

A Proof of de Bruijn Identity based on Generalized Price's Theorem

Jaume Riba, *Senior Member, IEEE* and Ferran de Cabrera, *Student Member, IEEE*
 Signal Theory and Communications Department, Technical University of Catalonia (SPCOM/UPC)
 {jaume.riba, ferran.de.cabrera}@upc.edu

Abstract—This paper shows that de Bruijn identity, which relates entropy with Fisher information, can be obtained as a particular case of an immediate generalization of Price's theorem, which is a tool used in the analysis of nonlinear memoryless systems with Gaussian inputs. It is shown that, while the general Price's theorem follows since the density of the perturbation satisfies the heat equation, the particular case of de Bruijn identity follows since the score function is zero-mean, which is the well-known condition that provides the insightful Cramér-Rao bound expression based on the negative second derivative of the log-likelihood function. The unified framework uses the characteristic function as a main tool and becomes a more intuitive alternative to the classical technical proof obtained by integrating by parts. Second-order Tsallis entropy is also briefly explored under this general framework.

Index Terms—De Bruijn identity, Price's theorem, Fisher information, differential entropy, Tsallis entropy, kernel methods, nonlinear signal processing.

I. INTRODUCTION

Nonlinear transformations of Gaussian variates is a classical subject of probability theory. By using partial differential equations, a useful theorem in the nonlinear statistical signal processing framework was provided by Price in [1]. In its original form, Price's theorem relates the expectation of a nonlinear function of a number of normally distributed random variates to the covariances of the variates. Several extensions of this theorem have been provided in the past in [2], [3] and [4]. One of the main utilities of Price's theorem is to simplify the computation of moments at the output of nonlinear devices, resorting to the fact that the derivatives of the nonlinear function are usually simpler than the original function. When this is the case, it is worth to first compute the sensitivity of the output to a small increment of the perturbation variance of the Gaussian input, and then integrate the result with respect to the variance in a second stage.

On the other hand, within the information theory field, de Bruijn identity [5]-[6] relates the derivative of the differential entropy (with respect to the variance of a Gaussian perturbation) with the nonparametric Fisher information, whose inverse is the Cramér-Rao lower bound for the variance of any unbiased estimator of a location parameter [7]. This connection between entropy and Fisher information quantities has attracted increased interest in the past few years, as it

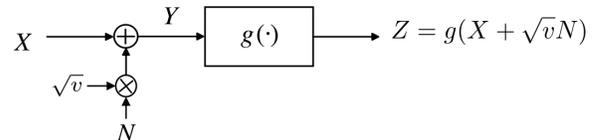


Fig. 1. The statistical model used to provide a unified framework for Price's theorem and de Bruijn identity. The purpose is to characterize the sensitivity of the expectation of a nonlinear device to the variance of an additive perturbation.

constitutes the theoretical foundation of important information theoretic inequalities [8]. Besides, it demonstrates utility in the analysis of estimation and iterative decoding schemes. In general, these two quantities are very important in information theory, statistics, statistical physics and signal processing [6], [9], [7]. Notably, the fundamental I-MMSE identity unveiled by Guo *et al.* was shown to be equivalent to de Bruijn identity [10]. Several extensions of the idea have been provided in [11], [12] and [13], and an extension to ϕ -entropies has been recently proposed in [14].

Historically, Price's theorem and de Bruijn identity (and their extensions) have followed independent routes. Both statements, however, share a very similar form of the involved differential equations. Therefore, a natural question arises of whether or not they can be related or put in a more general framework. This paper goes in this direction by trying to highlight the hidden implications and relationships between these apparently unconnected concepts. This is done first (Section II) by generalizing the applicability of Price's theorem (originally concerned with Gaussian variates) to any random variate with Gaussian perturbation. Once this immediate generalization is made, the derivation of de Bruijn identity (Section III) can be easily connected with the nonlinear signal processing analysis framework inherent in Price's theorem, and we show that this connection is possible as a result of the well-known zero-mean property of the score function. Finally, the simplicity of the approach motivates (Section IV) its extension to the second-order Tsallis entropy.

II. GENERALIZED PRICE'S THEOREM

Price's theorem is concerned with the computation of the expectation of nonlinear functions of normally distributed variates. Consider an additive Gaussian noise (AGN) channel defined as follows:

$$Y = X + \sqrt{v}N \quad (v > 0). \quad (1)$$

This work is supported by projects TEC2016-76409-C2-1-R (WINTER), Ministerio de Economía y Competitividad, Spanish National Research Plan, and 2017 SGR 578 - AGAUR, Catalan Government.

The overall theoretical framework used henceforth is captured by the model in Fig. 1, which is common in communications systems. In this model, X represents the useful information, N the Gaussian noise or interference random perturbation (independent from X), and $g(\cdot)$ a nonlinear memoryless device, which may correspond either to a physical system or to an estimator/detector, in both cases providing the (real) random variate Z .

Without loss of generality, we assume that N is zero-mean, as any mean component can be attributed to X , and unit-variance, as the variance of the true perturbation $\sqrt{v}N$ is generally established by v . The purpose is to characterize the sensitivity of Z to changes in the amount of perturbation. In particular, we are interested in the following quantity:

$$\frac{\partial}{\partial v} \mathbb{E} \{g(X + \sqrt{v}N)\} = \frac{\partial}{\partial v} \int_{-\infty}^{\infty} f_{Y|v}(y)g(y)dy, \quad (2)$$

where $f_{Y|v}(y)$ is the density of Y conditioned to the variance parameter v .

The integral in (2) is a scalar product between two functions. As established by Parseval's theorem, a scalar product is preserved by unitary transforms. Therefore, we define the Fourier transform (with sign reversal in the complex exponential) of the conditional density $f_{Y|v}(y)$ and of transformation $g(y)$, respectively, as:

$$F_{Y|v}(\omega) = \int_{-\infty}^{\infty} f_{Y|v}(y) \exp[j\omega y] dy, \quad (3)$$

$$G(\omega) = \int_{-\infty}^{\infty} g(y) \exp[j\omega y] dy. \quad (4)$$

Note that $F_{Y|v}(\omega)$ in (3) is, by definition, the (conditional) *characteristic function* of the random variate Y . As Y is the sum of two independent random variates, its *probability density function* (PDF) is the convolution of both PDFs. Therefore, the characteristic function of the sum is the product of characteristic functions of both terms, that is

$$F_{Y|v}(\omega) = F_{\sqrt{v}N}(\omega)F_X(\omega). \quad (5)$$

As a result of the independence property given in (5) and Parseval's theorem, we can write the right hand side of (2) as follows:

$$\begin{aligned} & \frac{1}{2\pi} \frac{\partial}{\partial v} \int_{-\infty}^{\infty} F_{Y|v}(\omega)G^*(\omega)d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial v} F_{\sqrt{v}N}(\omega) \right) F_X(\omega)G^*(\omega)d\omega. \end{aligned} \quad (6)$$

In the sequel we assume that the density of N fulfills the heat equation [5], [12]:

$$\frac{\partial}{\partial v} f_{\sqrt{v}N}(y) = \frac{1}{2} \frac{\partial^2}{\partial y^2} f_{\sqrt{v}N}(y). \quad (7)$$

Expressing (7) by means of the inverse Fourier transform (with sign reversal in the complex exponential) yields:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial v} F_{\sqrt{v}N}(\omega) \right) \exp[-j\omega y] d\omega \quad (8)$$

$$= \frac{1}{2\pi} \frac{1}{2} \frac{\partial^2}{\partial y^2} \int_{-\infty}^{\infty} F_{\sqrt{v}N}(\omega) \exp[-j\omega y] d\omega \quad (9)$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\frac{-\omega^2}{2} F_{\sqrt{v}N}(\omega) \right) \exp[-j\omega y] d\omega. \quad (10)$$

By identifying the integrands in (8) and (10) we obtain the heat equation in the transformed domain:

$$\frac{\partial}{\partial v} F_{\sqrt{v}N}(\omega) = \frac{-\omega^2}{2} F_{\sqrt{v}N}(\omega). \quad (11)$$

It is well-known that the normal density obeys the heat equation. Effectively, if N is normal, then

$$F_{\sqrt{v}N}(\omega) = \exp\left[-v \frac{\omega^2}{2}\right] \quad (12)$$

and, therefore,

$$\frac{\partial}{\partial v} F_{\sqrt{v}N}(\omega) = \frac{-\omega^2}{2} \exp\left[-v \frac{\omega^2}{2}\right] = \frac{-\omega^2}{2} F_{\sqrt{v}N}(\omega), \quad (13)$$

as stated in (11). As a result, following from (6) by using the property in (11), we have:

$$\begin{aligned} & \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\frac{-\omega^2}{2} F_{\sqrt{v}N}(\omega) \right) F_X(\omega)G^*(\omega)d\omega \\ &= \frac{1}{2\pi} \frac{1}{2} \int_{-\infty}^{\infty} (-\omega^2 G^*(\omega)) (F_{\sqrt{v}N}(\omega)F_X(\omega)) d\omega \\ &= \frac{1}{2\pi} \frac{1}{2} \int_{-\infty}^{\infty} (-\omega^2 G(\omega))^* F_{Y|v}(\omega)d\omega. \end{aligned} \quad (14)$$

The previous integral is a scalar product between functions $-\omega^2 G(\omega)$ and $F_{Y|v}(\omega)$. Therefore, by applying again the Parseval's theorem and using the fact that

$$\begin{aligned} & \frac{1}{2\pi} \int_{-\infty}^{\infty} (-\omega^2 G(\omega)) \exp[-j\omega y] d\omega \\ &= \frac{1}{2\pi} \frac{\partial^2}{\partial y^2} \int_{-\infty}^{\infty} G(\omega) \exp[-j\omega y] d\omega = \frac{\partial^2}{\partial y^2} g(y) = g''(y), \end{aligned} \quad (15)$$

the right hand side of (14) can be written as follows:

$$\frac{1}{2} \int_{-\infty}^{\infty} g''(y) f_{Y|v}(y) dy. \quad (16)$$

As a result we obtain the following generalization of Price's theorem:

Theorem 1. (Generalization of Price's theorem for the statistical analysis of nonlinear systems under additive Gaussian perturbation). Given the statistical model in Fig. 1, where X stands for any random variate with finite second-order moment, N is a zero-mean, unit-variance normal random variate, and $g(\cdot)$ is any memoryless nonlinear transformation, then

$$\frac{\partial}{\partial v} \mathbb{E} \{g(X + \sqrt{v}N)\} = \frac{1}{2} \mathbb{E} \{g''(X + \sqrt{v}N)\}, \quad (17)$$

where

$$g''(y) = \frac{\partial^2}{\partial y^2} g(y) \quad (18)$$

is the second derivative of the memoryless transformation.

Note that Theorem 1 reduces to the classical Price's theorem¹ for the particular setting of $X = 0$ in (17), in which case the input to the nonlinear device becomes normally distributed. The more general statement in Theorem 1 includes an arbitrary independent additive term X while maintaining the simplicity of the proof thanks to the use of the characteristic function. Effectively, the added independent component X only affects as a multiplicative term $F_X(\omega)$ in (5), and does not involve any complication in the rest of the mathematical development. The addition of this general term not only enhances the applicability of this theorem to the nonlinear signal processing field as an analysis tool (as exemplified in the sequel), but also provides a very direct link to de Bruijn identity (as seen in Section III). It should be noted that the stated generalization in Theorem 1 with respect to the original Price's statement in [1] is not strictly novel², as it comes essentially from the fact that a PDF given by the convolution of a probability density function with a Gaussian density itself satisfies the heat equation. Implicitly, this fact is pointed out by Stam in equation (5.1) of [5].

In the sequel we provide two examples of the applicability of Theorem 1 in the nonlinear signal processing field for non-Gaussian inputs. In both examples, we model X as a binary random variate with density function³

$$f_X(x) = \frac{1}{2}\delta(x - A) + \frac{1}{2}\delta(x + A), \quad (19)$$

which is common in communications, so that the density of Y is a Gaussian mixture:

$$f_{Y|v}(y) = \frac{1}{2\sqrt{2\pi v}} \times \left(\exp\left[\frac{-(y - A)^2}{2v}\right] + \exp\left[\frac{-(y + A)^2}{2v}\right] \right). \quad (20)$$

Example 1: (The linear rectifier and the *folded normal density*, [15]) Let $g(\cdot)$ be a linear rectifier characterized by the following nonlinear transformation: $g(y) = |y|$. In that case $Z = |X + N|$ follows the so-called folded normal density. The second derivative of the nonlinear device with respect to its input is:

$$g''(y) = \frac{\partial^2}{\partial y^2} g(y) = 2\delta(y). \quad (21)$$

Therefore, using Theorem 1, we have:

$$\begin{aligned} \frac{\partial}{\partial v} \mathbb{E}\{|Y|\} &= \frac{1}{2} \mathbb{E}\{2\delta(Y)\} \\ &= \int_{-\infty}^{\infty} \delta(y) f_{Y|v}(y) dy = f_{Y|v}(0) = \frac{1}{\sqrt{2\pi v}} \exp\left[\frac{-A^2}{2v}\right]. \end{aligned} \quad (22)$$

¹Note that, with respect to [1], (17) is solely concerned with the derivatives with respect to the variance. Although the theorem may also include the derivatives with respect to the correlation coefficients in the case of more than one random variate, this more general form is omitted here for the sake of concreteness. For further details, the reader is referred to equation (7) in [4], which includes both the variances and covariances in a single expression.

²We would like to thank the anonymous reviewer who raised this issue.

³Throughout this paper, $\delta(\cdot)$ is the Dirac delta function.

Similarly, with the traditional Price's theorem tool, the proposed generalization can be used to compute $\mathbb{E}\{|Y|\}$ by integration with respect to the variance of the perturbation:

$$\mathbb{E}\{|Y|\} = \int_0^v \frac{1}{\sqrt{2\pi\bar{v}}} \exp\left[\frac{-A^2}{2\bar{v}}\right] d\bar{v} + C, \quad (23)$$

where C is obtained by solving the case of $v = 0$ directly, which in that case easily leads to $C = \mathbb{E}\{|X|\} = A$. Using integration by parts in (23), the well-known mean of the folded normal density is obtained (see equation (7) in [15]).

Example 2: (The fourth-power) Let $g(\cdot)$ be a fourth-power device characterized by the following nonlinear transformation: $g(y) = y^4$. The second derivative of the nonlinear device with respect to its input is:

$$g''(y) = \frac{\partial^2}{\partial y^2} g(y) = 12y^2. \quad (24)$$

Therefore, using Theorem 1, we have:

$$\frac{\partial}{\partial v} \mathbb{E}\{Y^4\} = \frac{1}{2} \mathbb{E}\{12Y^2\} = 6(A^2 + v). \quad (25)$$

$\mathbb{E}\{Y^4\}$ can now be obtained by integration with respect to the variance of the perturbation:

$$\mathbb{E}\{Y^4\} = \int_0^v 6(A^2 + \bar{v}) d\bar{v} + C, \quad (26)$$

where C is obtained by solving the case of $v = 0$ directly, which would easily lead to $C = \mathbb{E}\{X^4\} = A^4$. Therefore, we have

$$\mathbb{E}\{Y^4\} = 3v^2 + 6A^2v + A^4. \quad (27)$$

Remark from the examples: It is worth highlighting what is common in the considered *Examples 1* and *2*: in both illustrations, the reason why Theorem 1 became specially useful was that the second derivative of the considered nonlinear function $g(\cdot)$ is mathematically much simpler than the original function.

III. DE BRUIJN IDENTITY BASED ON GENERALIZED PRICE'S THEOREM

Once we have obtained Theorem 1 and exemplified its utility in the nonlinear statistical signal processing field, we next move to its connection with de Bruijn identity.

The differential entropy of $Y = X + \sqrt{v}N$ is defined as:

$$h(X + \sqrt{v}N) = - \int_{-\infty}^{\infty} f_{Y|v}(y) \ln f_{Y|v}(y) dy. \quad (28)$$

Applying the chain rule, the derivative of the differential entropy with respect to the variance of the perturbation is:

$$\begin{aligned} \frac{\partial}{\partial v} h(X + \sqrt{v}N) &= - \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial v} f_{Y|v}(y) \right) \ln f_{Y|v}(y) dy \\ &\quad - \int_{-\infty}^{\infty} f_{Y|v}(y) \left(\frac{\partial}{\partial v} \ln f_{Y|v}(y) \right) dy. \end{aligned} \quad (29)$$

The last integral in (29) can be expressed as follows:

$$\int_{-\infty}^{\infty} f_{Y|v}(y) \left(\frac{\partial}{\partial v} \ln f_{Y|v}(y) \right) dy = \mathbb{E}\{\xi_v(Y)\}, \quad (30)$$

where

$$\xi_v(Y) = \frac{\partial}{\partial v} \ln f_{Y|v}(Y) \quad (31)$$

is the score function associated to the density $f_{Y|v}(y)$ with respect to its parameter v . A well-known result in the estimation field [7] is that the score function is zero-mean for any parameter (see Appendix), for which the second integral in (29) is null.

Let us define the variance of the perturbation as

$$v = \gamma + \epsilon, \quad (32)$$

where γ is the nominal variance and ϵ is some small incremental perturbation. With this consideration, and as a consequence of the zero-mean property of the score, we have from (29) that

$$\begin{aligned} \frac{\partial}{\partial v} h(X + \sqrt{v}N) &= \frac{\partial}{\partial \epsilon} h(X + \sqrt{\gamma + \epsilon}N) \Big|_{\epsilon=0} \\ &= - \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial \epsilon} f_{Y|\gamma+\epsilon}(y) \Big|_{\epsilon=0} \right) g(y) dy \quad (33) \end{aligned}$$

$$= \frac{-\partial}{\partial \epsilon} \left(\int_{-\infty}^{\infty} f_{Y|\gamma+\epsilon}(y) g(y) dy \right) \Big|_{\epsilon=0} = \frac{-\partial}{\partial \epsilon} \mathbb{E} \{g(Y)\} \Big|_{\epsilon=0},$$

where

$$g(y) = \ln f_{Y|\gamma}(y). \quad (34)$$

As a result of (33)&(34), we can now interpret the derivative of the differential entropy with respect to the perturbation variance in view of the generalized Price's theorem. To clarify this idea, Fig. 2 depicts the underlying statistical experiment in view of (33)&(34). It can be seen that, as a particular instance, the equivalent nonlinear device is now formed by the log-likelihood function evaluated at the nominal variance γ . By applying the generalized Price's theorem to that problem, we obtain:

$$\begin{aligned} \frac{\partial}{\partial v} h(X + \sqrt{v}N) &= \frac{-\partial}{\partial \epsilon} \mathbb{E} \{ \ln f_{Y|\gamma}(X + \sqrt{\gamma + \epsilon}N) \} \Big|_{\epsilon=0} \\ &= \frac{-1}{2} \mathbb{E} \left\{ \frac{\partial^2}{\partial Y^2} \ln f_{Y|v}(Y) \right\}, \quad (35) \end{aligned}$$

which yields the well-known de Bruijn identity:

$$\frac{\partial}{\partial v} h(X + \sqrt{v}N) = \frac{1}{2} J(X + \sqrt{v}N), \quad (36)$$

where $J(Y)$ is the Fisher information associated to $f_{Y|v}(Y)$ with respect to a location (or translation) parameter:

$$J(Y) = -\mathbb{E} \left\{ \frac{\partial^2}{\partial Y^2} \ln f_{Y|v}(Y) \right\}. \quad (37)$$

Note that the Fisher information can also be expressed as follows:

$$J(Y) = \mathbb{E} \left\{ \left(\frac{\partial}{\partial Y} \ln f_{Y|v}(Y) \right)^2 \right\}. \quad (38)$$

As it is well-known in the estimation field [7], the expressions of the Fisher information given in (37) and (38) are equivalent as a consequence of the zero-mean property of the score function with respect to any parameter. Interestingly, as it has

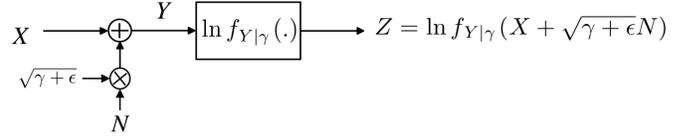


Fig. 2. The statistical model used to embrace the de Bruijn identity in the nonlinear signal processing analysis framework. De Bruijn identity comes directly from Price's theorem for the case that the nonlinear device in Fig. 1 is the log-likelihood function evaluated at the nominal perturbation variance.

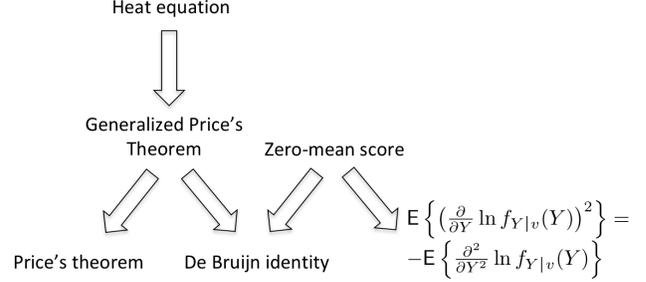


Fig. 3. Summary of the implications and properties considered in this paper.

been shown, the zero-mean property of the score is precisely the main property that made possible to derive the de Bruijn identity from the generalized Price's theorem.

Finally, a summary of the implications analyzed in this paper is depicted in Fig. 3, where de Bruijn identity is situated within a more general framework. Under this new vision, while Theorem 1 offers a general tool for the analysis of nonlinear systems in AGN channels, Price's theorem and de Bruijn identity become special cases.

IV. EXTENSION TO TSALLIS ENTROPY

Theorem 1 is used here to extend the analysis to other entropies. For illustration, we focus on Tsallis entropy [16] given by

$$S_q(X + \sqrt{v}N) = \frac{1}{q-1} \left(1 - \int_{-\infty}^{\infty} f_{Y|v}^q(y) dy \right), \quad (39)$$

where q is the entropic index. Applying the chain rule:

$$\frac{\partial}{\partial v} S_q(X + \sqrt{v}N) = \frac{-q}{q-1} \int_{-\infty}^{\infty} f_{Y|v}^{q-1}(y) \frac{\partial}{\partial v} f_{Y|v}(y) dy. \quad (40)$$

Using again the incremental perturbation idea introduced in (32), we have:

$$\begin{aligned} \frac{\partial}{\partial v} S_q(X + \sqrt{v}N) &= \frac{\partial}{\partial \epsilon} S_q(X + \sqrt{\gamma + \epsilon}N) \Big|_{\epsilon=0} \\ &= - \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial \epsilon} f_{Y|\gamma+\epsilon}(y) \Big|_{\epsilon=0} \right) g(y) dy \quad (41) \\ &= \frac{-\partial}{\partial \epsilon} \left(\int_{-\infty}^{\infty} f_{Y|\gamma+\epsilon}(y) g(y) dy \right) \Big|_{\epsilon=0} = \frac{-\partial}{\partial \epsilon} \mathbb{E} \{g(Y)\} \Big|_{\epsilon=0}, \end{aligned}$$

where

$$g(y) = \frac{q}{q-1} f_{Y|\gamma}^{q-1}(y). \quad (42)$$

Once more, but now without necessity of resorting to the zero-mean property of the score, we can interpret the derivative of

the Tsallis entropy with respect to the perturbation variance in view of the generalized Price's theorem. We focus on the case of $q = 2$ because it is the only case for which a simple expression is obtained. In that case, it is clear from (42) that the equivalent nonlinear device is formed by twice the likelihood function evaluated at the nominal variance γ , that is

$$g(y) = 2f_{Y|\gamma}(y). \quad (43)$$

By applying the generalized Price's theorem to this new problem we obtain:

$$\begin{aligned} & \frac{\partial}{\partial v} S_2(X + \sqrt{v}N) \\ &= \frac{-\partial}{\partial \epsilon} \mathbb{E} \left\{ 2f_{Y|\gamma}(X + \sqrt{\gamma + \epsilon}N) \right\} \Big|_{\epsilon=0} \\ &= -\mathbb{E} \left\{ \frac{\partial^2}{\partial Y^2} f_{Y|v}(Y) \right\}, \end{aligned} \quad (44)$$

which yields a de Bruijn-like identity for Tsallis entropy:

$$\frac{\partial}{\partial v} S_2(X + \sqrt{v}N) = J_S(X + \sqrt{v}N), \quad (45)$$

where $J_S(Y)$ is a Fisher-like information associated to $f_{Y|v}(Y)$ with respect to a location (or translation) parameter:

$$J_S(Y) = -\mathbb{E} \left\{ \frac{\partial^2}{\partial Y^2} f_{Y|v}(Y) \right\}. \quad (46)$$

Therefore, we find that the second order Tsallis entropy admits a de Bruijn-like identity where the likelihood function is involved instead of the log-likelihood function, and the $1/2$ constant is absent with respect to the original expression. Interestingly, the case of $q = 2$ provides a quadratic measure of entropy which involves the estimation of the squared norm of the PDF. In fact, the integral in (39) for $q = 2$ is also shared by the second-order Rényi's entropy [17], and it is used by some researchers [18], [19] to provide more robust signal processing approaches. The main advantage of the case $q = 2$ is that the squared norm of the PDF, also called information potential, can be estimated directly from independent data samples using kernel methods [18]. Therefore, the new identity in (45) may be of interest to provide new insights into this context.

V. CONCLUSIONS

The similarity between the partial derivatives involved in Price's theorem and de Bruijn identity have been explored in this paper. The connection has been established by providing an immediate generalization of Price's theorem, which was originally formulated for Gaussian inputs on nonlinear devices. Instead of pure Gaussian inputs, the generalization applies to AGN channels with any input distribution. To highlight the general form, some examples have been provided in the context of nonlinear statistical signal processing analysis. With the generalization to AGN channels, it has been shown that the de Bruijn identity can be derived as a particular case, and that this connection is possible as a consequence of the well-known zero-mean property of the score function. The general framework has also provided an identity for the second-order Tsallis entropy very similar to the de Bruijn identity. In

summary, the unveiled connection provides a more intuitive view and may perhaps suggest further extensions and insights to related branches of physics and engineering.

APPENDIX

The classical technical proof (see also [7], Eq. (3A.2)) of the zero-mean regularity condition of the *score function* reads:

$$\begin{aligned} \mathbb{E} \left\{ \frac{\partial}{\partial v} \ln f_{Y|v}(Y) \right\} &= \mathbb{E} \left\{ \frac{\frac{\partial}{\partial v} f_{Y|v}(Y)}{f_{Y|v}(Y)} \right\} \\ &= \int_{-\infty}^{\infty} f_{Y|v}(y) \frac{\frac{\partial}{\partial v} f_{Y|v}(y)}{f_{Y|v}(y)} dy = \int_{-\infty}^{\infty} \frac{\partial}{\partial v} f_{Y|v}(y) dy \\ &= \frac{\partial}{\partial v} \int_{-\infty}^{\infty} f_{Y|v}(y) dy = \frac{\partial}{\partial v} 1 = 0. \end{aligned} \quad (47)$$

In general, the order of differentiation and integration in (47) may be interchanged if the domain for which $f_{Y|v}(y)$ is zero does *not* depend on the parameter v . This requirement is clearly fulfilled in this paper, since a Gaussian perturbation with positive variance assures that $f_{Y|v}(y) > 0$ for all $v > 0$.

REFERENCES

- [1] R. Price, "A useful theorem for nonlinear devices having Gaussian inputs," *IEEE Trans. Inf. Theory*, vol. 4, pp. 69–72, June 1958.
- [2] R. Pawula, "A modified version of Price's theorem," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 285–288, 1967.
- [3] A. van den Bos, "Price's theorem for complex variates," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 286–287, 1996.
- [4] —, "Nonlinear statistical signal processing: Useful theorems and their application," in *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP)*, June 1999, pp. 603–606.
- [5] A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Inf. Control*, vol. 2, pp. 101–112, 1959.
- [6] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York: Wiley, 2006.
- [7] S. M. Kay, *Fundamentals of Statistical Signal Processing: estimation theory*. New York: Prentice-Hall, 1993.
- [8] O. Rioul, "Information theoretic proofs of entropy power inequalities," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 34–39, Jan. 2011.
- [9] B. R. Frieden, *Science from Fisher information: a unification*, Cambridge, U.K. ed. Cambridge, Univ. Press, 2004.
- [10] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.
- [11] D. Guo, "Relative entropy and score function: new information-estimation relationships through arbitrary additive perturbation," in *IEEE Int. Symposium on Information Theory*, 2009, pp. 814–818.
- [12] S. Park, E. Serpedin, and K. Qaraqe, "New perspectives, extensions and applications of de Bruijn identity," in *IEEE 13th International Workshop on Signal Processing Advances in Wireless Communications*, 2012, pp. 399–403.
- [13] O. Johnson and S. Guha, "A de Bruijn identity for discrete random variables," in *IEEE Int. Symposium on Information Theory*, 2017, pp. 898–902.
- [14] I. Valero, S. Zozor, and J. M. Brossier, "Generalization of the de Bruijn identity to general ϕ -entropies and ϕ -Fisher informations," *IEEE Trans. Inf. Theory*, vol. 64, no. 10, pp. 6743–6758, Oct. 2018.
- [15] M. Tsagris, C. Beneki, and H. Hassani, "On the folded normal distribution," *Mathematics*, vol. 2, no. 1, pp. 12–28, Feb. 2014 (arXiv:1402.3559v1).
- [16] C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics," *Journal of Statistical Physics*, vol. 52, pp. 479–487, 1988.
- [17] A. Rényi, "On measures of entropy and information," in *4th Berkeley Symp. Prob. Theory Maths. Stat. Berkeley, CA, USA*, 1961, pp. 547–561.
- [18] J. C. Principe, *Information Theoretic Learning: Rényi's Entropy and Kernel Perspectives*. New York: Springer, 2010.
- [19] F. de Cabrera, J. Riba, and G. Vázquez, "Entropy-based covariance determinant estimation," in *IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications*, May. 2017.