Dispersion Bound for the Wyner-Ahlsweede-Körner Network via Reverse Hypercontractivity on Types

Jingbo Liu
Dept. of Electrical Eng., Princeton University, NJ 08544
jingbo@princeton.edu

Abstract—Using the functional-entropic duality and the reverse hypercontractivity of the transposition semigroup, we lower bound the error probability for each joint type in the Wyner-Ahlsweede-Körner problem. Then by averaging the error probability over types, we lower bound the c-dispersion (which characterizes the second-order behavior of the weighted sum of the rates of the two compressors when a nonvanishing error probability is small) as the variance of the gradient of \( \inf_{P_{U|X}} \{ cH(Y|U) + I(U;X) \} \) with respect to \( Q_{XY} \), the per-letter side information and source distribution. On the other hand, using the method of types we derive a new upper bound of the rates of the two compressors when a nonvanishing characterizes the second-order behavior of the weighted sum of the rates of the two compressors when a nonvanishing error probability is small.

\[ n \ln P_{U|X} (\inf Q_{XY} | Y^n) \leq n \ln n \]

where \( P_{U|X} \) is the probability of \( \inf Q_{XY} | Y^n \).

We show that this bound is tight up to a constant factor, and that the second-order term \( \mathcal{O}(\sqrt{n}) \) is stable. In particular, our bound is tight in the case of the Gray-Wyner network, yielding a strong converse for WAK but not appearing to improve the second-order term.

I. INTRODUCTION

\[ Y^n \xrightarrow{Encoder 2} W_2 \xrightarrow{Decoder} \hat{Y}^n \]

\[ X^n \xrightarrow{Encoder 1} W_1 \]

Figure 1. Source coding with compressed side information

In the Wyner-Ahlsweede-Körner (WAK) problem [1][2], a source \( Y^n \) and a side information \( X^n \) are compressed separately as integers \( W_2 \) and \( W_1 \), respectively, and a decoder reconstructs \( Y^n \) as \( \hat{Y}^n \). Consider the discrete memoryless setting where the per-letter source distribution is \( Q_{XY} \), for any \( c > 0 \), define

\[ \phi_c(Q_{XY}) := \inf_{P_{U|X}} \{ cH(Y|U) + I(U;X) \} \quad (1) \]

where \( (U, X, Y) \sim P_{U|X} Q_{XY} \). The following strong converse result was proved in [2] using the blowing-up lemma: if the error probability \( P[Y^n \neq Y^n] \) is below some \( \epsilon \in (0,1) \), then

\[ \ln |W_1| + c \ln |W_2| \geq n \phi_c(Q_{XY}) - O(\sqrt{n} \ln^{1/2} n) \quad (2) \]

where the cardinality of the auxiliary can be bounded by \( |U| \leq |X| + 2 \). The first-order term in (2) is the precise single-letter characterization [1][2]. Note that for any \( c < 1 \), we have \( \phi_c(Q_{XY}) = cH(Y) \) by the data processing inequality. Moreover, \( \ln |W_1| + c \ln |W_2| \geq c \ln |W_1 \times W_2| \geq cnH(Y) \) + \( O(\sqrt{n}) \), which follows simply from a method of type analysis [3] of the single source compression problem. Therefore the only nontrivial case is \( c \geq 1 \).

While recent research has succeeded in studying the second-order rates for various single-user and selected multiuser problems (see e.g. [4][5]), it remained a formidable challenge to improve the second-order term in (2). Indeed, [5, Section 9.2.2, 9.2.3] listed it as a major open problem since previous converse techniques (e.g. simple method of types or meta-converse) appear insufficient for cases where the auxiliary random variable satisfies a Markov chain. Recently Watanabe [6] examined the converse bound obtained by taking limits in the Gray-Wyner network, yielding a strong converse for WAK but not appearing to improve the second-order term.

Recently, [7] proposed a new strong converse proof technique based on functional-entropic duality and reverse hypercontractivity, which bounds the second-order term in (2) as \( c \sqrt{n \ln \frac{1}{1-\epsilon}} \), for some \( C > 0 \) depending on the minimum probability in \( Q_{XY} \). This is the first time that an \( O(\sqrt{n}) \) second-order converse is proved for WAK. After the publication of [7], Zhou-Tan-Yu-Motani [8] and Oohama [9] improved a previous technique of Oohama and claimed that an \( O(\sqrt{n}) \) converse for WAK also follows from that technique, although a precise characterization of the prefactor appears out of reach.

The idea of [7] is roughly described as follows: first we note that an entropic quantity related to \( \phi_c \) has an equivalent functional version (23) which contains quantities such as \( \frac{1}{2} \ln dP \). If one directly takes \( f \) to be the indicator function of a decoding set, then generally \( \frac{1}{2} \ln dP = -\infty \) which is useless. However, using a machinery called reverse hypercontractivity, we design some “magic operator” \( \Lambda \) such that \( \frac{1}{2} \ln (Af) dP \geq \frac{1}{2} \ln dP \), and \( \frac{1}{2} \ln dP \) is the probability of correct decoding which we desired. For all source and channel networks where a strong converse was proved in [3], we can now bound the second-order term as \( C \sqrt{n \ln \frac{1}{1-\epsilon}} \).

In this paper, by applying the idea of [7] to each type class, we show the following lower bound on the \( c \)-dispersion, defined as the left side of (4):

\[ \lim \sup_{\epsilon \downarrow 0} \lim_{n \to \infty} \frac{\left[ \inf \{ \ln |W_1| + c \ln |W_2| \} - n \phi_c(Q_{XY}) \right]^2}{2n \ln \frac{1}{\epsilon}} \geq \text{Var}(\nabla \phi_c|Q_{XY}(X,Y)) \quad (3) \]

\[ = \text{Var}(E[c_{Y|U}(Y|U) + u_{U,X}(U;X)|XY]) \quad (4) \]

where the infimum is over codes for which \( P[E_n] \leq \epsilon \), \( (U,X,Y) \sim Q_{UXY} := Q_{U|X} Q_{XY} \), \( Q_{U|X} \) is any infimizer
for (1), and we used the notations
\[ i_{Y|U}(y|u) := \frac{1}{Q_{Y|U}(y|u)}, \quad \forall (y, u); \]
\[ i_{U:X}(u;x) := \frac{Q_{X|U}(x|u)}{Q_X(x)}, \quad \forall (u, x). \]
We can take \(|U| \leq |X| + 2\) [2]. We remark that the second-order bound \(C_{\sqrt{n} \ln \frac{1}{1-\epsilon}}\) in [7] does not give a nontrivial bound for the dispersion, whereas the present bound (3) is analogous to the dispersion formula in most other previously solved problems from the network information theory. On the achievability side, a previously published upper-bound on the \(\phi_0\) for (1), and we used the notations
\[ i_{Y|U}(y|u) := \frac{1}{Q_{Y|U}(y|u)}, \quad \forall (y, u); \]
\[ i_{U:X}(u;x) := \frac{Q_{X|U}(x|u)}{Q_X(x)}, \quad \forall (u, x). \]
We can take \(|U| \leq |X| + 2\) [2]. We remark that the second-order bound \(C_{\sqrt{n} \ln \frac{1}{1-\epsilon}}\) in [7] does not give a nontrivial bound for the dispersion, whereas the present bound (3) is analogous to the dispersion formula in most other previously solved problems from the network information theory. On the achievability side, a previously published upper-bound on the \(\phi_0\) for (1), and we used the notations
\[ i_{Y|U}(y|u) := \frac{1}{Q_{Y|U}(y|u)}, \quad \forall (y, u); \]
\[ i_{U:X}(u;x) := \frac{Q_{X|U}(x|u)}{Q_X(x)}, \quad \forall (u, x). \]
We can take \(|U| \leq |X| + 2\) [2]. We remark that the second-order bound \(C_{\sqrt{n} \ln \frac{1}{1-\epsilon}}\) in [7] does not give a nontrivial bound for the dispersion, whereas the present bound (3) is analogous to the dispersion formula in most other previously solved problems from the network information theory. On the achievability side, a previously published upper-bound on the \(\phi_0\) for (1), and we used the notations
\[ i_{Y|U}(y|u) := \frac{1}{Q_{Y|U}(y|u)}, \quad \forall (y, u); \]
\[ i_{U:X}(u;x) := \frac{Q_{X|U}(x|u)}{Q_X(x)}, \quad \forall (u, x). \]
We can take \(|U| \leq |X| + 2\) [2]. We remark that the second-order bound \(C_{\sqrt{n} \ln \frac{1}{1-\epsilon}}\) in [7] does not give a nontrivial bound for the dispersion, whereas the present bound (3) is analogous to the dispersion formula in most other previously solved problems from the network information theory. On the achievability side, a previously published upper-bound on the \(\phi_0\) for (1), and we used the notations
\[ i_{Y|U}(y|u) := \frac{1}{Q_{Y|U}(y|u)}, \quad \forall (y, u); \]
\[ i_{U:X}(u;x) := \frac{Q_{X|U}(x|u)}{Q_X(x)}, \quad \forall (u, x). \]
\[ \lim_{n \to \infty} \mathbb{P}[E_n] \leq Q\left( \frac{D}{\sqrt{n}} \right), \]  
where we defined \( V \) as (7).

### III. Discussion

It is instructive to compare our results with the known second-order rate for lossy compression of a single source (see e.g. [12]). In that problem, we are given a single source with per-letter distribution \( Q_x \), and a per-letter distortion \( d: \mathcal{U} \times \mathcal{X} \to \mathbb{R} \) on the reconstruction alphabet and the source alphabet. If \( P_{U|X} \) is an optimizer for \( \varphi(Q_x) := \inf_{P_{U|X}} \{ I(U; X) + \lambda \mathbb{E}[d(U; X)] \} \), then the stationarity condition implies that \( u_{U; X}(u; x) + \lambda d(u; x) \) is independent of \( u \). If \( P_{U|X} \) is not equal to \( \nabla \varphi(Q_x)(X) \), regardless of the choice of the optimal \( P_{U|X} \). It is known (e.g. [12]) that the dispersion equals \( \text{Var}(\nabla \varphi(Q_x)(X)) \).

Now in WAK, \( c_{WAK}/(y(u) + u_{U; X}(u; x)) \) is generally not independent of \( u \); 2) upon taking the expectation over \( u \) equals \( \nabla \varphi_e(Q)(x, y) \). Moreover, we have obtained an analogous dispersion formula in the converse part (Corollary 2).

### IV. Proof of the Converse

#### A. Proof of Theorem 1

Suppose that \( f: \mathbb{X}^n \to W_1 \), \( g: \mathbb{Y}^n \to W_2 \) are the encoders, and \( V: W_1 \times W_2 \to \mathbb{Y}^n \) denotes the decoder. For each \( w \in W_1 \), define the "correctly decodable set":

\[ B_w := \{ y^n : V(w, g(y^n)) = y^n \}. \]

Let \( P_{X^n Y^n} \) be the equiprobable distribution on \( \mathcal{T}_n(P_{XY}) \). By the assumption,

\[ \int P_{X^n Y^n} [B_{f(x^n)}|x^n] d P_{X^n}(x^n) \geq 1 - \varepsilon. \]  

Next, we lower bound the error probability using the functional inequality and reverse hypercontractivity approach. We introduce a "magic" linear operator \( \Lambda_{n,t}: \mathcal{H}_n(\mathbb{Y}) \to \mathcal{H}_n(\mathbb{Y}) \), apply it to the indicator function of a decodable set, and plug the resulting function into the functional inequality. We postpone the definition of \( \Lambda_{n,t} \) to (42) in Section V. The key properties we use are: for \( f \in \mathcal{H}_{\frac{n}{10n}}(\mathbb{Y}) \) and \( t = 1/\sqrt{n} \).

**Lower bound** \( P_{Y^n X^n}(\ln \Lambda_{n,t}f) \geq O(\sqrt{n}) \ln P_{Y^n X^n}(f) \).

**Upper bound** \( P_{Y^n}(\Lambda_{n,t}f) \leq \exp(O(\sqrt{n})) P_{Y^n}(f) \).

Now, for any \( t > 0 \),

\[ (1 - \varepsilon)(1 + \delta) \leq \int P_{Y^n X^n} [B_{f(x^n)}|x^n] d P_{X^n}(x^n) \]

\[ \sum_{w \in W_1} \int_{x^n : f(x^n) = w} P_{Y^n X^n} [B_w|x^n] d P_{X^n}(x^n) \]

\[ \leq |W_1| \int P_{Y^n X^n} [B_{w^*}|x^n] d P_{X^n}(x^n) \]

\[ \leq |W_1| \int \exp(cP_{Y^n X^n}(\ln \Lambda_{n,t}B_{w^*})) d P_{X^n} \]

\[ \leq \varepsilon^d |W_1| \int P_{Y^n}^{\varepsilon^d}[\Lambda_{n,t}B_{w^*}] \]

Here,

- (15) used Jensen’s inequality.
- For (17), we can clearly choose some \( w^* \in W_1 \) such that this line holds.
- (18) used the precise form of the lower bound stated above. This is the reverse hypercontractivity step.
- For (19), we defined

\[ d := \sup_{S_x^n} \{ c D(S_{Y^n} D(P_{Y^n}) - D(S_{X^n} D(P_{X^n})) \} \]

where \( S_x^n \to P_{Y^n X^n} \to S_{Y^n} \). A basic functional-entropic duality result (see e.g. [13]) is that

\[ d = \sup_{f \in \mathcal{H}_n(\mathbb{Y})} \left\{ \ln P_{X^n}(e^{cP_{Y^n X^n}(\ln f)} - c \ln P_{Y^n}(f) \right\} \]

which is the key functional-entropic duality step.

- (20) used the precise form of the upper bound.
- (21) used \( |B_{w^*}| \leq |W_2| \).

We thus obtain

\[ \ln |W_1| + c \ln |W_2| \]

\[ \geq -d + c \ln |\mathcal{T}_n(P_{Y^n})| \]

\[ - \inf_{t > 0} \left\{ \frac{nt}{\ln_{\min_x P_{X}(x)} + c \left( 1 + \frac{1}{t} \right) \ln \frac{1}{1 - \varepsilon} \right\} \]

\[ \leq -d - c \ln |\mathcal{T}_n(P_{Y^n})| + c \ln(1 - \varepsilon) \]

\[ - 2c \sqrt{\ln_{\min_x P_{X}(x)} \frac{1}{1 - \varepsilon}.} \]

Lemma 4 bounds \( -d - c \ln |\mathcal{T}_n(P_{Y^n})| \), and we are done.

**Remark 1.** From the proof we see that the result continues to hold if the \( Y \)-encoder is allowed to access the message of the \( X \)-encoder: \( g: \mathbb{Y} \times W_1 \to W_2 \).

**Remark 2.** We used Jensen inequality to get (17) from (14). In contrast, [2] used a reverse Markov inequality, essentially deducing from (14) that

\[ P_{X^n}[x^n : P_{Y^n X^n = x^n}[B_{w^*}] \geq 1 - \varepsilon] \geq \varepsilon^d \frac{|W_1|}{c} \]

which gives rise to a new parameter \( \varepsilon' \) to be optimized. It is possible to follow [26] with the functional approach as we did in [7]. However, proceeding with (17) is more natural and better manifests the simplicity and flexibility of the functional approach [7].

\[^{1}\text{It is interesting to note that the largest } c > 1 \text{ for which } d = 0 \text{ equals the reciprocal of the strong data processing constant.}\]
B. Single-letterization on Types

Given an $n$-type $P_{XY}$, let $P_{X^n|Y^n}$ be the equiprobable distribution on $T_n(P_{XY})$, and let $P_{Y^n|X^n}$ be the induced random transformation defined for measures supported on $T_n(P_X)$. Let

$$\psi_{c,n}(P_{XY}) := \inf_{S_X^n} \{ cH(S_{Y^n}) + D(S_{X^n}||P_X^n) \}. \quad (27)$$

Here, the infimum is over $S_{Y^n}$ supported on $T_n(P_X)$, and we have set $S_{Y^n} = S_{X^n} \rightarrow P_{Y^n|X^n} \rightarrow S_{Y^n}$.

**Lemma 4.** Given $Q_{XY}$ and $c \geq 1$, there exists $\lambda \in (0,1)$ and $E > 0$ such that for any $n$-type $P_{XY}$:

$$|P_{XY} - Q_{XY}| < \lambda,$$

$$\psi_{c,n}(P_{XY}) \geq \lambda \log \left( \frac{c}{|P_{XY} - Q_{XY}|} \right) - E \log n. \quad (28)$$

**Proof:** Under the assumption that $\phi_c$ has bounded second derivatives in a neighborhood of $Q_{XY}$, there exists $\lambda \in (0,1)$ and $E' > 0$ large enough such that

$$\phi_c(S_{X^n}) \geq \phi_c(P_{XY}) + \langle \nabla \phi_c | P_{XY} \rangle \cdot S_{XY} - P_{XY} \rangle - E' \|S_{XY} - P_{XY}\|^2 \quad (29)$$

for any $P_{XY}$: $|P_{XY} - Q_{XY}| \leq \lambda$ and any $S_{XY}$ in the probability simplex (the Taylor expansion proves (29) for $S_{XY}$ in a neighborhood of $P_{XY}$. Then using the boundedness of $\phi_c(\cdot)$ we can extend (29) to all $S_{XY}$ by choosing $E'$ large enough). Here $\|\cdot\|$ denotes the $\ell_2$ norm, although any norm admitting an inner product would work. Consider any $S_{XY}$ supported on $T_n(P_X)$, and put $S_{X^n|Y^n} = S_{X^n} P_{Y^n|X^n}$. Let $I$ be equiprobable on $\{1, \ldots, n\}$ and independent of $(X^n, Y^n)$ under $S$. Let $X_{ij}$ denote the coordinates excluding the $I$-th one. Next we introduce a notation: for any $(x, y)$, define

$$P_{XY}(x', y') := \frac{1}{n-1} \left[ n P_{XY}(x', y') - 1_{(x', y')=(x,y)} \right], \quad \forall (x', y').$$

That is, $P_{XY}$ denotes the $(n-1)$-type obtained by removing one pair $(x, y)$ from sequences of the type $n$-type $P_{XY}$. Using the chain rule and induction (detail omitted),

$$\psi_{c,n}(P_{XY}) \geq \phi_c(P_{XY}) + E[\psi_{c,n-1}(P_{XY} + \Delta)] \quad (30)$$

$$\geq \ldots \quad (31)$$

$$\geq \sum_{k=0}^{n-1} E[\phi_c(P_{XY} + \Delta_1 + \ldots + \Delta_k)] \quad (32)$$

where we defined the sequence $\Delta_1, \Delta_2, \ldots$ of random vectors in the following way: conditioned on $\Delta_1, \Delta_2, \ldots, \Delta_k$, denote $S_{XY} := P_{XY} + \sum_{k=1}^\infty \Delta_k$, and then $\Delta_{k+1} := S_{XY}^{\infty} - S_{XY}$ with probability $S_{XY}(x,y)$ for any $(x, y)$. Using (29), and noting that $\Delta_1 + \ldots + \Delta_k$ is zero mean martingale, we have

$$\psi_{c,n}(P_{XY}) \geq n \phi_c(P_{XY}) - E' \sum_{k=1}^{n-1} (n-k) E\|\Delta_k\|^2 \quad (33)$$

$$= n \phi_c(P_{XY}) - E' \sum_{k=1}^{n-1} (n-k) \cdot \frac{4}{(n-k)^2} \quad (34)$$

where (34) follows from the fact that $\|\Delta_k\| \leq |\Delta_k| \leq \frac{2}{n-k}$ with probability 1. Taking $E = 10E'$ completes the proof. ■

V. RHC for the Transposition Model

We construct the magic operator $\Lambda_{n,t}$ used in the converse.

A. The Transposition Model

Let $S = \{1, \ldots, n\}$. Consider a reversible Markov chain where the state space $\Omega$ consists of the $n!$ permutations of the sequence $(1,2,\ldots,n)$, and the generator is given by

$$L_n f := \frac{1}{n} \sum_{1 \leq i \neq j \leq n} (f \delta_{ij} - f) \quad (36)$$

for any real-valued function $f$ on $\Omega$, where $f \delta_{ij}$ denotes the composition of two mappings, and $\delta_{ij}$ denotes the transposition operator. That is, $\delta_{ij}$ switches the $i$-th and the $j$-th coordinates of a sequence for any $s^n \in \Omega$,

$$\delta_{ij} s^n_k := \begin{cases} s_i & k = j; \\ s_j & k = i; \\ s_k & \text{otherwise}. \end{cases} \quad (37)$$

As an alternative interpretation of this Markov chain, whenever a Poisson clock of rate $\frac{1}{n}$ clicks, an index pair $(i,j) \in \{1, \ldots, n\}^2$ is randomly selected and the corresponding coordinates are switched. Remark that the rate at which each coordinate changes its value roughly equals 1, which is the same as the semi-simple Markov Chain we used in [7]. Functional inequalities such as Poincaré, log-Sobolev, and modified log-Sobolev for such a Markov chain have been studied to bound its mixing time under various metrics. In particular, we recall the following upper bound on the modified log-Sobolev constant in [11], which was proved using a chain-rule and induction argument:

**Theorem 5 ([11]).** Let $P$ be the equiprobable distribution on $\Omega$. For any $n \geq 2$,

$$D(S||P) \leq -\mathbb{E} \left[ \left( L_n \log \frac{dS}{dP} \right) (X) \right], \quad \forall S \ll P, X \sim S.$$

It is known (e.g. [14, Theorem 1.11]) that a modified log-Sobolev inequality is equivalent to a reverse hypercontractivity of the corresponding Markov semigroup operator $e^{L_n t} := \sum_{k=0}^\infty t^k \frac{H_k}{k!}$. We thus have

**Corollary 6.** In the transposition model, for any $q < p < 1$, $t \geq \ln \frac{1-a}{1-p}$ and $f \in \mathcal{H}_+(\Omega)$,

$$\|e^{L_n t} f\|_{L^p(\Omega)} \geq \|f\|_{L^q(\Omega)}. \quad (38)$$

We remark that the norms in (38) are with respect to the equiprobable measure $P$. By taking the limits, we have $\|f\|_{L^q(\Omega)} = \exp (P(\ln f))$.\]
B. Reverse Hypercontractivity on Types

Now consider any finite $\mathcal{Y}$ and a Markov chain with state space $\mathcal{X}^n$. With a slight abuse of notation, let $L_n$ denote the generator of this new Markov chain. Let $P_Y$ be an $n$-type. Note that $T_n(P_Y)$ is invariant under transposition and hence also an invariant subspace for the chain. We now prove a reverse hypercontractivity for the Markov semigroup operator for this new chain. Pick any map $\phi : \mathcal{S} \to \mathcal{Y}$ such that $|\phi^{-1}(y)| = nP_Y(y)$ for each $y$. Then the extension $\phi^n$ defines a function $\Omega \to T_n(P_Y)$. Now for any $f \in \mathcal{H}_+(\mathcal{Y}^n)$, from (38) we have

$$\|e^{L_n t}(f \phi^n)\|_{L^p(\Omega)} \geq \|f \phi^n\|_{L^p(\Omega)}.$$

We claim that (39) is equivalent to

$$\|e^{L_n t}f\|_{L^p(T_n(P_Y))} \geq \|f\|_{L^p(T_n(P_Y))}.$$

Indeed, $\|f \phi^n\|_{L^p(\Omega)} = E_q^{\frac{1}{p}} \left[|f \phi^n(S^n)|^p\right] = E_q^{\frac{1}{p}} \left[|f^n(Y^n)|\right] = \|f\|_{L^p(T_n(P_Y))}$. Here, $L^p(T_n(P_Y))$ is with respect to the equiprobable measure on $T_n(P_Y)$, and so the value of $f$ on $\mathcal{Y}^n / T_n(P_Y)$ is immaterial. Moreover, from the definitions we can see that $\phi^n$ commutes with transposition, so $\left(e^{L_n t}(f \phi^n)\right)\left(s^n\right) = \left(e^{L_n t}f\right)\left(\phi^n(s^n)\right)$ for any $s^n \in \Omega$ and the left sides of (39) and (40) are therefore also equal by the same argument.

We remark that for $P_Y$ not concentrated on a $y \in \mathcal{Y}$ and as $n \to \infty$, we don’t lose too much tightness in the composition step argument, and the estimate in (40) is sharp. That is, the modified log-Sobolev constant is indeed of the constant order; the lower bound can be seen by taking linear functions in the corresponding Poincaré inequality, which is weaker than the modified log-Sobolev inequality.

C. Conditional Types: the Tensorization Argument

Let $\mathcal{X}$ and $\mathcal{Y}$ both be finite sets. For any $x^n \in \mathcal{X}^n$, define a linear operator $L_{x^n} : \mathcal{H}_+(\mathcal{Y}^n) \to \mathcal{H}_+(\mathcal{Y}^n)$ by $L_{x^n}f := \sum_{x \in \mathcal{X}} nP_X(x) \sum_{j=1}^n f_{x_j} = f_{\{i : x_i = x\}} - f$. Then we recall that $\hat{P}_{x^n}$ denotes the empirical distribution of $x^n$. Note that $L_{x^n}$ is the generator of the Markov chain where independently for each $x \in \mathcal{X}$, the length $nP_X(x)$ subsequence of $\mathcal{Y}^n$ with indices $\{i : x_i = x\}$ is the transposition model in Section V-B. Since $L_{x^n}$ is the sum of $|\mathcal{X}|$ generators for transposition models, the Markov semigroup operator $e^{L_{x^n}t}$ is a tensor product, which satisfies the reverse hypercontractivity with the same constant, by the tensorization property (see e.g. [14]). Therefore for any $n$-type $P_{XY}$, $x^n \in T_n(P_X)$, and $f : \mathcal{H}_+(\mathcal{Y}^n) \to \mathcal{H}_+(\mathcal{Y}^n)$,

$$\|e^{L_{x^n} t}f\|_{L^p(T_n(P_{Y|x}))} \geq \|f\|_{L^p(T_n(P_{Y|x}))}.$$

D. A Dominating Operator

The operator in (41) depends on $x^n$ and hence cannot be used directly in the proof of Theorem 1. We now find an upper bound which is independent of $x^n$. Define a linear operator $L_n : \mathcal{H}_+(\mathcal{Y}^n) \to \mathcal{H}_+(\mathcal{Y}^n)$ by $L_n f(y^n) = \frac{1}{n \min_j P_Y(y)} \sum_{1 \leq i,j \leq n} f(\sigma_{ij} y^n)$. Note that the summation includes the $i = j$ case, where $\sigma_{ij}$ becomes the identity. From the general formula $\left(\frac{d}{dt} e^{Lt} f\right) = Le^{Lt} f$ we can see a comparison property: since the matrix of $L_n$ entry-wisely dominate $L^n$, we have $e^{L_{x^n} t}f \geq e^{L_n t}f$ pointwise for any $t \geq 0$ and $f \in \mathcal{H}_+(\mathcal{Y}^n)$. Now consider

$$\Lambda_{n,t} := e^{L_{x^n} t} \quad \forall t > 0$$

which forms an operator semigroup (although not associated with a conditional expectation). Now $\Lambda_{n,t}$ is the operator we used in the proof of Theorem 1. Details of showing the lower bound and upper bound in the proof of Theorem 1 can be found in the extended version.

VI. ACKNOWLEDGEMENT

The author thanks Prof. Ramon van Handel for his generous scholarship and many inspiring discussions, including supplying reference [11] and the composition argument in (39), and Professor Sergio Verdú for his unceasing support and guidance on research along this line. This work is supported by NSF grants CCF-1350595, CCF-1016625, CCF-0939370, and DMS-1148711, by ARO Grants W911NF-15-1-0479 and W911NF-14-1-0094, and AFOSR FA9550-15-1-0180, and by the Center for Science of Information.

REFERENCES


