z) = \mathsf{s}_r^*(W;Z)$. Further the line-segment connecting the curve at r_W^* and p_W is on the lower convex envelope of the curve: $p(W=1) \mapsto H(Z) - \lambda H(W)$.

Proof: We claim that the lower convex envelope of $p(W=1) \mapsto H(Z) - \lambda H(W)$ consists of an initial convex part, then (possibly) a line segment and then a final convex part. The line segment part exists if the whole curve is not convex. This is depicted in Fig. **??**. To see this we use Lemma **??** to prove that the curve $\Pr(W=1) \mapsto H(Z) - \lambda H(W)$ has at most two inflexion points, and the second derivative is positive when $\Pr(W=1) = s \in \{0,1\}$. Further we also note that the first derivative is $-\infty$ at $s=0$ and $+\infty$ at $s=1$.

Therefore given a λ where the $\Pr(W=1) \mapsto H(Z) - \lambda H(W)$ is not completely convex, we obtain that $\lambda = \mathsf{s}^*(W;Z)$ for two values of $\Pr(W=1)$ corresponding to the points where the tangent (of the lower concave envelope) meets the curve. Here we have used Theorem **??** and an observation that the left and right end points of the line segment continuously move towards each other. The last observation is not hard to justify given the continuity of the curve in both s and λ.

When $\lambda = \mathsf{s}^*(W;Z)$ we know that one of the points where the tangent meet the curve is given by $\Pr(W=1) = s$. Let the other point be $\Pr(W=1) = r$. Then λ is characterized by these two sets of equations

$$\begin{aligned}
&(\bar{c}-d)\log_2\left(\frac{cw+\bar{d}\bar{w}}{\bar{c}w+d\bar{w}}\right) - \lambda \log_2 \frac{\bar{w}}{w} \\
&\quad = (\bar{c}-d)\log_2\left(\frac{cu+\bar{d}\bar{u}}{\bar{c}u+d\bar{u}}\right) - \lambda \log_2 \frac{\bar{u}}{u} \\
&\frac{1}{w-u}\big(H(\bar{c}w+d\bar{w}) - \lambda H(w) - (H(\bar{c}u+d\bar{u}) - \lambda H(u))\big) \\
&\quad = (\bar{c}-d)\log_2\left(\frac{cu+\bar{d}\bar{u}}{\bar{c}u+d\bar{u}}\right) - \lambda \log_2 \frac{\bar{u}}{u}.
\end{aligned}$$

This is equivalent to

$$\begin{aligned}
&\bar{c}\log_2\left(\frac{\bar{c}\bar{s}+ds}{\bar{c}\bar{r}+dr}\right) + (1-\bar{c})\log_2\left(\frac{c\bar{s}+\bar{d}s}{c\bar{r}+\bar{d}r}\right) \\
&\quad = \lambda \log \frac{\bar{s}}{\bar{r}},
\end{aligned} \tag{11}$$

$$\begin{aligned}
&d\log_2\left(\frac{\bar{c}\bar{s}+ds}{\bar{c}\bar{r}+dr}\right) + (1-d)\log_2\left(\frac{c\bar{s}+\bar{d}s}{c\bar{r}+\bar{d}r}\right). \\
&\quad = \lambda \log \frac{s}{r}
\end{aligned} \tag{12}$$

Multiplying the first equality above by $\bar{s}$, second by s, and taking their sum yields

$$D(\bar{c}\bar{s}+ds \| \bar{c}\bar{r}+dr) = \lambda D(\bar{s}\|\bar{r}). \tag{13}$$

Similarly multiplying (**??**) by $\bar{r}$, (**??**) by r, and taking their sum yields

$$D(\bar{c}\bar{r}+dr \| \bar{c}\bar{s}+ds) = \lambda D(\bar{r}\|\bar{s}). \tag{14}$$

Fig. 2. The typical behaviour of the curve $p(W=1) \mapsto H(Z) - \lambda H(W)$ and its lower convex envelope.

Since λ corresponds to both $s_p^*(W;Z)$ and $\mathsf{s}_r^*(W;Z)$ where $r_{WZ} := r_W p_{Z|W}$, it is clear that r_W is r_W^* as defined in the Theorem. The duality is now obvious using equations (**??**) and (**??**).

Lemma 1: The second derivative of the function $p(W=1) \mapsto H(Z) - \lambda H(W)$ has at most two zeros in the interval $[0,1]$. The second derivative has at most one zero if $c=0$ or $d=0$. Further, the first derivative of this function is negative at $p(W=1)=0$ and it is positive at $p(W=1)=1$.

Proof: Using the notation of Fig. **??**, we can write $H(Z) - \lambda H(W)$ as a function of $s = p(W=1)$. Let us call this function $f(s)$. Then the first derivative is

$$\lambda \log \frac{s}{1-s} - (1-d-c) \log \frac{s(1-d)+(1-s)c}{sd+(1-s)(1-c)}$$

If c and d are in $(0,1)$ the first derivative is $-\infty$ at $p(W=1)=0$ and it is $+\infty$ at $p(W=1)=1$. When c or d is in $\{0,1\}$ we can use continuity to conclude that the first derivative is negative at $p(W=1)=0$ and it is positive at $p(W=1)=1$.

The second derivative of f is equal to

$$\frac{\lambda}{s(1-s)} - \frac{(1-c-d)^2}{(s(1-d)+(1-s)c)(sd+(1-s)(1-c))}$$

This can be written as $\frac{A(s)}{B(s)}$ where $A(s)$ is a second degree polynomial. Hence it can have at most two zeros. If $c=0$, the second derivative will become of the form $\frac{A(s)}{sB(s)}$ where $A(s)$ is a first degree polynomial. Therefore it can have at most one zero. A similar statement holds when $d=0$.

B. Convexity of $\mathsf{s}^*(W;Z)$ in $p(z|w)$

Let us fix the input $p(w)$ and vary the channel $p(z|w)$. We claim that $\mathsf{s}^*(W;Z)$ is convex in $p(z|w)$ for a fixed $p(w)$. In this sense $\mathsf{s}^*(Z;W)$ resembles the mutual information $I(Z;W)$.

Remark: Since $1 - h((1-\sqrt{x})/2)$ is an increasing convex function, we get that $1 - h((1-\sqrt{\mathsf{s}^*(Z;W)})/2)$ is a convex function in the channel $p(z|w)$. Thus, we have two convex functions, namely $1 - h((1-\sqrt{\mathsf{s}^*(Z;W)})/2)$, and $I(Z;W)$. Conjecture 3 claims that one of these convex function is always above the other.

Proof: We use $\mathsf{s}^*(p(w), p(z|w))$ instead of $\mathsf{s}^*(Z;W)$ to emphasize the underlying pmfs.

Take $p_0(z|w)$, $p_1(z|w)$ and $p_2(z|w)$ such that

$$p_1(z|w) = \beta p_0(z|w) + (1-\beta) p_2(z|w).$$

For $i = 0, 1, 2$ define

$$p_i(z) = \sum_w p(w) p_i(z|w),$$

Observe that

$$p_1(z) = \beta p_0(z) + (1-\beta) p_2(z).$$

Let $r(w) \not\equiv p(w)$ be any other probability distribution and for $i = 0,1,2$ define $r_i(z) = \sum_w r(w) p_i(z|w)$. Observe that

$$r_1(z) = \beta r_0(z) + (1-\beta) r_2(z).$$

Now we have

$$\begin{aligned}
&\frac{D(r_1(z)\|p_1(z))}{D(r(w)\|p(w))} \\
&= \frac{D(\beta r_0(z) + (1-\beta) r_2(z) \| \beta p_0(z) + (1-\beta) p_2(z))}{D(r(w)\|p(w))} \\
&\le \frac{\beta D(r_0(z)\|p_0(z)) + (1-\beta) D(r_2(z)\|p_2(z))}{D(r(w)\|p(w))} \\
&= \beta \cdot \frac{D(r_0(z)\|p_0(z))}{D(r(w)\|p(w))} + (1-\beta) \cdot \frac{D(r_2(z)\|p_2(z))}{D(r(w)\|p(w))} \\
&\le \beta \mathsf{s}^*(p(w), p_0(z|w)) + (1-\beta) \mathsf{s}^*(p(w), p_2(z|w)).
\end{aligned}$$

Taking supremum over $r(w) \not\equiv p(w)$ completes the proof.

IV. Analytical proof of Conjecture ?? in special cases

Let us specify the joint distribution of (W,Z) in the following way (see Fig. **??**):

- W, Z take values in $\{0,1\}$
- $s := \mathsf{Pr}(W=1)$
- $c := \mathsf{Pr}(Z=1|W=0)$
- $d := \mathsf{Pr}(Z=1|W=1)$
- $t := \mathsf{Pr}(Z=1) = (1-s)c + s(1-d)$

Since we will deal only with binary-valued random variables in the rest of the paper, we abuse notation to write $\mathsf{s}^*(W;Z) = \mathsf{s}^*(s,c,d), \rho_m(W;Z) = \rho_m(s,c,d), I(W;Z) = I(s,c,d)$.

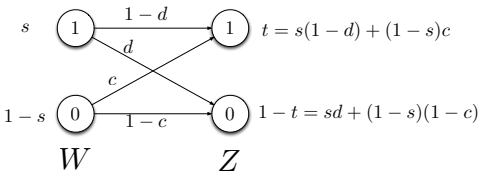

Fig. 3. Joint distribution of binary valued W, Z

Under this notation Conjecture **??** that for all $0 \leq s, c, d \leq 1$ the following inequality holds:

$$h\left(\frac{1-\sqrt{\mathsf{s}^*(s,c,d)}}{2}\right) + I(s,c,d) \leq 1. \quad (15)$$

Given $r \in [0,1]$, define $\bar{r} := 1 - r$ and $D(u\|v) := u\log_2\frac{u}{v} + \bar{u}\log_2\frac{\bar{u}}{\bar{v}}$. It suffices to restrict to the case where W, Z are not independent. This implies $0 < s < 1, c + d \neq 1$. We will assume these conditions hold in the rest of the paper.

Values of s^* for some special distributions are as follows:

- If $p_{Z|W}(z|w)$ is a binary symmetric channel, i.e. if $c = d$, and $s \neq \frac{1}{2}$, then

$$\mathsf{s}^*(s,c,c) = (1-2c)^2\frac{h'(s\bar{c}+\bar{c}s)}{h'(s)} \quad (16)$$

  where $h'(w) := \frac{d}{dw}h(w) = \log_2\frac{1-w}{w}$.
  Proof: The curve $s = p(W = 1) \mapsto H(Z) - \lambda H(W)$ is symmetric around $s = \frac{1}{2}$ i.e. it has the same value at s and $1-s$. The lower tangent to any such curve is always horizontal. Therefore, using Theorem **??**, the maximizer of $\mathsf{s}^*(s,c,d)$ occurs at $r = 1-s$. Substituting this value of r into Theorem **??** gives the desired result.
- If $p_{Z|W}(z|w)$ is a Z-channel, that is, if $c = 0$, then

$$\mathsf{s}^*(s,0,d) = \frac{\log_2(1-s\bar{d})}{\log_2(1-s)}. \quad (17)$$

  Proof: Using Lemma **??** for the case of $c = 0$, we can conclude that the curve $s = p(W = 1) \mapsto H(Z) - \lambda H(W)$ consists of an initial convex part and then (possibly) a line segment that connects to the end point of $(0,0)$. Using Theorem **??**, a simple calculation yields

$$\mathsf{s}^*(s,c,d) = \sup_{0\leq r\leq 1, r\neq s}\frac{D(\bar{r}c + r\bar{d}\|\bar{s}c + s\bar{d})}{D(r\|s)}.$$

We now prove Conjecture **??** for some special cases.

Theorem 3: Conjecture **??** (equiavlently (**??**)) holds when $c = d$.

Proof: For the case of $c = d$, we do have an exact formula for $\mathsf{s}^*(s,c,c)$, but we will only use the lower bound $\mathsf{s}^*(s,c,c) \geq \rho_m^2(s,c,c) = (1-2c)^2\frac{s(1-s)}{t(1-t)}$, where $t = s\bar{c} + \bar{s}c$. That is, it suffices to show that

$$h\left(\frac{1-|1-2c|\sqrt{\frac{s(1-s)}{t(1-t)}}}{2}\right) + h(t) - h(c) \leq 1. \quad (18)$$

By the standard transformation $\gamma := 1 - 2c, \sigma := 1 - 2s, \tau := 1 - 2t$, and observing that $\tau = \gamma\sigma$, this reduces to showing

$$h\left(\frac{1-|\gamma|\sqrt{\frac{1-\sigma^2}{1-\gamma^2\sigma^2}}}{2}\right) + h\left(\frac{1-\gamma\sigma}{2}\right) - h\left(\frac{1-\gamma}{2}\right) \leq 1, \quad (19)$$

for $-1 < \sigma < 1, -1 \leq \gamma \leq 1$.

Defining $\Lambda(u) := (1+u)\log_e(1+u) + (1-u)\log_e(1-u)$, we need to show

$$\Lambda(\gamma) \leq \Lambda(\gamma\sigma) + \Lambda\left(|\gamma|\sqrt{\frac{1-\sigma^2}{1-\gamma^2\sigma^2}}\right).$$

Since

$$(1-\gamma^2) = (1-(\gamma\sigma)^2)\left(1 - \left(|\gamma|\sqrt{\frac{1-\sigma^2}{1-\gamma^2\sigma^2}}\right)^2\right),$$

we only need to show that if $\Phi(v) := \Lambda(\sqrt{1-\exp(-v)})$, then for any $v_1, v_2 \geq 0$, $\Phi(v_1 + v_2) \leq \Phi(v_1) + \Phi(v_2)$. This follows by verifying that Φ is non-decreasing and concave.

Indeed the above result can also be obtained using the result stated below which generalizes the triples (s,c,d) for which the conjecture holds.

Theorem 4: Conjecture **??** holds for any triple (s,c,d) satisfying

$$\sqrt{1-\mathsf{s}^*(s,c,d)} + 2\sqrt{t\bar{t}} \leq 1 + 2\bar{s}\sqrt{c\bar{c}} + 2s\sqrt{d\bar{d}}. \quad (20)$$

Condition in (**??**) holds as long as (s,c,d) satisfies

$$\frac{\sqrt{\bar{s}c\bar{c} + sd\bar{d}}}{\sqrt{t\bar{t}}} + 2\sqrt{t\bar{t}} \leq 1 + 2\bar{s}\sqrt{c\bar{c}} + 2s\sqrt{d\bar{d}}. \quad (21)$$

Remark: Equation (**??**) holds when $c = d$ as it reduces to showing

$$\frac{\sqrt{c\bar{c}}}{\sqrt{t\bar{t}}} + 2\sqrt{t\bar{t}} \leq 1 + 2\sqrt{c\bar{c}},$$

which is true since $\sqrt{c\bar{c}} \leq \sqrt{t\bar{t}} \leq \frac{1}{2}$. Recall that when $c = d$ we have $t = s(1-c) + (1-s)c$.

Theorem **??** can be viewed as a special instance of the following strategy to solve Conjecture **??** which we state below. Theorem uses a majorization argument whose proof employs the following Lemma.

Lemma 2 (Lemma 1 in [?]): Let $x_0, ..., x_N$ and $y_0, ..., y_N$ be non-decreasing sequence of real numbers.

Let $\xi_0, ..., \xi_N$ be a sequence of real numbers such that for each k in the range $0 \leq k \leq N$,

$$\sum_{j=k}^{N} \xi_j x_j \geq \sum_{j=k}^{N} \xi_j y_j$$

with equality when $k = 0$. Then for any convex function Λ,

$$\sum_{j=0}^{N} \xi_j \Lambda(x_j) \geq \sum_{j=0}^{N} \xi_j \Lambda(y_j).$$

Remark: In [**?**] the above Lemma is stated for concave functions and the final inequality is reversed but the equivalence of the two statements is immediate.

Theorem 5: Suppose there is a bijection $g : [0,1] \mapsto [0, \frac{1}{2}]$ with $g^{-1} : [0, \frac{1}{2}] \to [0,1]$ denoting the inverse of g. Extend the inverse function to $g_e^{-1} : [0,1] \mapsto [0,1]$ according to $g_e^{-1}(x) := g^{-1}(\min\{x, 1-x\})$. If the following conditions hold:

1) $g(x)$ is increasing in x,
2) $h(g(x))$ is convex in x,
3) $1 + \bar{s} g_e^{-1}(c) + s g_e^{-1}(\bar{d}) \geq g_e^{-1}\left(\frac{1-\sqrt{\mathsf{s}^*(s,c,d)}}{2}\right) + g_e^{-1}(t)$,

then, Conjecture **??** is true for the chosen s, c, d.

Proof: The proof is a application of Lemma **??** to $\Lambda(x) = h(g(x))$. The details are presented below.

Let $x_1 = g_e^{-1}(c), x_2 = g_e^{-1}(\bar{d}), x_3 = 1$ and let $y_1 = g_e^{-1}(t), y_2 = 1 + \bar{s} g_e^{-1}(c) + s g_e^{-1}(\bar{d}) - g_e^{-1}(t)$. Further let $\tilde{x}_1, \tilde{x}_2$ be a rearrangement of x_1, x_2 in increasing order; and let $\tilde{y}_1, \tilde{y}_3$ be a rearrangement of y_1, y_2 in increasing order. Set $\tilde{y}_2 = \tilde{y}_1$. Allocate a weight $\bar{s}$ to x_2 and a weight s to x_1. Let ξ_1, ξ_2 denote the rearrangement of the weights s and $\bar{s}$ so that $\xi_1 \tilde{x}_1 + \xi_2 \tilde{x}_2 = \bar{s} x_1 + s x_2$.

Observe that the following holds:

$$\begin{aligned} \xi_1 \tilde{x}_1 + \xi_2 \tilde{x}_2 + x_3 &= \xi_1 \tilde{y}_1 + \xi_2 \tilde{y}_2 + \tilde{y}_3 && \text{By construction} \\ x_3 &\geq \tilde{y}_3 && \text{Since } x_3 = 1 \\ \xi_2 \tilde{x}_2 + x_3 &\geq \xi_2 \tilde{y}_2 + \tilde{y}_3. \end{aligned}$$

The last step follows since $\tilde{y}_1 = \tilde{y}_2 \geq \xi_1 \tilde{x}_1 + \xi_2 \tilde{x}_2 \geq \tilde{x}_1$. Further $\xi_1 \geq 0$ yields $\xi_1 \tilde{x}_1 \leq \xi_1 \tilde{y}_1$ and hence the desired inequality.

Observing that $h(g(g_e^{-1}(y))) = h(y)$ and that $h(g(x))$ is increasing in x, yields a proof of Conjecture **??** when the conditions on $g(x)$ stated in Theorem **??** hold.

We now prove Theorem **??**.

Proof (Theorem ??) :

Consider the function $g(\cdot) : [0,1] \mapsto [0, \frac{1}{2}]$ defined by

$$g(x) := \frac{1 - \sqrt{1-x^2}}{2}.$$

This function satisfies the conditions of Theorem **??**. A simple calculation shows that for this choice of $g(x)$ we obtain $g_e^{-1}(y) = 2\sqrt{y(1-y)}$. Further it is immediate that $g(x)$ is increasing in x for $x \in [0,1]$.

To verify convexity of $h(g(x))$ observe that

$$\begin{aligned} &\frac{1}{\log_2 e} \frac{d^2}{dx^2} h(g(x)) \\ &\quad = \log_e \left(\frac{1-g(x)}{g(x)}\right) g''(x) - \frac{g'(x)^2}{g(x)(1-g(x))} \\ &\quad = \log_e \left(\frac{1+\sqrt{1-x^2}}{1-\sqrt{1-x^2}}\right) \frac{1}{2\sqrt{1-x^2}} - \frac{1}{1-x^2}. \end{aligned}$$

Hence to show $h(g(x))$ is convex in x, it suffices to show that $\log_e \frac{1+a}{1-a} \geq 2a, a \in [0,1)$ which clearly holds by the Taylor series expansion of the left hand side which yields $\sum_{k \geq 1} \frac{2a^{2k-1}}{2k-1}$.

For this choice of $g(x)$ and the corresponding $g_e^{-1}(x)$ mentioned above, condition 3) in Theorem **??** is equivalent to the condition

$$\sqrt{1 - \mathsf{s}^*(s,c,d)} + 2\sqrt{t\bar{t}} \leq 1 + 2\bar{s}\sqrt{c\bar{c}} + 2s\sqrt{d\bar{d}}.$$

Thus from Theorem **??** we have,

$$h\left(\frac{1 - \sqrt{\mathsf{s}^*(s,c,d)}}{2}\right) + I(s,c,d) \leq 1.$$

This proves the first part or validity of Conjecture **??** when (**??**) holds.

Lower bounding $\mathsf{s}^*(s,c,d)$ by $\rho_m^2(s,c,d)$ yields (**??**). To this end, it is a simple exercie to note that

$$1 - \rho_m^2 = \frac{\bar{s}c\bar{c} + sd\bar{d}}{t\bar{t}}.$$

HISTORICAL REMARKS

Conjecture **??** was originally formulated by Kamath and Anantharam in an attempt to establish Conjecture **??**. It was then communicated to Gohari and Nair when all of them were collaborating to obtain the results in [**?**]. Bogdanov and Nair were independently working on Conjecture **??** and at that point had obtained a proof for the special setting $b = b'$ [**?**]. The results in Sections **??** and **??** are a result of the joint collaboration among the authors as a natural followup of their collaboration in [**?**]. There are a couple of other results along these lines that were obtained with Bogdanov that are not mentioned in this writeup but did help tune the intuition of the authors.

ACKNOWLEDGMENTS

Venkat Anantharam and Sudeep Kamath gratefully acknowledge the research support from the ARO MURI grant W911NF-08-1-0233, "Tools for the Analysis and Design of Complex Multi-scale Networks", from the NSF grant CNS-0910702, and from the NSF Science & Technology Center grant CCF-0939370, "Science of Information".

Chandra Nair wishes to thank Andrej Bogdanov for some insightful discussions and for some related results.

The work of C. Nair was partially supported by the following grants from the University Grants Committee of the Hong Kong Special Administrative Region, China: a) (Project No. AoE/E-02/08), b) GRF Project 415810. He also acknowledges the support from the Institute of Theoretical Computer Science and Communications (ITCSC) at The Chinese University of Hong Kong.

REFERENCES

[1] G. Kumar and T. Courtade, "Which Boolean Functions are Most Informative?", in *Proc. of IEEE ISIT*, Istanbul, Turkey, 2013.

[2] A. Bogdanov and C. Nair, *Personal Communication*, 2013.

[3] R. Ahlswede and P. Gács, "Spreading of sets in product spaces and hypercontraction of the Markov operator", *Annals of Probability*, vol. 4, pp. 925–939, 1976.

[4] Aline Bonami, "Étude des coefficients de Fourier des fonctions de $l^p(g)$", *Ann. Inst. Fourier (Grenoble)*, vol. 20, no. 2, pp. 335–402, 1971.

[5] William Beckner, "Inequalities in fourier analysis", *Ann. of Math.*, vol. 2, no. 1, pp. 159–182, 1975.

[6] V. Anantharam, A. Gohari, S. Kamath, and C. Nair, "On Maximal Correlation, Hypercontractivity, and the Data Processing Inequality studied by Erkip and Cover", *arXiv:1304.6133 [cs.IT]*, Apr. 2013.

[7] H.S. Witsenhausen, "On sequences of pairs of dependent random variables", *SIAM Journal on Applied Mathematics*, vol. 28, no. 1, pp. 100–113, January 1975.

[8] Bruce E. Hajek and Michael B. Pursley, "Evaluation of an achievable rate region for the broadcast channel", *IEEE Transactions on Information Theory*, vol. 25, no. 1, pp. 36–46, 1979.