# Reverse hypercontractivity region for the binary erasure channel

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

***Abstract*—In this paper, we obtain the reverse hypercontractive region for the pair of variables $(X, Y)$ where $X$ is a uniformly distributed binary random variable and $Y$ (a ternary random variable) is obtained by passing $X$ through a symmetric binary erasure channel (BEC), for a non-trivial range of parameters. The technique used builds on two recent results: $a)$ characterization of reverse hypercontractivity using information measures, and $b)$ computation of the forward hypercontractive region for the BEC.**

## I. Introduction

Reverse hypercontractive inequalities, like the (forward or regular) hypercontractive inequalities, are a family of inequalities that are studied in functional analysis; and have proven useful in mathematics [1], [2] and in computer science [3]. The exploration of the link between single-letterization (in information theory) and tensorization phenomenon has recently led information-theorists [4]–[9] to revisit connections between inequalities involving information measures and inequalities whose parameters tensorize such as hypercontractive or Brascamp-Lieb type inequalities.

Exact computation of the parameters (hypercontractive, reverse-hypercontractive, or Brascamp-Lieb) has been a challenging task with very few exact characterizations. Two well known cases where exact computations have been feasible are for jointly Gaussian random variables, and when $X$ is a uniform random variable and $Y$ is obtained from $X$ by passing it through a binary symmetric channel. These results are celebrated results with several different proofs provided over the years and have found applications in many areas. Last year, the authors computed the forward hypercontractive region for the binary erasure channel starting from the characterization using information measures; a computation prompted by the authors of [10].

In all the results mentioned above, it turns out that the hypercontractive or reverse-hypercontractive region matches the correlation bound (though in general it is known that these two are not the same regions). The computation in this paper shows a non-trivial exact characterization where the region is not given by the correlation bound.

In the discussion section towards the end of the paper, we will outline why the arguments, such as the one used here, may have a broader significance in a whole variety of problems, including those in multi-user information theory.

### A. Preliminaries

A pair of random variables $(X, Y)$ is said to be $(\lambda_1, \lambda_2)$-reverse-hypercontractive, for $\lambda_1, \lambda_2 \in (-\infty, 1)$, if

$$\mathrm{E}(f(X)g(Y)) \geq \|f(X)\|_{\lambda_1} \|g(Y)\|_{\lambda_2} \tag{1}$$

holds for all positive functions $f(\cdot), g(\cdot)$. In the above

$$\|Z\|_\lambda = \mathrm{E}\left(|Z|^\lambda\right)^{\frac{1}{\lambda}},\ \lambda \neq 0,$$

and $\|Z\|_0 = e^{\mathrm{E}(\log |Z|)}$. Reverse Hölder's inequality says that the above inequality holds when

$$\frac{1}{\lambda_1} + \frac{1}{\lambda_2} = 1,$$

and the monotonicity of the $\|Z\|_\lambda$ in $\lambda$ yields a trivial region of parameters where (1) always holds. This, for instance, includes the region $\lambda_1, \lambda_2 \in (-\infty, 0]$. Therefore the non-trivial region of the reverse-hypercontractive region is when at least one of the parameters $\lambda_1$ or $\lambda_2$ is strictly positive.

Our starting point is the following equivalent characterization of the reverse hypercontractive region extracted from a more general result last year by one of the authors.

**Theorem 1** ( [8]). *Depending on the regime of parameters of $\lambda_1, \lambda_2$, the following yields an equivalent characterization of* (1) *in terms of relative entropies.*

$(i)$ *When $\lambda_1, \lambda_2 \in (0, 1)$ equation* (1) *holds iff:*
*For any $q_X$ and $q_Y$ there exists $r_{XY}$ with $r_X = q_X$ and $r_Y = q_Y$ such that:*

$$\frac{1}{\lambda_1} D(q_X||p_X) + \frac{1}{\lambda_2} D(q_Y||p_Y) \geq D(r_{XY}||p_{XY})$$

$(ii)$ *When $0 < \lambda_1 < 1$ and $\lambda_2 < 0$ equation* (1) *holds iff:*
*For any $q_X$ there exists $r_{XY}$ with $r_X = q_X$ such that:*

$$\frac{1}{\lambda_1} D(q_X||p_X) + \frac{1}{\lambda_2} D(r_Y||p_Y) \geq D(r_{XY}||p_{XY})$$

$(iii)$ *When $\lambda_1 < 0$ and $0 < \lambda_2 < 1$ equation* (1) *holds iff:*
*For any $q_Y$ there exists $r_{XY}$ with $r_Y = q_Y$ such that:*

$$\frac{1}{\lambda_1} D(r_X||p_X) + \frac{1}{\lambda_2} D(q_Y||p_Y) \geq D(r_{XY}||p_{XY})$$

A necessary condition for $(X, Y)$ to be $(\lambda_1, \lambda_2)$-reverse-hypercontractive is presented in the following theorem.

**Theorem 2** (Correlation bound)**.** *A pair of random variables $(X,Y) \sim p(x,y)$ is $(\lambda_1,\lambda_2)$-reverse-hypercontractive, for $\lambda_1,\lambda_2 \in (-\infty,1)$, only if*

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

*where $\rho_m^2$ is the maximal correlation of $(X,Y)$.*

*Remark* 1. This should be a classical result and we present a proof for completeness.

*Proof.* Let $f(x) = 1 + \delta\hat{f}(x)$, and $g(y) = 1 + \delta\hat{g}(y)$, for some $\hat{f},\hat{g}$ satisfying $\mathrm{E}(\hat{f}(X)) = \mathrm{E}(\hat{g}(Y)) = 0$, and $\mathrm{E}(\hat{f}^2(X)) = \sigma_X^2, \mathrm{E}(\hat{g}^2(Y)) = \sigma_Y^2$. Note that $-|\rho_m\sigma_X\sigma_Y| = \inf \mathrm{E}(\hat{f}(X)\hat{g}(Y))$, where the infimum is taken over $\hat{f},\hat{g}$ satisfying the above conditions. Take $\delta$ small enough so that $f,g$ are positive functions. Routine calculations show that

$$\begin{aligned}\mathrm{E}(f(X)g(Y)) &= 1 + \delta^2\,\mathrm{E}(\hat{f}(X)\hat{g}(Y))\\ \|f(X)\|_{\lambda_1} &= 1 + \delta^2\sigma_X^2\frac{\lambda_1-1}{2} + O(\delta^3)\\ \|g(Y)\|_{\lambda_2} &= 1 + \delta^2\sigma_Y^2\frac{\lambda_2-1}{2} + O(\delta^3).\end{aligned}$$

Therefore if condition (1) holds for all $f(X), g(Y)$, then by taking $\delta \to 0$, we require

$$\mathrm{E}(\hat{f}(X)\hat{g}(Y)) \geq \frac{\lambda_1-1}{2}\sigma_X^2 + \frac{\lambda_2-1}{2}\sigma_Y^2 \quad \forall \hat{f}(X),\hat{g}(Y).$$

This can be rewritten as

$$\begin{aligned}&\frac{1}{2}\left(\sqrt{1-\lambda_1}|\sigma_X| - \sqrt{1-\lambda_2}|\sigma_Y|\right)^2\\ &\quad + \sqrt{(\lambda_1-1)(\lambda_2-1)}|\sigma_X\sigma_Y| + \mathrm{E}(\hat{f}(X)\hat{g}(Y)) \geq 0.\end{aligned}$$

Hence by taking infimum over $\hat{f},\hat{g}$ we require that

$$\begin{aligned}&\frac{1}{2}\left(\sqrt{1-\lambda_1}|\sigma_X| - \sqrt{1-\lambda_2}|\sigma_Y|\right)^2\\ &\quad + \left(\sqrt{(\lambda_1-1)(\lambda_2-1)} - |\rho_m|\right)|\sigma_X\sigma_Y| \geq 0,\ \forall \sigma_X,\sigma_Y.\end{aligned}$$

This clearly requires $(\lambda_1-1)(\lambda_2-1) \geq \rho_m^2$. □

Before we state this result, we state a well-known lemma (mentioned by Mossel to the authors) that already provides a simple characterization for pairs of random variables whose support is not the entire product space $\mathcal{X}\times\mathcal{Y}$.

**Lemma 1.** *Consider a pair of random variables $(X,Y) \sim p(x,y)$. Suppose there exists $(x_0,y_0) \in \mathcal{X}\times\mathcal{Y}$ such that $p(x_0,y_0) = 0$, then for no pair $(\lambda_1,\lambda_2) \in (0,1)\times(0,1)$ will $(X,Y)$ be $(\lambda_1,\lambda_2)$-reverse-hypercontractive.*

*Proof.* The simple argument is presented here for completeness. Consider $f(X)$ and $g(Y)$ defined by $f(x_0) = 1, f(x') = \epsilon\forall x' \neq x_0$; $g(y_0) = 1, g(y') = \epsilon\forall y' \neq y_0$. Note that

$$\mathrm{E}(f(X)g(Y)) = p(x_0,y_0) + O(\epsilon) = O(\epsilon).$$

On the other hand $\|f(X)\|_{\lambda_1} \geq p_X(x_0)^{\frac{1}{\lambda_1}}, \|g(Y)\|_{\lambda_2} \geq p_Y(y_0)^{\frac{1}{\lambda_2}}$. Taking $\epsilon \to 0$, we see that (1) is violated by a suitably small $\epsilon$. Note that since $x_0,y_0$ belong to the support, $p_X(x_0), p_Y(y_0) > 0$. □

## II. Binary Erasure Channel

Consider a uniform binary random variable $X \in \{0,1\}$ passed through a binary erasure channel BEC($\epsilon$) ($0<\epsilon<1$) producing the ternary output $Y \in \{0,E,1\}$ . Concretely $P(Y=0|X=0) = \mathrm{P}(Y=1|X=1) = 1-\epsilon$ and $P(Y=E|X=0) = \mathrm{P}(Y=E|X=1) = \epsilon$.

Let $p_{XY}^{BEC}$ denote the joint law. The correlation bound for this setting says that $(X,Y)$ is $(\lambda_1,\lambda_2)$ reverse-hypercontractive for $\lambda_1,\lambda_2 \in (-\infty,1)$ only if

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

The following main new result of this paper concerns characterizing the reverse-hypercontractive region for the binary erasure channel for certain range of parameters. (This leaves as undetermined only one of the three regimes (the third one) in Theorem 1.)

**Theorem 3.** *Let $(X,Y)$ be distributed according to $p_{XY}^{BEC}$ and $\lambda_1,\lambda_2 \in (-\infty,1)\setminus\{0\}$. When $\lambda_2 < 0$, $(X,Y)$ is $(\lambda_1,\lambda_2)$ reverse-hypercontractive if and only if*

$$\lambda_1 \leq \frac{\ln 2}{\ln 2 - \frac{\lambda_2-1}{\lambda_2}\ln[(1-\epsilon)2^{\frac{1}{\lambda_2-1}}+\epsilon]}.$$

*Proof.* When $\lambda_2 < 0$ and $\lambda_1 \leq \lambda_2'(:= \frac{\lambda_2}{\lambda_2-1})$ will belong to the reverse hypercontractive region trivially from the Reverse Hölder's inequality and the monotonicity of $\|Z\|_\lambda$ in $\lambda$.

From Theorem 1 we are left with determining the range of $\lambda_1 \in (\lambda_2',1)$ satisfying the following: for any $q_X$ there exists $r_{XY}$ with $r_X = q_X$ such that

$$\frac{1}{\lambda_1}D(q_X||p_X^{BEC}) + \frac{1}{\lambda_2}D(r_Y||p_Y^{BEC}) \geq D(r_{XY}||p_{XY}^{BEC}). \tag{2}$$

We will show that the above condition holds *if and only if*

$$\lambda_1 \leq \frac{\ln 2}{\ln 2 - \frac{\lambda_2-1}{\lambda_2}\ln[(1-\epsilon)2^{\frac{1}{\lambda_2-1}}+\epsilon]}. \tag{3}$$

(3) $\implies$ (2): If $r_{XY}$ is not absolutely continuous with respect to $p_{XY}^{BEC}$, $D(r_{XY}||p_{XY}^{BEC})$ will become $+\infty$, while $\frac{1}{\lambda_1}D(q_X||p_X^{BEC}) + \frac{1}{\lambda_2}D(r_Y||p_Y^{BEC})$ are finite; violating (2). Thus, it is sufficient to search over $r_{XY}$ that are absolute continuous with respect to $p_{XY}^{BEC}$.

Denote $q_X(X=0) = x$, $r_{XY}(X=0,Y=0) = r$, $r_{XY}(X=1,Y=1) = s$. Hence $r_{XY}(X=0,Y=E) = x-r$, $r_{XY}(X=1,Y=E) = 1-x-s$, since $r_X(X=0) = q_X(X=0) = x$.

Define $f(x,r,s)$ according to

$$\begin{aligned}f(x,r,s) := \frac{1}{\lambda_1}D(q_X||p_X^{BEC}) + \frac{1}{\lambda_2}D(r_Y||p_Y^{BEC})\\ -D(r_{XY}||p_{XY}^{BEC})\end{aligned}$$

We need to show that when $\lambda_2 < 0$ and $\lambda_1$ satifies (3) then

$$\min_{x\in[0,1]}\ \max_{\substack{0\leq r\leq x,\\ 0\leq s\leq 1-x}} f(x,r,s) \geq 0.$$

Define the function

$$g(x) := \max_{0\le r\le x, 0\le s\le 1-x} f(x,r,s).$$

Then suffices that $g(x)\ge 0$ for $x\in[0,1]$. A simple symmetry argument shows that $g(x)$ is symmetric about $x=\frac{1}{2}$.

The idea of the proof is as follows: we will show that $g(x)$ has 3 stationary points in the interval $x\in(0,1)$, with one of them being at $x=\frac{1}{2}$. When $(\lambda_1-1)(\lambda_2-1)\ge 1-\epsilon$, we will show that $g(x)$ is a local minimum at $x=\frac{1}{2}$, implying that the other two symmetric stationary points correspond to local maxima. Since $g(\frac{1}{2})=0$, it suffices to verify that the boundary condition, i.e. $g(0)\ge 0$. It will turn out that this boundary point is what yields (3), the critical condition in this case.

For a fixed $x\in(0,1)$, since $\lambda_2<0$, convexity of $D(p||q)$ in $p$ immediately implies that $f(x,r,s)$ is concave in $r,s$ (when viewed as a bivariate function). Further the derivatives at the boundary tend to infinite, implying that the maximum of $f(x,r,s)$ (for a fixed $x$) is attained strictly in the interior. Thus, from concavity, there is a unique pair of points $r_0(x)\in(0,x)$ and $s_0(x)\in(0,1-x)$ such that

$$g(x)=f(x,r_0(x),s_0(x)).$$

We will first analyze the interior stationary points of $g(x)$. If $x^*$ is a stationary point, then one can check that $f(x^*,r_0(x^*),s_0(x^*))$ is a stationary point of $f(x,r,s)$. This is just a consequence of $f(x,r,s)$ being sufficiently smooth and the details are omitted here.

Setting gradients to be zero, we have

$$\begin{aligned} \frac{1}{\lambda_1}\ln\frac{x}{1-x}-\ln\frac{x-r}{1-x-s}&=0,\\ \frac{1}{\lambda_2}\ln\frac{2\epsilon r}{(1-\epsilon)(1-r-s)}-\ln\frac{\epsilon r}{(1-\epsilon)(x-r)}&=0,\\ \frac{1}{\lambda_2}\ln\frac{2\epsilon s}{(1-\epsilon)(1-r-s)}-\ln\frac{\epsilon s}{(1-\epsilon)(1-x-s)}&=0. \end{aligned}$$

These equations are the same as those Lagrange conditions in forward hypercontractivity [11]. So via the same manipulations there (not repeated here), letting $y=\frac{2(x-r)}{1-r-s}$, we have

$$\frac{1-\epsilon}{\epsilon}y^{\lambda_2'-\lambda_1}+y^{1-\lambda_1}=\frac{1-\epsilon}{\epsilon}(2-y)^{\lambda_2'-\lambda_1}+(2-y)^{1-\lambda_1} \tag{4}$$

where $\lambda_2'$ is Hölder conjugate of $\lambda_2$. Further every solution of the gradients condition is in one-to-one correspondence, same equations as in [11], with a root of (4).

According to Lemma 2, under the condition $\lambda_1\le\frac{\ln 2}{\ln 2-\frac{\lambda_2-1}{\lambda_2}\ln[(1-\epsilon)2^{\frac{1}{\lambda_2-1}}+\epsilon]}$, equation (4) has only three roots $y=1-\gamma,1+\gamma,1$ for some $\gamma\in(0,1)$.

Correspondingly, the number of interior stationary points $f(x,r,s)$ is three given by: $x_1^*=\frac{1}{2}$; and two symmetric points $x_2^*=\frac{(1+\gamma)\epsilon+(1+\gamma)^{\lambda_2'}(1-\epsilon)}{2\epsilon+(1-\epsilon)[(1+\gamma)^{\lambda_2'}+(1-\gamma)^{\lambda_2'}]}>\frac{1}{2}$, and $x_3^*=1-x_2^*=\frac{(1-\gamma)\epsilon+(1-\gamma)^{\lambda_2'}(1-\epsilon)}{2\epsilon+(1-\epsilon)[(1+\gamma)^{\lambda_2'}+(1-\gamma)^{\lambda_2'}]}<\frac{1}{2}$.

Part $(i)$ of Lemma 2 establishes that $(\lambda_1-1)(\lambda_2-1)\ge 1-\epsilon$; and under this case we will show that $x^*=\frac{1}{2}$ is a local minimizer of $g(x)$. Then $x_2^*$ and $x_3^*$ cannot be a local minimizer of $g(x)$ as $g(x)$ is continuously differentiable on $(0,1)$. Thus, $x_2^*$ and $x_3^*$ cannot be global minimizers of $g(x)$.

To show $x^*=\frac{1}{2}$ is a local minimizer of $g(x)$, notice that $g(\frac{1}{2})=f(\frac{1}{2},\frac{1-\epsilon}{2},\frac{1-\epsilon}{2})=0$. So suffices to show that for $\delta>0$ arbitrarily small, $g(\frac{1}{2}+\delta)>0$.

One can verify that

$$\begin{aligned} &f(\frac{1}{2}+\delta,r_0(\frac{1}{2}+\delta),s_0(\frac{1}{2}+\delta))\\ &=2(\frac{1}{\lambda_1}-\frac{1-\lambda_2}{\epsilon-\lambda_2})\delta^2+O(\delta^3). \end{aligned}$$

which is strictly positive for small $\delta$ precisely when

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

Thus the global minimizer of $g(x)$ can only be one of the three points $\{0,\frac{1}{2},1\}$. By symmetry $g(0)=g(1)$. Now $g(0)=\max_{s\in[0,1]}f(0,0,s)$, where

$$\begin{aligned} f(0,0,s)=\frac{1}{\lambda_1}\ln 2+\frac{1}{\lambda_2}[s\ln\frac{2s}{1-\epsilon}+(1-s)\ln\frac{1-s}{\epsilon}]\\ -s\ln\frac{2s}{1-\epsilon}-(1-s)\ln\frac{2(1-s)}{\epsilon} \end{aligned}$$

Notice the above function is concave over $s$. By taking derivative over $s$, we get that the maximum point $s_0(0)=\frac{1-\epsilon}{1-\epsilon+2^{\frac{1}{1-\lambda_2}}\epsilon}$.

Thus $f(0,0,s_0(0))\ge 0$ is equivalent (after re-arranging) to

$$\lambda_1\le\frac{\ln 2}{\ln 2-\frac{\lambda_2-1}{\lambda_2}\ln[(1-\epsilon)2^{\frac{1}{\lambda_2-1}}+\epsilon]}.$$

This range, from first part of Lemma 2, also satisfies $(\lambda_1-1)(\lambda_2-1)>1-\epsilon$, implying that when (3) holds, $g(x)\ge 0$ for all $x\in[0,1]$ and hence (2) holds.

(2) $\Rightarrow$ (3): Let $q_X(X=0)=0$. If $r_{XY}$ is not absolutely continuous with respect to $p_{XY}^{BEC}$, $D(r_{XY}||p_{XY}^{BEC})$ will become $+\infty$, while $\frac{1}{\lambda_1}D(q_X||p_X^{BEC})$, $\frac{1}{\lambda_2}D(r_Y||p_Y^{BEC})$ are finite, which contradicts the condition. Suffices to consider the case when $r_{XY}$ is absolutely continuous with respect to $p_{XY}^{BEC}$.

As before denote $r(X=1,Y=1)=s,(0\le s\le 1)$. The condition $\frac{1}{\lambda_1}D(q_X||p_X^{BEC})+\frac{1}{\lambda_2}D(r_Y||p_Y^{BEC})\ge D(r_{XY}||p_{XY}^{BEC})$ for some $r_{XY}$ with $r_X=q_X$ leads to $f(0,0,s)\ge 0$ for some $s\in[0,1]$. But as mentioned in the previous section, this is equivalent to $f(0,0,s_0(0))\ge 0$, which leads to

$$\lambda_1\le\frac{\ln 2}{\ln 2-\frac{\lambda_2-1}{\lambda_2}\ln[(1-\epsilon)2^{\frac{1}{\lambda_2-1}}+\epsilon]}.$$

□

**Lemma 2.** *Let* $\lambda_2'<\lambda_1<1,\lambda_2<0$*. When* $\lambda_1\le\frac{\ln 2}{\ln 2-\frac{\lambda_2-1}{\lambda_2}\ln[(1-\epsilon)2^{\frac{1}{\lambda_2-1}}+\epsilon]}$*, the following hold:*

$(i)$ $(\lambda_1-1)(\lambda_2-1)\ge 1-\epsilon$*. Further the inequality is strict if* $\epsilon\in(0,1)$*.*

*(ii) The equation*

$$\frac{1-\epsilon}{\epsilon}x^{\lambda_2'-\lambda_1}+x^{1-\lambda_1}=\frac{1-\epsilon}{\epsilon}(2-x)^{\lambda_2'-\lambda_1}+(2-x)^{1-\lambda_1}$$

*has three roots $x=1-\gamma,1,1+\gamma$ for some $\gamma\in(0,1)$.*

*Proof.* Note that

$$(\lambda_1-1)(\lambda_2-1)\geq\frac{\frac{(\lambda_2-1)^2}{\lambda_2}\ln[(1-\epsilon)2^{\frac{1}{\lambda_2-1}}+\epsilon]}{\ln 2-\frac{\lambda_2-1}{\lambda_2}\ln[(1-\epsilon)2^{\frac{1}{\lambda_2-1}}+\epsilon]}.$$

Therefore it suffices to show that the right-hand-side is larger than $1-\epsilon$ when $\lambda_2<0$. Setting $r=\frac{1}{1-\lambda_2}\in(0,1)$ and substituting into the right-hand-side, it suffices to show that

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

This can be rearranged as

$$(1-\epsilon)+\epsilon 2^r\leq 2^{\frac{\epsilon r}{1-r+\epsilon r}}.$$

It is a rather immediate exercise to verify that the right-hand-side is strictly concave in $\epsilon$, for $\epsilon\in(0,1)$; and since equality holds at $\epsilon=0$ and $\epsilon=1$, we have the desired result. This establishes part $(i)$ of the lemma.

Proof of $(ii)$: Define the function $h(x)$

$$\begin{aligned}h(x)=&\frac{1-\epsilon}{\epsilon}x^{\lambda_2'-\lambda_1}+x^{1-\lambda_1}\\&-\frac{1-\epsilon}{\epsilon}(2-x)^{\lambda_2'-\lambda_1}-(2-x)^{1-\lambda_1}.\end{aligned}$$

Note that $h(1)=0$, $\lim_{x\downarrow 0}h(x)=+\infty$. Further $h(x)=-h(2-x)$. Part $(ii)$ follows by showing that there is only one root for $h(x)=0$ for $x\in(0,1)$.

Take the derivative with respect to $x$,

$$\begin{aligned}h'(x)=&\frac{(1-\epsilon)(\lambda_2'-\lambda_1)}{\epsilon}x^{\lambda_2'-\lambda_1-1}+(1-\lambda_1)x^{-\lambda_1}\\&+\frac{(1-\epsilon)(\lambda_2'-\lambda_1)}{\epsilon}(2-x)^{\lambda_2'-\lambda_1-1}+(1-\lambda_1)(2-x)^{-\lambda_1}.\end{aligned}$$

Note that

$$(1-\epsilon)\lambda_2'+\epsilon-\lambda_1>0\iff(\lambda_1-1)(\lambda_2-1)>1-\epsilon.$$

Thus $h'(1)=2\left(\frac{(1-\epsilon)\lambda_2'+\epsilon-\lambda_1}{\epsilon}\right)>0$ from part $(i)$. Hence $h(x)=0$ will have at least one root in $(0,1)$.

The claim that $h(x)=0$ has only one root in $(0,1)$ will follow by showing that $h(x)$ first decreases and then increases on $(0,1)$; in other words $h'(x)$ has only one root in $(0,1)$. Since $\lim_{x\downarrow 0}h'(x)=-\infty$, $h'(1)>0$, and $h'(x)$ is continuous on $(0,1]$, implies that there is at least one root at $x=1-y_0$ for $y_0\in(0,1)$ for $h'(x)$.

Setting $x=1-y$ and considering the Taylor Series expansion of $h'(x)$ w.r.t. $y$, about $y=0$, we obtain

$$\begin{aligned}h'(1-y)=2\sum_{k=0}^{\infty}\Bigg[&\frac{(1-\epsilon)(\lambda_2'-\lambda_1)}{\epsilon}\binom{\lambda_2'-\lambda_1-1}{2k}\\&+(1-\lambda_1)\binom{-\lambda_1}{2k}\Bigg]y^{2k}.\end{aligned}$$

Let $a_k=(1-\lambda_1)\binom{-\lambda_1}{2k}$ and $b_k=\frac{(1-\epsilon)(\lambda_1-\lambda_2')}{\epsilon}\binom{\lambda_2'-\lambda_1-1}{2k}$. Note that $a_k,b_k\geq 0$ and

$$h'(1-y)=2\sum_{k\geq 0}(a_k-b_k)y^{2k}.$$

Note that $a_0\geq b_0$ (from part $(i)$ or since this is $h'(1)$).

Suppose there exists $k_0$ such that $a_{k_0}\leq b_{k_0}$, then $a_k\leq b_k,\forall k\geq k_0$. This follows basically from an induction argument, since

$$\begin{aligned}a_{k+1}&=a_k\frac{(\lambda_1+2k)(\lambda_1+2k+1)}{(2k+1)(2k+2)}\\b_{k+1}&=b_k\frac{(\lambda_1+1-\lambda_2'+2k)(\lambda_1+1-\lambda_2'+2k+1)}{(2k+1)(2k+2)}.\end{aligned}$$

$1-\lambda_2'>0$ implies that once $b_k\geq a_k$, the inequality continues to hold for larger $k$. Since $h'(1-y)=0$ has a root in $(0,1)$, implies that $\exists\ m\geq 0$ such that $a_k\geq b_k,\forall k\leq m$ and $b_k\geq a_k,\forall k>m$.

Define $c_k=|a_k-b_k|$. Then

$$h'(1-y)=\sum_{k=0}^{m}c_ky^{2k}-\sum_{k\geq m+1}^{\infty}c_ky^{2k}$$

where $c_k\geq 0$ (with at least one $c_k$ in each range, $k\in[1:m]$ and $k\geq m+1$ being strictly positive). Let $y_0\in(0,1)$ be a root of $h'(1-y)=0$.

For $y>y_0>0$, note that

$$\begin{aligned}\sum_{k=0}^{m}c_ky^{2k}&<\left(\frac{y}{y_0}\right)^{2m}\sum_{k=0}^{m}c_ky_0^{2k}\\&=\left(\frac{y}{y_0}\right)^{2m}\sum_{k=m+1}^{\infty}c_ky_0^{2k}\\&<\sum_{k=m+1}^{\infty}c_ky^{2k}.\end{aligned}$$

The equality above is a consequence of $y_0$ being a root. Thus, no $y>y_0$ can be a root of $h'(1-y)=0$. Similarly, reversing inequalities above, for $0<y<y_0$, $y$ cannot be a root for $h'(1-y)=0$.

Thus $h'(x)=0$ has only one root in the interval $x\in(0,1)$, and as $\lim_{x\downarrow 0}h'(x)=-\infty$, $h'(1)>0$, due to the continuity of $h'(x)$, we have $h'(x)<0$ for $x\in(0,1-y_0)$ and $h'(x)>0$ for $x\in(1-y_0,1)$. Putting this together with $\lim_{x\downarrow 0}h(x)=+\infty$ and $h(1)=0$ implies that, $h(x)=0$ has precisely one root, say $x=1-\gamma$, in the interval $x\in(0,1)$. Since $h(1-y)$ is an odd function; the roots are given by $x=1-\gamma,1,1+\gamma$. This completes the proof of part $(ii)$. □

### *A. Binary Symmetric Channel*

The same technique that we employed can be used for the case of uniform binary input, $X$, passing through a binary symmetric channel to produce $Y$. In this case, a result due to Borrell [1] already shows that the correlation bound is tight and the technqiue developed here just provides another proof. Due to space limitations, we only provide an outline of this argument. As you will see, this case is considerably simpler than that of the erasure channel.

The information-measure characterization in Theorem 1 essentially reduces to checking that a certain min-max expression is non-negative. By analyzing each case (in Theorem 1) separately we can show, in a similar fashion, that any interior local minimum must be a stationary point.

Further by analyzing the first derivative conditions, we will arrive that all stationary points are in one-to-one correspondence with the set of $y$ satisfying

$$y^{-t(\frac{1-\theta}{1+\theta})^2} = \frac{(1+\theta y)^t\theta + (\theta+y)^t}{(\theta+y)^t\theta + (1+\theta y)^t}, \tag{5}$$

for some appropriately defined $t \in (-\infty, 0)$ and $\theta \in (0, \infty)$. This is identical to the forward analysis presented in [11] and the details are omitted. As shown again in the forward case, the above equation has a unique root $y = 1$ in $(0, \infty)$; when $\theta \in (0, \infty)\backslash\{1\}$. This shows that the unique interior stationary point is $r_{XY} = p_{XY}^{BSC}$. Contrary to the binary erasure channel, it turns out that the boundary points do not influence the reverse-hypercontractive region.

## III. Discussion

As shown in [12] hypercontractive region is same as the Gray-Wyner source coding region. In recent past a variety of computations of capacity regions (or achievable regions) have been performed in network information theory, mostly by the authors and/or their collaborators. All of them involve optimizing non-convex functions over probability spaces. The functions are linear combinations of information measures and usually satisfy the tensorization (or sub-additivity) property. The exact computations have been in small alphabet spaces, but nevertheless in all the cases, the global maximizer could be identified by a local analysis.

In many cases, for instance [13], there is only a single interior local optimizer; and sometimes it is a competition between the boundary and the interior point, [14]. However, in each case, the proofs are quite complicated and require careful analysis with very few re-use of specific results. There are some other similar problems (conjectures), for example the one in [15], where numerically, there do not exist any other local optimizer other than the conjectured ones. However, a rigorous mathematical proof is lacking for many of these settings.

It is possible that all the problems being considered belong to a larger sub-class of non-convex problems where a certain set of standard tools could be devised to isolate the global maximizers. This could have far reaching consequences: for instance a fast approximation algorithm for obtaining the hypercontractivity parameters (which would then make certain classes of problems, believed to be inapproximable better than a certain ratio, have better approximation guarantees using polynomial time algorithms), or even more massively, some type of justification for why the Replica heuristic from Statistical Physics gives accurate predictions. Therefore, these calculations even for special cases may help someone piece together a cohesive picture.

## IV. Conclusion

In this paper we derive the reverse-hypercontractivity region for a pair of variables distributed via the binary erasure channel with uniform input probability. The technique employed is essentially a local analysis (identifying local extremal points and comparing the function values between them). The key insight that enables us to do this effectively is that a certain Taylor series expansion has exactly one sign change in its coefficients, leading us to get a control on the number of stationary points. The analysis techniques used here may be of independent interest for a variety of problems where, numerically, a local analysis seems to provide a solution; but a proper analysis has not yet been completed.

## References


[1] C. Borell, "Positivity improving operators and hypercontractivity." *Mathematische Zeitschrift*, vol. 180, pp. 225–234, 1982. [Online]. Available: http://eudml.org/doc/173183

[2] E. Mossel, K. Oleszkiewicz, and A. Sen, "On reverse hypercontractivity," *Geometric and Functional Analysis*, vol. 23, no. 3, pp. 1062–1097, 2013. [Online]. Available: http://dx.doi.org/10.1007/s00039-013-0229-4

[3] E. Mossel, R. O'Donnell, O. Regev, J. E. Steif, and B. Sudakov, "Non-interactive correlation distillation, inhomogeneous markov chains, and the reverse bonami-beckner inequality," *Israel Journal of Mathematics*, vol. 154, no. 1, pp. 299–336, 2006. [Online]. Available: http://dx.doi.org/10.1007/BF02773611

[4] 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.

[5] C. Nair, "An extremal inequality related to hypercontractivity of gaussian random variables," *Information Theory and Applications Workshop*, 2014.

[6] ——, "Equivalent formulations of hypercontractivity using information measures," *International Zurich Seminar*, 2014.

[7] S. Kamath, "Reverse hypercontractivity using information measures," in *Proceedings of the 53rd Annual Allerton Conference on Communications, Control and Computing*, 2015.

[8] S. Beigi and C. Nair, "Equivalent characterization of reverse brascamp-lieb-type inequalities using information measures," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 1038–1042.

[9] J. Liu, T. A. Courtade, P. W. Cuff, and S. Verdú, "Brascamp-lieb inequality and its reverse: An information theoretic view," in *IEEE International Symposium on Information Theory, ISIT 2016, Barcelona, Spain, July 10-15, 2016*, 2016, pp. 1048–1052. [Online]. Available: http://dx.doi.org/10.1109/ISIT.2016.7541459

[10] V. Guruswami and J. Radhakrishnan, "Tight Bounds for Communication-Assisted Agreement Distillation," in *31st Conference on Computational Complexity (CCC 2016)*, ser. Leibniz International Proceedings in Informatics (LIPIcs), R. Raz, Ed., vol. 50. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016, pp. 6:1–6:17. [Online]. Available: http://drops.dagstuhl.de/opus/volltexte/2016/5845

[11] C. Nair and Y. N. Wang, "Evaluating hypercontractivity parameters using information measures," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 570–574.

[12] S. Beigi and A. Gohari, "On the duality of additivity and tensorization," *CoRR*, vol. abs/1502.00827, 2015. [Online]. Available: http://arxiv.org/abs/1502.00827

[13] Y. Geng, V. Jog, C. Nair, and Z. V. Wang, "An information inequality and evaluation of marton's inner bound for binary input broadcast channel," *IEEE Trans. Info. Theory (accepted)*.

[14] M. H. M. Costa and C. Nair, "Gaussian z-interference channel: around the corner," *Information Theory and Applications Workshop (ITA)*, 2016.

[15] M. Sefidgaran, A. Gohari, and M. R. Aref, "On korner-marton's sum modulo two problem," in *2015 Iran Workshop on Communication and Information Theory (IWCIT)*, May 2015, pp. 1–6.