Evaluating hypercontractivity parameters using information measures

Chandra Nair
Dept. of Information Engineering
The Chinese University of Hong Kong
Hong Kong

Yan Nan Wang
Dept. of Information Engineering
The Chinese University of Hong Kong
Hong Kong

***Abstract*—We use an equivalent characterization of hypercontractive parameters using relative entropy to compute the hypercontractive region for the binary erasure channel. A similar analysis also recovers the celebrated result for the binary symmetric channel, also called the Bonami-Beckner inequality.**

I. Introduction

Hypercontractive inequalities play an important role in many areas of mathematics and computer science. A pair of random variables (X,Y) is said to be (λ_1,λ_2)-hypercontractive, for $\lambda_1,\lambda_2 \in (1,\infty)$, if

$$\mathrm{E}(f(X)g(Y)) \leq \|f(X)\|_{\lambda_1}\|g(Y)\|_{\lambda_2}$$

holds for all non-negative functions $f(\cdot), g(\cdot)$. *Remark*: While the above definition is rather non-standard we use this equivalent parameterization (see [1] for a more standard notation that is easily mappable to the current one) since it makes our notation easier. In the above, we adopt the following notation for λ-th norm of random variables:

$$\|Z\|_\lambda := \mathrm{E}(|Z|^\lambda)^{\frac{1}{\lambda}}.$$

From Hölder's inequality and monotonicity of norm, it is immediate that if

$$\frac{1}{\lambda_1} + \frac{1}{\lambda_2} \leq 1$$

then the above inequality holds. We only consider finite valued random variables in this paper, though the standard machine (where finite valued random variables are called *simple functions*) enables the extension of the characterizations to families of general random variables.

There were several equivalent characterizations of the above condition using information measures derived in [2]. One of the characterization using divergence, stated below, can also be inferred from an earlier work [3]. The wording of the theorem below is adapted and modified from the sources to conform to the notation used in this paper.

Theorem 1 ([3], [2]). *Consider a pair of random variables (X,Y) distributed according to μ_{XY}. The following two assertions are equivalent:*

(i) *For all non-negative functions $f(\cdot), g(\cdot)$,*

$$\mathrm{E}(f(X)g(Y)) \leq \|f(X)\|_{\lambda_1}\|g(Y)\|_{\lambda_2}$$

(ii) *For every $\nu_{XY} (\ll \mu_{XY})$ we have*[1]

$$\frac{1}{\lambda_1}D(\nu_X\|\mu_X) + \frac{1}{\lambda_2}D(\nu_Y\|\mu_Y) \leq D(\nu_{XY}\|\mu_{XY}).$$

A necessary condition for (X,Y) to be (λ_1,λ_2)-hypercontractive is presented in the following theorem.

Theorem 2 (Correlation lower bound, [4]). *A pair of random variables is (λ_1,λ_2)-hypercontractive, for $\lambda_1,\lambda_2 \in (1,\infty)$, only if*

$$(\lambda_1-1)(\lambda_2-1) \geq \rho^2,$$

where ρ^2 is the maximal correlation between the pair (X,Y).

Determining the exact (or even sufficient) conditions for hypercontractivity has been a subject of active research in math. When (X,Y) is DSBS (doubly symmetric binary source) or when they are jointly Gaussian a complete characterization has been known since the 70s and these were celebrated results. In this paper we provide a complete characterization of the hypercontractive inequalities when a uniform X is passed through a binary erasure channel BEC(ϵ) to produce Y, when $0<\epsilon\leq\frac{1}{2}$, and a partial characterization of the hypercontractive inequalities when $\epsilon \in (\frac{1}{2},1)$. Furthermore our techniques can be used to derive an alternate proof of the result for the DSBS case. The proof startegy here has been motivated by that of Friedgut [5] where he establishes the DSBS case for $\lambda_1=\lambda_2$. Thus the results here also generalize his result.

II. Binary Erasure Channel

Consider a uniform binary random variable X passed through a binary erasure channel BEC(ϵ) producing the ternary output Y. Let μ_{XY}^{BEC} denote the joint law. The correlation inner bound for this setting says that (X,Y) is (λ_1,λ_2) hypercontractive for $\lambda_1,\lambda_2\in(1,\infty)$ only if

$$(\lambda_1-1)(\lambda_2-1)\geq 1-\epsilon.$$

The theorem below (main new result of this paper) determines the set of parameters for which correlation bound is tight, i.e. yields the hypercontractive region.

[1]The notation $\nu \ll \mu$ stands for absolute continuity of measure ν with respect to measure μ.

Theorem 3. *Let (X,Y) distributed according to μ_{XY}^{BEC} and $\lambda_1, \lambda_2 \in (1,\infty)$ satisfy $(\lambda_1-1)(\lambda_2-1)=1-\epsilon$. Then (X,Y) is (λ_1,λ_2)-hypercontractive, i.e. the correlation bound is tight, if and only if the following condition is satisfied:*

$$\epsilon - \frac{1}{2} \leq \frac{3}{2}(\lambda_2 - 1).$$

Remark 1. If $\epsilon \leq \frac{1}{2}$ then the correlation inner bound is tight; else it turns out to be tight only for a subset of the regime of parameters.

Proof. The proof is divided into two parts. In the first part, we will establish the result for $\lambda_2 \geq 2$ directly using the definition of hypercontractivity, by mimicking Janson's proof [6] for the DSBS case. For $\lambda_2 < 2$ we will use the equivalent characterization using divergences to provide a proof.

Case 1: $\lambda_2 \geq 2$. Let λ_1 be defined according to $(\lambda_1 - 1)(\lambda_2 - 1) = 1 - \epsilon$. We wish to show that for all functions $f(\cdot), g(\cdot)$ the inequality

$$\mathrm{E}(f(X)g(Y)) \leq \|f(X)\|_{\lambda_1} \|g(Y)\|_{\lambda_2}$$

holds. Observe that, by using Hölder's inequality,

$$\begin{aligned}\mathrm{E}(f(X)g(Y)) &= \mathrm{E}(\mathrm{E}(f(X)|Y)g(Y)) \\ &\leq \| \mathrm{E}(f(X)|Y)\|_{\lambda_2'} \|g(Y)\|_{\lambda_2}.\end{aligned}$$

Here $\lambda_2' \in (1,2]$ is the Hölder conjugate of λ_2. Hence showing (in fact this is an equivalent condition) the following suffices

$$\| \mathrm{E}(f(X)|Y)\|_{\lambda_2'} \leq \|f(X)\|_{\lambda_1}.$$

W.l.o.g. let $f(0) = 1 - \delta$, $f(1) = 1 + \delta$. Then the above inequality reduces to

$$\begin{aligned}&\left[\frac{1-\epsilon}{2}(1-\delta)^{\lambda_2'} + \frac{1-\epsilon}{2}(1+\delta)^{\lambda_2'} + \epsilon\right]^{\frac{1}{\lambda_2'}} \\ &\qquad \leq \left[\frac{1}{2}(1-\delta)^{\lambda_1} + \frac{1}{2}(1+\delta)^{\lambda_1}\right]^{\frac{1}{\lambda_1}}.\end{aligned}$$

That is, suffices that

$$1 + (1-\epsilon)\sum_{k=1}^{\infty}\binom{\lambda_2'}{2k}\delta^{2k} \leq \left(1 + \sum_{k=1}^{\infty}\binom{\lambda_1}{2k}\delta^{2k}\right)^{\frac{\lambda_2'}{\lambda_1}}$$

To get the above reduction we use the multiplicative formula extension of binomial co-efficients and the infinite power series

$$(1+x)^{\alpha} = 1 + \sum_{k=1}^{\infty}\binom{\alpha}{k}x^k, |x| < 1.$$

Substituting for λ_1 we see that $\frac{\lambda_2'}{\lambda_1} = \frac{\lambda_2}{\lambda_2 - \epsilon} > 1$. Since $(1+x)^a\ geq 1 + ax$ $(a > 1, x > 0)$, it suffices to show that

$$1 + (1-\epsilon)\sum_{k=1}^{\infty}\binom{\lambda_2'}{2k}\delta^{2k} \leq 1 + \frac{\lambda_2'}{\lambda_1}\sum_{k=1}^{\infty}\binom{\lambda_1}{2k}\delta^{2k}$$

Since $1 < \lambda_1 \leq \lambda_2' \leq 2$ the inequality is easily seen to be true by comparing the coefficients of δ^{2k} term by term (all terms are non-negative). Equality holds for $k=1$ and for all other powers it is an inequality, in general. (See Remark 2 at the end of next section.)

Case 2: $\lambda_2 < 2$. We use the equivalent characterization using divergences in this case. Again let $(\lambda_1 - 1)(\lambda_2 - 1) = 1 - \epsilon$. We wish to show that

$$\begin{aligned}&\max_{\nu_{XY} \ll \mu_{XY}^{BEC}} \frac{1}{\lambda_1}D(\nu_X \| \mu_X^{BEC}) + \frac{1}{\lambda_2}D(\nu_Y \| \mu_Y^{BEC}) \\ &\quad - D(\nu_{XY} \| \mu_{XY}^{BEC}) = \begin{cases} 0 & \epsilon - \frac{1}{2} \leq \frac{3}{2}(\lambda_2 - 1) \\ > 0 & \text{o.w.}\end{cases}\end{aligned}$$

It is easy to see that the maximum has to be an interior point by considering the behavior at the boundaries. This is primarily because the last term has an infinite slope at the boundaries and since $\lambda_1, \lambda_2 > 1$ this infinite slope cannot be completely canceled by the first two terms. We omit the details of this calculation here.

Thus the main part of the proof is to show that there is only one interior stationary point $\nu_{XY} = \mu_{XY}^{BEC}$ when $\epsilon - \frac{1}{2} \leq \frac{3}{2}(\lambda_2 - 1)$ and otherwise, $\nu_{XY} = \mu_{XY}^{BEC}$ is not even a local maximum.

For any (strictly) interior stationary points, the Lagrange conditions yield

$$k = \frac{1}{\lambda_1}\ln(\nu_{00} + \nu_{0E}) - \frac{1}{\lambda_2'}\ln\frac{\nu_{00}}{1-\epsilon} \tag{1a}$$

$$k = \frac{1}{\lambda_1}\ln(\nu_{11} + \nu_{1E}) - \frac{1}{\lambda_2'}\ln\frac{\nu_{11}}{1-\epsilon} \tag{1b}$$

$$\begin{aligned}k = &\frac{1}{\lambda_1}\ln(\nu_{00} + \nu_{0E}) + \frac{1}{\lambda_2}\ln(\nu_{0E} + \nu_{1E}) - \frac{1}{\lambda_2}\ln 2 \\ &- \ln \nu_{0E} + \frac{1}{\lambda_2'}\ln\epsilon\end{aligned} \tag{1c}$$

$$\begin{aligned}k = &\frac{1}{\lambda_1}\ln(\nu_{11} + \nu_{1E}) + \frac{1}{\lambda_2}\ln(\nu_{0E} + \nu_{1E}) - \frac{1}{\lambda_2}\ln 2 \\ &- \ln \nu_{1E} + \frac{1}{\lambda_2'}\ln\epsilon\end{aligned} \tag{1d}$$

Equating (1c) and (1d) yields

$$\frac{\nu_{0E}}{\nu_{1E}} = \left(\frac{\nu_{00} + \nu_{0E}}{\nu_{11} + \nu_{1E}}\right)^{\frac{1}{\lambda_1}}. \tag{2a}$$

Equating (1a) and (1c) yields

$$\nu_{00} = \frac{\nu_{0E}^{\lambda_2'}2^{\lambda_2'-1}}{(\nu_{0E} + \nu_{1E})^{\lambda_2'-1}}\frac{1-\epsilon}{\epsilon}. \tag{2b}$$

Equating (1b) and (1d) yields

$$\nu_{11} = \frac{\nu_{1E}^{\lambda_2'}2^{\lambda_2'-1}}{(\nu_{0E} + \nu_{1E})^{\lambda_2'-1}}\frac{1-\epsilon}{\epsilon}. \tag{2c}$$

Substituting for ν_{00} and ν_{11} using (2b) and (2c) in (2a), setting $x = \frac{2\nu_{0E}}{\nu_{0E} + \nu_{1E}} \in [0,2]$ and using $(\lambda_1 - 1)(\lambda_2 - 1) = 1 - \epsilon$ we obtain

$$\begin{aligned}&(1-\epsilon)x^{\frac{\epsilon}{\lambda_2 - 1}} + \epsilon x^{\frac{\epsilon - 1}{\lambda_2 - 1}} \\ &\qquad = (1-\epsilon)(2-x)^{\frac{\epsilon}{\lambda_2 - 1}} + \epsilon(2-x)^{\frac{\epsilon-1}{\lambda_2 - 1}}.\end{aligned}$$

From Lemma 1 we know that the above equation has exactly one solution, $x = 1$, when $(\epsilon - \frac{1}{2}) \leq \frac{3}{2}(\lambda_2 - 1)$. Thus under the above condition on (λ_2, ϵ) every interior stationary point must satisfy $\nu_{0E} = \nu_{1E}$. Further from (2b) and (2c) we can conclude that

$$\frac{\nu_{00}}{1-\epsilon} = \frac{\nu_{11}}{1-\epsilon} = \frac{\nu_{0E}}{\epsilon} = \frac{\nu_{1E}}{\epsilon},$$

implying that the only stationary point (hence global maximizer) is $\nu_{XY} = \mu_{XY}^{BEC}$, which yields a maximum value 0 as desired.

When $(\epsilon - \frac{1}{2})(\lambda_2 - 1) > \frac{3}{2}$, choose

$$\nu_{XY} = \begin{bmatrix} \frac{(1-\delta)^{\lambda_2'}(1-\epsilon)}{A} & \frac{\epsilon(1-\delta)}{A} & 0 \\ 0 & \frac{\epsilon(1+\delta)}{A} & \frac{(1-\epsilon)(1+\delta)^{\lambda_2'}}{A} \end{bmatrix}$$

where $A = 2\epsilon + (1-\epsilon)[(1+\delta)^{\lambda_2'} + (1-\delta)^{\lambda_2'}]$ is the normalizing constant. Taylor series expansion of the term

$$\frac{1}{\lambda_1} D(\nu_X \| \mu_X^{BEC}) + \frac{1}{\lambda_2} D(\nu_Y \| \mu_Y^{BEC}) - D(\nu_{XY} \| \mu_{XY}^{BEC})$$

around $\delta = 0$ yields an expansion

$$\frac{1}{24}\epsilon(1-\epsilon)(\lambda_2' - 1)^2((2\epsilon - 1)(\lambda_2' - 1) - 3)\delta^4 + O(\delta^6)$$

which is positive when

$$\epsilon - \frac{1}{2} > \frac{3}{2}(\lambda_2 - 1),$$

yielding that the maximum of the function is strictly positive under these parameter settings. □

Lemma 1. *For $x \in [0, 2]$, $\lambda_2 \in (1, 2)$, $\epsilon \in (0, 1)$ the following equation*

$$(1-\epsilon)x^{\frac{\epsilon}{\lambda_2 - 1}} + \epsilon x^{\frac{\epsilon-1}{\lambda_2-1}} = (1-\epsilon)(2-x)^{\frac{\epsilon}{\lambda_2-1}} + \epsilon(2-x)^{\frac{\epsilon-1}{\lambda_2-1}}.$$

has only one root at $x = 1$ if $(\epsilon - \frac{1}{2}) \leq \frac{3}{2}(\lambda_2 - 1)$.

Proof. Clearly $x = 1$ is a root of this equation. Denote $p - 1 = \frac{1}{\lambda_2 - 1}$. Note that $p \in (2, \infty)$. Define the function $g(x)$

$$g(x) = \frac{1-\epsilon}{\epsilon} x^{(p-1)\epsilon} + x^{(p-1)(\epsilon-1)} - \frac{1-\epsilon}{\epsilon}(2-x)^{(p-1)\epsilon} - (2-x)^{(p-1)(\epsilon-1)}$$

$g(1) = 0$, $\lim_{x \downarrow 0} g(x) = +\infty$. Further $g(x) = -g(2-x)$. The statement follows by showing $g(x)$ decreases over $(0, 1)$ if $(p-1)(\epsilon - \frac{1}{2}) \leq \frac{3}{2}$.
Take the derivative with respect to x,

$$g'(x) = (1-\epsilon)(p-1)[x^{p\epsilon - \epsilon - 1} - x^{p\epsilon - p - \epsilon} + (2-x)^{p\epsilon - \epsilon - 1} - (2-x)^{p\epsilon - p - \epsilon}]$$

Let $y = 1 - x$, where $y \in (0, 1)$ and let $r = p\epsilon - \epsilon - \frac{p+1}{2}$, then $g'(x) \leq 0$ is equivalent to

$$(1-y)^r[(1-y)^{-\frac{p-1}{2}} - (1-y)^{\frac{p-1}{2}}] \geq (1+y)^r[(1+y)^{\frac{p-1}{2}} - (1+y)^{-\frac{p-1}{2}}]).$$

Observe that $r \leq \frac{1}{2}$ is equivalent to $(\epsilon - \frac{1}{2}) \leq \frac{3}{2}(\lambda_2 - 1)$. So we are done if we show that the above inequality holds for any $r \leq \frac{1}{2}$ and $p > 2$. Further since $\left(\frac{1-y}{1+y}\right)^r$ decreases in r, it suffices to show the inequality for $r = \frac{1}{2}$ and $p > 2$. Substituting $r = \frac{1}{2}$ and rearranging, we wish to show

$$(1-y)^{-\frac{p}{2}+1} + (1+y)^{-\frac{p}{2}+1} \geq (1+y)^{\frac{p}{2}} + (1-y)^{\frac{p}{2}}.$$

Performing a Taylor series expansion, it suffices to show

$$2[1 + \sum_{k=1}^{\infty} \binom{1-\frac{p}{2}}{2k} y^{2k}] \geq 2[1 + \sum_{k=1}^{\infty} \binom{\frac{p}{2}}{2k} y^{2k}]$$

Note that the first term ($k = 1$) is equal for both sides and is positive (in the case that $p > 2$). For $k \geq 2$ it is immediate (by expanding the binomial term) that

$$\binom{1-\frac{p}{2}}{2k} \geq \max\left\{0, \binom{\frac{p}{2}}{2k}\right\}.$$

This completes the proof of the lemma. □

III. Binary Input Symmetric Output Channel

Consider a (X, Y) where X is binary and uniformly distributed, and Y is obtained via a channel $W(y|x)$ that satisfies a symmetry property, $W(Y = i|X = 1) = W(Y = -i|X = -1) = \mu_i$, for $-K \leq i \leq K$. This class contains both the Binary erasure channel and the binary symmetric channel. Let μ_{XY}^{BISO} denote the joint law of this pair of random variables. The correlation inner bound (a simple calculation) for this setting says that (X, Y) is (λ_1, λ_2)-hypercontractive *only if*

$$(\lambda_1 - 1)(\lambda_2 - 1) \geq \sum_{i=1}^{K} \frac{(\mu_i - \mu_{-i})^2}{\mu_i + \mu_{-i}}.$$

Proposition 1. *For any $\lambda_2 \geq 2$, the pair (X, Y) is (λ_1, λ_2)-hypercontractive for any pair of λ_1, λ_2 satisfying the correlation bound.*

Proof. The proof mimics the proof of Case 1 in the proof of Theorem 3. Following the approach we need to show that

$$\| \mathrm{E}(f(X)|Y) \|_{\lambda_2'} \leq \| f(X) \|_{\lambda_1}.$$

Further, by monotonicity of norm, it suffices to restrict to

$$(\lambda_1 - 1)(\lambda_2 - 1) = \sum_{i=1}^{K} \frac{(\mu_i - \mu_{-i})^2}{\mu_i + \mu_{-i}}.$$

W.l.o.g. let $f(-1) = 1 - \delta$, $f(1) = 1 + \delta$. Then the above inequality reduces to showing

$$\sum_{i=-K}^{K} \frac{\mu_i + \mu_{-i}}{2} \left(1 - \delta \frac{\mu_i - \mu_{-i}}{\mu_i + \mu_{-i}}\right)^{\lambda_2'} \leq \left[\frac{1}{2}(1-\delta)^{\lambda_1} + \frac{1}{2}(1+\delta)^{\lambda_1}\right]^{\frac{\lambda_2'}{\lambda_1}}.$$

Observing that $\frac{\lambda_2'}{\lambda_1} \geq 1$, taking the binomial expansion of both sides (as earlier) and using $(1+x)^a \geq 1 + ax, a > 1, x \geq 0$, it suffices to show

$$1 + \sum_{k=1}^{\infty} \binom{\lambda_2'}{2k} \delta^{2k} \left(\sum_{i=-K}^{K} \frac{\mu_i + \mu_{-i}}{2} \left(\frac{\mu_i - \mu_{-i}}{\mu_i + \mu_{-i}} \right)^{2k} \right) \leq 1 + \frac{\lambda_2'}{\lambda_1} \sum_{k=1}^{\infty} \binom{\lambda_1}{2k} \delta^{2k}.$$

Comparing term by term, we see that equality holds when $k = 1$ and the inequality holds for other terms since $k \geq 2$ implies

$$\sum_{i=-K}^{K} \frac{\mu_i + \mu_{-i}}{2} \left(\frac{\mu_i - \mu_{-i}}{\mu_i + \mu_{-i}} \right)^{2k} \leq \sum_{i=-K}^{K} \frac{\mu_i + \mu_{-i}}{2} \left(\frac{\mu_i - \mu_{-i}}{\mu_i + \mu_{-i}} \right)^{2}.$$

This completes the proof of the proposition. ☐

Remark 2. A key observation in the above argument is that when $1 < \lambda_1 \leq \lambda_2' \leq 2$, the terms $\binom{\lambda_1}{2k}$ and $\binom{\lambda_2'}{2k}$ are non-negative for any $k \geq 1$; $\rho^2 \binom{\lambda_2'}{2} = \frac{\lambda_2'}{\lambda_1} \binom{\lambda_1}{2}$ (where ρ^2 is the correlation); and for $j \geq 2$ the term $j - \lambda_2' \geq j - \lambda_1$ allowing one to conclude the term by term relation. This is essentially a borrow of the argument in [6] for the binary symmetric channel (BSC) scenario. We provide an alternate proof of BSC in the coming section. .

IV. Binary Symmetric Channel

Consider a uniformly distributed binary valued X and Y obtained by passing X through a BSC with crossover probability $\frac{1-\rho}{2}$. The hypercontractivity for this pair of $(X,Y) \sim \mu_{XY}^{BSC}$ has been established since the 70s and there are various proofs in the literature. The simplest one, according to the authors, is the one due to Janson [6]. This section yields yet another proof of the celebrated Bonami-Beckner inequality starting from the divergence characterization. Friedgut [5] established a proof along the very same lines for a particular choice $\lambda_1 = \lambda_2 = 1 + |\rho|$, and this proof generalizes the proof to all parameters.

The key result in BSC is that the correlation lower bound is tight.

Theorem 4 (Bonami-Beckner; alternate proof provided here)**.** *For (X,Y) distributed according μ_{XY}^{BSC}, the pair is (λ_1, λ_2) hypercontractive if*

$$(\lambda_1 - 1)(\lambda_2 - 1) \geq \rho^2.$$

Proof. When $\rho = 0$ the result is trivial and follows from the monotonicity of norm. Hence, we assume that $\rho \neq 0$. The proof mimics that of Case 2 of the BEC proof. We consider, w.l.o.g. the pair (λ_1, λ_2) satisfying $(\lambda_1 - 1)(\lambda_2 - 1) = \rho^2$. We are required to show that

$$\max_{\nu_{XY} \ll \mu_{XY}^{BSC}} \frac{1}{\lambda_1} D(\nu_X \| \mu_X^{BSC}) + \frac{1}{\lambda_2} D(\nu_Y \| \mu_Y^{BSC}) - D(\nu_{XY} \| \mu_{XY}^{BSC}) = 0.$$

It is rather elementary to see that the boundary points cannot be the maximizers; so we will only consider the interior points. The idea is to show that there is only one interior stationary point at $\nu_{XY} = \mu_{XY}^{BSC}$.

For any (strictly) interior stationary points, the Lagrange conditions yield

$$k = \frac{1}{\lambda_1} \ln(\nu_{00} + \nu_{01}) + \frac{1}{\lambda_2} \ln(\nu_{00} + \nu_{10}) - \ln \frac{\nu_{00}}{1+\rho} \quad (3a)$$

$$k = \frac{1}{\lambda_1} \ln(\nu_{00} + \nu_{01}) + \frac{1}{\lambda_2} \ln(\nu_{01} + \nu_{11}) - \ln \frac{\nu_{01}}{1-\rho} \quad (3b)$$

$$k = \frac{1}{\lambda_1} \ln(\nu_{10} + \nu_{11}) + \frac{1}{\lambda_2} \ln(\nu_{00} + \nu_{10}) - \ln \frac{\nu_{10}}{1-\rho} \quad (3c)$$

$$k = \frac{1}{\lambda_1} \ln(\nu_{10} + \nu_{11}) + \frac{1}{\lambda_2} \ln(\nu_{01} + \nu_{11}) - \ln \frac{\nu_{11}}{1+\rho} \quad (3d)$$

By considering equations (3a) and (3c); and (3b) and (3d) we obtain

$$\left(\frac{\nu_{00} + \nu_{01}}{\nu_{10} + \nu_{11}} \right)^{\frac{1}{\lambda_1}} = \frac{\nu_{00}}{\nu_{10}} \frac{1-\rho}{1+\rho} = \frac{\nu_{01}}{\nu_{11}} \frac{1+\rho}{1-\rho} = x \quad (4)$$

Similarly considering equations (3a) and (3b); and (3c) and (3d) we obtain

$$\left(\frac{\nu_{00} + \nu_{10}}{\nu_{01} + \nu_{11}} \right)^{\frac{1}{\lambda_2}} = \frac{\nu_{00}}{\nu_{01}} \frac{1-\rho}{1+\rho} = \frac{\nu_{10}}{\nu_{11}} \frac{1+\rho}{1-\rho} \quad (5)$$

Since $\nu_{00} + \nu_{01} + \nu_{10} + \nu_{11} = 1$, denoting $\theta = \frac{1-\rho}{1+\rho} \in (0,1) \cup (1,\infty)$ (since $\rho \neq 0$), elementary manipulations show that x satisfies the following equation

$$x^{\lambda_1 - 1} = \frac{(1+\theta x)^{\frac{1}{\lambda_2 - 1}} \theta + (\theta + x)^{\frac{1}{\lambda_2 - 1}}}{(\theta + x)^{\frac{1}{\lambda_2 - 1}} \theta + (1+\theta x)^{\frac{1}{\lambda_2 - 1}}}.$$

Since $(\lambda_1 - 1)(\lambda_2 - 1) = \rho^2 = \left(\frac{1-\theta}{1+\theta}\right)^2$, denoting by $t = \frac{1}{\lambda_2 - 1}$, we obtain that x satisfies

$$x^{t\left(\frac{1-\theta}{1+\theta}\right)^2} = \frac{(1+\theta x)^t \theta + (\theta + x)^t}{(\theta + x)^t \theta + (1+\theta x)^t}.$$

From Lemma 2, we could know that the equation above has only one root $x = 1$. Therefore there is exactly one stationary point, $\nu_{XY} = \mu_{XY}^{BSC}$. This ensures that the maximum of the divergence expression is zero and completes the proof. ☐

Lemma 2. *For any $t \in (0,\infty)$ and $\theta \in (0,1) \cup (1,\infty)$ the equation*

$$x^{t\left(\frac{1-\theta}{1+\theta}\right)^2} = \frac{(1+\theta x)^t \theta + (\theta + x)^t}{(\theta + x)^t \theta + (1+\theta x)^t}$$

has only one root at $x = 1$ for $x \in (0,\infty)$.

Proof. Let $x = e^h$ and define

$$g(h) = \ln\left((1+\theta e^h)^t \theta + (\theta + e^h)^t\right).$$

Taking logarithms of the equation in Lemma 2 and making above substitutions, we wish to show that

$$ht\left(\frac{1-\theta}{1+\theta}\right)^2 = g(h) - g(-h) - ht$$

has exactly one zero at $h = 0$. Define

$$r(h) = g(h) - g(-h) - ht - ht\left(\frac{1-\theta}{1+\theta}\right)^2.$$

We will show that $r'(h) \leq 0$ implying the desired result.

Note that

$$r'(h) = g'(h) + g'(-h) - t - t\left(\frac{1-\theta}{1+\theta}\right)^2.$$

Observe that

$$\begin{aligned} g'(h) &= t\left(\frac{\theta^2 e^h(1+\theta e^h)^{t-1} + e^h(\theta+e^h)^{t-1}}{(1+\theta e^h)^t\theta + (\theta+e^h)^t}\right) \\ &= t\left(1 - \theta\left(\frac{(1+\theta e^h)^{t-1} + (\theta+e^h)^{t-1}}{(1+\theta e^h)^t\theta + (\theta+e^h)^t}\right)\right). \end{aligned}$$

Substituting this into $r'(h)$, and after performing elementary manipulations, the condition $r'(h) \leq 0$ becomes equivalent to verifying

$$\begin{aligned} \frac{4}{(1+\theta)^2} \leq &\left(\frac{(1+\theta e^h)^{t-1} + (\theta+e^h)^{t-1}}{(1+\theta e^h)^t\theta + (\theta+e^h)^t}\right) \\ &+ e^h\left(\frac{(1+\theta e^h)^{t-1} + (\theta+e^h)^{t-1}}{(1+\theta e^h)^t + \theta(\theta+e^h)^t}\right). \end{aligned}$$

The above condition can be re-expressed as

$$\begin{aligned} &\left((1+\theta e^h)^{t-1} + (\theta+e^h)^{t-1}\right)\left((1+\theta e^h)^{t+1} + (\theta+e^h)^{t+1}\right) \\ &\geq \frac{4}{(1+\theta)^2}\left((1+\theta e^h)^t\theta + (\theta+e^h)^t\right) \\ &\quad \times \left((1+\theta e^h)^t + \theta(\theta+e^h)^t\right). \end{aligned}$$

Elementary algebraic manipulation reduces the above to

$$\begin{aligned} &\left(\frac{1-\theta}{1+\theta}\right)^2\left((1+\theta e^h)^t - (\theta+e^h)^t\right)^2 \\ &\quad + (1+\theta e^h)^{t-1}(\theta+e^h)^{t-1}(1+\theta e^h - \theta - e^h)^2 \geq 0, \end{aligned}$$

which trivially holds. Furthermore, equality holds only at $h = 0$ implying that $r(h) = 0$ only at $h = 0$. □

V. Discussion and Conclusion

The hypercontractivity parameters for pairs of binary random variables through a symmetric channel is derived in many regimes. We also obtain a proof of the Bonami-Beckner inequality (the BSC case). An interesting observation is that when correlation inner bound was tight, it turned out that the non-convex optimization problem had only one stationary point. For the case of binary input symmetric output channels we showed that the correlation inner bound is tight for $\lambda_2 \geq 2$. However numerical simulations indicate that perhaps the correlation inner bound is tight until $\lambda_2 \geq \frac{4}{3}$; indicating yet another example of binary erasure channel being the opposite extremal (the other one is BSC) case among the space of binary input symmetric output channels. We were led to investigating the uniqueness of stationary point after hearing Friedgut present his proof for a particular parameter of the BSC case. The determination of hypercontractivity parameter for the binary erasure channel was a question posed to the first author by Jaikumar Radhakrishnan and Venkat Guruswami during the Simon's institute semester long program in information theory.

For the binary erasure channel, one can extend the proof technique borrowed from [6] to the parameter regime $\lambda_2 \geq \frac{3}{2}$. However for rest of the regimes, the only proof we could obtain was using the divergence characterization.

Subsequent to the initial submission of this work the authors have also managed to characterize, using the same techniques, the reverse hypercontractive regions for various binary symmetric channels - some completely and some in certain parameter regimes.

Acknowledgment

The first author would like to thank Venkat Anantharam and Sudeep Kamath for interesting discussions, and for various joint attempts to recover the hypercontractivity result for the BSC case using the information measure characterizations. He would also like to thank the Simon's institute for the theory of Computing at UC Berkeley. Chandra Nair would like to acknowledge the support of GRF grants 2150785 and 2150829, as well as the Area of Excellence grant on Network Coding for financial support towards this research work.

References

[1] V. Anantharam, A. A. Gohari, S. Kamath, and C. Nair, "On maximal correlation, hypercontractivity, and the data processing inequality studied by erkip and cover," *CoRR*, vol. abs/1304.6133, 2013.

[2] C. Nair, "Equivalent formulations of hypercontractivity using information measures," *International Zurich Seminar*, 2014.

[3] E. A. Carlen and C. E. Dario, "Subadditivity of the entropy and its relation to Brascamp-Lieb type inequalities," *Geometric and Functional Analysis*, vol. 19, no. 2, pp. 373–405, 2009. [Online]. Available: http://dx.doi.org/10.1007/s00039-009-0001-y

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

[5] E. Friedgut, "An information-theoretic proof of a hypercontractive inequality," *ArXiv e-prints*, Apr. 2015.

[6] S. Janson, *Gaussian Hilbert Spaces*, ser. Cambridge Tracts in Mathematics. Cambridge University Press, 1997. [Online]. Available: http://books.google.com.hk/books?id=heCzfIclio0C