Non-interactive Simulation of Joint Distributions: The Hirschfeld-Gebelein-Rényi Maximal Correlation and the Hypercontractivity Ribbon

Sudeep Kamath and Venkat Anantharam
EECS Department
University of California, Berkeley, CA, USA
sudeep,ananth@eecs.berkeley.edu

Abstract— We consider the following problem: Alice and Bob observe sequences X^n and Y^n respectively where $\{(X_i, Y_i)\}_{i=1}^{\infty}$ are drawn i.i.d. from $P(x, y)$, and they output U and V respectively which is required to have a joint law that is close in total variation to a specified $Q(u, v)$. One important technique to establish impossibility results for this problem is the Hirschfeld-Gebelein-Rényi maximal correlation which was considered by Witsenhausen [1]. Hypercontractivity studied by Ahlswede and Gács [2] and reverse hypercontractivity recently studied by Mossel et al. [3] provide another approach for proving impossibility results. We consider the tightest impossibility results that can be obtained using hypercontractivity and reverse hypercontractivity and provide a necessary and sufficient condition on the source distribution $P(x, y)$ for when this approach subsumes the maximal correlation approach. We show that the binary pair source distribution with symmetric noise satisfies this condition.

I. Introduction

We consider the problem of simulation of one sample of a joint distribution by physically separated non-interacting agents observing i.i.d. copies of correlated random variables. Related problems have been well-studied in the literature. Wyner [4] studied the problem of simulating a joint distribution from shared randomness while Gács and Körner [5] studied the problem of extracting common randomness from correlated observations. Cuff studied communication requirements for simulating a channel [6]. Gohari and Anantharam generalized Cuff's formulation in [7] and Yassaee, Gohari and Aref recently solved this problem in [8]. Cuff, Permuter, Cover studied communication requirements for establishing dependence among nodes in a network setting [9].

Non-Interactive Correlation Distillation, a setup in which non-interacting agents have to each output a uniform random bit which agree with high probability, has been studied in [1], [10], [11]. In this paper, we propose a generalization of this problem.

Rows a) and b) of Table I summarize two different formulations of the problem of simulation of joint distributions while Row c) describes our formulation as a motivation to provide a converse to the formulation in Row b).

Definition: Let $\mathcal{X}, \mathcal{Y}, \mathcal{U}, \mathcal{V}$ denote finite sets. Given a *source distribution* $P(x, y)$ over $\mathcal{X} \times \mathcal{Y}$ and a *target distribution* $Q(u, v)$ over $\mathcal{U} \times \mathcal{V}$, we say that *non-interactive simulation* of $Q(u, v)$ using $P(x, y)$ is possible, if for any $\epsilon > 0$, there exists a positive integer n and functions $f : \mathcal{X}^n \mapsto \mathcal{U}, g : \mathcal{Y}^n \mapsto \mathcal{V}$ such that

$$d_{\mathrm{TV}}\left((f(X^n), g(Y^n)), (U, V)\right) \leq \epsilon$$

where $\{(X_i, Y_i)\}_{i=1}^{n}$ is a sequence of i.i.d. samples drawn from $P(x, y)$, (U, V) is drawn from $Q(u, v)$ and $d_{\mathrm{TV}}(\cdot, \cdot)$ is the total variation distance.

For a fixed $P(x, y)$, the set of distributions $Q(u, v)$ on $\mathcal{U} \times \mathcal{V}$ for which non-interactive simulation is possible can be shown to be the closure of the set of marginal distributions of (U, V) satisfying $U - X^k - Y^k - V$ for some k. However, this set of distributions seems to be very hard to characterize. The current paper focuses on outer bounds on this set or in other words, impossibility results for non-interactive simulation.

Note that the simulation problem specified in the above definition does not have any more generality if we allow the agents to use their own private randomness: Agents can obtain as much private randomness as desired by using extended observations that are non-overlapping in time, i.e. the agents observe $n_1 + n_2 + n_3$ symbols, they use $X_1, \ldots, X_{n_1}$ as their correlated observations, one agent uses $X_{n_1+1}, \ldots, X_{n_2}$ as her private randomness and the other agent uses $X_{n_2+1}, \ldots, X_{n_3}$ as his private randomness.

Note also that the notion of *simulation* we consider is distinct from the notion of *exact generation*. If we have a strategic setting, such as a distributed game, in which

Private Randomness $X^n \rightarrow$ Alice $\rightarrow U^n$ Common Randomness at fixed rate multiple rounds of rate-limited communication $Y^n \rightarrow$ Bob $\rightarrow V^n$ Private Randomness	**a)** This formulation was proposed by Gohari and Anantharam [7] as a generalization of Cuff's formulation [6]. Yassaee, Gohari, Aref [8] recently solved this problem completely. The task is for two agents to simulate i.i.d. samples of a specified joint distribution $P(x, y, u, v)$. Nature supplies i.i.d. copies of (X, Y) with the right marginal distribution as shown and the agents can use a certain rate of common randomness and certain rate-limited communication and an infinite stream of their own private randomness to accomplish the desired task.
Private Randomness $X^n \rightarrow$ Alice $\rightarrow U^{nR}$ $Y^n \rightarrow$ Bob $\rightarrow V^{nR}$ Private Randomness	**b)** In this formulation, two agents having access to their own infinite stream of private randomness observe n i.i.d. copies of samples generated according to a specified law $P(x, y)$ as shown and are required to output nR samples drawn from a distribution that is close (in total variation) to the the distribution constructed by taking i.i.d. copies of a specified law $Q(u, v)$. Let R^* be the supremum of all achievable rates. • When $(U, V) \sim Q(u, v)$ has $U = V \sim \mathrm{Ber}(1/2)$, we have $R^* = K(X; Y)$, the Gács-Körner common information [5] of X and Y. • When $(X, Y) \sim P(x, y)$ has $X = Y \sim \mathrm{Ber}(1/2)$, we have $\frac{1}{R^*} = C(U; V)$, the Wyner common information [4] of U and V. The problem of characterizing R^* is open for general distributions $P(x, y), Q(u, v)$ and indeed, so is the problem of characterizing when $R^* > 0$.
Private Randomness $X^n \rightarrow$ Alice $\rightarrow U$ $Y^n \rightarrow$ Bob $\rightarrow V$ Private Randomness	**c)** Since the problem of characterizing when $R^* > 0$ in formulation b) is also non-trivial, we propose a relaxed problem where two agents observe an arbitrary but finite number of samples drawn i.i.d. from $P(x, y)$ as shown and are required to output one random variable each with the requirement that the output distribution be close in total variation to a specified $Q(u, v)$. Clearly, if it is impossible to generate even a single sample, we obtain $R^* = 0$. We therefore, focus on impossibility results for this problem which will be relevant to formulation b) above. It is not clear if the converse is true, i.e. it is unclear whether the possibility of generating one sample implies that we may generate samples at a positive rate $R > 0$. When $(U, V) \sim Q(u, v)$ has $U = V \sim \mathrm{Ber}(1/2)$, the problem has recently come to be called Non-Interactive Correlation Distillation [10]. We therefore, call our formulation the problem of Non-Interactive Simulation of Joint Distributions.

TABLE I

DIFFERENT FORMULATIONS OF THE JOINT DISTRIBUTION SIMULATION PROBLEM

a player, represented by a number of distributed agents, is playing against an adversary, the agents would often need to generate a joint distribution exactly [12].

We will consider two examples to motivate the focus of this study.

A. Example 1

Let X be a uniform Bernoulli random variable, $X \sim \mathrm{Ber}(\frac{1}{2})$. Let Y be a noisy copy of X, i.e. $Y = X + N$ where $N \sim \mathrm{Ber}(\alpha)$ for $0 < \alpha < \frac{1}{2}$, is independent of X. We say that (X, Y) has the Doubly Symmetric Binary Source distribution with parameter α denoted $\mathrm{DSBS}(\alpha)$ following the notation of Wyner [4]. We consider $(U, V) \sim \mathrm{DSBS}(\beta)$ for $0 \leq \beta < \frac{1}{2}$. We ask whether non-interactive simulation of $\mathrm{DSBS}(0)$ using $\mathrm{DSBS}(\alpha)$ is possible. Witsenhausen answered this question in the negative in [1], thus significantly strengthening the result of Gács and Körner [5]. Witsenhausen established this by proving the tensorization of the Hirschfeld-Gebelein-Rényi maximal correlation, henceforth simply called the maximal correlation (both tensorization and maximal correlation are defined and discussed in Section II-A). Witsenhausen's approach easily allows us to conclude that if non-interactive simulation is possible, then the maximal correlation of the target distribution can be no more than that of the source distribution. The maxi-

mal correlation of a pair of binary random variables distributed as DSBS(α) equals $|1 - 2\alpha|$. Thus, for instance, if the non-interactive simulation of DSBS(β) using DSBS(α) is possible, with $0 \leq \alpha, \beta \leq \frac{1}{2}$, then we must have $\alpha \leq \beta$. It is easy to see in this case that if $\alpha \leq \beta$, then non-interactive simulation is indeed possible: one agent outputs the first bit of her observation while the other agent outputs a suitable noisy copy of his first bit, the noise realization created from his other $n-1$ observations. Thus, for $0 \leq \alpha, \beta \leq \frac{1}{2}$, non-interactive simulation of DSBS(β) using DSBS(α) is possible if and only if $\alpha \leq \beta$.

B. Example 2

Let $(X, Y) \sim$ DSBS(α) with $0 < \alpha < \frac{1}{2}$. Consider binary random variables (U, V) distributed as $Q(u, v)$ given by: $Q(0, 0) = 0, Q(0, 1) = Q(1, 0) = Q(1, 1) = \frac{1}{3}$. We ask if non-interactive simulation of $Q(u, v)$ using DSBS(α) is possible. The maximal correlation of a DSBS(α) source distribution is $1 - 2\alpha$ while that of $Q(u, v)$ is $\frac{1}{2}$. The approach of comparing maximal correlations of the source and target informs us that the inequality $1 - 2\alpha \leq \frac{1}{2}$, if violated, makes non-interactive simulation impossible. Thus, if $\frac{1}{4} < \alpha < \frac{1}{2}$, then non-interactive simulation is impossible. But what about the case when $0 < \alpha \leq \frac{1}{4}$? Can we come up with a suitable scheme to simulate $Q(u, v)$? The answer turns out to be no for each $0 < \alpha \leq \frac{1}{4}$ and can be proved using the so-called reverse hypercontractive inequalities [3]. The following inequality holds for $\{(X_i, Y_i)\}_{i=1}^{\infty}$ being i.i.d DSBS(α), and for arbitrary sets $S, T \subseteq \{0, 1\}^n$:

$$\Pr\left(X^n \in S, Y^n \in T\right) \geq \Pr\left(X^n \in S\right)^{\frac{1}{2\alpha}} \Pr\left(Y^n \in T\right)^{\frac{1}{2\alpha}}. \tag{1}$$

If non-interactive simulation of $Q(u, v)$ using DSBS(α) were possible, we should be able to find sets S, T such that $\Pr\left(X^n \in S\right) \approx \frac{1}{3}, \Pr\left(Y^n \in T\right) \approx \frac{1}{3}$ and $\Pr\left(X^n \in S, Y^n \in T\right) \approx 0$. Inequality (1) rules out this possibility. Thus, hypercontractivity or reverse hypercontractivity can provide impossibility results when the maximal correlation approach cannot. Is it true that one is always stronger than the other? We show indeed that the approach using hypercontractivity and reverse hypercontractivity subsumes the maximal correlation approach for the case when $P(x, y)$ is of the form DSBS(α). More generally, we give necessary and sufficient conditions on $P(x, y)$ for this subsumption. This arises from an inequality obtained by Ahlswede and Gács [2] in the hypercontractive case which we extend to the reverse hypercontractive case.

The rest of the paper is organized as follows. Section II discusses preliminaries on maximal correlation, hypercontractivity and reverse hypercontractivity. We present our main results in Section III. Section IV contains all the proofs. Finally, Section V discusses an interesting example of non-interactive simulation of a joint distribution of three random variables.

II. MAXIMAL CORRELATION AND THE HYPERCONTRACTIVITY RIBBON

In this paper, all sets are finite and all probability distributions are discrete and have finite support. For a finite set $\mathcal{X}$, let $\mathcal{F}_{\mathcal{X}}, \mathcal{F}_{\mathcal{X}}^{+}$ denote the set of all functions from $\mathcal{X}$ to $\mathbb{R}$ and to $\mathbb{R}_{\geq 0}$ respectively. We will also assume without loss of generality throughout the rest of the paper that the marginals of $P(x, y)$ and $Q(u, v)$ (denoted P_X, P_Y and Q_U, Q_V respectively) assign zero probability only to the null set.

A. Maximal Correlation and its properties

For jointly distributed random variables (X, Y), define their *maximal correlation* $\rho(X; Y) := \sup \mathbb{E} f(X) g(Y)$ where the supremum is over $f : \mathcal{X} \mapsto \mathbb{R}, g : \mathcal{Y} \mapsto \mathbb{R}$ such that $\mathbb{E} f(X) = \mathbb{E} g(Y) = 0$ and $\mathbb{E} \left(f(X)\right)^2 = \mathbb{E} \left(g(Y)\right)^2 = 1$ and with the convention that the supremum over the empty set evaluates to 0.

The following theorem was proved by Witsenhausen in [1]. Kumar has obtained simpler proofs of the same result [13], [14].

Theorem 2.1: (Witsenhausen [1]) If $(X_1, Y_1), (X_2, Y_2)$ are independent, then $\rho(X_1, X_2; Y_1, Y_2) = \max\{\rho(X_1; Y_1), \rho(X_2; Y_2)\}$. If $(X_1, Y_1), (X_2, Y_2)$ are i.i.d., then $\rho(X_1, X_2; Y_1, Y_2) = \rho(X_1; Y_1)$.

We mention here that Kumar [14] has obtained a natural modification of the maximal correlation that he has called n-ary Rényi correlation, for each $n \geq 2$. He shows that these quantities also tensorize and can be used for proving impossibility results for non-interactive simulation. We don't discuss these quantities in the current paper.

The following monotonicity lemma is immediate.

Lemma 2.2: If $\phi(X) = U, \psi(Y) = V$, then $\rho(X; Y) \geq \rho(U; V)$.

The following properties hold for the maximal correlation of two discrete valued random variables with finite support [15].

1) If $(X, Y) \sim$ DSBS(α), then $\rho(X; Y) = |1 - 2\alpha|$.
2) $\rho(X; Y) = 0$ if and only if X is independent of Y.
3) $\rho(X; Y) = 1$ if and only if the Gács-Körner common information $K(X; Y) > 0$, i.e. if and only if (X, Y) is *decomposable*.

Putting together Theorem 2.1 and Lemma 2.2 and using continuity of the maximal correlation in the joint distribution $Q(u, v)$ (continuity requires Q_U, Q_V assign zero probability only to the null set), we get the following corollary.

Corollary 2.3: Non-interactive simulation of $(U, V) \sim Q(u, v)$ using $(X, Y) \sim P(x, y)$ is possible only if $\rho(X; Y) \geq \rho(U; V)$.

B. Hypercontractivity ribbon and its properties

Definition: For any random variable W and real number $p \neq 0$, define $||W||_p := (\mathbb{E}|W|^p)^{1/p}$. Define $||W||_0 := \exp(\mathbb{E}\log|W|)$. For $p \leq 0$, $||W||_p = 0$ if $\Pr(|W| = 0) > 0$.

$||W||_p$ is continuous and non-decreasing in p. If W is non-constant, then $||W||_p$ is strictly increasing for $p \geq 0$. If in addition, $\Pr(|W| = 0) = 0$, then $||W||_p$ is strictly increasing for all p.

Definition: For a pair of random variables $(X, Y) \sim P(x, y)$ on $\mathcal{X} \times \mathcal{Y}$, define the operator $T_{X;Y} : \mathcal{F}_{\mathcal{Y}} \mapsto \mathcal{F}_{\mathcal{X}}$ as

$$(T_{X;Y}f)(x) := \mathbb{E}[f(Y)|X = x]. \tag{2}$$

Likewise, define $T_{Y;X} : \mathcal{F}_{\mathcal{X}} \mapsto \mathcal{F}_{\mathcal{Y}}$ as

$$(T_{Y;X}g)(y) := \mathbb{E}[g(X)|Y = y]. \tag{3}$$

Definition: For a pair of random variables $(X, Y) \sim P(x, y)$ on $\mathcal{X} \times \mathcal{Y}$, we define the *hypercontractivity ribbon*

$$\mathcal{R}_{X;Y} \subseteq \{(p, q) : 1 \leq q \leq p \text{ or } 1 \geq q \geq p\}$$

as follows:

- For $1 \leq q \leq p$, we have $(p, q) \in \mathbb{R}_{X;Y}$ if

$$||T_{X;Y}f(X)||_p \leq ||f(Y)||_q \quad \forall f \in \mathcal{F}_{\mathcal{Y}}; \tag{4}$$

- For $1 \geq q \geq p$, we have $(p, q) \in \mathbb{R}_{X;Y}$ if

$$||T_{X;Y}f(X)||_p \geq ||f(Y)||_q \quad \forall f \in \mathcal{F}_{\mathcal{Y}}^+. \tag{5}$$

Likewise, we can define $\mathcal{R}_{Y;X}$. These are both regions in $\mathbb{R}^2$ pinching to a point at $(1, 1)$ resembling a ribbon, explaining our choice of the name (see Fig. 1). $\mathcal{R}_{X;Y}$ and $\mathcal{R}_{Y;X}$ are intimately connected by a duality relationship which we will discuss later. $T_{X;Y}$ is contractive in the p-norm when $p \geq 1$ and inequality (4) is a hypercontractive inequality since $q \leq p$. $T_{X;Y}$ is reverse contractive for non-negative valued functions f under the p-pseudo-norm when $p \leq 1$, (the triangle inequality is violated) and inequality (5) is called a reverse hypercontractive inequality and has been studied in [3].

Definition: For any real $p \neq 0, 1$, define its *Hölder conjugate* p' by $\frac{1}{p} + \frac{1}{p'} = 1$. For $p = 0$, define $p' = 0$.

Remark: An equivalent definition of $\mathcal{R}_{X;Y}$ which does not use the definition of the operator $T_{X;Y}$ can be provided by observing how much the corresponding Hölder's and reverse Hölder's inequalities may be tightened.

- For $1 \leq q < p$, we have $(p, q) \in \mathbb{R}_{X;Y}$ iff

$$\mathbb{E}f(X)g(Y) \leq ||f(X)||_{p'}||g(Y)||_q$$
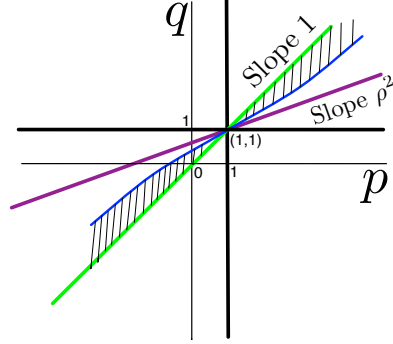$$\forall f \in \mathcal{F}_{\mathcal{X}}, g \in \mathcal{F}_{\mathcal{Y}}; \tag{6}$$

Fig. 1. The hypercontractivity ribbon $\mathcal{R}_{X;Y}$ is the shaded region. Also shown a straight line of slope $\rho^2 := \rho^2(X; Y)$ through $(1, 1)$.

- For $1 \geq q > p$, we have $(p, q) \in \mathbb{R}_{X;Y}$ iff

$$\mathbb{E}f(X)g(Y) \geq ||f(X)||_{p'}||g(Y)||_q$$
$$\forall f \in \mathcal{F}_{\mathcal{X}}^+, g \in \mathcal{F}_{\mathcal{Y}}^+; \tag{7}$$

- $(1, 1) \in \mathcal{R}_{X;Y}$.

To see the equivalence, observe that for $p > 1$, if (4) holds, then by Hölder's inequality, we get

$$\mathbb{E}f(X)g(Y) = \mathbb{E}f(X)(T_{X;Y}g)(X) \tag{8}$$
$$\leq ||f(X)||_{p'}||(T_{X;Y}g)(X)||_p \tag{9}$$
$$\leq ||f(X)||_{p'}||g(Y)||_q. \tag{10}$$

Conversely, if the inequality in (4) fails for some non-negative f, say $f = h$, then by choosing the function $e(X) = (T_{X;Y}h(X))^{p/p'}$, we have equality in Hölder's inequality as follows:

$$\mathbb{E}e(X)h(Y) = \mathbb{E}e(X)(T_{X;Y}h)(X) \tag{11}$$
$$= ||e(X)||_{p'}||(T_{X;Y}h)(X)||_p \tag{12}$$
$$> ||e(X)||_{p'}||h(Y)||_q, \tag{13}$$

since $||e(X)||_{p'} > 0$, thus producing the desired contradiction to (6). It suffices to consider non-negative f, since $-|f| \leq f \leq |f|$ holds pointwise and so $|T_{X;Y}f| \leq T_{X;Y}|f|$ holds pointwise so that if (4) fails for some f then it also fails for $|f|$. A similar equivalence can be observed for $p < 1$, using the reverse Hölder's inequality:

$$\mathbb{E}[WZ] \geq ||W||_{p'}||Z||_p, \tag{14}$$

which holds when $p < 1$ and W, Z are non-negative random variables. The contradiction is first observed for strictly positive functions with $p/p' := -1$ in the case $p = 0$ and then for non-negative functions by taking limits.

$\mathcal{R}_{X;Y}$ is closed and connected in $\mathbb{R}^2$. Moreover, $\{(p, q) : p = q\} \subseteq \mathcal{R}_{X;Y}$. So, $\mathcal{R}_{X;Y}$ is completely characterized by its other boundary, a continuous non-decreasing function $q_{X;Y}^* : \mathbb{R} \mapsto \mathbb{R}$ such that

- $q^*_{X;Y}(p) \leq p$ whenever $p \geq 1$, and $q^*_{X;Y}(p) \geq p$ whenever $p \leq 1$, so $q^*_{X;Y}(1) = 1$;
- $\mathcal{R}_{X;Y} = \{(p,q) : 1 \leq q^*_{X;Y}(p) \leq q \leq p\} \cup \{(p,q) : 1 \geq q^*_{X;Y}(p) \geq q \geq p\}$.

Hypercontractive inequalities and reverse hypercontractive inequalities tensorize [3].

Theorem 2.4: Suppose $(p,q) \in \mathcal{R}_{X_1;Y_1}$ and $(p,q) \in \mathcal{R}_{X_2;Y_2}$. If $(X_1,Y_1), (X_2,Y_2)$ are independent, then $(p,q) \in \mathcal{R}_{X_1,X_2;Y_1Y_2}$, so that $\mathcal{R}_{(X_1,X_2);(Y_1,Y_2)} = \mathcal{R}_{X_1;Y_1} \cap \mathcal{R}_{X_2;Y_2}$. If $(X_1,Y_1),(X_2,Y_2)$ are i.i.d., then $\mathcal{R}_{(X_1,X_2);(Y_1,Y_2)} = \mathcal{R}_{X_1;Y_1}$.

The following lemma provides a monotonicity property for the hypercontractivity ribbon [3].

Lemma 2.5: If $\phi(X) = U, \psi(Y) = V$, then $\mathcal{R}_{X;Y} \subseteq \mathcal{R}_{U;V}$.

Putting together Theorem 2.4 and Lemma 2.5 and using continuity of $q^*_{U;V}(p)$ for each p in the distribution of (U,V) (continuity requires Q_U, Q_V assign zero probability only to the null set), we get the following corollary.

Corollary 2.6: Non-interactive simulation of $(U,V) \sim Q(u,v)$ using $(X,Y) \sim P(x,y)$ is possible only if $\mathcal{R}_{X;Y} \subseteq \mathcal{R}_{U;V}$.

The following properties hold for the hypercontractivity ribbon for two discrete valued random variables with finite support [3].

1) If $(X,Y) \sim \text{DSBS}(\alpha)$, then
   $q^*_{X;Y}(p) - 1 = (1-2\alpha)^2(p-1)$ [10].
2) $q^*_{X;Y}(p) \equiv 1$ if and only if X and Y are independent, i.e. $I(X;Y) = 0$.
3) $q^*_{X;Y}(p) \equiv p$ if and only if $P(x,y)$ is decomposable, i.e. the Gács-Körner common information $K(X;Y) > 0$.
4) If $K(X;Y) = 0$ but $I(X;Y) > 0$, then for $p > 1$, we have the strict inequalities $1 < q^*_{X;Y}(p) < p$ [2].

C. Proving impossibility results for non-interactive simulation using the hypercontractivity ribbon $\mathcal{R}_{X;Y}$

While Corollary 2.6 describes the technique for proving impossibility results, it is worthwhile noting that this is equivalent to the techniques that were originally used to produce inequalities like (1).

Suppose that non-interactive simulation of $Q(u,v)$ using $P(x,y)$ is possible, i.e. suppose for any $\epsilon > 0$, there exists n and functions $\phi : \mathcal{X}^n \mapsto \mathcal{U}, \psi : \mathcal{Y}^n \mapsto \mathcal{V}$ so that $\phi(X^n) = \tilde{U}, \psi(Y^n) = \tilde{V}$ produces $(\tilde{U}, \tilde{V})$ satisfying $d_{\text{TV}}((\tilde{U}, \tilde{V}); (U,V)) \leq \epsilon$ when $(U,V) \sim Q(u,v)$ and $\{(X_i,Y_i)\}_{i=1}^n$ are generated i.i.d. from $P(x,y)$. Choose

$$f(x^n) = \sum_{u \in \mathcal{U}} \lambda_u \mathbb{1}_{[\phi(x^n)=u]}, \tag{15}$$

$$g(y^n) = \sum_{v \in \mathcal{V}} \mu_v \mathbb{1}_{[\psi(y^n)=v]}. \tag{16}$$

For $(p,q) \in \mathcal{R}_{X;Y}$, with $p > 1$, using (6), we obtain upon taking the limit as $\epsilon \to 0$,

$$\sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} \lambda_u \mu_v Q(u,v)$$
$$\leq \left(\sum_{u \in \mathcal{U}} \lambda_u^{p'} Q_U(u) \right)^{1/p'} \cdot \left(\sum_{v \in \mathcal{V}} \mu_v^q Q_V(v) \right)^{1/q}. \tag{17}$$

For $(p,q) \in \mathcal{R}_{X;Y}$, with $p < 1$, using (7), we obtain for non-negative $\{\lambda_u\}_{u \in \mathcal{U}}, \{\mu_v\}_{v \in \mathcal{V}}$, the inequality

$$\sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} \lambda_u \mu_v Q(u,v)$$
$$\geq \left(\sum_{u \in \mathcal{U}} \lambda_u^{p'} Q_U(u) \right)^{1/p'} \cdot \left(\sum_{v \in \mathcal{V}} \mu_v^q Q_V(v) \right)^{1/q}. \tag{18}$$

Indeed, (1) is a version of (18) with $(p,q) \in \mathcal{R}_{X;Y}$ for $(X,Y) \in \text{DSBS}(\alpha)$ given by $p = -\frac{2\alpha}{1-2\alpha}, q = 2\alpha$.

The inclusion $\mathcal{R}_{X;Y} \subseteq \mathcal{R}_{U;V}$ implies the collection of inequalities (17) for any choice of real $\{\lambda_u\}_{u \in \mathcal{U}}, \{\mu_v\}_{v \in \mathcal{V}}$ and the collection of inequalities (18) for any choice of non-negative $\{\lambda_u\}_{u \in \mathcal{U}}, \{\mu_v\}_{v \in \mathcal{V}}$. By an argument similar to the one proving equivalence of the two definitions of $\mathcal{R}_{X;Y}$, one can prove the reverse implication from the collection of inequalities (17), (18) to $\mathcal{R}_{X;Y} \subseteq \mathcal{R}_{U;V}$.

III. MAIN RESULTS

Theorem 3.1:

$$\rho(X;Y) \leq \inf_{(p,q) \in \mathcal{R}_{X;Y}, p \neq 1} \sqrt{\frac{q-1}{p-1}} = \inf_{p \neq 1} \sqrt{\frac{q^*_{X;Y}(p) - 1}{p-1}}. \tag{19}$$

Theorem 3.1 is obtained in [2] for the case of hypercontractive inequalities. We provide an alternate proof of the same result and derive it for the reverse hypercontractive inequalities. In the current form of the statement of Theorem 3.1, the maximal correlation is afforded a geometric meaning, namely its square is the slope of a straight line bound constraining the hypercontractivity ribbon (see Fig 1). Indeed, for $(X,Y) \sim \text{DSBS}(\alpha)$, the hypercontractivity ribbon is precisely the wedge obtained by the straight lines $p = q$, and $q - 1 = \rho(X;Y)^2(p-1)$ [10].

Theorem 3.2: The following are equivalent:

- For all (U,V), we have $\mathcal{R}_{X;Y} \subseteq \mathcal{R}_{U;V} \implies \rho(X;Y) \geq \rho(U;V)$.

-
$$\rho(X;Y) = \inf_{(p,q) \in \mathcal{R}_{X;Y}, p \neq 1} \sqrt{\frac{q-1}{p-1}}. \tag{20}$$

Theorem 3.2 states that Corollary 2.6 subsumes Corollary 2.3 for all $Q(u,v)$ if and only if (19) holds with equality.

Ahlswede and Gács [2] show that $\lim_{p \to \infty} \frac{q^*_{X;Y}(p)}{p}$ exists and equals a quantity $s^*(X;Y)$, defined as follows:

Consider finite sets $\mathcal{X}, \mathcal{Y}$ and let $P(x,y)$ be a joint distribution over the product $\mathcal{X} \times \mathcal{Y}$. Let $R(x)$ be an arbitrary probability distribution on $\mathcal{X}$. Let $\sum_X P_{Y|X} * R$ denote the probability distribution on $\mathcal{Y}$ whose probability mass at y is $\sum_{x \in \mathcal{X}} \frac{P(x,y)}{P_X(x)} R(x)$. If $(X,Y) \sim P(x,y)$, then we define $s^*(X;Y) = \sup_{R:R \neq P_X} \frac{D(\sum_X P_{Y|X} * R || P_Y)}{D(R || P_X)}$. Erkip and Cover consider a related quantity and show that it equals $\rho^2(X;Y)$ [16].

We prove the same result, also extending it to reverse hypercontractive inequalities by a simpler approach.

Theorem 3.3:
$$\lim_{p \to 1} \frac{q^*_{X;Y}(p) - 1}{p - 1} = s^*(Y;X). \tag{21}$$

Corollary 3.4 follows from Theorem 3.3 upon using a duality result connecting $\mathcal{R}_{X;Y}$ and $\mathcal{R}_{Y;X}$.

Corollary 3.4:
$$\lim_{p \to \infty} \frac{q^*_{X;Y}(p) - 1}{p - 1} = \lim_{p \to -\infty} \frac{q^*_{X;Y}(p) - 1}{p - 1} = s^*(X;Y). \tag{22}$$

Corollary 3.5 provides a sufficient condition for (20) to hold.

Corollary 3.5: If $\rho(X;Y) = \min\{\sqrt{s^*(X;Y)}, \sqrt{s^*(Y;X)}\}$, then

$$\forall \ (U,V), \ \mathcal{R}_{X;Y} \subseteq \mathcal{R}_{U;V} \implies \rho(X;Y) \geq \rho(U;V).$$

Note that from properties listed for the hypercontractivity ribbon, DSBS sources always satisfy the condition in Corollary 3.5.

IV. Proofs

Proof of Theorem 3.1:

The proof proceeds from a perturbative argument. Let (X,Y) distributed as $P(x,y)$. Fix functions $\phi : \mathcal{X} \mapsto \mathbb{R}, \psi : \mathcal{Y} \mapsto \mathbb{R}$ such that

$$\mathbb{E}\phi(X) = \mathbb{E}\psi(Y) = 0, \ \mathbb{E}\phi(X)^2 = \mathbb{E}\psi(Y)^2 = 1. \tag{23}$$

Fix $r > 0$. Define $f : \mathcal{X} \mapsto \mathbb{R}, g : \mathcal{Y} \mapsto \mathbb{R}$ by $f(x) = 1 + \frac{\sigma}{r}\phi(x), g(y) = 1 + \sigma r \psi(y)$. Note that for sufficiently small σ, the functions f, g take only positive values. Fix $(p,q) \in \mathbb{R}_{X;Y}$ with $p > 1$. Using (6), we have

$$\mathbb{E}[(1 + \frac{\sigma}{r}\phi(X))(1 + \sigma r \psi(Y))] \leq$$
$$\left(\mathbb{E}[(1 + \frac{\sigma}{r}\phi(X))^{p'}]\right)^{1/p'} \cdot \left(\mathbb{E}[(1 + \sigma r \psi(Y))^q]\right)^{1/q}. \tag{24}$$

For Z satisfying $\mathbb{E}Z = 0, \mathbb{E}Z^2 = 1$,

$$\left(\mathbb{E}[(1 + aZ)^l]\right)^{1/l}$$
$$= \left(1 + l \cdot a\mathbb{E}Z + \frac{l(l-1)}{2} \cdot a^2 \mathbb{E}Z^2 + O(a^3)\right)^{1/l}$$
$$= \left(1 + \frac{l-1}{2}a^2 + O(a^3)\right).$$

The first two terms of the expansion on both sides of (24) match. Comparing the coefficient of σ^2 on both sides, we get

$$\mathbb{E}\phi(X)\psi(Y) \leq \frac{p'-1}{2r^2} + \frac{(q-1)r^2}{2}.$$

Taking the supremum over all ϕ, ψ satisfying (23) and the infimum over all $r > 0$, we have

$$\rho(X;Y) \leq \sqrt{\frac{q-1}{p-1}}.$$

We can similarly prove the inequality in the case when $p < 1$. We get $\mathbb{E}\phi(X)\psi(Y) \geq -\sqrt{\frac{q-1}{p-1}}$ in this case and we replace ϕ by $-\phi$ and perform similar steps to get the desired. This completes the proof.

Proof of Theorem 3.2:

The if part of the statement follows immediately from Theorem 3.1. For the only if part, suppose that for $(X,Y) \sim P(x,y)$, we have for some $\delta > 0$,

$$\rho(X;Y) = \inf_{(p,q) \in \mathcal{R}_{X;Y}, p \neq 1} \sqrt{\frac{q-1}{p-1}} - \delta.$$

A classical result [10] states that for $(U,V) \sim \text{DSBS}(\epsilon)$,

$$\frac{q^*_{U;V}(p) - 1}{p - 1} = (1 - 2\epsilon)^2 = \rho(U;V)^2.$$

Choosing ϵ so that $\rho(U;V) = 1 - 2\epsilon = \inf_{(p,q) \in \mathcal{R}_{X;Y}, p \neq 1} \sqrt{\frac{q-1}{p-1}}$, we have $\rho(X;Y) < \rho(U;V)$ and $\mathcal{R}_{X;Y} \subseteq \mathcal{R}_{U;V}$. This completes the proof.

Proof of Theorem 3.3:

As noted earlier, the inequality (4) holds for all functions f only if it holds for all non-negative functions f. Now, for non-negative f, we always have

$$||T_{X;Y}f(X)||_1 = ||f(Y)||_1 \quad \forall f \in \mathcal{F}_{\mathcal{Y}}^+. \tag{25}$$

As in [3], we define for any non-negative random variable X, the function $\text{Ent}(X) := \mathbb{E}X \log X - \mathbb{E}X \cdot \log \mathbb{E}X$, where by convention $0 \log 0 := 0$. By strict convexity of the function $x \mapsto x \log x$, we get using Jensen's inequality that $\text{Ent}(X) \geq 0$ and equality holds if and only if X is a constant almost surely. Also, note that $\text{Ent}(\cdot)$ is homogenous, that is, $\text{Ent}(aX) = a\text{Ent}(X)$ for any $a \geq 0$.

Define $s := \sup \frac{\text{Ent}(T_{X;Y}f(X))}{\text{Ent}(f(Y))}$, where the supremum is taken over non-constant functions $f \in \mathcal{F}_{\mathcal{Y}}^+$. As P_Y assigns a positive probability to all elements of $\mathcal{Y}$, this rules out the possibility of a non-constant function f with $f(Y)$ being a constant almost surely.

If $m < s$, then $(1 + \tau, 1 + m\tau) \notin \mathcal{R}_{X;Y}$ for all sufficiently small $\tau > 0$. To see this, fix f_0 to be any (non-constant) function in $\mathcal{F}_{\mathcal{Y}}^+$ that satisfies

$$\frac{\text{Ent}(T_{X;Y}f_0(X))}{\text{Ent}(f_0(Y))} \geq m + \frac{\delta}{2}, \tag{26}$$

where $\delta := s - m$. Now,

$$\|f_0(Y)\|_{1+m\tau} = \|f_0(Y)\|_1$$
$$+ m\tau \, \mathrm{Ent}(f_0(Y)) + O(\tau^2), \quad (27)$$

$$\|T_{X;Y}f_0(X)\|_{1+\tau} = \|T_{X;Y}f_0(X)\|_1$$
$$+ \tau \, \mathrm{Ent}(T_{X;Y}f_0(X)) + O(\tau^2). \quad (28)$$

Putting together (25), (26), (27), (28), we get the existence of $\tau_0 > 0$ such that

$$\|T_{X;Y}f_0(X)\|_{1+\tau} > \|f_0(Y)\|_{1+m\tau} \quad \forall \tau : 0 < \tau \le \tau_0. \quad (29)$$

If $m > s$, then consider the set $\mathcal{H}$ of all functions $f : \mathcal{Y} \mapsto \mathcal{R}^+$ that satisfy $\|f\|_1 = 1$ and define $\tau(f) := \max\{\tau : 0 \le \tau \le 1, \|T_{X;Y}f(X)\|_{1+\tau} \le \|f(Y)\|_{1+m\tau}\}$. As $\tau(f)$ is continuous over the compact set $\mathcal{H}$, showing $\tau(f) > 0 \; \forall f \in \mathcal{H}$ would yield $\tau_1 := \inf_{f \in \mathcal{H}} \tau(f) > 0$. But that is obvious since for f constant, $\tau(f) = 1$ and for f non-constant, $\tau(f) > 0$ from (25), (26), (27), (28).

This gives $\|T_{X;Y}f(X)\|_{1+\tau} \le \|f(Y)\|_{1+m\tau}$ for all $f \in \mathcal{H}, 0 < \tau \le \tau_1$. By homogeneity of the p-norm, it follows that $\|T_{X;Y}f(X)\|_{1+\tau} \le \|f(Y)\|_{1+m\tau} \; \forall f \in \mathcal{F}_{\mathcal{Y}+}, 0 < \tau \le \tau_1$, thus proving that

$$\lim_{p \to 1^+} \frac{q^*_{X;Y}(p) - 1}{p - 1} = s. \quad (30)$$

Similarly, one can prove the same limit for $p \to 1^-$. The final step is to show $s = s^*(Y; X)$. For any distribution $R(\cdot)$ on $\mathcal{Y}$, that is not equal to $P_Y(\cdot)$ consider the non-constant function f given by $f(y) := \frac{R(y)}{P_Y(y)}$. This choice yields $\mathrm{Ent}(f(Y)) = D(R\|P_Y)$ and $\mathrm{Ent}(T_{X;Y}f(X)) = D(\sum_Y P_{X|Y} * R\|P_X))$ which gives $s \ge s^*(Y; X)$. Homogeneity of $\mathrm{Ent}(\cdot)$ then completes the proof.

Proof of Corollary 3.4:

The existence of the limit and its value both follow from Theorem 3.3 and the following well-known duality result that follows from the equivalent formulations of the hypercontractivity ribbon in inequalities (6), (7): For $1 < q < p$ or $1 > q > p$,

$$(p, q) \in \mathcal{R}_{X;Y} \iff (q', p') \in \mathcal{R}_{Y;X}. \quad (31)$$

Proof of Corollary 3.5:

This is an easy consequence of Theorems 3.2, 3.3 and Corollary 3.4.

V. Non-interactive Simulation of a Three Random Variable Joint Distribution

This section discusses an interesting example. Consider joint distributions $P(x, y, z), Q(u, v, w)$ with binary random variables X, Y, Z and U, V, W. Fix $0 < \epsilon < \frac{1}{2}$. Let $X \sim \mathrm{Ber}(\frac{1}{2})$ and $Y = X + N_1, Z = Y + N_2$ where $N_1, N_2 \sim \mathrm{Ber}(\epsilon)$ are independent of X with $P(N_1 = N_2 = 0) = 1 - \frac{3\epsilon}{2}, P(N_1 = 0, N_2 = 1) = P(N_1 = 1, N_2 = 0) = P(N_1 = N_2 = 1) = \frac{\epsilon}{2}$. Let $U \sim$

$\mathrm{Ber}(\frac{1}{2})$ and $V = U + N_3, W = V + N_4$ where $N_3, N_4 \sim \mathrm{Ber}(\epsilon)$ such that U, N_3, N_4 are independent. Note that $(X, Y), (Y, Z), (X, Z), (U, V), (V, W) \sim \mathrm{DSBS}(\epsilon)$ and $(U, W) \sim \mathrm{DSBS}(2\epsilon(1-\epsilon))$ as shown in the Fig. 2

Consider the problem where three agents try to simulate a triple joint distribution as follows. Agents A, B, C observe X^n, Y^n, Z^n respectively and output $\tilde{U}, \tilde{V}, \tilde{W}$, respectively which is required to be close in total variation to the target distribution (U, V, W) as shown.

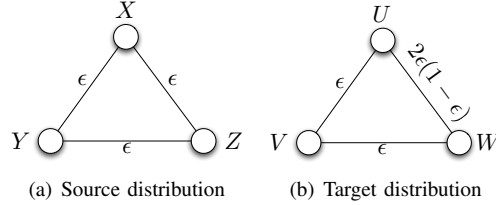(a) Source distribution   (b) Target distribution

Fig. 2. Three random variable simulation example

As discussed earlier, non-interactive simulation of a DSBS target distribution with parameter $q < \frac{1}{2}$ using a DSBS source distribution with parameter $p < \frac{1}{2}$ is possible if and only if the target distribution is more noisy, i.e. $p \le q$. Thus, for this example, each pair of agents can perform the marginal pair simulation desired of them. However, the three agents cannot simulate the desired triple joint distribution. Calculation shows

$$\rho(X, Z; Y) = \frac{1 - 2\epsilon}{\sqrt{1 - \epsilon}}, \quad (32)$$

$$\rho(U, W; V) = \frac{1 - 2\epsilon}{\sqrt{1 - 2\epsilon + 2\epsilon^2}}. \quad (33)$$

For $0 < \epsilon < \frac{1}{2}$, we have $1 - 2\epsilon + 2\epsilon^2 < 1 - \epsilon$, which gives $\rho(X, Z; Y) < \rho(U, W; V)$. This shows that even if agents A and C were merged into one agent $\tilde{A}$, then $\tilde{A}$ and B cannot achieve the desired non-interactive simulation.

References

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

[2] 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.

[3] E. Mossel, K. Oleszkiewicz, and A. Sen, "On reverse hypercontractivity", *preprint at arxiv: 1108.1210v1*, 2011.

[4] A.D. Wyner, "The common information of two dependent random variables", *IEEE Transactions On Information Theory*, vol. 21, no. 2, pp. 163–179, March 1975.

[5] P. Gács and J. Körner, "Common information is far less than mutual information", *Problems of Control and Information Theory*, vol. 2, no. 2, pp. 119–162, 1972.

[6] P. Cuff, "Communication requirements for generating correlated random variables", in *Proc. of IEEE ISIT*, Toronto, Canada, July 2008.

[7] A.A. Gohari and V. Anantharam, "Generating dependent random variables over networks", in *Proceedings of the IEEE Information Theory Workshop*, Paraty, Brazil, October 2011, pp. 698–702.

[8] M.H. Yassaee, A.A. Gohari, and M.R. Aref, "Channel simulation via interactive communications", *preprint at arxiv: 1203.3217*, 2012.

[9] P. Cuff, H. Permuter, and T. Cover, "Coordination capacity", *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4181–4206, September 2010.

[10] E. Mossel, R. O'Donnell, O. Regev, J. Steif, and B. Sudakov, "Non-interactive correlation distillation, inhomogeneous Markov chains, and the reverse Bonami-Beckner inequality", *preprint at arxiv: 0410560v1*, 2004.

[11] A. Bogdanov and E. Mossel, "On extracting common random bits from correlated sources", *preprint at arxiv: 1007.2315v2*, 2010.

[12] V. Anantharam and V. Borkar, "Common randomness and distributed control: A counterexample", *Systems and Control Letters*, vol. 56, no. 7-8, pp. 568–572, July 2007.

[13] Gowtham Kumar, "On sequences of pairs of dependent random variables: A simpler proof of the main result using SVD", *On webpage*, July 2010, `http://www.stanford.edu/~gowthamr/research/Witsenhausen_simpleproof.pdf`.

[14] Gowtham Kumar, "Binary Rényi Correlation", *On webpage*, July 2010, `http://www.stanford.edu/~gowthamr/research/binary_renyi_correlation.pdf`.

[15] A. Rényi, "On measures of dependence", *Acta. Math. Acad. Sci. Hung.*, vol. 10, pp. 441–451, 1959.

[16] E. Erkip and T. Cover, "The efficiency of investment information", *IEEE Transactions On Information Theory*, vol. 44, pp. 1026–1040, May 1998.