On the Accuracy of the High SNR Approximation of the Differential Entropy of Signals in Additive Gaussian Noise

Ramy H. Gohary and Halim Yanikomeroglu

Department of Systems and Computer Engineering, Carleton University, Ottawa, ON, Canada

Abstract—One approach for analyzing the high signal-to-noise ratio (SNR) capacity of non-coherent wireless communication systems is to ignore the noise component of the received signal in the computation of its differential entropy. In this paper we consider the error incurred by this approximation when the transmitter and the receiver have one antenna each, and the noise has a Gaussian distribution. For a general instance of this case, we show that the approximation error decays as 1/SNR. In addition, we consider the special instance in which the received signal corresponds to a signal transmitted over a channel with additive Gaussian noise and a Gaussian fading coefficient. For that case, we provide an explicit expression for the second order term of the Taylor series expansion of the differential entropy. To circumvent the difficulty that arises in the direct computation of that term, we invoke Schwartz's inequality to obtain an efficiently computable bound on it, and we provide examples that illustrate the utility of this bound.

Index Terms—High-SNR non-coherent capacity, differential entropy, sum and product of random variables, Lebesgue dominated convergence.

I. INTRODUCTION

Evaluating the capacity of a given channel involves the search for input distributions that maximize the mutual information between the random variables representing the transmitted signals and the signals observed by the receivers. When the transmitted and received signals assume continuous values, the mutual information can be expressed as the difference between the non-conditional differential entropy of the received signals and their entropy conditioned on the transmitted signals. In many cases, including non-coherent communication ones [1]–[4], the received signals are contaminated by additive noise. In those cases, the conditional differential entropy of the random variable representing the received signal is typically straightforward to compute. In contrast, computing the non-conditional counterpart of this entropy is rather difficult.

The main difficulty that arises in computing and maximizing the non-conditional differential entropy of the received signal follows from the fact that the received signals are composed of the sum of random variables of potentially unknown distributions. This difficulty usually renders direct computation and maximization of the non-conditional entropy intractable. An exception of this is the case considered in [5]. Therein, a closed form expression is given for the optimal distribution that maximizes the sum of n independent random variables with finite supports and distributions that are symmetric around zero. In more general situations, bounds are typically used to gain insight into the input distribution that maximizes the non-conditional differential entropy. For instance, a lower bound on the differential entropy of the sum of independent random variables can be obtained using entropy power inequalities [6], [7]. In a complementary fashion, upper bounds on the entropy of the sum of random variables have been derived under various assumptions. For instance, a tight upper bound on the differential entropy of the sum of two dependent random variables was provided in [8] when the distributions of these variables are log-concave. In [9] and [10] the difference, and the ratio of the entropy of the sum and difference of independent identically distributed random variables were considered

using results from additive combinatorics. Upper bounds on the discrete and differential entropy of the sum of independent random variables are provided in [11]. Other upper bounds are derived from the observation that the Gaussian distribution maximizes the differential entropy [1], [3].

The difficulty of computing and maximizing the non-conditional entropy of the received signal arises in the evaluation of the non-coherent capacity of multiple-input multiple-output (MIMO) communication systems [1], [3] and in systems with phase noise [4]. To obtain asymptotic results in the case of non-coherent MIMO communication systems, the analysis is restricted to asymptotically low and high SNR scenarios. For instance, at high SNR, the noise contribution to the non-conditional differential entropy of the random matrix representing the received signal is ignored and that entropy is approximated by the differential entropy of the random matrix representing the product of the input signal matrix and the channel matrix. This approach enables an approximate expression for the non-coherent capacity to be obtained and the asymptotically optimal input distribution to be determined.

Given the central role of ignoring the noise contribution in computing the non-conditional differential entropy at high SNRs, in this paper we investigate the accuracy of this approximation. In particular, we assess how the approximation error decays with the increase of the SNR. To answer this question we begin by considering an expression of the differential entropy of the sum of an arbitrary and a zero mean Gaussian-distributed random variable with variance given by the inverse of the SNR. We derive explicit expressions for the first two non-constant terms of the Taylor series expansion of the differential entropy of the received signal and we use these expressions to show that the approximation error decays as 1/SNR. Next, we consider a non-coherent wireless communication system in which the signal component of the received signal can be expressed as the product of a random variable with some unknown distribution and a Gaussian-distributed random variable representing the channel fading coefficient. For that case, we provide an explicit expression for the second order term of the Taylor series expansion of the differential entropy, and to circumvent the difficulty that arises in the direct computation of that term, we invoke Schwartz's inequality to obtain an efficiently computable bound on it. Finally, we provide examples that illustrate the utility of the obtained bound.

II. PRELIMINARIES

In the forthcoming analysis we will frequently exchange the order of the operations of integration and taking limits. Such an exchange is valid if the conditions of Lebesgue's dominated convergence theorem are satisfied [12]. For completeness, we state this theorem.

Theorem 1 (Lebesgue's dominated convergence theorem):
Suppose that $\{f_n\}$ is a sequence of measurable functions on a measurable space $\mathcal{X}$ such that $f(x) = \lim_{n \to \infty} f_n(x)$ exists for

every $x \in \mathcal{X}$. If there is a Lebesgue integrable function g such that

$$\int_{\mathcal{X}} |g| d\mu < \infty, \text{ and } |f_n(x)| \leq g(x), \forall n = 1, 2, \ldots; \ x \in \mathcal{X}, \quad (1)$$

where μ denotes a positive measure on $\mathcal{X}$, then $\lim_{n \to \infty} \int_{\mathcal{X}} f_n(x) d\mu = \int_{\mathcal{X}} f d\mu$.

Proof: The proof follows from Fatou's Lemma; for details, see, e.g., [12, Chapter 1]. ∎

To use this theorem to exchange the order of integration and taking limits, we note that the space, $\mathcal{X}$, that we will encounter is the open interval $(-\infty, \infty)$ and the functions that we will encounter can be readily verified to be bounded by Lebesgue integrable functions. For instance, such functions will include the integrands in the convolution of two probability density functions (pdfs), which are, therefore, non-negative and their integration over $\mathcal{X}$ is unity. Hence, for the cases that we will consider, Theorem 1 can be readily applied to exchange the order of integration and taking limits; see [13]. In our analysis, the measure $d\mu = dx$, the sequence of functions $\{f_n(x)\}$ will be parametrized by $n = \sqrt{\text{SNR}}$, and the limiting functions, $f(x)$, will assume the form of the Dirac delta function, $\delta(x)$, the unit doublet function, $\delta'(x)$, or the unit triplet function, $\delta''(x)$.

III. THE TAYLOR SERIES EXPANSION OF THE SUM OF AN ARBITRARY AND A GAUSSIAN-DISTRIBUTED RANDOM VARIABLE

Let X be an arbitrary random variable with a twice differentiable pdf $p_X(\cdot)$ with support $[a, b]$, i.e., $p_X(x) > 0$ for $x \in [a, b]$ and zero otherwise. Let Z be Gaussian-distributed with zero-mean and unit variance, that is, $p_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$.

In a point-to-point communication channel with additive Gaussian noise, the received signal can be expressed as $Y = X + tZ$, where $t = \frac{1}{\sqrt{\text{SNR}}}$. Our goal is to examine the Taylor series expansion of the (non-conditional) differential entropy of Y. For notational convenience, we will denote this entropy by $h_t(Y)$. For any given $t > 0$, the pdf of tZ is given by $\frac{1}{\sqrt{2\pi}t} e^{-\frac{z^2}{2t^2}}$, which implies that the pdf of Y is

$$p_Y(y; t) = \frac{1}{\sqrt{2\pi}t} \int_{-\infty}^{\infty} p_X(y - u) e^{-\frac{u^2}{2t^2}} du. \quad (2)$$

Notice that in this notation, t parametrizes a family of pdfs. Using this notation, we can write

$$h_t(Y) = -\int_{-\infty}^{\infty} p_Y(y; t) \log p_Y(y; t) dy. \quad (3)$$

Evaluating $h_t(Y)$ directly seems intractable. To circumvent this difficulty, we assume that $p_X(\cdot)$ is such that $h_t(Y)$ is analytic in t, which implies that all the derivatives of $h_t(Y)$ with respect to t exist and are finite. A class of practical instances in which $p_X(\cdot)$ satisfies this condition is identified in Section IV below. When this condition is satisfied, the Taylor series expansion of $h_t(Y)$ around 0 converges for all $t < 1$. To expand $h_t(Y)$, we begin by noting that $p_Y(y; 0) = p_X(y)$ and thereby $h_0(Y) = h(X)$. Hence, the Taylor series expansion can be expressed as

$$h_t(Y) = h(X) + \frac{dh_t(Y)}{dt}\Big|_{t=0} t + \frac{d^2 h_t(Y)}{dt^2}\Big|_{t=0} \frac{t^2}{2!} + \mathcal{O}(t^3). \quad (4)$$

To evaluate $\frac{dh_t(Y)}{dt}\Big|_{t=0}$ and $\frac{d^2 h_t(Y)}{dt^2}\Big|_{t=0}$, we use (2) and (3).

A. Evaluating the First Order Term

Using (3), we have

$$\frac{dh_t(Y)}{dt} = -\int_{-\infty}^{\infty} (1 + \log p_Y(y; t)) \frac{\partial p_Y(y; t)}{\partial t} dy. \quad (5)$$

Since herein we consider cases in which $h_t(Y)$ is analytic, it is continuous and evaluating $\frac{dh_t(Y)}{dt}\Big|_{t=0}$ is equivalent to evaluating $\lim_{t \searrow 0} \frac{dh_t(Y)}{dt}$. Using (2) and invoking Theorem 1, we can write

$$\frac{dh_t(Y)}{dt}\Big|_{t=0} = -\int_{-\infty}^{\infty} \left(1 + \lim_{t \searrow 0} \log p_Y(y; t)\right) \lim_{t \searrow 0} \frac{\partial p_Y(y; t)}{\partial t} dy$$
$$= -\int_{-\infty}^{\infty} \left(1 + \log p_X(y)\right) \lim_{t \searrow 0} \frac{\partial p_Y(y; t)}{\partial t} dy. \quad (6)$$

Using (2), we have

$$\frac{\partial p_Y(y; t)}{\partial t} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} p_X(y - u) \left(\frac{u^2}{t^4} - \frac{1}{t^2}\right) e^{-\frac{u^2}{2t^2}} du. \quad (7)$$

To evaluate $\lim_{t \searrow 0} \frac{\partial p_Y(y; t)}{\partial t}$, we use the following result:

Lemma 1:

$$\lim_{t \searrow 0} \frac{1}{t^2} e^{-\frac{u^2}{2t^2}} = \frac{1}{2} \lim_{t \searrow 0} \frac{u^2}{t^4} e^{-\frac{u^2}{2t^2}}. \quad (8)$$

Proof: The proof of this result follows from applying l'Hôpital's rule to the left hand side of (8) and the fact that the derivatives of the numerator and the denominator with respect to t are given by $\frac{u^2}{t^3} e^{-\frac{u^2}{2t^2}}$ and $2t$, respectively. ∎

Using Lemma 1 in (7) and taking the limit as $t \searrow 0$ yields

$$\lim_{t \searrow 0} \frac{\partial p_Y(y; t)}{\partial t} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} p_X(y - u) \lim_{t \searrow 0} \frac{1}{t^2} e^{-\frac{u^2}{2t^2}} du. \quad (9)$$

To evaluate the above limit, we will prove the following result:

Lemma 2:

$$\lim_{t \searrow 0} \frac{1}{t^2} e^{-\frac{u^2}{2t^2}} U(u) = -\delta'(u), \quad (10)$$

where $U(u)$ is the unit step function.

Proof: See Appendix A. ∎

We will now use Lemma 2 to show that $\frac{dh_t(Y)}{dt}\Big|_{t=0} = 0$. Since $\delta'(u)$ is an odd function, it follows that $\lim_{t \searrow 0} \frac{1}{t^2} e^{-\frac{u^2}{2t^2}} U(-u) = \delta'(u)$. Now, assuming that $p_X(x) < \infty$ for all $x \in (-\infty, \infty)$, the integration on the right hand side of (9) can be written as $\int_{-\infty}^{\infty} p_X(y - u) \lim_{t \searrow 0} \frac{1}{t^2} e^{-\frac{u^2}{2t^2}} (U(u) + U(-u)) du$, implying that

$$\lim_{t \searrow 0} \frac{\partial p_Y(y; t)}{\partial t} = 0. \quad (11)$$

Substituting this result in (6) yields

$$\frac{dh_t(Y)}{dt}\Big|_{t=0} = 0. \quad (12)$$

B. Evaluating the Second Order Term

The coefficient of the second order term is given by $\lim_{t \searrow 0} \frac{d^2 h_t(Y)}{dt^2}$. Differentiating both sides of (5), it can be readily verified that

$$\frac{d^2 h_t(Y)}{dt^2} = -\int_{-\infty}^{\infty} \left(\frac{1}{p_Y(y; t)} \left(\frac{\partial p_Y(y; t)}{\partial t}\right)^2 \right.$$
$$\left. + (1 + \log p_Y(y; t)) \frac{\partial^2 p_Y(y; t)}{\partial t^2}\right) dy. \quad (13)$$

Taking the limit as $t \searrow 0$ and using (11) along with the fact that the support of $p_X(\cdot)$ is $[a, b]$ yields

$$\lim_{t \searrow 0} \frac{d^2 h_t(Y)}{dt^2} = -\int_a^b (1 + \log p_X(y)) \lim_{t \searrow 0} \frac{\partial^2 p_Y(y; t)}{\partial t^2} dy.$$

Hence, to evaluate the integration of the second term in (13), we need to evaluate $\lim_{t\searrow 0}\frac{\partial^2 p_Y(y;t)}{\partial t^2}$. Differentiating both sides of (7) with respect to t and taking the limit as $t\searrow 0$ yields

$$\lim_{t\searrow 0}\frac{\partial^2 p_Y(y;t)}{\partial t^2} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}p_X(y-u)$$
$$\times \lim_{t\searrow 0}\Big(\frac{u^4}{t^7}-5\frac{u^2}{t^5}+\frac{2}{t^3}\Big)e^{-\frac{u^2}{2t^2}}\,du \quad (14)$$

The evaluation of this integral is simplified by the following lemma.

Lemma 3:

$$\lim_{t\searrow 0}\frac{1}{t^3}e^{-\frac{u^2}{2t^2}} = \frac{1}{3}\lim_{t\searrow 0}\frac{u^2}{t^5}e^{-\frac{u^2}{2t^2}}$$
$$= \frac{1}{15}\lim_{t\searrow 0}\frac{u^4}{t^7}e^{-\frac{u^2}{2t^2}}.$$

Proof: The proof of this result uses a technique similar to the one used in the proof of Lemma 1, and is omitted for brevity. ∎

Using Lemma 3, the limit in (14) yields

$$\lim_{t\searrow 0}\frac{\partial^2 p_Y(y;t)}{\partial t^2} = \sqrt{\frac{2}{\pi}}\int_{-\infty}^{\infty}p_X(y-u)\lim_{t\searrow 0}\frac{1}{t^3}e^{-\frac{u^2}{2t^2}}\,du. \quad (15)$$

To evaluate the limit in the integrand in (15), we prove the following lemma.

Lemma 4:

$$\sqrt{\frac{2}{\pi}}\lim_{t\searrow 0}\frac{1}{t^3}e^{-\frac{u^2}{2t^2}} = \delta''(u). \quad (16)$$

Proof: See Appendix B. ∎

Using Lemma 4 in (15) and using the fact that, for any twice differentiable function $f(x)$, $(f*\delta'')(x) = f''(x)$, where '$*$' denotes the convolution operation, yield

$$\lim_{t\searrow 0}\frac{\partial^2 p_Y(y;t)}{\partial t^2} = \frac{d^2 p_X(y)}{dy^2}. \quad (17)$$

Using integration by parts, we can write

$$\lim_{t\searrow 0}\frac{d^2 h_t(Y)}{dt^2} = -\int_a^b\big(1+\log p_X(y)\big)\frac{d^2 p_X(y)}{dy^2}dy$$
$$= \int_a^b\frac{1}{p_X(y)}\Big(\frac{dp_X(y)}{dy}\Big)^2 dy$$
$$- \big(1+\log p_X(y)\big)\frac{dp_X(y)}{dy}\Big|_a^b \quad (18)$$

C. The Leading Terms of Taylor Series—The General Case

Substituting from (12) and (18) in (4) yields, for any SNR strictly greater than 0 dB, t is strictly less than 1 and

$$h_t(Y) = h(X) + \frac{t^2}{2}\Big(\int_a^b\frac{1}{p_X(y)}\Big(\frac{dp_X(y)}{dy}\Big)^2 dy$$
$$- \big(1+\log p_X(y)\big)\frac{dp_X(y)}{dy}\Big|_a^b\Big) + \mathcal{O}(t^3). \quad (19)$$

IV. A SPECIAL CASE

A case of special interest that arises in non-coherent communication scenarios is when $X = RS$, where R and S are two independent random variables. For instance, in non-coherent communications, R represents the input signal and S represents the channel. When S is Gaussian-distributed the pdf of X possesses some favourable properties that can be exploited to simplify the expression in (19) and to yield further insight into information-theoretic high SNR analysis. We have the following result:

Lemma 5: Let $X = RS$, where R and S are random variables with a joint pdf $p_{R,S}(\cdot,\cdot)$, then the pdf of X is given by

$$p_X(x) = \int_0^{\infty}\frac{1}{r}\Big(p_{R,S}\Big(r,\frac{x}{r}\Big)+p_{R,S}\Big(-r,-\frac{x}{r}\Big)\Big)dr. \quad (20)$$

In particular, if S is statistically independent of R and is Gaussian-distributed with zero-mean and unit variance,

$$p_X(x) = \frac{1}{\sqrt{2\pi}}\int_0^{\infty}\frac{1}{r}\big(p_R(r)+p_R(-r)\big)e^{-\frac{x^2}{2r^2}}\,dr. \quad (21)$$

Proof: This lemma can be proved by considering the cumulative distribution function of X, $F_X(x) = \Pr\{RS\le x\}$. Expressing this probability in terms of $p_{R,S}(\cdot,\cdot)$ and taking the derivative with respect to r yields (20); see e.g., [14]. ∎

Immediate from (21) is that, when X is the product of two independent random variables, one of which is Gaussian, the support of $p_X(\cdot)$ is the entire real line $(-\infty,\infty)$. Hence, in (19) $a=-\infty$ and $b=\infty$ and from (21) $p_X(\infty) = p_X(-\infty) = 0$. Differentiating $p_X(x)$ in (21) with respect to x and using Theorem 1 to compute the limit as x tends to $\pm\infty$, it can be readily verified that $\lim_{x\to\pm\infty}\frac{dp_X(x)}{dx} = 0$. Hence, in this case (19) can be expressed as

$$h_t(Y) = h(X) + \frac{t^2}{2}\Big(\int_{-\infty}^{\infty}\frac{1}{p_X(y)}\Big(\frac{dp_X(y)}{dy}\Big)^2 dy$$
$$- \log p_X(y)\frac{dp_X(y)}{dy}\Big|_{-\infty}^{\infty}\Big) + \mathcal{O}(t^3).$$

To further simplify this expression, we will show that $\lim_{b\to\infty}\frac{dp_X(b)}{dx}\log p_X(b) = 0$. Since $p_X(b)\searrow 0$ as $b\to\infty$, we have $(1-p_X(b))<1$. Hence, Taylor series expansion can be used to write $\log p_X(b) = \log\big(1-(1-p_X(b))\big) = \sum_{n=1}^{\infty}\frac{(1-p_X(b))^n}{n}$. Using this, we have

$$\lim_{b\to\infty}\frac{dp_X(b)}{dx}\log p_X(b) = \sum_{n=1}^{\infty}\lim_{b\to\infty}\frac{dp_X(b)}{dx}\frac{(1-p_X(b))^n}{n} = 0.$$

A similar argument shows that $\lim_{a\to-\infty}\frac{dp_X(a)}{dx}\log p_X(a) = 0$. These results conform to the convention of setting $0\log 0 = 0$ [6] and imply that, for $p_X(x)$ in (21),

$$h_t(Y) = h(X) + \frac{t^2}{2}\int_{-\infty}^{\infty}\frac{1}{p_X(y)}\Big(\frac{dp_X(y)}{dy}\Big)^2 dy + \mathcal{O}(t^3). \quad (22)$$

A. An Upper Bound on the Second Order Error Term

Evaluating the coefficient of $t^2/2$ in (22) directly can be intractable for many distributions of R. To circumvent this difficulty, in this section we will provide a bound $\int_{-\infty}^{\infty}\frac{1}{p_X(y)}\Big(\frac{dp_X(y)}{dy}\Big)^2 dy$. When $p_X(x)$ is given by the expression in (21), the integrand can be expressed as

$$\frac{1}{p_X(y)}\Big(\frac{dp_X(y)}{dy}\Big)^2$$
$$= \frac{y^2}{\sqrt{2\pi}}\frac{\Big(\int_0^{\infty}\frac{1}{r^3}\big(p_R(r)+p_R(-r)\big)e^{-\frac{y^2}{2r^2}}\,dr\Big)^2}{\int_0^{\infty}\frac{1}{r}\big(p_R(r)+p_R(-r)\big)e^{-\frac{y^2}{2r^2}}\,dr}.$$

Using Schwartz's inequality, we have

$$\frac{1}{p_X(y)}\Big(\frac{dp_X(y)}{dy}\Big)^2 \le \frac{y^2}{\sqrt{2\pi}}\int_0^{\infty}\frac{1}{r^5}\big(p_R(r)+p_R(-r)\big)e^{-\frac{y^2}{2r^2}}\,dr,$$

and equality holds if r assumes a deterministic value in $[0,\infty)$. Substituting this result, the coefficient of $t^2/2$ in (22) can be readily seen to be bounded by

$$\int_{-\infty}^{\infty} \frac{y^2}{\sqrt{2\pi}} \int_0^{\infty} \frac{1}{r^5} \big(p_R(r) + p_R(-r) \big) e^{-\frac{y^2}{2r^2}} \, dr \, dy =$$
$$\int_0^{\infty} \frac{1}{r^2} \big(p_R(r) + p_R(-r) \big) dr. \quad (23)$$

Below, we provide a few examples to illustrate the error, $|h_t(Y) - h(X)|$, in approximating $h_t(Y)$ with $h(X)$.

V. EXAMPLES

Example 1: Let the input signal, R, be antipodal, i.e., $R = \pm a$ with equal probability. In this case, $p_R(r) = \frac{1}{2}\big(\delta(r+a) + \delta(r-a)\big)$, which yields

$$|h_t(Y) - h(X)| \leq \frac{t^2}{2a^2} + \mathcal{O}(t^3).$$

In this case, it can be readily verified that the bound derived from Schwartz's inequality is tight. □

Example 2: If R is χ-distributed with m degrees of freedom, its pdf is given by

$$p_R(r) = \frac{2}{2^{m/2} \sigma^m \Gamma(m/2)} r^{m-1} e^{-r^2/2\sigma^2} U(r), \quad (24)$$

where $\sigma > 0$, and $\Gamma(\cdot)$ is the standard Gamma function. This is the distribution of the square root of the sum of the squares of m independent zero mean Gaussian random variables with variance σ^2 each. Hence, for $m = 2$, the χ-distribution reduces to the Rayleigh distribution and, for $m = 3$, it reduces to the Maxwell distribution [14].

For $m \geq 3$, the right hand side of (23) can be readily calculated to yield that the error, $|h_t(Y) - h(X)|$, can be bounded by

$$|h_t(Y) - h(X)| \leq \left(\frac{2^{-(m/2-1)} \rho_0}{\sigma^2 \Gamma(m/2)} (m-4)(m-6) \cdots \right) t^2 + \mathcal{O}(t^3),$$
$$(25)$$

where $\rho_0 = \begin{cases} 1 & m \text{ even} \\ \sqrt{\frac{\pi}{2}} & m \text{ odd} \end{cases}$.

The coefficient of t^2 in (25) goes to zero as m goes to infinity. □

VI. CONCLUSION

In this paper we analyzed the dominant error term resulting from ignoring the noise component in computing the high-SNR entropy of the received signal of non-coherent single-input single-output communication systems. We have shown that this error decays with the inverse SNR and we provided an explicit expression and an upper bound for the dominant error term for the case in which the input signal is transmitted over a channel with Gaussian fading coefficients. Our analysis is supported by examples of input signal with various input distributions.

APPENDIX A
PROOF OF LEMMA 2

To prove Lemma 2 we use the χ-distribution given in (24) with $m = 2$, which corresponds to the case of a Rayleigh-distributed random variable. In this case,

$$p_\chi(x) = \frac{1}{\sigma^2} x e^{-x^2/2\sigma^2} U(x).$$

Taking the limit as $\sigma \searrow 0$ corresponds to the limiting case at which χ approaches a deterministic value at zero. Since $\int_{-\infty}^{\infty} p_\chi(x) dx = 1$, we can write [15, Appendix A]

$$\lim_{\sigma^2 \searrow 0} \frac{1}{\sigma^2} x e^{-x^2/2\sigma^2} U(x) = \delta(x).$$

Hence, $\lim_{\sigma^2 \searrow 0} \frac{1}{\sigma^2} e^{-x^2/2\sigma^2} U(x) = \frac{1}{x} \delta(x)$. Using the identity,

$$x\delta'(x) = -\delta(x)$$

yields the statement of the lemma.

APPENDIX B
PROOF OF LEMMA 4

The proof of this lemma uses an alternate representation of the Dirac delta function [15]. In particular, since the pdf of a Gaussian distributed random variable converges to the Dirac delta function as the variance goes to zero, we have

$$\delta(x) = \frac{1}{\sqrt{2\pi}} \lim_{\epsilon \searrow 0} \frac{1}{\epsilon} e^{-\frac{x^2}{2\epsilon^2}}.$$

A difficulty in proving this lemma arises because the order of differentiation and taking the limit cannot be exchanged unless uniform convergence to the limiting function is guaranteed [12]. To avoid this difficulty, we note that using l'Hôpital's rule, the limit on the right hand side can be expressed as $\lim_{\epsilon \searrow 0} \frac{x^2}{\epsilon^3} e^{-\frac{x^2}{2\epsilon^2}}$. Now, using the identity that $2\delta(x) = x^2 \delta''(x)$, we have

$$\sqrt{\frac{2}{\pi}} \lim_{\epsilon \searrow 0} \frac{1}{\epsilon^3} e^{-\frac{x^2}{2\epsilon^2}} = \delta''(x), \quad (26)$$

which completes the proof of the lemma.

REFERENCES

[1] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, pp. 359–383, Feb. 2002.
[2] R. H. Gohary and T. N. Davidson, "Non-coherent MIMO communication: Grassmannian constellations and efficient detection," *IEEE Trans. Inf. Theory*, vol. 55, pp. 1176–1205, Mar. 2009.
[3] W. Yang, G. Durisi, and E. Riegler, "On the capacity of large-MIMO block-fading channels," *IEEE J. Select. Areas Commun.*, vol. 31, pp. 117–132, Feb. 2013.
[4] G. Durisi, "On the capacity of the block-memoryless phase-noise channel," *IEEE Commun. Lett.*, vol. 16, pp. 1157–1160, Aug. 2012.
[5] E. Ordentlich, "Maximizing the entropy of a sum of independent bounded random variables," *IEEE Trans. Inf. Theory*, vol. 52, pp. 2176–2181, May 2006.
[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[7] M. Madiman and F. Ghassemi, "The entropy power of a sum is fractionally superadditive," in *Proc. IEEE Int. Symp. Inf. Theory*, (Seoul), pp. 295–298, June 2009.
[8] T. M. Cover and Z. Zhang, "On the maximum entropy of the sum of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. 40, pp. 1244–1246, Apr. 1994.
[9] A. Lapidoth and G. Pete, "On the entropy of the sum and of the difference of independent random variables," in *Proc. IEEE Conven. Elect. Electron. Eng. Israel (IEEEI)*, (Eilat, Israel), pp. 623–625, Nov. 2008.
[10] M. Madiman and I. Kontoyiannis, "The entropies of the sum and the difference of two IID random variables are not too different," in *Proc. IEEE Int. Symp. Inf. Theory*, (Austin, Texas), pp. 1369–1372, June 2010.
[11] M. Madiman, "On the entropy of sums," in *Proc. IEEE Inform. Theory Wkshp. (ITW)*, (Porto), pp. 303–307, May 2008.
[12] W. Rudin, *Real and Complex Analysis*. New York: McGraw-Hill, 3rd ed., 1987.
[13] A. Papoulis, *The Fourier Integral and its Applications*. New York: McGraw-Hill Co., 1962.
[14] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 2nd ed., 1984.
[15] S. Gasiorowicz, *Quantum Physics*. New York: Wiley, 1974.