香港中文大學
The Chinese University of Hong Kong
DEPARTMENTS OF MATHEMATICS AND INFORMATION ENGINEERING

ON THE LOG CONVEXITY OF FISHER INFORMATION

AUTHOR: GUAN YUNRUI(ID: 1155141425)
PROJECT SUPERVISOR: PROF. CHANDRA NAIR

Declaration

I declare that the assignment here submitted is original, contributed by me or members as a group, except for source material explicitly acknowledged. I also acknowledge that I am are aware of the University's policy and regulations on honesty in academic work, and of the disciplinary guidelines and procedures applicable to breaches of such policy and regulations, as contained in the University website `http://www.cuhk.edu.hk/policy/academichonesty/`.

Signed: GUAN YUNRUI Date: November 25, 2022

NOVEMBER 25, 2022

Abstract

McKean conjectured in 1966 that Gaussians optimize derivatives of entropy along the heat flow, which would then imply that the entropy (or rather the Fisher Information) would be completely monotonic in the heat flow parameter. A further weakening of the above conjecture is that the Fisher Information is log-convex in the heat flow parameter. Recently the scalar case of the conjecture was proven in the affirmative and in this report we seek to generalize the result to random vectors. We are able to obtain a proof for mutually independent random vectors, as well as prove the result for log-concave initial distributions. The latter result for log-concave initial distributions can also be obtained as a corollary of Toscani's work.

1 Introduction

1.1 Background

Let X be a random variable with finite variance and Z be a standard Gaussian being independent of X. Define $X_t = X + \sqrt{t}Z$ and μ_t^X denote its probability density function. McKean conjectured that the Fisher information $I(\mu_t^X)$ is a completely monotone function.

In [1] Theorem 1, it was proved that any completely monotone function $f(t)$ is log-convex with respect to t. With this result, we know that if McKean's conjecture is true, then the Fisher information must be log-convex.

The following conjecture can be found in [2].

Conjecture 1. *Log-Convexity Conjecture*

The function $I(\mu_t^X)$ is log-convex. That is,

$$I(\mu_t^X)\frac{d^2}{dt^2}I(\mu_t^X) \geq [\frac{d}{dt}I(\mu_t^X)]^2.$$

Related works are discussed as follows. In [3], the log-convexity property was proved to be true for any scalar random variable. A direct consequence of Theorem 1 in [4] is that if the distribution is log-concave, then the log-convexity of Fisher information holds.

This report is arranged as follows. In Section 1.2, we state some definitions and notations. Then we give some lemmas which are generalizations from [3]. In Section 2.1, we prove the log-convexity for a special case, when the random vector has independent components. In Section 2.2, we gave another proof for the Log-Convexity Conjecture for log-concave distribution.

1.2 Preliminary and definitions

Definition 1. *The Fisher Information is defined as*

$$I(\mu_t^X) = \int_{\mathbb{R}^l} \mu_t^X(x_1, ..., x_l) |\nabla \mu_t^X(x_1, ..., x_l)|^2 dx_1 ... dx_l$$

Now we define some notations which will be used later.

Definition 2. *Denote $v(x_1, ..., x_l) = \ln \mu_t^X(x_1, ..., x_l)$ which maps from $\mathbb{R}^l$ to $\mathbb{R}$. For ease of notation, we sometimes write it as $v(x)$.*

Define $v_k^{k_1,...,k_l}(x) = \frac{\partial^k}{\partial x_1^{k_1} ... \partial x_l^{k_l}} \ln \mu_t^X(x)$.

Definition 3. *Define $\langle \phi \rangle := \int_{\mathbb{R}^l} \phi \mu_t^X(x) dx$ to be the integration with respect to the probability measaure μ_t^X.*

1.2.1 Lemmas

In this subsection, we generalize some lemmas in Wang's thesis [3] from the scalar case to the vector case. The proofs of some of the lemmas below is deferred to the Appendix.

The follow lemma is a vector generalization of Lemma 4.1 in [3].

Lemma 1. *For differentiable ϕ that satisfies the following,*

$$\lim_{|x_j| \to \infty} v_{k-1}^{k_1,...,k_j-1,...,k_l} \phi \mu_t^X = 0$$

we have

$$\langle \phi v_k^{k_1,\dots,k_l} + \phi v_1^{e_j} v_{k-1}^{k_1,\dots,k_j-1,\dots,k_l} + v_{k-1}^{k_1,\dots,k_j-1,\dots,k_l} \frac{\partial}{\partial x_j}\phi \rangle = 0$$

where $e_j = (0, ..., 0, 1, 0, ..., 0)$ *with* 1 *on the j-th entry.*

Remark 1. *This lemma implies some equality constraints which are useful in the proofs and will be frequently used throughout this report.*

Proof. We perform integration by parts with respect to x_j.

$$\begin{aligned}
&\int_{\mathbb{R}^l} \phi v_k^{k_1,\dots,k_l} \mu_t^X dx = \int_{\mathbb{R}^{l-1}} \int_{\mathbb{R}} \phi v_k^{k_1,\dots,k_l} \mu_t^X dx_j dx \\
&= \int_{\mathbb{R}^{l-1}} (v_{k-1}^{k_1,\dots,k_j-1,\dots,k_l} \phi \mu_t^X)|_{-\infty}^{\infty} dx - \int_{\mathbb{R}^l} v_{k-1}^{k_1,\dots,k_j-1,\dots,k_l} \frac{\partial}{\partial x_j}(\phi \mu_t^X) dx \\
&= -\int_{\mathbb{R}^l} v_{k-1}^{k_1,\dots,k_j-1,\dots,k_l} (\frac{\partial}{\partial x_j}\phi \mu_t^X + \phi \frac{\partial}{\partial x_j}\mu_t^X) dx \\
&= -\int_{\mathbb{R}^l} v_{k-1}^{k_1,\dots,k_j-1,\dots,k_l} (\frac{\partial}{\partial x_j}\phi + \phi v_1^{e_j}) \mu_t^X dx
\end{aligned} \tag{1}$$

The last equality follows from

$$\mu_t^X v_1^{e_j} = \frac{\frac{\partial}{\partial x_j}\mu_t^X}{\mu_t^X} \mu_t^X = \frac{\partial}{\partial x_j}\mu_t^X$$

Then we get

$$\langle \phi v_k^{k_1,\dots,k_l} + \phi v_1^{e_j} v_{k-1}^{k_1,\dots,k_j-1,\dots,k_l} + v_{k-1}^{k_1,\dots,k_j-1,\dots,k_l} \frac{\partial}{\partial x_j}\phi \rangle = 0$$

as desired.

□

The following lemma is a vector generalization of the first part of Lemma 4.2 in [3].

Lemma 2. *Under the condition that* $\lim_{|x_i|\to\infty} \phi\mu_t^X v_1^{e_i} = 0$ *for all* i*, we have*

$$\frac{\partial}{\partial t}\langle\phi\rangle = \langle\frac{\partial}{\partial t}\phi - \frac{1}{2}\sum_{i=1}^{l}\frac{\partial}{\partial x_i}\phi \cdot v_1^{e_i}(x)\rangle$$

The following lemma is a vector generalization of the second part of Lemma 4.2 in [3].

Lemma 3. *We have*

$$2\frac{\partial}{\partial t}v_k = \frac{\partial^k}{\partial x_1^{k_1}...\partial x_l^{k_l}}[\sum_{i=1}^{l} v_2^{2e_i} + (v_1^{e_i})^2]$$

when $v_k = v_1^{e_j}$*,*

$$2\frac{\partial}{\partial t}v_1^{e_j} = \frac{\partial}{\partial x_j}[\sum_{i=1}^{l} v_2^{2e_i} + (v_1^{e_i})^2] = \sum_{i=1}^{l} v_3^{2e_i+e_j} + 2v_1^{e_i}v_2^{e_i+e_j}$$

For ease of notation, we write $v_k^{k_1,\ldots,k_l}(x) = v_{1\ldots12\ldots2\ldots l\ldots l}$ where the number of i is k_i. For example, we write $v_3^{1,2}(x) = v_{122}$.

The following three lemmas are about $I(\mu_t^X)$, $\frac{d}{dt}I(\mu_t^X)$ and $\frac{d^2}{dt^2}I(\mu_t^X)$. Similar results corresponding to one-dimensional case can be found in [3].

Lemma 4. *The Fisher information can be written as*

$$I(\mu_t^X) = \langle\sum_{i=1}^{l}(v_1^{e_i})^2\rangle = \langle\sum_{i=1}^{l} v_i^2\rangle$$

Lemma 5. *We can write the derivative of Fisher information w.r.t.* t *as*

$$\frac{d}{dt}I(\mu_t^X) = -\langle\sum_{j=1}^{l}\sum_{i=1}^{l}(v_2^{e_i+e_j})^2\rangle = -\langle\sum_{j=1}^{l}\sum_{i=1}^{l} v_{ij}^2\rangle$$

Lemma 6. *We can write the second order derivative of Fisher information w.r.t. t as*

$$\begin{aligned}\frac{d^2}{dt^2}I(\mu_t^X) =&\langle\sum_{j=1}^{l}\sum_{i=1}^{l}\sum_{w=1}^{l}(v_3^{e_w+e_i+e_j})^2 - 2v_2^{e_i+e_j}v_2^{e_w+e_j}v_2^{e_i+e_w}\rangle \\ =&\langle\sum_{j=1}^{l}\sum_{i=1}^{l}\sum_{w=1}^{l}(v_3^{e_w+e_i+e_j})^2 + 2v_2^{e_w+e_i}v_2^{e_w+e_j}v_1^{e_j}v_1^{e_i} + 2v_1^{e_i}v_2^{e_w+e_i}v_3^{e_w+2e_j} \\ &+ 2v_1^{e_i}v_3^{e_w+e_i+e_j}v_2^{e_w+e_j}\rangle\end{aligned} \tag{2}$$

The second order derivative of $\log I(\mu_t^X)$ is

$$\frac{d^2}{dt^2}\log I(\mu_t^X) = \frac{d}{dt}\left(\frac{\frac{d}{dt}I(\mu_t^X)}{I(\mu_t^X)}\right) = \frac{I(\mu_t^X)\frac{d^2}{dt^2}I(\mu_t^X) - [\frac{d}{dt}I(\mu_t^X)]^2}{I(\mu_t^X)^2}$$

Now, we can re-write condition for log-convexity of Fisher Information using the notation introduced previously. Log-convexity of Fisher information is equivalent to showing the following inequality i.e. $I(\mu_t^X)\frac{d^2}{dt^2}I(\mu_t^X) \geq [\frac{d}{dt}I(\mu_t^X)]^2$. We can re-write this as the following inequality:

$$\langle\sum_{j=1}^{l}\sum_{i=1}^{l}(v_2^{e_i+e_j})^2\rangle^2 \leq \langle\sum_{i=1}^{l}(v_1^{e_i})^2\rangle\langle\sum_{j=1}^{l}\sum_{i=1}^{l}\sum_{w=1}^{l}(v_3^{e_w+e_i+e_j})^2 - 2v_2^{e_i+e_j}v_2^{e_w+e_j}v_2^{e_i+e_w}\rangle \tag{3}$$

2 Proofs of Results

2.1 Independent Case: Mutually Independent Random Vector

In this section, we generalize Wang's result [3] to vector case with independent components.

Theorem 1. *The Cauchy-Schwarz inequality holds under* $\langle\sum_{i=1}^{n}\rangle$*. That is,*

$$\langle\sum_{i=1}^{n} x_i y_i\rangle^2 \leq \langle\sum_{i=1}^{n} x_i^2\rangle\langle\sum_{i=1}^{n} y_i^2\rangle$$

Proof. We present a proof for completeness. By Cauchy-Schwarz inequality, we have

$$|\sum_{i=1}^{n} x_i y_i| \leq (\sum_{i=1}^{n} x_i^2)^{\frac{1}{2}}(\sum_{i=1}^{n} y_i^2)^{\frac{1}{2}}$$

Integration on both sides, we get

$$\mathbb{E}[\sum_{i=1}^{n} x_i y_i] \leq \mathbb{E}[(\sum_{i=1}^{n} x_i^2)^{\frac{1}{2}}(\sum_{i=1}^{n} y_i^2)^{\frac{1}{2}}]$$

By Cauchy-Schwarz inequality for integral, we have $\mathbb{E}[XY] \leq \mathbb{E}[X^2]^{\frac{1}{2}}\mathbb{E}[Y^2]^{\frac{1}{2}}$.

Combine the above equations, we can conclude that

$$\mathbb{E}[\sum_{i=1}^{n} x_i y_i] \leq \mathbb{E}[(\sum_{i=1}^{n} x_i^2)^{\frac{1}{2}}(\sum_{i=1}^{n} y_i^2)^{\frac{1}{2}}] \leq \mathbb{E}[\sum_{i=1}^{n} x_i^2]^{\frac{1}{2}}\mathbb{E}[\sum_{i=1}^{n} y_i^2]^{\frac{1}{2}}$$

which is the desired result. $\square$

Now we consider the Independent setting (mutually independent random vector). In this setting, we can write the 2-dim distribution as $\mu_t^X(x_1, x_2) = \mu_t^{X_1}(x_1)\mu_t^{X_2}(x_2)$. This implies that

$$v_{12} = \frac{\partial^2}{\partial x_1 x_2}(\log \mu_t^{X_1} + \log \mu_t^{X_2}) = 0.$$

For n-dim distribution, $\mu_t^X(x_1, x_2, ..., x_n) = \prod_{j=1}^n \mu_t^{X_j}(x_j)$. Then similarly $v_{ij} = 0$ for all $i \neq j$.

Denote $v_i = v_{i,1}$, $v_{ii} = v_{i,2}$, $v_{iii} = v_{i,3}$. For example, $v_1 = v_{1,1}$, $v_{11} = v_{1,2}$, $v_{111} = v_{1,3}$ and $v_2 = v_{2,1}$, $v_{22} = v_{2,2}$, $v_{222} = v_{2,3}$.

Then we can write

$$I(\mu_t^X)\frac{d^2}{dt^2}I(\mu_t^X) \geq [\frac{d}{dt}I(\mu_t^X)]^2$$

as

$$\langle\sum_{i=1}^n v_{i,2}^2\rangle^2 \leq \langle\sum_{i=1}^n v_{i,1}^2\rangle\langle\sum_{i=1}^n v_{i,3}^2 - 2v_{i,2}^3\rangle = \langle\sum_{i=1}^n v_{i,1}^2\rangle\langle\sum_{i=1}^n v_{i,3}^2 + 2v_{i,1}^2 v_{i,2}^2 + 4v_{i,1}v_{i,2}v_{i,3}\rangle \quad (1)$$

Note: the equality follows from applying Lemma 1

$$\langle v_{i,3}^2 - 2v_{i,2}^3\rangle = \langle v_{i,3}^2 + 2v_{i,1}^2 v_{i,2}^2 + 4v_{i,1}v_{i,2}v_{i,3}\rangle$$

The following theorem is a generalization of the log-convexity result in [3] to n dimensional case when the random vector has independent components.

Theorem 2. *The log-convexity holds in this independent case. That is, when we can write $\mu_t^X(x_1, x_2, ..., x_n) = \prod_{j=1}^n \mu_t^{X_j}(x_j)$, we have the log-convexity.*

Proof. We prove the following.

$$\langle \sum_{i=1}^{2} v_{i,2}^2 \rangle^2 \leq \langle \sum_{i=1}^{2} v_{i,1}^2 \rangle \langle \sum_{i=1}^{2} v_{i,3}^2 + 2v_{i,1}^2 v_{i,2}^2 + 4v_{i,1}v_{i,2}v_{i,3} \rangle \tag{2}$$

In the following, we mimic Wang's approach [3].

We have the following equality constraints from Lemma 1:
$\langle \sum_{i=1}^{n} v_{i,2}^2 + v_{i,1}^2 v_{i,2} + v_{i,1}v_{i,3} \rangle = 0$ and $\langle \sum_{i=1}^{n} v_{i,1}^4 + 3v_{i,1}^2 v_{i,2} \rangle = 0$.

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

$$\langle \sum_{i=1}^{n} v_{i,2}^2 + v_{i,1}^2 v_{i,2} + v_{i,1}v_{i,3} \rangle - (\frac{1-\alpha}{3}) \langle \sum_{i=1}^{n} v_{i,1}^4 + 3v_{i,1}^2 v_{i,2} \rangle = 0$$

That is,

$$\langle \sum_{i=1}^{n} v_{i,2}^2 + \alpha v_{i,1}^2 v_{i,2} + v_{i,1}v_{i,3} - \frac{1-\alpha}{3} v_{i,1}^4 \rangle = 0$$

We can write it as

$$\langle \sum_{i=1}^{n} v_{i,2}^2 \rangle = \langle \sum_{i=1}^{n} v_{i,1}(-\alpha v_{i,1}v_{i,2} - v_{i,3} + \frac{1-\alpha}{3} v_{i,1}^3) \rangle$$

By Cauchy-Schwarz inequality, we get

$$\langle \sum_{i=1}^{n} v_{i,1}(-\alpha v_{i,1}v_{i,2} - v_{i,3} + \frac{1-\alpha}{3} v_{i,1}^3) \rangle^2 \leq \langle \sum_{i=1}^{n} v_{i,1}^2 \rangle \langle \sum_{i=1}^{n} (-\alpha v_{i,1}v_{i,2} - v_{i,3} + \frac{1-\alpha}{3} v_{i,1}^3)^2 \rangle$$

In order to show the log-convexity in this case, it remains to show that

$$\langle \sum_{i=1}^{n} (-\alpha v_{i,1}v_{i,2} - v_{i,3} + \frac{1-\alpha}{3} v_{i,1}^3)^2 \rangle \leq \langle \sum_{i=1}^{n} v_{i,3}^2 + 2v_{i,1}^2 v_{i,2}^2 + 4v_{i,1}v_{i,2}v_{i,3} \rangle$$

This corresponds to equation 4.8 in Wang's thesis [3].

We expand the square term.

$$
\begin{aligned}
&(-\alpha v_{i,1}v_{i,2} - v_{i,3} + \frac{1-\alpha}{3}v_{i,1}^3)^2 \\
=&\alpha^2 v_{i,1}^2 v_{i,2}^2 + v_{i,3}^2 + \frac{(1-\alpha)^2}{9}v_{i,1}^6 + 2\alpha v_{i,1}v_{i,2}v_{i,3} - 2\alpha v_{i,1}v_{i,2}\frac{1-\alpha}{3}v_{i,1}^3 - 2v_{i,3}\frac{1-\alpha}{3}v_{i,1}^3
\end{aligned} \tag{3}
$$

Then we rearrange the terms, and we want to prove

$$
\begin{aligned}
&\langle \sum_{i=1}^n v_{i,3}^2 + 2v_{i,1}^2v_{i,2}^2 + 4v_{i,1}v_{i,2}v_{i,3} - \alpha^2 v_{i,1}^2 v_{i,2}^2 - v_{i,3}^2 - \frac{(1-\alpha)^2}{9}v_{i,1}^6 \\
&- 2\alpha v_{i,1}v_{i,2}v_{i,3} + 2\alpha v_{i,1}v_{i,2}\frac{1-\alpha}{3}v_{i,1}^3 + 2v_{i,3}\frac{1-\alpha}{3}v_{i,1}^3 \rangle \geq 0
\end{aligned} \tag{4}
$$

We simplify this expression.

$$
\begin{aligned}
&\langle \sum_{i=1}^n (2-\alpha^2)v_{i,1}^2v_{i,2}^2 + (4-2\alpha)v_{i,1}v_{i,2}v_{i,3} - \frac{(1-\alpha)^2}{9}v_{i,1}^6 \\
&+ \frac{2}{3}\alpha(1-\alpha)v_{i,2}v_{i,1}^4 + \frac{2}{3}(1-\alpha)v_{i,3}v_{i,1}^3 \rangle \\
&\geq 0
\end{aligned} \tag{5}
$$

We use the following equality constraints.

$$
\langle \sum_{i=1}^n v_{i,1}^3 v_{i,3} + v_{i,2}v_{i,1}^4 + 3v_{i,1}^2v_{i,2}^2 \rangle = 0
$$

and

$$
\langle \sum_{i=1}^n v_{i,1}^6 + 5v_{i,1}^4 v_{i,2} \rangle = 0
$$

We need to show that there exists some $\alpha, \beta, \gamma \in \mathbb{R}$ such that

$$
\begin{aligned}
&\langle \sum_{i=1}^n (2-\alpha^2)v_{i,1}^2v_{i,2}^2 + (4-2\alpha)v_{i,1}v_{i,2}v_{i,3} - \frac{(1-\alpha)^2}{9}v_{i,1}^6 + \frac{2}{3}\alpha(1-\alpha)v_{i,2}v_{i,1}^4 + \frac{2}{3}(1-\alpha)v_{i,3}v_{i,1}^3 \rangle \\
&+ \beta\langle \sum_{i=1}^n v_{i,1}^3 v_{i,3} + v_{i,2}v_{i,1}^4 + 3v_{i,1}^2v_{i,2}^2 \rangle + \gamma\langle \sum_{i=1}^n v_{i,1}^6 + 5v_{i,1}^4 v_{i,2} \rangle \geq 0
\end{aligned} \tag{6}
$$

We pick $\alpha = 2$, $\beta = \frac{2}{3}$, $\gamma = \frac{2}{15}$ same as in Wang's theis. We need to prove

$$\begin{aligned}&\langle \sum_{i=1}^{n} -2v_{i,1}^2 v_{i,2}^2 - \frac{1}{9}v_{i,1}^6 - \frac{4}{3}v_{i,2}v_{i,1}^4 - \frac{2}{3}v_{i,3}v_{i,1}^3 \\ &+ \frac{2}{3}v_{i,1}^3 v_{i,3} + \frac{2}{3}v_{i,2}v_{i,1}^4 + 2v_{i,1}^2 v_{i,2}^2 + \frac{2}{15}v_{i,1}^6 + \frac{2}{3}v_{i,1}^4 v_{i,2}\rangle \geq 0\end{aligned} \tag{7}$$

By further simplification, we get

$$\langle \sum_{i=1}^{n} \frac{1}{45}v_{i,1}^6\rangle \geq 0 \tag{8}$$

It's clear that this inequality holds. Hence the log-convexity holds in this independent case. □

2.2 Special Case: Log-Concave Distribution

In [4], Theorem 1, the author proved that for any log-concave distribution, it holds that

$$\frac{d^2}{dt^2}\frac{1}{I(\mu_t^X)} \leq 0$$

This result implies that

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

which is stronger than log-convexity of Fisher information. Here we give another proof which would leads to the log-convexity directly. We first present the two-dimensional case, then generalize it to n dimensional case.

Here we mention that the convolution of two log-concave distributions is still log-concave. This result can be found in Proposition 3.5 of [5]. Then we know that, since the Gaussian distribution is log-concave, if X is log-concave, then μ_t^X is also log-concave.

2.2.1 Two dimensional case

Proposition 1. *For 2-dimensional case, the Log-Convexity Conjecture is true if X is log-concave.*

Proof. By Cauchy inequality, we have

$$\begin{aligned}\langle\sum_{j=1}^{l}\sum_{i=1}^{l}v_{ij}^2\rangle^2 &= \langle\sum_{i=1}^{l}v_i\sum_{j=1}^{l}(v_{ij}v_j+v_{ijj})\rangle^2 \\ &\le \langle\sum_{i=1}^{l}v_i^2\rangle\langle\sum_{i=1}^{l}(\sum_{j=1}^{l}v_{ij}v_j+v_{ijj})^2\rangle\end{aligned} \tag{9}$$

We can try to prove the following.

$$\begin{aligned}\langle v_{11}^2+v_{22}^2+2v_{12}^2\rangle^2 &\le \langle\sum_{i=1}^{l}v_i^2\rangle\langle\sum_{i=1}^{l}(\sum_{j=1}^{l}v_{ij}v_j+v_{ijj})^2\rangle \\ &\le \langle v_1^2+v_2^2\rangle\langle\sum_{j=1}^{l}\sum_{i=1}^{l}\sum_{w=1}^{l}v_{ijw}^2-2v_{ij}v_{jw}v_{iw}\rangle\end{aligned} \tag{10}$$

We want to show

$$\langle\sum_{i=1}^{l}(\sum_{j=1}^{l}v_{ij}v_j+v_{ijj})^2\rangle \le \langle\sum_{j=1}^{l}\sum_{i=1}^{l}\sum_{w=1}^{l}v_{ijw}^2-2v_{ij}v_{jw}v_{iw}\rangle$$

That is,

$$\begin{aligned}&\langle v_{111}^2-2v_{11}^3+v_{222}^2-2v_{22}^3+3(v_{112}^2+v_{122}^2-2v_{11}v_{12}^2-2v_{22}v_{12}^2) \\ &-(v_{11}v_1+v_{111}+v_{12}v_2+v_{122})^2-(v_{12}v_1+v_{112}+v_{22}v_2+v_{222})^2\rangle \ge 0\end{aligned} \tag{11}$$

Now we expand the square terms. $(v_{11}v_1+v_{111}+v_{12}v_2+v_{122})^2$ equals to

$$\begin{aligned}&v_{11}^2v_1^2+v_{111}^2+v_{12}^2v_2^2+v_{122}^2 \\ &+2(v_{111}v_{11}v_1+v_{12}v_2v_{122}+v_1v_{11}v_{12}v_2+v_1v_{11}v_{122}+v_{111}v_{12}v_2+v_{111}v_{122})\end{aligned} \tag{12}$$

Similarly,

$$\begin{aligned} &v_{22}^2 v_2^2 + v_{222}^2 + v_{21}^2 v_1^2 + v_{211}^2 \\ &+ 2(v_{222}v_{22}v_2 + v_{21}v_1v_{211} + v_2v_{22}v_{21}v_1 + v_2v_{22}v_{211} + v_{222}v_{21}v_1 + v_{222}v_{211}) \end{aligned} \tag{13}$$

By using $\langle v_{11}^3 + v_{11}^2 v_1^2 + 2v_{111}v_{11}v_1\rangle = 0$ and $\langle v_{22}^3 + v_{22}^2 v_2^2 + 2v_{222}v_{22}v_2\rangle = 0$, we have

$$\begin{aligned} &\langle v_{111}^2 - 2v_{11}^3 + v_{222}^2 - 2v_{22}^3 + 3(v_{112}^2 + v_{122}^2 - 2v_{11}v_{12}^2 - 2v_{22}v_{12}^2) \\ &- (v_{11}v_1 + v_{111} + v_{12}v_2 + v_{122})^2 - (v_{12}v_1 + v_{112} + v_{22}v_2 + v_{222})^2\rangle = \langle -v_{11}^3 - v_{22}^3 + \\ &2v_{112}^2 - 6v_{11}v_{12}^2 - v_1^2v_{12}^2 - 2v_1v_2v_{11}v_{12} - 2v_1v_{11}v_{122} - 2v_2v_{12}v_{111} - 2v_{111}v_{122} - 2v_1v_{12}v_{112} \\ &+ 2v_{221}^2 - 6v_{22}v_{21}^2 - v_2^2v_{21}^2 - 2v_2v_1v_{22}v_{21} - 2v_2v_{22}v_{211} - 2v_1v_{21}v_{222} - 2v_{222}v_{211} - 2v_2v_{21}v_{221}\rangle \end{aligned} \tag{14}$$

We first handle

$$2v_{112}^2 - 6v_{11}v_{12}^2 - v_1^2v_{12}^2 - 2v_1v_2v_{11}v_{12} - 2v_1v_{11}v_{122} - 2v_2v_{12}v_{111} - 2v_{111}v_{122} - 2v_1v_{12}v_{112}$$

By using equality constraints, we can get the following steps.

By $\langle v_{11}v_{12}^2 + v_{11}v_{12}v_1v_2 + v_1v_{12}v_{112} + v_1v_{11}v_{122}\rangle = 0$, we get

$$2v_{112}^2 - 4v_{11}v_{12}^2 - v_1^2v_{12}^2 - 2v_2v_{12}v_{111} - 2v_{111}v_{122}$$

By $\langle v_2v_{12}v_{111} + v_{122}v_{111} + v_{12}v_{1112}\rangle = 0$ and $\langle v_1v_{12}v_{112} + v_{112}v_{112} + v_{12}v_{1112}\rangle = 0$, we get $\langle v_2v_{12}v_{111} + v_{122}v_{111}\rangle = \langle v_1v_{12}v_{112} + v_{112}v_{112}\rangle$. This implies that $\langle 3v_{112}^2 - 2v_2v_{12}v_{111} - 2v_{111}v_{122}\rangle = \langle v_{112}^2 - 2v_1v_2v_{112}\rangle$. Based on this, we get

$$-2v_1v_{12}v_{112} - 4v_{11}v_{12}^2 - v_1^2v_{12}^2$$

Last, by $\langle v_{12}^2 v_{11} + v_1 v_{12} v_1 v_{12} + v_1 v_{12} v_{112} + v_1 v_{12} v_{112} \rangle = 0$ we get

$$-3v_{11}v_{12}^2$$

It follows that (by symmetry)

$$\begin{aligned}
&\langle 2v_{112}^2 - 6v_{11}v_{12}^2 - v_1^2 v_{12}^2 - 2v_1 v_2 v_{11} v_{12} - \\
&2v_1 v_{11} v_{122} - 2v_2 v_{12} v_{111} - 2v_{111} v_{122} - 2v_1 v_{12} v_{112} \\
&+ 2v_{221}^2 - 6v_{22} v_{21}^2 - v_2^2 v_{21}^2 - 2v_2 v_1 v_{22} v_{21} - \\
&2v_2 v_{22} v_{211} - 2v_1 v_{21} v_{222} - 2v_{222} v_{211} - 2v_2 v_{21} v_{221} \rangle \\
=&\langle -3v_{11} v_{12}^2 - 3v_{22} v_{12}^2 \rangle
\end{aligned} \tag{15}$$

Now we need to prove that

$$\langle -v_{11}^3 - v_{22}^3 - 3v_{11}v_{12}^2 - 3v_{22}v_{12}^2 \rangle \geq 0$$

holds if the distribution μ_t^X is log-concave. And this is clear because for log-concave distribution we have $v_{11} \leq 0$ and $v_{22} \leq 0$

Since log-concavity is closed under convolution and a Gaussian distribution is log-concave, we know that if X is log-concave, then μ_t^X is concave. The result follows.

□

Remark: It seems that this would imply, the use of Cauchy-Schwarz inequality in this specific way is not appropriate if we want to prove the general case of log-convexity.

2.2.2 Generalization

In this subsection, we generalize the above result to n-dim distribution.

Define A to be a $n \times n$ matrix with entries $A_{ij} = v_{ij}$.

Theorem 3. *If X is log-concave, then $A \preccurlyeq 0$. In this case, the Fisher Information for n-dim case is log-convex.*

Proof. To show the log-convexity, we need

$$\langle \sum_{j=1}^{l} \sum_{i=1}^{l} v_{ij}^2 \rangle^2 \leq \langle \sum_{i=1}^{l} v_i^2 \rangle \langle \sum_{j=1}^{l} \sum_{i=1}^{l} \sum_{w=1}^{l} v_{ijw}^2 - 2v_{ij}v_{jw}v_{iw} \rangle \tag{16}$$

Note that $\langle v_{ij}^2 + v_{ij}v_iv_j + v_iv_{ijj} \rangle = 0$. This implies that $\langle v_{ij}^2 \rangle = -\langle v_i(v_{ij}v_j + v_{ijj}) \rangle$.

We can write

$$\langle \sum_{j=1}^{l} \sum_{i=1}^{l} v_{ij}^2 \rangle = -\langle \sum_{i=1}^{l} v_i \sum_{j=1}^{l} (v_{ij}v_j + v_{ijj}) \rangle$$

By Cauchy inequality, we have

$$\begin{aligned} \langle \sum_{j=1}^{l} \sum_{i=1}^{l} v_{ij}^2 \rangle^2 &= \langle \sum_{i=1}^{l} v_i \sum_{j=1}^{l} (v_{ij}v_j + v_{ijj}) \rangle^2 \\ &\leq \langle \sum_{i=1}^{l} v_i^2 \rangle \langle \sum_{i=1}^{l} (\sum_{j=1}^{l} v_{ij}v_j + v_{ijj})^2 \rangle \end{aligned} \tag{17}$$

We want to show

$$\langle \sum_{i=1}^{l} (\sum_{j=1}^{l} v_{ij}v_j + v_{ijj})^2 \rangle \leq \langle \sum_{j=1}^{l} \sum_{i=1}^{l} \sum_{w=1}^{l} v_{ijw}^2 - 2v_{ij}v_{jw}v_{iw} \rangle$$

We expand the square terms.

$$\begin{aligned} &(\sum_{j=1}^{l} v_{ij}v_j + v_{ijj})^2 \\ = &\sum_{j=1}^{l} \sum_{w=1}^{l} v_{ij}v_jv_{iww} + v_{ijj}v_{iw}v_w + v_{ij}v_jv_{iw}v_w + v_{ijj}v_{iww} \end{aligned}$$

Hence we need to show that

$$\begin{aligned}&\langle \sum_{j=1}^{l}\sum_{i=1}^{l}\sum_{w=1}^{l} v_{ijw}^2 - 2v_{ij}v_{jw}v_{iw}\rangle \\ &-\langle \sum_{i=1}^{l}\sum_{j=1}^{l}\sum_{w=1}^{l} v_{ij}v_jv_{iww} + v_{ijj}v_{iw}v_w + v_{ij}v_jv_{iw}v_w + v_{ijj}v_{iww}\rangle \geq 0\end{aligned}$$

Rearrange the terms, we need to prove

$$\langle \sum_{j=1}^{l}\sum_{i=1}^{l}\sum_{w=1}^{l} v_{ijw}^2 - 2v_{ij}v_{jw}v_{iw} - v_{ij}v_jv_{iww} - v_{ijj}v_{iw}v_w - v_{ij}v_jv_{iw}v_w - v_{ijj}v_{iww}\rangle \geq 0 \quad (18)$$

By using equality constrains, the above inequality is equivalent to

$$\langle \sum_{j=1}^{l}\sum_{i=1}^{l}\sum_{w=1}^{l} -v_{ij}v_{jw}v_{iw}\rangle \geq 0 \quad (19)$$

We wish to have that $\sum_{j=1}^{l}\sum_{i=1}^{l}\sum_{w=1}^{l} v_{ij}v_{jw}v_{iw} \leq 0$

Note that

$$tr(A^3) = \sum_{j=1}^{l}\sum_{i=1}^{l}\sum_{w=1}^{l} v_{ij}v_{jw}v_{iw}$$

By assumption we know X is log-concave, so μ_t^X is log-concave. Then we know A is negative semidefinite, so A^3 is also negative semidefinite. This implies that $tr(A^3) \leq 0$. Hence we get $\sum_{j=1}^{l}\sum_{i=1}^{l}\sum_{w=1}^{l} v_{ij}v_{jw}v_{iw} \leq 0$ as desired. □

3 | Future Work and Conclusion

In this Chapter, we will talk about our plans for future work.

3.1 Future Work

Efforts have been made to prove the Log-Convexity Conjecture in the general case by following the setting in this report. We have done many calculations to see if a similar approach, which applies the Cauchy-Schwarz inequality, will work. Unfortunately, all attempts trying to cancel out terms failed.

In the future, our next objective could be to find other approaches to prove the Log-Convexity Conjecture, or to see if other classes of distributions could be proved to satisfy the Log-Convexity Conjecture, other than the class of log-concave distributions.

3.2 Conclusions

In this report, we worked on the Log-Convexity Conjecture. We proved a special case of the Log-Convexity Conjecture when the distribution has independent components. Then we gave another proof for the Log-Convexity Conjecture for log-concave distributions.

Bibliography

[1] A. Fink, "Kolmogorov-landau inequalities for monotone functions," *Journal of Mathematical Analysis and Applications*, vol. 90, no. 1, pp. 251–258, 1982.

[2] M. Ledoux, "Differentials of entropy and fisher information along heat flow: a brief review of some conjectures," 2022.

[3] Y. Wang, "Optimization of some non-convex functionals arising in information theory," Ph.D. dissertation, The Chinese University of Hong Kong (Hong Kong), 2021.

[4] G. Toscani, "A concavity property for the reciprocal of fisher information and its consequences on costa's epi," *Physica A: Statistical Mechanics and its Applications*, vol. 432, pp. 35–42, 2015.

[5] A. Saumard and J. A. Wellner, "Log-concavity and strong log-concavity: a review," *Statistics surveys*, vol. 8, p. 45, 2014.

A1 Appendix 1 - Calculations in Proof

A1.1 Proof for lemmas

Proposition 2. *Integration by parts:*

$$\int_{\Omega} \nabla f \cdot \nabla v dx = -\int_{\Omega} (\nabla \cdot \nabla f) v dx + \int_{\partial\Omega} v \cdot (\nabla f \cdot n) ds$$

where n is the outward normal vector on boundary $\partial\Omega$. Note that
$\nabla \cdot \nabla f = \sum_{i=1}^{k} \frac{\partial^2}{\partial x_i^2} f = \Delta f$.

A1.1.1 Proof of Lemma 2

Proof. By integration by parts, under certain limiting conditions, we have

$$\int_{\mathbb{R}^l} \phi \Delta \mu_t^X(x) dx = -\int_{\mathbb{R}^l} \nabla \phi \nabla \mu_t^X(x) dx + \int \phi \cdot (\nabla \mu_t^X(x) \cdot n) ds = -\int_{\mathbb{R}^l} \nabla \phi \nabla \mu_t^X(x) dx$$

Here we explain the last equality. Recall that we have

$$\mu_t^X v_1^{e_j} = \frac{\frac{\partial}{\partial x_j} \mu_t^X}{\mu_t^X} \mu_t^X = \frac{\partial}{\partial x_j} \mu_t^X$$

Therefore we can write

$$\nabla \mu_t^X = \begin{pmatrix} \mu_t^X v_1^{e_1} \\ \mu_t^X v_1^{e_2} \\ \ldots \\ \mu_t^X v_1^{e_l} \end{pmatrix}$$

For finite l, note that

$$\phi \cdot (\nabla \mu_t^X(x) \cdot n) \leq \sum_{i=1}^{l} \phi \mu_t^X v_1^{e_i}$$

As long as we have $\lim_{|x_i| \to \infty} \phi \mu_t^X v_1^{e_i} = 0$ for all i, we can get $\int \phi \cdot (\nabla \mu_t^X(x) \cdot n) ds = 0$.

By properties of heat equation, we have

$$\begin{aligned} \frac{\partial}{\partial t} \langle \phi \rangle &= \int_{\mathbb{R}^l} \frac{\partial}{\partial t} \phi \mu_t^X(x) dx \\ &= \langle \frac{\partial}{\partial t} \phi \rangle + \frac{1}{2} \int_{\mathbb{R}^l} \phi \Delta \mu_t^X(x) dx \\ &= \langle \frac{\partial}{\partial t} \phi \rangle - \frac{1}{2} \int_{\mathbb{R}^l} \nabla \phi \nabla \mu_t^X(x) dx \\ &= \langle \frac{\partial}{\partial t} \phi - \frac{1}{2} \sum_{i=1}^{l} \frac{\partial}{\partial x_i} \phi \cdot v_1^{e_i}(x) \rangle \end{aligned} \tag{1}$$

□

A1.1.2 Proof of Lemma 3

Proof. Note that $v_1^{e_i} = \frac{\frac{\partial}{\partial x_i} \mu_t^X}{\mu_t^X}$ and

$$v_2^{2e_i} = \frac{(\frac{\partial^2}{\partial x_i^2} \mu_t^X) \mu_t^X - (\frac{\partial}{\partial x_i} \mu_t^X)^2}{(\mu_t^X)^2}$$

$$v_2^{e_i + e_j} = \frac{(\frac{\partial^2}{\partial x_i \partial x_j} \mu_t^X) \mu_t^X - (\frac{\partial}{\partial x_i} \mu_t^X)(\frac{\partial}{\partial x_j} \mu_t^X)}{(\mu_t^X)^2}$$

$$v_1^{e_i} \frac{\partial}{\partial x_j} v_1^{e_i} = \frac{(\frac{\partial^2}{\partial x_i \partial x_j} \mu_t^X)(\frac{\partial}{\partial x_i} \mu_t^X)}{(\mu_t^X)^2} - \frac{(\frac{\partial}{\partial x_i} \mu_t^X)^2 (\frac{\partial}{\partial x_j} \mu_t^X)}{(\mu_t^X)^3}$$

So we have

$$\frac{\frac{\partial^2}{\partial x_i^2}\mu_t^X}{\mu_t^X} = v_2^{2e_i} + (v_1^{e_i})^2$$

$$\begin{aligned} 2\frac{\partial}{\partial t}v_k &= 2\frac{\partial^k}{\partial x_1^{k_1}...\partial x_l^{k_l}}[\frac{1}{\mu_t^X}\frac{\partial}{\partial t}\mu_t^X] \\ &= \frac{\partial^k}{\partial x_1^{k_1}...\partial x_l^{k_l}}[\frac{1}{\mu_t^X}\sum_{i=1}^{l}\frac{\partial^2}{\partial x_i^2}\mu_t^X] \\ &= \frac{\partial^k}{\partial x_1^{k_1}...\partial x_l^{k_l}}[\sum_{i=1}^{l} v_2^{2e_i} + (v_1^{e_i})^2] \end{aligned} \tag{2}$$

The second equality follows from chain rule and Heat equation property. □

A1.1.3 Proof of Lemma 4

Proof. $v_1^{e_i} = \frac{\frac{\partial}{\partial x_i}\mu_t^X}{\mu_t^X}$

$$\begin{aligned} I(\mu_t^X) &= \int_{\mathbb{R}^k} \frac{(\sum_{i=1}^{l}\frac{\partial}{\partial x_i}\mu_t^X(x_1,...,x_l))^2}{\mu_t^X(x_1,...,x_l)} dx_1...dx_l \\ &= \langle \frac{(\sum_{i=1}^{l}\frac{\partial}{\partial x_i}\mu_t^X)^2}{(\mu_t^X)^2} \rangle \\ &= \langle [\frac{\sum_{i=1}^{l}\frac{\partial}{\partial x_i}\mu_t^X}{\mu_t^X}]^2 \rangle \\ &= \langle (\sum_{i=1}^{l} v_1^{e_i})^2 \rangle \end{aligned} \tag{3}$$

□

A1.1.4 Proof of Lemma 5

Proof.

$$\begin{aligned}
\frac{d}{dt}I(\mu_t^X) &= \frac{d}{dt}\sum_{i=1}^{l}\langle (v_1^{0,\dots,1,\dots,0})^2\rangle \\
&= \langle \sum_{j=1}^{l}\frac{\partial}{\partial t}(v_1^{e_j})^2 - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\frac{\partial}{\partial x_i}(v_1^{e_j})^2 \cdot v_1^{e_i}\rangle \\
&= \langle \sum_{j=1}^{l} 2v_1^{e_j}\frac{\partial}{\partial t}v_1^{e_j} - \sum_{i=1}^{l}\sum_{j=1}^{l} v_1^{e_j}v_1^{e_i}v_2^{e_j+e_i}\rangle \qquad (4) \\
&= \langle \sum_{j=1}^{l} v_1^{e_j}\sum_{i=1}^{l}(v_3^{2e_i+e_j} + 2v_1^{e_i}v_2^{e_i+e_j}) - \sum_{i=1}^{l}\sum_{j=1}^{l} v_1^{e_j}v_1^{e_i}v_2^{e_j+e_i}\rangle \\
&= \langle \sum_{j=1}^{l}\sum_{i=1}^{l} v_1^{e_j}v_3^{2e_i+e_j} + v_1^{e_j}v_1^{e_i}v_2^{e_i+e_j}\rangle
\end{aligned}$$

consider $\phi = v_1^{e_j}, k = 3$ on lemma 4.1:

$$\langle v_1^{e_j}v_3^{2e_i+e_j} + v_2^{e_i+e_j}v_1^{e_i}v_1^{e_j} + v_2^{e_i+e_j}\frac{\partial}{\partial x_i}v_1^{e_j}\rangle = 0$$

Hence we have

$$\langle v_1^{e_j}v_3^{2e_i+e_j} + v_2^{e_i+e_j}v_1^{e_i}v_1^{e_j}\rangle = -\langle (v_2^{e_i+e_j})^2\rangle$$

It follows that

$$\frac{d}{dt}I(\mu_t^X) = -\langle \sum_{j=1}^{l}\sum_{i=1}^{l}(v_2^{e_i+e_j})^2\rangle$$

□

A1.1.5 Proof of Lemma 6

Proof. Recall that

$$2\frac{\partial}{\partial t}v_k = \frac{\partial^k}{\partial x_1^{k_1}\dots\partial x_l^{k_l}}[\sum_{i=1}^{l} v_2^{2e_i} + (v_1^{e_i})^2]$$

When $k=2$,

$$2\frac{\partial}{\partial t}v_2^{e_i+e_j}=\frac{\partial^2}{\partial x_i\partial x_j}[\sum_{w=1}^{l}v_2^{2e_w}+(v_1^{e_w})^2]=\sum_{w=1}^{l}[v_4^{2e_w+e_i+e_j}+2v_2^{e_w+e_j}v_2^{e_i+e_w}+2v_3^{e_i+e_j+e_w}v_1^{e_w}]$$

$$\begin{aligned}\frac{d^2}{dt^2}I(\mu_t^X)&=-\frac{d}{dt}\langle\sum_{j=1}^{l}\sum_{i=1}^{l}(v_2^{e_i+e_j})^2\rangle\\&=-\langle\sum_{j=1}^{l}\sum_{i=1}^{l}2v_2^{e_i+e_j}\frac{d}{dt}v_2^{e_i+e_j}-\frac{1}{2}\sum_{i=w}^{l}[\sum_{j=1}^{l}\sum_{i=1}^{l}2v_2^{e_i+e_j}v_3^{e_i+e_j+e_w}]v_1^{e_w}\rangle\\&=-\langle\sum_{j=1}^{l}\sum_{i=1}^{l}\sum_{w=1}^{l}v_2^{e_i+e_j}[v_4^{2e_w+e_i+e_j}+2v_2^{e_w+e_j}v_2^{e_i+e_w}+2v_3^{e_i+e_j+e_w}v_1^{e_w}]\\&\quad-v_2^{e_i+e_j}v_3^{e_i+e_j+e_w}v_1^{e_w}\rangle\end{aligned}\tag{5}$$

Recall that

$$\langle\phi v_k^{k_1,\ldots,k_l}+\phi v_1^{e_j}v_{k-1}^{k_1,\ldots,k_j-1,\ldots,k_l}+v_{k-1}^{k_1,\ldots,k_j-1,\ldots,k_l}\frac{\partial}{\partial x_j}\phi\rangle=0$$

Set $k=4$ and $\phi=v_2^{e_i+e_j}$.

$$\langle v_2^{e_i+e_j}v_4^{2e_w+e_i+e_j}+v_2^{e_i+e_j}v_1^{e_w}v_3^{e_w+e_i+e_j}+(v_3^{e_w+e_i+e_j})^2\rangle=0$$

So we have

$$\frac{d^2}{dt^2}I(\mu_t^X)=\langle\sum_{j=1}^{l}\sum_{i=1}^{l}\sum_{w=1}^{l}(v_3^{e_w+e_i+e_j})^2-2v_2^{e_i+e_j}v_2^{e_w+e_j}v_2^{e_i+e_w}\rangle$$

Now we consider to set $k=2$ and $\phi=v_2^{e_w+e_i}v_2^{e_w+e_j}$

$$\langle v_2^{e_w+e_i}v_2^{e_w+e_j}v_2^{e_i+e_j}+v_2^{e_w+e_i}v_2^{e_w+e_j}v_1^{e_j}v_1^{e_i}+v_1^{e_i}v_2^{e_w+e_i}v_3^{e_w+2e_j}+v_1^{e_i}v_3^{e_w+e_i+e_j}v_2^{e_w+e_j}\rangle=0$$

$$
\begin{aligned}
&\langle -2v_2^{e_i+e_j}v_2^{e_w+e_j}v_2^{e_i+e_w}\rangle \\
=&\langle 2v_2^{e_w+e_i}v_2^{e_w+e_j}v_1^{e_j}v_1^{e_i}+2v_1^{e_i}v_2^{e_w+e_i}v_3^{e_w+2e_j}+2v_1^{e_i}v_3^{e_w+e_i+e_j}v_2^{e_w+e_j}\rangle
\end{aligned} \tag{6}
$$

This proves the last equality. □