# Hypercontractivity and Information Theory

Chandra Nair

The Chinese University of Hong Kong
August 25, 2016

# Hypercontractive inequalities: an introduction

**Disclaimer**: If you are a mathematician

- Hypercontractivity is *usually* discussed using the language of Markov semi-groups
- In this talk, I will use conditional expectations (snapshot rather than a time-indexed family) to discuss hypercontractivity

# Hypercontractive inequalities: an introduction

**Disclaimer**: If you are a mathematician

- Hypercontractivity is *usually* discussed using the language of Markov semi-groups
- In this talk, I will use conditional expectations (snapshot rather than a time-indexed family) to discuss hypercontractivity

**Elementary result**

Conditional expectation (a Markov operator) is *contractive*

$$\|\mathrm{E}(X|Y)\|_p \leq \|X\|_p, \quad \forall p \geq 1,$$

where $\|X\|_p = \mathrm{E}(|X|^p)^{1/p}$.

# Hypercontractive inequalities: an introduction

**Disclaimer**: If you are a mathematician

- Hypercontractivity is *usually* discussed using the language of Markov semi-groups
- In this talk, I will use conditional expectations (snapshot rather than a time-indexed family) to discuss hypercontractivity

**Elementary result**

Conditional expectation (a Markov operator) is *contractive*

$$\|\mathrm{E}(X|Y)\|_p \leq \|X\|_p, \quad \forall p \geq 1,$$

where $\|X\|_p = \mathrm{E}(|X|^p)^{1/p}$.

**Hypercontractivity**

$(X, Y) \sim \mu_{XY}$ satisfies $(p, q)$-hypercontractivity $(1 \leq q \leq p)$ if

$$\|\mathrm{E}(g(Y)|X)\|_p \leq \|g(Y)\|_q \quad \forall g \geq 0.$$

# Background

Hypercontractive inequalities have been used in

- Quantum field theory
- Establish best constants in classical inequalities
- Bounds on semi-group kernels

# Background

Hypercontractive inequalities have been used in

- Quantum field theory
- Establish best constants in classical inequalities
- Bounds on semi-group kernels
- Boolean function analysis (KKL theorem on influences)

**This talk**: relation to (network) information theory

- equivalent characterizations
- why should information-theorists care
- why this relationship may interest mathematicians

# Part I

## Equivalent characterizations of hypercontractive inequalities using information measures

# Elementary exercises

**Definition**: $(X,Y) \sim \mu_{XY}$ is $(p,q)$-hypercontractive for $1 \leq q \leq p$ if

$$\|\mathrm{E}(g(Y)|X)\|_p \leq \|g(Y)\|_q \quad \forall g \geq 0.$$

# Elementary exercises

**Definition**: $(X,Y) \sim \mu_{XY}$ is $(p,q)$-hypercontractive for $1 \le q \le p$ if

$$\|\mathrm{E}(g(Y)|X)\|_p \le \|g(Y)\|_q \quad \forall g \ge 0.$$

**An equivalent condition**: $(X,Y) \sim \mu_{XY}$ is $(p,q)$-hypercontractive for $1 \le q \le p$ if and only if

$$\mathrm{E}(f(X)g(Y)) \le \|f(X)\|_{p'} \|g(Y)\|_q \quad \forall f, g \ge 0,$$

where $p' = \frac{p}{p-1}$, the Hölder conjugate.
**Proof**: An application of Hölder's inequality.

# Elementary exercises

**Definition**: $(X,Y) \sim \mu_{XY}$ is $(p,q)$-hypercontractive for $1 \leq q \leq p$ if

$$\|\mathrm{E}(g(Y)|X)\|_p \leq \|g(Y)\|_q \quad \forall g \geq 0.$$

**An equivalent condition**: $(X,Y) \sim \mu_{XY}$ is $(p,q)$-hypercontractive for $1 \leq q \leq p$ if and only if

$$\mathrm{E}(f(X)g(Y)) \leq \|f(X)\|_{p'} \|g(Y)\|_q \quad \forall f, g \geq 0,$$

where $p' = \frac{p}{p-1}$, the Hölder conjugate.
**Proof**: An application of Hölder's inequality.

**Tensorization property**: Let $(X_1, Y_1) \sim \mu^1_{XY}$ be independent of $(X_2, Y_2) \sim \mu^2_{XY}$, and let $(X_1, Y_1)$ and $(X_2, Y_2)$ be $(p,q)$-hypercontractive.

Then $((X_1, X_2), (Y_1, Y_2))$ is also $(p,q)$-hypercontractive.

# Elementary exercises continued...

**Define**: $r_p(X;Y) = \frac{1}{p} \times \{\inf q : (X,Y) \text{ is } (p,q)\text{-hypercontractive.}\}$

1. $r_p(X;Y)$ is decreasing in $p$.
2. The $p \to \infty$ limit of $r_p(X;Y)$ is given by

$$r_\infty(X;Y) = \inf \left\{ r : \mathrm{E}\left( e^{\mathrm{E}(\log g(Y)|X)} \right) \leq \|g(Y)\|_r \quad \forall g > 0 \right\}.$$

# Elementary exercises continued...

**Define**: $r_p(X;Y) = \frac{1}{p} \times \{\inf q : (X,Y) \text{ is } (p,q)\text{-hypercontractive.}\}$

1. $r_p(X;Y)$ is decreasing in $p$.
2. The $p \to \infty$ limit of $r_p(X;Y)$ is given by

$$r_\infty(X;Y) = \inf \left\{ r : \mathrm{E}\left( e^{\mathrm{E}(\log g(Y)|X)} \right) \leq \|g(Y)\|_r \quad \forall g > 0 \right\}.$$

A (slightly) non-trivial inequality: If $(X,Y)$ is $(p,q)$-hypercontractive then

$$\frac{q-1}{p-1} \geq \rho_m^2(X;Y),$$

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

- **Maximal correlation**: $\rho_m(X;Y) = \sup_{f,g} \mathrm{E}(f(X)g(Y))$ where $f, g$ satisfy $\mathrm{E}(f(X)) = 0 = \mathrm{E}(g(Y))$ and $\mathrm{E}(f^2(X)) = 1 = \mathrm{E}(g^2(Y))$.
- A proof follows using perturbations from constant functions along directions induced by the optimizers for maximal correlation.

# Equivalent characterizations

**Ahlswede-Gács '76**

$$r_\infty(X;Y) = \sup_{\nu_X \ll \mu_x} \frac{D(\nu_Y \| \mu_Y)}{D(\nu_X \| \mu_X)},$$

where $\nu_Y$ is the (output) distribution induced by operating the same channel $\mu_{Y|X}$ on the input distribution $\nu_X$.

**Remark**: Gács (independently) observed and used the hypercontraction of the Markov operator to study:
*Images of a set via a channel* or equivalently
*Region where measure concentrates when a noise operator is applied to a set*

# Equivalent characterizations

**Ahlswede-Gács '76**

$$r_\infty(X;Y) = \sup_{\nu_X \ll \mu_x} \frac{D(\nu_Y \| \mu_Y)}{D(\nu_X \| \mu_X)},$$

where $\nu_Y$ is the (output) distribution induced by operating the same channel $\mu_{Y|X}$ on the input distribution $\nu_X$.

**Remark**: Gács (independently) observed and used the hypercontraction of the Markov operator to study:
*Images of a set via a channel* or equivalently
*Region where measure concentrates when a noise operator is applied to a set*

**Anantharam-Gohari-Kamath-Nair '13**

$$r_\infty(X;Y) = \sup_{\nu_X \ll \mu_x} \frac{D(\nu_Y \| \mu_Y)}{D(\nu_X \| \mu_X)} = \sup_{U:U-X-Y} \frac{I(U;Y)}{I(U;X)}$$
$$= \inf \{\lambda : \mathsf{K}_X[H(Y) - \lambda H(X)]_\mu = H_\mu(Y) - \lambda H_\mu(X)\}$$

**Remark**: Our interest was motivated by the tensorization property (clear later)

# Entire regime, $p \geq 1$

The following conditions are equivalent:

1. $$\|\mathrm{E}(g(Y)|X)\|_p \leq \|g(Y)\|_q \quad \forall\ g \geq 0.$$

2. $$\mathrm{E}(f(X)g(Y)) \leq \|f(X)\|_{p'}\|g(Y)\|_q \quad \forall\ f, g \geq 0.$$

# Entire regime, $p \geq 1$

The following conditions are equivalent:

1. $$\|\mathrm{E}(g(Y)|X)\|_p \leq \|g(Y)\|_q \quad \forall\ g \geq 0.$$

2. $$\mathrm{E}(f(X)g(Y)) \leq \|f(X)\|_{p'}\|g(Y)\|_q \quad \forall\ f, g \geq 0.$$

3. Using relative entropies (Carlen – Cordero-Erasquin '09, Nair '14, Friedgut '15)
$$\frac{1}{p'}D(\nu_X\|\mu_X) + \frac{1}{q}D(\nu_Y\|\mu_Y) \leq D(\nu_{XY}\|\mu_{XY}) \quad \forall \nu_{XY} \ll \mu_{XY}.$$

# Entire regime, $p \geq 1$

The following conditions are equivalent:

1. $$\|\mathrm{E}(g(Y)|X)\|_p \leq \|g(Y)\|_q \quad \forall\, g \geq 0.$$

2. $$\mathrm{E}(f(X)g(Y)) \leq \|f(X)\|_{p'}\|g(Y)\|_q \quad \forall\, f, g \geq 0.$$

3. Using relative entropies (Carlen – Cordero-Erasquin '09, Nair '14, Friedgut '15)
$$\frac{1}{p'}D(\nu_X\|\mu_X) + \frac{1}{q}D(\nu_Y\|\mu_Y) \leq D(\nu_{XY}\|\mu_{XY}) \quad \forall \nu_{XY} \ll \mu_{XY}.$$

4. Using mutual information and auxiliary variables (Nair '14)
$$\frac{1}{p'}I(U;X) + \frac{1}{q}I(U;Y) \leq I(U;XY) \quad \forall \mu_{U|XY}.$$

5. Using convex envelopes (Nair '14)
$$\mathsf{K}_{XY}\left[\frac{1}{p'}H(X) + \frac{1}{q}H(Y) - H(XY)\right]_{\mu_{XY}} = \frac{1}{p'}H_\mu(X) + \frac{1}{q}H_\mu(Y) - H_\mu(XY).$$

# Some remarks on equivalence proof

**Functional form $\implies$ mutual information condition**

Use tensorization property:
$f(X^n) = 1_A$, where $A = \{x^n : (u_0^n, x^n) \text{ is jointly typical}\}$
$g(Y^n) = 1_B$, where $B = \{y^n : (u_0^n, y^n) \text{ is jointly typical}\}$

# Some remarks on equivalence proof

**Functional form $\implies$ mutual information condition**

Use tensorization property:
$f(X^n) = 1_A$, where $A = \{x^n : (u_0^n, x^n) \text{ is jointly typical}\}$
$g(Y^n) = 1_B$, where $B = \{y^n : (u_0^n, y^n) \text{ is jointly typical}\}$

**Mutual information condition $\implies$ relative entropy condition**

A (natural) perturbation argument

# Some remarks on equivalence proof

**Functional form $\implies$ mutual information condition**

Use tensorization property:
$f(X^n) = 1_A$, where $A = \{x^n : (u_0^n, x^n) \text{ is jointly typical}\}$
$g(Y^n) = 1_B$, where $B = \{y^n : (u_0^n, y^n) \text{ is jointly typical}\}$

**Mutual information condition $\implies$ relative entropy condition**

A (natural) perturbation argument

**Relative entropy condition $\implies$ functional form**

Let $\|f(X)\|_{p'} = \|g(Y)\|_q = 1$. Define $\nu_{XY} = \frac{1}{Z}\mu_{XY} f(X)g(Y)$.

# Some remarks on equivalence proof

**Functional form $\implies$ mutual information condition**

Use tensorization property:
$f(X^n) = 1_A$, where $A = \{x^n : (u_0^n, x^n) \text{ is jointly typical}\}$
$g(Y^n) = 1_B$, where $B = \{y^n : (u_0^n, y^n) \text{ is jointly typical}\}$

**Mutual information condition $\implies$ relative entropy condition**

A (natural) perturbation argument

**Relative entropy condition $\implies$ functional form**

Let $\|f(X)\|_{p'} = \|g(Y)\|_q = 1$. Define $\nu_{XY} = \frac{1}{Z}\mu_{XY} f(X)g(Y)$.

$$\begin{aligned} &D(\nu_{XY}\|\mu_{XY}) - \frac{1}{p'}D(\nu_X\|\mu_X) - \frac{1}{q}D(\nu_Y\|\mu_Y) \\ &= \log\frac{1}{Z} + \frac{1}{p'}\mathrm{E}_\nu\left(\log\frac{\mu_X f(X)^{p'}}{\nu_X}\right) + \frac{1}{q}\mathrm{E}_\nu\left(\log\frac{\mu_Y g(Y)^q}{\nu_Y}\right) \le \log\frac{1}{Z}. \end{aligned}$$

# Brascamp-Lieb-type inequalities

**Brascamp Lieb-type inequalities**

$(X_1, .., X_m) \sim \mu_{XY}$ is said to satisfy Brascamp-Lieb type inequalities with parameters $(\lambda_1, \lambda_2, \cdots, \lambda_m, C)$ with $\lambda_i \geq 0$ if

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

# Brascamp-Lieb-type inequalities

**Brascamp Lieb-type inequalities**

$(X_1, .., X_m) \sim \mu_{XY}$ is said to satisfy Brascamp-Lieb type inequalities with parameters $(\lambda_1, \lambda_2, \cdots, \lambda_m, C)$ with $\lambda_i \geq 0$ if

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

- Hypercontractivity is a special case of above, $C = 0$ and $m = 2$
- These parameters satisfy **tensorization** property
- Strengthen Hölder's inequality

# Equivalent characterizations: Brascamp-Lieb type inequalities

Let $X_1, .., X_m \sim \mu_{X_1,...,X_m}$.

The following conditions are equivalent:

1. $$\mathrm{E}(\prod_{i=1}^{m} f_i(X_i)) \le 2^C \prod_{i=1}^{m} \|f_i(X_i)\|_{\lambda_i} \quad \forall\ f_i \ge 0.$$

2. $$\|\mathrm{E}(\prod_{i=2}^{m} f_i(X_i)|X_1)\|_{\lambda_1'} \le 2^C \prod_{i=2}^{m} \|f_i(X_i)\|_{\lambda_i} \quad \forall\ f_i \ge 0. \quad \frac{1}{\lambda_1'} = 1 - \frac{1}{\lambda_1}.$$

3. Using relative entropies (Carlen – Cordero-Erasquin '09)
$$\sum_{i=1}^{m} \frac{1}{\lambda_i} D(\nu_{X_i} \| \mu_{X_i}) \le C + D(\nu_{X_1,..,X_m} \| \mu_{X_1,...,X_m}) \quad \forall \nu_{X_1,...,X_m} \ll \mu_{X_1,..,X_m}.$$

# Equivalent characterizations: Brascamp-Lieb type inequalities

Let $X_1, .., X_m \sim \mu_{X_1,...,X_m}$.

The following conditions are equivalent:

1. $$\mathrm{E}\Big(\prod_{i=1}^m f_i(X_i)\Big) \le 2^C \prod_{i=1}^m \|f_i(X_i)\|_{\lambda_i} \quad \forall\ f_i \ge 0.$$

2. $$\Big\|\mathrm{E}\Big(\prod_{i=2}^m f_i(X_i)\Big|X_1\Big)\Big\|_{\lambda_1'} \le 2^C \prod_{i=2}^m \|f_i(X_i)\|_{\lambda_i} \quad \forall\ f_i \ge 0. \quad \frac{1}{\lambda_1'} = 1 - \frac{1}{\lambda_1}.$$

3. Using relative entropies (Carlen – Cordero-Erasquin '09)
$$\sum_{i=1}^m \frac{1}{\lambda_i} D(\nu_{X_i}\|\mu_{X_i}) \le C + D(\nu_{X_1,..,X_m}\|\mu_{X_1,...,X_m}) \quad \forall \nu_{X_1,...,X_m} \ll \mu_{X_1,..,X_m}.$$

4. When $C = 0$ then it is also equivalent to (earlier proof immediately extends)
$$\sum_{i=1}^m \frac{1}{\lambda_i} I(U; X_i) \le I(U; X_1, ..., X_m) \quad \forall \mu_{U|X_1,..,X_m}.$$

# Ahlswede-Gacs type limit (special case)

**Interesting limit**: for information theorists

Let $\lambda_1' \to \infty$ and, $\lambda_i \to \infty$ such that $r_i = \frac{\lambda_i}{\lambda_1'}, i = 2, .., m$.

The functional characterization (Bracscamp-Lieb) reduces to

$$e^{\mathrm{E}(\sum_{i=2}^m \log f_i(X_i) | X_1)} \leq 2^C \prod_{i=2}^m \| f_i(X_i) \|_{r_i} \quad \forall f_i > 0,$$

# Ahlswede-Gacs type limit (special case)

**Interesting limit**: for information theorists

Let $\lambda_1' \to \infty$ and, $\lambda_i \to \infty$ such that $r_i = \frac{\lambda_i}{\lambda_1'}, i = 2, .., m$.

The functional characterization (Bracscamp-Lieb) reduces to

$$e^{\mathrm{E}(\sum_{i=2}^{m} \log f_i(X_i)|X_1)} \leq 2^C \prod_{i=2}^{m} \|f_i(X_i)\|_{r_i} \quad \forall f_i > 0,$$

Equivalent characterization of (Carlen – Cordero-Erasquin '09) reduces to

$$\sum_{i=2}^{m} \frac{1}{r_i} D(\nu_{X_i} \| \mu_{X_i}) \leq C + D(\nu_{X_1} \| \mu_{X_1}) \quad \forall \nu_{X_1} \ll \mu_{X_1}.$$

Here $\nu_{X_i} = \nu_{X_1} \odot \mu_{X_i|X_1}$, i.e. channels from $X_1$ to $X_i$ are preserved.

# Ahlswede-Gacs type limit (special case)

**Interesting limit**: for information theorists

Let $\lambda_1' \to \infty$ and, $\lambda_i \to \infty$ such that $r_i = \frac{\lambda_i}{\lambda_1'}, i = 2, .., m$.

The functional characterization (Bracscamp-Lieb) reduces to

$$e^{\mathrm{E}(\sum_{i=2}^m \log f_i(X_i)|X_1)} \leq 2^C \prod_{i=2}^m \|f_i(X_i)\|_{r_i} \quad \forall f_i > 0,$$

Equivalent characterization of (Carlen – Cordero-Erasquin '09) reduces to

$$\sum_{i=2}^m \frac{1}{r_i} D(\nu_{X_i} \| \mu_{X_i}) \leq C + D(\nu_{X_1} \| \mu_{X_1}) \quad \forall \nu_{X_1} \ll \mu_{X_1}.$$

Here $\nu_{X_i} = \nu_{X_1} \odot \mu_{X_i|X_1}$, i.e. channels from $X_1$ to $X_i$ are preserved.

**Remark**: Work by (Liu et. al. '16): derive above equivalence directly extending the technique of (Carlen – Cordero-Erasquin '09) and not as a limit.

# Definitions: Reverse Inequalities

## Reverse Hypercontractivity

$(X, Y) \sim \mu_{XY}$ is said to be $(\lambda_1, \lambda_2)$-reverse-hypercontractive if

$$\mathrm{E}(f(X)g(Y)) \geq \|f(X)\|_{\lambda_1} \|g(Y)\|_{\lambda_2} \quad \forall\, f(X), g(Y).$$

Interested in $\lambda_1, \lambda_2 \leq 1$ and $\frac{1}{\lambda_1} + \frac{1}{\lambda_2} \leq 1$. $\big($Notation: $\|Z\|_\lambda = \mathrm{E}(|Z|^\lambda)^{1/\lambda}$.$\big)$

# Definitions: Reverse Inequalities

## Reverse Hypercontractivity

$(X, Y) \sim \mu_{XY}$ is said to be $(\lambda_1, \lambda_2)$-reverse-hypercontractive if

$$\mathrm{E}(f(X)g(Y)) \geq \|f(X)\|_{\lambda_1} \|g(Y)\|_{\lambda_2} \quad \forall\, f(X), g(Y).$$

Interested in $\lambda_1, \lambda_2 \leq 1$ and $\frac{1}{\lambda_1} + \frac{1}{\lambda_2} \leq 1$. $\big($Notation: $\|Z\|_\lambda = \mathrm{E}(|Z|^\lambda)^{1/\lambda}.\big)$

## Reverse Brascamp-Lieb-type inequalities

$(X_1, .., X_m) \sim \mu_{XY}$ is said to satisfy reverse-Brascamp-Lieb type inequalities with parameters $(\lambda_1, \lambda_2, \cdots, \lambda_m, C)$ if

$$\mathrm{E}\Big(\prod_{i=1}^{m} f_i(X_i)\Big) \geq 2^C \prod_{i=1}^{m} \|f_i(X_i)\|_{\lambda_i} \quad \forall\, \{f_i\}.$$

# Definitions: Reverse Inequalities

## Reverse Hypercontractivity

$(X,Y) \sim \mu_{XY}$ is said to be $(\lambda_1, \lambda_2)$-reverse-hypercontractive if

$$\mathrm{E}(f(X)g(Y)) \geq \|f(X)\|_{\lambda_1} \|g(Y)\|_{\lambda_2} \quad \forall\, f(X), g(Y).$$

Interested in $\lambda_1, \lambda_2 \leq 1$ and $\frac{1}{\lambda_1} + \frac{1}{\lambda_2} \leq 1$. $\big($Notation: $\|Z\|_\lambda = \mathrm{E}(|Z|^\lambda)^{1/\lambda}.\big)$

## Reverse Brascamp-Lieb-type inequalities

$(X_1, .., X_m) \sim \mu_{XY}$ is said to satisfy reverse-Brascamp-Lieb type inequalities with parameters $(\lambda_1, \lambda_2, \cdots, \lambda_m, C)$ if

$$\mathrm{E}(\prod_{i=1}^m f_i(X_i)) \geq 2^C \prod_{i=1}^m \|f_i(X_i)\|_{\lambda_i} \quad \forall\, \{f_i\}.$$

- Reverse-Hypercontractivity is a special case of reverse-Brascamp-Lieb
- These parameters satisfy tensorization property

# Reverse Brascamp-Lieb-type inequalities

**Beigi-Nair '16**

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

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

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

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

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

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

**Recap**

**Saw**: hypercontractive inequalities can be equivalently characterized using information measures

**Recap**

**Saw**: hypercontractive inequalities can be equivalently characterized using information measures

**Part II**

Why should some information-theorists care?

# (Degraded) broadcast channel

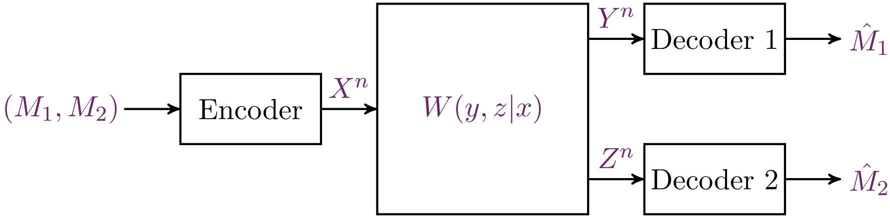


Figure 1: Discrete memoryless broadcast channel

- **Degraded**: A broadcast channel is degraded if $W(z|x) = \sum_y W'(z|y)W(y|x)$

# (Degraded) broadcast channel

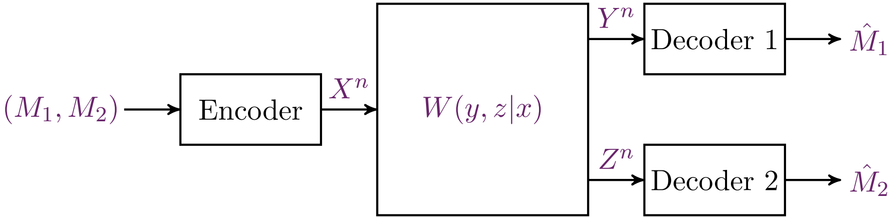


Figure 1: Discrete memoryless broadcast channel

- **Degraded**: A broadcast channel is degraded if $W(z|x) = \sum_y W'(z|y)W(y|x)$
- **Particular sub-setting**: $Y = X$

**Key Question**: What is the capacity region (or union of achievable rate pairs)?

# Capacity region characterization

(Cover '72, Gallager '74)

The capacity region, $\mathcal{C}$, is given by the union of rate pairs $(R_1, R_2)$ satisfying

$$R_2 \le I(U; Z)$$
$$R_1 \le H(X|U)$$

for some $U$ such that $U - X - Z$ is Markov.

# Capacity region characterization

(Cover '72, Gallager '74)

The capacity region, $\mathcal{C}$, is given by the union of rate pairs $(R_1, R_2)$ satisfying

$$\begin{aligned} R_2 &\leq I(U;Z) \\ R_1 &\leq H(X|U) \end{aligned}$$

for some $U$ such that $U - X - Z$ is Markov.

Gallager's converse proof:

- Single-letterization argument
- **Explicit identification** of auxiliary $U$ in terms of other variables induced by a given code

**Remark**: There are some important settings where single-letter achievable regions (in terms of auxiliaries) lack a converse, and where there is evidence to suggest that the achievable regions are optimal

# Capacity region characterization

(Cover '72, Gallager '74)

The capacity region, $\mathcal{C}$, is given by the union of rate pairs $(R_1, R_2)$ satisfying

$$R_2 \leq I(U;Z)$$
$$R_1 \leq H(X|U)$$

for some $U$ such that $U - X - Z$ is Markov.

Gallager's converse proof:

- Single-letterization argument
- **Explicit identification** of auxiliary $U$ in terms of other variables induced by a given code

**Remark**: There are some important settings where single-letter achievable regions (in terms of auxiliaries) lack a converse, and where there is evidence to suggest that the achievable regions are optimal

**Question**: Can we provide an alternate proof to the capacity region (single-letter expression) that does not involve explicit identification of auxiliaries

# Alternate converse

**Alternate characterization of capacity region**

$\max_{(R_1,R_2)\in\mathcal{C}} R_1 + \lambda R_2 = \max_{\mu_X} \lambda I_\mu(X;Z) + \mathsf{C}_X[H(X) - \lambda I(X;Z)]_\mu$.

## Remarks

- Supporting hyperplane characterization of a convex region
- Interested in $\lambda \geq 1$
- **Key**: Sub-additivity of $\mathsf{C}_X[H(X) - \lambda I(X;Z)]_\mu$ implies optimality (converse)

# Alternate converse

**Alternate characterization of capacity region**

$\max_{(R_1,R_2)\in\mathcal{C}} R_1 + \lambda R_2 = \max_{\mu_X} \lambda I_\mu(X;Z) + \mathsf{C}_X[H(X) - \lambda I(X;Z)]_\mu.$

**Remarks**

- Supporting hyperplane characterization of a convex region
- Interested in $\lambda \geq 1$
- **Key**: Sub-additivity of $\mathsf{C}_X[H(X) - \lambda I(X;Z)]_\mu$ implies optimality (converse)

**Lemma**

Sub-additivity of $\mathsf{C}_X[H(X) - \lambda I(X;Z)]_\mu$ is equivalent to *tensorization property* of $r_\infty(X;Z)$.

- **Proof**: follows from an equivalent characterization of $r_\infty(X;Z)$
- *Tensorization property* of hypercontractivity region: a simple exercise
- No identification of auxiliary variables

# Alternate converse

**Alternate characterization of capacity region**

$\max_{(R_1,R_2)\in\mathcal{C}} R_1 + \lambda R_2 = \max_{\mu_X} \lambda I_\mu(X;Z) + \mathsf{C}_X[H(X) - \lambda I(X;Z)]_\mu.$

## Remarks

- Supporting hyperplane characterization of a convex region
- Interested in $\lambda \geq 1$
- **Key**: Sub-additivity of $\mathsf{C}_X[H(X) - \lambda I(X;Z)]_\mu$ implies optimality (converse)

**Lemma**

Sub-additivity of $\mathsf{C}_X[H(X) - \lambda I(X;Z)]_\mu$ is equivalent to *tensorization property* of $r_\infty(X;Z)$.

- **Proof**: follows from an equivalent characterization of $r_\infty(X;Z)$
- *Tensorization property* of hypercontractivity region: a simple exercise
- No identification of auxiliary variables
- Our original interest in hypercontractivity came from its tensorization property

# Remarks

**Beigi-Gohari '15**

The entire hypercontractive region's tensorization property implies optimality of Gray-Wyner source coding problem

# Remarks

**Beigi-Gohari '15**

The entire hypercontractive region's tensorization property implies optimality of Gray-Wyner source coding problem

**Recall**: There are some important settings where single-letter achievable regions (in terms of auxiliaries) lack a converse, and where there is evidence to suggest that the achievable regions are optimal

- Two receiver discrete memoryless broadcast channel
- Gaussian interference channel
- Some sub-classes of broadcast channels with three or more receivers
- Sum-capacity of interference channels with very weak interference

Optimality in these settings would be implied by showing *sub-additivity* of certain functionals.

# Remarks

**Beigi-Gohari '15**

The entire hypercontractive region's tensorization property implies optimality of Gray-Wyner source coding problem

**Recall**: There are some important settings where single-letter achievable regions (in terms of auxiliaries) lack a converse, and where there is evidence to suggest that the achievable regions are optimal

- Two receiver discrete memoryless broadcast channel
- Gaussian interference channel
- Some sub-classes of broadcast channels with three or more receivers
- Sum-capacity of interference channels with very weak interference

Optimality in these settings would be implied by showing *sub-additivity* of certain functionals.

**Questions**

1. Are these sub-additivity questions equivalent to showing that certain functional inequalities satisfy a tensorization property?
2. Do the corresponding functional inequalities have an operational link with the corresponding coding questions?

Recap

**Saw**: Equivalent characterizations and tensorization property together imply optimality of single-letter regions in some settings.

Proposal is that this link is worth exploring to solve open problems and to understand existing results in a different light

### Recap

**Saw**: Equivalent characterizations and tensorization property together imply optimality of single-letter regions in some settings.

Proposal is that this link is worth exploring to solve open problems and to understand existing results in a different light

### Part III

Why may some mathematicians care?

# Background

Consider binary-valued random variables $X, Y$ distributed as follows:
$X$ is uniform, $W(y|x) \sim BSC\left(\frac{1+\rho}{2}\right)$.

Theorem (Bonami '70, Beckner '75)

*$(X, Y)$ is $(p, q)$-hypercontractive if and only if*

$$\frac{q-1}{p-1} \geq \rho^2.$$

*Shows tightness of the correlation lower bound.*

# Background

Consider binary-valued random variables $X, Y$ distributed as follows:
$X$ is uniform, $W(y|x) \sim BSC\left(\frac{1+\rho}{2}\right)$.

**Theorem (Bonami '70, Beckner '75)**

*$(X, Y)$ is $(p, q)$-hypercontractive if and only if*

$$\frac{q-1}{p-1} \geq \rho^2.$$

*Shows tightness of the correlation lower bound.*

A similar statement also holds for jointly Gaussian random variables (Gross '75)

## Remarks

- Exact characterization of optimal (or near optimal) hypercontractivity parameters has been done only in a few settings
- Typically arguments are non-trivial

**Idea**: Use equivalent characterizations to obtain new results.

# Results on $r_\infty(X;Y)$, the strong data processing constant

**Anantharam-Gohari-Kamath-Nair '13**

Consider binary-valued random variables $X, Y$ distributed as:

1. $\mathrm{P}(X=0) = \frac{1+s}{2}$, $W(y|x) \sim BSC\left(\frac{1+\rho}{2}\right)$, then

$$r_\infty(X;Y) = \frac{J\left(\frac{1+s\rho}{2}\right)}{J\left(\frac{1+s}{2}\right)}, \quad \text{where} \quad J(x) = \log\frac{1-x}{x}.$$

# Results on $r_\infty(X;Y)$, the strong data processing constant

**Anantharam-Gohari-Kamath-Nair '13**

Consider binary-valued random variables $X, Y$ distributed as:

1. $\mathrm{P}(X=0) = \frac{1+s}{2}$, $W(y|x) \sim BSC\left(\frac{1+\rho}{2}\right)$, then

$$r_\infty(X;Y) = \frac{J\left(\frac{1+s\rho}{2}\right)}{J\left(\frac{1+s}{2}\right)}, \quad \text{where} \quad J(x) = \log\frac{1-x}{x}.$$

2. $\mathrm{P}(X=1) = x$, $W(y|x) \sim Z(z)$, i.e. $W_{Y|X}(0|1) = z$, then

$$r_\infty(X;Y) = \frac{\log(1-x(1-z))}{\log(1-x)}.$$

**Remark**: Both of these immediately follow from the *convex envelope* equivalent characterization.

# Results on $r_\infty(X;Y)$, continued..

**Kamath-Nair '15**

Let $X_1, ..., X_n$ be a sequence of i.i.d. random variables and $S_m = \sum_{i=1}^m X_i$, $m \leq n$. Then,

$$r_\infty(S_n; S_m) \leq \frac{m}{n}, \text{ when } m \leq n.$$

Finite second moment, for instance, implies equality above.

**Remark**: This strengthens a result by (Dembo et. al. '01) that establish a similar result for correlation.

# Results on $r_\infty(X;Y)$, continued..

**Kamath-Nair '15**

Let $X_1, ..., X_n$ be a sequence of i.i.d. random variables and $S_m = \sum_{i=1}^m X_i$, $m \leq n$. Then,

$$r_\infty(S_n; S_m) \leq \frac{m}{n}, \text{ when } m \leq n.$$

Finite second moment, for instance, implies equality above.

**Remark**: This strengthens a result by (Dembo et. al. '01) that establish a similar result for correlation.

**Proof**: Given $U - S_n - S_m$ is Markov. W.l.o.g. can assume that $U - S_n - (X_1, ..., X_n)$ is Markov.
Let $\Phi(m) = I(U; S_m)$. Then since $I(U; S_n) = I(U; S_n, S_m, S_n - S_m, X_1^m)$, we have

$$0 = I(U; X_1^m | S_m, S_n - S_m) \geq I(U; X_1^m | S_m) \geq 0.$$

Hence $\Phi(m) = I(U; X_1^m)$ for all $m \leq n$.

# Results on $r_\infty(X;Y)$, continued..

**Kamath-Nair '15**

Let $X_1, ..., X_n$ be a sequence of i.i.d. random variables and $S_m = \sum_{i=1}^m X_i$, $m \le n$. Then,

$$r_\infty(S_n; S_m) \le \frac{m}{n}, \text{ when } m \le n.$$

Finite second moment, for instance, implies equality above.

**Remark**: This strengthens a result by (Dembo et. al. '01) that establish a similar result for correlation.

**Proof**: Given $U - S_n - S_m$ is Markov. W.l.o.g. can assume that
$U - S_n - (X_1, ..., X_n)$ is Markov.
Let $\Phi(m) = I(U; S_m)$. Then since $I(U; S_n) = I(U; S_n, S_m, S_n - S_m, X_1^m)$, we have

$$0 = I(U; X_1^m | S_m, S_n - S_m) \ge I(U; X_1^m | S_m) \ge 0.$$

Hence $\Phi(m) = I(U; X_1^m)$ for all $m \le n$.

$$\Phi(m+1) - \Phi(m) = I(U; X_{m+1} | X_1^m) \ge I(U; X_{m+1} | X_2^m) = \Phi(m) - \Phi(m-1).$$

The above convexity implies that $\frac{\Phi(m)}{m} \le \frac{\Phi(n)}{n}$ or equivalently $\frac{\Phi(m)}{\Phi(n)} \le \frac{m}{n}$.

# Results on $(p, q)$-hypercontractivity

Consider random variable $X, Y$ distributed as follows:
$X$ is uniform and binary, $W(y|x) \sim BEC(\epsilon)$.

**Theorem (Nair-Wang '16)**

*For BEC the correlation bound is tight, i.e. $(X, Y)$ is $(p, q)$-hypercontractive for $\frac{q-1}{p-1} = 1 - \epsilon$, if and only if the following condition is satisfied:*

$$\epsilon - \frac{1}{2} \leq \frac{3}{2}(q - 1).$$

# Results on $(p, q)$-hypercontractivity

Consider random variable $X, Y$ distributed as follows:
$X$ is uniform and binary, $W(y|x) \sim BEC(\epsilon)$.

**Theorem (Nair-Wang '16)**

*For BEC the correlation bound is tight, i.e. $(X, Y)$ is $(p, q)$-hypercontractive for $\frac{q-1}{p-1} = 1 - \epsilon$, if and only if the following condition is satisfied:*

$$\epsilon - \frac{1}{2} \leq \frac{3}{2}(q-1).$$

**Remarks**:

- Always holds when $\epsilon \leq \frac{1}{2}$
- Holds for all $\epsilon$ if $q \geq \frac{4}{3}$.

# Results on $(p, q)$-hypercontractivity

Consider random variable $X, Y$ distributed as follows:
$X$ is uniform and binary, $W(y|x) \sim BEC(\epsilon)$.

**Theorem (Nair-Wang '16)**

*For BEC the correlation bound is tight, i.e. $(X, Y)$ is $(p, q)$-hypercontractive for $\frac{q-1}{p-1} = 1 - \epsilon$, if and only if the following condition is satisfied:*

$$\epsilon - \frac{1}{2} \leq \frac{3}{2}(q - 1).$$

**Remarks**:

- Always holds when $\epsilon \leq \frac{1}{2}$
- Holds for all $\epsilon$ if $q \geq \frac{4}{3}$.

**Proof**:

- Uses the relative entropy characterization
- Approach: study the stationary points (unique in above region)
- Technique also yields another proof of Bonami's inequality for BSC.

### Recap

**Saw**: Equivalent characterizations help

- Compute the hypercontractivity parameters in several new settings
- Obtain new proofs of old results

**Recap**

**Saw**: Equivalent characterizations help

- Compute the hypercontractivity parameters in several new settings
- Obtain new proofs of old results

Thank You