On the equivalence between Stein identity and de Bruijn identity

Sangwoo Park, Erchin Serpedin, and Khalid Qaraqe
Electrical and Computer Engineering Department, Texas A&M University, College Station, TX 77843, USA
Email: swpark78@neo.tamu.edu, serpedin@ece.tamu.edu, kqaraqe@tamu.edu

Abstract—This paper illustrates the equivalence between two fundamental results: Stein identity, originally proposed in the statistical estimation realm, and de Bruijn identity, considered for the first time in the information theory field. Two distinctive extensions of de Bruijn identity are presented as well. For arbitrary but fixed input and noise distributions, the first-order derivative of differential entropy is expressed by means of a function of the posterior mean, while the second-order derivative of differential entropy is manifested in terms of a function of Fisher information. Several applications exemplify the utility of the proposed results.

I. INTRODUCTION

De Bruijn identity, which shows a link between differential entropy and Fisher information, was first mentioned and exploited by Stam in the context of building the first rigorous proof of the entropy power inequality (EPI), a fundamental result exploited by Shannon to prove a lower bound on the capacity of additive Gaussian noise channels. Recently, a renewed interest was manifested in the applications of de Bruijn identity in estimation and turbo (iterative) decoding schemes, and in relating the input-output mutual information with the minimum mean-square error (MMSE) of additive Gaussian and non-Gaussian noise channels [1], [2].

A large number of approaches have been applied to prove de Bruijn identity. Amongst them, herein paper, we consider the most direct approach because it shares one common technique, viz. integration by parts, with another well-known result—referred to as Stein identity [3], [4]. Stein identity points out a relationship between the expectations of a function of the posterior mean, while the second-order derivative of differential entropy is expressed by means of a function of Fisher information. Several applications exemplify the utility of the proposed results.

II. RELATIONSHIPS BETWEEN STEIN IDENTITY AND DE BRUIJN IDENTITY

To interconnect Stein identity and de Bruijn identity, first we will introduce three fundamental results.

Theorem 1 (De Bruijn’s Identity [8], [9]): Given the channel model (1), let \(X\) be an arbitrary RV with a finite second-order moment, and \(W\) be a Gaussian RV with zero mean and unit variance. Independence between RVs \(X\) and \(W\) is also RV \(W\) is Gaussian, the equivalence between the generalized Stein and de Bruijn identities is proved. In particular, when RV \(Y\) is Gaussian, not only Stein and de Bruijn identities are equivalent, but also they are equivalent to the heat equation identity [5].

The second major goal of this paper is to generalize de Bruijn identity in two distinctive ways. Considering the fixed noise distribution channel model (1), the first-order derivative of differential entropy of output signal \(Y\) will be expressed as a function of the posterior mean \(E_{X|Y}\cdot|\cdot\), while the second-order derivative of differential entropy of output signal \(Y\) will be represented explicitly in terms of Fisher information. Even though some of these relationships do not include Fisher information, they still show fundamental relationships among basic concepts in information theory and statistical estimation theory, and these relationships hold for arbitrary noise channels.

Several applications are briefly mentioned to illustrate the usefulness of the proposed new results. Costa’s entropy power inequality (EPI) is derived in a simpler and alternative way. Cramér-Rao lower bound (CRLB) [6], Bayesian Cramér-Rao lower bound (BCRLB) [7], and a new lower bound—tighter than the BCRLB— for the mean square error (MSE) in Bayesian estimation can be further derived by exploiting the proposed results. Even though some of the proposed applications have already been proved before, herein paper a series of novel relationships and perspectives are presented on these applications.

The rest of this paper is organized as follows. Preliminary results and relationships between Stein and de Bruijn identities are provided in Section II. Two distinctive extensions of de Bruijn identity are presented in Section III. In Section IV, several potential applications of the aforementioned results are briefly mentioned due to lack of space. Finally, conclusions are stated in Section V.
assumed. Then,
\[ \frac{d}{da} h(Y) = \frac{1}{2} J(Y), \]
where \( h(\cdot) \) stands for differential entropy, \( J(Y) = \mathbb{E}_Y [S_r(Y)^2] \), \( \mathbb{E}_r[\cdot] \) is the expectation with respect to \( Y \). \( S_r(Y) \) is defined as \( d \log f_r(y; a)/dy \), and \( \log \) denotes the natural logarithm.

**Proof:** See [9].

**Theorem 1 (Generalized Stein Identity [10]):** Let \( Y \) be an absolutely continuous random variable. If the probability density function (pdf) \( f_Y(y) \) satisfies \( \lim_{y \to \pm \infty} k(y) f_Y(y) = 0 \), and
\[ \frac{d}{dy} f_Y(y) \quad = \quad -\frac{d}{dy} k(y) + \frac{(\nu - t(y))}{k(y)} \]
for some function \( k(y) \), then
\[ \mathbb{E}_Y [r(Y)(t(Y) - \nu)] = \mathbb{E}_Y \left[ \frac{d}{dy} r(Y) k(Y) \right]. \tag{2} \]
for any function \( r(y) \) which satisfies \( \mathbb{E}_Y [r(Y)t(Y)] < \infty \), \( \mathbb{E}_Y [r(Y)^2] < \infty \), and \( \mathbb{E}_Y \left[ \left| k(Y) \frac{d}{dy} r(Y) \right| \right] < \infty \). In particular, when \( Y \) is a Gaussian RV with mean \( \mu_y \) and variance \( \sigma^2_y \), (2) is simplified to
\[ \mathbb{E}_Y [r(Y)(Y - \mu_y)] = \sigma^2_y \mathbb{E}_Y \left[ \frac{d}{dy} r(Y) \right]. \tag{3} \]
Equation (3) represents the well-known classic Stein identity.

**Proof:** See [10].

**Theorem 2 (Heat Equation Identity [5]):** Let \( Y \) be a Gaussian RV with mean \( \mu \) and variance \( 1 + a \). Assume \( g(y) \) is a twice continuously differentiable function, and that both \( g(y) \) and \( \frac{d}{dy} g(y) \) are\(^1 \) \( O(e^{\alpha |y|}) \) for some \( 0 \leq c < \infty \). Then,
\[ \frac{d}{da} \mathbb{E}_Y [g(Y)] = \frac{1}{2} \mathbb{E}_Y \left[ \frac{d^2}{dY^2} g(Y) \right]. \tag{4} \]

**Proof:** See [5].

Since Stein identity (3) is valid only for Gaussian RVs, we have to consider either a special case of de Bruijn identity in Theorem 1 or use a generalized version of Stein identity (2) to compare it with de Bruijn identity. Therefore, given (1) with \( W \) a Gaussian RV, the equivalence between Stein identity and de Bruijn identity is next established.

**Theorem 4:** Given the channel (1), let \( X \) be an arbitrary RV with a finite second-order moment, and \( W \) be a Gaussian RV with zero mean and unit variance. Independence between RVs \( X \) and \( W \) is also assumed. If \( r(y; a) \) is defined as
\[ r(y; a) = -\frac{d}{dy} \log f_r(y; a), \]
in (2), de Bruijn identity is equivalent to the generalized Stein identity. In particular, if \( X \) is also a Gaussian RV, and \( g(y; a) = -\log f_r(y; a) \) in (4), then de Bruijn, Stein, and heat equation identities are all equivalent to one another.

\(^1\) \( O(\cdot) \) denotes the limiting behavior of a function, i.e., \( g(y) = O(g(y)) \) if and only if there exist positive real numbers \( K \) and \( y^* \) such that \( g(y) \leq K|g(y)| \) for any \( y \) which is greater than \( y^* \).

**Proof:** Like the proof of Theorem 3 in [5], the equivalence is proved by showing that de Bruijn identity is derived from Stein identity, and vice versa. Before illustrating the details of the proof, Lemma 1 will be exploited.

**Lemma 1:** For RVs \( X \) and \( Y \) defined in (1), the following relations hold:
\[ \frac{d}{da} \log f_Y(y; a) \quad = \quad \frac{\mathbb{E}_X \left[ (y - X)^2 f_{Y|X}(y|X; a) \right]}{2a^2 f_Y(y; a)} \quad \text{and} \quad \frac{d}{da} \log f_Y(y; a) \quad = \quad \frac{\mathbb{E}_X \left[ (y - X) f_{Y|X}(y|X; a) \right]}{a f_Y(y; a)}. \tag{5} \]

Exploiting the generalized Stein identity for
\[ r(y; a) = -\frac{d}{dy} \log f_r(y; a), \quad k(y) = 1, \]
\[ t(y; a) = -\frac{d}{dy} f_r(y; a), \quad \text{and} \quad \nu = 0, \]
de Bruijn identity is proved via the following steps.

\[ \frac{1}{2} J(Y) = \frac{1}{2} \mathbb{E}_Y \left[ \frac{d}{dy} r(Y; a) \right] \]
\[ = -\frac{1}{2} \mathbb{E}_Y \left[ r(Y; a) \frac{d}{dy} f_Y(Y; a) \right] \]
\[ = -\mathbb{E}_X \left[ \int_{-\infty}^{\infty} (y - X) f_{Y|X}(y|X; a) \frac{d}{dy} \log f_Y(y; a) dy \right] \]
\[ = -\mathbb{E}_X \left[ \int_{-\infty}^{\infty} (y - u) \int_{-\infty}^{\infty} f_{X|Y}(y|u; a) \frac{d}{dy} \log f_Y(y; a) du \right]. \tag{A} \]

Using Stein identity, dominated convergence theorem (DCT), and Fubini’s theorem, the equalities in (6) are established. Adopting the change of variables \( y = u + \sqrt{aw} \), the term \( (A) \) is expressed as
\[ \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{w^2}{2} \right) \frac{d}{da} \log f_Y(y; a) \] 
\[ = -\frac{d}{da} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{w^2}{2} \right) \log f_Y(u; a) du. \tag{7} \]

Due to (5) and normality of conditional pdf \( f_{Y|X}(y|u; a) \), (7) is verified. Re-defining \( w \) as \( (y - u)/\sqrt{a} \), (6) is expressed as
\[ \int_{-\infty}^{\infty} f_X(u) \left( \int_{-\infty}^{\infty} f_{Y|X}(y|u; a) \frac{d}{dy} \log f_Y(y; a) dy \right) du \]
\[ = -\frac{d}{da} \int_{-\infty}^{\infty} f_{Y|X}(y|u; a) \log f_Y(y; a) du \]
\[ = \frac{d}{da} h(Y). \]
Therefore, de Bruijn identity:
\[
\frac{1}{2} J(Y) = \frac{d}{da} h(Y),
\]
is established from the generalized Stein identity.

Second, the generalized Stein identity is derived from de Bruijn identity as follows. First, define the function
\[
g(y; a) = \int_0^y r(u; a) du + q(a),
\]
and express its expectation as follows
\[
E_Y [g(Y; a)]
= \int_0^\infty f_Y(y; a) r(u; a) dy - \int_0^\infty f_Y(y; a) dy + q(a)
= \int_{-\infty}^{\infty} f_X(x) \int_{-\infty}^{\infty} \left(1 - \Phi_{Y|X} \left(\frac{u - x}{\sqrt{a}}\right)\right) r(u; a) du dX
- \int_{-\infty}^{\infty} f_X(x) \int_{-\infty}^{\infty} \Phi_{Y|X} \left(\frac{u - x}{\sqrt{a}}\right) r(u; a) du dX + q(a),
\]
where \( \Phi_{Y|X} (\cdot) \) denotes the standard conditional normal cumulative density function. Then, differentiate both sides of (8) with respect to parameter \( a \):
\[
\frac{d}{da} E_Y [g(Y; a)] = -E_Y \left[ \int_{-\infty}^{\infty} \Phi_{Y|X} \left(\frac{u - X}{\sqrt{a}}\right) r(u; a) du \right]
+ E_X \left[ \int_{-\infty}^{\infty} \left(1 - \Phi_{Y|X} \left(\frac{u - X}{\sqrt{a}}\right)\right) \frac{d}{da} r(u; a) du \right] - E_X \left[ \int_{-\infty}^{\infty} \Phi_{Y|X} \left(\frac{u - X}{\sqrt{a}}\right) \frac{d}{da} r(u; a) du \right] + \frac{d}{da} q(a). \tag{9}
\]
Since \( (d/da) g(y; a) = (d/da) \int_0^y r(u; a) du + (d/da) q(a), \) equations (B) and (C) are simplified as follows:
\[(B) - (C) = \int_{-\infty}^{\infty} f_Y(y; a) \frac{d}{da} g(y; a) dy - \frac{d}{da} q(a) = -\frac{d}{da} q(a). \]
Therefore, (9) is expressed as
\[
\frac{d}{da} E_Y [g(Y; a)] = -E_X \left[ \int_{-\infty}^{\infty} \frac{d}{da} \Phi_{Y|X} \left(\frac{u - X}{\sqrt{a}}\right) r(u; a) du \right]
- E_X \left[ \int_{-\infty}^{\infty} \frac{d}{da} \Phi_{Y|X} \left(\frac{u - X}{\sqrt{a}}\right) r(u; a) du \right]
- \frac{1}{2} \int_{-\infty}^{\infty} \frac{d}{da} f_Y(u; a) f_Y(u; a) r(u; a) du.
\]
Since
\[
\frac{d}{da} h(Y) = \frac{d}{da} E_Y [g(Y; a)] = \frac{1}{2} E_Y \left[ \frac{d}{dY} f_Y(Y; a) \frac{d}{dY} r(Y; a) \right]
\]
and
\[
\frac{1}{2} J(Y) = \frac{1}{2} E_Y \left[ \frac{d^2}{dY^2} g(Y; a) \right] = \frac{1}{2} E_Y \left[ \frac{d}{dY} r(Y; a) \right],
\]
from de Bruijn identity, we derive the generalized Stein identity as
\[
\frac{1}{2} E_Y \left[ \frac{d}{dY} f_Y(Y; a) \frac{d}{dY} r(Y; a) \right] = \frac{1}{2} E_Y \left[ \frac{d}{dY} r(Y; a) \right]
\]
\[
\iff E_Y [t(Y; a) r(Y; a)] = E_Y \left[ \frac{d}{dY} r(Y; a) \right],
\]
where \( t(y; a) = -\frac{d}{dY} f_Y(y; a) / f_Y(y; a), \) and \( \iff \) denotes equivalence between before and after the notation.

III. EXTENSION OF DE BRUIJN’S IDENTITY

The main ingredient for establishing de Bruijn identity is that Gaussian density functions satisfy the heat equation. However, general pdfs do not satisfy the heat equation. Therefore, to extend de Bruijn identity to additive non-Gaussian noise channels, a general relationship between differentials of a pdf with respect to \( y \) and \( a \) of the form:
\[
\frac{d}{da} f_Y|X(y|x; a) = -\frac{1}{2} \frac{d}{dy} \left((y - x) f_Y|X(y|x; a)\right), \tag{10}
\]
will be exploited. Simple calculations show that (10) can be obtained directly from equation (1). The relationship (10) represents the key ingredient in establishing a connection between the derivative of differential entropy and posterior mean, as illustrated by the following theorem.

**Theorem 5:** Consider the channel (1), where \( X \) and \( W \) are arbitrary RVs, independent of each other. Given the following assumptions:

i) \( \frac{d}{dy} E_X \left[ f_Y|X(y|x; a) \right] = \frac{d}{dy} f_Y|X(y|x; a) \), \( \frac{d}{da} \int f_Y(y; a) log f_Y(y; a) dy = \int \frac{d}{da} (f_Y(y; a) log f_Y(y; a)) dy \).

ii) \( \lim_{y \to -\infty} X f_Y|X(y|x; a) = E_X \left[ \lim_{y \to -\infty} X f_Y|X(y|x; a) \right] \), \( \lim_{y \to -\infty} y^2 f_Y(y; a) = 0 \).

iii) \( \left| E_X \left[ f_Y|X(y|x; a) \right] \right| < \infty \),

the first-order derivative of differential entropy can be expressed as
\[
\frac{d}{da} h(Y) = \frac{1}{2a} \left\{ 1 - E_Y \left[ \frac{d}{dY} E_X|Y [X|Y] \right] \right\}, \tag{11}
\]
where \( E_X|Y [\cdot] \) denotes the posterior mean.

**Proof:** This theorem is proved using (10), integration by parts, and DCT. Since the proof is similar to the one in Theorem 6, the details of the proof are delegated to [11].

For additive non-Gaussian noise channels, differential entropy appears not to be expressible in terms of Fisher information. Instead, differential entropy is expressed by a function of the posterior mean as shown in Theorem 5. Fortunately, several noise distributions of interest in communication problems satisfy the required assumptions in Theorem 5 (e.g., Gaussian, gamma, exponential, chi-square with restrictions on parameters, Rayleigh, etc). Therefore, Theorem 5 is quite
powerful. In particular, if the posterior mean \( \mathbb{E}_{X|Y}[X|Y] \) is expressed by a polynomial function of \( Y \), e.g., \( X \) and \( W \) are independent Gaussian RVs in (1) or RVs belonging to the natural exponential family of distribution [12], then (11) can be expressed in simpler forms. We would like also to mention that a result similar in nature to Theorem 5 was reported before in [2] using a different approach. Similarly to [2], a series of specialized applications of Theorem 5 could be envisaged.

Now, we consider the second-order derivative of differential entropy. One interesting property of the second-order derivative of differential entropy is that it can always be expressed as a function of Fisher information.

**Theorem 6:** Consider the channel (1), where \( X \) and \( W \) are two arbitrary RVs, independent of each other. Given the following assumptions:

i) \( \frac{d^2}{dy^2} \mathbb{E}_X [f_{Y|X}(y|X; a)] = \mathbb{E}_X \left[ \frac{d^2}{dy^2} f_{Y|X}(y|X; a) \right] \)

\( \frac{d^2}{dy^2} \mathbb{E}_X [f_{Y|X}(y|X; a)] = \mathbb{E}_X \left[ \frac{d^2}{dy^2} f_{Y|X}(y|X; a) \right] \)

\( \frac{d^2}{dy^2} \int f_{Y|X}(y|X; a) \mathbb{E}_X \left[ \frac{d^2}{dy^2} f_{Y|X}(y|X; a) \right] dy \)

\( \lim_{y \to \pm \infty} \mathbb{E}_X [f_{Y|X}(y|X; a)] = \mathbb{E}_X \lim_{y \to \pm \infty} f_{Y|X}(y|X; a) \)

\( \lim_{y \to \pm \infty} f_{Y|X}(y|X; a) = 0 \)

\[ \mathbb{E}_X \left[ \frac{X^2 f_{Y|X}(y|X; a)}{(f_{Y|X}(y|X; a))^2} \right] < \infty \]

the following equality holds:

\[ \frac{d^2}{ds^2} h(Y) = - \int_{-\infty}^{\infty} \log f_Y(y; a) \frac{d}{dy} f_Y(y; a) dy \]

where \( h(Y) = \mathbb{E}_Y[\log f_Y(Y)] \), and \( S_Y = d \log f_Y(Y)/da \).

**Proof:** Since the first-order derivative of differential entropy with respect to \( a \) is expressed as

\[ \frac{d}{da} h(Y) = - \int_{-\infty}^{\infty} \log f_Y(y; a) \frac{d}{da} f_Y(y; a) dy \]

we obtain the second-order derivative of the differential entropy with respect to \( a \) as follows:

\[ \frac{d^2}{da^2} h(Y) = - \int_{-\infty}^{\infty} \log f_Y(y; a) \frac{d}{da} f_Y(y; a) dy \] (12)

From (10), we derive an additional relationship between the second-order differentials with respect to \( y \) and \( a \):

\[ \frac{d^2}{da^2} f_Y(y; a) = \frac{1}{4a^2} \left\{ \frac{d^2}{dy^2} \mathbb{E}_X [(y - X)^2 f_{Y|X}(y|x; a)] \right\} \]

\[ + \frac{d}{dy} \mathbb{E}_X [(y - X) f_{Y|X}(y|x; a)] \] (13)

Upon substituting \( (d^2/da^2) f_Y(y; a) \) from (13) into (12), the second term of equation (12) takes the expression:

\[ - \int_{-\infty}^{\infty} \log f_Y(y; a) \frac{d}{dy} f_Y(y; a) dy \]

\[ = - \frac{1}{4a^2} \int_{-\infty}^{\infty} \log f_Y(y; a) \frac{d^2}{dy^2} \mathbb{E}_X [(y - X)^2 f_{Y|X}(y|x; a)] dy \]

\[ - \frac{1}{4a^2} \int_{-\infty}^{\infty} \log f_Y(y; a) \frac{d}{dy} \mathbb{E}_X [(y - X) f_{Y|X}(y|x; a)] dy \]

The terms \( (D) \) and \( (E) \) are further simplified as follows:

\[ (D) = - \frac{1}{4a^2} \int_{-\infty}^{\infty} \log f_Y(y; a) \mathbb{E}_X [(y - X)^2 f_{Y|X}(y|x; a)] dy \]

\[ - \mathbb{E}_Y [S_Y(Y^2) \mathbb{E}_X [(y - X)^2|Y]] \]

\[ = - \frac{1}{4a^2} \mathbb{E}_Y \left[ \frac{d}{dy} S_Y(Y) \mathbb{E}_X [(y - X)^2|Y] \right] \] (14)

\[ (E) = \frac{1}{4a^2} \mathbb{E}_Y \left[ \frac{d}{dy} S_Y(Y) \mathbb{E}_X [(y - X)^2|Y] \right] \]

and the proof is completed. Additional details of the proof are delegated to [11].

Although we have not enumerated all possible pdfs for Theorems 5 and 6, most of the pdfs that present an exponentially-decaying factor in their expression satisfy the assumptions required in Theorems 5 and 6, since such a condition is sufficient for the required interchange between limit and integral.

**IV. APPLICATIONS**

**A. Applications in Information Theory**

In information theory, entropy power inequality (EPI) is one of the most important inequalities since it helps to assess channel capacity results under different circumstances, e.g., the capacity of Gaussian MIMO broadcast channel and the secrecy capacity of Gaussian wire-tap channel. Therefore, various versions of EPI such as Costa’s EPI [13] and the extremal entropy inequality [14] were proposed by different authors. Herein section, we will prove Costa’s EPI, a stronger version of a classical EPI, using Theorems 1, 4, and 6.

**Lemma 2 (Costa’s EPI [13]):** For an arbitrary but fixed RV \( X \) with a finite second-order moment, and a Gaussian RV \( W \) with zero mean and unit variance,

\[ \frac{d^2}{da^2} N(Y) \leq 0, \] (16)

where \( Y = X + \sqrt{a} W \), \( a \) is a positive real number, \( X \) and \( W \) are independent of each other, and the entropy power \( N(Y) \) is defined as \( N(Y) = (1/2\pi e) \exp(2h(Y)) \).

**Proof:** Since

\[ \frac{d^2}{da^2} N(Y) = N(Y) \left( J(Y) + 2 \frac{d^2}{da^2} h(Y) \right) \]
by Theorem 1 (de Bruijn identity), and \( N(Y) \geq 0 \), proving (16) is equivalent to proving the following inequality:
\[
J(Y)^2 + 2 \frac{d^2}{da^2} h(Y) \leq 0.
\]
(17)
Since \( W \) is a Gaussian RV, Fisher information inequality (FII) [9], which is proved using Theorems 4 and 6, is equivalently expressed as
\[
J(Y) \leq \frac{J(X)}{1 + aJ(X)}.
\]
(18)
Define the function \( l(a) \) as follows
\[
l(a) = - \frac{J(X)}{1 + aJ(X)} + J(Y).
\]
Since \( J(Y) \) converges to \( J(X) \) as \( a \) approaches zero, \( l(0) = 0 \), and the following inequality holds for an arbitrary but fixed random variable \( X \) and arbitrary small non-negative real-valued \( \epsilon \):
\[
l(\epsilon) - l(0) = - \frac{J(X)}{1 + \epsilon J(X)} + J(X + \sqrt{\epsilon}W) \leq 0.
\]
(19)
Therefore,
\[
\frac{d}{da}l(\epsilon) \bigg|_{\epsilon=0} \leq 0,
\]
for an arbitrary but fixed random variable \( X \).
Since the inequality in (20) holds for an arbitrary random variable \( X \), we define \( X = 1 + \sqrt{\epsilon}W \), where \( X \) is an arbitrary but fixed random variable, \( W \) is a Gaussian random variable whose variance is identical to the variance of \( W \), and \( \tilde{X} \), \( \tilde{W} \), and \( W \) are independent of one another. Then, the inequality in (20) is equivalent to the following inequalities:
\[
\left( \frac{J(\tilde{X} + \sqrt{\epsilon}W)}{1 + \epsilon J(\tilde{X} + \sqrt{\epsilon}W)} \right)^2 + \frac{d}{da} \left( \frac{J(\tilde{X} + \sqrt{\epsilon}W)}{1 + \epsilon J(\tilde{X} + \sqrt{\epsilon}W)} \right) \bigg|_{\epsilon=0} \leq 0
\]
\[
\Leftrightarrow \left( \frac{J(\tilde{X} + \sqrt{\epsilon}W)}{1 + \epsilon J(\tilde{X} + \sqrt{\epsilon}W)} \right)^2 + \frac{d}{da} \left( \frac{J(\tilde{X} + \sqrt{\epsilon}W)}{1 + \epsilon J(\tilde{X} + \sqrt{\epsilon}W)} \right) \bigg|_{\epsilon=0} \leq 0
\]
\[
\Leftrightarrow \left( \frac{J(\tilde{X} + \sqrt{\epsilon}W)}{1 + \epsilon J(\tilde{X} + \sqrt{\epsilon}W)} \right)^2 + \frac{d}{da} \left( \frac{J(\tilde{X} + \sqrt{\epsilon}W)}{1 + \epsilon J(\tilde{X} + \sqrt{\epsilon}W)} \right) \bigg|_{\epsilon=0} \leq 0
\]
\[
\Leftrightarrow J(\tilde{X} + \sqrt{\epsilon}W)^2 + \frac{d}{da} J(\tilde{X} + \sqrt{\epsilon}W) \leq 0,
\]
(21)
where \( \Leftrightarrow \) denotes the equivalence between before and after the notation.
Since random variable \( \tilde{X} \) is arbitrary and \( a \) is an arbitrary non-negative real-valued number in equation (21), the proof is completed.

B. Applications in Other Areas

Due to space limitations, we only detailed Costa’s EPI as an application of the proposed results. However, the proposed results can be further exploited in a variety of other applications information theory, statistical signal processing, and wireless communications. First, both CRLB and BCRLE can be derived based on Theorems 1, 2, 4, 5, and 6 (for details see [11]). A new lower bound, tighter than BCRLE, for the mean square error for a Bayesian estimator can be also derived based on Theorems 1 and 6 [11]. Second, since Theorem 5 is equivalent to Theorem 1 in [2], Theorem 5 can be used for applications such as generalized EXIT charts and power allocation in systems with parallel non-Gaussian noise channels as mentioned in [2]. Finally, by Theorem 4, we showed the equivalence among Stein, de Bruijn, and heat equation identities. Therefore, a broad range of problems (in probability, decision theory, Bayesian statistics, economics and graph theory [5] that were traditionally established/analyzed via Stein identity) could be analyzed via de Bruijn identity (in the light of Theorems 4, 5 and 6).

V. CONCLUSIONS

This paper mainly revealed three important information-estimation theoretic results. First, we proved that Stein identity is equivalent to de Bruijn identity. Second, the first-order and second-order derivatives of differential entropy with respect to the parameter \( a \) were expressed in terms of the posterior mean and Fisher information, respectively. Finally, several applications of the aforementioned results were provided. The proposed applications illustrate that the newly developed results are useful not only in information theory but also in the estimation theory field and other fields.

REFERENCES