Equivalent characterization of reverse Brascamp-Lieb-type inequalities using information measures

Salman Beigi*†, Chandra Nair†
*School of Mathematics, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
†Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong

***Abstract*—We derive an equivalent characterization, using information measures, for a class of reverse Brascamp-Lieb type inequalities. These inequalities contain, in particular, the family of reverse hypercontractive inequalities.**

I. Introduction

A. Background

Hypercontractive inequalities have played an important role in many areas such as analysis, probability theory, theoretical computer science, and information theory. The relationship between hypercontractive inequalities, images of a set, and information measures were first explored by Ahlswede and Gacs [1]. Several equivalent characterizations of hypercontractive inequalities using information measures were developed in [2], [3], thus extending the work of Ahlswede and Gacs. A more general class of inequalities called Brascamp-Lieb type inequalities were considered independently by Carlen and Cordero-Erausquin in [4] and one of the equivalent characterizations of hypercontractive inequalities using Kullback-Liebler divergences in [3] can be inferred from the above paper.

Reverse hypercontractive inequalities form another related family of inequalities which have been well studied in analysis, theoretical computer science, and probabilty theory. A straightforward application of the ideas in [3], as well as some personal communication, enabled Kamath [5] to derive an equivalent characterization of a restricted parameter range of reverse hypercontractive inequalities using information measures. In this paper we derive an equivalent characterization of reverse Brascamp-Lieb type inequalities using Kullback-Liebler divergences, thus completing the reverse counterpart of the results in [4]. Our results include, as a special case, all parameters of the reverse hypercontractive inequalities. We establish the equivalent characterizations in finite alphabet spaces omitting their rather straightforward extensions via standard techniques of real analysis to more general spaces.

B. Forward hypercontractive and Brascamp-Lieb-type inequalities

One viewpoint of hypercontractive inequalities is to consider them as strengthening of Hölder's inequality. For a random variable X, positive function f and any $r \neq 0$ we denote

$$\|f(X)\|_r = \mathrm{E}[f(X)^r]^{1/r}.$$

For any pair of random variables (X, Y), Hölder's inequality states that, for all functions $f(X), g(Y)$,

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

for any $\lambda_1, \lambda_2 \in (1, \infty)$ such that $\frac{1}{\lambda_1} + \frac{1}{\lambda_2} = 1$.

A hypercontractive inequality is a stronger form of Hölder's inequality, which states that a pair of random variables (X, Y) is (λ_1, λ_2)-hypercontractive, for $\lambda_1, \lambda_2 \in (1, \infty)$ with $\frac{1}{\lambda_1} + \frac{1}{\lambda_2} \geq 1$, if

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

for all functions $f(X), g(Y)$. By monotonicity of $\|f(X)\|_r$ in r, Hölder's inequality allows one to conclude that every (X, Y) is (λ_1, λ_2)-hypercontractive if $\frac{1}{\lambda_1} + \frac{1}{\lambda_2} \leq 1$. Thus hypercontractive inequalities are stronger than Hölder's inequalities.

Carlen and Cordero-Erausquin considered a further generalization along the lines of Brascamp-Lieb [6] inequalities.

Definition 1. Finite valued random variables $X_1, \ldots, X_m$ jointly distributed according to some probability mass function $\mu(\cdot)$ are said to satisfy the *Brascamp-Lieb type inequality* with parameters $(\lambda_1, ..., \lambda_m), \lambda_i \geq 0, 1 \leq i \leq m$ and C, if for all positive functions $f_1, \ldots, f_m$ we have

$$\mathrm{E}[f_1(X_1)f_2(X_2)\cdots f_m(X_m)] \geq 2^C \prod_{i=1}^m \|f_i(X_i)\|_{\lambda_i}.$$

Carlen and Cordero-Erausquin consider random variables with densities and in this paper we consider random variables with finite alphabets just as in [1] and [3]. In the following theorem we restate the main result of [4] in terms of finite alphabets. Logarithms are considered to base 2 in this paper.

Theorem 1 (Theorem 2.1 in [4]). *Let $X_1, ..., X_m$ be finite valued random variables and let μ denote the joint probability mass function of these variables. Let μ_i, $1 \leq i \leq m$, denote the marginal distribution of X_i. Let $\lambda_1, ..., \lambda_m$ be non-negative numbers. Then for any $C \in \mathbb{R}$ the following two assertions are equivalent:*

(i) *For any non-negative functions $f_1, .., f_m$ we have*

$$\mathrm{E}\left[\prod_{i=1}^m f_i(X_i)\right] \leq 2^C \prod_{i=1}^m \|f_i(X_i)\|_{\lambda_i}.$$

(ii) *For every probability mass function ν of $X_1, ..., X_m$ with marginals ν_i, $1 \leq i \leq m$, we have*

$$\sum_{i=1}^{m} \frac{1}{\lambda_i} D(\nu_i \| \mu_i) \leq C + D(\nu \| \mu).$$

C. Reverse hypercontractive and Brascamp-Lieb type inequalities

Reverse hypercontractive inequalities arise as stronger forms of reverse Hölder's inequality. For any pair of random variables (X, Y), reverse Hölder's inequality states that, for all strictly positive functions $f(X), g(Y)$,

$$\mathrm{E}[f(X)g(Y)] \geq \|f(X)\|_{\lambda_1} \|g(Y)\|_{\lambda_2},$$

for any $\lambda_1, \lambda_2 \in (-\infty, 1) \setminus \{0\}$ such that $\frac{1}{\lambda_1} + \frac{1}{\lambda_2} = 1$.

Definition 2. Finite valued random variables $X_1, \ldots, X_m$ jointly distributed according to some probability mass function $\mu(\cdot)$ are said to satisfy the reverse-Brascamp-Lieb type inequality with parameters $(\lambda_1, ..., \lambda_m)$ and C, if for all positive functions $f_1, \ldots, f_m$ we have

$$\mathrm{E}[f_1(X_1)f_2(X_2)\cdots f_m(X_m)] \geq 2^C \prod_{i=1}^{m} \|f_i(X_i)\|_{\lambda_i}.$$

Setting $m = 2, C = 0$ in the above definition yields the *reverse hypercontractive inequalities*.

II. Main Result

We start this section by proving that the reverse-Brascamp-Lieb type inequalities defined earlier satisfies tensorization.

Lemma 1 (Tensorization)**.** *Let $(X_{11}, \ldots, X_{m1})$ and $(X_{12}, ..., X_{m2})$ be independent and identically distributed tuples of random variables. Then, $(X_{11}, \ldots, X_{m1})$, and then $(X_{12}, ..., X_{m2})$, satisfy the reverse-Brascamp-Lieb type inequality with parameters $(\lambda_1, ..., \lambda_m)$ and C, if and only if the tuple $\big((X_{11}, X_{12}), \ldots, (X_{m1}, X_{m2})\big)$ satisfies the reverse-Brascamp-Lieb type inequality with parameters $(\lambda_1, ..., \lambda_m)$ and $2C$.*

Proof. $(\Rightarrow)$ Denote the probability mass function of $(X_{1j}, \ldots, X_{mj})$, $j = 1, 2$, by μ_j. Let $f_i(X_{i1}, X_{i2})$, $1 \leq i \leq m$, be arbitrary positive functions. Observe that

$$\begin{aligned}
&\mathrm{E}\left[\prod_{i=1}^{m} f_i(X_{i1}, X_{i2})\right] \\
&= \sum_{\mathbf{x}_1} \mu_1(\mathbf{x}_1)\, \mathrm{E}_{\mu_2}\left[\prod_{i=1}^{m} f_i(x_{i1}, X_{i2})\right] \\
&\geq 2^C\, \mathrm{E}_{\mu_1}\left[\prod_{i=1}^{m} \big\| \mathrm{E}[f_i(X_{i1}, X_{i2})|X_{i1}]\big\|_{\lambda_i}\right] \\
&\geq 2^{2C} \prod_{i=1}^{m} \| \mathrm{E}(f_i(X_{i1}, X_{i2}))\|_{\lambda_i},
\end{aligned}$$

where the summation is over $\mathbf{x}_1 = (x_{11}, ..., x_{m1})$, and the inequalities follow from our assumption that μ_1 and μ_2 satisfies the reverse-Brascamp-Lieb type inequality with parameters $(\lambda_1, ..., \lambda_m)$ and C.

$(\Leftarrow)$ Let $g_1(X_{11}), \ldots, g_m(X_{m1})$ be positive functions which violate the reverse-Brascamp-Lieb type inequality with parameters $(\lambda_1, ..., \lambda_m)$ and C. Then the functions $f_i(X_{i1}, X_{i2}) = g_i(X_{i1})g_i(X_{i2})$, $1 \leq i \leq m$, violate the reverse-Brascamp-Lieb type inequality with parameters $(\lambda_1, ..., \lambda_m)$ and $2C$. □

The main result of this paper is Theorem 2 which yields a similar characterization as in Theorem 1, albeit in the reverse direction. The proof mainly uses standard notions in information theory such as typicality, types (empirical distributions) and concentration of measure. As such we will be skimpy on standard parts of the arguments. A reader may consult [7] for a text-book treatment of these standard *type-counting* arguments.

Theorem 2. *Let $X_1, ..., X_m$ be finite valued random variables and let μ denote their joint probability mass function with marginals μ_i, $1 \leq i \leq m$. Let $\lambda_1, ..., \lambda_m$ be non-zero numbers. Let $S_+ = \{i : \lambda_i > 0\}$ be the set containing the indices of the positive λ_i's. Then for any $C \in \mathbb{R}$ the followings are equivalent:*

(i) *For all positive functions $f_1, .., f_m$ we have*

$$\mathrm{E}\left[\prod_{i=1}^{m} f_i(X_i)\right] \geq 2^C \prod_{i=1}^{m} \|f_i(X_i)\|_{\lambda_i}. \tag{1}$$

(ii) *For all probability mass functions ν_i for $i \in S_+$, there exists a probability mass function ν, consistent with the given marginals $\nu_i, i \in S_+$ such that*

$$\sum_{i=1}^{m} \frac{1}{\lambda_i} D(\nu_i \| \mu_i) \geq C + D(\nu \| \mu).$$

For $i \notin S_+$, ν_i is the marginal induced by the p.m.f. ν.

Proof. $(i) \Rightarrow (ii)$ We use the tensorization property of the reverse-Brascamp-Lieb type inequality stated in Lemma 1.

Consider n independent and identically distributed copies of $(X_1, \ldots, X_m)$, i.e., consider random variables $(X_1^n, \ldots, X_m^n)$ with the probability mass function

$$\tilde{\mu}(X_1^n, \ldots, X_m^n) = \prod_{j=1}^{n} \mu(X_{1j}, ..., X_{mj}).$$

Given probability mass functions $\nu_i, i \in S_+$, define the functions

$$f_i(x_{i1}, ..., x_{in}) = \begin{cases} 1 & (x_{i1}, ..., x_{in}) \in \mathcal{T}_{\epsilon_n}^{(n)}(\nu_i) \\ 0 & \text{otherwise.} \end{cases}$$

Here $\mathcal{T}_{\epsilon_n}^{(n)}(\nu_i)$ denotes the set of ϵ_n-typical sequences with respect to the distribution ν_i, where $\{\epsilon_n\}_{n=1}^{\infty}$ is a sequence of positive numbers with $\epsilon_n \to 0$ and $\epsilon_n \sqrt{n} \to \infty$ as n tends to infinity.

Standard calculations (Sanov's theorem [8]) show that

$$\|f_i(X_i^n)\|_{\lambda_i} = 2^{-n \frac{1}{\lambda_i} D(\nu_i \| \mu_i) + o(n)}. \tag{2}$$

Given a tuple x_i^n, let $\pi_i(x_i^n)$ denote the type (empirical distribution induced by the sequence x_i^n). Given a type π_i, denote by

$$\tilde{\mu}_i(\pi_i) = \sum_{x_i^n : \pi_i(x_i^n)=\pi_i} \tilde{\mu}_i(x_i^n) = 2^{-nD(\pi_i\|\mu_i)+o(n)}$$

the total (marginal) probability of all sequences that have type π_i. Then, for every $i \notin S_+$ define

$$f_i(x_{i1}, ..., x_{in}) = \tilde{\mu}_i(\pi_i(x_i^n))^{-\frac{1}{\lambda_i}}.$$

This function assigns the same value to all sequences belonging to a particular type class. Now, elementary calculations along with the fact that there are only a polynomial number of types yield for every $i \notin S_+$,

$$\begin{aligned}\|f_i(X_i^n)\|_{\lambda_i} &= \left(\sum_{\pi_i} \sum_{x_i^n:\pi_i(x_i^n)=\pi_i} \tilde{\mu}(x_i^n)\tilde{\mu}_i(\pi_i)^{-1}\right)^{\frac{1}{\lambda_i}} \\ &= 2^{o(n)}, \end{aligned} \quad (3)$$

where the sum over all sequences is first partitioned into types π_i.

For a tuple $\mathbf{x} = (x_1^n, \ldots, x_m^n)$ denote by $\pi(\mathbf{x})$ its type (which is an empirical probability mass function on the alphabet space of $(X_1, .., X_m)$), and as before, denote by $\pi_i(x_i^n)$ the type on the i-th marginal. Then from the above equations it is easy to verify that

$$f_1(x_1^n) \cdots f_m(x_m^n) = \prod_{i \notin S_+} 2^{n\frac{1}{\lambda_i}D(\pi_i\|\mu_i)+o(n)},$$

if the types $\pi_i(x_i^n)$, for $i \in S_+$, are ϵ_n-close to ν_i, and $f_1(x_1^n) \cdots f_m(x_m^n) = 0$ otherwise. Therefore, we have

$$\begin{aligned}&\mathrm{E}\left[f_1(X_1^n) \cdots f_m(X_m^n)\right] \\ &= \sum_{\pi} \sum_{\mathbf{x}:\pi(\mathbf{x})=\pi} \tilde{\mu}(\mathbf{x}) \prod_{i\notin S_+} 2^{n\frac{1}{\lambda_i}D(\pi_i\|\mu_i)+o(n)}, \\ &= \sum_{\pi} 2^{-n\left(D(\pi\|\mu) - \sum_{i\notin S_+} \frac{1}{\lambda_i}D(\pi_i\|\mu_i)\right)+o(n)}\end{aligned}$$

where the outer sum is over types π for which π_i, for $i \in S_+$ is ϵ_n-close to ν_i. Since the number of types grows polynomially in n, we can further say that there exists a type π^n, with π_i^n, for $i \in S_+$ being ϵ_n-close to ν_i, such that

$$\begin{aligned}&\mathrm{E}[f_1(X_1^n) \cdots f_m(x_m^n)] \\ &= 2^{-n\left(D(\pi^n\|\mu) - \sum_{i\notin S_+} \frac{1}{\lambda_i}D(\pi_i^n\|\mu_i)\right)+o(n)}.\end{aligned}$$

Substituting this in the given inequality in (1) for the above choice of functions, using equations (2) and (3), and letting $n \to \infty$ we conclude that there exists a distribution ν, consistent with the given marginals $\nu_i, i \in S_+$, such that

$$\sum_{i=1}^m \frac{1}{\lambda_i} D(\nu_i\|\mu_i) \geq C + D(\nu\|\mu).$$

$(ii) \Rightarrow (i)$ Let $g_1(X_1), \ldots, g_m(X_m)$ be positive functions which violate the reverse-Brascamp-Lieb type inequality with parameters $(\lambda_1, ..., \lambda_m)$ and C. Thus there exists $\delta > 0$ such that

$$\mathrm{E}\left[g_1(X_1) \cdots g_m(X_m)\right] < 2^{C-\delta} \prod_{i=1}^m \|g_i(X_i)\|_{\lambda_i}.$$

We normalize the functions so that $\|g_i(X_i^n)\|_{\lambda_i} = 1$. We then have

$$\mathrm{E}\left[g_1(X_1) \cdots g_m(X_m)\right] < 2^{C-\delta}.$$

Letting $f_i(X_i^n) = g_i(X_{i1}) \cdots g_i(X_{in})$, $1 \leq i \leq m$ we then find that

$$\mathrm{E}\left[f_1(X_1^n) f_2(X_2^n) \cdots f_m(X_m^n)\right] < 2^{n(C-\delta)}. \quad (4)$$

For every $1 \leq i \leq m$ define

$$\nu_i(x_i) = \mu_i(x_i) g_i(x_i)^{\lambda_i},$$

and

$$\tilde{\nu}_i(x_i^n) = \tilde{\mu}_i(x_i^n) f_i(x_i^n)^{\lambda_i} = \prod_{j=1}^n \nu_i(x_{ij}).$$

Observe that, by the normalization $\|g_i(X_i^n)\|_{\lambda_i} = 1$, both ν_i and $\tilde{\nu}_i$ are probability mass functions. Now by (4) we have

$$\sum_{\mathbf{x}} \tilde{\mu}(\mathbf{x}) \prod_{i=1}^m \left(\frac{\tilde{\nu}_i(x_i^n)}{\tilde{\mu}_i(x_i^n)}\right)^{\frac{1}{\lambda_i}} < 2^{n(C-\delta)},$$

where the summation runs over all tuples $\mathbf{x} = (x_1^n, .., x_m^n)$. Again dividing into types we get

$$\begin{aligned}&\sum_{\pi} \sum_{\mathbf{x}:\pi(\mathbf{x})=\pi} \tilde{\mu}(\mathbf{x}) \prod_{i=1}^m \left(\frac{\tilde{\nu}_i(x_i^n)}{\tilde{\mu}_i(x_i^n)}\right)^{\frac{1}{\lambda_i}} \\ &= \sum_{\pi} 2^{-n\left(D(\pi\|\mu) - \sum_{i=1}^m \frac{1}{\lambda_i}(D(\pi_i\|\mu_i) - D(\pi\|\nu_i))\right)+o(n)} \\ &< 2^{n(C-\delta)}.\end{aligned}$$

Therefore for every type π we have that

$$D(\pi\|\mu) + C - \delta > \sum_{i=1}^m \frac{1}{\lambda_i} \left(D(\pi_i\|\mu_i) - D(\pi_i\|\nu_i)\right).$$

Now consider a π such that its marginals π_i satisfy $\sum_{i\in S_+} D(\pi_i\|\nu_i) = o(1)$. (Such empirical distributions clearly exist). We then have for every such π the inequality

$$D(\pi\|\mu) + \sum_{i\notin S_+} \frac{1}{\lambda_i} D(\pi_i\|\nu_i) + C - \delta > \sum_{i=1}^m \frac{1}{\lambda_i} D(\pi_i\|\mu_i) + o(1),$$

and then

$$D(\pi\|\mu) + C - \delta > \sum_{i=1}^m \frac{1}{\lambda_i} D(\pi_i\|\mu_i) + o(1),$$

since λ_i is negative for $i \notin S_+$. Letting $n \to \infty$ we find that for every distribution $\hat{\nu}$ consistent with marginals $\nu_i, i \in S_+$,

$$D(\hat{\nu}\|\mu) + \sum_{i\notin S_+} \frac{1}{\lambda_i} D(\hat{\nu}_i\|\nu_i) + C - \delta > \sum_{i=1}^k \frac{1}{\lambda_i} D(\pi_i\|\mu_i),$$

which is a clear violation of our assumption in (ii). □

A. Remarks to Theorem 2

(a) $C \leq 0$ is a necessary condition; to see this, take the constant functions.

(b) Let $C = 0$, $\lambda_1 \in (0,1)$ and $\lambda_i < 0$, for $2 \leq i \leq m$ such that $\sum_{i=1}^n \frac{1}{\lambda_i} = 1$. For these parameters, the fact that reverse Brascamp-Lieb type inequalities hold with parameters $(\lambda_1, ..., \lambda_m)$ and $C = 0$ is a consequence of reverse-Hölder's inequality. This fact combined with the monotonicity of norms yield a trivial range of parameters for which reverse Brascamp-Lieb type inequalities hold. It is an easy exercise to see how this can be inferred from the equivalent characterization using divergences. (Hint: think of X_1 as the input and the rest of X_i's, $2 \leq i \leq m$, as the output of a fixed channel. For the given ν_1 consider the induced output distributions and use data-processing inequality for divergences).

(c) Exact computation of the set of parameters $(C, \lambda_1, \dots, \lambda_m)$ for which the reverse Brascamp-Lieb type inequality holds, is a hard problem in general. The equivalent characterizations for the forward hypercontractivity regime has helped to obtain new results [9], [10] in this direction, and to recover some old results. In the same vein, it is hoped that these new characterizations will help in the evaluation of this set of parameters. Towards this computation the following observation may be relevant.
The equivalent chracterization in Theorem 2 states that $X_1, \dots, X_m$ jointly distributed according to some probability mass function μ satisfies a reverse-Brascamp-Lieb type inequality with parameters $(\lambda_1, ..., \lambda_m)$ and C if

$$\min_{\nu_i, \in S_+} \max_{\nu} \sum_{i=1}^m \frac{1}{\lambda_i} D(\nu_i \| \mu_i) - D(\nu \| \mu) - C \geq 0$$

where the outer minimization is over the marginal distributions $\nu_i, i \in S_+$ and the inner maximization is over distributions ν that are consistent with the marginals ν_i. It is easy to verify that if one fixes $\nu_i, i \in S_+$, then

$$\sum_{i=1}^m \frac{1}{\lambda_i} D(\nu_i \| \mu_i) - D(\nu \| \mu) - C$$

is concave in ν, (recall that when $i \notin S_+$, $\lambda_i < 0$). Thus at least the inner maximizers can be computed efficiently using standard convex optimization techniques.

(d) There are two novel aspects to the proof of Theorem 2 as compared to the proof of the equivalent setting in the forward regime, or of its straightforward extension by Kamath to a regime of the reverse hypercontractivity. The idea of using indicator functions of typical sets as a way to relate norms to divergences can be found in [3]; however indicator functions behave poorly when $\lambda_i < 0$. The choice of functions for negative λ_i used in the proof is a key new element of this proof. The second key element is that both directions of implications use tensorization as well as standard concentration ideas. Previously, tensorization was only used in one direction, and the other direction followed by a multiway equivalence that also involved auxiliary random variables and a perturbation argument. The new argument is markedly concise.

We have the following immediate corollary to Theorem 2. Traditional proofs of such convexity results use non-trivial interpolation techniques. (see Riesz-Thorin Theorem).

Corollary 1. *The set of tuples $(C, \frac{1}{\lambda_1}, .., \frac{1}{\lambda_m})$ such that $(X_1, .., X_m)$ satisfies the reverse Brascamp-Lieb type inequalities, with non-negative parameters $(\lambda_1, ..., \lambda_m)$, is convex.*

Proof. Supppose $(X_1, ..., X_m)$ satisfies reverse Brascamp-Lieb type inequalities with parameters $(C, \frac{1}{\lambda_1}, .., \frac{1}{\lambda_m})$ and $(\hat{C}, \frac{1}{\hat{\lambda}_1}, .., \frac{1}{\hat{\lambda}_m})$, where $(\lambda_1, ..., \lambda_m)$ and $(\hat{\lambda}_1, ..., \hat{\lambda}_m)$ are non-negative. Since $S_+ = \{1, .., m\}$ we know that for any set of marginals $\nu_1, ..., \nu_m$, there exists distributions ν and $\hat{\nu}$ consistent with the marginals such that

$$\sum_{i=1}^m \frac{1}{\lambda_i} D(\nu_i \| \mu_i) \geq C + D(\nu \| \mu)$$
$$\sum_{i=1}^m \frac{1}{\hat{\lambda}_i} D(\nu_i \| \mu_i) \geq \hat{C} + D(\hat{\nu} \| \mu).$$

Hence, we have for any $\alpha \in [0, 1]$

$$\begin{aligned}
&\sum_{i=1}^m \left(\alpha \frac{1}{\lambda_i} + (1-\alpha) \frac{1}{\hat{\lambda}_i} \right) D(\nu_i \| \mu_i) \\
&\quad \geq \alpha C + (1-a)\hat{C} + \alpha D(\nu \| \mu) + (1-\alpha) D(\hat{\nu} \| \mu) \\
&\quad \geq \alpha C + (1-a)\hat{C} + D\left((\alpha \nu + (1-\alpha)\hat{\nu}) \| \mu \right).
\end{aligned}$$

The last inequality follows from the convexity of relative entropy $D(\nu \| \mu)$ with respevct to ν. □

III. Conclusion

We present an equivalent characterization of reverse Brascamp-Lieb type inequalities using divergence measures. The ideas used in this proof are elementary concentration ideas concerning typical sets and types.

Acknowledgment

Chandra Nair would like to thank Elchanal Mossel and Thomas Holenstein who prompted the author to think about reverse-hypercontractive-inequalities. He would also like to thank Sudeep Kamath who initially mentioned the possibility of extending the result in [3] to the reverse direction; and Prof. Venkat Anantharam who mentioned that we should write up the results for reverse Brascamp-Lieb type inequalities rather than just reverse hypercontractive inequalities.

He would also like to thank the Simon's institute for the theory of Computing at UC Berkeley, as well as 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] 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.

[2] 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.

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

[4] 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. [Online]. Available: http://dx.doi.org/10.1007/s00039-009-0001-y

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

[6] H. J. Brascamp and E. H. Lieb, "Best constants in young's inequality, its converse, and its generalization to more than three functions," *Advances in Mathematics*, vol. 20, no. 2, pp. 151 – 173, 1976. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0001870876901845

[7] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Orlando, FL, USA: Academic Press, Inc., 1982.

[8] I. Sanov, *On the probability of large deviations of random variables*, 1958.

[9] S. Kamath and C. Nair, "The strong data processing constant for sums of i.i.d. random variables," in *Information Theory (ISIT), 2015 IEEE International Symposium on*, June 2015, pp. 2550–2552.

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