Reverse Hypercontractivity using Information Measures

Sudeep Kamath
EE Department,
Princeton University,
Princeton, NJ USA
sukamath@princeton.edu

***Abstract*—We provide an equivalent description of reverse hypercontractivity using inequalities among information measures.**

***Index Terms*—reverse hypercontractivity**

I. Introduction

Hypercontractivity and reverse hypercontractivity have had important applications in mathematics, physics, and computer science [1]–[8]. In recent years, a number of results in different areas within information theory and computer science have critically utilized reverse hypercontractivity [9]–[13]. The basic workhorse in the application of reverse hypercontractivity in many of these results has been a lower bound on the joint probability of two high-dimensional random variables (say X^n, Y^n of dimension n) falling in high-dimensional sets (say S, T respectively) in terms of their respective marginal probabilities, of the form

$$P(X^n \in S, Y^n \in T) \geq P(X^n \in S)^{\alpha} P(Y^n \in T)^{\beta} \ , \tag{1}$$

for suitable exponents $\alpha, \beta > 0$.

In this paper, we study connections between reverse hypercontractivity and inequalities among information measures. The first explicit connection between information measures and hypercontractivity was discovered by Ahlswede and Gács [14], who related the strong data processing constant to the limiting chordal slope of hypercontractivity. This result was extended in great generality to the entire regime of hypercontractivity by Nair in [15]. Carlen and Cordero-Erausquin in [16, Thm. 2.1] have established a similar equivalence for the closely-related Brascamp-Lieb type inequalities.

In this paper, we show that the regime in which reverse hypercontractivity provides the useful lower bounds on joint probabilities in terms of marginal probabilities as in (1), has an equivalent description purely using inequalities among information measures. We call this regime 'reverse hypercontractivity in the second quadrant'. Our main result can be viewed as a natural extension of the program of studying equivalence of hypercontractivity and information measures [14]–[16] to reverse hypercontractivity. The novel technical contribution of this work lies in the implication from the information measure inequalities to the reverse hypercontractive inequalities which is obtained by studying the geometric properties of the extremal functions satisfying reverse hypercontractivity.

The hope is that these equivalence results may help strengthen the bridge between information theory and functional analysis, probability theory, and theoretical computer science. This can facilitate the use of tools from one field to provide new insights or ideas in another. Evidence of this has already surfaced in Nair's recent proof [17] of Gaussian hypercontractivity [3] using the interpretation of hypercontractivity as inequalities among information measures [15].

Other related works. For the doubly symmetric binary source distribution $(X, Y) \sim \mathrm{DSBS}(\epsilon)$ defined as X, Y being marginally uniform random variables taking values in $\{0, 1\}$ with $P(Y \neq X | X = 0) = P(Y \neq X | X = 1) = \epsilon$, the hypercontractivity region is explicitly known, and the corresponding hypercontractive inequality is sometimes referred to as the Bonami-Beckner inequality [1], [2], [5]. It is to be noted that the works [18]–[20] provide information-theoretic proofs of the hypercontractivity region corresponding to this specific joint distribution. In contrast, our result in this paper provides an information measure inequality interpretation to the *phenomenon* of reverse hypercontractivity itself, providing an equivalent description of reverse hypercontractivity for arbitary joint distributions over finite alphabets. While this interpretation may help understand the reverse hypercontractivity region for specific joint distributions such as the doubly symmetric binary source and possibly others, the present paper does not explore this direction.

The rest of this paper is organized as follows. Sec. II sets up notation and preliminaries of reverse hypercontractivity. Sec. III contains our main result and its proof. Sec. IV observes some simple applications of the main result and concludes with some remarks and open questions.

II. Preliminaries

Notation 1. In this paper, we let $\mathcal{X}, \mathcal{Y}, \mathcal{U}$ denote arbitrary finite sets and generic elements in $\mathcal{X}, \mathcal{Y}, \mathcal{U}$ will be denoted by lower-case x, y, u respectively. E.g. $\forall x$ will be understood to mean $\forall x \in \mathcal{X}$. $\sum_y h(y)$ will mean $\sum_{y \in \mathcal{Y}} h(y)$.

Notation 2. We will use p, q, r to denote probability distributions on $\mathcal{X}$ or $\mathcal{Y}$ or $\mathcal{X} \times \mathcal{Y}$. The exact set will be specified by subscripts, with or without dummy arguments. E.g. $p_Y(y)$, $q_{X,Y}(x, y)$, and r_X are probability distributions on $\mathcal{Y}, \mathcal{X} \times \mathcal{Y}$, and $\mathcal{X}$ respectively. $p_X > 0$ will be understood to mean $p_X(x) > 0,\ \forall x \in \mathcal{X}$, and $p_{X,Y} > 0$ will mean $p_{X,Y}(x, y) > 0\ \forall x \in \mathcal{X}, \forall y \in \mathcal{Y}$. Furthermore, $p_{X,Y}(x, y)$ is a fixed reference distribution throughout and all expectations will be computed according to it.

Notation 3. We use f, g to denote real-valued functions, the domain being specified by the context. $\forall f$ will mean the set of all real-valued functions with the given domain and $g \geq 0$ means g takes non-negative values.

Definition 1. For two probability distributions r_Z, p_Z on the same finite alphabet, let

$$D(r_Z||p_Z) := \sum_z r_Z(z) \log \frac{r_Z(z)}{p_Z(z)}$$

denote their relative entropy[1]. For random variables W, Z taking values in finite alphabets, let

$$\begin{aligned} I(W; Z) &:= D(p_{W,Z}||p_W \times p_Z) \\ &= \sum_{w,z} p_{W,Z}(w, z) \log \frac{p_{W,Z}(w, z)}{p_W(w)p_Z(z)}, \end{aligned}$$

denote the mutual information between W and Z.

Definition 2. For any real-valued random variable W with finite support, and any real number p, define

$$||W||_p := \begin{cases} \left(\mathbb{E}|W|^p\right)^{1/p}, & p \neq 0; \\ \exp\left(\mathbb{E} \log |W|\right) & p = 0, \end{cases} \tag{2}$$

with the understanding that for $p \leq 0$, $||W||_p = 0$ if $\Pr(|W| = 0) > 0$.

Definition 3. Define the Hölder conjugate of $p \neq 0, 1$ by $p' := \frac{p}{p-1}$. For $p = 0$, define $p' = 0$. For $0 < p < 1$, we have $p' < 0$.

Definition 4. For a pair of random variables $(X, Y) \sim p_{X,Y}(x, y)$ on $\mathcal{X} \times \mathcal{Y}$, we say (X, Y) is

- (p, q)-hypercontractive if $1 \leq q \leq p$, and
$$||\mathbb{E}[g(Y)|X]||_p \leq ||g(Y)||_q \quad \forall g \geq 0; \tag{3}$$
- (p, q)-reverse hypercontractive if $1 \geq q \geq p$, and
$$||\mathbb{E}[g(Y)|X]||_p \geq ||g(Y)||_q \quad \forall g \geq 0. \tag{4}$$

[1] All logarithms in this paper are natural logarithms.

Define the hypercontractivity ribbon as the set of pairs (p, q) for which (X, Y) is hypercontractive or reverse hypercontractive.

Fig. 1 shows an illustration of a hypercontractivity ribbon. In this paper, we will concern ourselves with 'reverse hypercontractivity in the second quadrant', i.e. in the region $p < 0 < q < 1$ which is shaded in the figure. We will show that we can understand this region purely using inequalities among information measures.

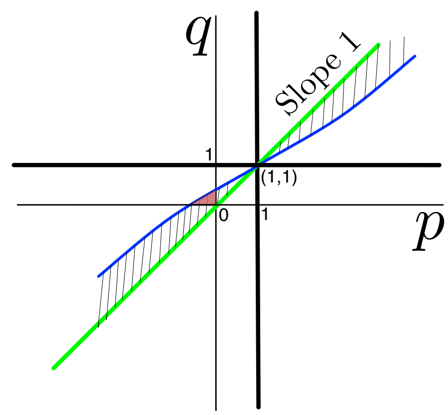

Fig. 1. Hypercontractivity ribbon: the region in the second quadrant $p < 0 < q < 1$ is highlighted by shading

III. Main Result

Consider a joint distribution $p_{X,Y}$. Suppose there exist x_0, y_0 so that $p_{X,Y}(x_0, y_0) = 0$ but $p_X(x_0), p_Y(y_0) > 0$. Then, we cannot have reverse hypercontractivity in the second quadrant, i.e. there do not exist $p < 0 < q < 1$ so that (4) holds, because we may choose $g(y) = \mathbb{1}_{[y=y_0]}$ for which $||g(Y)||_q > 0$ and $||\mathbb{E}[g(Y)|X]||_p = 0$. We will therefore restrict ourselves to joint distributions $p_{X,Y}(x, y) > 0,\ \forall x \in \mathcal{X}, y \in \mathcal{Y}$. In this case, it is easy to show by a simple argument that that there must be *some* reverse hypercontractivity in the second quadrant, that is, there exist some $p < 0 < q < 1$ so that (4) holds (eg. see [10, Proposition 8.1]).

Our main result below shows that when $p_{X,Y} > 0$, reverse hypercontractivity in the second quadrant can be alternatively described by inequalities among information measures.

Theorem 1. *Fix a joint distribution $p_{X,Y}$ so that $p_{X,Y} > 0$. Fix $0 < a, b \leq 1$, so that $a' = \frac{a}{a-1} < 0$. The following are equivalent:*

(i) $||\mathbb{E}[g(Y)|X]||_{a'} \geq ||g(Y)||_b \ \ \forall\ g \geq 0;$

(ii) $\mathbb{E}f(X)g(Y) \geq ||f(X)||_a||g(Y)||_b \;\; \forall\, f, g \geq 0$;

(iii) $\inf_{q_{X,Y,U}\in\mathcal{J}(X,Y,U)} D(q_{X,Y,U}||p_{X,Y}\times p_U)$

$\leq \frac{1}{a}I(U;X)+\frac{1}{b}I(U;Y)\;\forall\, U$*(finite valued)*;

(iv) $\inf_{q_{X,Y}\in\mathcal{K}(r_X,r_Y)} D(q_{X,Y}||p_{X,Y})$

$\leq \frac{1}{a}D(r_X||p_X)+\frac{1}{b}D(r_Y||p_Y) \;\; \forall\, r_X, r_Y$,

where the infimum in (iii) *is over all joint distributions belonging to* $\mathcal{J}(X,Y,U)$ *defined as the set of all* $q_{X,Y,U}$ *which have pairwise marginals specified by* $q_{X,U} = p_{X,U}$ *and* $q_{Y,U} = p_{Y,U}$*; and the infimum in* (iv) *is over all joint distributions belonging to* $\mathcal{K}(r_X, r_Y)$ *defined as the set of all* $q_{X,Y}$ *which have marginals specified by* r_X, r_Y*, i.e.* $q_X = r_X, q_Y = r_Y$.

Remark 1. $\inf_{q_{X,Y,U}\in\mathcal{J}(X,Y,U)} D(q_{X,Y,U}||p_{X,Y}\times p_U)$ can also be equivalently written as:

$$\inf_{q_{X,Y,U}\in\mathcal{J}(X,Y,U)} \left[I_q(\tilde{X},\tilde{Y};\tilde{U}) + D(q_{X,Y}||p_{X,Y})\right],$$

where I_q denotes mutual information for the joint triple distribution $(\tilde{X},\tilde{Y},\tilde{U}) \sim q_{X,Y,U}$.

Remark 2. $\inf_{q_{X,Y}\in\mathcal{K}(r_X,r_Y)} D(q_{X,Y}||p_{X,Y})$ appears as the error exponent of hypothesis testing when structurally restricted to comparator-based tests [21].

The proof follows along the arguments of [15] for the analogous equivalence of hypercontractivity with inequalities among information measures. The main contribution is in establishing $(iv) \implies (ii)$ where we use the geometric properties of the extremal functions in (ii) and the structure of the optimizing joint distribution in the left hand side of (iv).

Proof of Theorem 1. The equivalence between (i) and (ii) is a simple exercise using reverse Hölder's inequality. We will prove $(ii) \implies (iii) \implies (iv) \implies (ii)$.

$\mathbf{(ii)} \implies \mathbf{(iii)}$: Fix any joint distribution $(U,X,Y) \sim p_{U,X,Y}(u,x,y)$ so that the marginal distribution of the pair (X,Y) is $p_{X,Y}$ and so that U takes values in the finite alphabet $\mathcal{U}$. For $z^n = (z_1,\ldots,z_n)$, let $N(z|z^n)$ denote $|\{i : 1 \leq i \leq n, z = z_i\}|$. Fix a sequence u^n so that

$$\left|N(u|u^n) - np_U(u)\right| \leq |\mathcal{U}| \;\forall u \in \mathcal{U}.$$

Define sets using the above fixed u^n as follows:

$$A_n := \{x^n : \left|N(u,x|u^n,x^n) - np_{U,X}(u,x)\right| \leq \sqrt{n}\log(n)p_{U,X}(u,x) \;\forall u \in \mathcal{U}, \forall x \in \mathcal{X}\}, \quad (5)$$

$$B_n := \{y^n : \left|N(u,y|u^n,y^n) - np_{U,Y}(u,y)\right| \leq \sqrt{n}\log(n)p_{U,Y}(u,y) \;\forall u \in \mathcal{U}, \forall y \in \mathcal{Y}\}. \quad (6)$$

A counting argument (e.g. [22, Chapter 2]) can be used to obtain:

$$P(X^n \in A_n) = \exp(-nI(U;X) + o(n)), \quad (7)$$

$$P(Y^n \in B_n) = \exp(-nI(U;Y) + o(n)), \quad (8)$$

$$P(X^n \in A_n, Y^n \in B_n) = \exp(-n \inf_{q_{X,Y,U}\in\mathcal{J}(X,Y,U)} D(q_{X,Y,U}||p_{X,Y}\times p_U) + o(n)). \quad (9)$$

We use the tensorization property of reverse hypercontractivity [10], [23]. Assume that (ii) is true. For $\{(X_i,Y_i)\}_{i=1}^n$ drawn i.i.d. from $p_{X,Y}$, we have

$$\mathbb{E}f(X^n)g(Y^n) \geq ||f(X^n)||_a||g(Y^n)||_b \;\; \forall\, f, g \geq 0. \quad (10)$$

Then, for any integer n and any sets $C_n \subseteq \mathcal{X}^n, D_n \subseteq \mathcal{Y}^n$, using functions $f(x^n) = \mathbb{1}_{[x^n\in C_n]}, g(y^n) = \mathbb{1}_{[y^n\in D_n]}$ in (10), we get

$$P(X^n \in C_n, Y^n \in D_n) \geq P(X^n \in C_n)^{\frac{1}{a}} P(Y^n \in D_n)^{\frac{1}{b}}. \quad (11)$$

Choose $C_n = A_n$ from (5), $D_n = B_n$ from (6), use (7), (8), (9), take logarithms on both sides, divide by n, and take the limit $n \to \infty$ to get (iii).

$\mathbf{(iii)} \implies \mathbf{(iv)}$: Suppose that (iii) is true. Fix any probability measures r_X, r_Y, and consider a family of joint triples $(U,X,Y) = (U_\epsilon,X,Y) \sim p_{X,Y,U}$ indexed by ϵ and well-defined for all sufficiently small ϵ satisfying $0 \leq \epsilon < \epsilon_0 = \min_{x,y}\frac{p_{X,Y}(x,y)}{r_X(x)r_Y(y)}$ specified for $\epsilon > 0$ as follows:

$$P(U_\epsilon = 1) = \epsilon, \qquad P(U_\epsilon = 0) = 1-\epsilon,$$

$$P(X=x, Y=y|U_\epsilon = 1) = r_X(x)r_Y(y),$$

$$P(X=x, Y=y|U_\epsilon = 0) = \frac{p_{X,Y}(x,y) - \epsilon r_X(x)r_Y(y)}{1-\epsilon},$$

and as $P(x = x, Y = y, U_\epsilon = 0) = p_{X,Y}(x,y)$ for $\epsilon = 0$.

Fix any $q_{X,Y,U} \in \mathcal{J}(X,Y,U_\epsilon)$. Note that $q_U = p_U$. Thus,

$$\begin{aligned} &D(q_{X,Y,U}||p_{X,Y}\times p_U) \\ =&p_U(1)D(q_{X,Y|U=1}||p_{X,Y}) + p_U(0)D(q_{X,Y|U=0}||p_{X,Y}) \\ \geq&\quad \epsilon D(q_{X,Y|U=1}||p_{X,Y}) \\ \geq&\quad \epsilon \inf_{q_{X,Y}\in\mathcal{K}(r_X,r_Y)} D(q_{X,Y}||p_{X,Y}), \end{aligned}$$

where the last inequality follows from the fact that $q_{X|U=1} = p_{X|U=1} = r_X(x), q_{Y|U=1} = p_{Y|U=1} = r_Y(y)$, since the pairwise marginals of (X,U) and (Y,U) are the same in $q_{X,Y,U}$ and $p_{X,Y,U}$. This gives

$$\inf_{q_{X,Y,U}\in\mathcal{J}(X,Y,U)} D(q_{X,Y,U}||p_{X,Y}\times p_U) \geq \epsilon \inf_{q_{X,Y}\in\mathcal{K}(r_X,r_Y)} D(q_{X,Y}||p_{X,Y}). \quad (12)$$

Using this inequality with (iii) leads to the fact that the continuously differentiable function ϕ defined by

$$\phi(\epsilon) := \frac{1}{a} I(U_\epsilon : X) + \frac{1}{b} I(U_\epsilon; Y) - \epsilon \inf_{q_{X,Y} \in \mathcal{K}(r_X, r_Y)} D(q_{X,Y} || p_{X,Y}), \quad (13)$$

is non-negative for $0 \le \epsilon \le \epsilon_0$. We find that $\phi(0) = 0$, so we must have $\phi'(0) \ge 0$. Writing the latter out explicitly gives us (iv).

(iv) $\implies$ **(ii)** : This is the novel contribution in the paper. We will use the lemma below which is proved in the Appendix.

Lemma 1. *Suppose $p_{X,Y} > 0$. Then, the following are equivalent.*

(iv) $\inf_{q_{X,Y} \in \mathcal{K}(r_X, r_Y)} D(q_{X,Y} || p_{X,Y}) \le \frac{1}{a} D(r_X || p_X) + \frac{1}{b} D(r_Y || p_Y) \quad \forall\, r_X, r_Y.$

(iv') $D(r_{X,Y} || p_{X,Y}) \le \frac{1}{a} D(r_X || p_X) + \frac{1}{b} D(r_Y || p_Y)$ $\forall\, r_{X,Y}$ *of the form*

$$r_{X,Y}(x, y) = \frac{p_{X,Y}(x, y) f(x) g(y)}{\mathbb{E} f(X) g(Y)},$$

where $f, g \ge 0$, and $\mathbb{E} f(X) g(Y) > 0$.

We show $(iv) \implies (ii)$ by showing the contrapositive. Suppose (ii) is false, so that there exist $f, g \ge 0$ that satisfy $\mathbb{E} f(X) g(Y) < ||f(X)||_a ||g(Y)||_b$. This means

$$\sup_{f,g \ge 0, \mathbb{E} f(X) g(Y) = 1} ||f(X)||_a ||g(Y)||_b > 1. \quad (14)$$

Fix any f^*, g^* that achieve the supremum above; the supremum is easily argued to be attained since sets $\mathcal{X}, \mathcal{Y}$ are finite[2]. As $\mathbb{E} f^*(X) g^*(Y) = 1$, neither f^* nor g^* can be identically zero. Since $p_{X,Y}(x, y) > 0$ for all $x \in \mathcal{X}, y \in \mathcal{Y}$, we have $\mathbb{E}[g^*(Y) | X = x] > 0$ for all $x \in \mathcal{X}$, $\mathbb{E}[f^*(X) | Y = y] > 0$ for all $y \in \mathcal{Y}$. For the fixed $g = g^*$, f^* is a local (and global) maximum for the corresponding constrained concave maximization problem, i.e.

$$f^* = \operatorname*{arg\,max}_{f \ge 0, \mathbb{E} f(X) g^*(Y) = 1} ||f(X)||_a. \quad (15)$$

By using Lagrange multipliers, this means that:

$$f^*(x) = \frac{\mathbb{E}[g^*(Y) | X = x]^{-\frac{1}{1-a}}}{||\mathbb{E}[g^*(Y)|X]||_{-\frac{a}{1-a}}^{-\frac{a}{1-a}}}. \quad (16)$$

Similarly, for the fixed $f = f^*$, g^* is a local (and global) maximum for the corresponding constrained concave maximization problem and by using Lagrange multipliers, we must have:

$$g^*(y) = \frac{\mathbb{E}[f^*(X) | Y = y]^{-\frac{1}{1-b}}}{||\mathbb{E}[f^*(X)|Y]||_{-\frac{b}{1-b}}^{-\frac{b}{1-b}}}. \quad (17)$$

Now, define

$$r^*_{X,Y}(x, y) := \frac{p_{X,Y}(x, y) f^*(x) g^*(y)}{\mathbb{E} f^*(X) g^*(Y)} = p_{X,Y}(x, y) f^*(x) g^*(y), \quad (18)$$

$$r^*_X(x) = p_X(x) f^*(x) \mathbb{E}[g^*(Y) | X = x], \quad (19)$$

$$r^*_Y(y) = p_Y(y) g^*(y) \mathbb{E}[f^*(X) | Y = y]. \quad (20)$$

Then,

$$\frac{1}{a} D(r^*_X || p_X) + \frac{1}{b} D(r^*_Y || p_Y) - D(r^*_{X,Y} || p_{X,Y})$$
$$= \sum_{x,y} r^*_{X,Y}(x, y) \left(\frac{1}{a} \log \frac{r^*_X(x)}{p_X(x)} + \frac{1}{b} \log \frac{r^*_Y(y)}{p_Y(y)} - \log \frac{r^*_{X,Y}(x, y)}{p_{X,Y}(x, y)} \right) \quad (21)$$
$$= \log ||\mathbb{E}[f^*(X)|Y]||_{-\frac{b}{1-b}} ||\mathbb{E}[g^*(Y)|X]||_{-\frac{a}{1-a}}, \quad (22)$$

where the last equality follows from using (16), (17), (18), (19), and (20).

However, from reverse Hölder's inequality, we have

$$1 = \mathbb{E} f^*(X) g^*(Y) \ge ||f^*(X)||_a ||\mathbb{E}[g^*(Y)|X]||_{-\frac{a}{1-a}},$$
$$1 = \mathbb{E} f^*(X) g^*(Y) \ge ||g^*(Y)||_b ||\mathbb{E}[f^*(X)|Y]||_{-\frac{b}{1-b}},$$

and since $||f^*(X)||_a ||g^*(Y)||_b > 1$, the above two inequalities necessitate $||\mathbb{E}[f^*(X)|Y]||_{-\frac{b}{1-b}} ||\mathbb{E}[g^*(Y)|X]||_{-\frac{a}{1-a}} < 1$. This means that the right hand side of (22) must be negative and hence, (iv') is false. Using Lemma 1, (iv) is false too. □

IV. Applications and Concluding Remarks

A. Comparison with the analogous result for hypercontractivity

Nair [15] showed the following equivalence between hypercontractivity and inequalities between information measures:

Theorem 2. *Fix a joint distribution $p_{X,Y}$ so that $p_X, p_Y > 0$. Fix $a, b \ge 1$. The following are equivalent:*

(v) $||\mathbb{E}[g(Y)|X]||_{a'} \le ||g(Y)||_b \quad \forall\, g \ge 0$;

(vi) $\mathbb{E} f(X) g(Y) \le ||f(X)||_a ||g(Y)||_b \quad \forall\, f, g \ge 0$;

(vii) $I(U; X, Y) \ge \frac{1}{a} I(U; X) + \frac{1}{b} I(U; Y)$ $\forall\, U$ *(finite valued)*;

$(viii)$ $D(r_{X,Y} || p_{X,Y}) \ge \frac{1}{a} D(r_X || p_X) + \frac{1}{b} D(r_Y || p_Y)$ $\forall\, r_{X,Y} << p_{X,Y}$,

[2] Indeed, the optimization problem $\sup_{f,g \ge 0, \sum_x f(x) = \sum_y g(y) = 1} \frac{||f(X)||_a ||g(Y)||_b}{\mathbb{E} f(X) g(Y)}$ has the same optimum value as (14) and the supremum is attained here since the objective function is continuous and the space over which we are optimizing is compact. The maximizer for this problem is easily scaled to give a maximizer for (14).

Our proof of Theorem 1 is similar to the proof of Theorem 2 in [15], except for the implication $(viii) \implies (vi)$. By choosing $r_{X,Y}(x,y) = \frac{p_{X,Y}(x,y)f(x)g(y)}{\mathbb{E}f(X)g(Y)}$ in $(viii)$ with $f,g \geq 0, \mathbb{E}f(X)g(Y) > 0$, it is possible to rewrite $(viii)$ as:

$$\begin{aligned}&\log \frac{||f(X)||_a ||g(Y)||_b}{\mathbb{E}f(X)g(Y)}\\ \geq &\frac{1}{a} D\left(\frac{p_X(x)f(x)\mathbb{E}[g(Y)|X=x]}{\mathbb{E}f(X)g(Y)} \middle|\middle| \frac{p_X(x)f(x)^a}{||f(X)||_a^a} \right)\\ &+ \frac{1}{b} D\left(\frac{p_Y(y)g(y)\mathbb{E}[f(X)|Y=y]}{\mathbb{E}f(X)g(Y)} \middle|\middle| \frac{p_Y(y)g(y)^b}{||g(Y)||_b^b} \right),\end{aligned}$$

which implies (vi) from non-negativity of relative entropy. In the case of reverse hypercontractivity, the inequality is reversed and such a deduction is impossible, thus we needed more work to prove $(iv) \implies (ii)$. Indeed, for reverse hypercontractivity, the rewrite as above can be used to show $(ii) \implies (iv)$ instead. This also gives an alternate proof of $(ii) \implies (iv)$ without going through typical sets and large deviation principles.

B. *Non-interactive simulation of joint distributions using Boolean functions*

Let $\{(X_i, Y_i)\}_{i=1}^{\infty}$ be drawn i.i.d. from the doubly symmetric binary source $\mathrm{DSBS}(\epsilon)$. Suppose we want to provide constraints on the space of possible joint distributions that can be created by Boolean functions $b, b' : \{0,1\}^n \to \{0,1\}$ as $(W,Z) = (b(X^n), b'(Y^n))$, for some n. This problem arises for instance, when attacking the following weaker version of a conjecture of Kumar and Courtade [24] that was considered in [25].

Conjecture 1.

$$I(b(X^n); b'(Y^n)) \leq I(X_1; Y_1) = 1 - h(\epsilon), \tag{23}$$

where $h(x) = x \log_2 \frac{1}{x} + (1-x) \log_2 \frac{1}{1-x}$.

[25] considered the following approach to prove Conjecture 1. The strong data processing constant for the joint distribution $p_{W,Z}$ is defined as follows: For any test distribution $r_W(w)$, let $r_Z(z)$ be computed as the output of the channel $p_{Z|W}$ under input $r_W : r_Z(z) = \sum_w r_W(w) p_{Z|W}(z|w)$. Then, define $\mathsf{s}^*(W;Z) := \sup_{r_W \not\equiv p_W} \frac{D(r_Z||p_Z)}{D(r_W||p_W)}$. They observed using the fact that s^* enjoys the properties of tensorization and data processing inequality that

$$\begin{aligned}\mathsf{s}^*(W;Z) &= \mathsf{s}^*(b(X^n); b'(Y^n))\\ &\leq \mathsf{s}^*(X^n; Y^n) \text{ (data processing)}\\ &= \mathsf{s}^*(X_1; Y_1) \text{ (tensorization)}\\ &= (1-2\epsilon)^2 \text{ (from [14])}\end{aligned}$$

They then, observed using numerical computation that for binary valued random variables W, Z, the following seems to hold:

$$\mathsf{s}^*(W;Z) \leq (1-2\epsilon)^2 \implies I(W;Z) \leq 1 - h(\epsilon).$$

While an analytical proof of the above statement is still missing, it provides a formulaic approach to Conjecture 1. However, using the entire hypercontractivity ribbon can provide sharper bounds on the space of distributions, than simply using s^* which is only the limiting chordal slope of hypercontractivity [14].

The hypercontractivity ribbon corresponding to $(X_1, Y_1) \sim \mathrm{DSBS}(\epsilon)$ is given by [5], [23]: $(p,q) \in \mathcal{R}(X_1; Y_1)$ iff

$$\frac{q-1}{p-1} \geq (1-2\epsilon)^2 \ . \tag{24}$$

Using tensorization and data processing, we can show that $(W,Z) = (b(X^n), b'(Y^n))$ must satisfy

$$\left\{ (p,q) : 1 \leq q \leq p, \frac{q-1}{p-1} \geq (1-2\epsilon)^2 \right\} \subseteq \mathcal{R}(W;Z), \tag{25}$$

$$\left\{ (p,q) : 1 \geq q \geq p, \frac{q-1}{p-1} \geq (1-2\epsilon)^2 \right\} \subseteq \mathcal{R}(W;Z). \tag{26}$$

Utilizing the equivalence in Thm. 1 and Thm. 2, we can rewrite the above conditions using information measures. For test distributions r_W, r_Z, abbreviate $D(r_W||p_W), D(r_Z||p_Z)$ and $\inf_{q_{W,Z} \in \mathcal{K}(r_W, r_Z)} D(q_{W,Z}||p_{W,Z})$ by D_1, D_2 and D respectively. Then, denoting $(x)^+ = \max\{x, 0\}$, calculations show that (25), (26) are respectively equivalent to:

$$\sup_{r_W, r_Z} \left(\frac{\sqrt{D_1 D_2} - \sqrt{(D-D_1)(D-D_2)}}{D} \right)^+ \leq |1-2\epsilon|, \tag{27}$$

$$\sup_{r_W, r_Z} \left(\frac{\sqrt{(D-D_1)(D-D_2)} - \sqrt{D_1 D_2}}{D} \right)^+ \leq |1-2\epsilon| \ . \tag{28}$$

A computer search suggests that either of the constraints (27), (28) individually appear to suffice to establish $I(W;Z) \leq 1 - h(\epsilon)$. Indeed, it can be readily checked that (27) is a stronger constraint on (W,Z) than $\mathsf{s}^*(W;Z) \leq (1-2\epsilon)^2$ that was used in [25]. Inequalities (27) and (28) can thus provide a stronger formulaic approach to establish Conjecture 1 analytically.

C. Isoperimetric inequalities

If (X, Y) is (a', b)-reverse hypercontractive as in Theorem 1, then (11) says that for any n and any sets $C_n \subseteq \mathcal{X}^n, D_n \subseteq \mathcal{Y}^n$, the inequality

$$P(X^n \in C_n, Y^n \in D_n) \geq P(X^n \in C_n)^{\frac{1}{a}} P(Y^n \in D_n)^{\frac{1}{b}}. \tag{29}$$

A natural question to ask is: How tight is the above inequality? This was discussed in [9, Sec. 3.2, Sec. 3.4] which considered the doubly symmetric binary source distribution (X, Y). Since the reverse hypercontractivity region for this joint distribution is explicitly known [see (24)], they showed by suitable choices of C_n, D_n that (29) cannot be improved in terms of the exponents on the right hand side, thus establishing (29) (with optimization of exponents with knowledge of marginal probabilities on the right) as an isoperimetric inequality.

Since inequality (29) shows up as an intermediate step in our proof of $(ii) \implies (iii)$ as part of the equivalence proof $(ii) \implies (iii) \implies (iv) \implies (ii)$, it follows that this fact is true for any joint distribution (X, Y) on finite alphabets, i.e. the inequality (29) holds for all n and all sets $C_n \subseteq \mathcal{X}^n, D_n \subseteq \mathcal{Y}^n$ if and only if the exponents on the right hand side are $\frac{1}{a}$ and $\frac{1}{b}$ respectively for some $0 < a, b \leq 1$, and (X, Y) is (a', b)-reverse hypercontractive.

D. Other regimes of reverse hypercontractivity

Is there an equivalent understanding of reverse hypercontractivity using inequalities among information measures for the other regimes of reverse hypercontractivity namely $0 < p, q < 1$ and $p, q < 0$ as shown in Fig. 1, that is, in the first and third quadrants. A positive answer to this question was announced recently [26].

Acknowledgements

I would like to thank Jingbo Liu, Chandra Nair, and Himanshu Tyagi for useful discussions.

References

[1] Aline Bonami, "Ensembles $\Lambda(p)$ dans le dual de D^∞", *Ann. Inst. Fourier*, vol. 18, no. 2, pp. 193–204, 1968.

[2] Aline Bonami, "Étude des coefficients de Fourier des fonctions de $L^p(G)$", *Ann. Inst. Fourier (Grenoble)*, vol. 20, no. 2, pp. 335–402, 1970.

[3] Edward Nelson, "Construction of quantum fields from Markoff fields", *J. Functional Analysis*, vol. 12, pp. 97–112, 1973.

[4] L. Gross, "Logarithmic Sobolev Inequalities", *Amer. J. Math.*, vol. 97, pp. 1061–1083, 1975.

[5] William Beckner, "Inequalities in Fourier analysis", *Ann. of Math.*, vol. 102, no. 1, pp. 159–182, 1975.

[6] J. Kahn, G. Kalai, and N. Linial, "The influence of variables on Boolean functions", in *Proc. of 29th Annual Symposium on Foundations of Computer Science*, 1988.

[7] E. Friedgut, "Boolean functions with low average sensitivity", *Combinatorica*, vol. 18, pp. 27–36, 1998.

[8] E. Mossel, R. O'Donnell, and K. Oleszkiewicz, "Noise stability of functions with low influences: Invariance and Optimality", in *Proceedings of the 46th Annual Symposium on Foundations of Computer Science*, 2005.

[9] 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.

[10] E. Mossel, K. Oleszkiewicz, and A. Sen, "On Reverse Hypercontractivity", *Geometric and Functional Analysis*, vol. 23, no. 3, pp. 1062–1097, 2011.

[11] Y. Polyanskiy, "Hypothesis testing via a comparator and hypercontractivity", http://people.lids.mit.edu/yp/homepage/data/htstruct_journal.pdf, 2012, [Online].

[12] E. Mossel and M. Racz, "A quantitative Gibbard-Satterthwaite theorem without neutrality", in *Proc. of the 44th Annual Symposium on Theory of Computing*, 2012.

[13] V. Guruswami and E. Lee, "Strong Inapproximability Results on Balanced Rainbow-Colorable Hypergraphs Contact", http://eccc.hpi-web.de/report/2014/043/, 2014, [Online].

[14] 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.

[15] C. Nair, "Equivalent formulations of hypercontractivity using information measures", in *IZS seminar*, February 2014.

[16] E.A. Carlen and D. Cordero-Erausquin, "Subadditivity of the entropy and its relation to Brascamp-Lieb type inequalities", *Geometric and Functional Analysis*, vol. 19, no. 2, pp. 373–405, 2009.

[17] C. Nair, "An extremal inequality related to hypercontractivity of Gaussian random variables", in *ITA workshop*, February 2014.

[18] E. Friedgut and V. Rödl, "Proof of a hypercontractive estimate via entropy", *Israel J. Math.*, vol. 125, no. 1, pp. 369–380, 2001.

[19] E. Blais and L. Tan, "Hypercontractivity via the entropy method", *Theory of Computing*, vol. 9, no. 29, pp. 889–896, 2013.

[20] E. Friedgut, "An information-theoretic proof of a hypercontractive inequality", *arXiv:1504.01506 [math.PR]*, Apr. 2015.

[21] Y. Polyanskiy, "Hypothesis testing via a comparator", in *International Symposium on Information Theory*, MIT, Cambridge, USA, July 2012.

[22] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, New York: Academic, 1981.

[23] Christer Borell, "Positivity improving operators and hypercontractivity", *Math. Z.*, vol. 180, no. 2, pp. 225–234, 1982.

[24] G. Kumar and T. Courtade, "Which Boolean functions are most informative?", in *Proc. of IEEE ISIT*, Istanbul, Turkey, 2013.

[25] V. Anantharam, A. Gohari, S. Kamath, and C. Nair, "On hypercontractivity and the mutual information between Boolean functions", in *Proc. of the 51st Annual Allerton Conference on Communications, Control and Computing*, Monticello, Illinois, October 2013.

[26] S. Beigi and C. Nair, "Untitled manuscript", (Personal communication).

[27] B. Kalantari, I. Lari, F. Ricca, and B. Simeone, "On the complexity of general matrix scaling and entropy minimization via the RAS algorithm", *Mathematical Programming, Series A*, vol. 112, pp. 371–401, 2008.

[28] R. Sinkhorn, "A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices", *Ann. Math. Stat.*, vol. 35, no. 2, pp. 876–879, 1964.

Appendix

A. Proof of Lemma 1

Proof. Fix a pair of probability distributions $r_X(x)$ and $r_Y(y)$ and consider the following convex optimization problem:

$$\begin{aligned}
\underset{q_{X,Y}(x,y)}{\text{minimize}} \quad & D(q_{X,Y}||p_{X,Y}) \\
\text{subject to} \quad & q_{X,Y}(x,y) \geq 0 \ \forall x \in \mathcal{X}, y \in \mathcal{Y}, \\
& \sum_y q_{X,Y}(x,y) = r_X(x) \ \forall x \in \mathcal{X}, \\
& \sum_x q_{X,Y}(x,y) = r_Y(y) \ \forall y \in \mathcal{Y}.
\end{aligned} \tag{30}$$

As $p_{X,Y} > 0$, the objective function is finite for all probability distributions $q_{X,Y}$. The feasible set of $q_{X,Y}$ that satisfies stated conditions is compact and non-empty since it contains $q_{X,Y}(x,y) = r_X(x)r_Y(y)$. Thus, from strict convexity of $D(\cdot||p_{X,Y})$, the minimum of the convex problem (30) is finite and attained at a unique optimizer $q^*_{X,Y}(x,y)$.

Suppose for now that $r_X > 0$ and $r_Y > 0$. For suitable values of dual variables $\lambda(x,y), \mu(x), \nu(y) \in [-\infty, \infty]$, this optimizer must satisfy the Karush-Kuhn-Tucker (KKT) conditions:

Stationarity:

$$1 + \log \frac{q^*_{X,Y}(x,y)}{p_{X,Y}(x,y)} = \mu(x) + \nu(y) + \lambda(x,y), \quad \forall x, y,$$

Primal Feasibility:

$$\begin{aligned}
q^*_{X,Y}(x,y) &\geq 0, & \forall x, y, \\
q^*_X(x) &= r_X(x), & \forall x, \\
q^*_Y(y) &= r_Y(y), & \forall y,
\end{aligned}$$

Dual Feasibility:

$$\lambda(x,y) \geq 0, \quad \forall x, y,$$

Complementary Slackness:

$$\lambda(x,y) q^*_{X,Y}(x,y) = 0, \quad \forall x, y.$$

It is easily argued that $q^*_{X,Y}(x,y) > 0 \ \forall x \in \mathcal{X}, y \in \mathcal{Y}$.[3] By complementary slackness, this means $\lambda(x,y) = 0 \ \forall x, y$ and the form of the minimizer is given by:

$$q^*_{X,Y}(x,y) = p_{X,Y}(x,y) e^{\mu(x)-1} e^{\nu(y)} ,$$

where $-\infty < \mu(x), \nu(y) < \infty$ for all $x \in \mathcal{X}, y \in \mathcal{Y}$. Thus, $q^*(x,y)$ is of the form $\frac{p_{X,Y}(x,y)f(x)g(y)}{\mathbb{E}f(X)g(Y)}$ for suitable f, g.

[3]Indeed, if $q^*_{X,Y}(x,y) = 0$ for some (x,y), consider the family of distributions parametrized by λ described by $q^\lambda_{X,Y}(x,y) := (1-\lambda)q^*_{X,Y}(x,y) + \lambda r_X(x) r_Y(y)$ for $\lambda \in [0,1]$. All of these distributions are present in the feasible set of (30) and $q^\lambda_{X,Y} > 0$ for all $\lambda > 0$. Furthermore, the derivative of the function $D(q^\lambda_{X,Y}||p_{X,Y})$ at $\lambda = 0$ is $-\infty$. This means there exists a $\lambda > 0$ where the objective function takes a smaller value at $q^\lambda_{X,Y}$ than at $q^0_{X,Y} = q^*_{X,Y}$, which contradicts the assumption that $q^*_{X,Y}$ is a minimizer of (30).

Now, suppose $r_X(x) = 0$ for $x \in A \subseteq \mathcal{X}$ and $r_Y(y) = 0$ for some $y \in B \subseteq \mathcal{Y}$, where at least one of A or B is non-empty. Then it is clear that for any feasible $q_{X,Y}$ in (30), we will need $q_{X,Y}(x,y) = 0$ if $x \in A$ or if $y \in B$. If $x \notin A$ and $y \notin B$, we can argue again that $q^*_{X,Y}(x,y) > 0$ as before. In this case too, using Lagrange multipliers, we find $q^*_{X,Y}(x,y) = \frac{p_{X,Y}(x,y)f(x)g(y)}{\mathbb{E}f(X)g(Y)}$ for suitable f, g where $f(x) = 0$ for $x \in A$, $g(y) = 0$ for $y \in B$.

All of the above then means $(iv') \implies (iv)$.

On the other hand, for any functions $f, g \geq 0$, neither identically zero, we may fix

$$r_X(x) = \sum_y \frac{p_{X,Y}(x,y)f(x)g(y)}{\mathbb{E}f(X)g(Y)}$$

$$r_Y(y) = \sum_x \frac{p_{X,Y}(x,y)f(x)g(y)}{\mathbb{E}f(X)g(Y)}$$

and consider the convex optimization problem in (30). Then, $q^\dagger_{X,Y}(x,y) := \frac{p_{X,Y}(x,y)f(x)g(y)}{\mathbb{E}f(X)g(Y)}$ is feasible and satisfies all the KKT conditions. Since the constraints are affine and the objective function is convex, any solution to the KKT conditions is necessarily an optimizer, so $q^\dagger_{X,Y} = q^*_{X,Y}$ and we have $(iv) \implies (iv')$. □

Remark 3. To check (iv) of Lemma 1, one must search over the space of pairs of distributions (r_X, r_Y) which has dimension $|\mathcal{X}|+|\mathcal{Y}|-2$. To check (iv') of Lemma 1 one must search over the space of functions $f, g \geq 0$ satisfying $\sum_x f(x) = \sum_y g(y) = 1$, which also has dimension $|\mathcal{X}|+|\mathcal{Y}|-2$. However, checking the former is more difficult computationally since there is no explicit closed form formula for the infimum in (iv) in terms of r_X, r_Y. Computing this infimum is a convex problem and may be solved efficiently using convex solvers. A popular algorithm for this computation in practice is the alternating minimization algorithm also known as the RAS algorithm (see [27] for details about this and other algorithms). However, by choosing the specific form for the r_X and r_Y in (iv'), we get to bypass the computationally intensive step that calculates the infimum. Therefore, with a computational purpose in mind, (iv') is preferable to (iv).

In this paper though, the utility of (iv') arises from the explicit structural form of the optimizer which helps the analysis.

Remark 4. Lemma 1 is similar in spirit to Sinkhorn's theorem [28]. We however, chose to provide a full proof of the lemma in this paper to keep it self-contained.