On weighted Fisher information matrix properties

M. Kelbert, Y. Suhov, S. Yasaei Sekeh

Abstract

In this paper, we review Fisher information matrices properties in weighted version and discuss inequalities/bounds on it by using reduced weight functions. In particular, an extended form of the Fisher information inequality previously established in [6] is given. Further, along with generalized De-Bruijn’s identity, we provide new interpretation of the concavity for the entropy power.

1 Introduction

Discussion on the quality of a given parametric probabilistic system with respect to the inference of its unknown parameters has been the intention of various studies in information theory and plays an important role in the analysis and design of the signal processes systems. Having explore the Fisher information measure, [7, 8], associated with estimation method, provides a useful theoretical tool to describe the propagation of information through systems.

The aim of this paper is to establish an extended (weighted) version of Fisher information matrix (FIM) and investigate a number of its properties. Indeed, the definitions presented in [10] are also obtained by an unified method which is based on non-negative weight function (WF) $\varphi$. Hence the novel version is called weighted Fisher information matrix (WFIM). Moreover, several straightforward facts, of the WFIM has been derived in [10]. For instance, invoking WFIM, the multivariate weighted Cramér-Rao and Kullback inequalities were established. In fact this evidences the need of analyzing the WFIM with more details. As reflection on weighted features in information aspects and relevant determinant bounds/inequalities we address the reader to [17, 18, 19] in which in contrast to previous works the quality/importance, known as utility/weight, of the occurrence of an event has not been ignored.

In addition of mirroring similar bounds/inequalities of FIM established in [20, 15, 21], along with a list of theoretical results, in current paper we also concentrate on the connection between WFI and the weighted form of certain divergence measures, such as KL-divergence and mutual information, at which the obtained properties in this work could be helpful in a branch of research in coding theorem and machine learning fields.

The standard De-Bruijn’s identity is the fundamental relation between differential entropy and Fisher information (FI), and as such, is used to prove the entropy power inequality from the corresponding FI. This identity in terms of the standard Gaussian additive noise, has been stated together with proof in scalar case, [13, 9], and also the vector case by [3]. As it can be seen in the literature, the De-Bruijn’s identity relies on a diffusion equation.

2010 Mathematics Subject Classification: 60A10, 60B05, 60C05

Key words and phrases: weight function, weighted Fisher information matrix, weighted information correlation, convexity, concavity, De-Bruijn’s identity, entropy power.
satisfied by the Gaussian distribution and is obtained by integration by parts in the scalar version and Green’s identity for the vector case. Note that in [11, 10], De-Bruijn’s identity has been represented due to the nonstandard Gaussian noise. We devote the ultimate part of this article, section 5, to focus on systems with noises and propose a generalized (weighted) form of De-Bruijn’s identity in vector case which is fulfilled under assumption on the WF $\varphi$.

Let us begin with basic definitions:

**Definition 1.1** Let $X = (X_1, \ldots, X_n)$ be a random $1 \times n$ vector with probability density function (PDF) $f_\theta(x) = f_X(x; \theta)$, $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, where $\theta = (\theta_1, \ldots, \theta_m) \in \mathbb{R}^m$ is a parameter vector. Suppose that dependence $\theta \mapsto f_\theta$ is $C^1$. The $m \times m$ weighted Fisher information matrix (WFIM) $J_\varphi^w(X; \theta)$, with a given $WF x \in \mathbb{R}^n \mapsto \varphi(x) \geq 0$, is defined by

$$J_\varphi^w(f; \theta) := J_\varphi^w(X; \theta) = \mathbb{E}\left[\varphi(X)S(X, \theta)^T S(X, \theta)\right] = \int \frac{\varphi(x)}{f_\theta(x)} \left(\frac{\partial f_\theta(x)}{\partial \theta}\right)^T \frac{\partial f_\theta(x)}{\partial \theta} 1 (f_\theta(x) > 0) \ dx,$$

assuming the integrals are absolutely convergent. Here and below, $\frac{\partial}{\partial \theta}$ stands for the $1 \times m$ gradient in $\theta$ and $S(X, \theta) = 1(f_\theta(x) > 0) \frac{\partial}{\partial \theta}\log f_\theta(x)$ denotes the score vector. It can easily be seen that $J_\varphi^w(f; \theta)$ is a positive-definite matrix, for non-null vector $u = (u_1, \ldots, u_n) \in \mathbb{R}^n$:

$$u^T J_\varphi^w(f; \theta) u = \mathbb{E}\left[\varphi S^T S\right] u^T = \mathbb{E}\left[\varphi u S^T S u^T\right] = \mathbb{E}\left[\varphi (S^T u)^2\right] \geq 0.$$

In what follows, a non-negativity for a matrix $A \succeq 0$ means a positive-definite matrix and inequality $A \preceq B$ shows that the matrix $B - A$ is a positive-definite matrix.

Next, let $(X, Y)$ be a pair of RVs with a joint PDF $f_\theta(x, y) = f_{X,Y}(x, y; \theta)$ and conditional PDF $f_\theta(y|x) = f_{Y|X}(x, y; \theta) := \frac{f_\theta(x, y)}{f_\theta(x)}$. Given a joint $WF (x, y) \in \mathbb{R}^n \times \mathbb{R}^n \mapsto \varphi(x, y) \geq 0$, we set:

$$J_\varphi^w(X, Y; \theta) = \mathbb{E}\left[\varphi(X, Y) \left(\frac{\partial \log f_\theta(X, Y)}{\partial \theta}\right)^T \frac{\partial \log f_\theta(X, Y)}{\partial \theta} 1 (f_\theta(X, Y) > 0)\right] = \int \frac{\varphi(x, y)}{f_\theta(x, y)} \left(\frac{\partial f_\theta(x, y)}{\partial \theta}\right)^T \frac{\partial f_\theta(x, y)}{\partial \theta} 1 (f_\theta(x, y) > 0) \ dx dy$$

and

$$J_\varphi^w(Y|X; \theta) = \mathbb{E}\left[\varphi(Y|X) \left(\frac{\partial \log f_\theta(Y|X)}{\partial \theta}\right)^T \frac{\partial \log f_\theta(Y|X)}{\partial \theta} 1 (f_\theta(Y|X) > 0)\right] = \int f_\theta(x, y) \varphi(y|x) \left(\frac{\partial \log f_\theta(y|x)}{\partial \theta}\right)^T \frac{\partial \log f_\theta(y|x)}{\partial \theta} 1 (f_\theta(y|x) > 0) \ dx dy.$$

Following the notations in [16], consider an $m \times m$ matrix $S_\varphi = S_\varphi(f_X, y)$ and a $1 \times m$ vector $B_\varphi = B_\varphi(x, f_Y|X)$:

$$B_\varphi = \mathbb{E}_{Y|X=x} \left[\varphi(x, Y) \frac{\partial \log f_\theta(Y|X)}{\partial \theta}\right] \quad \text{and} \quad S_\varphi = \mathbb{E} \left\{ \left(\frac{\partial \log f_\theta(X)}{\partial \theta}\right)^T B_\varphi(X) + B_\varphi(X)^T \left(\frac{\partial \log f_\theta(X)}{\partial \theta}\right) \ 1 (f_\theta(X) > 0) \right\}.$$
Definition 1.2 Given two functions, \( x \in \mathbb{R}^n \mapsto f_{\theta_1}(x) \geq 0 \) and \( x \in \mathbb{R}^n \mapsto f_{\theta_2}(x) \geq 0 \), the term weighted KL-divergence (of \( f_{\theta_1} \) from \( f_{\theta_2} \)) with WF \( \varphi \) is defined by

\[
D_\varphi^w(f_{\theta_1} \| f_{\theta_2}) = \mathbb{E}_{f_{\theta_2}} \left[ \varphi(X) \log \frac{f_{\theta_1}(X)}{f_{\theta_2}(X)} \right] = \int \varphi(x) f_{\theta_1}(x) \log \frac{f_{\theta_1}(x)}{f_{\theta_2}(x)} \, dx. \tag{1.5}
\]

Further, the mutual WE between \( X_1 \sim f_{\theta_1} \) and \( X_2 \sim f_{\theta_2} \) with joint PDF \( f_{\theta_2} \), is given by

\[
i^w(X_1; X_2) = D_\varphi^w(f_{\theta_2} \| f_{\theta_1} \otimes f_{\theta_2}) = \mathbb{E} \left\{ \varphi(X_1, X_2) \log \frac{f_{\theta_1}(X_1) f_{\theta_2}(X_2)}{f_{\theta_2}(X_1) f_{\theta_1}(X_2)} \right\}. \tag{1.6}
\]

We motivate our subsequent development with the following example.

Example 1.1 (Weighted KL-divergence in exponential families:) Consider an exponential family density \( f_\theta \) with the canonical parameter \( \theta \), sufficient statistics vector \( T(x) \) and cumulant generating function \( A(\theta) \):

\[
f_\theta(x) = h(x) \exp \left\{ \langle \theta, T(x) \rangle - A(\theta) \right\}.
\]

Set

\[
A_\varphi(\theta) = \log \int \varphi(x) h(x) \exp \left\{ \langle \theta, T(x) \rangle \right\} \, dx.
\]

A straightforward calculation implies that for any \( \theta_1 \) and \( \theta_2 \), the weighted KL-divergence is obtained by

\[
D_\varphi^w(f_{\theta_1} \| f_{\theta_2}) = \left( A(\theta_2) - A(\theta_1) \right) \mathbb{E}_{f_{\theta_1}} \left[ \varphi(X) \right] - \langle e^{A_\varphi(\theta_1) - A(\theta_1)} \nabla A_\varphi(\theta_1), \theta_2 - \theta_1 \rangle
\]

\[
= e^{A_\varphi(\theta_1) - A(\theta_1)} \left( A(\theta_2) - A(\theta_1) - \left\langle \nabla A_\varphi(\theta_1), \theta_2 - \theta_1 \right\rangle \right). \tag{1.7}
\]

Here and below, we use symbol \( \nabla \) for nabla operator as a vector whose components are the partial derivative operators or spatial gradient \( 1 \times m \) vectors, \( \frac{\partial}{\partial \theta_i} \).

Further, implement Taylor expansion for \( A(\theta_2) \) and its first order expansion around \( \theta_1 \), hence \( (1.7) \) becomes

\[
D_\varphi^w(f_{\theta_1} \| f_{\theta_2}) = e^{A_\varphi(\theta_1) - A(\theta_1)} \left( \left\langle \nabla A(\theta_1), \theta_2 - \theta_1 \right\rangle + \frac{1}{2} \left\langle (\theta_2 - \theta_1), \Delta A(\theta_1)(\theta_2 - \theta_1) \right\rangle + O(\|\theta_2 - \theta_1\|^3) \right), \tag{1.8}
\]

where \( \Delta \) stands for Laplacian:

\[
\Delta g(\theta) = \sum_{i=1}^n \frac{\partial^2}{\partial \theta_i^2} g(\theta).
\]

In addition, by recalling \( (1.5) \) for exponential families, after some computations one yields:

\[
J_\varphi^w(f; \theta) = e^{A_\varphi(\theta) - A(\theta)} \left( \left\langle \nabla A_\varphi(\theta), \nabla A_\varphi(\theta) - \nabla A(\theta) \right\rangle + \Delta A_\varphi(\theta) \right). \tag{1.9}
\]
Plug out $\Delta A(\theta)$ from (1.4) and insert it into (1.8). So, we can write

$$D^w_{\varphi}(f_{\theta_1} \parallel f_{\theta_2}) = \frac{1}{2} \left\langle (\theta_2 - \theta_1), J^w_{\varphi}(f_1; \theta_1) (\theta_2 - \theta_1) \right\rangle + e^{A_{\varphi}(\theta_1) - A(\theta_1)} \left( \left\langle \nabla A(\theta_1), \nabla A_{\varphi}(\theta_1); \theta_2 - \theta_1 \right\rangle - \frac{1}{2} \left\langle (\theta_2 - \theta_1), (\nabla A(\theta_1) - \nabla A_{\varphi}(\theta_1)) (\theta_2 - \theta_1) \right\rangle \right)$$

(1.10)

The relation (1.10) suggests that for given WF $\varphi$, we may approximate the weighted KL-divergence via the quadratic form of WFIM, $\|A(\theta) - A_{\varphi}(\theta)\|^2$ and also the first order expansion of $A(\theta_2) - A_{\varphi}(\theta_2)$ around $\theta_1$. We also remark in passing that particularly when $\varphi \equiv 1$, the KL-divergence is roughly quadratic for exponential family models, where the quadratic form is given by the FIM.

**Theorem 1.1** (Connection between WFIM and weighted KL-divergence measures.) For smooth families $\{f_{\theta}\}_{\theta \in \Theta}$ and given WF $\varphi$, we get

$$D^w_{\varphi}(f_{\theta_1} \parallel f_{\theta_2}) = \frac{1}{2} \left\langle (\theta_2 - \theta_1), J^w_{\varphi}(X; \theta_1) (\theta_2 - \theta_1) \right\rangle + \left\langle E_{\theta_1} [\varphi(X) \nabla \log f_{\theta_1}(X)], \theta_2 - \theta_1 \right\rangle - \frac{1}{2} \left\langle (\theta_2 - \theta_1), E_{\theta_1} [\varphi(X) \Delta f_{\theta_1}(X)] (\theta_2 - \theta_1) \right\rangle + o\left(\|\theta_1 - \theta_2\|^2\right) E_{\theta_1} [\varphi(X)].$$

(1.11)

**Proof:** By virtue of a Taylor expansion of $\log f_{\theta_1}$ around $\theta_1$, we obtain

$$\log f_{\theta_2} = \log f_{\theta_1} + \left\langle \nabla \log f_{\theta_1}, \theta_1 - \theta_2 \right\rangle + \frac{1}{2} \left\langle (\theta_1 - \theta_2), \Delta \log f_{\theta_1}(\theta_1 - \theta_2) \right\rangle + O_x(\|\theta_1 - \theta_2\|^3).$$

(1.12)

Here $O_x(\|\theta_1 - \theta_2\|^3)$ denotes the reminder term which is a hidden dependence on $x$. Multiple both sides of (1.12) by $\varphi$, then by taking expectations and assuming that we can interchange differentiation and expectation appropriately, we write

$$E_{\theta_1} [\varphi \log f_{\theta_1}] = E_{\theta_1} [\varphi \log f_{\theta_1}] + \left\langle E_{\theta_1} [\varphi \nabla \log f_{\theta_1}], \theta_1 - \theta_2 \right\rangle + \frac{1}{2} \left\langle (\theta_1 - \theta_2), E_{\theta_1} [\varphi \Delta \log f_{\theta_1}](\theta_1 - \theta_2) \right\rangle + E_{\theta_1} [\varphi O_x(\|\theta_1 - \theta_2\|^3)].$$

(1.13)

Also, it can be seen that

$$\Delta \log f_{\theta_1} = \frac{\Delta f_{\theta_1}}{f_{\theta_1}} - \frac{\|\nabla f_{\theta_1}\|^2}{f_{\theta_1}^2},$$

Consequently,

$$E_{\theta_1} [\varphi \Delta \log f_{\theta_1}] = E_{\theta_1} [\varphi \frac{\Delta f_{\theta_1}}{f_{\theta_1}}] - J^w_{\varphi}(f_{\theta_1}).$$

(1.14)

Substitute (1.14) into (1.13) and assume that the reminder $R := O(\|\theta_1 - \theta_2\|^3)$ is uniform enough in $X$ that $E[R] = o(\|\theta_1 - \theta_2\|^2)$. Therefore the claimed result, i.e. (1.11), is achieved.

Subsequently in the present work, we analyze more properties of WFIM in order to understand this aspect more precisely. This all would help us to employ WFIM with better perspective in the course of estimation/approximation of the weighted KL-divergence which is useful in applications and has been studied in [10].
2 Convexity and chain rule

In this section, we verify the convexity property for the WFIM which we may require in future works. Also, we focus on the chain rule which has been established earlier in [21].

**Theorem 2.1** (Convexity of the WFIM; cf. [21], Theorem 3, [4] Theorem p. 591.) The WFIM is convex in \( f_1(x; \theta), f_2(x; \theta) \), non-negative function \( \varphi(x) \geq 0, \lambda_1, \lambda_2 \in [0, 1] \) with \( \lambda_1 + \lambda_2 = 1 \),

\[
J^w_{\varphi}(\lambda_1 f_1 + \lambda_2 f_2; \theta) \leq \lambda_1 J^w_{\varphi}(f_1; \theta) + \lambda_2 J^w_{\varphi}(f_2; \theta).
\]

(2.1)

In (2.1) equality occurs when \( \lambda_1, \lambda_2 \) vanishes or if \( f_1 \) and \( f_2 \) coincides modulo \( \varphi \).

**Proof:** Recall Definition 1.1, assume that there exists at least one \( x \), say \( x_0 \), such that

\[
\left( \frac{\partial}{\partial \theta} \left( \lambda_1 f_1(x_0; \theta) + \lambda_2 f_2(x_0; \theta) \right) \right) ^T \left( \frac{\partial}{\partial \theta} \left( \lambda_1 f_1(x_0; \theta) + \lambda_2 f_2(x_0; \theta) \right) \right) / (\lambda_1 f_1(x_0; \theta) + \lambda_2 f_2(x_0; \theta)) > \lambda_1 \left( \frac{\partial f_2(x_0; \theta)}{\partial \theta} \right) ^T \left( \frac{\partial f_2(x_0; \theta)}{\partial \theta} \right) / f_2(x_0; \theta),
\]

where after simplification equals the following:

\[
f_1 f_2 \left( \frac{\partial f_1(x_0; \theta)}{\partial \theta} \right) ^T \left( \frac{\partial f_2(x_0; \theta)}{\partial \theta} \right) + f_1 f_2 \left( \frac{\partial f_2(x_0; \theta)}{\partial \theta} \right) ^T \left( \frac{\partial f_2(x_0; \theta)}{\partial \theta} \right) > f_2^2 \left( \frac{\partial f_1(x_0; \theta)}{\partial \theta} \right) ^T \left( \frac{\partial f_1(x_0; \theta)}{\partial \theta} \right) + f_1^2 \left( \frac{\partial f_1(x_0; \theta)}{\partial \theta} \right) ^T \left( \frac{\partial f_1(x_0; \theta)}{\partial \theta} \right).
\]

This inequality may be written in the form

\[
0 > \left( f_2 \frac{\partial f_1(x_0; \theta)}{\partial \theta} - f_1 \frac{\partial f_2(x_0; \theta)}{\partial \theta} \right) ^T \left( f_2 \frac{\partial f_1(x_0; \theta)}{\partial \theta} - f_1 \frac{\partial f_2(x_0; \theta)}{\partial \theta} \right).
\]

This is contradiction, hence for all \( x \in \mathbb{R}^n \) we have

\[
\left( \frac{\partial}{\partial \theta} \left( \lambda_1 f_1(x; \theta) + \lambda_2 f_2(x; \theta) \right) \right) ^T \left( \frac{\partial}{\partial \theta} \left( \lambda_1 f_1(x; \theta) + \lambda_2 f_2(x; \theta) \right) \right) / (\lambda_1 f_1(x; \theta) + \lambda_2 f_2(x; \theta)) \leq \lambda_1 \left( \frac{\partial f_2(x; \theta)}{\partial \theta} \right) ^T \left( \frac{\partial f_2(x; \theta)}{\partial \theta} \right) / f_2(x; \theta),
\]

Multiple the both sides of above inequality in \( \varphi \) and integrate over \( \mathbb{R}^n \) concludes the proof. 

**Theorem 2.2** (Chain rule for the WFIM; cf. [21], Theorem 6.) Given a set of random variables \( X := X^n_t = (X_1, \ldots, X_n) \) and WF \( x := x^n_t \mapsto \varphi(x^n_t) \geq 0, set \)

\[
\psi_k := \psi_k(x^n_t) = \int \varphi(x) f_2^n(x_{k+1} | x^n_t) dx_{k+1},
\]

\[
B^k = \mathbb{E}_{X^n_{k+1} | X^n_t = x^n_t} \left[ \varphi(x^n_t, X_{k+1}^n) \frac{\partial \log f_2^n(X_{k+1}^n | x^n_t)}{\partial \theta} \right],
\]

\[
S^k = \mathbb{E} \left\{ \left( \frac{\partial \log f_2^n(X^n_t)}{\partial \theta} \right) ^T B^k (X^n_t) + B^k (X^n_t) ^T \left( \frac{\partial \log f_2^n(X^n_t)}{\partial \theta} \right) \right\} \mathbb{1} \left( f_2^n(X^n_t) > 0 \right).
\]

Then

\[
J^w_{\varphi}(X^n_t; \theta) = \sum_{k=1}^n J^w_{\varphi_k} (X_k^n | X_{k-1}^{n-1}; \theta) + S^k.
\]

(2.2)
In particular, if \( n \) random variables are iid, define
\[
\psi_k^w(x_1^{k+1}) = \int \phi_k(x) \, d\mathbf{x}_2^k, \\
\psi_1^w(x_1) = \int \phi_1(x) (f(x_2))^{n-1} \, d\mathbf{x}_2^n = \int \phi_2(x_1, x_2) (f(x_2))^{n-1} \, dx_2.
\]

Moreover, set
\[
B_k^w = (n-k) \int \phi_k(x_1^{k+1}) (f(x_{k+1}))^{n-k} \log f(x_{k+1}) \, dx_{k+1}, \\
S_k^w = k \int (f(x_1))^k \left[ \left( \frac{\partial \log f(x_1)}{\partial \theta} \right)^T B_k^w(x_1^k) \right] \, dx_1^n.
\]

We derive
\[
J^w_n(X; \theta) = n J^w_{\psi_1}(X_1; \theta) + \sum_{k=1}^{n} S_k^w.
\]

**Proof:** By virtue of the chain rule for WFI, Lemma 4.3, \([16]\), we can write
\[
J^w_n(X_1, \ldots, X_n; \theta) = J^w_{\psi_1}(X_1; \theta) + J^w_{\psi_2}(X_2, \ldots, X_n | X_1; \theta) + S_2^w = J^w_{\psi_1}(X_1; \theta) + J^w_{\psi_2}(X_2 | X_1; \theta) + S_2^w + S_2^w
\]
\[
+ J^w_{\psi_3}(X_3, \ldots, X_n | X_1, X_2; \theta)
\]
\[
\vdots
\]
\[
= \sum_{k=1}^{n} J^w_{\psi_k}(X_k | X_1^{k-1}; \theta) + S_k^w.
\]
This yields the equality \((2.2)\). \(\Box\)

### 3 Weighted Fisher information inequality

As we said in the introduction, one of our main purposes of this work is to construct the weighted version of the well-known Fisher information inequality, denoted by WFII; cf \([6]\).

Let \( \mathbf{Z} = (\mathbf{X}, \mathbf{Y}) \) be a pair of independent RVs \( \mathbf{X} \) and \( \mathbf{Y} \), in \( \mathbb{R}^n \), with sample values \( \mathbf{z} = (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^n \) and marginal PDFs \( f_1, f_2 \), respectively. Let \( f_{\mathbf{Z}|\mathbf{X}+\mathbf{Y}} \) stands the conditional PDF as
\[
f_{\mathbf{Z}|\mathbf{X}+\mathbf{Y}}(\mathbf{x}, \mathbf{y}|\mathbf{z}) = \frac{f_1(\mathbf{x}) f_2(\mathbf{y}) 1(\mathbf{x} + \mathbf{y} = \mathbf{z})}{\int_{\mathbb{R}^n} f_1(\mathbf{v}) f_2(\mathbf{z} - \mathbf{v}) \, dv} \, dv.
\]
Given a WF \( \mathbf{z} = (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^n \rightarrow \phi(\mathbf{z}) \geq 0 \), we employ the following list of reduced WFs:
\[
\theta(\mathbf{z}) = \int \phi(\mathbf{v}, \mathbf{z} - \mathbf{v}) f_{\mathbf{Z}|\mathbf{X}+\mathbf{Y}}(\mathbf{v}, \mathbf{z} - \mathbf{v}|\mathbf{z}) \, dv, \\
\theta_1(\mathbf{x}) = \int \phi(\mathbf{x} + \mathbf{y}, \mathbf{y}) f_2(\mathbf{y}) \, dy, \quad \theta_2(\mathbf{y}) = \int \phi(\mathbf{x} + \mathbf{y}, \mathbf{f}_1(\mathbf{x}) \, dx.
\]

In addition, we define matrices \( \mathbf{M}_\phi \) and \( \mathbf{G}_\phi \):
\[
\mathbf{M}_\phi = \int \phi(\mathbf{x}, \mathbf{y}) f_1(\mathbf{x}) f_2(\mathbf{y}) \left( \frac{\partial \log f_1(\mathbf{x})}{\partial \theta} \right)^T \left( \frac{\partial \log f_2(\mathbf{y})}{\partial \theta} \right) 1(f_1(\mathbf{x}) f_2(\mathbf{y}) > 0) \, dx \, dy, \quad (3.2)
\]
and

\[ G_\varphi = J_{\theta_1}^\varphi(X)^{-1} \mathcal{M}_\varphi \left( J_{\theta_2}^\varphi(Y) \right)^{-1}. \]  

(3.3)

Observe that

\[ J_{\theta_1}^\varphi(X) = E_X \left\{ \theta_1(X) \left( \frac{\partial \log f_1(X)}{\partial \varphi} \right) \left( \frac{\partial \log f_1(X)}{\partial \theta} \right)^T 1(f_1(X) > 0) \right\}. \]

(3.4)

Swapping \( X \) with \( Y \), the \( J_{\theta_2}^\varphi(Y) \) is given. Now by using the Schwarz inequality, we derive

\[ \mathcal{M}_\varphi^2 \leq J_{\theta_1}^\varphi(X) J_{\theta_2}^\varphi(Y), \quad \text{or} \quad \mathcal{M}_\varphi G_\varphi \leq I, \]

(3.5)

with equality iff \( f_X^{(1)}(x) = \frac{\partial \log f_1(x)}{\partial \varphi} \propto \frac{\partial \log f_2(y)}{\partial \theta} = f_Y^{(1)}(y) \).

We need the following well-known expression for the inverse of a block matrix

\[
\begin{pmatrix}
C_{11} & C_{21} \\
C_{12} & C_{22}
\end{pmatrix}^{-1} = \begin{pmatrix}
C_{11}^{-1} + D_{12}D_{22}^{-1}D_{21} & -D_{12}D_{22}^{-1} \\
-D_{12}D_{22}^{-1} & D_{22}^{-1}
\end{pmatrix}
\]

(3.6)

where

\[ D_{22} = C_{22} - C_{21}C_{11}^{-1}C_{12}, \quad D_{12} = C_{11}^{-1}C_{12}, \quad D_{21} = C_{21}C_{11}^{-1}. \]

In this section, we introduce matrices

\[ C_{11} = J_{\theta_1}^\varphi(X), \quad C_{22} = J_{\theta_2}^\varphi(Y), \] and \( C_{12} = C_{21} = \mathcal{M}_\varphi, \)

The WFI approach offers the following WFII:

**Theorem 3.1 (Weighted Fisher information inequality (WFII), cf. [20], Theorem 1.)** Let \( X \) and \( Y \) be statistically independent random variables. Recalling notations (3.1), (3.2) and (3.3) assume that \( f_X^{(1)}(x) \) is not a proportion of \( f_Y^{(1)}(y) \). Set

\[
\Xi := \Xi_{\theta_1, \theta_2}(X, Y) = \mathcal{M}_\varphi^{-1}J_{\theta_1}^\varphi(X)G_\varphi \left( I - \mathcal{M}_\varphi G_\varphi \right)^{-1} \mathcal{M}_\varphi \left[ G_\varphi J_{\theta_2}^\varphi(Y)G_\varphi - J_{\theta_2}^\varphi(Y) \right] \\
+ G_\varphi \left( I - \mathcal{M}_\varphi G_\varphi \right)^{-1} \mathcal{M}_\varphi G_\varphi J_{\theta_2}^\varphi(Y) \left[ \mathcal{M}_\varphi^{-1}G_\varphi \right] - G_\varphi J_{\theta_2}^\varphi(Y)G_\varphi - G_\varphi.
\]

(3.7)

Then

\[
J_{\theta_1}^\varphi(X + Y) \leq \left( I - \mathcal{M}_\varphi G_\varphi \right) \left\{ \left( J_{\theta_1}^\varphi(X) \right)^{-1} + \left( J_{\theta_2}^\varphi(Y) \right)^{-1} + \Xi_{\theta_1, \theta_2}(X, Y) \right\}^{-1}.
\]

(3.8)

**Proof:** We use the same methodology as in Theorem 1 from [20]. Recalling Corollary 4.8, (iii) in [10] substitute \( P := I = [1, 1] \). Therefore for \( Z = (X, Y) \), \( J_{\theta_2}(Z) \) is a \( m \times m \) matrix with main diagonals \( J_{\theta_1}^\varphi(X) \) and \( J_{\theta_2}^\varphi(Y) \) and antidiagonals \( \mathcal{M}_\varphi: \)

\[
\left( J_{\theta_2}^\varphi(Z) \right)^{-1} = \begin{pmatrix}
J_{\theta_1}^\varphi(X) & \mathcal{M}_\varphi \\
\mathcal{M}_\varphi & J_{\theta_2}^\varphi(Y)
\end{pmatrix}^{-1}.
\]

(3.9)

where \( J_{\theta_1}^\varphi(X) \), \( J_{\theta_2}^\varphi(Y) \) and \( \mathcal{M}_\varphi \) have been introduced in (3.4) and (3.2), respectively. Define

\[
\delta := J_{\theta_2}^\varphi(Y) - \mathcal{M}_\varphi \left( J_{\theta_1}^\varphi(X) \right)^{-1} \mathcal{M}_\varphi = \left( I - \mathcal{M}_\varphi G_\varphi \right) J_{\theta_2}^\varphi(Y) \\
\Rightarrow \delta^{-1} = \left( J_{\theta_2}^\varphi(Y) \right)^{-1} \left( I - \mathcal{M}_\varphi G_\varphi \right)^{-1}.
\]

(3.10)
Here $G_\varphi$ is as before in (3.3). Thus, owing to the (3.3), particularly $P = [1, 1]$, we can write

$$P \left( J_\theta^w (Z) \right)^{-1} P^T = \left( J_\theta^w (X) \right)^{-1} + \left( J_\theta^w (X) \right)^{-1} M_\varphi \delta^{-1} M_\varphi \left( J_\theta^w (X) \right)^{-1}$$

$$- \left( J_\theta^w (X) \right)^{-1} M_\varphi \delta^{-1} - \delta^{-1} M_\varphi \left( J_\theta^w (X) \right)^{-1} + \delta^{-1} \quad (3.11)$$

Substituting $G_\varphi$, (3.3), and $\delta^{-1}$, (3.10), in above expression, we have

$$P \left( J_\theta^w (Z) \right)^{-1} P^T$$

$$= \left( J_\theta^w (X) \right)^{-1} + G_\varphi \left( I - M_\varphi G_\varphi \right)^{-1} M_\varphi \left( J_\theta^w (X) \right)^{-1} - G_\varphi \left( I - M_\varphi G_\varphi \right)^{-1}$$

$$- \left( J_\theta^w (Y) \right)^{-1} \left( I - M_\varphi G_\varphi \right)^{-1} M_\varphi \left( J_\theta^w (X) \right)^{-1} + \left( J_\theta^w (Y) \right)^{-1} \left( I - M_\varphi G_\varphi \right)^{-1} \quad (3.12)$$

Consequently by simplifying (3.12), one yields

$$P \left( J_\theta^w (Z) \right)^{-1} P^T = \left\{ \left( J_\theta^w (X) \right)^{-1} + \left( J_\theta^w (Y) \right)^{-1} + \Xi_{\theta_1, \theta_2} (X, Y) \right\} \left( I - M_\varphi G_\varphi \right)^{-1} \quad (3.13)$$

Here $\Xi$ stands as in (3.7). By using Corollary 3.4, (iii) from [10], we obtain the property claimed in (3.8):

$$J_\theta^w (X + Y) \leq \left\{ \left( J_\theta^w (X) \right)^{-1} + \left( J_\theta^w (Y) \right)^{-1} + \Xi_{\theta_1, \theta_2} (X, Y) \right\} \left( I - M_\varphi G_\varphi \right)^{-1} \quad .$$

This completes the proof.

4 Upper bound for WIC, Data processing inequality

In the following, we begin with the definition the $m \times m$ relative weighted Fisher information matrix, denoted by $D_\varphi^w (f_X \| f_Y)$, and subsequently the weighted information correlation, written as $C_\varphi^w (f_X, f_Y; \theta)$. In agreement with [21], analogously to the weighted KL-divergence [16] and Definition 1.2, also known as the weighted relative entropy, the relative WFM for given WF $\varphi \geq 0$, has been designed by replacing the ratio of two intervening density functions into Eqn. (1.5) as is established in the following definition.

**Definition 4.1** Consider two random $1 \times n$ vectors $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$ with probability density functions (PDFs) respectively $f_X (x; \theta)$ and $f_Y (y; \theta)$, $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, where $\theta = (\theta_1, \ldots, \theta_m) \in \mathbb{R}^m$ is a parameter vector. Set

$$K(X, Y; \theta) = \frac{\partial}{\partial \theta} \log \frac{f_X (x; \theta)}{f_Y (x; \theta)}$$

For given WF $(x) \in \mathbb{R}^n \rightarrow \varphi (x) \geq 0$, the $m \times m$ relative weighted Fisher information matrix is defined by

$$D_\varphi^w (f_X \| f_Y) = \mathbb{E}_X \left[ \varphi (X) K(X, Y; \theta) K(X, Y; \theta)^T \right] . \quad (4.1)$$
Further, let \( f_{X,Y}(x,y;\theta) \) be the joint PDF RV \((X,Y)\), the weighted information correlation (WIC) is introduced by

\[
C^w_\varphi(f_{X,Y};\theta) = \int \varphi(x,y) f_{X,Y}(x,y;\theta) \left( \frac{\partial \log f_X(x;\theta)}{\partial \theta} \right)^T \left( \frac{\partial \log f_Y(y;\theta)}{\partial \theta} \right) \, dx \, dy. \tag{4.2}
\]

**Lemma 4.1** (Upper bound for the WIC; cf. [21], Theorem 7.) Consider \( C^w_\varphi(f_{X,Y};\theta) \) as entries of the WIC matrix. Then we have

\[
(C^w_\varphi(f_{X,Y};\theta))_{ij}^2 \leq J^w_{\varphi_1}(f_X)_{ii} \, J^w_{\varphi_2}(f_Y)_{ii}, \tag{4.3}
\]

where \( J^w_{\varphi_1}(f_X)_{ii} \) and \( J^w_{\varphi_2}(f_Y)_{ii} \) are also entries of WFIMs in terms of \( X \) and \( Y \) with WFs

\[
\varphi_1(x) = \int \varphi(x,y) f_Y(y) \, dy, \quad \varphi_2(y) = \int \varphi(x,y) f_X(x) \, dx. \tag{4.4}
\]

**Proof:** Bound (4.3) follows with the analogue method as Theorem 7 in [21] using the non-negativity of expression below for every possible \( a \):

\[
\int \varphi(x,y) f_{X,Y}(x,y;\theta) \left( a \frac{\partial \log f_X(x)}{\partial \theta_i} + \frac{\partial \log f_Y(y)}{\partial \theta_i} \right)^2 \, dx \, dy \geq 0. \tag*{\blacksquare}
\]

**Lemma 4.2** For given WF \( \varphi \),

\[
C^w_\varphi(f_{X,Y};\theta) = 0. \tag{4.5}
\]

if at least one of the following conditions holds true:

1. Either \( f_\theta(x) \) or \( f_\theta(y) \) does not depend on \( \theta \).
2. RVs \( X \) and \( Y \) are independent and particularly \( \varphi(x,y) = \varphi_X(x) \varphi_Y(y) \), such that one of the \( E_X[\varphi_X], E_Y[\varphi_Y] \) does not depend on \( \theta \).

In this stage of work, for pair RVs \((X,Y)\) the relative WFIM, (4.1), by using the ration between their joint PDF and the product of their marginals, is called as the \( m \times m \) weighted mutual Fisher information matrix, is also defined.

**Definition 4.2** Let \( X, Y \) be two RVs, with PDFs \( f_\theta(x) = f_X(x;\theta) \), \( f_\theta(y) = f_Y(y;\theta) \) and join PDF \( f_\theta(x,y) = f_{X,Y}(x,y;\theta) \). The \( m \times m \) mutual weighted Fisher information matrix for given WF \( \varphi \) is defined by

\[
m^w_\varphi(f_{X,Y};\theta) := m^w_\varphi(X,Y;\theta)
= E_{X,Y} \left[ \varphi(X,Y) \left( \frac{\partial \log f_{X,Y}}{\partial \theta} \right) \right] \left( \frac{\partial \log f_{X,Y}}{\partial \theta} \right)^T.
\tag{4.6}
\]

Obviously, the mutual WFIM is positive-definite matrix, that’s \( m^w_\varphi(f_{X,Y};\theta) \geq 0 \).

**Theorem 4.1** For given WF \( \varphi(x,y) \geq 0 \), introduce

\[
\Theta_\varphi(X,Y;\theta) = E_{X,Y} \left[ \varphi(X,Y) \left( \frac{\partial \log f_\theta(x)}{\partial \theta} \right) \right]^T \left( \frac{\partial \log f_\theta(x)}{\partial \theta} \right) - C^w_\varphi(f_{X,Y};\theta). \tag{4.7}
\]
And
\[ \overline{\phi}(X, Y; \theta) = E_{X,Y} \left[ \phi(X, Y) \left( \frac{\partial}{\partial \theta} \log f_\theta(Y) \right) ^T \frac{\partial}{\partial \theta} \log f_\theta(X, Y) \right] - (C_\phi^w(f_{X,Y}; \theta))^T. \] (4.8)

Then
\[ m_\phi^w(f_{X,Y}; \theta) = J_\phi^w(f_{X|Y}; \theta) + J_\phi^w(f_{X}; \theta) - \Theta_\phi(X, Y; \theta) - (\Theta_\phi(X, Y; \theta))^T. \] (4.9)

Note that here \( \phi_1 \) and \( \phi_2 \) are as defined in \([4.4]\).

**Proof:** We begin with
\[ \left( \frac{\partial}{\partial \theta} \log \left( \frac{f_\theta(X, Y)}{f_\theta(X)f_\theta(Y)} \right) \right) ^T \frac{\partial}{\partial \theta} \log f_\theta(X, Y) = \left( \frac{\partial}{\partial \theta} \log f_\theta(X|Y) \right) ^T \frac{\partial}{\partial \theta} \log f_\theta(X,Y) + \left( \frac{\partial}{\partial \theta} \log f_\theta(X) \right) ^T \frac{\partial}{\partial \theta} \log f_\theta(X,Y). \] (4.10)

Furthermore, observe that
\[ E_{X,Y} \left[ \phi(X, Y) \left( \frac{\partial}{\partial \theta} \log f_\theta(X) \right) ^T \frac{\partial}{\partial \theta} \log f_\theta(X|Y) \right] = E_{X,Y} \left[ \phi(X, Y) \left( \frac{\partial}{\partial \theta} \log f_\theta(Y) \right) ^T \frac{\partial}{\partial \theta} \log f_\theta(X, Y) \right] - C_\phi^w(f_{X,Y}; \theta). \]

Thus, the bound in (4.9) then follows. \( \blacksquare \)

**Lemma 4.3** (Conditioning increases Information in weighted version; cf. \([21]\), Theorem 12.) Assume \( f_\theta(y|x) \) depends on \( \theta \) while \( f_\theta(x) \) does not. For given WF \( \phi \geq 0 \), recall \( \phi_2 \) from \([4.2]\) and set
\[ \Theta_\phi^*(X, Y; \theta) = E_{X,Y} \left[ \phi(X, Y) \left( \frac{\partial}{\partial \theta} \log f_\theta(X) \right) ^T \frac{\partial}{\partial \theta} \log f_\theta(X, Y) \right]. \]

We have
\[ J_\phi^w(f_{Y|X}; \theta) \geq \Theta_\phi^*(X, Y; \theta) + (\Theta_\phi^*(X, Y; \theta))^T - J_\phi^w(f_{X}; \theta). \] (4.11)

**Proof:** Following Lemma 4.2 since \( f_\theta(x) \) does not depend on \( \theta \) we get \( C_\phi^w(f_{X,Y}; \theta) = 0 \). Hence owing to Theorem 4.1, \( 4.9 \), the inequality (4.11) is proved. \( \blacksquare \)

**Theorem 4.2** (Data processing inequality for the mutual WFIM; cf. \([21]\), Theorem 13.) Consider a Markov chain \( X \rightarrow Y \rightarrow T \) created by triple RVs \( (X, Y, T) \), such that among all PDFs only \( f_\theta(t|y) \) depends on \( \theta \). Then for given WF \( \phi \),
\[ m_\psi^w(X,T; \theta) \leq m_\psi^w(Y,T; \theta). \] (4.12)

Here
\[ \psi_{XT} := \psi_{X,T}(x, t) = \int \phi(x, y,t) f_\theta(y|x,t) \, dy, \quad \text{and} \]
\[ \psi_{YT} := \psi_{Y,T}(y, t) = \int \phi(x, y,t) f_\theta(x|y,t) \, dx. \]
Proof. We prove the assertion (4.12) in five steps:
Step 1.
\[
m_w^N(X, Y, T; \theta) = J_w^N(X, Y | T; \theta) + J_{wXY}^w(X, Y; \theta) - \Theta_w^w(X, Y, T; \theta) - \left(\Theta_w(X, Y, T; \theta)\right)^T.
\] (4.13)
This claim is verified by observing that
\[
\Theta_w(X, Y, T; \theta) = \Theta_w(X, Y, T; \theta) - C_w^w(X, Y, T; \theta),
\]
where
\[
\Theta_w^w(X, Y, T; \theta) = E \left[\varphi(X, Y, T) \left(\frac{\partial}{\partial \theta} \log f_\theta(X, Y)\right)^T \left(\frac{\partial}{\partial \theta} \log f_\theta(X, Y, T)\right)\right].
\] (4.14)
Step 2. By recalling conditional chain rule, cf. [16], Lemma 4.3, we get
\[
\begin{align*}
J_w^w(X, Y | T; \theta) &= J_w^w(X | Y, T; \theta) + J_{wYT}^w(Y | T) + S_{YT}(\theta),
\end{align*}
\] (4.15)
where
\[
\begin{align*}
B_{YT}(\theta) &:= B_{Y, T}(y, t; \theta) = \int \varphi(x, y, t) \frac{\partial}{\partial \theta} \log f_\theta(x | y, t) \, dx, \\
S_{YT}(\theta) &= E \left\{ \left(\frac{\partial}{\partial \theta} \log f_\theta(Y | T)\right)^T B_{YT} + (B_{YT})^T \frac{\partial}{\partial \theta} \log f_\theta(Y | T) \right\}.
\end{align*}
\] (4.16)
Step 3.
\[
J_{wXY}^w(X, Y; \theta) = J_{wXY}^w(X | Y) + J_{wY}^w(Y) + S_Y(\theta),
\] (4.17)
similarly, here we define \(S_Y(\theta)\) as above \(S_{YT}(\theta)\) by replacing \(f_\theta(Y)\) and \(B_{YT}(\theta)\) in \(f_\theta(Y | T)\), \(B_{YT}(\theta)\), respectively. Moreover note that
\[
B_Y(\theta) = \int \psi_{XY}(x, y) \frac{\partial}{\partial \theta} \log f_\theta(x | y) \, dx
= \int \varphi(x, y, t) \left(\frac{f_\theta(x, y | t) f_\theta(y)}{f_\theta(x, y)}\right) \frac{\partial}{\partial \theta} \log f_\theta(x | y) \, dx dt.
\]
Step 4.
\[
\begin{align*}
J_w^w(X | Y, T; \theta) + J_{wXY}^w(X | Y; \theta) &= m_w^w(X | T | Y; \theta) + \Theta_w^w(X, T | Y; \theta) \\
&- C_w^w(X, T | Y; \theta) + \left(\Theta_w^w(X, T | Y; \theta)\right)^T - \left(C_w^w(X, T | Y; \theta)\right)^T.
\end{align*}
\] (4.18)
It can be prejudiced that \(\Theta_w^w(X, T | Y; \theta)\) and \(C_w^w(X, T | Y; \theta)\) stand as the conditional form of \(\Theta_w^w\) in [11,14] and the WCI, respectively.
Step 5. Analogous with step 3, we can write
\[
\begin{align*}
J_w^{wYT}(Y | T; \theta) + J_w^w(Y; \theta) &= m_w^{wYT}(Y, T; \theta) + \Theta_w^{wYT}(Y, T; \theta) \\
&- C_w^{wYT}(Y, T; \theta) + \left(\Theta_w^{wYT}(Y, T; \theta)\right)^T - \left(C_w^{wYT}(Y, T; \theta)\right)^T.
\end{align*}
\] (4.19)
Combining Steps 1-5 and the fact that only \(f_\theta(t | y)\) depends on \(\theta\), one derives
\[
m_w^w(X, Y, T; \theta) = m_w^w(X, T | Y; \theta) + m_{wXT}^w(Y, T; \theta).
\] (4.18)
Similarly it can be obtained
\[
m_w^w(X, Y, T; \theta) = m_w^w(Y, T | X; \theta) + m_{wXT}^w(X, T; \theta).
\] (4.19)
Consequently, since \(X, T\) are independent given \(Y\), thus \(m_w^w(X, T | Y; \theta) = 0\). Further, \(m_w^w(Y, T | X; \theta) \geq 0\). This implies the claimed result [11,12].
5 The WFIM and additive Gaussian noise

In the final section of the paper, we first illustrate a multivariate representation theorem due to Lemma 1 from [12], in which provides the needed connection between weighted conditional variance and WFIM via considering a system with additive Gaussian noise involved. Furthermore, as interesting achievement, an extended version for De-Bruijn’s identity is proposed.

5.1 Extended De-Bruijn’s identity

**Theorem 5.1** Consider additive RV $Y = X + Z$, such that $n$-Gaussian vector $Z$ follows mean vector zero, covariance matrix $\Sigma$ and is independent of $X$. Introduce matrices

\[ V_\varphi(X) = E\left[ \varphi\left( X - E[X]\right)^T (X - E(X)) \right], \]

\[ E_\varphi = E\left[ \varphi\left( Y - E[X|Y]\right)^T (X - E[X|Y]) \right], \quad E_\varphi = E_\varphi + E_\varphi^T. \quad (5.1) \]

The WFIM of RV $Y$ is merely

\[ J^\varphi_W(Y) = (\Sigma^{-1})^T \left\{ E\left[ \varphi Z_\Sigma^T Z_\Sigma \right] + E_\varphi - V_\varphi(X|Y) \right\} \Sigma^{-1}. \quad (5.2) \]

**Proof:** As in [12], we deal with the conditional representation of the score function for the additive RVs, so we have

\[ \nabla \log f_Y(y) = E\left[ \nabla \log f_{N_0}(Z_\Sigma|X + Z_\Sigma = y) \right] = E\left[ -Z_\Sigma^T \Sigma^{-1} Y = y \right] = \left( E[X|Y = y] - y \right)^T \Sigma^{-1}. \quad (5.3) \]

Therefore, owing to the definition of WFIM, one yields

\[ J^\varphi_W(Y) = (\Sigma^{-1})^T \left\{ E\left[ \varphi Z_\Sigma^T Z_\Sigma \right] + E_\varphi - V_\varphi(X|Y) \right\} \Sigma^{-1}. \quad (5.4) \]

On the other hand, it can be derived

\[ E\left[ \varphi Z_\Sigma^T (X - E[X|Y]) \right] = E_\varphi - V_\varphi(X|Y), \quad (5.5) \]

by denoting $E_\varphi, V_\varphi$ as before in (5.1). Next, substitute (5.5) into the last term of (5.4) the desired results is achieved. ■

**Theorem 5.2** Consider independent RVs $X$ and $Z$ with finite covariance and mean vector zero. For given WF $\varphi \geq 0$ and scalar $\gamma$ set

\[ \varphi_\gamma := \varphi_\gamma(x,z) = \varphi(x + \sqrt{\gamma}z, z), \quad (5.6) \]
then
\[ i_\psi^w(X + \sqrt{\gamma}Z; Z) = \frac{\gamma}{2} \mathbb{E}_Z \left\{ Z J_{\psi,Z}^w(X) Z^T \right\} + \sqrt{\gamma} \mathbb{E}_Z \left\{ Z \mathbb{E}_X \left[ \varphi_{\gamma} (\nabla \log f(X))^T \right] \right\} - \frac{\gamma}{2} \mathbb{E}_Z \left\{ Z Z^T \mathbb{E}_X \left[ \varphi_{\gamma} \frac{\Delta f(X)}{f(X)} \right] \right\} + o(\gamma) \mathbb{E}_Z \{ \mathbb{E}_X[\varphi_{\gamma}] \}. \] (5.7)

Particularly, if \( Z_1 \in \mathbb{R}^n \sim N(0, I) \) takes values \( z_1 \), the WF \( \varphi(x, z_1) = x z_1^T \), hence \( \varphi = x z_1^T + \sqrt{\gamma} z_1 z_1^T \), then \( 5.7 \) becomes:
\[
\begin{align*}
&i_\psi^w(X + \sqrt{\gamma}Z_1; Z_1) \\
&= \sqrt{\gamma} n \left\{ \frac{\gamma}{2} \text{tr} J(X) + \mathbb{E}_X \left[ X(\nabla \log f(X))^T \right] \right\} - \frac{\gamma}{2} \mathbb{E}_X \left[ \frac{\Delta f(X)}{f(X)} \right] + o(\gamma) \right\}.
\end{align*}
\]
Here \( J(X) \) denotes the WFI when \( \varphi \equiv 1 \).

**Proof:** By the definition of mutual WE, (1.6), between RVs \( Y = X + \sqrt{\gamma}Z \) with density \( g(y) \), and \( Z \) having PDF \( f_Z \) let \( \theta = \sqrt{\gamma} \), hence one gets
\[
\begin{align*}
i_\psi^w(X + \sqrt{\gamma}Z; Z) &= \int f_Z(z) \left\{ \varphi(y, z) f_X(y - \theta z) \log \frac{f_X(y - \theta z)}{g(y)} \right\} dy dz \\
&= \int f_Z(z) \left\{ \varphi(u + \theta z, z) f_X(u) \log \frac{f_X(u)}{g(u + \theta z)} \right\} dy dz = \mathbb{E}_Z \left\{ D_{\psi\theta}^w(g_0 \parallel g_0) \right\}.
\end{align*}
\] (5.8)

where the reduced WF \( \varphi_\theta \) is given as \( \varphi_{\gamma} \) replacing \( \gamma = \theta^2 \). Note that the second integral in (5.8) is achieved by changing the RV \( u = y - \theta z \) and moreover \( g_\theta = g_{X+\theta Z}(u + \theta z) \) is the parameterized family of densities of RV \( U \), and therefore \( g_0 = f_X(u) \). Now call Theorem (1.1) for scalar parameter \( \theta \) we can write
\[
\begin{align*}
i_\psi^w(X + \sqrt{\gamma}Z; Z) &= \frac{\theta^2}{2} \mathbb{E}_Z \left\{ J_{\psi\theta}^w(g_0) \right\} + \theta \mathbb{E}_Z \left\{ Z \mathbb{E}_X \left[ \varphi_{\theta} (\nabla \log g_0)^T \right] \right\} \\
&= \frac{\theta^2}{2} \mathbb{E}_Z \left\{ Z Z^T \mathbb{E}_X \left[ \varphi_{\theta} \frac{\Delta g_0}{g_0} \right] \right\} + o(\theta^2) \mathbb{E}_Z \{ \mathbb{E}_X[\varphi_{\theta}] \}. \tag{5.9}
\end{align*}
\]
Here \( J_{\psi\theta}^w(g_0) \) stands the parametric WFI of \( U \) about \( \theta = 0 \). On the other hand, using the first order Taylor expansion about \( \theta = 0 \) we verify
\[
g_\theta(u) = g_{X+\theta Z}(u + \theta z) = f_X(u) + \theta z \left( \nabla f_X(u) \right)^T + o(\theta).
\]
Thus, for \( \theta = 0 \) we compute score function:
\[
S_0(u) = \frac{\partial}{\partial \theta} \log g_\theta(u) \big|_{\theta = 0} = z \left( \nabla \log g_0(u) \right)^T.
\]
Consequently,
\[
\begin{align*}
J_{\psi\theta}^w(g_0) &= \mathbb{E}_U \left[ \varphi_{\theta} S_0(u) (S_0(u))^T \right] \\
&= z \mathbb{E}_U \left[ \varphi_{\theta} \left( \nabla \log g_0(U) \right)^T \nabla \log g_0(U) \right] z^T \\
&= z J_{\psi\theta}^w(X) z^T.
\end{align*}
\]
Plugging this expression into (5.9) gives (5.7) as required. \( \blacksquare \)
Definition 5.1 Given a function $x \in \mathbb{R}^n \mapsto \varphi(x) \geq 0$, and an RV $X$ with a PDF $f$, the weighted entropy (WE) of $X$ with WF $\varphi$ is defined by

$$h^w_\varphi(X) = -\mathbb{E}_X[\varphi(X) \log f(X)] = -\int \varphi(x) f(x) \log f(x) \, dx.$$  \hspace{1cm} (5.10)

Moreover, we propose the notion of weighted entropy power (WEP) as follows:

$$N^w_\varphi(X) := N^w_\varphi(f) = \exp \left\{ \frac{2}{n} h^w_\varphi(X) \right\}. \hspace{1cm} (5.11)$$

Theorem 5.3 (Extended De-Bruijn’s identity), cf. [14]: Let $X \sim f$ be a RV in $\mathbb{R}^n$, with a PDF $f \in C^2$. For standard Gaussian RV $Z_I \sim N(0, I)$ independent of $X$ and given $\gamma > 0$, let $Y = X + \sqrt{\gamma} Z_I$ having PDF $g$. Let $V_r$ be the $n$-sphere of radius $r$ centered at the origin and having surface denoted by $S_r$. Assume that for given WF $\varphi$,

$$\lim_{r \to \infty} \int_{S_r} \varphi(y) \log g(y) \left( \nabla g(y) \right) dS_r = 0,$$  \hspace{1cm} (5.12)

is fulfilled. Then

$$\frac{d}{d\gamma} h^w_\varphi(Y) = \frac{1}{2} \text{tr} J^w_\varphi(Y) - \frac{1}{2} \mathbb{E} \left\{ \varphi \frac{\Delta g(Y)}{g(Y)} \right\} + \mathcal{R}(\gamma). \hspace{1cm} (5.13)$$

Here

$$\mathcal{R}(\gamma) = \mathbb{E} \left[ \nabla \varphi \log g(Y) \left( \nabla \log g(Y) \right)^T \right]. \hspace{1cm} (5.14)$$

Proof: Going back to the RV $Y$, it follows PDF:

$$g(y) = \frac{1}{(2\pi\gamma)^{n/2}} \int f(v) \exp \left( -\frac{\|y - v\|^2}{2\gamma} \right) \, dv, \hspace{0.5cm} y \in \mathbb{R}^n. \hspace{1cm} (5.15)$$

Differentiating both sides of (5.15) with respect to $\gamma$ yields

$$\frac{d}{d\gamma} g(y) = \frac{1}{2} \Delta g(y). \hspace{1cm} (5.16)$$

Subsequently, one gets

$$\frac{d}{d\gamma} h^w_\varphi(Y) = -\int \varphi(y) \frac{d}{d\gamma} g(y) \log g(y) \, dy - \int \varphi(y) \frac{d}{d\gamma} g(y) \, dy = -\frac{1}{2} \int \varphi(y) \Delta g(y) \log g(y) \, dy - \frac{1}{2} \int \varphi(y) \Delta g(y) \, dy. \hspace{1cm} (5.17)$$

Now let us call Green’s identity. Suppose $\rho(x)$ and $\eta(x)$ are twice continuously differentiable functions in $\mathbb{R}^n$ and $V$ is a domain bounded by a piecewise smooth, closed and oriented surface $S$ in $\mathbb{R}^n$. Then

$$\int_V \rho(x) \Delta \eta(x) \, dx = \int_S \rho(x) \nabla \eta(x) \, dS - \int_V \nabla \rho(x) \left( \nabla \eta(x) \right)^T \, dx. \hspace{1cm} (5.18)$$

Here $\nabla \varphi$ denotes the gradient of $\eta$, $dS$ denotes the differential area vector, and $\nabla \eta \cdot dS$ is the inner produce of these two vectors.
Let now \( V_r \) be the \( n \) ball of radius \( r \) centered at the origin and \( S_r \) be its boundary (an \( n \)-sphere of radius \( r \)). Then we use Green’s identity on \( V_r \) and \( S_r \), with \( \rho(y) = \varphi(y) \log g(y) \), and \( \eta(y) = g(y) \). Then take the limit as \( r \to \infty \). Using the assumption (5.12) we can write:

\[
\frac{d}{d\gamma} h_\varphi^w(Y) = \frac{1}{2} \int \nabla \varphi(y) (\nabla g(y))^T \log g(y) dy + \frac{1}{2} \int \varphi(y) \nabla \log g(y) (\nabla g(y))^T dy - \frac{1}{2} \int \varphi(y) \Delta g(y) dy.
\] (5.19)

This concludes the proof. ■

**Remark 5.1** For RV \( Y = X + \sqrt{\gamma} Z \), \( Z \) as before, with PDF \( g \), (5.15), we obtain:

\[
\frac{d}{d\gamma} E \varphi(Y) = \frac{d}{d\gamma} \int \varphi(y) g(y) dy = \frac{1}{2} \int \varphi(y) \Delta g(y) dy = \frac{1}{2} E \left[ \varphi(Y), \frac{\Delta g(Y)}{g(Y)} \right].
\] (5.20)

Therefore (5.13) becomes

\[
\frac{d}{d\gamma} h_\varphi^w(Y) = \frac{1}{2} \text{tr} J_\varphi^w(Y) - \frac{1}{2} \frac{d}{d\gamma} E \varphi(Y) + \frac{R(\gamma)}{2}.
\] (5.21)

Now we shall compute the second derivative of WEP. Denote

\[
\Lambda(\gamma) = \frac{2}{n} \frac{d}{d\gamma} h_\varphi^w(Y) \left[ E[\varphi(Y)] \right]
\] (5.22)

and calculate

\[
\frac{d^2}{d\gamma^2} \exp \left\{ \left( \frac{2}{n} \frac{h_\varphi^w(Y)}{E[\varphi(Y)]} \right) \right\}
= \exp \left\{ \left( \frac{2}{n} \frac{h_\varphi^w(Y)}{E[\varphi(Y)]} \right) \right\} \left[ \left( \frac{2}{n} \frac{h_\varphi^w(Y)}{E[\varphi(Y)]} \right)^2 + \left( \frac{2}{n} \frac{d}{d\gamma} \frac{h_\varphi^w(Y)}{E[\varphi(Y)]} \right) \right] (5.23)
\]

5.2 The entropy power (EP) is a concave function

In the literature, several elegant proofs, employing for instance the Fisher information inequality \[6, 5\] or basic properties of mutual information \[13\], has been proposed in order to express that the entropy power is a concave function. In the spirit of WEP, we shall present a new proof for the fact of concave EP. Regarding this, let us apply (5.18) to \( \varphi \equiv 1 \). Then we have

\[
\left. \frac{d}{d\gamma} n \frac{h_\varphi^w(Y)}{J(Y)} \right|_{\gamma=0} \geq 1
\] (5.24)

**Theorem 5.4** Suppose that \( \forall \gamma \in (0, 1) \)

\[
\int f_Y(x) | \ln f_Y(x) | dx < \infty, \int |\nabla f_Y(y) \ln f_Y(y)| dy < \infty
\]
and for \( \forall \gamma \in (0, 1) \)

\[
\frac{d}{d\gamma} \frac{n}{J(Y)} \geq 1 + \epsilon.
\]  

(5.25)

Then \( \exists \delta = \delta(\epsilon) \) such that \( \forall \) weight functions \( \varphi \) such that \( \exists c > 0: |\varphi - c| < \delta, |\nabla \varphi| < \delta \)

the entropy power (EP)

\[
\exp \left( \frac{2}{n} \frac{h^w_\varphi(Y)}{E[\varphi(Y)]} \right)
\]

(5.26)

is a concave function of \( \gamma \). Under assumption

\[
\frac{d}{d\gamma} n J(Y) \bigg|_{\gamma=0} \geq 1 + \epsilon
\]

(5.27)

the EP is a concave function of \( \gamma \) in a small enough neighborhood of \( \gamma = 0 \).

**Proof:** It is sufficient to check that

\[
\frac{d}{d\gamma} \psi(\gamma) \geq 1
\]

(5.28)

where

\[
\psi(\gamma) = \left( \frac{2}{n} \frac{d}{d\gamma} \frac{h^w_\varphi(Y)}{E[\varphi(Y)]} \right)^{-1} = \Lambda(\gamma)^{-1}.
\]

(5.29)

Compute the derivative in (5.29) using (3.24) and obtain

\[
\psi(\gamma) = \frac{n}{J(Y)} + o(\delta).
\]

(5.30)

Indeed, by a straightforward calculation

\[
\psi(Y) = \frac{n\langle E[\varphi(Y)] \rangle^2}{\frac{d}{d\gamma} h^w_\varphi(Y) E[\varphi(Y)] - h^w_\varphi(Y) \frac{d}{d\gamma} E[\varphi(Y)]} + \frac{1}{2} \mathcal{R}(\gamma).
\]

W.l.o.g. assume that \( c = 1 \), and write

\[
\frac{d}{d\gamma} h^w_\varphi(Y) = \frac{1}{2} \text{tr} J^w_\varphi(Y) - \frac{1}{2} \frac{d}{d\gamma} E[\varphi(Y)] + \frac{1}{2} \mathcal{R}(\gamma).
\]

Clearly,

\[
1 - \delta < E[\varphi(Y)] < 1 + \delta, |\text{tr} J^w_\varphi(Y) - \text{tr} J(Y)| < \delta
\]

Next,

\[
\frac{d}{d\gamma} E[\varphi(Y)] = \frac{1}{2} \int \varphi(y) \Delta f_Y(y) dy
\]

and using the Stokes formula one bounds this term by \( \delta \). Finally,

\[
|\mathcal{R}(\gamma)| \leq \int |\nabla \varphi(y)^T \nabla f_Y(y) \ln f_Y(y)| dy < \delta
\]

leading to the claimed result. ■

**Acknowledgement**

MK thanks the Higher School of Economics for the support in the framework of the Global Competitiveness Programme. YS thanks the Office of the Rector, University of Sao Paulo (USP) for the financial support during the academic year 2013-4. YS thanks Math Department, Penn State University, USA for the hospitality and support during the academic years 2014-6. SYS thanks the CAPES PNPD-UFSCAR Foundation for the financial support in the year 2014-5. SYS thanks the Federal University of Sao Carlos, Department of Statistics, for hospitality during the year 2014-5.
References

[1] M. Belis and S. Guiasu. A Quantitative and qualitative measure of information in cybernetic systems. *IEEE Trans. on Inf. Theory*, **Vol 14** (1968), 593–594.

[2] A. Clim. Weighted entropy with application. *Analele Universității București, Matematică*, **Anul LVII** (2008), 223-231.

[3] M. H. M. Costa and T. M. Cover. On the similarity of the entropy power inequality and the Brunn-Minkowski inequality. *IEEE Trans. Inform. Theory*, **Vol 30** (1984), no. 6, 837–839.

[4] M. Cohen. The Fisher information and convexity. *IEEE Transaction on Information Theory*, **Vol 14** (1968), 591–592.

[5] A. Dembo. Simple proof of the concavity of the entropy power with respect to added Gaussian noise. *IEEE Trans. Inform. Theory*, **Vol 35** (1989), no. 4, 887–888.

[6] A. Dembo, T. M. Cover and J. A. Thomas. Information theoretic inequalities. *IEEE Trans. Inform. Theory*, **Vol 37** (1991), 1501–1518.

[7] R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philosoph. Trans. Roy. Soc. London*, **Vol 222** (1922), 309–368.

[8] R. A. Fisher. Theory of statistical estimation. *Rroc. Cambridge Philosoph. Soc.*, **Vol 22** (1925), no.5, 700–725.

[9] S. Guiasu. Weighted entropy. *Report on Math. Physics*, **2** (1971), 165–179.

[10] O. Johnson. A consitional power entropy inequality for dependent variables. *IEEE Trans. Inform. Theory*, **Vol 50** (2004), no. 8, 1581–1583.

[11] O. Johnson and Y. Suhov. Entropy and random vectors. *J. Statist. Phys.*, **104** (2001), no. 3, 145–165.

[12] T. Liu and P. Viswanath. Two Proofs of the Fisher Information Inequality via Data Processing Arguments. arXiv preprint cs/0604028 (2006).

[13] O. Rioul. Information theoretic proofs of entropy power inequalities. *IEEE Trans. Inform. Theory*, **Vol 57** (2011), n0 1, 33–55.

[14] A. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Inform. Contr.*, **Vol 2** (1959), 101–102.

[15] M. Stein, A. Mezghani and J. A. Nossek. A lower bound for the Fisher information measure. *Signal Processing Letters, IEEE*, **Vol 21** (2014), no 7, 796–799.

[16] Y. Suhov, I. Stuhl, S. Yasaei Sekeh and M. Kelbert. Basic inequalities for weighted entropies. *Aequationes Mathematicae* (2015), In Press.

[17] Suhov, Y.; Yasaei Sekeh, S.; Stuhl, I. Weighted Gaussian entropy and determinant inequalities entropy. arXiv: 1502.02188

[18] Suhov, Y.; Yasaei Sekeh, S.; Kelbert, M. Entropy-power inequality for weighted entropy. arXiv: 1502.02188
[19] Yasaei Sekeh, S. Extended inequalities for weighted Renyi entropy involving generalized Gaussian densities. arXiv: 1509.02190

[20] R. Zamir. A proof of the Fisher information inequality via a data processing argument. IEEE Transaction on Information Theory, 44, No. 3 (1998), 1246–1250.

[21] P. Zerges. Fisher information properties. Entropy, 17 (2015), 4918–4939.

Mark Kelbert: Math Dept, University of Swansea, UK; Moscow Higher School of Economics, RF

Yuri Suhov: DPMMS, University of Cambridge, UK; Math Dept, Penn State University, PA, USA; IPIT RAS, Moscow, RF

Salimeh Yasaei Sekeh: Stat Dept, Federal University of São Carlos, SP, Brazil