Generalization of the de Bruijn’s identity to general $\phi$-entropies and $\phi$-Fisher informations

Irene Valero Toranzo, Steeve Zozor and Jean-Marc Brossier

Abstract

In this paper, we propose generalizations of the de Bruijn’s identities based on extensions of the Shannon entropy, Fisher information and their associated divergences or relative measures. The foundation of these generalizations are the $\phi$-entropies and divergences of the Csiszár’s class (or Salicrú’s class) considered within a multidimensional context, included the monodimensional case, and for several type of noisy channels characterized by a more general probability distribution beyond the well-known Gaussian noise. It is found that the gradient and/or the hessian of these entropies or divergences with respect to the noise parameters give naturally rise to generalized versions of the Fisher information or divergence, which are named as the $\phi$-Fisher information (divergence). The obtained identities can be viewed as further extensions of the classical de Bruijn’s identity. Analogously, it is shown that a similar relation holds between the $\phi$-divergence and a extended mean-square error, named $\phi$-mean square error, for the Gaussian channel.

Index Terms

Communication channels, $\phi$-entropy and $\phi$-divergences, $\phi$-Fisher information, generalized de Bruijn’s identities.

I. INTRODUCTION

The goal of this paper is to extend the de Bruijn’s identity, relating two quantities of information, namely the differential Shannon entropy of the output of a Gaussian channel, and its Fisher information [1]. These two quantities are very important in information theory, in statistics, in statistical physics and in signal processing [2], [3], [4], [5], [6], [7], [8], [9], [10], [11].

The study of the notion of information related to a random variable (r.v.), or to a parameter attached to a r.v., is a huge long outstanding field of investigation. The sense attributed to “information” is closely linked to its field of application. The most usual measures used to quantify such an information can be viewed to be the vertices of a triangle, as symbolically depicted in figure [2] and are

1. V. Toranzo is with both the GIPSA-Lab, Image and Signal Processing Departement, 11 rue des Mathématiques, 38402 St Martin d’Hères, France and the Departamento de Física Atómica, Molecular y Nuclear, Universidad de Granada, 18071-Granada, Spain (ivtoranzo@ugr.es)
2. S. Zozor and J.-M. Brossier are with the GIPSA-Lab, Image and Signal Processing Departement, 11 rue des Mathématiques, 38402 St Martin d’Hères, France (steeve.zozor@gipsa-lab.grenoble-inp.fr; jean-marc.brossier@gipsa-lab.grenoble-inp.fr)
• The moments of a $d$-dimensional r.v. $X$, typically

$$
E[f(X)] = \int_{\Omega} f(x) \, p_X(x) \, dx
$$

for some function $f$ (independent of the pdf), where $p_X$ stands for the probability density function (pdf) of $X$ and $\Omega \subset \mathbb{R}^d$ its support; For $f(x) = x$, the mean $m_X$ describes where the pdf is centered and for $f(x) = (x - E[X]) \cdot (x - E[X])^t$ where $\cdot^t$ stands for the transposition the covariance matrix $C_X$ of $X$ describes the spread of the pdf around its mean: in some sense, these are two “information measures” regarding the pdf. A typical associated measure of interest is the Mean-Square Error (MSE) of an estimator $\hat{\theta}(X)$ of a parameter $\theta$, built using an observed variable $X$ parametrized by $\theta$,

$$
\text{MSE}(\hat{\theta}) = E \left[ (\hat{\theta} - \theta) \left( \hat{\theta} - \theta \right)^t \right]
$$

This quantity is widely used in estimation in order to assess the quality of an estimator for instance (its trace gives the “power” of the estimation error).

• The differential Shannon entropy of a r.v. is defined as \cite{12, 2, 13}

$$
H(X) = -\int_{\Omega} p_X(x) \log(p_X(x)) \, dx,
$$

and taking the exponential, one obtains the quantity known as the entropy power

$$
N(X) = \frac{1}{2\pi e} \exp \left( \frac{2}{d} H(X) \right),
$$

which is generally viewed as a measure of uncertainty. Indeed for any invertible (deterministic) matrix $A$ and any (deterministic) vector $b$ one has $N(AX + b) = |A|^2 N(x)$ (where $|\cdot|$ stands for the absolute value of the determinant). Thus, when $|A|$ goes to 0, $AX + b$ tends to be deterministic and its uncertainty goes to 0. At the opposite, when $|A|$ goes to the infinity, the law of $X$ tends to be highly dispersed and the uncertainty tends to be infinite. $H$ can also be viewed as the “information” brought by an observation or outcome. This quantity was naturally introduced in the context of communication, and the associate measure of particular interest is the mutual information between two random variables, $I(X;Y) = H(X) + H(Y) - H(X,Y)$, i.e.,

$$
I(X;Y) = \int_{\Omega} p_{X,Y}(x,y) \log \left( \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)} \right) \, dx \, dy
$$

This measure is fundamental as it quantifies the information transmitted through a communication channel while the maximal input-output information gives the channel capacity. The mutual information can be written through the Kullback-Leibler divergence, also called relative entropy \cite{2},

$$
D_{\text{KL}}(p||q) = \int_{\Omega} p(x) \log \left( \frac{p(x)}{q(x)} \right) \, dx
$$

that is a kind of distance of a pdf $p$ to a pdf $q$ that serves as reference: $D_{\text{KL}}(p||q) \geq 0$ with equality if and only if $p = q$ almost everywhere, but it is not symmetric and does not satisfy the triangle inequality \cite{2}.

\footnote{In this paper, vectors are column vectors.}
• The last “vertex” of the informational triangle given figure \(2\) is the Fisher information matrix relatively to a
\(n\)-dimensional parameter \(\theta\) attached to a r.v. \(X\) [1, 4, 3].
\[
J_\theta(X) = \int_\Omega \left[\nabla_\theta \log(p_X(x))\right] \left[\nabla_\theta \log(p_X(x))\right]^t p_X(x) \, dx
\]
where \(\nabla_\theta f = \left[\frac{\partial f}{\partial \theta_1}, \ldots, \frac{\partial f}{\partial \theta_n}\right]^t\) denotes the gradient of \(f\) versus \(\theta = [\theta_1, \ldots, \theta_n]^t\). Function \(\nabla_\theta \log(p_X)\)
is known as the score function (versus \(\theta\)) of the pdf. This matrix is highly popular in the estimation field as
it quantifies the information on \(\theta\) carried by the r.v. \(X\). As we will see in a few lines, it allows to bound the
variance of an estimator. When \(\theta\) is a location parameter (for example the mean of the variable), the gradient
in \(\theta\) can be replaced by a gradient in \(x\), the Fisher is then known as the nonparametric Fisher information
matrix, simply denoted by \(J(X)\).

Although they come from different scientific fields (probability theory, digital communications, estimation, . . .)
these quantities are generally related to each other, very often by inequalities, as symbolically represented by the
“edges” of the triangle in figure \(2\). Among the classical ones, given in [14], [2] for instance, let us mention some of them :
• The moment-entropy relations \(N(X) \leq |C_X|\) where \(\cdot|\cdot\) denotes the determinant. This relation is also detailed
and extended in a series of papers by Lutwak et al. [15], [16], [17], [18] or by Bercher [19], [20].
• The Cramér-Rao inequality that links the variance of a r.v. — or of an estimator — to the Fisher information,
\(C_X - J(X)^{-1} \geq 0\) and \(\text{MSE}(\hat{\theta}) - J_\theta(X)^{-1} \geq 0\) (in the unbiased context), where \(A \geq 0\) means that matrix
\(A\) is positive [4]. This inequality also gave rise to extensions [16], [18], [19], [21], [20].
• The Fisher information appears to be the curvature of the Kullback-Leibler divergence: for a pdf parametrized
by \(\theta \in \Theta\), for a given \(\theta_0 \in \Theta\), the second-order Taylor series expansion versus \(\theta\) in \(\theta = \theta_0\) writes \(D_{kl}(p_\theta||p_{\theta_0}) = \frac{1}{2}(\theta - \theta_0)^t J_{\theta_0}(X)(\theta - \theta_0) + o(\|\theta - \theta_0\|^2)\) [2], [13].
• The Stam’s inequalities lower bound the product between the entropy power and the trace or the determinant
of the Fisher information matrix [1, 22, 14, 2], \(N(X)\text{Tr}(J(X)) \geq d\) where \(\text{Tr}\) stands for the trace operator
and \(N(X)|J(X)|^{\frac{1}{2}} \geq 1\). As for the previous inequality, the Stam’s one were also extended by Lutwak or by
Bercher [16], [18], [19], [20].
• The two following relations we are precisely interested in here, due to de Bruijn and Guo et al. respectively, are
remarkable since they link two information measures by identities rather than inequalities. They deal with the
Gaussian channel, as depicted in figure [1] where \(G\) is a zero-mean standard Gaussian noise independent of the
input \(X\). Under some regularity assumptions, the de Bruijn’s identity links the variation of the entropy of the
output’s pdf with respect to the noise variance, and its Fisher information [1]. The Guo-Shamai-Verdú relation
links the variations of the input-output mutual information with respect to the input power and the MMSE of
the estimation of \(X\) from the output \(Y\), \(\text{MMSE}(X|Y) = \text{MSE}(E[X|Y])\) (see [4]). For the Gaussian scalar
context, these relations are recalled in figure [1] Several alternative formulations exists in terms of Kullback-Leibler
divergence versus Fisher divergence [23], [24].

These relations are precisely at the heart of our paper. Our goal is to generalize them outside the usual
\[ \sqrt{\varepsilon} G \sim N(0, \varepsilon) \]
\[ X \xrightarrow{+} Y = X + \sqrt{\varepsilon} G \]

de Bruijn: \[ \frac{d}{d\varepsilon} H(Y) = \frac{1}{2} J(Y) \]

\[ G \sim N(0, 1) \]
\[ \sqrt{s} X \xrightarrow{+} Y = \sqrt{s} X + G \]

Guo et al.: \[ \frac{d}{ds} I(X; Y) = \frac{1}{2} \text{MMSE}(X|Y) \]

Fig. 1. The Gaussian channel, where the input \( X \) is corrupted by a Gaussian noise \( G \). (a): In the de Bruijn’s approach, the variation of the entropy is characterized versus the noise variance \( \varepsilon \). (b): In the Guo’s approach, the noise variance is fixed and the pre-amplification \( \sqrt{s} \) of the input can vary: the variations of the mutual information is characterized versus the Minimal Mean-Square Error of the estimation of \( X \) using \( Y \).

Fig. 2. Classical “informational triangle” that schematically depicts the hugely used information measures (at the vertices), and the classical inequalities and identities that links these measures (at the edges). The Gaussian law is central since either the identities concerns the Gaussian channel, or the inequalities are saturated in the Gaussian context.

“Gauss-Shannon-Fisher” context.
In these relationships, the Gaussian play a central role since all the above-mentioned inequalities are saturated for Gaussian random variables, while the identities concern the Gaussian channel.

The de Bruijn’s identity is very important as, for instance, it was in the elements involved in the proof of the entropy power inequality [25], [1], [14], [2], and in the proof of the above-mentioned Stam’s inequality as well [25], [14], [2]. All these inequalities can also serve as a basis to prove the central limit theorem [26], [23], [27], [24].

Because the de Bruijn’s identity or its Guo’s version expresses the variations of the output entropy of the Gaussian channel (or mutual input-output information), it finds natural applications in communications. Indeed, as stressed
in [28] and the series of papers by the same team, the de Bruijn identity thus allows to assess the behavior of a
canal versus variation of the noise amplitude, and thus its robustness faced to noise. The divergence version is also
used to assess the behavior of such a channel subject to a mismatch between an assumed input and a true one [29],
[30]. This identity and some possible extensions showed also its importance through various applications, as for
instance given by Park et al. [31], [32], Brown et al. in [33] or Guo et al. in [34], among others. We can mention for
instance, the derivation of Cramér-Rao lower bounds from a Bayesian perspective (BCRLB) or from a frequentist
point of view, min-max optimal training sequences for channel estimation and synchronization in the presence of
unknown noise distribution, applications for turbo (iterative) decoding schemes, generalized EXIT charts and power
allocation in systems with parallel non-Gaussian noise channels, application in graph theory.

While Shannon entropy is widely used in communication, there is currently a re-emergence of the use of more
general entropic tools, in particular Rényi’s and Havrda-Charvát-Daróczy-Tsallis’s entropies [35], [36], [37]. These
generalized entropies find applications in various domains such as in statistical physics [37], [38], [39], [40],
[41], [42], in multifractal analysis [43] or in signal processing [2], [44], [45]. As the Kullback-Leibler divergence
quantifies the “distance” between a pdf relatively to another known as reference, other divergences can also quantify
such a distance, in particular that of the class of Csiszár (or Ali-Silvey) [46], [47] given later on in definition 2 and
denoted $D_\phi$. As previously mentioned, the generalization of such entropies, together with some generalizations of
the moments, gave rise later on to generalizations of the moment-entropy inequalities [15], [16], [17], [18], [19],
[20].

To generalize the Fisher information, one can imagine to start from the definition 2, eq. (8) given later on of the
$\phi$-divergences and to make a second order Taylor expansion of $D_\phi(p_\theta||p_{\theta_0})$ in $\theta = \theta_0$ as for the Kullback-Leibler
divergence. However, for Csiszár’s divergences sufficiently smooth, the curvature coincides again with the Fisher
information [48], showing the strength of this last quantity. This direction is thus not relevant to generalized the
Fisher information. Nevertheless, in spite of the fundamental character of this measure, following pioneer works
from Boeke or Vajda [49], [48], generalizations of the Fisher information began to appear. These extensions were
construct intrinsically from the Rényi’s entropies and then used to extend information-theoretic results on the “edges”
of the “informational triangle” of figure 2 such that the Cramér-Rao inequality [16], [18], [19], [21], [20] or the
Stam’s inequality [16], [18], [19], [20] in the Rényi context. Although not presented as a generalization of the
Fisher divergence, one can find precisely a quantity in [29] that appears as such a generalization. It came from of
a possible generalization of the de Bruijn’s identity in the scalar context. We will see later on that our proposed
generalizations of this identity in terms of divergence makes in fact appear the expression of [29] th. 15]. Both
generalizations of the informational measures gave rise to generalizations of their links, or were built to obtain such
generalizations.

Although many parts of the informational triangle of figure 2 were generalized, as far as we know, a few
generalizations of the de Bruijn’s identity were proposed. In [29], Guo proposed a version by extending their
previous version in terms of mutual information and MMSE in the scalar context, through Csiszár’s divergences. A
generalization of the Shannon mutual information–MMSE version in the non-Gaussian context was also proposed
by the same author [34]. Finally, one can mention a generalization of the identity for a law satisfying a nonlinear
heat equation [50], [51]. But in this non linear context, connecting the extended de Bruijn’s identity to a noisy
communication channel fails.

In our work, we are interested in answering the following questions. (i) What happens in terms of robustness
of the Gaussian channel if we use general divergences (or relative entropies) to characterize the system? (ii) Are
there equivalent results for more general channels rather than the Gaussian channel? The main result of the paper
is that the de Bruijn’s identity extends both to general divergences rather than the Kullback-Leibler one, and to
more general channels rather than the Gaussian one. In these cases, particular quantities appear which we name as
\(\phi\)-Fisher information and \(\phi\)-Fisher divergences and we will show that these extensions contain special cases, such
as the usual Fisher information, the \(\alpha\)-Fisher gain [52], or a recently defined Jensen-Fisher divergence [53]. As the
Rényi’s entropies showed its importance in various field of applications in particular in signal processing [2], [44],
[45], extending the de Bruijn identity in such a context, and far beyond this last one, open perspectives in these
applications in the light of the proposed extensions.

The known results and the extensions proposed here are summarized in the following table.

| Channel                | Shannon Fisher | \(\phi\)-entropies | \(\phi\)-Fisher |
|------------------------|----------------|--------------------|----------------|
| Gaussian channel       | Stam [1]       | Guo (scalar) [29]  |                |
|                        | Barron (scalar) [23] | Sec. 3             |                |
|                        | Johnson (scalar) [24] |                |                |
| Cauchy channel         | Johnson (scalar) [23] | Sec. 3 & 4         |                |
| Lévy channel           | Johnson (scalar) [24] | Sec. 3 (scalar)    |                |
| 2nd order PDE channels | Sec. 3 & 4     | Sec. 3 & 4         |                |

The paper is organized as follows. In section II the notation and assumptions used throughout the paper are shown.
Then, we will recall the definition of the \(\phi\)-entropies, \(\phi\)-divergences with their associated \(\phi\)-Fisher informations and
\(\phi\)-Fisher divergences, respectively. In section III we will reformulate the usual de Bruijn’s relation related to the
scalar Gaussian channel [1] in terms of the more general \(\phi\)-divergences due to Csiszár [46] or to Ali-Silvey [47] (see
also [54], [55]). In the same section we will go beyond the Gaussian noisy channel and consider noises characterized
by a more general pdf. We show two instances of the generalized version extending relations proposed by Johnson
for Cauchy or Lévy channels [24]. In section IV we will go a step further, proposing multivariate extensions in
which both the spatial coordinates and the noise parameter are vectors. We will then show in section IV that this
generalization encompasses both the multivariate de Bruijn’s identity [1], [2], [13], [14], the Guo’s one [56], as
well as other extensions due to Guo et al. [56], Palomar & Verdú [28] or Johnson [24], §5.3.

II. Definitions and notations

A. Notations and assumptions

Throughout the paper use the following notations and assumptions (except when specified or when additional assumptions are required):

- The probability laws are assumed to admit a density with respect to the Lebesgue measure.
- The probability density function (pdf) is denoted \( p \) when dealing with entropies, and \( p_1 \) and \( p_0 \) when dealing with divergences and are defined over sets \( \Omega \subseteq \mathbb{R}^d, \Omega_0 \subseteq \mathbb{R}^d \) and \( \Omega_1 \subseteq \mathbb{R}^d \) respectively, where \( d \in \mathbb{N}^* \) (multivariate context).
- The pdfs are supposed to be parametrized by a (common) vectorial parameter \( \theta \in \Theta \subseteq \mathbb{R}^n, n \in \mathbb{N}^* \).
- The “states” spaces \( \Omega, \Omega_0 \) and \( \Omega_1 \) are assumed to be independent of \( \theta \).
- We assume that \( \Omega_0 \subset \Omega_1 \), that is \( p_0(x) = 0 \Rightarrow p_1(x) = 0 \) (the probability measure attached to \( p_1 \) is absolutely continuous with respect to that attached to \( p_0 \)).
- Densities \( p \) and \( p_0 \) are assumed to vanish in the boundary of \( \Omega \) and \( \Omega_0 \), respectively.
- When necessary, densities \( p, p_0 \) and \( p_1 \) are assumed differentiable or twice differentiable with respect to \( \theta \) and/or with respect to \( x \).
- The notation \( \cdot^t \) denotes the transposition operation of a vector or a matrix, \( \text{Tr} \) is the trace operator and \( |\cdot| \) denotes the absolute value of the determinant of a matrix.
- The gradient or jacobian vs \( \theta \) of a function \( f: \Omega \rightarrow \mathbb{R}^k \) is defined as \( \nabla_\theta f = \left[ \frac{\partial f}{\partial \theta} \right]_{i,j} \) so that for \( n = 1 \), \( \nabla_\theta f = \frac{\partial f^t}{\partial \theta} \). The gradient or jacobian vs \( x \) is defined similarly via the partial derivative vs \( x \).
- The Hessian matrix vs \( \theta \) of function \( f: \Omega \rightarrow \mathbb{R} \) is defined as \( \mathcal{H}_\theta f = \left[ \frac{\partial^2 f}{\partial \theta_i \partial \theta_j} \right]_{i,j} \), so that for \( n = 1 \), \( \mathcal{H}_\theta f = \frac{\partial^2 f}{\partial \theta^2} \). The Hessian vs \( x \) is defined similarly via the second order partial derivative vs \( x \).
- The logarithm function will be denoted \( \log \), without specifying its base; the choice has no importance, provided the same one is considered for all the quantities that interplay.
- The entropic functional \( \phi: [0; +\infty) \rightarrow \mathbb{R} \) we will introduce in few lines needs to be convex. In the whole paper, we additionally assume that it is of class \( C^2 \), so that the convexity writes \( \phi'' \geq 0 \) (\( \cdot^t \) and \( \cdot'' \) denote the first and the second derivative, respectively).

B. Definitions

To extend the de Bruijn’s identity to generalized \( \phi \)-entropies and \( \phi \)-divergences, we need first to introduce these quantities, and the extensions of the Fisher information as well. What we will call \( \phi \)-Fisher information and \( \phi \)-Fisher divergences are quantities that appear naturally when \( \phi \)-entropies/divergences are used instead of the Shannon or Kulback-Leibler divergence to characterize the channels depicted in figure I (and extensions in the non-Gaussian context).
Let us start with the definition of the $\phi$-entropies and of the $\phi$-divergences of the Csiszár’s class \cite{46} (see also Salicrú \cite{57}):  

**Definition 1** ($\phi$-entropies). Here, we assume additionally\footnote{This condition is necessary (but not sufficient) to insure the convergence of the integral.} that $\phi(0) = 0$. The $\phi$-entropies of a pdf $p$ are then defined as  

$$H_\phi(p) = -\int_\Omega \phi(p(x)) \, dx.$$  

(7)

where $\phi$ is the so-called entropic functional.

Famous particular cases of $\phi$-entropies are, among many others \cite{6, 7, 57}:

- The Shannon entropy \cite{12}, given by $\phi(l) = l \log(l)$;
- The Havrda-Charvát \cite{35}, or Daróczy \cite{30} or Tsallis \cite{38} entropies, denoted in the sequel HCDT\footnote{It is worth to point out that the Rényi entropies, $R_\alpha(p) = \frac{1}{1-\alpha} \log \left( \int_\Omega p^\alpha(x) \, dx \right)$, is closely connected to the HCDT entropies $T_\alpha$ since $R_\alpha(p) = \frac{1}{1-\alpha} \log(1 - (1-\alpha)T_\alpha(p))$.} obtained for $\phi(l) = \frac{l^\alpha - l}{\alpha - 1}$, $\alpha > 0$ (one can even consider the situation $\alpha \leq 0$ when $\Omega$ is bounded);
- The Kaniadakis entropies \cite{58}, given by $\phi(l) = \frac{l^{1+\alpha} - l^{1-\alpha}}{2\alpha}$, $-1 < \alpha < 1$.

**Definition 2** ($\phi$-divergences (Csiszár \cite{46}, Ali-Silvey \cite{47})). The $\phi$-divergences between two pdfs $p_1$ and $p_0$, or relative $\phi$-entropies, relatively to pdf $p_0$, are defined as\footnote{One often finds a more general definition under the form $h\left(-\int_\Omega \phi\left(\frac{p_1(x)}{p_0(x)}\right) p_0(x) \, dx\right)$ where $h$ is an increasing function. We restrict here to $h = \text{Id}$ the identity, so that some usual $\phi$-divergences are a monotonous function of the divergences defined here. Note also that in \cite{46}, in the scalar context, the integration is over $\mathbb{R}$ using the convention $0 \phi(0/0) = 0$; moreover to avoid the restriction $\Omega_1 \subseteq \Omega_0$, Csiszár also imposes the convention $0 \phi(a/0) = a \lim_{a \to +\infty} \phi(u)/u$ \cite{60}. We let the reader.}

$$D_\phi(p_1 \parallel p_0) = \int_{\Omega_0} \phi\left(\frac{p_1(x)}{p_0(x)}\right) p_0(x) \, dx.$$  

(8)

Well-known cases of such divergences are the following \cite{6, 7, 57, 59}:

- The Kullback-Leibler divergence \cite{2, 46} given by $\phi(l) = l \log(l)$;
- The exponential of the so-called Rényi’s divergences and a linear function of the Hellinger’s divergences (or simply the Hellinger integral) \cite{2, 60, 46, 55} for $\phi(l) = l^\alpha$, $\alpha > 1$ (see also Tsallis \cite{38} or Havrda & Charvát \cite{35});
- The Jensen-Shannon divergence \cite{2, 46, 55}, for $\phi(l) = \frac{l}{2} \log l - \frac{1}{2} \log \frac{l+1}{2}$,
- Vajda divergences \cite{48, 55}, given by $\phi(l) = |l - 1|^\alpha$, $\alpha \geq 1$ (including the total variation divergence for $\alpha = 1$, and the Pearson divergence for $\alpha = 2$).

Such divergences have many common properties, and among them, assuming additionally\footnote{This is not a restriction since for any convex function $\tilde{\phi}$ defined on $\mathbb{R}^*_+$, function $\phi(x) = \tilde{\phi}(x) - \tilde{\phi}(1)x$ remains convex and $D_\phi(p_1 \parallel p_0) = D_{\tilde{\phi}}(p_1 \parallel p_0) - \tilde{\phi}(1)$ is only affected by a shift.} that $\phi(1) = 0$, from the Jensen inequality such $\phi$-divergences are nonnegative, and zero if and only if $p_1 = p_0$ (a.e.) \cite{46}. We let the reader
to references [6], [7], [55] for a brief panorama and for some applications of divergences in signal processing, physics and statistics.

Let us now turn to the generalization of the second information quantity appearing in the de Bruijn’s identity, namely the Fisher information.

**Definition 3** (ϕ-Fisher information matrix). We define the ϕ-Fisher information matrix of a pdf \( p \) relatively to a parameter \( \theta \) by

\[
J_\theta^{(\phi)}(p) = \int_{\Omega} \left[ \nabla_\theta \log p(x) \right] \left[ \nabla_\theta \log p(x) \right]^T \left[ \frac{p(x)}{x} \right]^2 \frac{\phi''(p(x))}{x} \, dx
\]

As an illustration we now show some particular cases that already exist in the literature.

- Obviously, in the Shannon context \( \phi(p) = p \log p \), so that \( p^2 \phi''(p) = p \); one recovers the usual Fisher information matrix \( J \).
- In the context of the HCDT entropies, \( \phi(p) = \frac{p^\alpha - p}{\alpha - 1} \) and thus \( p^2 \phi''(p) = \alpha p^\alpha \). It appears that the ϕ-Fisher information matrices of definition [3] coincide with the \( q \)-Fisher information matrices proposed recently by Johnson and Vignat [50, def. 3.2] (where their \( q \) and our \( \alpha \) are related by \( \alpha = 2q - 1 \) and up to a normalization coefficient) or with the \( (2, \lambda) \)-Fisher information matrices introduced by Lutwak et al. [18 eqs. (13)-(18)] (where their \( \lambda \) and our \( \alpha \) are related by \( \alpha = 2\lambda - 1 \) and up to a factor \( \alpha \); see also [16, eq. (7)] in the scalar context).

**Definition 4** (ϕ-Fisher divergence matrices). We define the ϕ-Fisher divergence matrices between two pdfs \( p_1 \) and \( p_0 \), relatively to parameter \( \theta \) and the reference pdf \( p_0 \) by

\[
J_\theta^{(\phi)}(p_1||p_0) = \int_{\Omega} \left[ \nabla_\theta \log \left( \frac{p_1(x)}{p_0(x)} \right) \right] \left[ \nabla_\theta \log \left( \frac{p_1(x)}{p_0(x)} \right) \right]^T \left[ \frac{p_1(x)}{p_0(x)} \right]^2 \phi'' \left( \frac{p_1(x)}{p_0(x)} \right) \, p_0(x) \, dx
\]

When \( \theta \) is a location parameter \( \nabla_\theta \equiv \nabla_x \) and the ϕ-Fisher information and Fisher divergence matrices reduce to the corresponding nonparametric ones, denoted \( J^{(\phi)}(p) \) and \( J^{(\phi)}(p_1||p_0) \), respectively.

Some particular cases of ϕ-Fisher divergences were proposed in the literature, in specific contexts, as follows:

- For the entropic function \( \phi \) of the Kullback-Leibler divergence, \( (p_1/p_0)^2 \phi''(p_1/p_0)p_0 = p_1 \) and thus the ϕ-Fisher divergence that corresponds to the same function \( \phi \) is the usual Fisher divergence [2], [24], [13].
- In the Rényi context, with the Rényi index \( \alpha \) (or HCDT), \( (p_1/p_0)^2 \phi''(p_1/p_0)p_0 \propto p_1^\alpha p_0^{1-\alpha} \) is a geometric mean of densities \( p_1 \) and \( p_0 \), leading, up to a normalization factor, to the \( \alpha \)-Fisher gain introduced by Hammad in [52].
- Note finally that for the Jensen-Shannon context, \( (p_1/p_0)^2 \phi''(p_1/p_0)p_0 \propto \frac{p_1 p_0}{p_1 + p_0} \) is an harmonic mean leading to a very recently defined Jensen-Fisher divergence \( J^{(JS)}(p_1||p_0) \) by Sánchez-Moreno et al. [53]. In [53], \( J^{(JS)}(p_1||p_0) \) was introduced by pure analogy with the Jensen-Shannon divergence under the form

\[
J^{(JS)}(p_1||p_0) = \frac{1}{2} J \left( p_0 \left| p_0 + p_1 \right| \frac{p_0 + p_1}{2} \right) + \frac{1}{2} J \left( p_1 \left| p_0 + p_1 \right| \frac{p_0 + p_1}{2} \right) = \frac{1}{2} J(p_0) + \frac{1}{2} J(p_1) - J \left( \frac{p_0 + p_1}{2} \right)
\]
and was used for physical description purposes.

Both these matrices are symmetric positive definite and vanish if and only if $p_1 = p_0$ a.e. Moreover, as for the usual Fisher divergence, one can also define the (scalar) $\phi$-Fisher divergences as the trace of the $\phi$-Fisher divergence matrices. Thus, obviously, the $\phi$-Fisher divergences are nonnegative.

Note that, as shown in [52, eq. (26)] for the usual Fisher matrices, both the $\phi$-divergences and the $\phi$-Fisher divergence matrices are invariant by the same biunivocal transformation of both $p_1$ and $p_0$.

As already evoked in the introduction, a generalization of the de Bruijn’s identity in the scalar context for the $\phi$-divergences and Gaussian channel\(^6\) has been made by Guo in [29] where, although no notion of Fisher information is explicitly mentioned, the derivative of the $\phi$-divergences with respect to the noise parameter is linked with the nonparametric $\phi$-Fisher information, this last quantity being expressed in terms of the difference of score functions $\nabla \log(p)$.

III. EXTENSION OF THE SCALAR DE BRUIJN’S IDENTITY TO $\phi$-ENTROPIES AND $\phi$-FISHER INFORMATIONS

In this section, we focus on the scalar context for both the state $x$ and the parameter $\theta$, i.e., $d = n = 1$. This restriction allows to increment by a first step the de Bruijn’s identity, while the general case (including the scalar one) will be the object of the next section. In this section, we will assume that the quantities that interplay (entropies, divergences, Fisher informations) exist. This assumption requires conditions on the pdfs and on the entropic functional $\phi$ that cannot be given in a general setting; they must be studied case by case.

Let us consider firstly the Gaussian channel as in the de Bruijn’s primal version, as done by Guo’s in some sense in its extension [29], before generalizing the result for a class of more general noises.

A. Gaussian noise

The key point of the de Bruijn’s identity for the Gaussian channel is that the pdf $p$ of the output follows the heat equation

$$\frac{\partial p}{\partial \theta} = \frac{1}{2} \frac{\partial^2 p}{\partial x^2}. \quad (11)$$

Reproducing the same steps than for the usual de Bruijn identity, writing the $\phi$-entropies of $p$, performing the derivative of this quantity once versus the parameter $\theta$, and using the heat equation, one obtains the following extension that we name $\phi$-de Bruijn’s identity,

\(^6\)More precisely, a general channel is considered, as in figure [1(a), where $\epsilon \to 0$. In this limit, the result lies on the heat equation followed by the output pdf, and the channel can be viewed as approximately Gaussian (in the second order and provided the noise has a finite variance).
Proposition 1 (φ-de Bruijn’s identity). Consider a pdf $p$ satisfying the heat equation (11), such that $\frac{\partial}{\partial \theta} \phi(p)$ is $\theta$-locally uniformly integrable\(^7\) and such that both $\phi(p)$ and $\frac{\partial}{\partial x} \phi(p)$ vanish in the boundary of $\Omega$. Then its $\phi$-entropy and $\phi$-Fisher information fulfill
\[
\frac{d}{d\theta} H_\phi(p) = \frac{1}{2} J^{(\phi)}(p). \tag{12}
\]

Proof: This case is a particular case of proposition 5, section IV, proved in appendix I-A.

In the Shannon context, one recovers the original de Bruijn’s identity [1].

Note that from the assumptions that $p$ vanishes in the boundary of $\Omega$ together with $\phi(0) = 0$, the vanishing assumption of $\phi(p)$ is indeed not a strong restriction. The other assumptions have to be studied case by case, given the explicit form of $p$ and $\phi$.

A particular situation of the general one depicted in this proposition, widely used in communication theory, occurs when considering the output for a Gaussian noisy channel, $Y = X + \sqrt{\theta} G$ since the pdf of the output satisfies the heat equation [26], [23] (see also appendix III-A). Clearly, in such a case, the regularity conditions stated in the propositions for pdf $p_Y$ imply conditions on the input pdf, depending on the entropic functional $\phi$. For instance, in the Shannon case there were shown to be true by Barron, provided that the input has a finite variance [26] Lemma 6.3. In the general $\phi$ context, the steps of Barron are more difficult to apply. However, it is shown in appendix III-B that $\phi(p_Y)$ vanishes in the boundary of the domain. Moreover, assuming that there exists some $k \in (0 ; 1)$ such that $u_k \phi'(u) \to 0$ when $u \to 0$, the vanishing property of $\frac{\partial}{\partial y} \phi(p_Y)$ in the boundary is also insured (see appendix III-B). This last condition on $\phi'$ is not very restrictive, applying for the entropies frequently used, such that the Shannon entropy, the HCDT entropy (provided $k > 1 - \alpha$) or the Kaniadakis entropy (provided that $k > \kappa$), in others.

As done for the Kullback-Leibler divergence in [24], this proposition can be recast in terms of $\phi$-divergences as follows.

Proposition 2 (φ-de Bruijn’s identity in terms of divergences). Let $p_0$ and $p_1$ parametrized by the same parameter $\theta$, both satisfying the heat equation (11), such that $\frac{\partial}{\partial \theta} \left[ p_0 \phi \left( \frac{p_1}{p_0} \right) \right]$ is $\theta$-locally uniformly integrable, and such that both $p_0 \phi \left( \frac{p_1}{p_0} \right)$ and $\nabla_x \left[ p_0 \phi \left( \frac{p_1}{p_0} \right) \right]$ vanish in the boundary of $\Omega$. Then, their $\phi$-divergences and $\phi$-Fisher divergences satisfy
\[
\frac{d}{d\theta} D_\phi(p_1 \parallel p_0) = -\frac{1}{2} J^{(\phi)}(p_1 \parallel p_0). \tag{13}
\]

Proof: This case is a particular case of proposition 6, section IV, proved in appendix I-B.

Again, a particular situation arises in the context of the Gaussian noisy channel. As previously mentioned the pdf of the output of this Gaussian channel satisfies the heat equation. For instance, in a mismatch context, considering

\(^7\)By this terminology, we express that for any compact $K \subset \Theta$, this partial derivative is integrable vs $x$ on $\Omega$, uniformly vs $\theta \in K$. This allows to interchange integration and derivation vs $\theta$ [61 § 63]. In practice, the sufficient condition that $\left| \frac{\partial}{\partial \theta} \phi(p) \right| \leq g$ for any $\theta \in K$ with $g$ integrable and independent of $\theta$ is often used, invoking thus the dominated convergence theorem together with the mean value theorem.
that $X_0$ is the assumed input of the channel, while the true input is $X_1$, and noting $p_0$ and $p_1$ the pdfs of the respective outputs, the $\phi$-divergence measures a kind of distance between the assumed output pdf $p_0$ (that serves as the reference) and the true one $p_1$. Hence, the $\phi$-Fisher information gives the variation of this mismatch measure with respect to the noise amplitude. Since $J(\phi) \geq 0$, the proposition states that the consequence of the mismatch decreases with $\theta$, the rate of decreasing being precisely given by this $\phi$-Fisher information. If $X_0 = 0$, the divergence measures the decrease of the distance to a Gaussian as $\theta$ increases, which is a key point used in some proofs of the central limit theorem when dealing with the Kullbach-Leibler divergence [23], [24], [27]. It has also been shown that in the limit $\theta \to 0$, the proposition apply for non Gaussian noises with finite variance and in the small amplitude noise limit $\theta \to 0$ since in this limit the pdf also satisfy the heat equation (and thus for the output pdf as well) [29]. As shown in this last reference, the $\phi$-Fisher information can be viewed as a mean-square distance between the outputs’ pdfs, but averaged over a “deformed” distribution instead of the reference one. Finally, anew in the Shannon context, one recovers the original de Bruijn’s identity formulation in terms of divergences of [23], [24] (there, the reference pdf, $p_0$, is a Gaussian of variance $\theta$ and $p_1$ as the output pdf). Let us finally mention that in the case of Jensen-Shannon divergence and in the scalar case, eq. (28) reduces to the Sanchez-Moreno et al. version of such de Bruijn’s identity [53, eq. (7)].

Note that once again, the conditions of the proposition are to be studied case by case according to the considered entropic functional $\phi$ and the pdfs of the inputs $X_0$ and $X_1$ as well.

B. Beyond the Gaussian noise: extension to more general scalar non-Gaussian channels

Here, we extend propositions 1 and 2 to a more general set up in which the channel noise is non-Gaussian. Indeed, although the most common noise in nature is of Gaussian type, there exists others whose probability distribution does not follow the heat equation but still have associated a partial differential equation (PDE) which, in turn, is the clincher to obtain de Bruijn-type identities. We will consider the general case as well as two particular cases of non-Gaussian noises, Lévy and Cauchy, whose corresponding PDE have a similar structure to that of the heat equation [24].

For both versions of the de Bruijn’s identity, the key point is that $p_0$ and $p_1$ follow the same second order PDE given by

$$
\alpha_1(\theta) \frac{\partial}{\partial \theta} p(x) + \alpha_2(\theta) \frac{\partial^2}{\partial x^2} p(x) = \frac{\partial}{\partial x} (\beta_1(x, \theta) p(x)) + \beta_2(\theta) \frac{\partial^2}{\partial x^2} p(x)
$$

(14)

Note that the PDE (14) reduces to a Fokker-Planck equation [62], when $\alpha_2 = 0$ and $\alpha_1 = 1$, where $-\beta_1$ is the drift and where $2\beta_2$ is the diffusion, that is state-independent in this case, the heat equation being a particular case ($\beta_1 = 0$ and with $\beta_2$ being constant).

Now, propositions 1 and 2 can be generalized one step further as follows:
Proposition 3 (Generalized scalar $\phi$-de Bruijn identity). Let a pdf $p$ satisfying the PDE (14) where the drift $\beta_1$ is state-independent ($\beta_1(x, \theta) = \beta_1(\theta)$), such that both $\frac{\partial}{\partial \theta} \phi(p)$ and $\frac{\partial^2}{\partial \theta^2} \phi(p)$ are $\theta$-locally uniformly integrable, and such that both $\phi(p)$ and $\frac{\partial}{\partial x} \phi(p)$ vanish in the boundary of $\Omega$. Then, the $\phi$-entropies and $\phi$-Fisher information of pdf $p$ satisfy the identity

$$
\alpha_1(\theta) \frac{d}{d\theta} H_\phi(p) + \alpha_2(\theta) \frac{d^2}{d\theta^2} H_\phi(p) = \beta_2(\theta) J_\phi(p) - \alpha_2(\theta) J_\phi^1(p).
$$

Proof: This is a particular case of proposition [5] section [IV] proved in appendix [I-A].

When $\alpha_2 = 0$, the $\theta$-local uniform integrability of $\frac{\partial^2}{\partial \theta^2} \phi(p)$ is unnecessary and, similarly, if $\beta_2 = 0$, no condition on $\frac{\partial}{\partial x} \phi(p)$ is required (see the proof of the proposition).

Condition $\phi(p) \to 0$ in the boundary of $\Omega$ is not restrictive due to the assumption $\phi(0) = 0$ and the vanishing assumption of $p$ in the boundary of $\Omega$. The other regularity conditions stated in the proposition imply conditions on the pdf $p$, depending on the entropic functional $\phi$, and must be studied case by case. To this end, one can follow the steps of Barron [26], [23] recalled and slightly extended in appendix [III-B] as a guidance.

As in the heat equation context, this proposition can again be recast in terms of divergences as follows:

Proposition 4 (Generalized scalar $\phi$-de Bruijn identity in terms of divergences). Let $p_0$ and $p_1$ two pdfs, with the same parameter $\theta$, both satisfying PDE (14) and such that both $\frac{\partial}{\partial \theta} \left[ p_0 \phi \left( \frac{p_1}{p_0} \right) \right]$ and $\frac{\partial^2}{\partial \theta^2} \left[ p_0 \phi \left( \frac{p_1}{p_0} \right) \right]$ are $\theta$-locally uniformly integrable, and such that both $\beta_1 p_0 \phi \left( \frac{p_1}{p_0} \right)$ and $\frac{\partial}{\partial x} \left[ p_0 \phi \left( \frac{p_1}{p_0} \right) \right]$ vanishes in the boundary of $\Omega$. Then, the $\phi$-divergences and $\phi$-Fisher divergences of pdf $p_1$ with respect to $p_0$ fulfill the relation

$$
\alpha_1(\theta) \frac{d}{d\theta} D_\phi(p_1||p_0) + \alpha_2(\theta) \frac{d^2}{d\theta^2} D_\phi(p_1||p_0) = \alpha_2(\theta) J_\phi^1(p_1||p_0) - \beta_2(\theta) J_\phi(p_1||p_0).
$$

Proof: This case is again a particular case of proposition [6] section [IV] proved in appendix [I-B].

As for the entropic version of the proposition, when $\alpha_2 = 0$, the $\theta$-local uniform integrability of $\frac{\partial}{\partial \theta} \left[ p_0 \phi \left( \frac{p_1}{p_0} \right) \right]$ is unnecessary, and similarly, if $\beta_2 = 0$, no condition on $\frac{\partial}{\partial x} \left[ p_0 \phi \left( \frac{p_1}{p_0} \right) \right]$ is required (see the proof of the proposition).

Note that here again, the conditions of the proposition are to be studied case by case according to the considered entropic functional $\phi$ and the pdfs of the inputs $p_0$ and $p_1$ as well.

It is first interesting to note that these identities apply again in the context of a noisy channel as in figure [4](a), but where the noise pdf satisfies PDE (14) in the context of state-independent $\beta_1$. Indeed, in this case, writing the output pdf as a convolution between the input and noise pdfs, and provided that the last is regular enough, one can show that the output pdf also satisfy PDE (14) (see the steps given in the Gaussian case, appendix [III-A]). Thus, these identities apply to the output of such a general channel, or to the output pdf relatively to the noise

\[8\]The first and second derivative vs $\theta$ and vs $x$ must be $\theta$-locally uniformly integrable and $x$-locally uniformly integrable respectively. Noting that the output pdf can be obtained as a convolution between the pdf of the input and that of the noise, a sufficient condition is that the partial and second order partial derivatives of the noise are (locally) uniformly bounded. Thus, the integrand are dominated by an integrable function proportional to the input pdf, with a coefficient independent of the parameter ($\theta$ or $x$).
pdf, provided the conditions required by the propositions are satisfied (these ones impose conditions on the input that can only be studied when the pdf of the noise is explicitly known). In other words, these results include and generalize the standard de Bruijn’s identity \[26\] as well as the Guo-Shamai-Verdu’s relations \[56\], \[29\] either to non-Gaussian noise, or to \(\phi\)-entropies and divergences, or both. They also include and generalize identities derived by Johnson in \[24\] for Cauchy and Lévy channels in the context of Kullback-Leibler divergence as we will show in few lines. The first version of the Guo-Shamai-Verdu’s identity is also recovered and extended by considering the output pdf and the pdf of the output conditionally to the input in the context of figure 1-(b).

**a) Lévy channel:** Consider again the channel of the form fig. 1 but now subject to Lévy noise with scale parameter \(\theta^2\), i.e., \(\theta^2 L\) where \(L\) is a standard Lévy r.v.. \(\theta^2 L\) has then the pdf

\[
p(x) = \frac{\theta \exp\left(-\frac{x^2}{2\theta^2}\right)}{\sqrt{2\pi \theta^2}}
\]

defined on \(\mathbb{R}_+\) \[63\]. One can easily see that both this pdf, and more specially, the pdf of the output \(Y\) satisfy the parabolic differential equation \[24\] (see also the steps used in the Gaussian case recalled appendix III-A)

\[
\frac{\partial^2}{\partial \theta^2} p(x) = 2 \frac{\partial}{\partial x} p(x)
\] (17)

As an immediate consequence of proposition 4, the \(\phi\)-entropy of the output and its \(\phi\)-Fisher information are linked by the relation

\[
\frac{d^2}{d\theta^2} H_\phi(p) = J^{(\phi)}_\theta(p)
\] (18)

Similarly for two output pdfs \(p_0\) and \(p_1\) of the Lévy channel (for instance when the input is respectively of Lévy and arbitrary), their \(\phi\)-divergences and \(\phi\)-Fisher divergences satisfy the relation

\[
\frac{d^2}{d\theta^2} D_\phi(p_1 \parallel p_0) = J^{(\phi)}_\theta(p_1 \parallel p_0)
\] (19)

The identity directly links the curvature of the \(\phi\)-entropies (resp. \(\phi\)-divergences) with the \(\phi\)-Fisher information (resp. \(\phi\)-Fisher divergences). For a Lévy distributed input vs an arbitrary input and in the context of Kullback-Leibler divergence, relation (19) is precisely that obtained by Johnson in \[24\, Th. 5.5\]. Again, to study some of the conditions required by the proposition, in the entropy context, one can follow the sketch appendix III-B.

**b) Cauchy channel:** Consider again the channel fig. 1 but now subject to Cauchy noise with scale parameter \(\theta\), i.e., \(\theta C\) where \(C\) is a standard Cauchy r.v.. \(\theta C\) has the pdf

\[
p(x) = \frac{\theta}{\pi(\theta^2+x^2)}
\]

so that both this pdf and, specially, the pdf of the output, satisfy the Laplace (elliptic differential) equation \[24\] (the very same steps used in the Gaussian case recalled appendix III-A allows to this conclusion)

\[
\frac{\partial^2}{\partial \theta^2} p(x) = -\frac{\partial^2}{\partial x^2} p(x).
\] (20)

Thus, as a consequence of proposition 3 the \(\phi\)-entropy and \(\phi\)-Fisher information of the output are linked by the relation

\[
\frac{d^2}{d\theta^2} H_\phi(p) = J^{(\phi)}_\theta(p) + J^{(\phi)}_\theta(p)
\] (21)

Similarly for two output pdfs \(p_0\) and \(p_1\) of the Cauchy channel (for instance when the input is respectively of Cauchy and arbitrary),

\[
\frac{d^2}{d\theta^2} D_\phi(p_1 \parallel p_0) = J^{(\phi)}_\theta(p_1 \parallel p_0) + J^{(\phi)}_\theta(p_1 \parallel p_0)
\] (22)
Note again that, now, the identity directly links the curvature of the $\phi$-entropies (resp. divergences) with the sum of the parametric and nonparametric $\phi$-Fisher informations (resp. divergences). Here again, for a Cauchy distributed input vs an arbitrary input, relation (22) reduces to that obtained by Johnson in [24, Th. 5.6]. Dealing with entropies, following the very same steps that in appendix III-B allows to conclude that the boundary conditions required by the proposition are satisfied.

IV. FROM THE SCALAR CASE TO THE MULTIDIMENSIONAL CONTEXT

In this section, we generalize the previous results to the general multivariate context, both for the state $x$ ($d \geq 1$) and parameter $\theta$ ($n \geq 1$). To this aim, as for the previous section, the approach relies on pdf satisfying a second order PDE with the same form than eq. (14). But since the gradient operators lead to vectors (or matrices in the context of Jacobian matrices) and the Hessian operators lead to matrices, one have to introduce operators in order to sum quantities with the same dimension.

More precisely, we consider pdf $p$, parametrized by a vector $\theta$ satisfying the following PDE,

$$
\mathcal{L}_1(\nabla_\theta p(x)) + \mathcal{L}_2(\mathcal{H}_\theta p(x)) = \mathcal{K}_1(\nabla_x [\beta_1(x,\theta) p(x)]) + \mathcal{K}_2(\mathcal{H}_x p(x))
$$

where $\mathcal{L}_i$ and $\mathcal{K}_i$ are linear operators acting on vectors or matrices, dependent on $\theta$ or not but independent on the state $x$,

- $\mathcal{L}_1 : \mathbb{R}^n \rightarrow \mathbb{R}^k$ and $\mathcal{L}_2 : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^k$,
- $\beta_1 : \mathbb{R}^d \times \mathbb{R}^n \rightarrow \mathbb{R}^l$,
- $\mathcal{K}_1 : \mathbb{R}^{d \times l} \rightarrow \mathbb{R}^k$ and $\mathcal{K}_2 : \mathbb{R}^{d \times d} \rightarrow \mathbb{R}^k$

for some $l \in \mathbb{N}$ and $k \in \mathbb{N}$. For instance, an operator $\mathcal{K}_i$ and/or $\mathcal{L}_i$ can be the trace operator, a right and/or left product by a matrix (possibly dependent of $\theta$), extraction of a subvector or of a submatrix, etc.

To get an idea on pdfs satisfying a PDE of the form eq. (23), consider the Gaussian pdf

$$
p(x) = \frac{1}{(2\pi)^{d/2}|R|^{1/2}} \exp \left( -\frac{1}{2\theta} x^t R^{-1} x \right)
$$

parametrized by the scalar $\theta$. Differentiating in $\theta$ on one hand, differentiating twice in $x$ on the other hand, and using the identity $\text{Tr}(uv^t) = u^t v$, one easily shows that $p$ satisfies the PDE

$$
\nabla_\theta p = \text{Tr} (R \mathcal{H}_x p)
$$

Here, $\mathcal{L}_2 = 0$, $\mathcal{L}_1 = I$ is the identity, $\mathcal{K}_1 = 0$ and $\mathcal{K}_2(M) = \text{Tr}(RM)$ for any $M \in \mathbb{R}^{d \times d}$.

Note that, in the particular context of a state-independent $\beta_1$, if the input noise pdf of a channel as in figure I satisfies a PDE of the form eq. (23), the pdf of the output satisfies the same PDE. Thus, in this case we are in situation to generalize the multivariate versions of the de Bruijn identities related to a noisy communication channel.

As for the scalar case, the gradient and Hessian have to be $\theta$- and $x$-locally uniformly integrable. Again, it is sufficient that these quantities are $\theta$ and $x$-locally uniformly bounded.
If $\theta$ is scalar, PDE (23) encompasses the multivariate Fokker-Planck equation with state-independent diffusion when $\mathcal{L}_1 = 0$, $\mathcal{L}_2 = \mathcal{I}$, $-\beta_1 : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ being the drift, $\mathcal{K}_1(M) = \text{Tr}(M)$ for any matrix $M \in \mathbb{R}^{d \times d}$, and, denoting $\mathbb{I}^i = [1 \ldots 1]$, $\mathcal{K}_2(M) = \frac{1}{2} \mathbb{I}^i D(\theta) \mathbb{I}^i$ for any matrix $M \in \mathbb{R}^{d \times d}$ and for a symmetric positive definite matrix $D : \mathbb{R} \to \mathbb{R}^{d \times d}$ being the diffusion tensor [62].

Finally, as mentioned in the previous propositions, the scalar PDE (14) is a particular example where $\mathcal{L}_i (i = 1, 2)$ reduces to multiplication by $\alpha_i(\theta)$, $\mathcal{K}_2$ reduces to a multiplication by $\beta_2(\theta)$ and $\mathcal{K}_1 = \mathcal{I}$.

In the multivariate context introduced here above, we can now generalize the de Bruijn’s identities in terms of $\phi$-entropies and $\phi$-divergences. From the generalizations, we will then exhibits three particular examples, recovering existing identities of the literature.

A. Multivariate general de Bruijn’s identities

**Proposition 5** (Generalized multivariate $\phi$-de Bruijn’s relation). Let $p$ be a pdf that fulfills the PDE (23) with $\beta_1$ state-independent ($\beta_1(x, \theta) = \beta_1(\theta)$), such that both $\nabla_\theta \phi(p)$ and $\mathcal{H}_\theta \phi(p)$ are $\theta$-locally uniformly integrable, and such that both $\phi(p)$ and $\nabla_x \phi(p)$ vanish in the boundary of $\Omega$. Then, its $\phi$-entropies and $\phi$-Fisher information matrices satisfy the relation

$$\mathcal{L}_1(\nabla_\theta \mathcal{H}_\phi(p)) + \mathcal{L}_2(\mathcal{H}_\theta \mathcal{H}_\phi(p)) = \mathcal{K}_2 \left( J^{(\phi)}(p) \right) - \mathcal{L}_2 \left( J^{(\phi)}_\theta(p) \right)$$

**Proof:** See Appendix I-A

Here again, when $\mathcal{L}_2 = 0$ the local integrability of $\mathcal{H}_\theta \phi(p)$ is unnecessary and similarly, if $\mathcal{K}_2 = 0$, the gradient $\nabla_x \phi(p)$ does not need to vanish in the boundary of $\Omega$ (see the proof of the proposition).

As for the previous scalar extensions of the de Bruijn’s identities, the proposition can be recast in terms of divergences as follows,

**Proposition 6** (Generalized multivariate $\phi$-de Bruijn’s relations in terms of divergences). Let two pdfs $p_1$ and $p_0$ be parametrized by a same vector $\theta$ and satisfying PDE (23), such that both $\nabla_\theta \left[ p_0 \phi \left( \frac{p_1}{p_0} \right) \right]$ and $\mathcal{H}_\theta \left[ p_0 \phi \left( \frac{p_1}{p_0} \right) \right]$ are locally integrable, and such that both $\beta_1 p_0 \phi \left( \frac{p_1}{p_0} \right)$ and $\nabla_x \left[ p_0 \phi \left( \frac{p_1}{p_0} \right) \right]$ vanish in the boundary of $\Omega$. Then, the $\phi$-divergences and $\phi$-Fisher divergence matrices satisfy the relation

$$\mathcal{L}_1(\nabla_\theta D_\phi(p_1||p_0)) + \mathcal{L}_2(\mathcal{H}_\theta D_\phi(p_1||p_0)) = \mathcal{L}_2 \left( J^{(\phi)}_\theta(p_1||p_0) \right) - \mathcal{K}_2 \left( J^{(\phi)}(p_1||p_0) \right)$$

**Proof:** See Appendix I-B

Once again, when $\mathcal{L}_2 = 0$ the local integrability of $\mathcal{H}_\theta \left[ p_0 \phi \left( \frac{p_1}{p_0} \right) \right]$ is unnecessary. Similarly, if $\mathcal{K}_2 = 0$, the gradient $\nabla_x \left[ p_0 \phi \left( \frac{p_1}{p_0} \right) \right]$ does not need to vanish in the boundary of $\Omega$ (see proof).
It appears that (26) looks somewhat similar to an extended version of the de Brujin’s identity proposed by Johnson & Vignat in the context of Rényi’s entropies [50, eq. (11)]. However, our extension cannot recover their version since (26) is based on densities satisfying the second order linear partial differential equation (14), while the version of [50] lies on a nonlinear extension of the heat equation (called $q$-heat equation, involving a so-called $q$-Fisher information) as mentioned in the introduction. But, as previously mentioned, the extension proposed in [50] cannot be related to channel as in figure 1 so easily. Indeed, if the noise satisfied the nonlinear differential equation leading to their extension, due to the nonlinear aspect, the output cannot satisfy this equation.

Note again that the conditions required to apply the last two propositions have to be studied case by case, given $p$, $p_0$, $p_1$ and $\phi$; again, such a study can be inspired by that of Barron [26], [23] recalled and slightly extended in appendix III-B.

B. Particular cases

1) Gaussian channel: Let us consider the Gaussian channel depicted in figure 1-(a), where the noise is now $\sqrt{\theta}G$, with $G$ a Gaussian vector with zero-mean and covariance matrix $R$. It is straightforward to show that both noise pdf and output pdf $p$ follows PDE (23) with $L_1 = 1$, $L_2 = 0$, $K_1 = 0$ and $K_2(M) = \frac{1}{2} \text{Tr}(RM)$, i.e., the multidimensional version of the heat equation given eq. (24). Then, one obtains the extended vector version of the de Brujin’s identity:

$$\frac{\partial}{\partial \theta} H_{\phi}(p) = \frac{1}{2} \text{Tr} \left( RJ^{(\phi)}(p) \right)$$

(27)

In the Shannon entropy context, the usual versions [28], [24] are obviously recovered, and of course, in the scalar case and for the Shannon entropy, the initial de Brujin’s identity, as presented by Stam in [1] is naturally recovered.

Moreover, the divergence version of de Brujin’s relation, also given in [24] in the scalar context and Shannon entropies, writes

$$\frac{\partial}{\partial \theta} D_{\phi}(p_1||p_0) = -\frac{1}{2} \text{Tr} \left( RJ^{(\phi)}(p_1||p_0) \right),$$

(28)

If $p_1$ is the pdf of the output of the Gaussian channel and $p_0$ the Gaussian pdf with the same covariance than the noise, this result can be interpreted as a convergence of the output to the Gaussian as $\theta$ increases since $J^{(\phi)} \geq 0$ and $R \geq 0$ implies a decrease of $D_{\phi}$. Hence, the $\phi$-Fisher divergence associated to the $\phi$-divergence (i.e., with the same entropic functional $\phi$) gives the speed of convergence. For that reason, it is not surprising that this relation was implied (in the scalar Shannon context), in some way, in a proof of the central limit theorem (see [24], [23], [27], [64], [65] or references cited in).

Note also that another way of thinking consists in considering two similar channels, with respective input $X_0$ and $X_1$, and of output respectively $Y_0$ and $Y_1$ with pdfs $p_0$ and $p_1$. Thus, instead of working with the output and the noise, one can wish to compare the two different outputs leading to a tendency of convergence of the two outputs’ pdfs as $\theta$ increases, with a convergence rate given by the corresponding $\phi$-Fisher divergence. Again, through this
point of view, the $\phi$-Fisher divergence allows to assess the behavior of the channel versus a mismatch between an assumed input and a true one.

As far as we know, these interpretations and the consequences in terms of central limit theorem will let open the question of the interpretation of the general de Bruijn’s relations (27) and (28).

As mentioned when dealing with the scalar context, following the steps of Barron [26], [23], we show in appendix III-A that the pdf of the output of the multivariate Gaussian channel also satisfies a multivariate heat equation and in appendix III-B that in this multivariate context the boundary conditions are also satisfied (under the weak assumption of the existence of a $k \in (0; 1)$ such that $u^k \phi'(u) \to 0$ when $u \to 0$, dealing with the second one).

2) Cauchy channel: Let us consider the Cauchy channel, where the channel noise is $\theta C$, where $C$ has the characteristic matrix $R$ and its associated density is given by

$$
p(x) = \frac{\Gamma\left(\frac{d+1}{2}\right)}{\pi^{\frac{d+1}{2}} |R|^\frac{1}{2}} \left(\theta^2 + x^t R^{-1} x\right)^{-\frac{d+1}{2}}
$$

which follows PDE (23) with $L_1 = 0$, $L_2 = I$, $K_1 = 0$ and $K_2 = -I$, i.e.,

$$\frac{\partial^2}{\partial \theta^2} p(x) = -\text{Tr} \left( R \mathcal{H}_x p(x) \right). \quad (29)$$

Again, following the very same steps than that of Barron [26], [23], recalled in appendix III-A allows to easily show that the pdf output of a multivariate Cauchy channel satisfies the same PDE than the Cauchy noise.

Thus, assuming that the pdfs satisfying eq. (29) also satisfy the condition required by proposition 5, the $\phi$-informational quantities of these pdfs satisfy the relation

$$\frac{d^2}{d \theta^2} \mathcal{H}_\phi(p) = \text{Tr} \left( R J^{(\phi)}(p) \right) + \text{Tr} \left( R J^{(\phi)}_\theta(p) \right). \quad (30)$$

Note that following the steps given in appendix III-B for the Gaussian channel, one can also easily show that in the multivariate Cauchy context, the boundary conditions of the proposition are also satisfied (under the same assumption for $u^k \phi'(u)$). Thus, for both the Cauchy pdf, or that of the output of a Cauchy channel.

Similarly, for two pdfs $p_0$ and $p_1$ satisfying eq. (29), for instance the pdf of the output of the Cauchy channel and a Cauchy distribution with the same characteristic matrix, or the pdfs of two different outputs (e.g., in the mismatch context) their $\phi$-divergences and $\phi$-Fisher divergences satisfy

$$\frac{d^2}{d \theta^2} D_\phi(p_1 \parallel p_0) = \text{Tr} \left( R J^{(\phi)}(p_1 \parallel p_0) \right) + \text{Tr} \left( R J^{(\phi)}_\theta(p_1 \parallel p_0) \right). \quad (31)$$

C. Extended Guo-Shamai-Verdu’s and extended Palomar-Verdu’s relations.

As we will see now, the Guo’s relation of [29] as well as the scalar and vectorial variations given in [56], [34], [28], [64] are particular cases of proposition 6.

First of all, one can notice that when parameter $\theta$ is matricial for instance, by a vectorization of this matrix, such a case can be treated through the formalism adopted in this section. Moreover, to conserve the structure of the quantities, the vectors or matrices that appear through the gradient or Hessian can be rearranged in tensors.
(“de-vectorization”). In this paragraph we only need to differentiate real-valued function of matricial argument
\[ M = [m_{ij}]_{i,j} \] for which \( \nabla_M f \) is the matrix of the partial derivatives, \( \nabla_M f = \left[ \frac{\partial f}{\partial m_{ij}} \right]_{i,j} \). Thus, we do need to introduce a complicate tensorial formalism. Note in particular that if \( f = \left[ M \right] \) (de-vectorization). In this paragraph we only need to differentiate real-valued function of matricial argument \( \theta \) is \( \nabla_{HB} \) multiplication is matricial, of the form \( e.g., \). Thus, we obtain \( \nabla_M f = [\nabla_x g] u^t \).

Now, we consider again the Gaussian channel of figure [1](b), but where the input \( X \) is a random vector and the multiplication is matricial, of the form \( H B X \), and where the noise \( N \) is Gaussian independent of \( X \), zero-mean and of covariance matrix \( R \), i.e., the output is \( Y = H B X + N \). In such a communication model, matrix \( H \) represents the transmission channel (filtering, etc.) while matrix \( B \) represents a pre-treatment of the data before sending them to the channel (e.g., beamforming), or can model the covariance matrix of the input (e.g., that would be \( B B^t \)). \( H \) and \( B \) are matrices that can be rectangular.

In [28], the authors are interested in the relationship between \( \nabla_H I(X;Y) \) or \( \nabla_B I(X;Y) \) and the MMSE matrix. Such relations allows to study the robustness of the information transmission vs the channel or vs the pre-treatment. These results can be recovered and extended thanks to the following proposition:

**Proposition 7.** Let us consider a multivariate Gaussian channel of the same form than figure [1](b), of input \( X \) put in form by the multiplication with a matrix \( \theta \), and corrupted by an independent channel noise \( N \), zero-mean, of covariance matrix \( R \) that does not depend of \( \theta \), i.e., \( Y = \theta X + N \). Assume that \( \nabla_{\theta} \left( \text{p} \right) \phi \left( \frac{p_{Y|X=x}}{p_Y} \right) \) is \( \theta \)-locally uniformly integrable, that both \( \phi' \left( \frac{p_{Y|X=x}}{p_Y} \right) \) and \( \phi'' \left( \frac{p_{Y|X=x}}{p_Y} \right) \) vanish in the boundary of \( \Omega \) and that \( \nabla_{\theta} \left( \text{D} \phi \left( \text{p}_{Y|X=x} \text{p}_Y \right) \right) \phi \left( \frac{p_{Y|X=x}}{p_Y} \right) \) is \( \theta \)-locally uniformly integrable. Thus, the generalized \( \phi \)-mutual information \( \text{D}_{\phi} (p_{X,Y} | \text{p}_X \text{p}_Y) \) satisfies the relation

\[
(\nabla_{\theta} \text{D}_{\phi} (p_{X,Y} | \text{p}_X \text{p}_Y)) \theta^t = R^{-1} \theta \text{MSE}_{\phi}(X|Y) \theta^t
\]  

with

\[
\text{MSE}_{\phi}(X|Y) = \int_{\Omega^2} (x - \mathbb{E}[X|Y = y])^2 \left( \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)} \right)^2 \phi'' \left( \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)} \right) p_X(x) p_Y(y) \, dx \, dy
\]

Again, \( \text{MSE}_{\phi}(X|Y) \) can be interpreted as a generalized \( \phi \)-mean-square error matrix, as defined in [23] for the classical case.

**Proof:** The proof is detailed in appendix [I]. It lies on the fact that the conditional pdfs \( p_{Y|X=x} \) and \( p_Y \) satisfy the same PDE (23). The result is thus almost a direct consequence of proposition 6.

Now, for \( \theta = HB \) together with the fact that for any scalar function \( g(H) \) it holds that \( \nabla_H g = \nabla_{\theta} g B^t \), we derive the relation

\[
\nabla_H \text{D}_{\phi} (p_{X,Y} | \text{p}_X \text{p}_Y) H^t = R^{-1} H B \text{MSE}_{\phi}(X|Y) B^t H^t
\]

that is nothing but [28 eq. (21)] up to the right multiplication by \( H^t \). Similarly, from \( \nabla_B g = H^t \nabla_{\theta} g \) we obtain

\[
\nabla_B \text{D}_{\phi} (p_{X,Y} | \text{p}_X \text{p}_Y) B^t H^t = H^t R^{-1} H B \text{MSE}_{\phi}(X|Y) B^t H^t
\]

that is nothing but [28 eq. (22)] up to the right multiplication by \( H^t \).
It is left as a future investigation the study of the simplification of Eq. (32) in the multivariate case since in this case it is not feasible to simply eliminate $\theta^t$ from both sides of this equation, given that $\theta$ is a matrix. Just in the case in which $\theta$ is tall (including square) and has full rank, it is possible to multiply both sides of Eq. (32) by $\theta$ and then (as the resulting square matrix is invertible) to simplify each side by $(\theta^t\theta)^{-1}$. This is for instance always true in the scalar case.

Note that in the scalar context of figure [1] the result of Guo [29, Th. 3] is thus recovered in the Shannon case and extended to $\phi$-divergences noting that for $\theta = \sqrt{s}$ and noting that, $\frac{\partial}{\partial s} = \frac{\partial}{\partial \theta} \frac{\partial}{\partial s} = \frac{1}{2\sqrt{s}} \frac{\partial}{\partial \theta}$.

V. CONCLUSIONS

In this paper we have proposed multidimensional generalizations of the standard de Bruijn’s identity obtained via the so-called $\phi$-entropies and divergences of Csiszárho (also Salicrú) class, within a scalar and vectorial framework. We first showed that, in the scalar case and for the Gaussian noisy channel, the derivative of the $\phi$-entropy (divergence) can be written as a generalized version of the Fisher information (divergence) and that these relations can be considered as a natural extension of the classical de Bruijn’s relation where both the noise and output pdfs follow the heat equation. Then, we have proposed a further step by considering non-Gaussian noises of pdfs governed by more general linear second-order PDE than the heat equation, both in the scalar and multivariate context (for both the state and the parameter). We thus obtained extended versions of the de Bruijn’s identity as well as extensions of the Guo-Shamai-Verdú’s relation that link the gradient of the $\phi$-mutual information to the generalized $\phi$-mean-square error.

The physical interpretation of the extended de Bruijn’s identities remains open as well as their potential implications and applications. Nevertheless, we believe that the extensions shown can broaden the perspective on the usual applications in statistics, estimation, communication theory or signal processing in a wider sense [2], [44], [45], [31], [32], [33], [34], [29], [30], [27], [28].

ACKNOWLEDGEMENTS

I. V. Toranzo acknowledges the support of the Spanish Ministerio de Educación under the program FPU 2014 and the BioTic International Mobility Program for PhD students. This work has also been partially supported by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01) funded by the French program Investissement d’avenir.

APPENDIX I

PROOF OF THE GENERALIZED MULTIVARIATE $\phi$-DE BRUIJN’S IDENTITIES.

In the proof of the propositions, we will very often use the divergence theorem, under various forms. In order to have it in mind, we recall it here:

**Lemma** (Divergence theorem). Let $\Omega \subset \mathbb{R}^d$ be a region in space and $\partial \Omega$ its boundary. Consider a vector field $f : \Omega \rightarrow \mathbb{R}^d$. Then, the volume integral of the divergence, $\text{div} f = \sum_i \frac{\partial f_i}{\partial x_i}$ over $\Omega$ is related to the surface integral
of $f$ over the boundary $\partial \Omega$ (assumed piecewise smooth) through
\[
\int_{\Omega} \text{div} f \, d\omega = \int_{\partial \Omega} f^t n \, ds,
\] (36)
where $f = [f_1 \cdots f_d]^t$ and where $n : \partial \Omega \rightarrow \mathbb{R}^d$ is the normal vector to the surface $\partial \Omega$.

When applied to $f = c\psi$ where $c$ is an arbitrary constant $d$-dimensional vector and $\psi : \Omega \rightarrow \mathbb{R}$, it leads to
\[
\int_{\Omega} \nabla \psi \, d\omega = \int_{\partial \Omega} \psi n \, ds
\] (37)

Finally, applying this last equation to any component of $\nabla \psi$ one obtains
\[
\int_{\Omega} \mathcal{H} \psi \, d\omega = \int_{\partial \Omega} \nabla \psi n^t \, ds
\] (38)

A. Formulation in terms of the $\phi$-entropies

Remind first that here $\beta_1(x, \theta) = \beta_1(\theta)$ so that this function, being independent on $x$, can be inserted in operator $\mathcal{K}_1$ (dependent only on $\theta$). In other words, without loss of generality, we can consider here that $\beta_1 = 1$.

We start from definition 1, eq. (7), of the $\phi$-entropies and perform the derivative respect to parameter $\theta$,
\[
\nabla \theta H_\phi = -\nabla \theta \int_{\Omega} \phi(p) \, dx = -\int_{\Omega} \nabla \theta [\phi(p)] \, dx
\]
where the argument $x$ of the functions are omitted for readability purposes, as well as the argument of the $\phi$-entropies. The interchange between the derivative and the integral follows from the $\theta$-local uniform integrability assumption \[61\, \S \, 63\]. This gives the expression
\[
\nabla \theta H_\phi = -\int_{\Omega} (\nabla \theta p) \phi'(p) \, dx
\] (39)

Differentiating again versus $\theta$, using again the $\theta$-local uniform integrability assumption to differentiate under the integral, we obtain
\[
\mathcal{H}_\theta H_\phi = -\int_{\Omega} \left[ (\mathcal{H}_\theta p) \phi'(p) + (\nabla \theta p)(\nabla \theta p)^t \phi''(p) \right] \, dx
\]
that is, from definition 3, eq. (9), of the $\phi$-Fisher information,
\[
\mathcal{H}_\theta H_\phi = -\int_{\Omega} (\mathcal{H}_\theta p) \phi'(p) \, dx - J^{(\phi)}_\theta
\] (40)
($J^{(\phi)}_\theta$ is supposed to exist, thus the integral of the sum can be separated as the sum of the integrals).

Then, we use successively the linearity of the operators \[10\] $\mathcal{L}_i$, the PDE (23) satisfied by $p$, relation $\mathcal{H}_x \phi(p) \phi'(p) = \mathcal{H}_x \phi(p) - (\nabla_x p)(\nabla_x p)^t \phi''(p)$, the linearity of operators $\mathcal{K}_i$, together with definition 3 of the nonparametric $\phi$-fisher

\[10\]The linear operators acting on matrices, elements of a finite dimensional Hilbert spaces, can be written as finite linear combination of the elements of the matrices, and thus, due to the finiteness, can be permuted with the integration (see for instance \[66\]).
information to obtain,
\[
\mathcal{L}_1(\nabla_\theta H_\phi) + \mathcal{L}_2(\mathcal{H}_\theta H_\phi) = -\int_\Omega \left[ \mathcal{L}_1(\nabla_\theta p) + \mathcal{L}_2(\mathcal{H}_\theta p) \right] \phi'(p) \, dx - \mathcal{L}_2 \left( J_\theta^{(\phi)} \right)
\]

Because \(\phi(p)\) vanishes on the boundary of \(\Omega\), from the formulation \((37)\) of the divergence theorem, we have
\[
\int_\Omega \nabla_x [\phi(p)] \, dx = 0
\]

Thus, from linearity of \(\mathcal{K}_1\), since necessarily \(\mathcal{K}_1(0) = 0\), we can conclude that
\[
\int_\Omega \mathcal{K}_1(\nabla_x \phi(p)) \, dx = 0
\]

Similarly, as \(\nabla_x [\phi(p)]\) vanishes on the boundary of \(\Omega\), one obtains from formulation \((38)\) of the divergence theorem together with the linearity of \(\mathcal{K}_2\) that
\[
\int_\Omega \mathcal{K}_2(\mathcal{H}_x \phi(p)) \, dx = 0
\]

which finishes the proof.

**B. Formulation in terms of \(\phi\)-divergences**

First of all, let us mention the following useful expression that we will often utilize in the sequel to make shorter the algebra,
\[
\nabla \left( \frac{p_1}{p_0} \right) = \frac{1}{p_0} \left( \nabla p_1 - \frac{p_1}{p_0} \nabla p_0 \right) = \frac{p_1}{p_0} \nabla \log \left( \frac{p_1}{p_0} \right) \tag{41}
\]

where the derivative \(\nabla\) can be either vs \(x\), or vs \(\theta\).

We start now the proof by first computing the derivative of the \(\phi\)-divergences given in definition \((2)\) eq. \((8)\) with respect to \(\theta\). Using the \(\theta\)-local uniform integrability assumption allowing to interchange the derivative with the integral, we obtain the expression
\[
\nabla_\theta D_\phi = \int_\Omega \left[ (\nabla_\theta p_0) \phi \left( \frac{p_1}{p_0} \right) + \left( \nabla_\theta p_1 - \frac{p_1}{p_0} \nabla_\theta p_0 \right) \phi' \left( \frac{p_1}{p_0} \right) \right] \, dx. \tag{42}
\]

Similarly, expressing the Hessian of \(p_0 \phi(p_1/p_0)\) starting from its gradient, using the \(\theta\)-local uniform integrability of the gradient of \(p_0 \phi(p_1/p_0)\) to interchange the differentiation in \(\theta\) and the integral in \((42)\), using relation \((41)\) to simplify the notation and the definition \((4)\) eq. \((9)\), we obtain the Hessian
\[
\mathcal{H}_\theta D_\phi = \int_\Omega \left[ (\mathcal{H}_\theta p_0) \phi \left( \frac{p_1}{p_0} \right) + \left( \mathcal{H}_\theta p_1 - \frac{p_1}{p_0} \mathcal{H}_\theta p_0 \right) \phi' \left( \frac{p_1}{p_0} \right) \right] \, dx + J_\theta^{(\phi)}. \tag{43}
\]

Then, (i) one combines eqs. \((42)\) and \((43)\), (ii) one uses the linearity of operators \(\mathcal{L}_i\) to interchange them with the integrations, (iii) one uses PDE \((23)\) satisfied by both \(p_0\) and \(p_1\), (iv) one observes that
\[
\nabla_x [\beta_1 p_1] \phi \left( \frac{p_1}{p_0} \right) + \left( \nabla_x [\beta_1 p_1] - \frac{p_1}{p_0} \nabla_x [\beta_1 p_0] \right) \phi' \left( \frac{p_1}{p_0} \right) = \nabla_x \left[ \beta_1 p_0 \phi \left( \frac{p_1}{p_0} \right) \right]
\]
and (v) that,
\[ \mathcal{H}_x p_0 \left( \frac{p_1}{p_0} \right) + \left( \mathcal{H}_x p_1 - \frac{p_1}{p_0} \mathcal{H}_x p_0 \right) \phi' \left( \frac{p_1}{p_0} \right) = \mathcal{H}_x \left[ p_0 \phi' \left( \frac{p_1}{p_0} \right) - \frac{p_1^2}{p_0^2} \left[ \nabla_x \log \left( \frac{p_1}{p_0} \right) \right] \right] \left[ \nabla_x \log \left( \frac{p_1}{p_0} \right) \right] \phi'' \left( \frac{p_1}{p_0} \right) \]
and (vi) definition \( \mathcal{H} \) eq. (4) of the nonparametric \( \phi \)-Fisher matrix to obtain
\[ \mathcal{L}_1 (\nabla \theta D_\phi) + \mathcal{L}_2 (\nabla \theta D_\phi) = \int_{\Omega} \left[ K_1 \left( \nabla_x \left[ \beta_1 p_0 \phi' \left( \frac{p_1}{p_0} \right) \right] \right) + K_2 \left( \mathcal{H}_x \left[ p_0 \phi' \left( \frac{p_1}{p_0} \right) \right] \right) \right] dx - \mathcal{K}_2 \left( J^{(\phi)} \right) + \mathcal{L}_2 \left( J^{(\phi)} \right) \]

Again, the vanishing property of \( \beta_1 p_0 \phi' \left( \frac{p_1}{p_0} \right) \) on the boundary of \( \Omega \) together with expression (37) of the divergence theorem, and the vanishing assumption of \( \nabla_x \left[ p_0 \phi' \left( \frac{p_1}{p_0} \right) \right] \) on the boundary, together with expression (38) of the divergence theorem, allow to see that the remaining integral term is zero, thus finishing the proof.

Appendix II

Proof the Generalized Multivariate Guo’s Identities.

First of all, note that if \( f(M) = g(x) \) with \( x = Mu \), from \( \frac{\partial f}{\partial m_{ij}} = \sum_k \frac{\partial g}{\partial x_k} \frac{\partial x_k}{\partial m_{ij}} \) we obtain \( \nabla_M f = [\nabla_x g] u^t \).

Now, \( N \) being independent of \( X \) we have \( p_{Y|X=x}(y) = p_N(y - \theta x) \) and from the expression of this Gaussian law together with the fact that \( R \) is not parametrized by \( \theta \), one easily shows that \( p_{Y|X=x} \) satisfies the PDE
\[ (\nabla_\theta p_{Y|X=x}) \theta^t = -\left( \nabla_y [y p_{Y|X=x}] \right)^t - \left( \mathcal{H}_y p_{Y|X=x} \right) R. \] \( \tag{44} \)

Note now that \( p_Y(y) = \int_{\Omega} p_{Y|X=x}(y) p_X(x) dx \). Following the same steps that in [26, Lemma 6.1], detailed in appendix [III-A] one shows that the gradient and the Hessian vs \( y \) and the integration in \( x \) can be interchanged and that, provided that \( X \) admits a second order moment \( [\nabla^2 \int x y^{t-1} p_X(x) dx ] = [\nabla^2 \int x y^{t-1} p_X(x) dx ] \) integration vs \( x \) and gradient in \( \theta \) can also be interchanged. Thus, multiplying eq. (44) by \( p_X(x) \) and integrating over \( x \) allows to show that \( p_Y \) satisfies this same PDE (44).

This PDE is of the form (23) with \( \mathcal{L}_2 = 0, \mathcal{L}_1(M) = M \theta^t, \beta_1(y, \theta) = y, \mathcal{K}_1(v) = -v^t \) and \( \mathcal{K}_2(M) = -MR \). Hence, immediately from proposition [5] we obtain
\[ \left[ \nabla_\theta D_\phi(p_{Y|X=x}||p_Y) \right] \theta^t = J^{(\phi)}(p_{Y|X=x}||p_Y) R \]
\( \tag{45} \)

Now, noting that \( \nabla_y p_{Y|X=x}(y) = -R^{-1}(y - \theta x) p_{Y|X=x}(y) \), one can also deduce that
\[ \nabla_y p_Y(y) = -\int_{\Omega} R^{-1}(y - \theta x) p_{Y|X=x}(y) p_X(x) dx = -R^{-1}(y - \theta E[X|Y = y]) p_Y(y) \]
leading to the following expression for the difference of the score functions
\[ \nabla_y \left[ \log \left( \frac{p_{Y|X=x}(y)}{p_Y(y)} \right) \right] = R^{-1} \theta (x - E[X|Y = y]) \]
and thus
\[ J^{(\phi)}(p_{Y|X=x}||p_Y) R = R^{-1} \theta \left( \int_{(x - E[X|Y = y])} (x - E[X|Y = y])^t \left( \frac{p_{Y|X=x}(y)}{p_Y(y)} \right)^2 \phi'' \left( \frac{p_{Y|X=x}(y)}{p_Y(y)} \right) p_Y(y) dy \right) \theta^t \]
The proof of proposition [7] finishes plugging this expression in (45), again noting that \( \frac{p_{Y|X=x}}{p_Y} = \frac{p_{X,Y}}{p_X p_Y} \) so that \( D_\phi(p_{X,Y}||p_{X}p_Y) = \int_{\Omega} D_\phi(p_{Y|X=x}||p_Y) p_X(x) dx \), multiplying both sides of (45) by \( p_X(x) \) and integrating over \( \Omega \).

\[ \text{This requirement is not necessary in the scalar context.} \]
APPENDIX III

VARIOUS ELEMENTS ON THE CONDITIONS NEEDED IN SOME OF THE PROPOSITIONS

The conditions set in the propositions are used to interchange derivation with respect to a parameter and integration, thanks to the dominated convergence theorem and the mean value theorem. Considering that the pdf and the considered entropic functionals are sufficiently regular, these conditions are probably not very restrictive.

In what follows, we give some elements dealing with the Gaussian channel, that can serve as a guidance for more general situations.

A. The pdf of the output of the Gaussian channel follows the heat equation.

It is shown for instance in [26], [23] in the scalar case that pdf of the output $Y = X + \sqrt{\theta} N$ of the Gaussian channel follows the heat equation. The same approach naturally applies when $N$ is a multivariate Gaussian with covariance matrix $R$ [24]. We recall here the main steps in the multivariate context, the scalar one being a particular case. The principle consists in writing the output pdf as a convolution, to derive it versus the parameter or the state, to interchange derivation and integrals, and thus to use the heat equation followed by the gaussian pdf (channel noise pdf). The very same steps are used in the multivariate Cauchy and can clearly serves as a basis to treat the generalized case.

In what follows, to simplify the notations, we write

$$\alpha = (2\pi)^{-\frac{d}{2}} |R|^{-\frac{1}{2}}$$

and

$$u = (y-x)^t R^{-1} (y-x) \geq 0$$

so that the pdf of the noise writes

$$p_N(y-x) = \alpha \theta^{-\frac{d}{2}} e^{\frac{u^2}{2\theta}}.$$  

We need then to be able to interchange derivation vs $\theta$ and the integration that gives the output pdf, and similarly for integration and Hessian vs $y$.

1) $\frac{\partial}{\partial \theta} \int_{\Omega} p_X(x) p_N(y-x) \, dx = \int_{\Omega} p_X(x) \frac{\partial}{\partial \theta} p_N(y-x) \, dx$: A direct calculus gives

$$\frac{\partial}{\partial \theta} p_N(y-x) = \frac{\alpha}{2} \theta^{-\frac{d}{2}-2} (u - \theta d) e^{\frac{-u^2}{2\theta}}.$$ 

A short study of this function versus $u \geq 0$ allows to prove that

$$\left| \frac{\partial}{\partial \theta} p_N(y-x) p_X(x) \right| \leq \frac{\alpha d}{2 \theta^{\frac{d}{2}+1}} p_X(x)$$

which is integrable. In other words $\frac{\partial}{\partial \theta} p_N(y-x) p_X(x)$ is $\theta$-locally dominated in $\mathbb{R}_+^*$ by an integrable function, which allows to conclude thanks to the dominated convergence theorem.

2) $\nabla_y \int_{\Omega} p_X(x) p_N(y-x) \, dx = \int_{\Omega} p_X(x) \nabla_y p_N(y-x) \, dx$: Direct algebra leads to

$$\left\| R^{\frac{1}{2}} \nabla_y p_N(y-x) \right\|^2 = \alpha^2 \theta^{-d-2} u e^{\frac{-u^2}{\theta}}.$$ 

where $R^{\frac{1}{2}}$ is the (unique) symmetric definite positive matrix, square root of $R$. Studying this function versus $u$, it is straightforward to show that

$$\left\| R^{\frac{1}{2}} \nabla_y p_N(y-x) \right\| \leq \alpha e^{-\frac{d}{2}} \theta^{-\frac{d+1}{2}}.$$
Thus, since $\nabla_y p_N(y - x) = R^{-\frac{1}{2}} R^{\frac{1}{2}} \nabla_x p_N(y - x)$ and the definition of the matrix 2-norm \[67\]^{12}

$$\| \nabla_y p_N(y - x)p_X(x) \| \leq \frac{\| R^{-\frac{1}{2}} \|_F^2}{\sqrt{\theta e (2\pi \theta)^{\frac{1}{2}} |R|^{\frac{1}{2}}}} p_X(x)$$

which is integrable. Again, $\nabla_y p_X(x)p_N(y - x)$ is dominated by an integrable function, allowing integration and derivation interchange.

3) $\mathcal{H}_y \int y \mathcal{H}_y p_N(y - x) dx = \int y \mathcal{H}_y p_N(y - x) dx$: Immediately from the pdf $p_N$ one has

$$R^{\frac{1}{2}} \mathcal{H}_y p_N(y - x) R^{\frac{1}{2}} = \alpha \theta^{-\frac{d}{2} - 2} \exp \left( \frac{-u}{\theta^2} \right) \left[ \exp \left( \frac{-u + R^{-\frac{1}{2}}(y - x)(y - x)^t R^{-\frac{1}{2}}}{\theta^2} \right) \right].$$

Multiplying this expression by its transposition and taking the trace to obtain its Frobenius norm $\| \cdot \|_F$ \[67\], one has thus

$$\left\| R^{\frac{1}{2}} \mathcal{H}_y p_N(y - x) R^{\frac{1}{2}} \right\|_F^2 = \alpha^{-d} \theta^{-\frac{d}{4}} \left( \theta u^2 - 2 \theta u + d \theta^2 \right) \exp \left( \frac{-u}{\theta^2} \right).$$

A short study of this function vs $u \geq 0$ shows that,

$$\left\| R^{\frac{1}{2}} \mathcal{H}_y p_N(y - x) R^{\frac{1}{2}} \right\|_F \leq \alpha \theta^{-\frac{d}{4}} \theta^{-\frac{d}{4} - 1}.$$

From \[67\] p. 279 stating that $\| AB \|_F \leq \| A \|_F \| B \|_F$, one obtains from $\mathcal{H}_y p_N(y - x) = R^{-\frac{1}{2}} \left( R^{\frac{1}{2}} \mathcal{H}_y p_N(y - x) R^{\frac{1}{2}} \right) R^{-\frac{1}{2}}$

$$\| \mathcal{H}_y p_X(x)p_N(y - x) \|_F \leq \frac{\sqrt{\theta}}{\sqrt{\theta} (2\pi \theta)^{\frac{1}{2}} |R|^{\frac{1}{2}}} \| p_X(x) \|,$$

which implies that $\mathcal{H}_y p_X(x)p_N(y - x) = \nabla_y \left( \nabla_y p_X(x)p_N(y - x) \right)$ is dominated by an integrable function, allowing again to finish the proof.

**B. Boundary conditions.**

1) $\phi(p_Y)$ vanishes in the boundary of $\Omega_Y$: Note first that $\Omega_Y = \mathbb{R}^d$ since $p_N > 0$ so that $p_X(x)p_N(y - x) \geq 0$ cannot be identically zero. Furthermore,

$$|p_X(x)p_N(y - x)| \leq p_X(x) \sup_{y \in \mathbb{R}^d} p_N(y) = \frac{1}{(2\pi \theta)^{\frac{1}{2}} |R|^{\frac{1}{2}}} p_X(x).$$

Hence, since $p_X(x)p_N(y - x)$ is dominated by an integrable function, one can evoke the dominate convergence theorem to conclude that

$$\lim_{\|y\| \to \infty} p_Y(y) \left( \int \Omega p_X(x) \lim_{\|y\| \to \infty} p_N(y - x) dx \right) = 0$$

i.e., $p_Y$ vanishes in the boundary of $\Omega_Y$. Together with $\phi(0) = 0$, $\phi(p_Y)$ vanishes in the boundary of $\Omega_Y$.

\[12\] Note that from $\| \cdot \|_2 \leq \| \cdot \|_F$ \[67\], one can replace the 2-norm by the Frobenius norm in the inequality.
2) $\nabla_y [\phi(p_Y)]$ vanishing in the boundary of $\Omega_Y$ under weak conditions: Remind that we assume here that there exists a $k \in (0; 1)$ such that $\lim_{u \to 0} u^k \phi'(u) = 0$. This weak condition is sufficient to insure the vanishing property of $\nabla_y [\phi(p_Y)]$.

To show this, let us write

$$\nabla_y [\phi(p_Y)] = [\nabla_y p_Y] \phi'(p_Y) = \frac{\nabla_y p_Y}{p_Y} p_Y^k \phi'(p_Y)$$

From the assumption on $\phi'$, since $p_Y$ goes to 0 in the boundary of $\Omega_Y$, the quantity $p_Y^k \phi'(p_Y)$ vanishes in the boundary of $\Omega_Y$.

Then, from $k < 1$ one applies the inverse H"older inequality [68, th. 189] to $p_Y(y) = \int_{\Omega} p_X(x) p_N(y - x) \, dx$ viewed as a scalar product between 1 and $p_N(y - x)$ of kernel $p_X$, leading to

$$p_Y(y) \geq \left( \int_{\Omega} p_X(x) \, dx \right)^\frac{1}{k} \left( \int_{\Omega} p_X(x) [p_N(y - x)]^k \, dx \right)^\frac{k}{2}$$

where $\frac{1}{k} + \frac{1}{k^*} = 1$ ($k^*$ is thus negative). We have already seen that for $p_Y$ we can permute $\nabla_y$ and the integral, hence by this permutation, bounding the norm of the integral by the integral of the norm, and from the previous inequality, one obtains

$$\frac{\|\nabla_y p_Y(y)\|}{[p_Y(y)]^k} \leq \sup_{x \in \Omega} \frac{\|\nabla_y p_N(y - x)\|}{[p_N(y - x)]^k} \frac{\int_{\Omega} p_X(x) \|\nabla_y p_N(y - x)\| \, dx}{\int_{\Omega} p_X(x) [p_N(y - x)]^k \, dx}$$

Now, a rapid study of

$$\left\| R^{\frac{k}{2}} \nabla_y p_N(y) \right\|_{[p_N(y)]^k}^2 = \alpha^{2(1-k)} \theta^{-(1-k)d/2} u \exp \left( -\frac{(1-k)u}{\theta} \right)$$

versus $u$ allows to show that

$$\frac{\|\nabla_y p_N(y)\|}{[p_N(y)]^k} \leq \frac{\|R^{-\frac{k}{2}}\|_2}{\sqrt{(1-k) e \theta (2\pi \theta)^{(1-k)d/2} |R|^{-\frac{k}{2}}}}$$

As a conclusion, $\sup_{y \in \mathbb{R}^d} \|\nabla_y p_N(y)\|_{[p_N(y)]^k}$ is finite, which finishes the proof.
REFERENCES

[1] A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2(2):101–112, June 1959.

[2] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Hoboken, New Jersey, 2nd edition, 2006.

[3] B. R. Frieden. *Science from Fisher Information: A Unification*. Cambridge University Press, Cambridge, UK, 2004.

[4] S. M. Kay. *Fundamentals for Statistical Signal Processing: Estimation Theory*. vol. 1. Prentice Hall, Upper Saddle River, NJ, 1993.

[5] C. Vignat and J.-F. Bercher. Analysis of signals in the Fisher-Shannon information plane. *Physics Letters A*, 312(1-2):27–33, June 2003.

[6] M. Basseville. Distance measures for signal processing and pattern recognition. *Signal Processing*, 18(4):349–369, December 1989.

[7] M. Basseville. Divergence measures for statistical data processing – an annotated bibliography. *Signal Processing*, 93(4):621–633, April 2013.

[8] A. O. Hero III, B. Ma, O. J. J. Michel, and J. Gorman. Application of entropic spanning graphs. *IEEE Signal Processing Magazine*, 19(5):85–95, September 2002.

[9] G. A. Darbellay and D. Wuertz. The entropy as a tool for analysing statistical dependences in financial time series. *Physica A*, 287(3-4):429–439, December 2000.

[10] W. Ebeling, L. Molgedey, J. Kurths, and U. Schwarz. Entropy, complexity, predictability and data analysis of time series and letter sequences. In *Theory of Disaster*. Springer Verlag, Berlin, A. Bundle and H.-J. Schellnhuber edition, 2000.

[11] J. S. Dehesa, S. López-Rosa, and D. Manzano. Entropy and complexity analyses of D-dimensional quantum systems. In K. D. Sen, editor, *Statistical Complexities: Application to Electronic Structure*. Springer, Berlin, 2010.

[12] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:623–656, October 1948.

[13] O. Rioul. *Théorie de l’information et du codage*. Lavoisier, Paris, 2007.

[14] A. Dembo, T. M. Cover, and J. A. Thomas. Information theoretic inequalities. *IEEE Transactions on Information Theory*, 37(6):1501–1518, November 1991.

[15] E. Lutwak, D. Yang, and G. Zhang. Moment-entropy inequalities. *The Annals of Probability*, 32(1B):757–774, January 2004.

[16] E. Lutwak, D. Yang, and G. Zhang. Cramér-Rao and moment-entropy inequalities for Rényi entropy and generalized Fisher information. *IEEE Transactions on Information Theory*, 51(2):473–478, February 2005.

[17] E. Lutwak, D. Yang, and G. Zhang. Moment-entropy inequalities for a random vector. *IEEE Transactions on Information Theory*, 53(4):1603–1607, April 2007.

[18] E. Lutwak, S. Lv, D. Yang, and G. Zhang. Extension of Fisher information and Stam’s inequality. *IEEE transactions on information theory*, 58(3):1319–1327, March 2012.

[19] J.-F. Bercher. On a $(\beta,q)$-generalized Fisher information and inequalities involving $q$-Gaussian distributions. *Journal of Mathematical Physics*, 53(6):063303, June 2012.

[20] J.-F. Bercher. On multidimensional generalized Cramér-Rao inequalities, uncertainty relations and characterizations of generalized $q$-Gaussian distributions. *Journal of Physics A*, 46(9):095303, March 2013.

[21] J.-F. Bercher. On generalized Cramér-Rao inequalities, generalized Fisher information and characterizations of generalized $q$-Gaussian distributions. *Journal of Physics A*, 45(25):255303, June 2012.

[22] A. Dembo. Information inequalities and uncertainty principles. Technical Report Cov NSF 75, Department of Statistics, Stanford University, California, USA, Department of Statistics, Stanford University, California, USA, July 1990.

[23] A. R. Barron. Entropy and the central limit theorem. *The Annals of Probability*, 14(1):336–342, January 1986.

[24] O. Johnson. *Information Theory and The Central Limit Theorem*. Imperial college Press, London, 2004.

[25] N. M. Blachman. The convolution inequality for entropy powers. *IEEE Transactions on Information Theory*, 11(2):267–271, April 1965.

[26] A. R. Barron. Monotonic central limit theorem for densities. Technical report no. 50, Department of Statistics, Stanford University, March 5 1984.

[27] O. Johnson and A. Barron. Fisher information inequalities and the central limit theorem. *Probability Theory and Related Fields*, 129(3):391–409, July 2004.

[28] D. P. Palomar and S. Verdú. Gradient of mutual information in linear vector Gaussian channels. *IEEE Transactions on Information Theory*, 52(1):141–154, January 2006.
[29] D. Guo. Relative entropy and score function: New information-estimation relationships through arbitrary additive perturbation. In *IEEE International Symposium on Information Theory*, pages 814–818, Seoul, South Korea, June 28 – July 3 2009.

[30] S. Verdú. Mismatched estimation and relative entropy. *IEEE Transactions on Information Theory*, 56(8):3712–3720, August 2010.

[31] S. Park, E. Serpedin, and K. Qaraqe. New perspectives, extensions and applications of de Bruijn identity. In *Proceedings of the IEEE 13th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC’12)*, pages 399–403, Cesme, Turkey, 17-20 June 2012.

[32] S. Park, E. Serpedin, and K. Qaraqe. On the equivalence between Stein and de Bruijn identities. *IEEE Transactions on Information Theory*, 58(12):7045–7067, December 2012.

[33] L. Brown, A. DasGupta, L. R. Haff, and W. E. Strawderman. The heat equation and Stein’s identity: Connections, applications. *Journal of Statistical Planning and Inference*, 136(7):2254 –2278, July 2006.

[34] D. Guo, S. Shamai, and S. Verdú. Additive non-gaussian noise channels: Mutual information and conditional mean estimation. In *IEEE International Symposium on Information Theory*, pages 719–723, Adelaide, SA, Australia, 4-9 september 2005.

[35] J. Havrda and F. Charvát. Quantification method of classification processes: Concept of structural α-entropy. *Kybernetika*, 3:30–35, 1967.

[36] Z. Daróczy. Generalized information functions. *Information and Control*, 16(1):36–51, March 1970.

[37] C. Tsallis. Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, 52(1-2):479–487, July 1988.

[38] C. Tsallis. Nonextensive statistics: theoretical, experimental and computational evidences and connections. *Brazilian Journal of Physics*, 29(1):1–35, March 1999.

[39] M. T. Martin, A. R. Plastino, and A. Plastino. Tsallis-like information measures and the analysis of complex signals. *Physica A*, 275(1-2):262–271, January 2000.

[40] M. Fortes and A. Plastino. Generalized entropy as measure of quantum uncertainty. *Physica A*, 225(3-4):412–430, April 1996.

[41] P. Jizba. Information theory and generalized statistics. In H.-T. Elze, editor, *Decoherence and Entropy in Complex Systems: Selected Lectures from DICE (Decoherence, Information, Complexity and Entropy; Piombino, Italy, September 2-6, 2002)*, volume 633 of *Lecture Notes in Physics*, pages 362–376, Heidelberg, 2003. Springer Verlag.

[42] P. Jizba and T. Arimitsu. The world according to Rényi: thermodynamics of multifractal systems. *Annals of Physics*, 312(1):17–59, July 2004.

[43] D. Harte. *Multifractals: Theory and applications*. Chapman & Hall / CRC, Boca Raton, 1st edition, 2001.

[44] L. L. Campbell. A coding theorem and Rényi’s entropy. *Information and Control*, 8(4):423–429, August 1965.

[45] J.-F. Bercher. Source coding with escort distributions and Rényi entropy bounds. *Physics Letters A*, 373(36):3235–3238, August 2009.

[46] I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, 2:299–318, 1967.

[47] S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society B*, 28(1):131–142, 1966.

[48] I. Vajda. χα-divergence and generalized fisher’s information. In *Transactions of the 6th Prague Conference on Information Theory, Statistics, Decision Functions and Random Processes*, pages 873–886, 1973.

[49] D. E. Boekee. An extension of the Fisher information measure. In I. Csiszár and P. Elias., editors, *Topics in information theory. Proc. 2nd Colloquium on Information Theory*, volume 16, pages 113–123, Keszthely, Hungary, 25-29 August 1975 1977.

[50] O. Johnson and C. Vignat. Some results concerning maximum Rényi entropy distributions. *Annales de l’Institut Henri Poincare (B) Probability and Statistics*, 43(3):339–351, May-June 2007.

[51] J.-F. Bercher. Some properties of generalized Fisher information in the context of nonextensive thermostatistics. *Physica A*, 392(15):3140–3154, August 2013.

[52] P. Hammad. Mesure d’ordre α de l’information au sens de Fisher. *Revue de Statistique Appliquée*, 26(1):73–84, 1978.

[53] P. Sánchez-Moreno, A. Zarzoo, and J. S. Dehesa. Jensen divergence based on Fisher’s information. *Journal of Physics A*, 45(12):125305, March 2012.

[54] S.-I. Amari and H. Nagaoka. *Methods of Information Geometry*. Oxford University Press, Rhode Island, 2000.

[55] F. Liese and I. Vajda. On divergence and informations in statistics and information theory. *IEEE Transactions on Information Theory*, 52(10):4394–4412, October 2006.
[56] D. Guo, S. Shamai, and S. Verdú. Mutual information and minimum mean-square error in Gaussian channels. *IEEE Transactions on Information Theory*, 51(4):1261–1282, April 2005.

[57] M. Salicrú. Funciones de entropía asociada a medidas de Csiszár. *Qüestió*, 11(3):3–12, 1987.

[58] G. Kaniadakis. Non-linear kinetics underlying generalized statistics. *Physica A*, 296(3-4):405–425, july 2001.

[59] P. Kumar and S. Chhina. A symmetric information divergence measure of the Csiszár’s f-divergence class and its bounds. *Computers and Mathematics with Applications*, 49(4):575–588, February 2005.

[60] A. Rényi. On measures of entropy and information. *in Proceeding of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, 1:547–561, 1961.

[61] F. S. Woods. *Advanced Calculus*. Ginn and company, Boston, 1934.

[62] H. Risken. *The Fokker-Planck Equation, Methods of Solution and Applications*. Springer Verlag. Heidelberg, 2nd edition, 1989.

[63] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes. Stochastic Models with infinite Variance*. Chapman & Hall, New-York, 1994.

[64] A. M. Tulino and S. Verdú. Monotonic decrease of the non-Gaussianness of the sum of independent random variables: A simple proof. *IEEE Transactions on Information Theory*, 52(9):4295–4297, September 2006.

[65] M. Madiman and A. Barron. The monotonicity of information in the central limit theorem and entropy power inequalities. In *IEEE International Symposium on Information Theory (ISIT’06)*, pages 1021–1025, Seattle, USA, 9-14 July 2006.

[66] J. K. Hunter and B. Nachtergaele. *Applied Analysis*. World Scientific Publishing, Singapore, 2001.

[67] C. D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, Philadelphia, PA, 2000.

[68] G. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, Cambridge, UK, 2nd edition, 1952.