Concavity of output relative entropy for channels with binary inputs




Devon Ding [1] Ken Lau [2] Chandra Nair [2] Dustin Wang [2]

[1] Dept. of Computer Science and Engg., The Chinese University of Hong Kong

[2] Dept. of Information Engg., The Chinese University of Hong Kong

IEEE ISIT 2021

Motivation

This talk is motivated by the study of non-convex optimization problems in network information theory.

Motivation

Motivation - 1 (Distributed Source Coding)

Let (X^n, Y^n) be a sequence of discrete random variables that are generated i.i.d. according to P_{XY}, denoted as 2-DMS, and the decoder would like to recover Y^n.

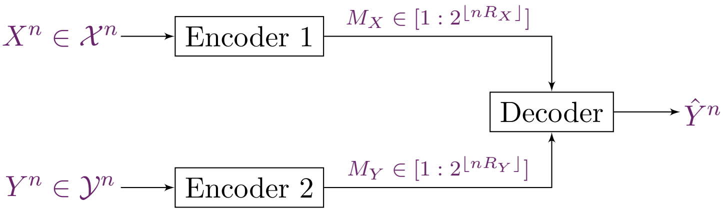


Figure 1: Distributed Source Coding With One Helper

Motivation

Motivation - 1 (Distributed Source Coding)

Let (X^n, Y^n) be a sequence of discrete random variables that are generated i.i.d. according to P_{XY}, denoted as 2-DMS, and the decoder would like to recover Y^n.

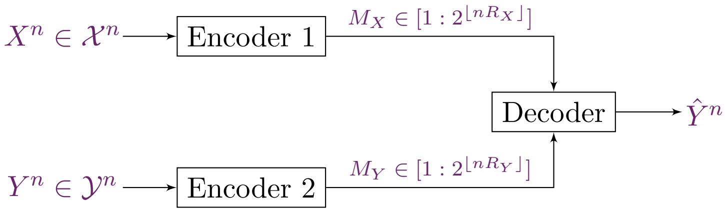


Figure 1: Distributed Source Coding With One Helper

A rate pair (R_X, R_Y) is achievable if there exists a sequence of (n, R_X, R_Y)-codes such that the $\Pr(\hat{Y}^n \neq Y^n) \to 0$ as $n \to \infty$.

Source Coding Problem [Ahlswede-Körner, '75]

What rate pairs (R_X, R_Y) are achievable under the above setting?

Motivation - 1 (Distributed Source Coding)

Source Coding Problem - Optimal Rate Region [Ahlswede-Körner, '75]

Given a 2-DMS P_{XY}, the optimal rate region is given by,

$$R_X \geq I(U;X)$$
$$R_Y \geq H(Y|U),$$

for some distribution $P_{U|X}$ such that $U \to X \to Y$ forms a Markov chain. Moreover, it suffices to consider $|\mathcal{U}| \leq |\mathcal{X}| + 1$.

Motivation - 1 (Distributed Source Coding)

Source Coding Problem - Optimal Rate Region [Ahlswede-Körner, '75]

Given a 2-DMS P_{XY}, the optimal rate region is given by,

$$R_X \geq I(U;X)$$
$$R_Y \geq H(Y|U),$$

for some distribution $P_{U|X}$ such that $U \to X \to Y$ forms a Markov chain. Moreover, it suffices to consider $|\mathcal{U}| \leq |\mathcal{X}| + 1$.

Evaluation of the region: using weighted sum-rates

The minimum λ-sum rate $\lambda R_X + R_Y$ is,

$$\begin{aligned}
&\min_{p_{U|X}} H(Y|U) + \lambda I(U;X) \\
&\quad = \lambda H(X) - \max_{p_{U|X}} \{\lambda H(X|U) - H(Y|U)\} \\
&\quad = \lambda H(X) - \mathfrak{C}[\lambda H(X) - H(Y)](P_X).
\end{aligned}$$

Non-trivial regime: $\lambda \in (0, 1)$.

The above optimization problem is a non-convex optimization problem in general.

Motivation - 1 (Distributed Source Coding)

Upper Concave Envelope and Duality (Fenchel)

Define $f(\Phi_X) := \lambda H(X) - H(Y)$.
$\mathfrak{C}[f]$ is the upper concave envelope of the function f.

Upper Concave Envelope

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

Motivation - 1 (Distributed Source Coding)

Upper Concave Envelope and Duality (Fenchel)

Define $f(\Phi_X) := \lambda H(X) - H(Y)$.
$\mathfrak{C}[f]$ is the upper concave envelope of the function f.

Upper Concave Envelope

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

From an optimization perspective

- Computing the concave envelope is not easy (in general)
- Essentially the difficulty can be translated into that of computing the dual function

Motivation - 1 (Distributed Source Coding)

Upper Concave Envelope and Duality (Fenchel)

Define $f(\Phi_X) := \lambda H(X) - H(Y)$.
$\mathfrak{C}[f]$ is the upper concave envelope of the function f.

Upper Concave Envelope

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

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 is

$$f^\dagger(d_X) := \sup_{\Phi_X} \left\{ f(\Phi_X) - \sum_{x \in \mathcal{X}} d_x \Phi_X(x) \right\}.$$

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

Motivation - 1 (Distributed Source Coding)

Upper Concave Envelope and Duality (Fenchel)

Define $f(\Phi_X) := \lambda H(X) - H(Y)$.
$\mathfrak{C}[f]$ is the upper concave envelope of the function f.

Upper Concave Envelope

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

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 is

$$f^\dagger(d_X) := \sup_{\Phi_X} \left\{ f(\Phi_X) - \sum_{x \in \mathcal{X}} d_x \Phi_X(x) \right\}.$$

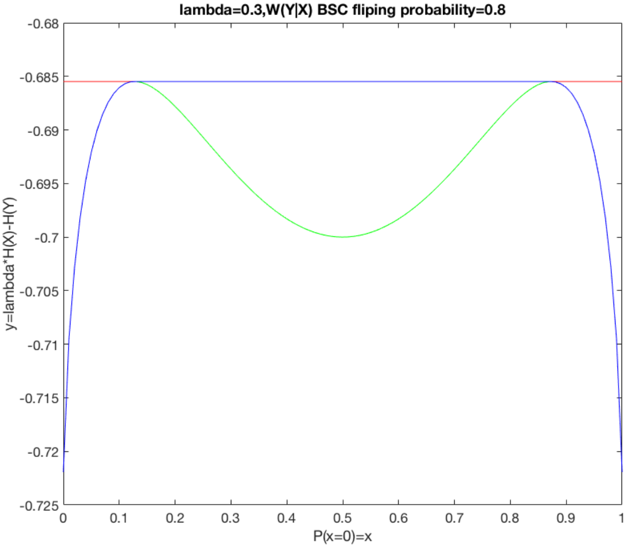
$$\mathfrak{C}_{\Phi_X}[f](P_X) = \inf_{d_X} \left\{ f^\dagger(d_X) + \sum_{x \in \mathcal{X}} d_x P_X(x) \right\}$$

The dual of the dual yields the upper concave envelope.

Motivation - 1 (Distributed Source Coding)

Plot of Upper Concave Envelope

Simple Observation: Suffices to consider gradients (dual variables) at interior distributions P_X.



Motivation - 1 (Distributed Source Coding)

Dual representation

$$f(\Phi_X) = \lambda H(\Phi_X) - H((W\Phi)_Y).$$

Let P_X be an interior distribution.

The gradient induced here, say d_X, is given by

$$\begin{aligned} d_X &= \left.\frac{\partial f(\Phi_X)}{\partial \Phi_X}\right|_{P_X} \\ &= -\lambda \ln P_X + \sum_y W_{Y|X} \ln P_Y. \end{aligned}$$

Motivation - 1 (Distributed Source Coding)

Dual representation

$$f(\Phi_X) = \lambda H(\Phi_X) - H((W\Phi)_Y).$$

Let P_X be an interior distribution.

The gradient induced here, say d_X, is given by

$$\begin{aligned} d_X &= \left.\frac{\partial f(\Phi_X)}{\partial \Phi_X}\right|_{P_X} \\ &= -\lambda \ln P_X + \sum_y W_{Y|X} \ln P_Y . \end{aligned}$$

$$f^\dagger(d_X) := \sup_{\Phi_X} \left\{ f(\Phi_X) - \sum_{x\in\mathcal{X}} d_x \Phi_X(x) \right\}$$

Substituting the above d_X into $f^\dagger$, we have,

$$f^\dagger(d_X) = -\inf_{\Phi_X} \{\lambda D(\Phi_X||P_X) - D((W\Phi)_Y||(WP)_Y)\}.$$

Related Work

[Ahlswede-Körner, '74] showed that given 2-to-2 DMC $W_{Y|X}$, the following function is convex in c.

$$f_1(c) := \inf_{H(X) \geq c} H(Y)$$

Related Work

[Ahlswede-Körner, '74] showed that given 2-to-2 DMC $W_{Y|X}$, the following function is convex in c.

$$f_1(c) := \inf_{H(X) \geq c} H(Y)$$

[Witsenhausen-Wyner, '74] showed that given any discrete memoryless channel (DMC) $W_{Y|X}$ and an input distribution P_X, for any Markov chain $U \to X \to Y$ where U is discrete r.v., the function

$$f_2(c) := \inf_{H(X|U)=c} H(Y|U)$$

is convex in c.

Related Work

[Ahlswede-Körner, '74] showed that given 2-to-2 DMC $W_{Y|X}$, the following function is convex in c.

$$f_1(c) := \inf_{H(X) \geq c} H(Y)$$

[Witsenhausen-Wyner, '74] showed that given any discrete memoryless channel (DMC) $W_{Y|X}$ and an input distribution P_X, for any Markov chain $U \to X \to Y$ where U is discrete r.v., the function

$$f_2(c) := \inf_{H(X|U) = c} H(Y|U)$$

is convex in c.

They explicitly calculated the function $\inf_{H(X|U)=c} H(Y|U)$ for BSC, BEC and Z channels (essentially using Fenchel duality).

Related Work

[Ahlswede-Körner, '74] showed that given 2-to-2 DMC $W_{Y|X}$, the following function is convex in c.

$$f_1(c) := \inf_{H(X) \geq c} H(Y)$$

[Witsenhausen-Wyner, '74] showed that given any discrete memoryless channel (DMC) $W_{Y|X}$ and an input distribution P_X, for any Markov chain $U \to X \to Y$ where U is discrete r.v., the function

$$f_2(c) := \inf_{H(X|U)=c} H(Y|U)$$

is convex in c.

They explicitly calculated the function $\inf_{H(X|U)=c} H(Y|U)$ for BSC, BEC and Z channels (essentially using Fenchel duality).

This work: we are looking at (essentially) the same problem from an optimization perspective for arbitrary 2-to-n channels.

Motivation - 2

Given $\lambda \in (0,1)$, $W_{Y|X}$, and μ_X.
View W as a Markov operator T, where for $f(Y)$ we have $Tf = \mathrm{E}(f|X)$.
Define

$$c_\lambda = \inf\{c : \mathrm{E}(\exp(T \ln f)) \le \|f\|_\lambda e^{\frac{c}{\lambda}}, \quad \forall f(Y) > 0\}.$$

Also define

$$m_\lambda := \inf_{\nu_X : \mu_X \ll \nu_X} \lambda D(\nu_X \| \mu_X) - D((W\nu)_Y \| (W\mu)_Y).$$

Motivation - 2

Given $\lambda \in (0,1)$, $W_{Y|X}$, and μ_X.
View W as a Markov operator T, where for $f(Y)$ we have $Tf = \mathrm{E}(f|X)$.
Define

$$c_\lambda = \inf\{c : \mathrm{E}(\exp(T \ln f)) \le \|f\|_\lambda e^{\frac{c}{\lambda}}, \quad \forall f(Y) > 0\}.$$

Also define

$$m_\lambda := \inf_{\nu_X : \mu_X \ll \nu_X} \lambda D(\nu_X \| \mu_X) - D((W\nu)_Y \| (W\mu)_Y).$$

Theorem (Corollary of Ahlswede-Gacs '76)

$m_\lambda = -c_\lambda$.

Motivation - 2

Given $\lambda \in (0,1)$, $W_{Y|X}$, and μ_X.
View W as a Markov operator T, where for $f(Y)$ we have $Tf = \mathrm{E}(f|X)$.
Define

$$c_\lambda = \inf\{c : \mathrm{E}(\exp(T \ln f)) \leq \|f\|_\lambda e^{\frac{c}{\lambda}}, \quad \forall f(Y) > 0\}.$$

Also define

$$m_\lambda := \inf_{\nu_X : \mu_X \ll \nu_X} \lambda D(\nu_X \| \mu_X) - D((W\nu)_Y \| (W\mu)_Y).$$

Theorem (Corollary of Ahlswede-Gacs '76)

$m_\lambda = -c_\lambda$.

- Related to strong data processing inequalities
- Original Motivation: This is the simplest case of a similar functional appearing in the evaluation of Marton's inner bound for broadcast channels

Main Problem

Non-Convex Optimization Problem

$$\min_{\Phi_X : \Phi_X \ll P_X} \{\lambda D(\Phi_X \| P_X) - D((W\Phi)_Y \| (WP)_Y)\}$$

- $\lambda \in (0, 1)$
- $W_{Y|X}$: a discrete memoryless channel
- P_X: a fixed "source/input" distribution
- $(WP)_Y$: channel induced output distribution

Main Problem

Non-Convex Optimization Problem

$$\min_{\Phi_X : \Phi_X \ll P_X} \{\lambda D(\Phi_X \| P_X) - D((W\Phi)_Y \| (WP)_Y)\}$$

- $\lambda \in (0, 1)$
- $W_{Y|X}$: a discrete memoryless channel
- P_X: a fixed "source/input" distribution
- $(WP)_Y$: channel induced output distribution

- Both $D(\Phi_X \| P_X)$ and $D((W\Phi)_Y \| (WP)_Y)$ are convex in Φ_X.
  The difference is non-convex in general.

Main Problem

Non-Convex Optimization Problem

$$\min_{\Phi_X:\Phi_X \ll P_X} \{\lambda D(\Phi_X \| P_X) - D((W\Phi)_Y \| (WP)_Y)\}$$

- $\lambda \in (0, 1)$
- $W_{Y|X}$: a discrete memoryless channel
- P_X: a fixed "source/input" distribution
- $(WP)_Y$: channel induced output distribution

- Both $D(\Phi_X \| P_X)$ and $D((W\Phi)_Y \| (WP)_Y)$ are convex in Φ_X.
  The difference is non-convex in general.
- Idea: We want to make $D(\Phi_X \| P_X)$ linear in parameter t and hope that $D((W\Phi)_Y \| (WP)_Y)$ becomes concave.

Main Problem

Non-Convex Optimization Problem

$$\min_{\Phi_X : \Phi_X \ll P_X} \{\lambda D(\Phi_X \| P_X) - D((W\Phi)_Y \| (WP)_Y)\}$$

- $\lambda \in (0, 1)$
- $W_{Y|X}$: a discrete memoryless channel
- P_X: a fixed "source/input" distribution
- $(WP)_Y$: channel induced output distribution

- Both $D(\Phi_X \| P_X)$ and $D((W\Phi)_Y \| (WP)_Y)$ are convex in Φ_X.
  The difference is non-convex in general.
- Idea: We want to make $D(\Phi_X \| P_X)$ linear in parameter t and hope that $D((W\Phi)_Y \| (WP)_Y)$ becomes concave.
- This parameterization is motivated by [Vishnoi-Sra-Yildiz '18] for positive definite matrices where they reformulate the problem of computing the Brascamp-Lieb constant into convex optimization.

Reparametrization

Given P_X, since the input space is binary, denote $\Phi_X(0) = \phi, P_X(0) = p$, we have

$$D(\Phi_X \| P_X) = \phi \ln \frac{\phi}{p} + \bar{\phi} \ln \frac{\bar{\phi}}{\bar{p}}.$$

We divide the space into two intervals: $[0, p]$ and $[p, 1]$

Reparametrization

Given P_X, since the input space is binary, denote $\Phi_X(0) = \phi, P_X(0) = p$, we have

$$D(\Phi_X \| P_X) = \phi \ln \frac{\phi}{p} + \bar{\phi} \ln \frac{\bar{\phi}}{\bar{p}}.$$

We divide the space into two intervals: $[0, p]$ and $[p, 1]$

- When $\phi \in [0, p]$, $D(\Phi_X \| P_X)$ is decreasing in ϕ. In this range, we define Φ_t by

$$D(\phi_t \| p) := (1-t) D(0 \| p) + t D(p \| p) = (1-t) \ln \left(\frac{1}{1-p} \right).$$

Reparametrization

Given P_X, since the input space is binary, denote $\Phi_X(0) = \phi, P_X(0) = p$, we have

$$D(\Phi_X \| P_X) = \phi \ln \frac{\phi}{p} + \bar{\phi} \ln \frac{\bar{\phi}}{\bar{p}}.$$

We divide the space into two intervals: $[0, p]$ and $[p, 1]$

- When $\phi \in [0, p]$, $D(\Phi_X \| P_X)$ is decreasing in ϕ. In this range, we define Φ_t by

$$D(\phi_t \| p) := (1 - t) D(0 \| p) + t D(p \| p) = (1 - t) \ln \left(\frac{1}{1 - p} \right).$$

- When $\phi \in [p, 1]$, $D(\Phi_X \| P_X)$ is increasing in ϕ. In this range, we define Φ_t by

$$D(\phi_t \| p) := (1 - t) D(p \| p) + t D(1 \| p) = t \ln \left(\frac{1}{p} \right).$$

Reparametrization

Given P_X, since the input space is binary, denote $\Phi_X(0) = \phi, P_X(0) = p$, we have

$$D(\Phi_X \| P_X) = \phi \ln \frac{\phi}{p} + \bar{\phi} \ln \frac{\bar{\phi}}{\bar{p}}.$$

We divide the space into two intervals: $[0, p]$ and $[p, 1]$

- When $\phi \in [0, p]$, $D(\Phi_X \| P_X)$ is decreasing in ϕ. In this range, we define Φ_t by

$$D(\phi_t \| p) := (1-t) D(0 \| p) + t D(p \| p) = (1-t) \ln \left(\frac{1}{1-p} \right).$$

- When $\phi \in [p, 1]$, $D(\Phi_X \| P_X)$ is increasing in ϕ. In this range, we define Φ_t by

$$D(\phi_t \| p) := (1-t) D(p \| p) + t D(1 \| p) = t \ln \left(\frac{1}{p} \right).$$

When p_X is uniform, then in $[0, \frac{1}{2}]$, we have

$$1 - H_2(\phi_t) = D\left(\phi_t \middle\| \frac{1}{2} \right) = (1-t).$$

Hence $\phi_t = H_2^{-1}(t)$. (parameterization used in Mrs. Gerber's lemma)

Main Result

Consider a 2-to-n channel $W_{Y|X}$ as follows.

$$W_{Y|X} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ \vdots & \vdots \\ a_n & b_n \end{bmatrix}$$

Here $W_{ij} = P(Y = i|X = j)$. Let $\Phi_{X,t} := (\phi_t, \overline{\phi_t})$ w.r.t. t, where $\overline{\phi_t} = 1 - \phi_t$.

Main Result

Consider a 2-to-n channel $W_{Y|X}$ as follows.

$$W_{Y|X} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ \vdots & \vdots \\ a_n & b_n \end{bmatrix}$$

Here $W_{ij} = P(Y = i|X = j)$. Let $\Phi_{X,t} := (\phi_t, \overline{\phi_t})$ w.r.t. t, where $\overline{\phi_t} = 1 - \phi_t$.

Theorem

Under the previous reparameterization, $D((W\Phi_{X,t})_Y \| (WP)_Y)$ is concave in t, if $p = p^*$ where p^* is the unique solution to

$$\sum_{i=1}^{n} \frac{(a_i - b_i)a_i b_i}{(pa_i + \bar{p}b_i)^2} = 0. \tag{1}$$

Main Result

Consider a 2-to-n channel $W_{Y|X}$ as follows.

$$W_{Y|X} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ \vdots & \vdots \\ a_n & b_n \end{bmatrix}$$

Here $W_{ij} = P(Y = i|X = j)$. Let $\Phi_{X,t} := (\phi_t, \overline{\phi_t})$ w.r.t. t, where $\overline{\phi_t} = 1 - \phi_t$.

Theorem

Under the previous reparameterization, $D((W\Phi_{X,t})_Y \| (WP)_Y)$ is concave in t, if $p = p^*$ where p^* is the unique solution to

$$\sum_{i=1}^{n} \frac{(a_i - b_i)a_i b_i}{(pa_i + \bar{p}b_i)^2} = 0. \tag{1}$$

This generalizes Mrs. Gerber's lemma to other channels.

Main Result

Consider a 2-to-n channel $W_{Y|X}$ as follows.

$$W_{Y|X} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ \vdots & \vdots \\ a_n & b_n \end{bmatrix}$$

Here $W_{ij} = P(Y = i|X = j)$. Let $\Phi_{X,t} := (\phi_t, \overline{\phi_t})$ w.r.t. t, where $\overline{\phi_t} = 1 - \phi_t$.

Theorem

Under the previous reparameterization, $D((W\Phi_{X,t})_Y \| (WP)_Y)$ is concave in t, only if $p = p^*$ where p^* is the unique solution to

$$\sum_{i=1}^{n} \frac{(a_i - b_i)a_i b_i}{(pa_i + \bar{p}b_i)^2} = 0.$$

Sufficiency of $p = p^*$

For convenience, we define $q_i = a_i p^* + b_i \bar{p^*}$ and $\psi_i = a_i \phi + b_i \bar{\phi}$.

$$D((W\Phi_{X,t})_Y \| (WP)_Y) = \sum_{i=1}^{n} \psi_i \ln \frac{\psi_i}{q_i}$$

Sufficiency of $p = p^*$

For convenience, we define $q_i = a_i p^* + b_i \bar{p}^*$ and $\psi_i = a_i \phi + b_i \bar{\phi}$.

$$D((W\Phi_{X,t})_Y \| (WP)_Y) = \sum_{i=1}^{n} \psi_i \ln \frac{\psi_i}{q_i}$$

For concavity: We want $\frac{d^2}{dt^2} D((W\Phi_{X,t})_Y \| (WP)_Y) \leq 0$.

We get (using the linearity of parameterization)

$$\frac{d^2}{dt^2} D((W\Phi_{X,t})_Y \| (WP)_Y) \overset{(a)}{=} \sum_{i=1}^{n} \left[-(a_i - b_i) \frac{\phi'^2}{\phi\bar{\phi}} \ln \frac{\psi_i}{q_i} \left(\ln \frac{\phi \bar{p}^*}{\bar{\phi} p^*} \right)^{-1} + \frac{(a_i - b_i)^2 \phi'^2}{\psi_i} \right]$$

Sufficiency of $p = p^*$

For convenience, we define $q_i = a_i p^* + b_i \bar{p^*}$ and $\psi_i = a_i \phi + b_i \bar{\phi}$.

$$D((W\Phi_{X,t})_Y \| (WP)_Y) = \sum_{i=1}^{n} \psi_i \ln \frac{\psi_i}{q_i}$$

For concavity: We want $\frac{d^2}{dt^2} D((W\Phi_{X,t})_Y \| (WP)_Y) \leq 0$.

We get (using the linearity of parameterization)

$$\frac{d^2}{dt^2} D((W\Phi_{X,t})_Y \| (WP)_Y) \overset{(a)}{=} \sum_{i=1}^{n} \left[-(a_i - b_i) \frac{\phi'^2}{\phi\bar{\phi}} \ln \frac{\psi_i}{q_i} \left(\ln \frac{\phi \bar{p^*}}{\bar{\phi} p^*} \right)^{-1} + \frac{(a_i - b_i)^2 \phi'^2}{\psi_i} \right]$$

$\frac{d^2}{dt^2} D((W\Phi_{X,t})_Y \| (WP)_Y) \leq 0$ is equivalent to

$$f(\phi; p^*) := \sum_{i=1}^{n} \left[-(a_i - b_i) \ln \frac{\psi_i}{q_i} + (a_i - b_i)^2 \frac{\phi\bar{\phi}}{\psi_i} \ln \frac{\phi \bar{p^*}}{\bar{\phi} p^*} \right] \begin{cases} \geq 0, & 0 \leq \phi \leq p^*; \\ \leq 0, & p^* \leq \phi \leq 1. \end{cases}$$

Sufficiency of $p = p^*$

To show

$$f(\phi; p^*) := \sum_{i=1}^{n} \left[-(a_i - b_i) \ln \frac{\psi_i}{q_i} + (a_i - b_i)^2 \frac{\phi\bar{\phi}}{\psi_i} \ln \frac{\phi \bar{p}^*}{\bar{\phi} p^*} \right] \begin{cases} \geq 0, & 0 \leq \phi \leq p^*; \\ \leq 0, & p^* \leq \phi \leq 1. \end{cases}$$

Sufficiency of $p = p^*$

To show

$$f(\phi; p^*) := \sum_{i=1}^{n} \left[-(a_i - b_i) \ln \frac{\psi_i}{q_i} + (a_i - b_i)^2 \frac{\phi\bar{\phi}}{\psi_i} \ln \frac{\phi\bar{p}^*}{\bar{\phi}p^*} \right] \begin{cases} \geq 0, & 0 \leq \phi \leq p^*; \\ \leq 0, & p^* \leq \phi \leq 1. \end{cases}$$

Observe that $f(p^*; p^*) = 0$.

Suffices to show that $f(\phi; p^*)$ is decreasing on $0 \leq \phi \leq 1$.

Let $g(\phi) := \sum_{i=1}^{n} \frac{a_i b_i (a_i - b_i)}{(a_i \phi + b_i \bar{\phi})^2}$, then $\frac{d}{d\phi} f(\phi; p^*) = g(\phi) \ln \frac{\phi \bar{p}^*}{\bar{\phi} p^*}$.

$f(\phi; p^*)$ is decreasing on $0 \leq \phi \leq 1$ is equivalent to

$$g(\phi) \begin{cases} \geq 0, & 0 \leq \phi \leq p^*; \\ \leq 0, & p^* \leq \phi \leq 1. \end{cases}$$

Sufficiency of $p = p^*$

To show

$$f(\phi; p^*) := \sum_{i=1}^{n} \left[-(a_i - b_i) \ln \frac{\psi_i}{q_i} + (a_i - b_i)^2 \frac{\phi\bar{\phi}}{\psi_i} \ln \frac{\phi\bar{p}^*}{\bar{\phi}p^*} \right] \begin{cases} \geq 0, & 0 \leq \phi \leq p^*; \\ \leq 0, & p^* \leq \phi \leq 1. \end{cases}$$

Observe that $f(p^*; p^*) = 0$.

Suffices to show that $f(\phi; p^*)$ is decreasing on $0 \leq \phi \leq 1$.

Let $g(\phi) := \sum_{i=1}^{n} \frac{a_i b_i (a_i - b_i)}{(a_i \phi + b_i \bar{\phi})^2}$, then $\frac{d}{d\phi} f(\phi; p^*) = g(\phi) \ln \frac{\phi\bar{p}^*}{\bar{\phi}p^*}$.

$f(\phi; p^*)$ is decreasing on $0 \leq \phi \leq 1$ is equivalent to

$$g(\phi) \begin{cases} \geq 0, & 0 \leq \phi \leq p^*; \\ \leq 0, & p^* \leq \phi \leq 1. \end{cases}$$

Observe that $g(p^*) = 0$. (definition of p^*).

Suffices to show that $g(\phi)$ is decreasing on $0 \leq \phi \leq 1$.

Sufficiency of $p = p^*$

To show

$$f(\phi; p^*) := \sum_{i=1}^{n} \left[-(a_i - b_i) \ln \frac{\psi_i}{q_i} + (a_i - b_i)^2 \frac{\phi\bar{\phi}}{\psi_i} \ln \frac{\phi \bar{p}^*}{\bar{\phi} p^*} \right] \begin{cases} \geq 0, & 0 \leq \phi \leq p^*; \\ \leq 0, & p^* \leq \phi \leq 1. \end{cases}$$

Observe that $f(p^*; p^*) = 0$.

Suffices to show that $f(\phi; p^*)$ is decreasing on $0 \leq \phi \leq 1$.

Let $g(\phi) := \sum_{i=1}^{n} \frac{a_i b_i (a_i - b_i)}{(a_i \phi + b_i \bar{\phi})^2}$, then $\frac{d}{d\phi} f(\phi; p^*) = g(\phi) \ln \frac{\phi \bar{p}^*}{\bar{\phi} p^*}$.

$f(\phi; p^*)$ is decreasing on $0 \leq \phi \leq 1$ is equivalent to

$$g(\phi) \begin{cases} \geq 0, & 0 \leq \phi \leq p^*; \\ \leq 0, & p^* \leq \phi \leq 1. \end{cases}$$

Taking derivative of $g(\phi)$ gives

$$\frac{d}{d\phi} g(\phi) = -\sum_{i=1}^{n} \frac{2(a_i - b_i)^2 a_i b_i}{(a_i \phi + b_i \bar{\phi})^3} < 0 \qquad \square$$

Necessity of $p = p^*$

Assume concavity, i.e. $\frac{d^2}{dt^2} D((W\Phi_{X,t})_Y \| (WP)_Y) \leq 0$

$$f(\phi;p) := \sum_{i=1}^{n} \left[-(a_i - b_i) \ln \frac{\psi_i}{q_i} + (a_i - b_i)^2 \frac{\phi\bar{\phi}}{\psi_i} \ln \frac{\phi\bar{p}}{\bar{\phi}p} \right] \begin{cases} \geq 0, & 0 \leq \phi \leq p; \\ \leq 0, & p \leq \phi \leq 1. \end{cases}$$

Taylor expansion around $\phi = p$ gives

$$f(p+\varepsilon;p) = \left. \frac{\partial^2 f(\phi;p)}{\partial \phi^2} \right|_{\phi=p} \varepsilon^2 + O(\varepsilon^3).$$

Note that we used $f(p;p) = \left. \frac{\partial f(\phi;p)}{\partial \phi} \right|_{\phi=p} = 0$.

Necessity of $p = p^*$

Assume concavity, i.e. $\frac{d^2}{dt^2} D((W\Phi_{X,t})_Y \| (WP)_Y) \le 0$

$$f(\phi; p) := \sum_{i=1}^{n} \left[-(a_i - b_i) \ln \frac{\psi_i}{q_i} + (a_i - b_i)^2 \frac{\phi\bar{\phi}}{\psi_i} \ln \frac{\phi\bar{p}}{\bar{\phi}p} \right] \begin{cases} \ge 0, & 0 \le \phi \le p; \\ \le 0, & p \le \phi \le 1. \end{cases}$$

Taylor expansion around $\phi = p$ gives

$$f(p+\varepsilon; p) = \left. \frac{\partial^2 f(\phi; p)}{\partial \phi^2} \right|_{\phi = p} \varepsilon^2 + O(\varepsilon^3).$$

Note that we used $f(p; p) = \left. \frac{\partial f(\phi; p)}{\partial \phi} \right|_{\phi = p} = 0$.

The concavity condition forces that

$$\left. \frac{\partial^2 f(\phi; p)}{\partial \phi^2} \right|_{\phi = p} = \frac{1}{p\bar{p}} g(p) = 0.$$

Necessity of $p = p^*$

Assume concavity, i.e. $\frac{d^2}{dt^2} D((W\Phi_{X,t})_Y \| (WP)_Y) \leq 0$

$$f(\phi; p) := \sum_{i=1}^{n} \left[-(a_i - b_i) \ln \frac{\psi_i}{q_i} + (a_i - b_i)^2 \frac{\phi\bar{\phi}}{\psi_i} \ln \frac{\phi\bar{p}}{\bar{\phi}p} \right] \begin{cases} \geq 0, & 0 \leq \phi \leq p; \\ \leq 0, & p \leq \phi \leq 1. \end{cases}$$

Taylor expansion around $\phi = p$ gives

$$f(p + \varepsilon; p) = \left. \frac{\partial^2 f(\phi; p)}{\partial \phi^2} \right|_{\phi = p} \varepsilon^2 + O(\varepsilon^3).$$

Note that we used $f(p; p) = \left. \frac{\partial f(\phi; p)}{\partial \phi} \right|_{\phi = p} = 0.$

The concavity condition forces that

$$\left. \frac{\partial^2 f(\phi; p)}{\partial \phi^2} \right|_{\phi = p} = \frac{1}{p\bar{p}} g(p) = 0. \qquad g(p) = 0 \iff p = p^*$$

Therefore, $\frac{d^2}{dt^2} D((W\Phi_{X,t})_Y \| (WP)_Y) \leq 0$ only if $p = p^*$.

When $p \neq p^*$

We still have one-sided concavity when $p \neq p^*$.

Theorem (one-sided concavity for $p \neq p^*$)

- *If* $p > p^*$, $D((W\Phi_{X,t})_Y \| (WP)_Y)$ *is concave on* $\phi \in [p, 1]$.
- *If* $p < p^*$, $D((W\Phi_{X,t})_Y \| (WP)_Y)$ *is concave on* $\phi \in [0, p]$.

When $p \neq p^*$

We still have one-sided concavity when $p \neq p^*$.

Theorem (one-sided concavity for $p \neq p^*$)

- *If* $p > p^*$, $D((W\Phi_{X,t})_Y \| (WP)_Y)$ *is concave on* $\phi \in [p, 1]$.
- *If* $p < p^*$, $D((W\Phi_{X,t})_Y \| (WP)_Y)$ *is concave on* $\phi \in [0, p]$.

Unfortunately, it is not necessarily concave in the remaining segment.

When $p \neq p^*$

We still have one-sided concavity when $p \neq p^*$.

Theorem (one-sided concavity for $p \neq p^*$)

- *If* $p > p^*$, $D((W\Phi_{X,t})_Y \| (WP)_Y)$ *is concave on* $\phi \in [p, 1]$.
- *If* $p < p^*$, $D((W\Phi_{X,t})_Y \| (WP)_Y)$ *is concave on* $\phi \in [0, p]$.

Unfortunately, it is not necessarily concave in the remaining segment.

Natural question: Are there other parameterizations of ϕ_X that makes the functional $\lambda D(\Phi_{X,t} \| P_X) - D((W\Phi_{X,t})_Y \| (WP)_Y)$ convex for $p \neq p^*$.

When $p \neq p^*$

We still have one-sided concavity when $p \neq p^*$.

Theorem (one-sided concavity for $p \neq p^*$)

- *If* $p > p^*$, $D((W\Phi_{X,t})_Y \| (WP)_Y)$ *is concave on* $\phi \in [p, 1]$.
- *If* $p < p^*$, $D((W\Phi_{X,t})_Y \| (WP)_Y)$ *is concave on* $\phi \in [0, p]$.

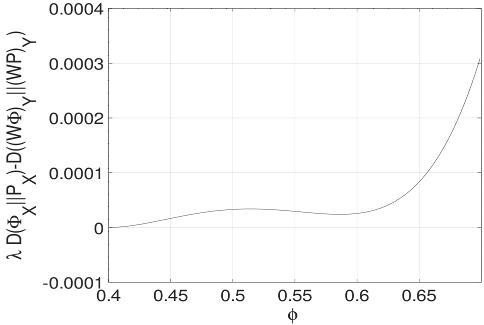
Unfortunately, it is not necessarily concave in the remaining segment.

Natural question: Are there other parameterizations of ϕ_X that makes the functional $\lambda D(\Phi_{X,t} \| P_X) - D((W\Phi_{X,t})_Y \| (WP)_Y)$ convex for $p \neq p^*$.

Answer: No

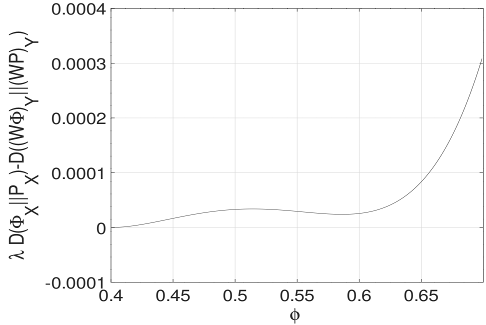
Lack of parametrization

Consider a BSC with $\epsilon = 0.3$, when $p = 0.4 \neq p^* = 0.5$ and $\lambda = 0.1584$.



Lack of parametrization

Consider a BSC with $\epsilon = 0.3$, when $p = 0.4 \neq p^* = 0.5$ and $\lambda = 0.1584$.



- Any parametrization using a submersion (differentiable map) will map strict local maximizers to strict local maximizers
- No convex function can have an interior local maximizer.
- Impossible to reparameterize in the regime $[0.4, 1]$ into a convex function.

Beyond binary inputs

Given a m-to-n channel W, suppose $\boldsymbol{\phi}(t) \in \Delta^{m-1}$ is an interval parametrized by $t \geq 0$ s.t. all coordinates are fixed except for two. W.l.o.g., we let

$$\boldsymbol{\phi}(t) := (\phi(t), \alpha - \phi(t), \phi_3, ...\phi_n)$$

where $\phi_i, i = 3, 4, ..., n$ are constants and $\sum_{i=3}^{n} \phi_i = 1 - \alpha$.

Given a m-to-n channel W, suppose $\boldsymbol{\phi}(t) \in \Delta^{m-1}$ is an interval parametrized by $t \geq 0$ s.t. all coordinates are fixed except for two. W.l.o.g., we let

$$\boldsymbol{\phi}(t) := (\phi(t), \alpha - \phi(t), \phi_3, ...\phi_n)$$

where $\phi_i, i = 3, 4, ..., n$ are constants and $\sum_{i=3}^{n} \phi_i = 1 - \alpha$.

Corollary

Then there exists a P_X on this interval and a similar $\Phi_{X,t}$ of this interval that makes

$$\lambda D(\Phi_{X,t} \| P_X) - D((W\Phi_{X,t})_Y \| (WP)_Y)$$

convex in t.

Beyond binary inputs

Given a m-to-n channel W, suppose $\boldsymbol{\phi}(t) \in \Delta^{m-1}$ is an interval parametrized by $t \geq 0$ s.t. all coordinates are fixed except for two. W.l.o.g., we let

$$\boldsymbol{\phi}(t) := (\phi(t), \alpha - \phi(t), \phi_3, ...\phi_n)$$

where $\phi_i, i = 3, 4, ..., n$ are constants and $\sum_{i=3}^{n} \phi_i = 1 - \alpha$.

Corollary

Then there exists a P_X on this interval and a similar $\Phi_{X,t}$ of this interval that makes

$$\lambda D(\Phi_{X,t} \| P_X) - D((W\Phi_{X,t})_Y \| (WP)_Y)$$

convex in t.
Previous one-sided concavity result also generalizes to this setting.

Beyond binary inputs

Given a m-to-n channel W, suppose $\boldsymbol{\phi}(t) \in \Delta^{m-1}$ is an interval parametrized by $t \geq 0$ s.t. all coordinates are fixed except for two. W.l.o.g., we let

$$\boldsymbol{\phi}(t) := (\phi(t), \alpha - \phi(t), \phi_3, ...\phi_n)$$

where $\phi_i, i = 3, 4, ..., n$ are constants and $\sum_{i=3}^{n} \phi_i = 1 - \alpha$.

Corollary

Then there exists a P_X on this interval and a similar $\Phi_{X,t}$ of this interval that makes

$$\lambda D(\Phi_{X,t} \| P_X) - D((W\Phi_{X,t})_Y \| (WP)_Y)$$

convex in t.

Proof: Effectively reduces to a new 2-to-n channel.

Issues with generalization to higher alphabets

Can we find a path from ϕ_X to (some) p_X^* so that under a suitable parametrization

$$\lambda D(\Phi_{X,t} \| P_X) - D((W\Phi_{X,t})_Y \| (WP)_Y)$$

is convex in t.

Issues with generalization to higher alphabets

Can we find a path from ϕ_X to (some) p_X^* so that under a suitable parametrization

$$\lambda D(\Phi_{X,t} \| P_X) - D((W\Phi_{X,t})_Y \| (WP)_Y)$$

is convex in t.

- In binary input, the path was fixed (since the space is a line).
- In higher alphabets, there are many possible choices for paths between two points (even in some fixed partition of the space)

Issues with generalization to higher alphabets

Can we find a path from ϕ_X to (some) p^*_X so that under a suitable parametrization

$$\lambda D(\Phi_{X,t} \| P_X) - D((W\Phi_{X,t})_Y \| (WP)_Y)$$

is convex in t.

Consider the following 3-SC:

$$W(y|x) = \begin{bmatrix} 0.55 & 0.15 & 0.15 \\ 0.15 & 0.55 & 0.15 \\ 0.15 & 0.15 & 0.55 \end{bmatrix}.$$

A natural guess for p^* is $P_X = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ by symmetry.

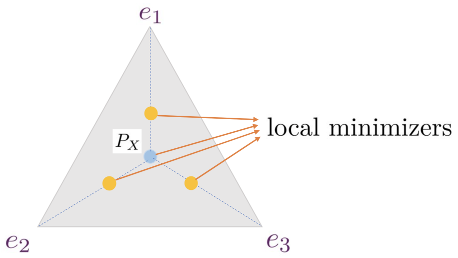
Issues with generalization to higher alphabets

Can we find a path from ϕ_X to (some) p_X^* so that under a suitable parametrization

$$\lambda D(\Phi_{X,t} \| P_X) - D((W\Phi_{X,t})_Y \| (WP)_Y)$$

is convex in t.

When $\lambda = 0.309$, there are has four local minimizers.



Remarks and future directions

What can be done: potential future directions

- Perhaps it is possible to restrict the locations of the local minimizers
- Establish some properties of local minimizers

Remarks and future directions

What can be done: potential future directions

- Perhaps it is possible to restrict the locations of the local minimizers
- Establish some properties of local minimizers

Related Work

- n-SC: all local minimizers lie on the paths connecting the center $P_X = \frac{1}{n}\mathbf{1}_n$ and e_i's (the vertices).
- This is a one-dimensional space. And there is at most one local minimizer on each path excluding the center.

Remarks and future directions

What can be done: potential future directions

- Perhaps it is possible to restrict the locations of the local minimizers
- Establish some properties of local minimizers

Related Work

- n-SC: all local minimizers lie on the paths connecting the center $P_X = \frac{1}{n}\mathbf{1}_n$ and e_i's (the vertices).
- This is a one-dimensional space. And there is at most one local minimizer on each path excluding the center.

Such results may be obtainable using the ideas here.

- They could be useful for designing algorithms
- They could be useful in establishing capacity regions

Remarks and future directions

What can be done: potential future directions

- Perhaps it is possible to restrict the locations of the local minimizers
- Establish some properties of local minimizers

Related Work

- n-SC: all local minimizers lie on the paths connecting the center $P_X = \frac{1}{n}\mathbf{1}_n$ and e_i's (the vertices).
- This is a one-dimensional space. And there is at most one local minimizer on each path excluding the center.

Such results may be obtainable using the ideas here.

- They could be useful for designing algorithms
- They could be useful in establishing capacity regions

Thank you for watching our presentation