Log-convexity of Fisher information along heat flow

Michel Ledoux
Institut de Mathématiques de Toulouse
University of Toulouse – Paul-Sabatier
Toulouse, France
Email: ledoux@math.univ-toulouse.fr

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

Abstract

This paper establishes the log-convexity of Fisher information for scalar random variables along the heat flow, thus resolving a conjecture posed in [1]. The convexity result can also be interpreted along similar lines as the convexity of $H_2(p * H_2^{-1}(u))$ in u, established in [2], where $H_2(x) = -x\log_2(x) - (1-x)\log_2(1-x)$ denotes the binary entropy function.

I. Introduction

A. Motivation

The primary motivation for this work comes from optimization problems of the following type, that occur often times in evaluation of achievable rate regions or outer bounds to the capacity regions in network information theory settings. Let $T_{Y|X}$ denote a channel that maps input distributions μ_X into output distributions $\mu_Y = T\mu_X$. If X and Y takes values in a finite alphabet space, then consider the problem of computing the maximum, over μ_X, of

$$F_\lambda(\mu_X) := \lambda H(\mu_X) - H(T\mu_X),$$

where $H(\mu_X) = -\sum_{x\in\mathcal{X}} \mu_X(x)\log\mu_X(x)$ denotes the Shannon entropy of X; and $\lambda \geq 0$ is a fixed constant. When $\lambda \geq 1$, it is immediate from the data-processing inequality that the functional $F_\lambda(\mu_X)$ is concave in μ_X. However for $\lambda \in [0,1)$, this is not necessarily true. In particular for $\lambda = 0$, $F_0(\mu_X)$ is convex in μ_X. Therefore, from a optimization perspective, computing the optimizers of $F_\lambda(\mu_X)$ becomes a non-convex optimization problem at least for some values of λ in the range $[0,1)$.

When the channel $T_{Y|X}$ is the binary-symmetric-channel (BSC), say with crossover probability p, consider the following reparameterization of μ_X, defined by $\mu_X(0) = H_2^{-1}(u)$, where $H_2^{-1} : [0,1] \mapsto [0,\frac{1}{2}]$ denotes the inverse binary entropy function. Under this reparameterization, for $BSC(p)$, observe that

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

It was shown in [2] that $H_2(p*H_2^{-1}(u))$ is convex in u and hence $\lambda u - H_2(p*H_2^{-1}(u))$ is a concave function in u for any λ. Therefore this non-linear parameterization converted the non-convex optimization problem to a convex-optimization problem. It is also worth remarking that the convexity of $H_2(p*H_2^{-1}(u))$ was developed by Wyner and Ziv in the context of evaluating the superposition-coding region for a degraded binary symmetric broadcast channel.

Additive White Gaussian Noise channels are in many ways the continuous analogue of Binary Symmetric Channels. Therefore it is natural to see if there is an analogous result in the additive Gaussian noise setting, where under a suitable parameterization of μ_X, $h(\mu_X)$ - the differential entropy - becomes linear in the parameter and $h(T_G\mu_X)$ becomes convex in the parameter, where T_G refers to the Markov operator corresponding to the channel with additive Gaussian noise W.

For distributions on binary alphabets, there is only one degree of freedom and hence the parameterization of $\mu_X(0) = H_2^{-1}(u)$ is forced on us, if we wish to make $H_2(\mu_X)$ linear. In the continuous world we assume that μ_X evolves along the heat flow, i.e. $X_t := X + \sqrt{t}Z, t > 0$, where Z is the standard Gaussian and independent of X. Therefore we seek a parameterization $t = \phi(u)$ such that $h(X + \sqrt{\phi(u)}Z)$ is linear in u and investigate whether, the output entropy, $h(\mu_Y) = h(X + \sqrt{\phi(u)}Z + W)$ is convex in u, where W is some Gaussian independent of X and Z. Let μ_t^X denote the distribution of $X_t = X + \sqrt{t}Z$. A bit of algebra immediately shows that this question is equivalent to asking whether the Fisher information $I(\mu_t^X)$ is log-convex in t, for all random variables X (see Remark 2).

A second motivation for this work comes from [1], where the authors study the signs of the higher order derivatives of $g_X(t) := h(\mu_t^X)$. Let $g_X^{(k)}(t) := \frac{\partial^k}{\partial t^k} h(\mu_t^X)$. It had been known earlier that $g_X^{(1)}(t) \geq 0$, $g_X^{(2)}(t) \leq 0$; and in [1] the authors showed that $g_X^{(3)}(t) \geq 0$, $g_X^{(4)}(t) \leq 0$ using techniques in [3], which was in turn motivated by calculations of Bakry. The authors further conjectured that $g_X^{(k)}(t) \geq 0$, if k is odd and $g_X^{(k)}(t) \leq 0$ if k is even; or equivalently that $I(\mu_t^X) = 2g_X^{(1)}(t)$ is a completely monotone function of t, for all X. Of course, such a conjecture is also implicit in the 1966 paper [4] by McKean. They also made a weaker conjecture (Conjecture 2 in [1]) that $I(\mu_t^X)$ is log-convex in t, a statement that would follow from the complete monotonicity of $I(\mu_t^X)$.

The main result of this paper is establishing that $I(\mu_t^X)$ is log-convex in t, thus resolving affirmatively Conjecture 2 in [1]. We do this by extending the ideas developed in [1] and [5].

B. Preliminaries

Given a random variable X on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $\mathbb{R}$, let the cumulative distribution function of X be $\tilde{F}(x) := Pr(X \leq x)$, $x \in \mathbb{R}$. For Z some independent standard Gaussian random variable with mean zero and variance one, consider $X_t := X + \sqrt{t}Z$, $t > 0$, with probability density function $f_t(x)$ with respect to the Lebesgue measure on $\mathbb{R}$. The density $f_t(x)$, $x \in \mathbb{R}$, can be written as

$$f_t(x) = \int_{\mathbb{R}} \frac{-z}{\sqrt{2\pi t}} e^{-\frac{z^2}{2t}} \tilde{F}(x-z) dz.$$

It is well-known in literature, e.g., [6], that the probability density function $f_t(x)$ of X_t is always upper bounded by $1+t$, strictly positive and infinitely differentiable with respect to $x \in (-\infty, \infty)$ and $t \in (0, \infty)$, and satisfy that

$$\lim_{|x| \to \infty} \frac{\partial^n f_t(x)}{\partial x^n} = 0, \forall n \in \mathbb{Z}_+.$$

Besides, $f_t(x)$ also satisfies the heat equation, see, e.g., [7].

$$\frac{\partial}{\partial t} f_t(x) = \frac{1}{2} \frac{\partial^2}{\partial x^2} f_t(x). \tag{1}$$

The differential entropy of X_t, $h(X_t)$, $t > 0$, is defined as

$$h(X_t) = -\int_{\mathbb{R}} f_t(x) \ln f_t(x) dx.$$

When X has a finite variance P, $h(X_t)$ exists and is maximized by X following a Gaussian distribution with variance P.

The Fisher information of X_t is defined as

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

One can verify that the Fisher information $I(\mu_t^X), t > 0$, always exists and is infinitively differentiable with respect to $t \in (0, \infty)$, see, e.g., [1].

The Fisher information $I(\mu_t^X)$ is closely related to the differential entropy of X_t via the de Bruijin's identity when X has a finite variance, see, e.g., [8]

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

Conjecture 2 in [1] postulates that $\ln I(\mu_t^X)$ is convex in $t > 0$. In this paper, a proof to this conjecture is presented along the lines of the arguments in [1] and [5].

C. Notations and previous results

For convenience of writing, we will suppress the dependence on t and write $v(x) := \ln f_t(x), t > 0$, and $v_k(x) := \frac{\partial^k \ln f_t(x)}{\partial x^k}, k \in \mathbb{Z}_+$, i.e., $v_k(x)$ is the k-th derivative of v as a function of $x \in \mathbb{R}$. Well-definedness of $v_k(x)$ for any $k \in \mathbb{Z}_+$ follows from the known properties of $f_t(x)$.

Proposition 1 (Proposition 2 in [1])**.** *For any $r, m_i, k_i \in \mathbb{Z}_+$,*

$$\int_{\mathbb{R}} \left| \prod_{i=1}^{r} v_{k_i}^{m_i}(x) \right| f_t(x) dx < \infty,$$

and

$$\lim_{|x| \to \infty} \left| \prod_{i=1}^{r} v_{k_i}^{m_i}(x) \right| f_t(x) = 0.$$

We define $\langle \varphi \rangle := \int_{\mathbb{R}} \varphi f_t(x) dx$ to denote the integration with respect to the probability measure $f_t(x)$. Under this notation

$$I(\mu_t^X) = \langle v_1^2 \rangle. \tag{3}$$

The following lemma is needed in our proof.

Lemma 1 (Lemma 3 in [5])**.** *For $k \geq 2$, let $\varphi(x)$ be some function continuously differentiable with respect to x satisfying that $\lim_{|x|\to\infty} \varphi v_{k-1} f_t = 0$, then*

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

One can see that this lemma follows from the basic integration by parts property. We present the short proof here for being self-contained.

Proof.

$$\begin{aligned}
\langle \varphi v_k + \varphi v_1 v_{k-1} + \frac{\partial \varphi}{\partial x} v_{k-1} \rangle &= \int_{\mathbb{R}} \left(\varphi v_k f_t + \varphi v_{k-1} \frac{\partial f_t}{\partial x} + \frac{\partial \varphi}{\partial x} v_{k-1} f_t \right) dx \\
&\overset{(a)}{=} \int_{\mathbb{R}} \left(\frac{\partial}{\partial x} \varphi v_{k-1} f_t \right) dx \\
&= \varphi v_{k-1} f_t |_{-\infty}^{\infty} \\
&\overset{(b)}{=} 0.
\end{aligned}$$

Equality (a) follows from the integration by parts property, and equality (b) follows from the condition that $\lim_{|x|\to\infty} \varphi v_{k-1} f_t = 0$. ∎

Notice that by Proposition 1 we could choose φ in Lemma 1 to be in the form of $\prod_{i=1}^{r} v_{k_i}^{m_i}(x)$, where $r, m_i, k_i \in \mathbb{Z}_+$.

Lemma 2 ([1], [5])**.** *Let $\varphi(x)$ be some function continuously differentiable with respect to x satisfying that $\lim_{|x|\to\infty} \varphi v_1 f_t = 0$. For $k \geq 0$, the following hold:*

$$\begin{aligned}
\frac{\partial}{\partial t} v_k &= \frac{1}{2} \left(v_{k+2} + \sum_{i=0}^{k} \binom{k}{i} v_{i+1} v_{k-i+1} \right), \\
\frac{\partial}{\partial t} \langle \varphi \rangle &= \langle \frac{\partial}{\partial t} \varphi - \frac{1}{2} \frac{\partial \varphi}{\partial x} v_1 \rangle.
\end{aligned}$$

Proof. The proof idea is to interchange integral and derivatives by Proposition 1 and the Dominated Convergence Theorem, and the calculations follow from the following observations (for details, see Appendix A in [5]). We present the outline here for being rather self-contained.

$$\begin{aligned}
2 \frac{\partial}{\partial t} v_k &= 2 \frac{\partial}{\partial t} \left(\frac{\partial^k}{\partial x^k} \ln f_t(x) \right) \\
&= 2 \frac{\partial^k}{\partial x^k} \left(\frac{\partial}{\partial t} \ln f_t(x) \right) \\
&\overset{(a)}{=} \frac{\partial^k}{\partial x^k} \left(\frac{\frac{\partial^2}{\partial x^2} f_t(x)}{f_t(x)} \right) \\
&= \frac{\partial^k}{\partial x^k} \left(v_2 + v_1^2 \right) \\
&\overset{(b)}{=} v_{k+2} + \sum_{i=0}^{k} \binom{k}{i} v_{i+1} v_{k-i+1}.
\end{aligned}$$

Equality (a) is due to the heat equation (1) and (b) can be established by mathematical induction.

For the second part, observe that

$$\begin{aligned}
\frac{\partial}{\partial t} \langle \varphi \rangle &= \langle \frac{\partial}{\partial t} \varphi \rangle + \int_{\mathbb{R}} \varphi \frac{\partial f_t}{\partial t} dx \\
&\overset{(a)}{=} \langle \frac{\partial}{\partial t} \varphi \rangle + \frac{1}{2} \int_{\mathbb{R}} \varphi \frac{\partial^2 f_t}{\partial x^2} dx \\
&\overset{(b)}{=} \langle \frac{\partial}{\partial t} \varphi \rangle - \frac{1}{2} \int_{\mathbb{R}} \frac{\partial \varphi}{\partial x} \frac{\partial f_t}{\partial x} dx \\
&= \langle \frac{\partial}{\partial t} \varphi \rangle - \frac{1}{2} \langle \frac{\partial \varphi}{\partial x} v_1 \rangle.
\end{aligned}$$

Equality (a) is again due to the heat equation (1) and (b) follows from integration by parts and the assumption that $\lim_{|x|\to\infty} \varphi v_1 f_t = 0$. $\square$

One can compute the derivatives of the Fisher information $I(\mu_t^X)$ with respect to t as following, see [9] and [5].

Lemma 3 ([1], [5]). *For $t > 0$, Fisher information $I(\mu_t^X)$ and its derivatives up to second order can be expressed as:*

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

Proof. In the interest of being self-contained, we outline the proof via applications of Lemmas 2 and 1. Observe that

$$\begin{aligned}
\frac{d}{dt} I(\mu_t^X) &= \frac{d}{dt} \langle v_1^2 \rangle \\
&\overset{(a)}{=} \langle 2v_1 \frac{\partial v_1}{\partial t} - v_2 v_1^2 \rangle \\
&\overset{(b)}{=} \langle v_1(v_3 + 2v_1 v_2) - v_2 v_1^2 \rangle \\
&\overset{(c)}{=} -\langle v_2^2 \rangle.
\end{aligned}$$

Here $(a), (b)$ follow from Lemma 2, and (c) follows from Lemma 1 by setting $\varphi = v_1$ and $k = 3$. Similarly, note that

$$\begin{aligned}
\frac{d^2}{dt^2} I(\mu_t^X) &= -\frac{d}{dt} \langle v_2^2 \rangle \\
&\overset{(a)}{=} \langle -2v_2 \frac{\partial v_2}{\partial t} + v_2 v_3 v_1 \rangle \\
&\overset{(b)}{=} \langle -v_2(v_4 + 2v_1 v_3 + 2v_2^2) + v_2 v_3 v_1 \rangle \\
&\overset{(c)}{=} \langle v_3^2 - 2v_2^3 \rangle \\
&\overset{(d)}{=} \langle v_3^2 + 2v_1^2 v_2^2 + 4v_1 v_2 v_3 \rangle.
\end{aligned}$$

Here $(a), (b)$ follow from Lemma 2, (c) follows from Lemma 1 by setting $\varphi = v_2$ and $k = 4$, and (d) follows from Lemma 1 by setting $\varphi = v_2^2$ and $k = 2$. $\square$

Remark 1. There are several equivalent ways of expressing $\frac{d^2}{dt^2} I(\mu_t^X)$ using Lemma 2. For instance, [5] expressed it as $\langle v_3^2 - 2v_2^3 \rangle$. We choose this particular representation, $\langle v_3^2 + 2v_1^2 v_2^2 + 4v_1 v_2 v_3 \rangle$, as it turns out to be useful to prove the log-convexity of Fisher information.

II. MAIN RESULT

Theorem 1. *Let X be a random variable on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $\mathbb{R}$, and Z some independent standard Gaussian random variable. Consider $X_t := X + \sqrt{t} Z, t > 0$, with probability density function $f_t(x)$ with respect to the Lebesgue measure on $\mathbb{R}$.*

The Fisher information of X_t is log-convex in t, i.e.

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

is convex in t.

Proof. Log-convexity of Fisher information is equivalent to showing

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

Using Lemma 3, this is equivalent to showing

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

In Lemma 1, the choices that $k = 2, \varphi = v_2$ and that $k = 2, \varphi = v_1^2$ will lead to the following two equalities respectively

$$\langle v_2^2 + v_1^2 v_2 + v_1 v_3 \rangle = 0 \tag{5}$$
$$\langle v_1^4 + 3v_1^2 v_2 \rangle = 0. \tag{6}$$

Consequently, for any $\alpha \in \mathbb{R}$ we have

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

The Cauchy-Schwarz inequality yields,

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

Thus to show inequality (4), it suffices to show that

$$\langle (v_3 + \alpha v_1 v_2 - \frac{1-\alpha}{3} v_1^3)^2 \rangle \leq \langle v_3^2 + 2 v_1^2 v_2^2 + 4 v_1 v_2 v_3 \rangle \tag{7}$$

holds for some $\alpha \in \mathbb{R}$. Expanding, (7) is equivalent to

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

In Lemma 1, the choices that $k = 3, \varphi = v_1^3$ and that $k = 2, \varphi = v_1^4$ will lead to the following two equalities respectively.

$$\langle v_1^3 v_3 + v_2 v_1^4 + 3 v_1^2 v_2^2 \rangle = 0$$
$$\langle v_1^6 + 5 v_1^4 v_2 \rangle = 0.$$

Thus proving inequality (7) for some $\alpha \in \mathbb{R}$ is equivalent to proving the following inequality

$$\begin{aligned} \langle (2-\alpha^2) v_1^2 v_2^2 + (4 - 2\alpha) v_1 v_2 v_3 - \frac{1}{9}(1-\alpha)^2 v_1^6 + \frac{2}{3}(1-\alpha) v_1^3 v_3 + \frac{2}{3}\alpha(1-\alpha) v_1^4 v_2 \rangle \\ + \beta \langle v_1^3 v_3 + v_2 v_1^4 + 3 v_1^2 v_2^2 \rangle + \gamma \langle v_1^6 + 5 v_1^4 v_2 \rangle \geq 0 \end{aligned} \tag{8}$$

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

We successively choose the values α, β, γ to eliminate the terms whose signs are not clear: first set $\alpha = 2$ to get rid of $\langle v_1 v_2 v_3 \rangle$, then $\beta = \frac{2}{3}$ to eliminate $\langle v_1^3 v_3 \rangle$, and finally $\gamma = \frac{2}{15}$ to handle $\langle v_1^4 v_2 \rangle$. With these choices, the above inequality (8) reduces to $\frac{1}{45}\langle v_1^6 \rangle \geq 0$, which holds trivially. □

Remark 2. Let $\phi(u)$, with $\phi(0) = 0$ and $\phi(1) = 1$, be the uniquely defined increasing function of u such that $h(X + \sqrt{\phi(u)}Z)$ is linear in u. Then we have

$$0 = \frac{d^2}{du^2} h(X + \sqrt{\phi(u)}Z) = \frac{1}{2}\left(\frac{d^2\phi(u)}{du^2} I(\mu^X_{\phi(u)}) + \left(\frac{d\phi(u)}{du}\right)^2 \frac{d}{d\phi(u)} I(\mu^X_{\phi(u)}) \right).$$

Now, showing that $\frac{d^2}{du^2} h(X + \sqrt{\phi(u)}Z + W) \geq 0$, for $W \sim \mathcal{N}(0, \sigma^2)$ independent of (X, Z), is equivalent to showing that

$$0 \leq \frac{1}{2}\left(\frac{d^2\phi(u)}{du^2} I(\mu^{X+W}_{\phi(u)}) + \left(\frac{d\phi(u)}{du}\right)^2 \frac{d}{d\phi(u)} I(\mu^{X+W}_{\phi(u)}) \right).$$

This can be rewritten using the equality above as requiring

$$\frac{\frac{d}{d\phi(u)} I(\mu^{X+W}_{\phi(u)})}{I(\mu^{X+W}_{\phi(u)})} \geq \frac{\frac{d}{d\phi(u)} I(\mu^{X}_{\phi(u)})}{I(\mu^{X}_{\phi(u)})}.$$

Since $I(\mu^{X+W}_{\phi(u)}) = I(\mu^X_{\phi(u_1)})$ for some $u_1 \geq u$, the above inequality is equivalent to showing that

$$\frac{\frac{d}{dt} I(\mu^X_t)}{I(\mu^X_t)}$$

is increasing in t or equivalently, that $\log I(\mu^X_t)$ is convex in t. Thus, the result we showed can be considered as a continuous analogue of the convexity result for BSC established by Wyner and Ziv.

III. Summary and Future Work

We resolved a conjectured about log-convexity of Fisher information stated in [1], in the scalar case. Our investigations also stemmed from understanding the behaviour of certain non-convex optimization problems arising in network information theory.

A. 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

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

where $\mathbf{X}$ and $\mathbf{Z}(\sim \mathcal{N}(0, I_d))$ are independent random vectors taking values in $\mathbb{R}^d$. If $\mathbf{X}$ has independent components, then an application of the Cauchy-Schwarz inequality immediately implies affirmatively the inequality above.

While the techniques applied in the scalar case do have natural extensions to the vector case, preliminary investigations by the authors indicate that these extensions seem insufficient to establish the log-convexity for vector valued random variables.

B. Generalization of convexity of the output entropy

Let us consider a channel given by

$$\mathbf{Y} = A\mathbf{X} + \mathbf{Z}$$

where A is an $l \times d$ (channel-gain) matrix, $\mathbf{X}$ is the input, and $\mathbf{Z}(\sim \mathcal{N}(0, I_l))$ is the additive Gaussian noise. Then one can ask for flows in the space of input distributions, say characterized by X_t, where $h(X_t)$ is linear in t and $h(Y_t)$ is convex in t.

An interesting such flow exists in the space of Gaussian vectors. Let $\mathbf{X}_0 \sim \mathcal{N}(0, K_0)$ and $\mathbf{X}_1 \sim \mathcal{N}(0, K_1)$ be two Gaussian random vectors with $K_0, K_1 \succ 0$. 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 $X_t \sim \mathcal{N}(0, K_t)$. Note that this is a continuous path that connects the distribution of X_0 to that of X_1. Further, observe that $h(X_t)$ is linear in t. It follows from the seminal work in [10], and is well-known, that

$$h(Y_t) = \log |AK_tA^T + I|$$

is convex in t.

From the perspective of non-convex optimization problems that arise in the computation of achievable regions or outer bound in network information theory, it will be very helpful to find similar flows in a more general setting, i.e. outside the space of Gaussian vectors and more generally for larger class of channels. Such results may also be useful in showing the uniqueness of local maximizers in such settings as is observed in settings such as the MIMO Gaussian broadcast channels.

REFERENCES

[1] F. Cheng and Y. Geng, "Higher order derivatives in costa's entropy power inequality," *IEEE Transactions on Information Theory*, vol. 61, no. 11, pp. 5892–5905, 2015.

[2] 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.

[3] C. Villani, "A short proof of the "concavity of entropy power"," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1695–1696, 2000.

[4] H. P. McKean, "Speed of approach to equilibrium for kac's caricature of a maxwellian gas," *Archive for Rational Mechanics and Analysis*, vol. 21, no. 5, pp. 343–367, jan 1966.

[5] X. Zhang, V. Anantharam, and Y. Geng, "Gaussian optimality for derivatives of differential entropy using linear matrix inequalities †," *Entropy (Basel, Switzerland)*, vol. 20, no. 3, March 2018. [Online]. Available: https://europepmc.org/articles/PMC7512698

[6] Y. Geng and C. Nair, "The capacity region of the two-receiver gaussian vector broadcast channel with private and common messages," *IEEE Transactions on Information Theory*, vol. 60, no. 4, pp. 2087–2104, 2014.

[7] A. Stam, "Some inequalities satisfied by the quantities of information of fisher and shannon," *Information and Control*, vol. 2, no. 2, pp. 101 – 112, 1959. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0019995859903481

[8] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley Interscience, 1991.

[9] M. Ledoux, "Heat flow derivatives and minimum mean-square error in gaussian noise," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3401–3409, 2016.

[10] F. Kubo and T. Andô, "Means of positive linear operators," *Mathematische Annalen*, vol. 246, pp. 205–224, 1980.