Concavity of output relative entropy for channels with binary inputs

Qinghua (Devon) Ding
Dept. of Computer Science and Engg.
The Chinese University of Hong Kong
Sha Tin, N.T., Hong Kong
Email: qhding@cse.cuhk.edu.hk

Chin Wa (Ken) Lau, Chandra Nair and Yan Nan Wang
Dept. of Information Engg.
The Chinese University of Hong Kong
Sha Tin, N.T., Hong Kong
Email: {kenlau,chandra,dustin}@ie.cuhk.edu.hk

***Abstract*—We generalize a convexity result due to Wyner and Ziv to channels with binary inputs and arbitrary outputs. This results in a convex reformulation of some non-convex optimization problems that arise naturally in multi-user information theory.**

I. Introduction

The optimality of certain achievable rate regions for communication settings in multiuser information theory, such as the Marton's region for the two-receiver broadcast channel, can be verified by establishing that product distributions are the global maximizers of a corresponding non-convex functional on product spaces, [1]. A functional satisfying the above property is said to satisfy global tensorization. As stated in [2] a curious connection has been repeatedly observed between functionals that satisfy global tensorization and those that satisfy a so-called local tensorization property. One way to reconcile this apparent relationship is to determine if all the local maximizers of non-convex functionals that satisfy global tensorization are also product distributions.

On a related note, information inequalities concerning non-convex functionals have also been established [3] by determining all the local maximizers. Additionally, certain non-convex functionals, such as the one arising in the capacity region computation of the vector Gaussian channel [4] is shown to have a unique local maximum. Inspired by these observations, we seek to understand the geometric structure of certain information functionals and determine its set of local extremizers. The family considered in this paper can be considered as an elementary but non-trivial sub-class of functionals. The results in this paper extend the celebrated convexity result, sometimes referred to as Mrs. Gerber's lemma, of Wyner and Ziv to a broader family of channels.

Given a conditional distribution $W_{Y|X}$, a reference distribution P_X, and a non-negative parameter λ we will be investigating non-convex optimization problems of the form

$$\min_{\Phi_X} \left\{ \lambda D(\Phi_X \| P_X) - D((W\Phi)_Y \| (WP)_Y) \right\}, \qquad (1)$$

and see if these problems can be reparameterized into convex optimization problems. In the above expression $D(\Phi_X \| P_X)$ denotes the relative entropy, and the logarithms are assumed to be with respect to base e. If such a reparameterization exists, then any local minimizer would also be a global minimizer (similar to the observation in the MIMO Gaussian broadcast channel). The main idea is to choose a parameterization of Φ_X so that $D(\Phi_X \| P_X)$ is linear in the parameter and determine whether the output relative entropy, $D((W\Phi)_Y \| (WP)_Y)$, is concave. This approach is motivated by geodesically convex reformulations of the Brascamp-Lieb constants in [5].

A. Motivation

Consider the following optimization scenarios originating in multiuser information theory.

(i) In the Ahlswede-Korner source coding problem [6], to compute the minimal weighted sum-rate, one is faced with the following optimization problem: Given a conditional distribution $W_{Y|X}$ and an input distribution Φ_X, one seeks to compute the value of the following optimization problem (parameterized by λ, $\lambda \geq 0$):

$$\min_{U: U \to X \to Y} H(Y|U) + \lambda I(U;X).$$

(ii) In the degraded broadcast channel, to compute the maximum weighted sum-rate $R_Z + \lambda R_Y$, one seeks to compute the value of the following optimization problem (parameterized by λ, $0 \leq \lambda \leq 1$):

$$\max_{U,X: U \to X \to Y \to Z} I(U;Z) + \lambda I(X;Y|U).$$

Both of these problems result in the computation of the lower convex envelope with respect to Φ_X for the functionals $H(Y) - \lambda H(X)$ and $H(Z) - \lambda H(Y)$, respectively. Observe that in the latter case, the channel $W_{Z|Y}$ is fixed, and in the former case the conditional distribution $W_{Y|X}$ is fixed. Note that, when $\lambda = 0$ both functionals are concave in Φ_X and when $\lambda = 1$ both functionals are convex in Φ_X. For $\lambda \in (0,1)$ (the interesting regime), the function is not necessarily convex or concave. Therefore the computation of the lower convex envelope does not reduce to a convex optimization problem and apriori the functionals may have multiple local minimizers. Hence it is natural to ask if there is a subset of the above family of problems for which under a suitable reparameterization, the problem reduces to a convex optimization problem.

Characterization of the lower convex envelope can be done via Fenchel duality by computing its supporting hyperplanes. To this end we seek to compute the minimum of

$$G(P_X) := \min_{\Phi_X} \{\lambda D(\Phi_X \| P_X) - D((W\Phi)_Y \| (WP)_Y)\}$$
$$= \min_{\Phi_X} \left\{ H_\Phi(Y) - \lambda H_\Phi(X) - \sum_x a_x \Phi_X(x) \right\},$$

where $(W\Phi)_Y$ denotes the distribution on Y induced by the input distribution Φ_X and the channel $W_{Y|X}$, $H_\Phi(X)$ denotes the Shannon entropy of X when $X \sim \Phi_X$, and $a_x = \sum_y W(y|x) \ln \frac{(WP)_y}{P(x)^\lambda}$. Thus $G(P_X)$ denotes the Fenchel dual for the convex envelope of $H(Y) - \lambda H(X)$, with $a_x = \sum_y W(y|x) \ln \frac{(WP)_y}{P(x)^\lambda}$ being the dual variables. This is one way in which optimization problems of the type described in (1) arise in multiuser information theory.

Another motivation for such optimization problems lie in determining the optimal constants for Strong-Data-Processing inequalities and in turn to determining limiting hypercontractivity parameters [7]. It has been shown in [8] that given $P_X, W_{Y|X}$, the inequality

$$I(U;Y) - \eta I(U;X) \le 0,$$

holds for all $U : U \to X \to Y$ is Markov, if and only if, the inequality

$$\min_{\Phi_X} \{\eta D(\Phi_X \| P_X) - D((W\Phi)_Y \| (WP)_Y)\} \ge 0$$

holds. Note that the range of η depends on P_X. One can also define a similar η that holds for all input distributions P_X (and thus depends only on the channel $W_{Y|X}$) to be

$$\eta_W := \min\{\eta : I(U;Y) - \eta I(U;X) \le 0, \quad \forall p_{UX} : U \to X \to Y \text{ is Markov}\}$$

or equivalently (see Exercise 15.12 in [9])

$$\eta_W := \min\{\eta : \lambda D(\Phi_X \| P_X) - D((W\Phi)_Y \| (WP)_Y) \ge 0, \quad \forall \Phi_X, P_X\}.$$

It has recently been shown [10] that for any $W_{Y|X}$ it suffices to consider P_X having support on two alphabets and $\Phi_X \ll P_X$ to compute η_W.

Remark 1. In light of this result, the case of $\mathcal{X}$ being binary takes particular significance while considering the family of optimization problems of the form

$$\min_{\Phi_X} \{\lambda D(\Phi_X \| P_X) - D((W\Phi)_Y \| (WP)_Y)\}, \quad (2)$$
$$\min_{P_X, \Phi_X} \{\lambda D(\Phi_X \| P_X) - D((W\Phi)_Y \| (WP)_Y)\}. \quad (3)$$

B. A convexity result due to Wyner and Ziv

While trying to compute the superposition coding region of a degraded binary-symmetric broadcast channel (see item (ii) in the Motivation), Wyner and Ziv showed that for any $\alpha \in [0, \frac{1}{2}]$, the function $H_2(\alpha * H_2^{-1}(u))$ is convex in u, where $H_2 : [0, \frac{1}{2}] \mapsto [0, \ln 2]$ is binary entropy function given by $H_2(x) = -x \ln x - (1-x) \ln(1-x)$ and $H_2^{-1} : [0, \ln 2] \mapsto [0, \frac{1}{2}]$ is its inverse. Here $a * b = a(1-b) + b(1-a)$ denotes a two-point convolution.

We can interpret this result alternately as the following: Let $W_{Y|X}$ be the binary symmetric channel with crossover probability α. Let P_X be the uniform distribution and parameterize $\Phi_{X,t} = (H_2^{-1}(t), 1 - H_2^{-1}(t))$. Now observe that under this parameterization, $D(\Phi_{X,t} \| P_X) = \ln 2 - t$ is linear in t, and $D((W\Phi_{X,t})_Y \| (WP)_Y) = \ln 2 - H_2(\alpha * H_2^{-1}(t))$ is concave in t. Therefore the function

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

is convex in t, reducing the computation of (1) to a convex optimization problem. Note that $\Phi_{X,t}$ determines a path along the binary simplex such that $D(\Phi_{X,t} \| P_X)$ is linear in t and $D((W\Phi_{X,t})_Y \| (WP)_Y)$ is concave in t.

Thus the question we seek to address is: given any channel $W_{Y|X}$, a reference distribution P_X, and an initial distribution Φ_X, is it possible to parameterize the path from Φ_X to P_X according to $\Phi_{X,t}$, where $\Phi_{X,0} = \Phi_X$ and $\Phi_{X,1} = P_X$, with the property that $D(\Phi_{X,t} \| P_X)$ is linear in t and $D((W\Phi_{X,t})_Y \| (WP)_Y)$ is concave in t. We will answer this question for channels with binary inputs and arbitray output cardinalities. As stated in Remark 1, the case of binary inputs (and outputs of arbitrary cardinality) is particularly useful when computing η_W for channels with arbitrary input alphabets.

We first present our results for channels with binary outputs as we have slightly stronger results (see Proposition 1) in this setting. Our main results are presented in Theorem 1 and Theorem 2; these results generalize the convexity of $H_2(\alpha * H_2^{-1}(u))$.

II. Channels with binary inputs and binary outputs

Let us denote a binary-input binary-output channel as

$$W_{Y|X} = \begin{bmatrix} a & b \\ \bar{a} & \bar{b} \end{bmatrix}. \quad (4)$$

Here the matrix entry $W_{ij} = P(Y = i | X = j)$, $\bar{a} = 1 - a, \bar{b} = 1 - b$. Let us denote $\Phi_{X,t} = (\phi(t), 1 - \phi(t))$ and $P_X = (p, 1-p)$ to characterize the parameterized path and the reference distribution. Further we denote, for $a, b \in [0,1]$,

$$D_2(a \| b) := a \ln \frac{a}{b} + (1-a) \ln \frac{1-a}{1-b}$$

to be the relative entropy between the two-point distributions characterized by $(a, 1-a)$ and $(b, 1-b)$ respectively. We also use $\bar{\phi}$ to represent $1 - \phi$ for brevity. We also assume that the reference measure satisfies $p > 0$; otherwise $D_2(\phi \| 0) = \infty$ for all $\phi \neq 0$.

Note that $D_2(\phi(t) \| p)$ is monotonically increasing (resp. decreasing) when $\phi(t) \ge p$ (resp. $\phi(t) \le p$). Hence if we enforce the linear dependence of input divergence on t, we obtain

$$\frac{d^2}{dt^2} D_2(\phi \| p) = \phi'' \ln \frac{\phi \bar{p}}{\bar{\phi} p} + \frac{\phi'^2}{\phi \bar{\phi}} = 0. \quad (5)$$

where $\phi' = \frac{d\phi}{dt}$ and $\phi'' = \frac{d^2\phi}{dt^2}$.

Imposing the boundary conditions $\phi(0) = p$ and $\phi(1) = 1$ (resp. $\phi(0) = 0$ and $\phi(1) = p$), then $\phi(t)$ can be uniquely determined (due to the monotonicity of $D_2(\phi(t)\|p)$). Concretely, for $\phi(t) \geq p$, $\phi(t)$ is the unique solution of

$$D_2(\phi(t)\|p) = t \ln \frac{1}{p},$$

and for $\phi(t) \leq p$, $\phi(t)$ is the unique solution of

$$D_2(\phi(t)\|p) = (1-t) \ln \frac{1}{1-p}.$$

Remark 2. Note that this reparameterization $\phi(t)$ generalizes the parameterization $H_2^{-1}(t)$ employed by Wyner and Ziv for the binary symmetric channel.

Let $(WP)_Y = (ap + b\bar{p}, \bar{a}p + \bar{b}\bar{p})$ and $(W\Phi)_Y = (a\phi + b\bar{\phi}, \bar{a}\phi + \bar{b}\bar{\phi})$. We define $q \triangleq ap + b\bar{p}$, and $\psi \triangleq a\phi + b\bar{\phi}$. Now we can calculate the second order derivative $\frac{d^2}{dt^2} D_2(\psi\|q)$ as following:

$$\begin{aligned} \frac{d^2}{dt^2} D_2(\psi\|q) =& \psi'' \ln \frac{\psi\bar{q}}{\bar{\psi}q} + \frac{\psi'^2}{\psi\bar{\psi}} \\ \overset{(a)}{=}& -(a-b)\frac{\phi'^2}{\phi\bar{\phi}} \frac{\ln\frac{\psi\bar{q}}{\bar{\psi}q}}{\ln\frac{\phi\bar{p}}{\bar{\phi}p}} + (a-b)^2 \frac{\phi'^2}{\psi\bar{\psi}} \end{aligned} \tag{6}$$

where $\psi' = \frac{d\psi}{dt}$ and $\psi'' = \frac{d^2\psi}{dt^2}$. Equality (a) follows from equation (5). Suppose $\phi' \neq 0$, then concavity of $D((W\Phi_{X,t})_Y\|(WP)_Y)$ is equivalent to $\frac{d^2}{dt^2} D_2(\psi\|q) \leq 0$. This, in turn, is equivalent to

$$\begin{aligned} f(\phi; p) :=& (a-b)^2 \phi\bar{\phi} \ln \frac{\phi\bar{p}}{\bar{\phi}p} - (a-b)\psi\bar{\psi} \ln \frac{\psi\bar{q}}{\bar{\psi}q} \\ & \begin{cases} \geq 0, & \phi \leq p; \\ \leq 0, & \phi \geq p; \end{cases} \end{aligned} \tag{7}$$

since $\ln \frac{\phi\bar{p}}{\bar{\phi}p} \leq 0$ (resp. ≥ 0) when $\phi \leq p$ (resp. $\phi \geq p$).

Remark that this condition (7) now does not depend on t. One may calculate the derivatives of $f(\phi; p)$ w.r.t. ϕ as follows.

$$\begin{aligned} \frac{d}{d\phi} f(\phi; p) =& (a-b)^2 \left[(1-2\phi) \ln \frac{\phi\bar{p}}{\bar{\phi}p} - (1-2\psi) \ln \frac{\psi\bar{q}}{\bar{\psi}q} \right] \\ \frac{d^2}{d\phi^2} f(\phi; p) =& (a-b)^2 \left[-2 \ln \frac{\phi\bar{p}}{\bar{\phi}p} + \frac{(1-2\phi)}{\phi\bar{\phi}} \right] \\ & - (a-b)^3 \left[-2 \ln \frac{\psi\bar{q}}{\bar{\psi}q} + \frac{(1-2\psi)}{\psi\bar{\psi}} \right] \\ \frac{d^3}{d\phi^3} f(\phi; p) =& -\frac{(a-b)^2}{\phi^2\bar{\phi}^2} + \frac{(a-b)^4}{\psi^2\bar{\psi}^2}. \end{aligned} \tag{8}$$

We will show that $\frac{d^2}{d\phi^2} f(\phi; p)$ is decreasing w.r.t. ϕ in the following lemma.

Lemma 1. *The second-order derivative $\frac{d^2}{d\phi^2} f(\phi; p)$ is monotonically decreasing in $\phi \in [0, 1]$.*

Proof. Suffices to show $\frac{d^3}{d\phi^3} f(\phi; p) \leq 0$, which is equivalent to $\psi^2\bar{\psi}^2 \geq (a-b)^2\phi^2\bar{\phi}^2$ or

$$(\psi\bar{\psi} + (a-b)\phi\bar{\phi})(\psi\bar{\psi} - (a-b)\phi\bar{\phi}) \geq 0.$$

When $0 \leq a, b \leq 1$, we have

$$\begin{aligned} & \psi\bar{\psi} - (a-b)\phi\bar{\phi} \\ =& (a\phi + b\bar{\phi})(\bar{a}\phi + \bar{b}\bar{\phi}) - (a-b)\phi\bar{\phi} \\ =& a\bar{a}\phi^2 + b\bar{b}\bar{\phi}^2 + (a\bar{b} + b\bar{a} - (a-b))\phi\bar{\phi} \\ =& a\bar{a}\phi^2 + b\bar{b}\bar{\phi}^2 + 2b\bar{a}\phi\bar{\phi} \geq 0 \end{aligned}$$

Similarly, we have

$$\psi\bar{\psi} + (a-b)\phi\bar{\phi} = a\bar{a}\phi^2 + b\bar{b}\bar{\phi}^2 + 2a\bar{b}\phi\bar{\phi} \geq 0.$$

This proves the required inequality. □

Theorem 1. *Consider a binary channel represented as Equation* (4)*, with $a \neq b$ and $a, b \in (0, 1)$. Assume that the input distribution $\Phi_{X,t}$ is reparametrized according to Equation* (5)*, then the output relative entropy $D((W\Phi_{X,t})_Y\|(WP)_Y)$ is concave w.r.t. t under such a reparametrization, if and only if p is equal to*

$$p^* := \frac{\sqrt{b\bar{b}}}{\sqrt{b\bar{b}} + \sqrt{a\bar{a}}}.$$

Proof. We will first show that $p = p^*$ is necessary. Calculate the Taylor expansion of $f(\phi; p)$ at $\phi = p$, and observe that $f(p; p) = 0$ and $\frac{d}{d\phi} f(\phi; p)|_{\phi=p} = 0$, we have $f(p + \epsilon; p) = \frac{\epsilon^2}{2} \frac{d^2}{d\phi^2} f(\phi; p)|_{\phi=p} + O(\epsilon^3)$. Hence to satisfy the condition in (7). i.e. for

$$f(\phi; p) \quad \begin{cases} \geq 0, & \phi \leq p; \\ \leq 0, & \phi \geq p; \end{cases}$$

we must have $\frac{d^2}{d\phi^2} f(\phi; p)|_{\phi=p} = 0$. By Equation (8), this is equivalent to

$$(a-b)\frac{1-2q}{q\bar{q}} - \frac{1-2p}{p\bar{p}} = 0.$$

One can solve above equation explicitly and the only feasible solution is $p = p^*$. Hence $p = p^*$ is necessary.

To show that it is sufficient, assume $p = p^*$. From Lemma 1, we have that $\frac{d^2}{d\phi^2} f(\phi; p^*)$ is decreasing w.r.t. ϕ. Since $\frac{d^2}{d\phi^2} f(\phi; p^*)|_{\phi=p^*} = 0$, then $\frac{d^2}{d\phi^2} f(\phi; p^*) \leq 0$ for $\phi \geq p^*$. This implies that $\frac{d}{d\phi} f(\phi; p^*)$ is decreasing for $\phi \geq p^*$. As $\frac{d}{d\phi} f(\phi; p^*)|_{\phi=p^*} = 0$, we have $\frac{d}{d\phi} f(\phi; p^*) \leq 0$ for $\phi \geq p^*$. Consequently $f(\phi; p^*)$ is decreasing for $\phi \geq p^*$. Finally, as $f(p^*; p^*) = 0$, we obtain $f(\phi; p^*) \leq 0$ when $\phi \geq p^*$. The analysis for $\phi \leq p^*$ is similar. This completes the proof. □

Remark 3. This theorem implies that for the binary symmetric channel, the only P_X for which we have the concavity of $D((W\Phi_{X,t})_Y\|(WP)_Y)$ with respect to t is the uniform distribution.

When $p \neq p^*$, the next proposition establishes a one-sided concavity result for the output relative entropy.

Proposition 1. *In the same setting as Theorem 1, if $p > p^*$, $D((W\Phi_{X,t})_Y\|(WP)_Y)$ is concave for $p \leq \phi \leq 1$. Similarly, if $p < p^*$, $D((W\Phi_{X,t})_Y\|(WP)_Y)$ is concave for $0 \leq \phi \leq p$.*

Proof. We will prove the claim when $p > p^*$. The case where $p < p^*$ is analogous. We will show that $\frac{d^2}{d\phi^2} f(\phi;p)|_{\phi=p}$ is decreasing w.r.t. p first. By Equation (8), we have

$$g(p) := \frac{d^2}{d\phi^2} f(\phi;p)|_{\phi=p} = (a-b)^2 \frac{1-2p}{p\bar{p}} - (a-b)^3 \frac{1-2q}{q\bar{q}}.$$

Since $q = ap + b\bar{p}$, we deduce that

$$\begin{aligned} \frac{d}{dp} g(p) &= (a-b)^2 \Big(-\frac{1}{p^2} - \frac{1}{\bar{p}^2} \Big) - (a-b)^4 \Big(-\frac{1}{q^2} - \frac{1}{\bar{q}^2} \Big) \\ &= -(a-b)^2 \frac{b(2(a-b)p+b)}{p^2 q^2} - (a-b)^2 \frac{\bar{a}(2(a-b)\bar{p}+\bar{a})}{\bar{p}^2 \bar{q}^2} \quad (9) \\ &= -(b-a)^2 \frac{\bar{b}(2(b-a)p+\bar{b})}{p^2 \bar{q}^2} - (b-a)^2 \frac{a(2(b-a)\bar{p}+a)}{\bar{p}^2 q^2} \quad (10) \end{aligned}$$

Therefore, irrespective of the sign of $a-b$ (see (9) or (10)), we have $\frac{d}{dp} g(p) \le 0$. Since $g(p^*) = 0$ and $g(p)$ is decreasing w.r.t p, $g(p) \le 0$ when $p \ge p^*$. Moreover, by Lemma 1, we have $\frac{d^2}{d\phi^2} f(\phi;p)$ is decreasing w.r.t ϕ and hence $\frac{d^2}{d\phi^2} f(\phi;p) \le g(p) \le 0$ for all $\phi \ge p$. Since $f(p;p) = 0$ and $\frac{d}{d\phi} f(\phi;p)|_{\phi=p} = 0$, we have $f(\phi;p) \le 0$ for all $\phi \ge p$. This implies $D((W\Phi_{X,t})_Y \| (WP)_Y)$ is concave with t when $p \le \phi \le 1$. ☐

III. Concavity over a 2-to-n Channel

We now generalize our result from binary outputs to 2-to-n channels for arbitrary finite output dimension n. To do so, we follow the same approach to find the p such that when we make the input divergence linear in t, the output divergence becomes concave in t. The key difference is that one is unable to explicitly identify the p^*, We denote the channel as

$$W(y|x) = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ \vdots & \vdots \\ a_n & b_n \end{bmatrix} \quad (11)$$

Here the matrix entry $W_{ij} = P(Y = i | X = j)$. The differential equation that makes the input divergence $D_2(\phi\|p)$ linear is same as the binary channel case, as is shown in Equation (5). However, the expression for the output divergence $D((W\Phi_{X,t})_Y \| (WP)_Y)$ is different. Define $q_i = a_i p + b_i \bar{p}$ and $\psi_i = a_i \phi + b_i \bar{\phi}$. Denote $\boldsymbol{\psi} = (\psi_1, \cdots, \psi_n)$, $\boldsymbol{q} = (q_1, \cdots, q_n)$ and $D((W\Phi_{X,t})_Y \| (WP)_Y) = D(\boldsymbol{\psi} \| \boldsymbol{q})$. We have

$$\begin{aligned} D(\boldsymbol{\psi}\|\boldsymbol{q}) &= \sum_{i=1}^n \psi_i \ln \frac{\psi_i}{q_i} \\ \frac{d^2}{dt^2} D(\boldsymbol{\psi}\|\boldsymbol{q}) &= \sum_{i=1}^n \left(\psi_i'' \ln \frac{\psi_i}{q_i} + \frac{\psi_i'^2}{\psi_i} \right) \\ &= \sum_{i=1}^n \left[(a_i - b_i)\phi'' \ln \frac{\psi_i}{q_i} + \frac{(a_i-b_i)^2 \phi'^2}{\psi_i} \right] \quad (12) \\ &\overset{(a)}{=} \sum_{i=1}^n \Bigg[-(a_i - b_i) \frac{\phi'^2}{\phi\bar{\phi}} \ln \frac{\psi_i}{q_i} \left(\ln \frac{\phi\bar{p}}{\bar{\phi}p} \right)^{-1} \\ &\quad + \frac{(a_i-b_i)^2 \phi'^2}{\psi_i} \Bigg] \end{aligned}$$

where $\psi_i' = \frac{d\psi_i}{dt}$ and $\psi_i'' = \frac{d^2\psi_i}{dt^2}$.

Here we used Equation (5) in equality (a). Requiring, the output relative entropy to be concave, i.e. the second-order derivative to be non-positive, is then equivalent to

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

Taking derivatives of $f(\phi;p)$ w.r.t. ϕ, we have

$$\begin{aligned} \frac{d}{d\phi} f(\phi;p) &= \ln \frac{\phi\bar{p}}{\bar{\phi}p} \sum_{i=1}^n \frac{(a_i-b_i)^2(-a_i\phi^2 + b_i\bar{\phi}^2)}{(a_i\phi + b_i\bar{\phi})^2} \\ &\overset{(a)}{=} \ln \frac{\phi\bar{p}}{\bar{\phi}p} \sum_{i=1}^n \frac{a_i b_i (a_i - b_i)}{(a_i\phi + b_i\bar{\phi})^2}. \end{aligned} \quad (13)$$

Equality (a) involves a bit of algebraic manipulations along with the observation that $\sum_{i=1}^n a_i = \sum_{i=1}^n b_i = 1$. The second derivative can be expressed in terms of the first derivative according to

$$\begin{aligned} \frac{d^2}{d\phi^2} f(\phi;p) &= \frac{1}{\phi\bar{\phi}} \sum_{i=1}^n \frac{a_i b_i (a_i - b_i)}{(a_i\phi + b_i\bar{\phi})^2} - 2 \ln \frac{\phi\bar{p}}{\bar{\phi}p} \sum_{i=1}^n \frac{(a_i-b_i)^2 a_i b_i}{(a_i\phi + b_i\bar{\phi})^3} \\ &= \frac{1}{\phi\bar{\phi}} \left(\frac{d}{d\phi} f(\phi;p) \right) \left(\ln \frac{\phi\bar{p}}{\bar{\phi}p} \right)^{-1} - 2 \ln \frac{\phi\bar{p}}{\bar{\phi}p} \sum \frac{(a_i-b_i)^2 a_i b_i}{(a_i\phi + b_i\bar{\phi})^3}. \end{aligned} \quad (14)$$

Finally the third derivative can be expressed as

$$\begin{aligned} \frac{d^3}{d\phi^3} f(\phi;p) &= \left(\frac{1}{\bar{\phi}^2} - \frac{1}{\phi^2} \right) \sum_{i=1}^n \frac{a_i b_i (a_i - b_i)}{(a_i\phi + b_i\bar{\phi})^2} \\ &\quad + 6 \ln \frac{\phi\bar{p}}{\bar{\phi}p} \sum \frac{(a_i-b_i)^3 a_i b_i}{(a_i\phi + b_i\bar{\phi})^4} - \frac{4}{\phi\bar{\phi}} \sum \frac{a_i b_i (a_i-b_i)^2}{(a_i\phi + b_i\bar{\phi})^3}. \end{aligned}$$

We can now generalize Theorem 1 to 2-to-n channels case. Denote $\boldsymbol{a} = (a_1, \cdots, a_n), \boldsymbol{b} = (b_1, \cdots, b_n)$.

Theorem 2. *For a 2-to-n channel represented as Equation* (11). *we reparameterize the input distribution according to Equation* (5). *If*

$$(a_i - b_i) a_i b_i = 0 \quad \forall i,$$

then the output relative entropy $D((W\Phi_{X,t})_Y \| (WP)_Y)$ is linear w.r.t. t. Else, the output relative entropy $D((W\Phi_{X,t})_Y \| (WP)_Y)$ is concave w.r.t. t under such reparameterization, if and only if $p = p^*$ where p^* is the unique solution to

$$\sum_{i=1}^n \frac{(a_i - b_i) a_i b_i}{(p a_i + \bar{p} b_i)^2} = 0. \quad (15)$$

Proof. If $(a_i - b_i) a_i b_i = 0 \ \ \forall i$, then from (13) we have that $f(\phi;p)$ is a constant in ϕ, and setting $\phi = p$, implies that $f(\phi;p) = 0$ for all ϕ. This implies that $\frac{d^2}{dt^2} D(\boldsymbol{\psi}\|\boldsymbol{q}) = 0$ and hence the output relative entropy $D((W\Phi_{X,t})_Y \| (WP)_Y)$ is linear w.r.t. t.

Now assume that there exists some i such that $(a_i - b_i)a_ib_i \neq 0$. Let $g(p) := \sum_{i=1}^n \frac{(a_i-b_i)a_ib_i}{(a_ip+b_i\bar{p})^2}$. Observe that $g(p)$ is decreasing since

$$\frac{d}{dp}g(p) = -\sum_{i=1}^n \frac{2(a_i-b_i)^2 a_ib_i}{(a_ip+b_i\bar{p})^3} < 0,$$

for all $p \in (0,1)$. Note that $g(0) = \sum_{i=1}^n \frac{(a_i-b_i)^2}{b_i} \geq 0$, $g(1) = -\sum_{i=1}^n \frac{(a_i-b_i)^2}{a_i} \leq 0$ (along with the observation that $\sum_{i=1}^n a_i = \sum_{i=1}^n b_i = 1$). Therefore we conclude that $g(p) = 0$ has a unique root p^* over $[0,1]$.

Since $f(p;p) = \frac{d}{d\phi}f(p;p)|_{\phi=p} = 0$ for any p, by considering the Taylor expansion at $\phi = p$ we see that the condition

$$f(\phi;p) \quad \begin{cases} \geq 0, & \phi \leq p; \\ \leq 0, & \phi \geq p; \end{cases}$$

forces $\frac{d^2}{d\phi^2}f(\phi;p)|_{\phi=p} = 0$. Therefore from Equation (14), require $g(p) = 0$ or that $p = p^*$ is necessary.

We now argue that the above condition is also sufficient. By Equation (14), we have $f(p^*;p^*) = \frac{d}{d\phi}f(\phi;p^*)|_{\phi=p^*} = \frac{d^2}{d\phi^2}f(\phi;p^*)|_{\phi=p^*} = 0$, and

$$\frac{d^3}{d\phi^3}f(\phi;p^*)|_{\phi=p^*} = -\frac{4}{p^*\bar{p}^*}\sum_{i=1}^n \frac{(a_i-b_i)^2 a_ib_i}{(a_ip^*+b_i\bar{p}^*)^3} < 0. \quad (16)$$

Using Lemma 2 completes the proof. □

Lemma 2. *Consider a real function $f(\phi) : (0,1) \to \mathbf{R}$ and assume $f \in \mathbf{C}^4$, i.e. four times differentiable,and satisfies the following properties:*

1) *$f(p) = f'(p) = f''(p) = 0$, and $f'''(p) < 0$ for some $p \in (0,1)$;*
2) *$f''(\phi) = a(\phi) \cdot f'(\phi) + b(\phi)$, where $a(\phi) > 0$ and $b(\phi) \leq 0$ for $\phi \in (p,1)$; while $a(\phi) < 0$ and $b(\phi) \geq 0$ for $\phi \in (0,p)$.*

Then we have $f(\phi) \leq 0$ for $\phi \in (p,1)$; and $f(\phi) \geq 0$ for $\phi \in (0,p)$.

Proof. From the Taylor expansion of $f'(\phi)$ at p, we have $f'(\phi) = \frac{f'''(p)}{2}(\phi-p)^2 + O((\phi-p)^2)$. Since $f'''(p)$ is strictly less than zero, then there must exist some positive constant $q \in (p,1)$, such that for $p < \phi \leq q$, we have $f'(\phi) < 0$. Suppose there is some $s \in (q,1)$, such that $f'(s) > 0$. $f'(p) = 0, f'(\phi) < 0$, $f'(s) > 0$ and $f'(\phi)$ is continuous over $\phi \in (p,s)$ imply that the minimum of $f'(\phi)$ over $\phi \in [p,s]$ exists and must be attained by some interior minimizer $\phi_0 \in (p,s)$, and $f'(\phi_0) < 0$. Also we have $f''(\phi_0) = 0$ by local optimality conditions for interior minimizers. Since $a(\phi) > 0$ and $b(\phi) \leq 0$ for $\phi \in (p,1)$, we obtain

$$0 = f''(\phi_0) = a(\phi_0) \cdot f'(\phi_0) + b(\phi_0) \leq a(\phi_0) \cdot f'(\phi_0) < 0.$$

Contradiction arises! Hence such an s cannot exist. This guarantees $f'(\phi) \leq 0$ for $\phi \in (p,1)$ and therefore $f(\phi) \leq 0$ for $\phi \in (p,1)$. The other side $\phi \in (0,p)$ can be proved by similar arguments. □

We then give an alternate proof of the sufficiency part in Theorem 2 as the following, without applying the above Lemma 11.

Alternate Proof of sufficiency. We note that $\frac{d}{d\phi}f(\phi;p^*) = g(\phi)\ln\frac{\phi\bar{p}^*}{\bar{\phi}p^*}$. Here $g(\phi) = \sum_{i=1}^n \frac{(a_i-b_i)a_ib_i}{(a_i\phi+b_i\bar{\phi})^2}$ as defined in the previous proof. We know that $g(\phi)$ is decreasing over $\phi \in [0,1]$, and hence $g(\phi) \leq 0$ for $\phi \geq p^*$. Also note that $\ln\frac{\phi\bar{p}^*}{\bar{\phi}p^*} \geq 0$ for $\phi \geq p^*$. Then we have $\frac{d}{d\phi}f(\phi;p^*) \leq 0$ for $\phi \geq p^*$, which further guarantees $f(\phi;p^*) \leq 0$ for $\phi \geq p^*$. The other side ($\phi \leq p^*$) can be proved analogously. □

Conclusion and Further Work

We generalized the convexity result of Wyner and Ziv [11] for the binary symmetric channel to channels with binary inputs. This allows us to reformulate certain non-convex optimization problems as convex optimization problems. More importantly, it shows that for such optimization problems any local extremizer is also a global extremizer.

It is worth trying to generalize these results to non-binary alphabets. The main issue is in the fact that there are multiple paths that connect two distributions on a probability simplex. For differential entropies and AWGN channels, such a result (along the heat flow) has recently been obtained in [12]. We also hope that some of the techniques that we used to prove the concavity in the case where output alphabet is non-binary can be useful in generalizing such results.

References

[1] V. Anantharam, A. Gohari, and C. Nair, "On the evaluation of marton's inner bound for two-receiver broadcast channels," *IEEE Transactions on Information Theory*, vol. 65, no. 3, pp. 1361–1371, March 2019.

[2] C. Nair, "On Marton's achievable region: Local tensorization for product channels with a binary component," in *2020 Information Theory and Applications Workshop (ITA)*, 2020.

[3] Y. Geng, V. Jog, C. Nair, and Z. V. Wang, "An information inequality and evaluation of Marton's inner bound for binary input broadcast channels," *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4095–4105, 2013.

[4] H. Weingarten, Y. Steinberg, and S. S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936–3964, Sept 2006.

[5] S. Sra, N. K. Vishnoi, and O. Yildiz, "On Geodesically Convex Formulations for the Brascamp-Lieb Constant," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*, ser. Leibniz International Proceedings in Informatics (LIPIcs), E. Blais, K. Jansen, J. D. P. Rolim, and D. Steurer, Eds., vol. 116. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018, pp. 25:1–25:15. [Online]. Available: http://drops.dagstuhl.de/opus/volltexte/2018/9429

[6] R. F. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Info. Theory*, vol. IT-21(6), pp. 629–637, November, 1975.

[7] R. Ahlswede and P. Gács, "Spreading of sets in product spaces and hypercontraction of the markov operator," *The Annals of Probability*, pp. 925–939, 1976.

[8] V. Anantharam, A. Gohari, S. Kamath, and C. Nair, "On hypercontractivity and a data processing inequality," in *2014 IEEE International Symposium on Information Theory (ISIT'2014)*, Honolulu, USA, Jun. 2014, pp. 3022–3026.

[9] I. Csiszar and J. Korner, *Information theory: Coding theorems for discrete memoryless systems*. Cambridge University Press, 1 2011.

[10] O. Ordentlich and Y. Polyanskiy, "Strong data processing constant is achieved by binary inputs," 2020.

[11] A. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications: Part I," *IEEE Trans. Inform. Theory*, vol. IT-19, no. 6, pp. 769–772, Nov 1973.

[12] M. Ledoux, C. Nair, and Y. Wang, *Log-convexity of Fisher information along heat flow*, 2021, available at http://chandra.ie.cuhk.edu.hk/pub/papers/NIT/Log-cvx.pdf.