Optimization of Some Non-Convex Functionals Arising in Information Theory

WANG Yannan (Dustin)

Dept. of Information Engg., The Chinese University of Hong Kong

June 2021

Introduction

A non-convex functional family of interest

Given a vector of discrete random variables $X^n := (X_1, \cdots, X_n)$ taking values in $\otimes_{i=1}^n \mathcal{X}_i$ and $d_{X^n} : \otimes_{i=1}^n \mathcal{X}_i \to \mathbb{R}$ some arbitrary vector:

$$G(d_{X^n}) := \max_{p_{X^n}} \left(\sum_{S \subset [1:n]} \alpha_S H(X_S) - E_{p_{X^n}}(d_{X^n}) \right),$$

where S is a subset of $[1 : n]$, X_S denotes the set $\{X_i : i \in S\}$, and $\alpha_S \in \mathbb{R}$ depends on S.

Introduction

A non-convex functional family of interest

Given a vector of discrete random variables $X^n := (X_1, \cdots, X_n)$ taking values in $\otimes_{i=1}^n \mathcal{X}_i$ and $d_{X^n} : \otimes_{i=1}^n \mathcal{X}_i \to \mathbb{R}$ some arbitrary vector:

$$G(d_{X^n}) := \max_{p_{X^n}} \left(\sum_{S \subset [1:n]} \alpha_S H(X_S) - E_{p_{X^n}}(d_{X^n}) \right),$$

where S is a subset of $[1:n]$, X_S denotes the set $\{X_i : i \in S\}$, and $\alpha_S \in \mathbb{R}$ depends on S. The non-convexity arises since α_S can be positive or negative.

Introduction

A non-convex functional family of interest

Given a vector of discrete random variables $X^n := (X_1, \cdots, X_n)$ taking values in $\otimes_{i=1}^n \mathcal{X}_i$ and $d_{X^n} : \otimes_{i=1}^n \mathcal{X}_i \to \mathbb{R}$ some arbitrary vector:

$$G(d_{X^n}) := \max_{p_{X^n}} \left(\sum_{S \subset [1:n]} \alpha_S H(X_S) - E_{p_{X^n}}(d_{X^n}) \right),$$

where S is a subset of $[1:n]$, X_S denotes the set $\{X_i : i \in S\}$, and $\alpha_S \in \mathbb{R}$ depends on S. The non-convexity arises since α_S can be positive or negative.

Why is this functional family interesting?

- The evaluation of certain achievable rate regions or bounds to the capacity region canonically involves functionals of the above form.
- If the global optimizers of a natural "product"-extension of a functional (corresponding to an achievable rate region) are product distributions, then the achievable rate regions can be shown to be optimal in many settings.

Lossless source coding with one helper

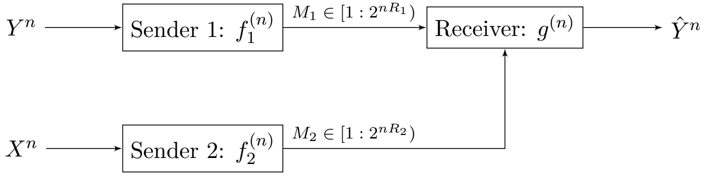


Figure 1: Lossless source coding with one helper

What is the optimal rate region $\mathscr{R}(p_{XY})$?

Lossless source coding with one helper

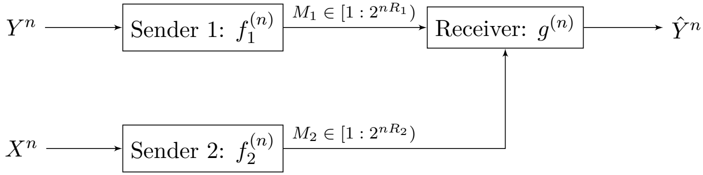


Figure 1: Lossless source coding with one helper

What is the optimal rate region $\mathscr{R}(p_{XY})$?

Theorem 1.1, [Ahlswede-Körner 1975; Wyner 1975]

Let $(X, Y) \sim p_{XY}$ be a discrete memoryless source. The optimal rate region $\mathscr{R}(p_{XY})$ for loseless source coding of Y with a helper observing X is the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\geq H(Y|U), \\ R_2 &\geq I(U; X) \end{aligned}$$

for some conditional pmf $p_{U|X}$, where $|\mathcal{U}| \leq |\mathcal{X}| + 1$.

Lossless source coding with one helper

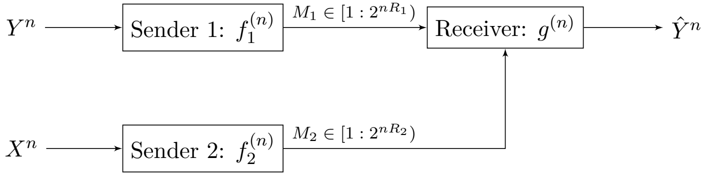


Figure 1: Lossless source coding with one helper

What is the optimal rate region $\mathscr{R}(p_{XY})$?
The optimal rate region is always convex by time-sharing argument.

Evaluation of the region: using weighted sum rates (supporting hyperplanes)

$$\begin{aligned}\min_{(R_1,R_2)\text{ achievable}} R_1 + \gamma R_2 &= \min_{p_{U|X}} H(Y|U) + \gamma I(U;X)\\ &= \gamma H(X) + \min_{p_{U|X}} \left[H(Y|U) - \gamma H(X|U)\right]\\ &= \gamma H(X) - \mathfrak{C}_{q_X}\left[\gamma H(X) - H(Y)\right](p_X)\end{aligned}$$

Non-trivial regime: $\gamma \in (0,1)$. It becomes a non-convex optimization problem.

Upper Concave Envelope and Duality (Fenchel)

Define $f(q_X) := \gamma H(X) - H(Y)$.

Upper Concave Envelope and Lower Convex Envelope (Example on next slide)

$$\mathfrak{C}_{q_X}[f] := \inf \{g : g \text{ is concave w.r.t. } q_X \text{ and } g(q_X) \geq f(q_X), \forall q_X\}$$
$$\mathfrak{K}_{q_X}[f] := -\mathfrak{C}_{q_X}[-f]$$

Fenchel's Dual Representation

Given $d_X = (d_x, x \in \mathcal{X})$ a real-valued vector of length $|\mathcal{X}|$, the Fenchel-dual of the function, $f(q_X)$, is

$$f^\dagger(d_X) := \sup_{q_X} \{\gamma H(X) - H(Y) - E_{q_X}(d_X)\}.$$

The dual variables d_x define hyperplanes, and $f^\dagger(d_X)$ is convex in d_X.

Upper Concave Envelope and Duality (Fenchel)

Define $f(q_X) := \gamma H(X) - H(Y)$.

Upper Concave Envelope and Lower Convex Envelope (Example on next slide)

$$\mathfrak{C}_{q_X}[f] := \inf \{g : g \text{ is concave w.r.t. } q_X \text{ and } g(q_X) \geq f(q_X), \forall q_X\}$$
$$\mathfrak{K}_{q_X}[f] := -\mathfrak{C}_{q_X}[-f]$$

Fenchel's Dual Representation

Given $d_X = (d_x, x \in \mathcal{X})$ a real-valued vector of length $|\mathcal{X}|$, the Fenchel-dual of the function, $f(q_X)$, is

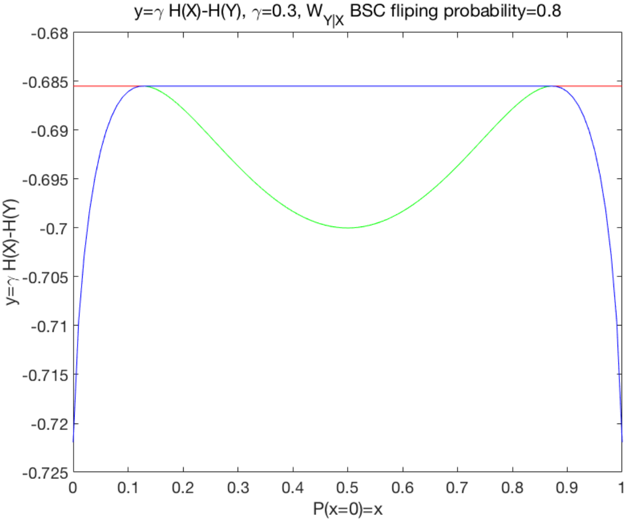
$$f^\dagger(d_X) := \sup_{q_X} \{\gamma H(X) - H(Y) - E_{q_X}(d_X)\}.$$

$$\mathfrak{C}_{q_X}[f](p_X) = \inf_{d_X} \left\{ f^\dagger(d_X) + \sum_{x \in \mathcal{X}} d_x p_X(x) \right\}$$

- The dual of the dual yields the upper concave envelope.
- Computing the dual of $f^\dagger(d_X)$ is a convex-optimization problem. Therefore, the main difficulty lies in computing the dual function, $f^\dagger(d_X)$.

Plot of Upper Concave Envelope

Simple Observation: Suffices to determine the extremal distributions, that is, the set of p_X satisfying $\mathfrak{C}_{q_X}[f](p_X) = f(p_X)$.
Example: Consider $\mathrm{P}(X=0)=x, \mathrm{P}(X=1)=1-x$.
$f(x) = 0.3H(X) - H(Y) = 0.3H_2(x) - H_2(0.8x + 0.2(1-x))$.



Outline

Outline

Outline

Definitions of Hypercontractivity

Hypercontractive Region Evaluation

Norm

$$||Z||_\lambda := E(|Z|^\lambda)^{\frac{1}{\lambda}}, \lambda \neq 0, \quad \text{(normalized } \lambda \text{ moment)}; \quad ||Z||_0 := e^{E(\log |Z|)}.$$

Forward hypercontractivity

A pair of random variables (X, Y) is said to be (λ_1, λ_2) forward hypercontractive, for $\lambda_1, \lambda_2 \in (1, \infty)$, if

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

holds for all non-negative functions $f(\cdot), g(\cdot)$.

Definitions of Hypercontractivity

Hypercontractive Region Evaluation

Norm

$$||Z||_\lambda := E(|Z|^\lambda)^{\frac{1}{\lambda}}, \lambda \neq 0, \quad \text{(normalized } \lambda \text{ moment)}; \quad ||Z||_0 := e^{E(\log |Z|)}.$$

Forward hypercontractivity

A pair of random variables (X, Y) is said to be (λ_1, λ_2) forward hypercontractive, for $\lambda_1, \lambda_2 \in (1, \infty)$, if

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

holds for all non-negative functions $f(\cdot), g(\cdot)$.

Reverse hypercontractivity

A pair of random variables (X, Y) is said to be (λ_1, λ_2) reverse hypercontractive, for $\lambda_1, \lambda_2 \in (-\infty, 1)$, if

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

holds for all positive functions $f(\cdot), g(\cdot)$.

Known hypercontractivity parameters

Binary Symmetric Channel (BSC) with uniform input: [Bonami 1970; Borell 1982]

Consider a uniformly distributed binary valued X and Y obtained by passing X through a BSC with crossover probability $\frac{1-\rho}{2}$. (X,Y) is (λ_1,λ_2) forward (reverse) hypercontractive if and only if

$$(\lambda_1-1)(\lambda_2-1)\geq\rho^2$$

Gaussian: [Gross 1975, Borell 1982]

Let $(X,Y)\sim N\left(0,\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right)$, (X,Y) is (λ_1,λ_2) forward (reverse) hypercontractive if and only if

$$(\lambda_1-1)(\lambda_2-1)\geq\rho^2$$

These hypercontractive parameters have found applications in theoretical computer science, see [Kahn-Kalai-Linial 1988; Mossel-O'Donnell-Rubinfeld-Servedio 2006].

Known hypercontractivity parameters

Binary Symmetric Channel (BSC) with uniform input: [Bonami 1970; Borell 1982]

Consider a uniformly distributed binary valued X and Y obtained by passing X through a BSC with crossover probability $\frac{1-\rho}{2}$. (X,Y) is (λ_1,λ_2) forward (reverse) hypercontractive if and only if

$$(\lambda_1-1)(\lambda_2-1)\geq\rho^2$$

Gaussian: [Gross 1975, Borell 1982]

Let $(X,Y)\sim N\left(0,\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right)$, (X,Y) is (λ_1,λ_2) forward (reverse) hypercontractive if and only if

$$(\lambda_1-1)(\lambda_2-1)\geq\rho^2$$

These hypercontractive parameters have found applications in theoretical computer science, see [Kahn-Kalai-Linial 1988; Mossel-O'Donnell-Rubinfeld-Servedio 2006].

A natural question (thanks: V. Guruswami and J. Radhakrishnan): What is the hypercontractive region for binary erasure channel with uniform inputs?

Equivalent characterizations of forward hypercontractivity

Hypercontractive Region Evaluation

Equivalent characterizations of hypercontractivity [Nair 2014]

Let $(X, Y) \sim p_{XY}$. The following assertions are equivalent:

1. For all non-negative functions $f(\cdot), g(\cdot)$,

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

2. For every $q_{XY} (\ll p_{XY})$ we have (also appeared in [Carlen-Cordero-Erausquin 2009])

$$\frac{1}{\lambda_1} D(q_X || p_X) + \frac{1}{\lambda_2} D(q_Y || p_Y) \leq D(q_{XY} || p_{XY})$$

3. For every extension $p_{U|XY}$ such that $I(U; XY) > 0$ we have

$$\frac{1}{\lambda_1} I(U; X) + \frac{1}{\lambda_2} I(U; Y) \leq I(U; XY)$$

4. Let $\mathfrak{K}[f]_x$ represents the lower convex envelope of the function f evaluated at x.

$$\mathfrak{K}\left[\frac{1}{\lambda_1} H(X) + \frac{1}{\lambda_2} H(Y) - H(XY)\right]_{p_{XY}} = \frac{1}{\lambda_1} H(X) + \frac{1}{\lambda_2} H(Y) - H(XY)$$

Equivalent characterizations of reverse hypercontractivity [Beigi-Nair 2016]

Hypercontractive Region Evaluation

Denote the reverse hypercontractive region of (λ_1, λ_2) for a pair of random variables (X, Y) distributed according to p_{XY} as $R^r(X, Y)$.

Equivalent characterizations of reverse hypercontractivity

1. The pair (λ_1, λ_2) with $0 < \lambda_1 < 1, 0 < \lambda_2 < 1$ belongs to $R^r(X, Y)$ if and only if for any q_X and q_Y there exists r_{XY} with $r_X = q_X$ and $r_Y = q_Y$ such that:

$$\frac{1}{\lambda_1} D(q_X||p_X) + \frac{1}{\lambda_2} D(q_Y||p_Y) \geq D(r_{XY}||p_{XY})$$

2. The pair (λ_1, λ_2) with $0 < \lambda_1 < 1, \lambda_2 < 0$ belongs to $R^r(X, Y)$ if and only if for any q_X there exists r_{XY} with $r_X = q_X$ such that:

$$\frac{1}{\lambda_1} D(q_X||p_X) + \frac{1}{\lambda_2} D(r_Y||p_Y) \geq D(r_{XY}||p_{XY})$$

3. The pair (λ_1, λ_2) with $\lambda_1 < 0, 0 < \lambda_2 < 1$ belongs to to $R^r(X, Y)$ if and only if for any q_Y there exists r_{XY} with $r_Y = q_Y$ such that:

$$\frac{1}{\lambda_1} D(r_X||p_X) + \frac{1}{\lambda_2} D(q_Y||p_Y) \geq D(r_{XY}||p_{XY})$$

Evaluation of (Reverse) Hypercontractivity Parameters

Hypercontractive Region Evaluation

Information Theory

- Related to determining *extremal distributions* in multiuser information theory

Gray-Wyner Source Coding

Hypercontractive Region Evaluation

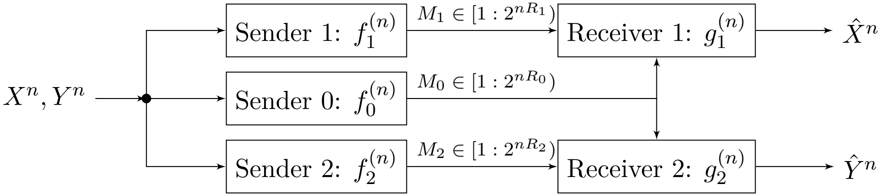


Figure 2: Gray-Wyner Source Coding Setting

Theorem 2.1 [Gray-Wyner 1974]

The optimal rate region $\mathscr{R}(p_{XY})$ is the set of rate triplets (R_0, R_1, R_2) such that

$$\begin{aligned} R_0 \geq & I(X, Y; V), \\ R_1 \geq & H(X|V), \\ R_2 \geq & H(Y|V) \end{aligned} \tag{1}$$

for some conditional pmf $p_{V|XY}$ with $|\mathcal{V}| \leq |\mathcal{X}||\mathcal{Y}| + 2$.

Gray-Wyner Source Coding Setting

Hypercontractive Region Evaluation

Evaluation of the region: using (γ_1, γ_2) weighted sum rates

$$\begin{aligned} &\min R_0 + \gamma_1 R_1 + \gamma_2 R_2 \\ =&\min I(XY;V) + \gamma_1 H(X|V) + \gamma_2 H(Y|V) \\ =&H(XY) + \mathfrak{K}[\gamma_1 H(X) + \gamma_2 H(Y) - H(XY)]_{p_{XY}} \end{aligned}$$

Observations [Beigi-Gohari 2015]:

- Tensorization of forward hypercontractivity $\Leftrightarrow$ Optimality of single letter expression of Gray-Wyner System
- $\{p_{XY} : p_{XY}$ is $(\frac{1}{\gamma_1}, \frac{1}{\gamma_2})$ forward hypercontractive$\} \equiv$ the set of extremal distributions p_{XY} for computing the lower convex envelope in the Gray-Wyner System

Outline

Main result 1: forward hypercontractive region

Binary Erasure Channel (BEC) with uniform input: [Nair-Wang 2016]

Consider a uniformly distributed binary valued X passed through a BEC with erasure probability ϵ producing the ternary output Y. For $\lambda_1, \lambda_2 \in (1, \infty)$,

- when $\epsilon - \frac{1}{2} \leq \frac{3}{2}(\lambda_2 - 1)$, the forward hypercontractive region for (X, Y) is characterized by $(\lambda_1 - 1)(\lambda_2 - 1) \geq 1 - \epsilon$;
- When $\epsilon - \frac{1}{2} > \frac{3}{2}(\lambda_2 - 1)$, the forward hypercontractive region for (X, Y) is strictly inside $(\lambda_1 - 1)(\lambda_2 - 1) \geq 1 - \epsilon$.

Main result 1: forward hypercontractive region

Binary Erasure Channel (BEC) with uniform input: [Nair-Wang 2016]

Consider a uniformly distributed binary valued X passed through a BEC with erasure probability ϵ producing the ternary output Y. For $\lambda_1, \lambda_2 \in (1, \infty)$,

- when $\epsilon - \frac{1}{2} \leq \frac{3}{2}(\lambda_2 - 1)$, the forward hypercontractive region for (X, Y) is characterized by $(\lambda_1 - 1)(\lambda_2 - 1) \geq 1 - \epsilon$;
- When $\epsilon - \frac{1}{2} > \frac{3}{2}(\lambda_2 - 1)$, the forward hypercontractive region for (X, Y) is strictly inside $(\lambda_1 - 1)(\lambda_2 - 1) \geq 1 - \epsilon$.

Remarks:

- $(\lambda_1 - 1)(\lambda_2 - 1) \geq 1 - \epsilon$ is tight when $\epsilon \leq \frac{1}{2}$;
- $(\lambda_1 - 1)(\lambda_2 - 1) \geq 1 - \epsilon$ is tight when $\lambda_2 \geq \frac{4}{3}$.

Forward hypercontractive region for BEC with uniform input

Define $\lambda_2' := \frac{\lambda_2}{\lambda_2 - 1}$, the Hölder conjugate of λ_2.
When $\epsilon = 0.2$,

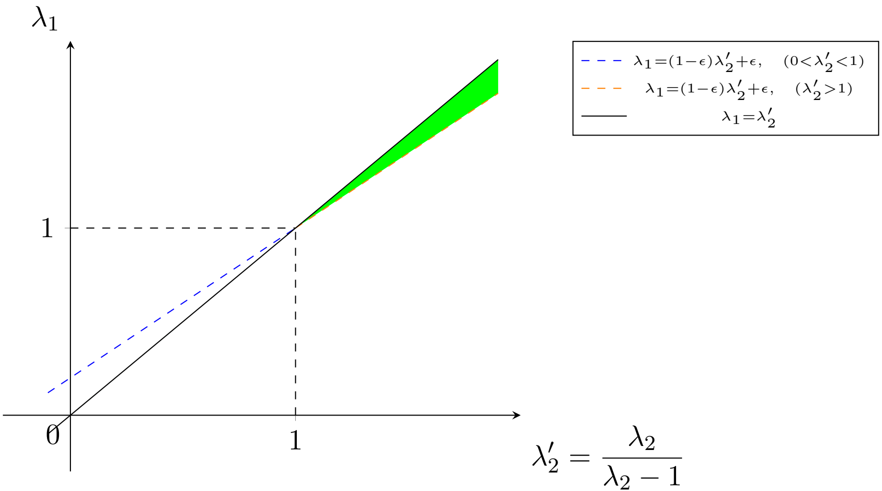


Figure 3: Forward hypercontractive region: $\epsilon = 0.2$

Main result 2: reverse hypercontractive region

Binary Erasure Channel (BEC) with uniform input: [Nair-Wang 2017]

Consider a uniformly distributed binary valued X passed through a BEC with erasure probability ϵ producing the ternary output Y. When $\lambda_2 < 0$, (X, Y) is (λ_1, λ_2) reverse hypercontractive if and only if

$$\lambda_1 \leq \frac{\ln 2}{\ln 2 - \frac{\lambda_2 - 1}{\lambda_2} \ln[(1-\epsilon) 2^{\frac{1}{\lambda_2 - 1}} + \epsilon]}$$

Remarks:

- The region $(\lambda_1 - 1)(\lambda_2 - 1) \geq 1 - \epsilon$ is not tight.

Reverse hypercontractive region for BEC with uniform input

Define $\lambda_2' := \frac{\lambda_2}{\lambda_2 - 1}$, the Hölder conjugate of λ_2.
When $\epsilon = 0.2$,

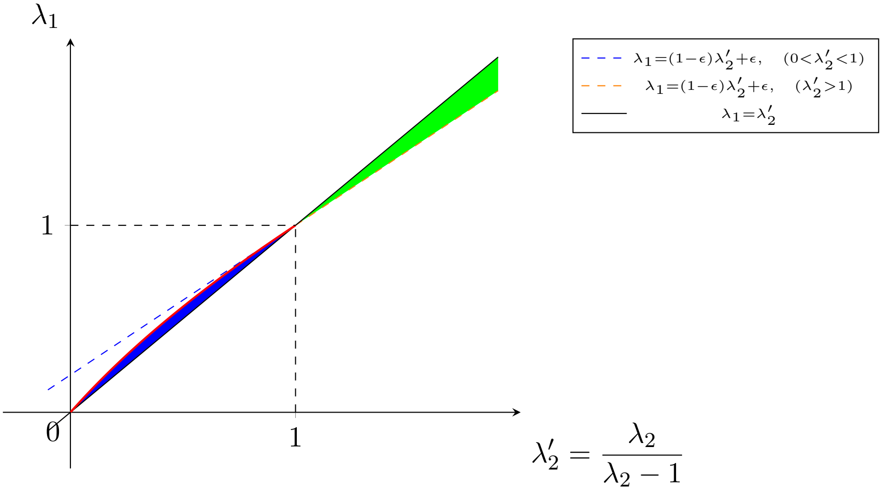


Figure 4: Reverse hypercontractive region: $\epsilon = 0.2$

Proof sketch for forward hypercontractive region

Case 1: $\lambda_2 \geq 2$

- Mimic Janson's proof technique for BSC [Janson 1997]: Denote $\lambda_2' := \frac{\lambda_2}{\lambda_2 - 1}$, by Hölder's inequality, suffices to show that given $(\lambda_1 - 1)(\lambda_2 - 1) \geq 1 - \epsilon$,

$$\| \mathrm{E}(f(X)|Y)\|_{\lambda_2'} \leq \|f(X)\|_{\lambda_1}$$

  W.l.o.g let $f(0) = 1 - \delta, f(1) = 1 + \delta$, compare the coefficients of Taylor expansion around $\delta = 0$ for both sides.

Proof sketch for forward hypercontractive region

Case 1: $\lambda_2 \geq 2$

- Mimic Janson's proof technique for BSC [Janson 1997]: Denote $\lambda_2' := \frac{\lambda_2}{\lambda_2 - 1}$, by Hölder's inequality, suffices to show that given $(\lambda_1 - 1)(\lambda_2 - 1) \geq 1 - \epsilon$,

$$\| \mathrm{E}(f(X)|Y)\|_{\lambda_2'} \leq \|f(X)\|_{\lambda_1}$$

  W.l.o.g let $f(0) = 1 - \delta, f(1) = 1 + \delta$, compare the coefficients of Taylor expansion around $\delta = 0$ for both sides.

- Works for binary input symmetric output channel with binary uniform inputs X and Y is obtained via a symmetric channel $W_{Y|X}$.

$$W_{Y|X}(Y = i|X = 1) = W_{Y|X}(Y = -i|X = -1) = p_i, \forall - K \leq i \leq K, K \in \mathbb{N}_+$$

Proof sketch for forward hypercontractive region

Case 2: $1 < \lambda_2 < 2$, denote the joint distribution of binary erasure channel with uniform input as $p_{XY}^{BEC(\epsilon)}$.

- Needs to show when $(\lambda_1 - 1)(\lambda_2 - 1) = 1 - \epsilon$,

$$\max_{q_{XY} \ll p_{XY}^{BEC(\epsilon)}} \frac{1}{\lambda_1} D(q_X || p_X^{BEC(\epsilon)}) + \frac{1}{\lambda_2} D(q_X || p_Y^{BEC(\epsilon)}) - D(q_{XY} || p_{XY}^{BEC(\epsilon)})$$
$$= \begin{cases} 0 & \text{if } \epsilon - \frac{1}{2} \leq \frac{3}{2}(\lambda_2 - 1) \\ > 0 & otherwise \end{cases}$$

- Non-convex optimization problem: Maximum happens in the interior.
- Trivial stationary point $q_{XY} = p_{XY}^{BEC(\epsilon)}$.

Proof sketch for forward hypercontractive region

Case 2: $1 < \lambda_2 < 2$, denote the joint distribution of binary erasure channel with uniform input as $p_{XY}^{BEC(\epsilon)}$.

- Needs to show when $(\lambda_1 - 1)(\lambda_2 - 1) = 1 - \epsilon$,

$$\max_{q_{XY} \ll p_{XY}^{BEC(\epsilon)}} \frac{1}{\lambda_1} D(q_X || p_X^{BEC(\epsilon)}) + \frac{1}{\lambda_2} D(q_X || p_Y^{BEC(\epsilon)}) - D(q_{XY} || p_{XY}^{BEC(\epsilon)})$$
$$= \begin{cases} 0 & \text{if } \epsilon - \frac{1}{2} \leq \frac{3}{2}(\lambda_2 - 1) \\ > 0 & otherwise \end{cases}$$

- Non-convex optimization problem: Maximum happens in the interior.
- Trivial stationary point $q_{XY} = p_{XY}^{BEC(\epsilon)}$.
- Proof idea: the stationary points of this 3-variable function are restricted on a 1-parameter path.

Proof sketch for forward hypercontractive region

Case 2: $1 < \lambda_2 < 2$, set $1-\delta = \frac{2q_{0E}}{q_{0E}+q_{1E}}$, from the first order conditions, any (strictly) interior stationary points can be parameterized in

$$\begin{aligned} q_{XY} &= [q_{00}, q_{0E}, q_{1E}, q_{11}] \\ &= \left[\frac{(1-\delta)^{\lambda_2'}(1-\epsilon)}{A}, \frac{\epsilon(1-\delta)}{A}, \frac{\epsilon(1+\delta)}{A}, \frac{(1-\epsilon)(1+\delta)^{\lambda_2'}}{A}\right] \end{aligned}$$

where $A = 2\epsilon + (1-\epsilon)[(1+\delta)^{\lambda_2'} + (1-\delta)^{\lambda_2'}]$ is the normalizing constant and δ satisfies that

$$(1-\epsilon)(1-\delta)^{\frac{\epsilon}{\lambda_2-1}} + \epsilon(1-\delta)^{\frac{\epsilon-1}{\lambda_2-1}} = (1-\epsilon)(1+\delta)^{\frac{\epsilon}{\lambda_2-1}} + \epsilon(1+\delta)^{\frac{\epsilon-1}{\lambda_2-1}}.$$

Proof sketch for forward hypercontractive region

Case 2: $1 < \lambda_2 < 2$, set $1-\delta = \frac{2q_{0E}}{q_{0E}+q_{1E}}$, from the first order conditions, any (strictly) interior stationary points can be parameterized in

$$\begin{aligned} q_{XY} &= [q_{00}, q_{0E}, q_{1E}, q_{11}] \\ &= \left[\frac{(1-\delta)^{\lambda_2'}(1-\epsilon)}{A}, \frac{\epsilon(1-\delta)}{A}, \frac{\epsilon(1+\delta)}{A}, \frac{(1-\epsilon)(1+\delta)^{\lambda_2'}}{A}\right] \end{aligned}$$

where $A = 2\epsilon + (1-\epsilon)[(1+\delta)^{\lambda_2'} + (1-\delta)^{\lambda_2'}]$ is the normalizing constant and δ satisfies that

$$(1-\epsilon)(1-\delta)^{\frac{\epsilon}{\lambda_2-1}} + \epsilon(1-\delta)^{\frac{\epsilon-1}{\lambda_2-1}} = (1-\epsilon)(1+\delta)^{\frac{\epsilon}{\lambda_2-1}} + \epsilon(1+\delta)^{\frac{\epsilon-1}{\lambda_2-1}}.$$

- When $(\epsilon - \frac{1}{2}) \le \frac{3}{2}(\lambda_2 - 1)$ and $\lambda_2 \in (1,2)$, this equation has only one root at $\delta = 0$, implying only one stationary point $q_{XY} = p_{XY}^{BEC(\epsilon)}$;

Proof sketch for forward hypercontractive region

Case 2: $1 < \lambda_2 < 2$, set $1-\delta = \frac{2q_{0E}}{q_{0E}+q_{1E}}$, from the first order conditions, any (strictly) interior stationary points can be parameterized in

$$\begin{aligned} q_{XY} &= [q_{00}, q_{0E}, q_{1E}, q_{11}] \\ &= \left[\frac{(1-\delta)^{\lambda_2'}(1-\epsilon)}{A}, \frac{\epsilon(1-\delta)}{A}, \frac{\epsilon(1+\delta)}{A}, \frac{(1-\epsilon)(1+\delta)^{\lambda_2'}}{A}\right] \end{aligned}$$

where $A = 2\epsilon + (1-\epsilon)[(1+\delta)^{\lambda_2'} + (1-\delta)^{\lambda_2'}]$ is the normalizing constant and δ satisfies that

$$(1-\epsilon)(1-\delta)^{\frac{\epsilon}{\lambda_2-1}} + \epsilon(1-\delta)^{\frac{\epsilon-1}{\lambda_2-1}} = (1-\epsilon)(1+\delta)^{\frac{\epsilon}{\lambda_2-1}} + \epsilon(1+\delta)^{\frac{\epsilon-1}{\lambda_2-1}}.$$

- When $(\epsilon - \frac{1}{2}) \le \frac{3}{2}(\lambda_2 - 1)$ and $\lambda_2 \in (1,2)$, this equation has only one root at $\delta = 0$, implying only one stationary point $q_{XY} = p_{XY}^{BEC(\epsilon)}$;
- When $(\epsilon - \frac{1}{2}) > \frac{3}{2}(\lambda_2 - 1)$, Taylor series expansion around $\delta = 0$ gives that

$$\frac{1}{\lambda_1} D(q_X \| p_X^{BEC(\varepsilon)}) + \frac{1}{\lambda_2} D(q_Y \| p_Y^{BEC(\varepsilon)}) - D(q_{XY} \| p_{XY}^{BEC(\varepsilon)}) > 0$$

Proof sketch for reverse hypercontractive region

When $\lambda_2 < 0$, need to determine (λ_1, λ_2) such that

$$\min_{q_X} \max_{r_{XY}} \frac{1}{\lambda_1} D(q_X || p_X^{BEC(\epsilon)}) + \frac{1}{\lambda_2} D(q_Y || p_Y^{BEC(\epsilon)}) - D(r_{XY} || p_{XY}^{BEC(\epsilon)}) \geq 0$$

Write $q_X(0) = x$, $r_{XY}(0,0) = r$, $r_{XY}(1,1) = s$ and denote above 3-variable function as $f(x, r, s)$. Define, for $x \in [0, 1]$

$$g(x) := \max_{r,s: r \in [0,x], s \in [0,1-x]} f(x, r, s).$$

Wish to determine (λ_1, λ_2) (with $\lambda_2 < 0$) such that $g(x) \geq 0, \forall x \in [0, 1]$.

Proof sketch for reverse hypercontractive region

When $\lambda_2 < 0$, need to determine (λ_1, λ_2) such that

$$\min_{q_X} \max_{r_{XY}} \frac{1}{\lambda_1} D(q_X || p_X^{BEC(\epsilon)}) + \frac{1}{\lambda_2} D(q_Y || p_Y^{BEC(\epsilon)}) - D(r_{XY} || p_{XY}^{BEC(\epsilon)}) \geq 0$$

Write $q_X(0) = x$, $r_{XY}(0,0) = r$, $r_{XY}(1,1) = s$ and denote above 3-variable function as $f(x,r,s)$. Define, for $x \in [0,1]$

$$g(x) := \max_{r,s: r \in [0,x], s \in [0,1-x]} f(x,r,s).$$

Wish to determine (λ_1, λ_2) (with $\lambda_2 < 0$) such that $g(x) \geq 0, \forall x \in [0,1]$.
Easy direction: From above, we require $g(0) \geq 0$. This implies that

$$\lambda_1 \leq \frac{\ln 2}{\ln 2 - \frac{\lambda_2 - 1}{\lambda_2} \ln[(1-\epsilon) 2^{\frac{1}{\lambda_2 - 1}} + \epsilon]}$$

Proof sketch for reverse hypercontractive region

When $\lambda_2 < 0$, need to determine (λ_1, λ_2) such that

$$\min_{q_X} \max_{r_{XY}} \frac{1}{\lambda_1} D(q_X || p_X^{BEC(\epsilon)}) + \frac{1}{\lambda_2} D(q_Y || p_Y^{BEC(\epsilon)}) - D(r_{XY} || p_{XY}^{BEC(\epsilon)}) \geq 0$$

Write $q_X(0) = x$, $r_{XY}(0,0) = r$, $r_{XY}(1,1) = s$ and denote above 3-variable function as $f(x, r, s)$. Define, for $x \in [0, 1]$

$$g(x) := \max_{r,s: r \in [0,x], s \in [0,1-x]} f(x, r, s).$$

Wish to determine (λ_1, λ_2) (with $\lambda_2 < 0$) such that $g(x) \geq 0, \forall x \in [0, 1]$.
Easy direction: From above, we require $g(0) \geq 0$. This implies that

$$\lambda_1 \leq \frac{\ln 2}{\ln 2 - \frac{\lambda_2 - 1}{\lambda_2} \ln[(1-\epsilon) 2^{\frac{1}{\lambda_2 - 1}} + \epsilon]}$$

Non-trivial direction: we show that

- $g(x)$ is symmetric along $x = \frac{1}{2}$. (Easy - by symmetry of the (X, Y)-distribution)
- $g(x)$ is convex at $x = \frac{1}{2}$ and $g'\left(\frac{1}{2}\right) = 0, g\left(\frac{1}{2}\right) = 0$. (Easy)
- $g(x)$ has only one stationary point, i.e., $g'(x) = 0$, between $(0, \frac{1}{2})$. (Needs to use the 1-parameter path)

Related Open Questions

To conclude,

1. In our proofs, *local analysis* suffices to compute the hypercontractive region.
2. The critical behavior happens at the boundary for reverse hypercontractivity.

Related Open Questions

To conclude,

1. In our proofs, *local analysis* suffices to compute the hypercontractive region.
2. The critical behavior happens at the boundary for reverse hypercontractivity.

How to determine hypercontractive parameters for a general joint distribution?

- In other words, does the functional

$$\frac{1}{\lambda_1}H(X) + \frac{1}{\lambda_2}H(Y) - H(XY) - E_{p_{XY}}(d_{XY})$$

  have *nice* geometric properties (or low dimensional reparametrizations) that allow such local arguments to work?
- If so, can we devise an algorithm to efficiently approximate the hypercontractivity parameters?

Outline

1. Hypercontractive Region Evaluation
    - Introduction to Hypercontractivity
    - Main Results on Hypercontractivity
2. Lower Bounds on Distributed Source Coding
    - Körner and Marton's Modulo Two Sum Problem
    - Alternative Proofs to Quadratic Gaussian CEO Problem and Distributed Source Coding Problem
3. Log-Convexity of Fisher Information
    - Motivations
    - Proof to log-convexity of Fisher information

Outline

Lossless source coding with two helpers

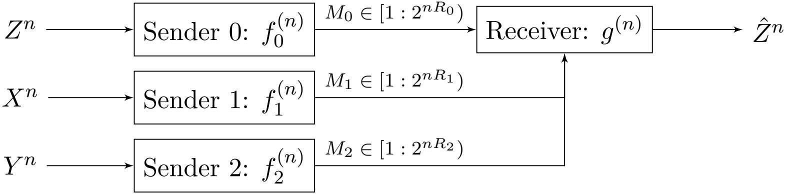


Figure 5: Lossless source coding with two helpers

- The optimal rate region is unknown for a general p_{XYZ}.
- Consider the projection $R_0 = 0$:

Lossless source coding with two helpers

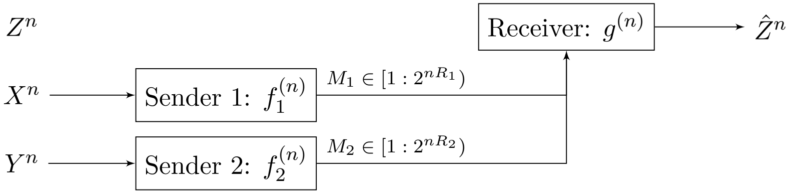


Figure 5: Lossless source coding with two helpers

- The optimal rate region is unknown for a general p_{XYZ}.
- Consider the projection $R_0 = 0$:

Slepian-Wolf region [Slepian-Wolf 1973]

When p_{XYZ} satisfies that $Z = (X, Y)$, the optimal rate region is given by (achieved by random binning)

$$\begin{aligned} R_1 &\geq H(X|Y) \\ R_2 &\geq H(Y|X) \\ R_1 + R_2 &\geq H(XY) \end{aligned}$$

Lossless source coding with two helpers

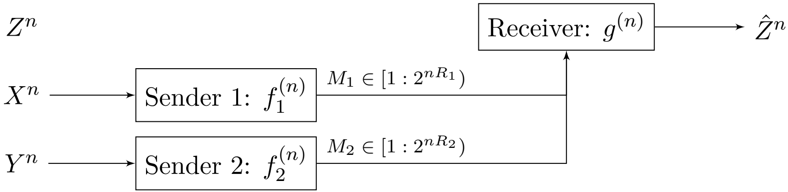


Figure 5: Lossless source coding with two helpers

- The optimal rate region is unknown for a general p_{XYZ}.
- Consider the projection $R_0 = 0$:

Körner-Marton region [Körner-Marton 1979]

When X, Y binary and p_{XYZ} satisfies that $Z = X \oplus Y$, a rate pair (R_1, R_2) is achievable by random linear codes if

$$R_1 \geq H(Z)$$
$$R_2 \geq H(Z)$$

The optimal rate region for $Z = X \oplus Y$ is unknown for a general p_{XY}. This is referred to as Körner and Marton's modulo two sum problem.

Known results on the optimal rate region

Exercise 16.23 in [Csiszár-Körner 2011][1]

When p_{XY} satisfies that $H(Z) \geq \min\{H(X), H(Y)\}$, the optimal rate region for $Z = X \oplus Y$ in $GF(2)$ is given by Slepian-Wolf region:

$$\begin{aligned} R_1 \geq & H(X|Y), \\ R_2 \geq & H(Y|X), \\ R_1 + R_2 \geq & H(XY). \end{aligned}$$

[1] I. Csiszár and J. Körner, Information theory: Coding theorems for discrete memoryless systems

Known results on the optimal rate region

Exercise 16.23 in [Csiszár-Körner 2011][1]

When p_{XY} satisfies that $H(Z) \geq \min\{H(X), H(Y)\}$, the optimal rate region for $Z = X \oplus Y$ in $GF(2)$ is given by Slepian-Wolf region:

$$\begin{aligned} R_1 &\geq H(X|Y), \\ R_2 &\geq H(Y|X), \\ R_1 + R_2 &\geq H(XY). \end{aligned}$$

Theorem 1 in [Körner-Marton 1979]

When p_{XY} follows binary symmetric channel with uniform inputs, the optimal rate region for $Z = X \oplus Y$ in $GF(2)$ is given by Körner-Marton region:

$$\begin{aligned} R_1 &\geq H(Z), \\ R_2 &\geq H(Z). \end{aligned}$$

[1] I. Csiszár and J. Körner, Information theory: Coding theorems for discrete memoryless systems

Known results on the optimal rate region

Exercise 16.23 in [Csiszár-Körner 2011][1]

When p_{XY} satisfies that $H(Z) \geq \min\{H(X), H(Y)\}$, the optimal rate region for $Z = X \oplus Y$ in $GF(2)$ is given by Slepian-Wolf region:

$$\begin{aligned} R_1 \geq& H(X|Y), \\ R_2 \geq& H(Y|X), \\ R_1 + R_2 \geq& H(XY). \end{aligned}$$

Theorem 1 in [Körner-Marton 1979]

When p_{XY} follows binary symmetric channel with uniform inputs, the optimal rate region for $Z = X \oplus Y$ in $GF(2)$ is given by Körner-Marton region:

$$\begin{aligned} R_1 \geq& H(Z), \\ R_2 \geq& H(Z). \end{aligned}$$

This part: more distributions p_{XY} are discovered for optimality of Slepian-Wolf coding scheme and Körner-Marton coding scheme on weighted sum rate.

[1] I. Csiszár and J. Körner, Information theory: Coding theorems for discrete memoryless systems

Achievable region and lower bound

Ahlswede-Han achievable region: [Ahlswede-Han 1973]

When $Z = X \oplus Y$, a rate pair (R_1, R_2) is achievable via a combination of random linear codes and random binning if

$$
\begin{aligned}
R_1 &\geq I(U;X|V) + H(Z|UV)\\
R_2 &\geq I(V;Y|U) + H(Z|UV)\\
R_1 + R_2 &\geq I(UV;XY) + 2H(Z|UV)
\end{aligned}
$$

for some U and V that satisfy the Markov chain $U \to X \to Y \to V$.

Achievable region and lower bound

Ahlswede-Han achievable region: [Ahlswede-Han 1973]

When $Z = X \oplus Y$, a rate pair (R_1, R_2) is achievable via a combination of random linear codes and random binning if

$$\begin{aligned} R_1 &\geq I(U;X|V) + H(Z|UV) \\ R_2 &\geq I(V;Y|U) + H(Z|UV) \\ R_1 + R_2 &\geq I(UV;XY) + 2H(Z|UV) \end{aligned}$$

for some U and V that satisfy the Markov chain $U \to X \to Y \to V$.

Cut-set lower bound: [Körner-Marton 1979]

Any achievable rate pair (R_1, R_2) for the modulo sum problem must satisfy

$$\begin{aligned} R_1 &\geq H(Z|Y) = H(X|Y) \\ R_2 &\geq H(Z|X) = H(Y|X) \\ R_1 + R_2 &\geq H(Z). \end{aligned}$$

Main result: A lower bound

A lower bound on modulo sum problem [Nair-Wang 2020]

Any achievable rate pair (R_1, R_2) for the modulo sum problem must satisfy the following constraints for any $\lambda \geq 1$:

$$R_1 + \lambda R_2 \geq H(XY) + \min_{U \to X \to Y} \lambda H(Z|U) - H(Y|U)$$
$$\lambda R_1 + R_2 \geq H(XY) + \min_{V \to Y \to X} \lambda H(Z|V) - H(X|V)$$

Remark: From [Nair 2013]

$$\min_{U \to X \to Y} \lambda H(Z|U) - H(Y|U) = -\mathfrak{C}_{q_X}[H(Y) - \lambda H(Z)]\big|_{p(x)},$$

where $\mathfrak{C}_x[f]\big|_{x_0}$ denotes the upper concave envelope of the function $f(x)$ with respect to x evaluated at $x = x_0$.

Proof sketch

For $\lambda \geq 1$, any "good" sequence of codes will require that

$$
\begin{aligned}
& n(R_1 + \lambda R_2) + n(1+\lambda)\varepsilon_n \\
& \geq I(M_1 M_2; X^n Y^n) + (\lambda - 1) H(M_2 | M_1) + (1+\lambda) H(Z^n | M_1 M_2)
\end{aligned}
$$

Proof sketch

For $\lambda \geq 1$, any "good" sequence of codes will require that

$$
\begin{aligned}
& n(R_1 + \lambda R_2) + n(1+\lambda)\varepsilon_n \\
& \geq I(M_1 M_2; X^n Y^n) + (\lambda - 1) H(M_2|M_1) + (1+\lambda) H(Z^n|M_1 M_2) \\
& = H(X^n Y^n) - \underset{\sim\sim\sim}{H(X^n Y^n M_1 M_2)} + (\lambda-1) H(M_1 M_2) - (\lambda - 1) H(M_1) \\
& \qquad + (1+\lambda) H(Z^n M_1 M_2) - \lambda H(M_1 M_2) \\
& \overset{(a)}{=} H(X^n Y^n) + \lambda H(Z^n M_1 M_2) + H(Z^n M_1 M_2) - \underset{\sim\sim\sim}{H(Z^n Y^n M_1 M_2)} - H(M_1 M_2) \\
& \qquad - (\lambda - 1) H(M_1) \\
& \overset{(b)}{=} H(X^n Y^n) + \lambda H(Z^n M_1) + \underline{\lambda H(M_2|M_1 Z^n)} - \underset{\sim\sim\sim}{H(Y^n M_1 M_2)} + \underline{I(Z^n; Y^n|M_1 M_2)} \\
& \qquad - (\lambda - 1) H(M_1) \\
& \geq n H(XY) + \lambda H(Z^n M_1) - \underset{\sim\sim\sim}{H(Y^n M_1)} - (\lambda-1) H(M_1) \\
& = n H(XY) + \lambda H(Z^n|M_1) - H(Y^n|M_1)
\end{aligned}
$$

- Step (a) uses $H(X^n Y^n M_1 M_2) = H(Z^n X^n Y^n M_1 M_2) = H(Z^n Y^n M_1 M_2)$.
- Step (b) uses $I(Z^n; Y^n|M_1 M_2) = H(Z^n M_1 M_2) + H(Y^n M_1 M_2) - H(M_1 M_2) - H(Z^n Y^n M_1 M_2)$.

Sinlge-letterize the lower bound

Lemma 3.1: Körner-Marton identity (4.14) in [Körner-Marton 1977]

Let $\lambda \geq 1$ and let (X^n, Y^n) be i.i.d distributed according to $p(x, y)$ where X, Y take values in a finite field. Let Z^n be obtained as $Z_i = X_i \oplus Y_i, i = 1, .., n$, i.e. the component-wise modulo sum on the field. Then for any $\lambda \geq 1$ the following holds:

$$\begin{aligned} &\min_{\hat{U}:\hat{U}\to X^n \to Y^n} \lambda H(Z^n|\hat{U}) - H(Y^n|\hat{U}) \\ &= n \left(\min_{U:U\to X\to Y} \lambda H(Z|U) - H(Y|U) \right). \end{aligned}$$

Sinlge-letterize the lower bound

Lemma 3.1: Körner-Marton identity (4.14) in [Körner-Marton 1977]

Let $\lambda \geq 1$ and let (X^n, Y^n) be i.i.d distributed according to $p(x,y)$ where X, Y take values in a finite field. Let Z^n be obtained as $Z_i = X_i \oplus Y_i, i = 1, .., n$, i.e. the component-wise modulo sum on the field. Then for any $\lambda \geq 1$ the following holds:

$$\min_{\hat{U}:\hat{U}\to X^n \to Y^n} \lambda H(Z^n|\hat{U}) - H(Y^n|\hat{U})$$
$$= n\left(\min_{U:U\to X\to Y} \lambda H(Z|U) - H(Y|U)\right).$$

- Evaluating the weighted sum rate lower bounds are non-convex optimization problems:

$$R_1 + \lambda R_2 \geq H(XY) + \min_{U\to X\to Y} \lambda H(Z|U) - H(Y|U)$$
$$\lambda R_1 + R_2 \geq H(XY) + \min_{V\to Y\to X} \lambda H(Z|V) - H(X|V)$$

- Next: When will this lower bound match Körner-Marton region or Slepian-Wolf region in terms of weighted sum rates?

Application to binary alphabets GF(2): Previous results

Notation: $P(X=0)=x, P(Y=0|X=0)=c, P(Y=1|X=1)=d.$

Previous results: Optimal weighted sum rates $R_1+\lambda R_2$.

When $(c-\frac{1}{2})(d-\frac{1}{2})\le 0 (\Leftrightarrow H(Z)\ge H(Y))$,

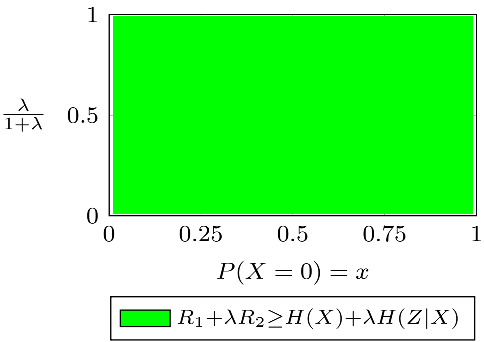


Figure 6: When is Slepian-Wolf region optimal

When $c=d$,

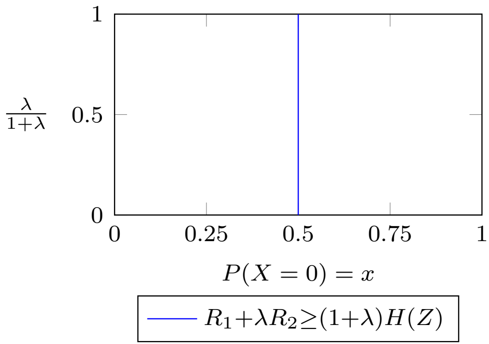


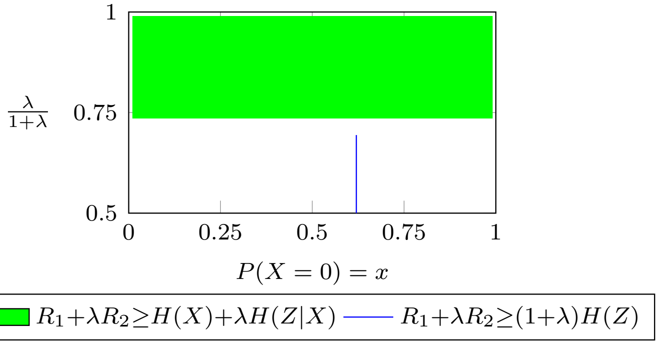
Figure 7: When is Körner-Marton region optimal

Application to binary alphabets GF(2): New Results

Notation: $P(X=0)=x, P(Y=0|X=0)=c, P(Y=1|X=1)=d.$

Our work: When $(c-\frac{1}{2})(d-\frac{1}{2})>0(\Leftrightarrow H(Z)<H(Y))$,

Example 1: $c=0.9, d=0.6$

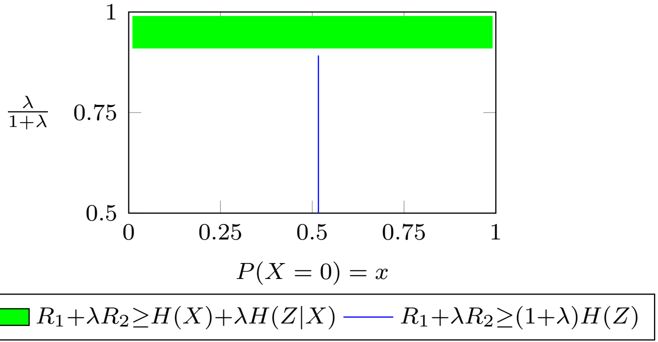


Application to binary alphabets GF(2): New Results

Notation: $P(X=0)=x, P(Y=0|X=0)=c, P(Y=1|X=1)=d.$

Our work: When $(c-\frac{1}{2})(d-\frac{1}{2})>0(\Leftrightarrow H(Z)<H(Y))$,

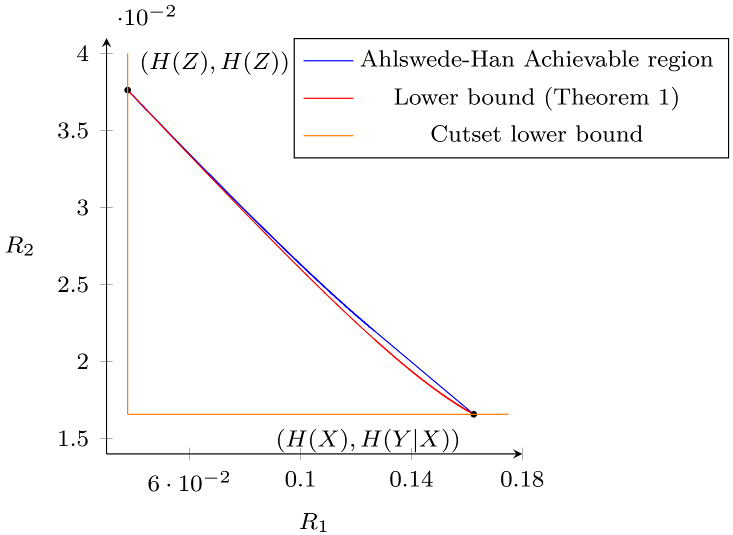
Example 2: $c=0.7, d=0.6$



Comparison of the bounds

In [Ahlswede-Han 1983], Ahlswede and Han chose the following p_{XY} given by

$$p_{XY} = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} 0.003920 & 0.019920 \\ 0.976080 & 0.000080 \end{bmatrix}$$



Outline

Quadratic Gaussian CEO Problem

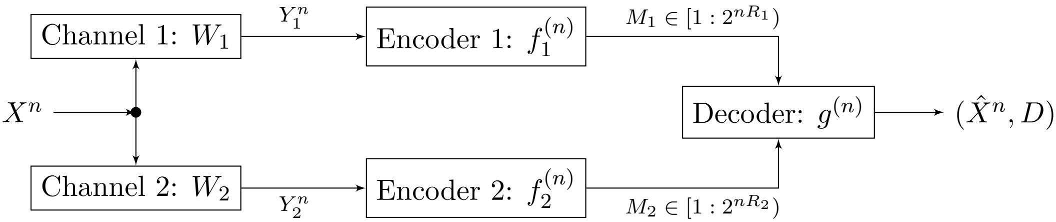


Figure 8: generalized CEO distributed source coding

Distortion criterion: $\limsup_{n\to\infty} \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} d(X_i, \hat{X}_i)\right) \le D$.

- Quadratic Gaussian CEO problem: $d(X_i, \hat{X}_i) = (X_i - \hat{X}_i)^2$; $Y_1 = X + Z_1, Z_1 \perp X, Z_1 \sim N(0, N_1)$; $Y_2 = X + Z_2, Z_2 \perp X, Z_2 \sim N(0, N_2)$.
- Berger-Tung coding scheme [Berger 1978; Tung 1978; Prabhakaran-Tse-Ramachandran 2004] is shown to be optimal by Oohama [Oohama 2005].

Quadratic Gaussian CEO Problem

A lower bound to generalized CEO problem

Consider previous generalized CEO distributed source coding on X, Y_1, Y_2 satisfying that $Y_1 \to X \to Y_2$ with the distortion criterion $\limsup_{n\to\infty} \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} d(X_i, \hat{X}_i)\right) \le D$. For any $\lambda \ge 1$, any achievable triple (R_1, R_2, D) must satisfy that

$$\begin{aligned} R_1 + \lambda R_2 \ge & H(XY_1) + \lambda H(Y_2|X) + (\lambda - 1)\max\left\{H(X|U_1WQ) - H(X|\hat{X}Q), 0\right\} \\ & - H(X|\hat{X}Q) - H(Y_1|XU_1WQ) - \lambda H(Y_2|XU_2WQ) \\ R_2 + \lambda R_1 \ge & H(XY_2) + \lambda H(Y_1|X) + (\lambda - 1)\max\left\{H(X|U_2WQ) - H(X|\hat{X}Q), 0\right\} \\ & - H(X|\hat{X}Q) - H(Y_2|XU_2WQ) - \lambda H(Y_1|XU_1WQ) \end{aligned}$$

subject to the constraints

$$\begin{aligned} & U_1 \leftarrow QWY_1 \leftarrow QWX \to QWY_2 \to U_2 \\ & QW \perp XY_1Y_2 \\ & \hat{X} \leftarrow QWU_1U_2 \to XY_1Y_2 \\ & E[d(X, \hat{X})] \le D. \end{aligned}$$

Proof sketch: $W_i = X^{n/i}, U_{1i} = M_1 Y_1^{i-1}, U_{2i} = M_2 Y_2^{i-1}$. Q is the uniform distribution over $i = 1, \cdots, n$. This is in a similar spirit as the lower bound [Wagner-Anantharam 2008].

Quadratic Gaussian CEO Problem

A lower bound to generalized CEO problem

Consider previous generalized CEO distributed source coding on X, Y_1, Y_2 satisfying that $Y_1 \to X \to Y_2$ with the distortion criterion $\limsup_{n\to\infty} \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} d(X_i, \hat{X}_i)\right) \leq D$. For any $\lambda \geq 1$, any achievable triple (R_1, R_2, D) must satisfy that

$$\begin{aligned}
R_1 + \lambda R_2 \geq & H(XY_1) + \lambda H(Y_2|X) + (\lambda - 1)\max\left\{H(X|U_1WQ) - H(X|\hat{X}Q), 0\right\} \\
& - H(X|\hat{X}Q) - H(Y_1|XU_1WQ) - \lambda H(Y_2|XU_2WQ) \\
R_2 + \lambda R_1 \geq & H(XY_2) + \lambda H(Y_1|X) + (\lambda - 1)\max\left\{H(X|U_2WQ) - H(X|\hat{X}Q), 0\right\} \\
& - H(X|\hat{X}Q) - H(Y_2|XU_2WQ) - \lambda H(Y_1|XU_1WQ)
\end{aligned}$$

subject to the constraints

$$\begin{aligned}
& U_1 \leftarrow QWY_1 \leftarrow QWX \to QWY_2 \to U_2 \\
& QW \perp XY_1Y_2 \\
& \hat{X} \leftarrow QWU_1U_2 \to XY_1Y_2 \\
& E[d(X, \hat{X})] \leq D.
\end{aligned}$$

Rotation techniques in [Geng-Nair 2014] can show that Gaussian U_1, U_2 and constant Q, W minimizes above weighted sum rate lower bound.

Quadratic Gaussian Distributed Source Coding

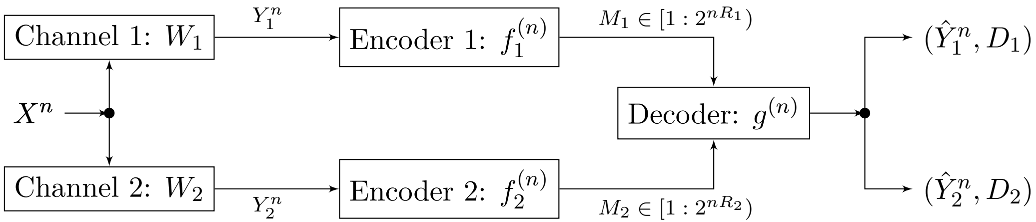


Figure 9: Generalized Distributed Source Coding with auxiliary source structure

$$\limsup_{n\to\infty} \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^n d(Y_{1i}, \hat{Y}_{1i})\right) \le D_1, \limsup_{n\to\infty} \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^n d(Y_{2i}, \hat{Y}_{2i})\right) \le D_2.$$

- Quadratic Gaussian distributed source coding problem: $d(X_i, \hat{X}_i) = (X_i - \hat{X}_i)^2$, $(Y_1, Y_2) \sim N\left(\vec{0}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$.
- Berger-Tung coding scheme [Berger 1978; Tung 1978; Prabhakaran-Tse-Ramachandran 2004] is shown to be optimal by [Wagner-Tavildar-Viswanath 2008].
- Assume there exists some auxiliary source X such that $Y_1 \leftarrow X \rightarrow Y_2$.

Quadratic Gaussian Distributed Source Coding

A lower bound to generalized distributed source coding

Consider the generalized quadratic distributed source coding on 2-DMS (Y_1, Y_2) with the distortion criterion $\limsup_{n\to\infty} \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} d(Y_{1i}, \hat{Y}_{1i})\right) \le D_1$ and $\limsup_{n\to\infty} \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} d(Y_{2i}, \hat{Y}_{2i})\right) \le D_2$. Assume there exists some auxiliary source X such that Y_1 and Y_2 are obtained by passing X through some discrete memoryless channel W_1 and W_2 respectively. For any $\lambda \ge 1$, any achievable triple (R_1, R_2, D_1, D_2) must satisfy that

$$\begin{aligned} R_1 + \lambda R_2 \ge & H(XY_1) + \lambda H(Y_2|X) - H(X|\hat{Y}_1\hat{Y}_2Q) - H(Y_1|XU_1QW) \\ & + (\lambda-1)\max\left\{H(X|U_1QW) - H(X|\hat{Y}_1\hat{Y}_2Q), 0\right\} - \lambda H(Y_2|XU_2QW) \\ R_2 + \lambda R_1 \ge & H(XY_2) + \lambda H(Y_1|X) - H(X|\hat{Y}_1\hat{Y}_2Q) - H(Y_2|XU_2QW) \\ & + (\lambda-1)\max\left\{H(X|U_2QW) - H(X|\hat{Y}_1\hat{Y}_2Q), 0\right\} - \lambda H(Y_1|XU_1QW) \end{aligned}$$

subject to the constraint

$$\begin{aligned} & U_1 \leftarrow QWY_1 \leftarrow QWX \rightarrow QWY_2 \rightarrow U_2 \\ & QW \perp XY_1Y_2 \\ & \hat{Y}_1\hat{Y}_2 \leftarrow QWU_1U_2 \rightarrow XY_1Y_2 \\ & \mathrm{E}[d(Y_1,\hat{Y}_1)] \le D_1, \mathrm{E}[d(Y_2,\hat{Y}_2)] \le D_2 \end{aligned}$$

.

This is also in a similar spirit as the lower bound [Wagner-Anantharam 2008].

Outline

Outline

Motivation 1: Non-convex optimization problem

A non-convex optimization problem

Let $W_{Y|X}$ denote a channel that maps input random variable X with distribution μ_X into the output random variable Y with distribution μ_Y. Consider the non-convex optimization problem, that is, computing the maximum over μ_X of

$$F_\lambda(\mu_X) := \lambda H(X) - H(Y)$$

where $0 \leq \lambda \leq 1$ is some fixed constant.

Motivation 1: Non-convex optimization problem

A non-convex optimization problem

Let $W_{Y|X}$ denote a channel that maps input random variable X with distribution μ_X into the output random variable Y with distribution μ_Y. Consider the non-convex optimization problem, that is, computing the maximum over μ_X of

$$F_\lambda(\mu_X) := \lambda H(X) - H(Y)$$

where $0 \leq \lambda \leq 1$ is some fixed constant.

Mrs Gerber's Lemma [Wyner-Ziv 1973]

When the channel $W_{Y|X}$ is the binary symmetric channel with flipping probability p, under the reparametrization of μ_X, defined by $\mu_X(u) = H_2^{-1}(u)$,

$$F_\lambda(u) = \lambda u - H_2(p * H_2^{-1}(u)).$$

is concave in u for any λ. Here $a * b := a(1-b) + (1-a)b$.

Question: Is there an analogous result in the additive Gaussian noise channel setting, that is, $Y = X + W$ where $W \sim N(0, \sigma^2)$?

Motivation 1: Non-convex optimization problem

- In the continuous world, to make moving direction limited in one dimension, we assume that

$$
\begin{aligned}
X_t :=& X + \sqrt{t}Z, t > 0, Z \sim N(0,1) \\
Y_t :=& X_t + W, W \sim N(0, \sigma^2)
\end{aligned}
$$

Motivation 1: Non-convex optimization problem

- In the continuous world, to make moving direction limited in one dimension, we assume that

$$X_t := X + \sqrt{t}Z, t > 0, Z \sim N(0, 1)$$
$$Y_t := X_t + W, W \sim N(0, \sigma^2)$$

- μ_t^X: the probability density function of $X + \sqrt{t}Z$. μ_t^X satisfies the heat flow equation: $\frac{\partial \mu_t^X}{\partial t} = \frac{1}{2}\frac{\partial^2 \mu_t^X}{\partial x^2}$ with initial condtion $\mu_0^X(x) \equiv f(x)$, where $f(x)$ is the probability density function of X.
- The differential entropy $h(X) := -\int_{\mathbb{R}} f(x) \ln f(x) dx$.

Motivation 1: Non-convex optimization problem

- In the continuous world, to make moving direction limited in one dimension, we assume that

$$\begin{aligned} X_t :=& X + \sqrt{t}Z, t > 0, Z \sim N(0,1) \\ Y_t :=& X_t + W, W \sim N(0,\sigma^2) \end{aligned}$$

- μ_t^X: the probability density function of $X + \sqrt{t}Z$. μ_t^X satisfies the heat flow equation: $\frac{\partial \mu_t^X}{\partial t} = \frac{1}{2}\frac{\partial^2 \mu_t^X}{\partial x^2}$ with initial condtion $\mu_0^X(x) \equiv f(x)$, where $f(x)$ is the probability density function of X.
- The differential entropy $h(X) := -\int_{\mathbb{R}} f(x) \ln f(x) dx$.
- Want a parametrization $t = \phi(u)$ such that $h(X + \sqrt{\phi(u)}Z)$ is linear in u and the output entropy, $h(\mu_Y) = h(X + \sqrt{\phi(u)}Z + W)$ is convex in u.
  A bit of algebra immediately shows that this question is equivalent to asking whether the Fisher information

$$I(\mu_t^X) := \int_{\mathbb{R}} \left(\frac{\partial}{\partial x}\mu_t^X(x)\right)^2 \mu_t^X(x) dx$$

  is log-convex in t, for all random variables X.

Motivation 2: Completely monontone and log-convexity

- Let X be a random variable with a finite variance P. Let $g_X^{(0)}(t) := h(\mu_t^X)$, $g_X^{(k)}(t) := \frac{\partial^k}{\partial t^k} h(\mu_t^X)$. De Bruijin's identity tells that

$$I(\mu_t^X) = 2\frac{\partial}{\partial t} h(\mu_t^X) = 2g_X^{(1)}(t)$$

- Let $Z \sim N(0, P)$. In Section 12 of [McKean 1966], McKean observes that for any $t \geq 0$, $g_Z^{(0)}(t) \geq g_X^{(0)}(t) \geq 0$, $g_Z^{(1)}(t) \leq g_X^{(1)}(t) \leq 0$, and $g_Z^{(2)}(t) \geq g_X^{(2)}(t) \geq 0$. Therefore he conjectured that

Mckean's conjecture [McKean 1966]

$$\forall X, \forall t \geq 0, (-1)^k g_Z^{(k)}(t) \geq (-1)^k g_X^{(k)}(t) \geq 0, \forall k \geq 3.$$

Motivation 2: Completely monontone and log-convexity

- Let X be a random variable with a finite variance P. Let $g_X^{(0)}(t) := h(\mu_t^X)$, $g_X^{(k)}(t) := \frac{\partial^k}{\partial t^k} h(\mu_t^X)$. De Bruijin's identity tells that

$$I(\mu_t^X) = 2\frac{\partial}{\partial t} h(\mu_t^X) = 2g_X^{(1)}(t)$$

- Let $Z \sim N(0, P)$. In Section 12 of [McKean 1966], McKean observes that for any $t \geq 0$, $g_Z^{(0)}(t) \geq g_X^{(0)}(t) \geq 0$, $g_Z^{(1)}(t) \leq g_X^{(1)}(t) \leq 0$, and $g_Z^{(2)}(t) \geq g_X^{(2)}(t) \geq 0$. Therefore he conjectured that

Mckean's conjecture [McKean 1966]

$$\forall X, \forall t \geq 0, (-1)^k g_Z^{(k)}(t) \geq (-1)^k g_X^{(k)}(t) \geq 0, \forall k \geq 3.$$

- In [Cheng-Geng 2015], $g_X^{(3)}(t) \geq 0$, and $g_X^{(4)}(t) \leq 0$ for any $t \geq 0$ are established. They made a <u>weaker</u> conjecutre that $(-1)^k g_X^{(k)}(t) \geq 0$. In other words, $I(\mu_t^X) = 2g_X^{(1)}(t)$ is a completely monotone function of t, for all X.

Motivation 2: Completely monontone and log-convexity

An alternate characterization of completely monotone function:

Bernstein's theorem

Let $g(t) : [0, \infty) \to [0, \infty)$ be a continuous and infinitely differentiable function. The following are equivalent:

- g is completely monotone: $\forall n \in \mathbb{N}, \forall t > 0, (-1)^n g^{(n)}(t) \geq 0$;
- g is the Laplace transform of a finite Borel measure ν in $\mathbb{R}_+$:

$$\forall x \in \mathbb{R}_+, g(x) = \int_0^\infty e^{-xt} d\nu(t).$$

- Via this theorem, one can show that any completely monotone function $g(t)$ is log-convex with respect to t, see [Fink 1982].
- If $I(\mu_t^X)$ is a completely monotone function with respect to t, then $\ln I(\mu_t^X)$ is convex with respect to t.

This part: We established that $I(\mu_t^X)$ is log-convex in t, thus resolving affirmatively Conjecture 2 in [Cheng-Geng 2015].

Outline

1. Hypercontractive Region Evaluation
   - Introduction to Hypercontractivity
   - Main Results on Hypercontractivity
2. Lower Bounds on Distributed Source Coding
   - Körner and Marton's Modulo Two Sum Problem
   - Alternative Proofs to Quadratic Gaussian CEO Problem and Distributed Source Coding Problem
3. Log-Convexity of Fisher Information
   - Motivations
   - Proof to log-convexity of Fisher information

Notations and Previous Results

Proof techniques:

- The ideas and techiniques starts from a short proof to the ”concavity of entropy power” (Costa’s EPI) by C. Villani [Villani-2000], which is in turn motivated by calculations of Bakry and Emery [Bakry-Emery 1985].
- Later, [Cheng-Geng 2015] and [Zhang-Anantharam-Geng 2018] followed the work and developed these tools and notations.

Notations and Previous Results

Proof techniques:

- The ideas and techiniques starts from a short proof to the "concavity of entropy power" (Costa's EPI) by C. Villani [Villani-2000], which is in turn motivated by calculations of Bakry and Emery [Bakry-Emery 1985].
- Later, [Cheng-Geng 2015] and [Zhang-Anantharam-Geng 2018] followed the work and developed these tools and notations.

Notations:

- $v(x) := \ln \mu_t^X(x), t > 0$, and $v_k(x) := \frac{\partial^k \ln \mu_t^X(x)}{\partial x^k}, k \in \mathbb{Z}_+$,
- $\langle \varphi \rangle := \int_{\mathbb{R}} \varphi \mu_t^X(x) dx$

Key idea: Under these notations, our problems can be rewritten as inequalities in terms of $\langle \prod_{i=1}^r v_{k_i}^{m_i} \rangle$, where $r, m_i, k_i \in \mathbb{Z}_+$.

Notations and Previous Results

Proof techniques:

- The ideas and techiniques starts from a short proof to the "concavity of entropy power" (Costa's EPI) by C. Villani [Villani-2000], which is in turn motivated by calculations of Bakry and Emery [Bakry-Emery 1985].
- Later, [Cheng-Geng 2015] and [Zhang-Anantharam-Geng 2018] followed the work and developed these tools and notations.

Notations:

- $v(x) := \ln \mu_t^X(x), t > 0$, and $v_k(x) := \frac{\partial^k \ln \mu_t^X(x)}{\partial x^k}, k \in \mathbb{Z}_+$,
- $\langle \varphi \rangle := \int_{\mathbb{R}} \varphi \mu_t^X(x) dx$

Key idea: Under these notations, our problems can be rewritten as inequalities in terms of $\langle \prod_{i=1}^r v_{k_i}^{m_i} \rangle$, where $r, m_i, k_i \in \mathbb{Z}_+$.

Integration by parts formula, Lemma 3 in [Zhang-Anantharam-Geng 2018]

For $k \geq 2$, let $\varphi(x)$ be any "reasonably smooth" function.

$$\langle \varphi v_k + \varphi v_1 v_{k-1} + \frac{\partial \varphi}{\partial x} v_{k-1} \rangle = 0.$$

- φ could be chosen in the form of $\prod_{i=1}^r v_{k_i}^{m_i}(x)$.
- This gives the linear dependence relationships among the terms $\langle \prod_{i=1}^r v_{k_i}^{m_i} \rangle$.

Villani's proof to Costa's EPI

Fisher information and its derivatives

For $t > 0$, Fisher information $I(\mu_t^X)$ and its derivatives up to second order are:

$$\begin{aligned} I(\mu_t^X) &= \langle v_1^2 \rangle, \\ \frac{d}{dt} I(\mu_t^X) &= -\langle v_2^2 \rangle, \\ \frac{d^2}{dt^2} I(\mu_t^X) &= \langle v_3^2 + 2v_1^2 v_2^2 + 4 v_1 v_2 v_3 \rangle. \end{aligned}$$

Villani's proof to Costa's EPI

Fisher information and its derivatives

For $t > 0$, Fisher information $I(\mu_t^X)$ and its derivatives up to second order are:

$$\begin{aligned} I(\mu_t^X) &= \langle v_1^2 \rangle, \\ \frac{d}{dt} I(\mu_t^X) &= -\langle v_2^2 \rangle, \\ \frac{d^2}{dt^2} I(\mu_t^X) &= \langle v_3^2 + 2v_1^2 v_2^2 + 4 v_1 v_2 v_3 \rangle. \end{aligned}$$

Costa's EPI in scalar case, [Costa 1985]

For any random variable X and $Z \perp X, Z \sim N(0,1)$, $e^{2h(X+\sqrt{t}Z)}$ is concave in $t \geq 0$.

Villani's proof to Costa's EPI

Fisher information and its derivatives

For $t > 0$, Fisher information $I(\mu_t^X)$ and its derivatives up to second order are:

$$\begin{aligned} I(\mu_t^X) &= \langle v_1^2 \rangle, \\ \frac{d}{dt} I(\mu_t^X) &= -\langle v_2^2 \rangle, \\ \frac{d^2}{dt^2} I(\mu_t^X) &= \langle v_3^2 + 2v_1^2 v_2^2 + 4 v_1 v_2 v_3 \rangle. \end{aligned}$$

Costa's EPI in scalar case, [Costa 1985]

For any random variable X and $Z \perp X, Z \sim N(0,1)$, $e^{2h(X+\sqrt{t}Z)}$ is concave in $t \geq 0$.

Proof [Villani-2000]:

Computing second derivative of $e^{2h(X+\sqrt{t}Z)}$ with respect to t yields:

$$e^{2h(X+\sqrt{t}Z)} \left[-\langle v_2^2 \rangle + \langle v_1^2 \rangle^2 \right] \leq 0 \overset{\langle v_2 + v_1^2 \rangle = 0}{\Longleftrightarrow} -\langle v_2^2 \rangle + \langle v_2 \rangle^2 \leq 0$$

Main result

Log-convexity of Fisher information in scalar case [Ledoux-Nair-Wang 2020]

Let X be a random variable on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $\mathbb{R}$, and $Z \perp X, Z \sim N(0,1)$. Consider $X_t := X + \sqrt{t}Z, t > 0$, with probability density function $\mu_t^X(x)$ with respect to the Lebesgue measure on $\mathbb{R}$. The Fisher information of X_t is log-convex in t, i.e.

$$\ln I(\mu_t^X) = \ln \int_{\mathbb{R}} \left(\frac{\partial}{\partial t} \ln \mu_t^X(x) \right)^2 \mu_t^X(x) dx$$

is convex in t.

Main result

Log-convexity of Fisher information in scalar case [Ledoux-Nair-Wang 2020]

Let X be a random variable on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $\mathbb{R}$, and $Z \perp X, Z \sim N(0,1)$. Consider $X_t := X + \sqrt{t}Z, t > 0$, with probability density function $\mu_t^X(x)$ with respect to the Lebesgue measure on $\mathbb{R}$. The Fisher information of X_t is log-convex in t, i.e.

$$\ln I(\mu_t^X) = \ln \int_{\mathbb{R}} \left(\frac{\partial}{\partial t} \ln \mu_t^X(x) \right)^2 \mu_t^X(x) dx$$

is convex in t.

Proof sketch: Log-convexity of Fisher information is equivalent to

$$\left(\frac{d}{dt} I(\mu_t^X) \right)^2 \leq I(\mu_t^X) \frac{d^2}{dt^2} I(\mu_t^X).$$

In terms of $\langle \prod_{i=1}^r v_{k_i}^{m_i} \rangle$, it is equivalent to showing

$$\langle v_2^2 \rangle^2 \leq \langle v_1^2 \rangle \langle v_3^2 + 2v_1^2 v_2^2 + 4v_1 v_2 v_3 \rangle.$$

Proof continued:

In integration by parts formula, choosing $k = 2, \varphi = v_2$ and $k = 2, \varphi = v_1^2$ will lead to:

$$\begin{cases} \langle v_2^2 \rangle = -\langle v_1^2 v_2 + v_1 v_3 \rangle \\ \langle v_1^2 v_2 \rangle = -\frac{1}{3} \langle v_1^4 \rangle. \end{cases}$$

$$\Rightarrow \langle v_2^2 \rangle = -\langle v_1 v_3 + \alpha v_1^2 v_2 - \frac{1-\alpha}{3} v_1^4 \rangle, \forall \alpha \in \mathbb{R}$$

$$\Rightarrow \langle v_2^2 \rangle = -\langle v_1 (v_3 + \alpha v_1 v_2 - \frac{1-\alpha}{3} v_1^3) \rangle, \forall \alpha \in \mathbb{R}$$

$$\overset{\text{Cauchy-Schwartz}}{\Longrightarrow} \langle v_2^2 \rangle^2 \leq \langle v_1^2 \rangle \langle (v_3 + \alpha v_1 v_2 - \frac{1-\alpha}{3} v_1^3)^2 \rangle, \forall \alpha \in \mathbb{R}.$$

Proof continued:

In integration by parts formula, choosing $k = 2, \varphi = v_2$ and $k = 2, \varphi = v_1^2$ will lead to:

$$\begin{cases} \langle v_2^2 \rangle = -\langle v_1^2 v_2 + v_1 v_3 \rangle \\ \langle v_1^2 v_2 \rangle = -\frac{1}{3}\langle v_1^4 \rangle. \end{cases}$$

$$\Rightarrow \langle v_2^2 \rangle = -\langle v_1 v_3 + \alpha v_1^2 v_2 - \frac{1-\alpha}{3} v_1^4 \rangle, \forall \alpha \in \mathbb{R}$$

$$\Rightarrow \langle v_2^2 \rangle = -\langle v_1 (v_3 + \alpha v_1 v_2 - \frac{1-\alpha}{3} v_1^3) \rangle, \forall \alpha \in \mathbb{R}$$

$$\xRightarrow{\text{Cauchy-Schwartz}} \langle v_2^2 \rangle^2 \leq \langle v_1^2 \rangle \langle (v_3 + \alpha v_1 v_2 - \frac{1-\alpha}{3} v_1^3)^2 \rangle, \forall \alpha \in \mathbb{R}.$$

To show

$$\langle v_2^2 \rangle^2 \leq \langle v_1^2 \rangle \langle v_3^2 + 2 v_1^2 v_2^2 + 4 v_1 v_2 v_3 \rangle.$$

Suffices to show that

$$\langle (v_3 + \alpha v_1 v_2 - \frac{1-\alpha}{3} v_1^3)^2 \rangle \leq \langle v_3^2 + 2 v_1^2 v_2^2 + 4 v_1 v_2 v_3 \rangle$$

holds for some $\alpha \in \mathbb{R}$.
We prove it for $\alpha = 2$ by integration by parts formula and some calculation.

Open: Generalization of log-convexity to higher dimensions

- One clear question that is definitely worth addressing is to determine whether the log-convexity of Fisher information along the heat flow also holds for random vectors.

- In particular we ask, whether

$$\left(\frac{d^3h(X^n+\sqrt{t}Z^n)}{dt^3}\right)\left(\frac{dh(X^n+\sqrt{t}Z^n)}{dt}\right) \geq \left(\frac{d^2h(X^n+\sqrt{t}Z^n)}{dt^2}\right)^2$$

  where X^n and $Z^n(\sim \mathcal{N}(0, I_n))$ are independent random vectors taking values in $\mathbb{R}^n$.

- If X^n has independent components, then an application of the Cauchy-Schwartz inequality immediately implies affirmatively the inequality above.

Summary: Take-aways

In this talk, we studied several non-convex optimization problems:

- Hypercontractive region evaluation for the binary erasure channel
- Distributed source coding
  - Modulo sum problem: Obtained improved lower bounds
  - Gaussian setting: Alternate proofs of optimality
- Log-convexity of Fisher information: Resolved the log-convexity of Fisher Information conjecture.

Summary: Take-aways

In this talk, we studied several non-convex optimization problems:

- Hypercontractive region evaluation for the binary erasure channel
- Distributed source coding
  - Modulo sum problem: Obtained improved lower bounds
  - Gaussian setting: Alternate proofs of optimality
- Log-convexity of Fisher information: Resolved the log-convexity of Fisher Information conjecture.

Thank you! Any questions are welcome!

Supplementary slides to Introduction: Channel coding

$$M \in [1:2^{nR}] \longrightarrow \boxed{\text{sender: } f^{(n)}} \xrightarrow{X^n} \boxed{\text{DMC: } W_{Y|X}} \xrightarrow{Y^n} \boxed{\text{receiver: } g^{(n)}} \longrightarrow \hat{M} \in [1:2^{nR}]$$

Figure 10: Point-to-point communication channel model

Theorem (Shannon 1948)

The capacity of a DMC $W_{Y|X}$ is given by

$$\mathscr{C}(W_{Y|X}) = \{R \geq 0 : R \leq \max_{p_X} I(X;Y)\}. \tag{2}$$

Supplementary slides to Introduction: Channel coding

$$M \in [1:2^{nR}] \longrightarrow \boxed{\text{sender: } f^{(n)}} \xrightarrow{X^n} \boxed{\text{DMC: } W_{Y|X}} \xrightarrow{Y^n} \boxed{\text{receiver: } g^{(n)}} \longrightarrow \hat{M} \in [1:2^{nR}]$$

Figure 10: Point-to-point communication channel model

- $\{R \geq 0 : R < \max_{p_X} I(X;Y)\}$ is an achievable rate region for DMC $W_{Y|X}$;
- Optimality: Suffices to show for any DMC $W_{Y|X}$ and its product $W_{Y|X}^{\otimes 2}$

$$\max_{p_{X_1 X_2}} I(X_1 X_2; Y_1 Y_2) = 2 \max_{p_X} I(X;Y)$$

- It belongs to the non-convex functional family: Let $c > 0$,

$$\max_{p_X} I(X;Y) = \lim_{c \to \infty} \max_{p_{XY}} I(X;Y) - c \sum_{x \in \mathcal{X}} p_X(x) D(p_{Y|X=x} || W_{Y|X=x})$$

Supplementary slides to Hypercontractivity: $g(x)$ has only one stationary point between $(0, \frac{1}{2})$

- Recall $g(x) := \max_{r,s:r\in[0,x],s\in[0,1-x]} f(x,r,s)$.
  Hence any stationary point of $g(x)$ will be a stationary point of $f(x,r,s)$.

- Let $y = \frac{2(x-r)}{1-r-s}$; we know from forward hypercontractivity proof

  the stationary points of $f(x,r,s)$ are in 1-1 correspondence with the roots of

$$\frac{1-\epsilon}{\epsilon} y^{\lambda_2'-\lambda_1} + y^{1-\lambda_1} = \frac{1-\epsilon}{\epsilon}(2-y)^{\lambda_2'-\lambda_1} + (2-y)^{1-\lambda_1}.$$

  Hence suffices to show that there is exactly *one root* of above equation for $y \in (0,1)$.
- This can be shown by taylor expansion and a key observation on the sign change patterns of the coefficients.

Supplementary slides to modulo sum problem: Single-letterize the lower bound on

Lemma 3.1

Let $\lambda \geq 1$ and let (X^n, Y^n) be i.i.d distributed according to $p(x,y)$ where X, Y take values in a finite field. Let Z^n be obtained as $Z_i = X_i \oplus Y_i, i = 1, .., n$, i.e. the component-wise modulo sum on the field. Then for any $\lambda \geq 1$ the following holds:

$$\begin{aligned} &\min_{\hat{U}:\hat{U}\to X^n \to Y^n} \lambda H(Z^n|\hat{U}) - H(Y^n|\hat{U}) \\ &= n \left(\min_{U:U\to X\to Y} \lambda H(Z|U) - H(Y|U) \right). \end{aligned}$$

Supplementary slides to modulo sum problem: Single-letterize the lower bound on

Lemma 3.1

Let $\lambda \geq 1$ and let (X^n, Y^n) be i.i.d distributed according to $p(x,y)$ where X, Y take values in a finite field. Let Z^n be obtained as $Z_i = X_i \oplus Y_i, i = 1, .., n$, i.e. the component-wise modulo sum on the field. Then for any $\lambda \geq 1$ the following holds:

$$\begin{aligned}
&\min_{\hat{U}:\hat{U}\to X^n\to Y^n} \lambda H(Z^n|\hat{U}) - H(Y^n|\hat{U}) \\
&= n\left(\min_{U:U\to X\to Y} \lambda H(Z|U) - H(Y|U)\right).
\end{aligned}$$

Proof sketch

- Taking i.i.d. copies of the minimizer of the right hand side, left hand side is at most the value of right hand side.

Supplementary slides to modulo sum problem: Single-letterize the lower bound on

Lemma 3.1

Let $\lambda \geq 1$ and let (X^n, Y^n) be i.i.d distributed according to $p(x, y)$ where X, Y take values in a finite field. Let Z^n be obtained as $Z_i = X_i \oplus Y_i, i = 1, .., n$, i.e. the component-wise modulo sum on the field. Then for any $\lambda \geq 1$ the following holds:

$$\min_{\hat{U}:\hat{U}\to X^n \to Y^n} \lambda H(Z^n|\hat{U}) - H(Y^n|\hat{U})$$
$$= n\left(\min_{U:U\to X\to Y} \lambda H(Z|U) - H(Y|U)\right).$$

Proof sketch

- Taking i.i.d. copies of the minimizer of the right hand side, left hand side is at most the value of right hand side.
- The other direction follows from Markov chain $(\hat{U}, Y_{i+1}^n, Z^{i-1}) \to X_i \to (Y_i, Z_i)$ and Körner-Marton identity.

Supplementary slides to modulo sum problem: Conditions for lower bound to be tight

Lemma 3.2

The lower bound for the weighted sum-rate $R_1 + \lambda R_2$, for $\lambda \geq 1$ given in Theorem 1 is optimal, i.e. matches the weighted sum-rate of the optimal rate region, if either of the following conditions hold:

(i) $\mathfrak{C}_{\mu(x)}[H(Y) - \lambda H(Z)]\big|_{p(x)} = H(Y) - \lambda H(Z)$ and Y is independent of Z,

(ii) $\mathfrak{C}_{\mu(x)}[H(Y) - \lambda H(Z)]\big|_{p(x)} = H(Y|X) - \lambda H(Z|X)$.

Further if condition (i) holds for some $\lambda_1 > 1$, then it will also hold for $1 \leq \lambda \leq \lambda_1$; and if condition (ii) holds for some $\lambda_2 \geq 1$, then it will also hold for $\lambda \geq \lambda_2$.

Supplementary slides to modulo sum problem: Conditions for lower bound to be tight

Lemma 3.2

The lower bound for the weighted sum-rate $R_1 + \lambda R_2$, for $\lambda \geq 1$ given in Theorem 1 is optimal, i.e. matches the weighted sum-rate of the optimal rate region, if either of the following conditions hold:

(i) $\mathfrak{C}_{\mu(x)}[H(Y) - \lambda H(Z)]\big|_{p(x)} = H(Y) - \lambda H(Z)$ and Y is independent of Z,

(ii) $\mathfrak{C}_{\mu(x)}[H(Y) - \lambda H(Z)]\big|_{p(x)} = H(Y|X) - \lambda H(Z|X)$.

Further if condition (i) holds for some $\lambda_1 > 1$, then it will also hold for $1 \leq \lambda \leq \lambda_1$; and if condition (ii) holds for some $\lambda_2 \geq 1$, then it will also hold for $\lambda \geq \lambda_2$.

Remark:

A relatively easier condition to verify is the convexity of $H(Y) - \lambda H(Z)$ with respect to the distribution of X.

Supplementary slides to Körner-Marton Problem: Application to binary alphabets GF(2)

Notation: We will parameterize the space of distributions over pairs of binary alphabets, $p(x, y)$ as follows:
$P(X = 0) = x, P(Y = 0|X = 0) = c, P(Y = 1|X = 1) = d.$

Supplementary slides to Körner-Marton Problem: Application to binary alphabets GF(2)

Notation: We will parameterize the space of distributions over pairs of binary alphabets, $p(x, y)$ as follows:
$P(X = 0) = x, P(Y = 0|X = 0) = c, P(Y = 1|X = 1) = d.$

Proposition 3.1: Optimality of Slepian-Wolf region

The optimal weighted sum-rate of the capacity region is given by the Slepian Wolf region if any of the following conditions hold:

(i) For any λ, if $(c - \frac{1}{2})(d - \frac{1}{2}) \leq 0$, or

(ii) $\lambda \geq \left(\frac{c-\bar{d}}{c-d}\right)^2$, $c \neq d$, and $(c - \frac{1}{2})(d - \frac{1}{2}) > 0$.

where $\bar{d} = 1 - d$.

Supplementary slides to Körner-Marton Problem: Application to binary alphabets GF(2)

Notation: We will parameterize the space of distributions over pairs of binary alphabets, $p(x, y)$ as follows:
$P(X = 0) = x, P(Y = 0|X = 0) = c, P(Y = 1|X = 1) = d.$

Proposition 3.1: Optimality of Slepian-Wolf region

The optimal weighted sum-rate of the capacity region is given by the Slepian Wolf region if any of the following conditions hold:

(i) For any λ, if $(c - \frac{1}{2})(d - \frac{1}{2}) \leq 0$, or

(ii) $\lambda \geq \left(\frac{c-\bar{d}}{c-d}\right)^2$, $c \neq d$, and $(c - \frac{1}{2})(d - \frac{1}{2}) > 0$.

where $\bar{d} = 1 - d$.

Remarks:

(i) The condition (i) above is already known and stated as exercise 16.23 page 390 of Csiszár and Körner's book. One can verify that that $H(Z) \geq H(Y)$ is equivalent to $(c - \frac{1}{2})(d - \frac{1}{2}) \leq 0$.

(ii) Note that an equivalent proposition can also be stated for the alternate parameterization: $P(Y = 0) = y, P(X = 0|Y = 0) = \hat{c}, P(X = 1|Y = 1) = \hat{d}$.

Supplementary slides to modulo sum problem: Application to binary alphabets GF(2)

Proposition 3.2: Optimality of Körner-Marton region

Let $P(X=0)=x, P(Y=0|X=0)=c, P(Y=1|X=1)=d$ where $x=\frac{\sqrt{d\bar{d}}}{\sqrt{d\bar{d}}+\sqrt{c\bar{c}}}$. The optimal weighted sum-rate of the capacity region is given by the Körner-Marton region, i.e. using linear codes, if any of the following conditions hold:

(i) For any λ, if $c=d$, or

(ii) $1 \leq \lambda \leq \lambda_1$, $c \neq d$, and $(c-\frac{1}{2})(d-\frac{1}{2})>0$, where λ_1 is the larger root of the quadratic equation

$$\lambda^2(c-d)^2+\lambda\big(2(c-d)(c-\bar{d})-4d\bar{d}(c-\bar{c})^2\big)+(c-\bar{d})^2=0.$$

where $\bar{d}=1-d, \bar{c}=1-c$.

Supplementary slides to modulo sum problem: Application to binary alphabets GF(2)

Proposition 3.2: Optimality of Körner-Marton region

Let $P(X=0)=x, P(Y=0|X=0)=c, P(Y=1|X=1)=d$ where $x=\frac{\sqrt{d\bar{d}}}{\sqrt{d\bar{d}}+\sqrt{c\bar{c}}}$. The optimal weighted sum-rate of the capacity region is given by the Körner-Marton region, i.e. using linear codes, if any of the following conditions hold:

- (i) For any λ, if $c=d$, or
- (ii) $1\le\lambda\le\lambda_1$, $c\neq d$, and $(c-\frac{1}{2})(d-\frac{1}{2})>0$, where λ_1 is the larger root of the quadratic equation

$$\lambda^2(c-d)^2+\lambda\big(2(c-d)(c-\bar{d})-4d\bar{d}(c-\bar{c})^2\big)+(c-\bar{d})^2=0.$$

where $\bar{d}=1-d, \bar{c}=1-c$.

Remarks:

- (i) As long as $(c-\frac{1}{2})(d-\frac{1}{2})>0$ and $x=\frac{\sqrt{d\bar{d}}}{\sqrt{d\bar{d}}+\sqrt{c\bar{c}}}$, the optimal sum-rate will be given by the Körner-Marton region, i.e. using linear codes.
- (ii) An equivalent Proposition can also be stated for the alternate parameterization: $P(Y=0)=y, P(X=0|Y=0)=\hat{c}, P(X=1|Y=1)=\hat{d}$.

Supplementary slides to modulo sum problem: to higher alphabet fields

Example 1

For $GF(3)$, one instance of $p(x, y)$ satisfying that Z is independent of Y and $\mathfrak{C}_{\mu(x)}[H(Y) - H(Z)]\big|_{p(x)} = H(Y) - H(Z)$ is given by the following distribution:

$$p(x, y) = \begin{bmatrix} 0.08 & 0.06 & 0.18 \\ 0.08 & 0.18 & 0.06 \\ 0.24 & 0.06 & 0.06 \end{bmatrix}$$

Supplementary slides to modulo sum problem: to higher alphabet fields

Example 1

For $GF(3)$, one instance of $p(x, y)$ satisfying that Z is independent of Y and $\mathfrak{C}_{\mu(x)}[H(Y) - H(Z)]\big|_{p(x)} = H(Y) - H(Z)$ is given by the following distribution:

$$p(x, y) = \begin{bmatrix} 0.08 & 0.06 & 0.18 \\ 0.08 & 0.18 & 0.06 \\ 0.24 & 0.06 & 0.06 \end{bmatrix}$$

Example 2

One instance of $p(x, y)$ satisfying $\mathfrak{C}_{\mu(x)}[H(Y) - H(Z)]\big|_{p(x)} = H(Y|X) - H(Z|X)$ is given by the following distribution:

$$p(x, y) = \begin{bmatrix} 0.02 & 0.02 & 0.48 \\ 0.02 & 0.06 & 0.16 \\ 0.06 & 0.02 & 0.16 \end{bmatrix}$$

Supplementary slides to Log-convexity of Fisher Information: Proof for $\alpha = 2$

Expanding what we want to show yields

$$\langle (2-\alpha^2)v_1^2v_2^2 + (4-2\alpha)v_1v_2v_3 - \frac{1}{9}(1-\alpha)^2v_1^6 + \frac{2}{3}(1-\alpha)v_1^3v_3 + \frac{2}{3}\alpha(1-\alpha)v_1^4v_2\rangle \geq 0.$$

Supplementary slides to Log-convexity of Fisher Information: Proof for $\alpha = 2$

Expanding what we want to show yields

$$\langle (2-\alpha^2)v_1^2v_2^2 + (4-2\alpha)v_1v_2v_3 - \frac{1}{9}(1-\alpha)^2v_1^6 + \frac{2}{3}(1-\alpha)v_1^3v_3 + \frac{2}{3}\alpha(1-\alpha)v_1^4v_2\rangle \geq 0.$$

In integration by parts formula, choosing $k=3, \varphi = v_1^3$ and that $k=2, \varphi = v_1^4$ gives

$$\begin{aligned} \langle v_1^3v_3 + v_2v_1^4 + 3v_1^2v_2^2\rangle &= 0 \\ \langle v_1^6 + 5v_1^4v_2\rangle &= 0. \end{aligned}$$

Supplementary slides to Log-convexity of Fisher Information: Proof for $\alpha = 2$

Expanding what we want to show yields

$$\langle (2-\alpha^2)v_1^2v_2^2 + (4-2\alpha)v_1v_2v_3 - \frac{1}{9}(1-\alpha)^2v_1^6 + \frac{2}{3}(1-\alpha)v_1^3v_3 + \frac{2}{3}\alpha(1-\alpha)v_1^4v_2\rangle \geq 0.$$

In integration by parts formula, choosing $k=3, \varphi = v_1^3$ and that $k=2, \varphi = v_1^4$ gives

$$\begin{aligned} \langle v_1^3v_3 + v_2v_1^4 + 3v_1^2v_2^2\rangle &= 0 \\ \langle v_1^6 + 5v_1^4v_2\rangle &= 0. \end{aligned}$$

Proving above inequality for some $\alpha \in \mathbb{R}$ is equivalent to proving the following inequality

$$\begin{aligned} \langle (2-\alpha^2)v_1^2v_2^2 + (4-2\alpha)v_1v_2v_3 - \frac{1}{9}(1-\alpha)^2v_1^6 + \frac{2}{3}(1-\alpha)v_1^3v_3 + \frac{2}{3}\alpha(1-\alpha)v_1^4v_2\rangle \\ +\beta\langle v_1^3v_3 + v_2v_1^4 + 3v_1^2v_2^2\rangle + \gamma\langle v_1^6 + 5v_1^4v_2\rangle \geq 0 \end{aligned}$$

for some $\alpha, \beta, \gamma \in \mathbb{R}$.

We successively choose $\alpha = 2$, $\beta = \frac{2}{3}$, and $\gamma = \frac{2}{15}$. Above reduces to $\frac{1}{45}\langle v_1^6\rangle \geq 0$.

Supplementary slides to Log-convexity of fisher information Open Problem: Generalization of convexity of the output entropy

- Consider a channel given by

$$Y^m = AX^n + Z^m,$$

  where A is an $m \times n$ (channel-gain) matrix, X^n is the input, and $Z^m(\sim N(0, I_m))$ is the additive Gaussian noise.
- What are the flows in the space of input distributions, say characterized by X_t^n, where $h(X_t^n)$ is linear in t and $h(Y_t^m)$ is convex in t?
- An interesting such flow exists in the space of Gaussian vectors [Kubo-Andô 1980]. Let $X_0^n \sim N(0, K_0)$ and $X_1^n \sim N(0, K_1)$. Define

$$K_t = K_0^{\frac{1}{2}} \left(K_0^{-\frac{1}{2}} K_1 K_0^{-\frac{1}{2}} \right)^t K_0^{\frac{1}{2}},$$

  and $X_t^n \sim N(0, K_t)$. Then $h(X_t^n)$ is linear in t and $h(Y_t^m) = \log |AK_tA^T + I|$ is convex in t.
- Question: Does similar flows exist in a more general setting, i.e. outside the space of Gaussian vectors and more generally for larger class of channels?

.