

A New Tight Upper Bound on the Entropy of Sums

Jihad Fahs * and Ibrahim Abou-Faycal

Received: 18 August 2015; Accepted: 14 December 2015; Published: 19 December 2015
Academic Editor: Raúl Alcaraz MartínezDepartment of Electrical and Computer Engineering, American University of Beirut,
Beirut 1107 2020, Lebanon; Ibrahim.Abou-Faycal@aub.edu.lb

* Correspondence: jjf03@mail.aub.edu; Tel.: +961-79-140666

Abstract: We consider the independent sum of a given random variable with a Gaussian variable and an infinitely divisible one. We find a novel tight upper bound on the entropy of the sum which still holds when the variable possibly has an infinite second moment. The proven bound has several implications on both information theoretic problems and infinitely divisible noise channels' transmission rates.

Keywords: entropy of sums; upper bound; infinite variance; infinitely divisible; differential entropy

1. Introduction

Information inequalities have been investigated since the foundation of information theory. Two such important ones are due to Shannon [1]:

- The first one is an upper bound on the entropy of Random Variables (RV)s having a finite second moment by virtue of the fact that Gaussian distributions maximize entropy under a second moment constraint (the (differential) entropy $h(Y)$ of a random variable Y having a probability density function $p(y)$ is defined as:

$$h(Y) = - \int_{-\infty}^{+\infty} p(y) \ln p(y) dy,$$

whenever the integral exists). For any RVs X and Z having respectively finite variances σ_X^2 and σ_Z^2 , we have

$$h(X + Z) \leq \frac{1}{2} \ln 2\pi e (\sigma_X^2 + \sigma_Z^2). \quad (1)$$

- The second one is a lower bound on the entropy of independent sums of RVs and commonly known as the Entropy Power Inequality (EPI). The EPI states that given two real independent RVs X, Z such that $h(X), h(Z)$ and $h(X + Z)$ exist, then (Corollary 3, [2])

$$N(X + Z) \geq N(X) + N(Z), \quad (2)$$

where $N(X)$ is the *entropy power* of X and is equal to

$$N(X) = \frac{1}{2\pi e} e^{2h(X)}.$$

While Shannon proposed Equation (2) and proved it locally around the normal distribution, Stam [3] was the first to prove this result in general followed by Blachman [4] in what is considered to be a simplified proof. The proof was done via the usage of two information identities:

- 1- The Fisher Information Inequality (FII): Let X and Z be two independent RVs such that the respective Fisher informations $J(X)$ and $J(Z)$ exist (the Fisher information $J(Y)$ of a random variable Y having a probability density function $p(y)$ is defined as:

$$J(Y) = \int_{-\infty}^{+\infty} \frac{1}{p(y)} p'^2(y) dy,$$

whenever the derivative and the integral exist). Then

$$\frac{1}{J(X + Z)} \geq \frac{1}{J(X)} + \frac{1}{J(Z)}. \tag{3}$$

- 2- The de Bruijn’s identity: For any $\epsilon > 0$,

$$\frac{d}{d\epsilon} h(X + \sqrt{\epsilon}Z) = \frac{\sigma^2}{2} J(X + \sqrt{\epsilon}Z), \tag{4}$$

where Z is a Gaussian RV with mean zero and variance σ^2 independent of X .

Rioul proved that the de Bruijn’s identity holds at $\epsilon = 0^+$ for any finite-variance RV Z (Proposition 7, p. 39, [5]).

The remarkable similarity between Equations (2) and (3) was pointed out in Stam’s paper [3] who in addition, related the entropy power and the Fisher information by an “uncertainty principle-type” relation:

$$N(X)J(X) \geq 1, \tag{5}$$

which is commonly known as the Isoperimetric Inequality for Entropies (IIE) (Theorem 16, [6]). Interestingly, equality holds in Equation (5) whenever X is Gaussian distributed and in Equations (1)–(3) whenever X and Z are independent Gaussian.

When it comes to upper bounds, a bound on the discrete entropy of the sum exists [7]:

$$H(X + Z) \leq H(X) + H(Z). \tag{6}$$

In addition, several identities involving discrete entropy of sums were shown in [8,9] using the Plünnecke-Ruzsa sumset theory and its analogy to Shannon entropy. Except for Equation (1), that holds for finite variance RVs, the differential entropy inequalities provided in some sense a lower bound on the entropy of sums of independent RVs. Equation (6) does not always hold for differential entropies, and unless the variance is finite, if we start with two RVs X and Z having respectively finite differential entropies $h(X)$ and $h(Z)$, one does not have a clear idea on how much the growth of $h(X + Z)$ will be. The authors in [10] deferred this to the fact that discrete entropy has a functional submodularity property which is not the case for differential entropy. Nevertheless, the authors were able to derive various useful inequalities. Madiman [11] used basic information theoretic relations to prove the submodularity of the entropy of independent sums and found accordingly upper bounds on the discrete and differential entropy of sums. Though, in its general form, the problem of upper bounding the differential entropy of independent sums is not always possible (proposition 4, [2]), several results are known in particular settings. Cover *et al.* [12] solved the problem of maximizing the differential entropy of the sum of dependent RVs having the same marginal log-concave densities. In [13], Ordentlich found the maximizing probability distribution for the differential entropy of the independent sum of n finitely supported symmetric RVs. For “sufficiently convex” probability distributions, an interesting reverse EPI was proven to hold in (Theorem 1.1, p. 63, [14]).

In this study, we find a tight upper bound on the (differential) entropy of the independent sum of a RV X not necessarily having a finite variance with an infinitely divisible variable having a Gaussian component. The proof is based on the infinite divisibility property with the application of the FII,

de Bruijn’s identity, a proven concavity result of the differential entropy and a novel “de Bruijn type” identity. We use convolutions along small perturbations to upper bound some relevant information theoretic quantities as done in [15] where some moment constraints were imposed on X which is not the case here. The novel bound presented in this paper is, for example, useful when studying Gaussian channels or when the additive noise is modeled as a combination of Gaussian and Poisson variables [16]. It has several implications which are listed in Section 2 and can be possibly used for variables with infinite second moments. Even when the second moment of X is finite, in some cases our bound can be tighter than Equation (1).

2. Main Result

We consider the following scalar RV,

$$Y = X + Z, \tag{7}$$

where $Z = Z_1 + Z_2$ is a composite infinitely divisible RV that is independent of X and where:

- $Z_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ is a Gaussian RV with mean μ_1 and positive variance σ_1^2 .
- Z_2 is an infinitely divisible RV with mean μ_2 and finite (possibly zero) variance σ_2^2 that is independent of Z_1 .

We note that since Z_1 is absolutely continuous with bounded Probability Density Function (PDF), then so are $Z = Z_1 + Z_2$ and $Y = X + Z$ for any RV X [17]. In addition, we define the set \mathcal{L} of distribution functions $F(x)$ that have a finite logarithmic moment:

$$\mathcal{L} = \left\{ F : \int \ln(1 + |X|) dF(x) < +\infty \right\}.$$

Let $E[\cdot]$ denote the expectation operator. Using the identity $\ln(1 + |x + n|) \leq \ln(1 + |x|) + \ln(1 + |n|)$,

$$E[\ln(1 + |Y|)] = E[\ln(1 + |X + Z|)] \leq E[\ln(1 + |X|)] + E[\ln(1 + |Z|)]. \tag{8}$$

Since Z has a bounded PDF with finite variance then necessarily it has a finite logarithmic moment, and by virtue of Equation (8), if $X \in \mathcal{L}$ then $Y \in \mathcal{L}$.

Under this finite logarithmic constraint, the differential entropy of X is well defined and is such that $-\infty \leq h(X) < +\infty$ (Proposition 1, [5]) and that of Y exists and is finite (Lemma 1, [5]). Also, when $X \in \mathcal{L}$, the identity $I(X + Z; Z) = h(X + Z) - h(X)$ always holds (Lemma 1, [5]).

The main result of this work stated in Theorem 1 is a novel upper bound on $h(Y)$ whenever $X \in \mathcal{L}$ with finite differential entropy $h(X)$ and finite Fisher information $J(X)$.

Theorem 1. *Let $X \in \mathcal{L}$ having finite $h(X)$ and $J(X)$. The differential entropy of $X + Z$ is upper bounded by:*

$$h(X + Z) \leq h(X) + \frac{1}{2} \ln(1 + \sigma_1^2 J(X)) + \sigma_2^2 \min \left\{ \sup_x \frac{D(p_X(u-x) \| p_X(u))}{x^2}; \frac{1}{2\sigma_1^2} \right\}, \tag{9}$$

where the Kullback-Leibler divergence between probability distributions p and q is denoted $D(p \| q)$. In the case where $Z \sim \mathcal{N}(\mu_1, \sigma_1^2)$, i.e. $\sigma_2^2 = 0$, we have

$$h(X + Z) \leq h(X) + \frac{1}{2} \ln(1 + \sigma_1^2 J(X)), \tag{10}$$

and equality holds if and only if both X and Z are Gaussian distributed.

We defer the proof of Theorem 1 to Section 3. The rest of this section is dedicated to the implications of identity Equation (10) which are fivefold:

- 1- While the usefulness of this upper bound is clear for RVs X having an infinite second moment for which Equation (1) fails, it can in some cases, present a tighter upper bound than the one provided by Shannon for finite second moment variables X . This is the case, for example, when $Z \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and X is a RV having the following PDF:

$$p_X(x) = \begin{cases} f(x+a) & -1-a \leq x \leq 1-a \\ f(x-a) & -1+a \leq x \leq 1+a, \end{cases}$$

for some $a > 0$ and where

$$f(x) = \begin{cases} \frac{3}{4}(1+x)^2 & -1 \leq x \leq 0 \\ \frac{3}{4}(1-x)^2 & 0 < x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

The involved quantities related to X are easily computed and they evaluate to the following: $E[X] = 0$, $E[X^2] = a^2 + \frac{1}{10}$, $J(X) = 12$ and $h(X) = \ln \frac{4}{3} + \frac{2}{3}$, for which Equation (10) becomes

$$h(X+Z) \leq h(X) + \frac{1}{2} \ln(1 + \sigma_1^2 J(X)) = \ln \frac{4}{3} + \frac{2}{3} + \frac{1}{2} \ln(1 + 12\sigma_1^2), \tag{11}$$

and Equation (1) becomes

$$h(X+Z) \leq \frac{1}{2} \ln 2\pi e (\sigma_X^2 + \sigma_1^2) = \frac{1}{2} \ln 2\pi e + \frac{1}{2} \ln \left(a^2 + \frac{1}{10} + \sigma_1^2 \right). \tag{12}$$

Comparing Equations (11) and (12), it can be seen that our upper bound is independent of a whereas the Shannon bound increases to ∞ and gets looser as a increases. This is explained by the fact that our bound is location independent and depends only on the PDF of X whereas the Shannon bound is location dependent via the variance of X .

- 2- Theorem 1 gives an analytical bound on the change in the transmission rates of the linear Gaussian channel function of an input scaling operation. In fact, let X be a RV satisfying the conditions of Theorem 1 and $Z \sim \mathcal{N}(\mu_1, \sigma_1^2)$. Then aX satisfies similar conditions for some positive scalar a . Hence

$$h(aX+Z) \leq h(aX) + \frac{1}{2} \ln(1 + \sigma_1^2 J(aX)) = h(X) + \ln a + \frac{1}{2} \ln \left(1 + \frac{\sigma_1^2}{a^2} J(X) \right),$$

where we used the fact that $h(aX) = h(X) + \ln a$ and $J(aX) = \frac{1}{a^2} J(X)$. Subtracting $h(Z)$ from both sides of the equation gives

$$I(aX+Z; X) - I(X+Z; X) \leq \frac{1}{2} \ln(a^2 + \sigma_1^2 J(X)).$$

It can be seen that the variation in the transmissions rates is bounded by a logarithmically growing function in a . This is a known behavior of the optimal transmission rates that are achieved by Gaussian inputs. A similar behavior is observed whenever $Z = Z_1 + Z_2$.

- 3- If the EPI is regarded as being a lower bound on the entropy of sums, Equation (10) can be considered as its upper bound counterpart whenever one of the variables is Gaussian. In fact using both of these inequalities gives:

$$N(X) + N(Z) \leq N(Y) \leq N(X) + N(Z) [N(X)J(X)]. \tag{13}$$

It can be seen that the sandwich bound is efficient whenever the IIE in Equation (5) evaluated for the variable X is close to its lower bound of 1.

- 4- The result of Theorem 1 is more powerful than the IIE in Equation (5). Indeed, using the fact that $h(Z) \leq h(X + Z)$, inequality Equation (10) gives the looser inequality:

$$h(Z) \leq h(X) + \frac{1}{2} \ln \left(1 + \sigma^2 J(X) \right),$$

which implies that

$$N(X)J(X) \geq \frac{\sigma^2 J(X)}{1 + \sigma^2 J(X)}, \tag{14}$$

where we used the fact that $h(Z) = \frac{1}{2} \ln (2\pi e\sigma^2)$. Since the left hand side of Equation (14) is scale invariant, then

$$N(aX)J(aX) = N(X)J(X) \geq \frac{\sigma^2 J(X)}{a^2 + \sigma^2 J(X)},$$

for any positive scalar a . Now, letting $a \rightarrow 0$ will yield Equation (5).

- 5- Finally, in the context of communicating over a channel, it is well-known that, under a second moment constraint, the best way to “fight” Gaussian noise is to use Gaussian inputs. This follows from the fact that Gaussian variables maximize entropy under a second moment constraint. Conversely, when using a Gaussian input, the worst noise in terms of minimizing the transmission rates is also Gaussian. This is a direct result of the EPI and is also due to the fact that Gaussian distributions have the highest entropy and therefore are the worst noise to deal with. If one were to make a similar statement where instead of the second moment, the Fisher information is constrained, *i.e.*, if the input X is subject to a Fisher information constraint: $J(X) \leq A$ for some $A > 0$, then the input minimizing the mutual information of the additive white Gaussian channel is Gaussian distributed. This is a result of the EPI in Equation (2) and the IIE in Equation (5). They both reduce in this setting to

$$\arg \min_{X:J(X) \leq A} h(X + Z) \sim \mathcal{N} \left(0; \frac{1}{A} \right).$$

Reciprocally, for a Gaussian input, what is the noise that maximizes the mutual information subject to a Fisher information constraint? this problem can be formally stated as follows: If $X \sim \mathcal{N}(0; p)$, find

$$\arg \max_{Z:J(Z) \leq A} h(X + Z).$$

An intuitive answer would be Gaussian since it has the minimum entropy for a given Fisher information. Indeed, Equation (10) provides the answer:

$$I(Y; X) \leq \frac{1}{2} \ln (1 + pJ(Z)),$$

is maximized whenever $Z \sim \mathcal{N} \left(0; \frac{1}{A} \right)$.

3. Proof of the Upper Bound

3.1. Concavity of Differential Entropy

Let U be an infinitely divisible RV with characteristic function $\phi_U(\omega)$ (the characteristic function $\phi(\omega)$ of a probability distribution function $F_U(u)$ is defined by:

$$\phi_U(\omega) = \int_{\mathbb{R}} e^{i\omega u} dF_U(u), \quad \omega \in \mathbb{R},$$

which is the Fourier transform of $F_U(u)$ at $-\omega$. For each real $t \geq 0$, denote by $F_t(\cdot)$ the unique probability distribution (Theorem 2.3.9, p. 65, [18]) with characteristic function:

$$\phi_t(\omega) = e^{t \ln \phi_U(\omega)}, \tag{15}$$

where $\ln(\cdot)$ is the principal branch of the logarithm. For the rest of this paper, we denote by U_t a RV with characteristic function $\phi_t(\omega)$ as defined in Equation (15). Note that U_0 is deterministically equal to 0 (i.e., distributed according to the Dirac delta distribution) and U_1 is distributed according to U . The family of probability distributions $\{F_t(\cdot)\}_{t \geq 0}$ forms a continuous convolution semi-group in the space of probability measures on \mathbb{R} (see Definition 2.3.8 and Theorem 2.3.9, [18]) and hence one can write:

$$U_s + U_t = U_{s+t} \quad \forall s, t \geq 0,$$

where U_s and U_t are independent.

Lemma 1. *Let U be an infinitely divisible RV and $\{U_t\}_{t \geq 0}$ an associated family of RVs distributed according to Equation (15) and independent of X . The differential entropy $h(X + U_t)$ is a concave function in $t \geq 0$.*

In the case of a Gaussian-distributed U , the family $\{U_t\}_{t \geq 0}$ has the same distribution as $\{\sqrt{t}U\}_t$, and it is already known that the entropy (and actually even the entropy power) of Y is concave in t ((Section VII, p. 51, [5]) and [19]).

Proof. We start by noting that $h(X + U_t)$ is non-decreasing in t . For $0 \leq s < t$,

$$h(X + U_t) = h(X + U_s + U_{t-s}) \geq h(X + U_s),$$

where U_t, U_s and U_{t-s} are three independent instances of RVs in the family $\{U_t\}_{t \geq 0}$. Next we show that $h(X + U_t)$ is midpoint concave: Let $U_t, U_s, U_{(t+s)/2}$ and $U_{(t-s)/2}$ be independent RVs in the family $\{U_t\}_{t \geq 0}$. For $0 \leq s < t$,

$$\begin{aligned} h(X + U_t) - h(X + U_{(t+s)/2}) &= h(X + U_{(t+s)/2} + U_{(t-s)/2}) - h(X + U_{(t+s)/2}) \\ &= I(X + U_{(t+s)/2} + U_{(t-s)/2}; U_{(t-s)/2}) \end{aligned} \tag{16}$$

$$\begin{aligned} &\leq I(X + U_s + U_{(t-s)/2}; U_{(t-s)/2}) \\ &= h(X + U_{(t+s)/2}) - h(X + U_s) \end{aligned} \tag{17}$$

where Equation (16) is the definition of the mutual information and Equation (17) is the application of the data processing inequality to the Markov chain $U_{(t-s)/2} - (X + U_s + U_{(t-s)/2}) - (X + U_{(t+s)/2} + U_{(t-s)/2})$. Therefore,

$$h(X + U_{(t+s)/2}) \geq \frac{1}{2}[h(X + U_t) + h(X + U_s)],$$

and the function is midpoint concave for $t \geq 0$. Since the function is non-decreasing, it is Lebesgue measurable and midpoint concavity guarantees its concavity. \square

An interesting implication of Lemma 1 is that $h(X + U_t)$ as a function of t is below any of its tangents. Particularly,

$$h(X + U_t) \leq h(X) + t \frac{dh(X + U_t)}{dt} \Big|_{t=0}. \tag{18}$$

3.2. Perturbations along U_t : An Identity of the de-Bruijn Type

Define the non-negative quantity

$$C_X = \sup_{x \neq 0} \frac{D(p_X(u-x) \| p_X(u))}{x^2}, \tag{19}$$

which is possibly infinite if the supremum is not finite. We first note that C_X has the following interesting properties:

- It was found by Verdu [20] to be equal to the channel capacity per unit cost of the linear average power constrained additive noise channel where the noise is independent of the input and is distributed according to X .
- Using the above interpretation, one can infer that for independent RVs X and W ,

$$C_{X+W} \leq C_X. \tag{20}$$

This inequality may be formally established using convexity of relative entropy (Theorem 2.7.2, [7]): Since $D(p\|q)$ is convex in the pair (p, q) ,

$$D(p_{X+W}(u-x)\|p_{X+W}(u)) \leq D(p_X(u-x)\|p_X(u)),$$

for independent RVs X and W .

- Using Kullback’s well-known result on the divergence (Section 2.6, [21]),

$$C_X \geq \lim_{x \rightarrow 0^+} \frac{D(p_X(u-x)\|p_X(u))}{x^2} = \frac{1}{2}J(X).$$

- Whenever the supremum is at “0”,

$$C_X = \frac{1}{2}J(X),$$

which is the case for RVs X whose density functions have heavier tails than the Gaussian (p.1020, [20]).

Next we prove the following lemma,

Lemma 2. *Let U be an infinitely divisible real RV with finite variance σ_U^2 and $\{U_t\}_{t \geq 0}$ the associated family of RVs distributed according to Equation (15). Let X be an independent real variable satisfying the following:*

- X has a positive PDF $p_X(x)$.
- The integrals $\left\{ \int_{\mathbb{R}} |\omega|^k |\phi_X(\omega)| d\omega \right\}_k$ are finite for all $k \in \mathbb{N} \setminus \{0\}$.
- $C_X = \sup_{x \neq 0} \frac{D(p_X(u-x)\|p_X(u))}{x^2}$ is finite.

For $t \geq 0$, the right derivative of $h(X + U_t)$

$$\left. \frac{d^+}{dt} h(X + U_t) \right|_{t=\tau} = \lim_{t \rightarrow 0^+} \frac{h(X + U_{\tau+t}) - h(X + U_\tau)}{t} \leq \sigma_U^2 C_X. \tag{21}$$

Before we proceed to the proof, we note that by Lemma 1, $h(X + U_\tau)$ is concave in the real-valued τ and hence it is everywhere right and left differentiable.

Proof. Using Equation (15), the characteristic function of the independent sum $X + U_{\tau+t}$ for $\tau \geq 0$ and small $t > 0$ is:

$$\begin{aligned} \phi_{X+U_{\tau+t}}(\omega) &= \phi_{X+U_\tau}(\omega)\phi_{U_t}(\omega) = \phi_{X+U_\tau}(\omega) \exp [t \ln \phi_U(\omega)] \\ &= \phi_{X+U_\tau}(\omega) + [\phi_{X+U_\tau}(\omega) \ln \phi_U(\omega)] t + o(t). \end{aligned}$$

Taking the inverse Fourier transform yields,

$$p_{X+U_{\tau+t}}(y) = p_{X+U_\tau}(y) + t \mathcal{F}^{-1} \{ \phi_{X+U_\tau}(\omega) \ln \phi_U(\omega) \} (-y) + o(t), \tag{22}$$

where \mathcal{F}^{-1} denotes the inverse distributional Fourier transform operator. Equation (22) holds true when $\mathcal{F}^{-1} \left\{ \phi_{X+U_\tau}(\omega) \ln^k \phi_U(\omega) \right\} (y)$ exists for all $k \in \mathbb{N} \setminus \{0\}$, which is the case. Indeed, using the Kolmogorov representation of the characteristic function of an infinitely divisible RV (Theorem 7.7, [22]), we know that there exists a unique distribution function $H_U(x)$ associated with U such that

$$\ln [\phi_U(\omega)] = \sigma_U^2 \int_{\mathbb{R}} \left(e^{i\omega x} - 1 - i\omega x \right) \frac{1}{x^2} dH_U(x). \tag{23}$$

Furthermore, since $|e^{i\omega x} - 1 - i\omega x| \leq \frac{\omega^2 x^2}{2}$ (p.179, [22]),

$$|\ln [\phi_U(\omega)]| \leq \sigma_U^2 \int_{\mathbb{R}} |e^{i\omega x} - 1 - i\omega x| \frac{1}{x^2} dH_U(x) \leq \frac{\sigma_U^2}{2} \omega^2,$$

which implies that

$$\begin{aligned} \int_{\mathbb{R}} \left| \phi_{X+U_\tau}(\omega) \ln^k \phi_U(\omega) \right| d\omega &\leq \int_{\mathbb{R}} |\phi_X(\omega)| |\phi_{U_\tau}(\omega)| |\ln \phi_U(\omega)|^k d\omega \\ &\leq \left(\frac{\sigma_U^2}{2} \right)^k \int_{\mathbb{R}} |\phi_X(\omega)| \omega^{2k} d\omega, \end{aligned}$$

which is finite under the conditions of the lemma and hence $\mathcal{F}^{-1} \left\{ \phi_{X+U_\tau}(\omega) \ln^k \phi_U(\omega) \right\} (y)$ exists. Using the definition of the derivative, Equation (22) implies that:

$$\left. \frac{d p_{X+U_t}(y)}{dt} \right|_{t=\tau} = \mathcal{F}^{-1} \left\{ \phi_{X+U_\tau}(\omega) \ln \phi_U(\omega) \right\} (-y).$$

Using the Mean Value Theorem: for some $0 < h(t) < t$,

$$\begin{aligned} \frac{h(X + U_{\tau+t}) - h(X + U_\tau)}{t} &= - \int_{\mathbb{R}} \frac{p_{X+U_{\tau+t}}(y) \ln p_{X+U_{\tau+t}}(y) - p_{X+U_\tau}(y) \ln p_{X+U_\tau}(y)}{t} dy \\ &= - \int_{\mathbb{R}} \left(1 + \ln p_{X+U_{\tau+h(t)}}(y) \right) \left. \frac{d p_{X+U_t}(y)}{dt} \right|_{\tau+h(t)} dy \\ &= - \int_{\mathbb{R}} \left(1 + \ln p_{X+U_{\tau+h(t)}}(y) \right) \mathcal{F}^{-1} \left\{ \phi_{X+U_{\tau+h(t)}}(\omega) \ln \phi_U(\omega) \right\} (-y) dy \\ &= - \int_{\mathbb{R}} \ln p_{X+U_{\tau+h(t)}}(y) \mathcal{F}^{-1} \left\{ \phi_{X+U_{\tau+h(t)}}(\omega) \ln \phi_U(\omega) \right\} (-y) dy, \tag{24} \end{aligned}$$

where Equation (24) is justified by the fact that $\phi_X(0) = \phi_U(0) = 1$. We proceed next to evaluate the inverse Fourier transform in the integrand of Equation (24):

$$\begin{aligned} \mathcal{F}^{-1} \left\{ \phi_{X+U_{\tau+h(t)}}(\omega) \ln \phi_U(\omega) \right\} (-y) &= \frac{1}{2\pi} \int_{\mathbb{R}} \phi_{X+U_{\tau+h(t)}}(\omega) \ln \phi_U(\omega) e^{-i\omega y} d\omega \\ &= \frac{\sigma_U^2}{2\pi} \int_{\mathbb{R}} \phi_{X+U_{\tau+h(t)}}(\omega) \int_{\mathbb{R}} \left(e^{i\omega x} - 1 - i\omega x \right) \frac{1}{x^2} dH_U(x) e^{-i\omega y} d\omega \tag{25} \end{aligned}$$

$$\begin{aligned} &= \frac{\sigma_U^2}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} \phi_{X+U_{\tau+h(t)}}(\omega) \left(e^{i\omega x} - 1 - i\omega x \right) e^{-i\omega y} d\omega \frac{1}{x^2} dH_U(x) \tag{26} \\ &= \sigma_U^2 \int_{\mathbb{R}} \left(p_{X+U_{\tau+h(t)}}(y-x) - p_{X+U_{\tau+h(t)}}(y) - x p'_{X+U_{\tau+h(t)}}(y) \right) \frac{1}{x^2} dH_U(x), \end{aligned}$$

where the last equation is due to standard properties of the Fourier transform and Equation (25) is due to Equation (23). The interchange in Equation (26) is justified by Fubini. In fact $|e^{i\omega x} - 1 - i\omega x| \leq \frac{\omega^2 x^2}{2}$ and

$$\begin{aligned} & \int_{\mathbb{R}} \int_{\mathbb{R}} \left| \phi_{X+U_{\tau+h(t)}}(\omega) \left(e^{i\omega x} - 1 - i\omega x \right) e^{-i\omega y} \frac{1}{x^2} \right| d\omega dH_U(x) \\ & \leq \int_{\mathbb{R}} \int_{\mathbb{R}} |\phi_X(\omega)| \frac{\omega^2}{2} d\omega dH_U(x) \leq \int_{\mathbb{R}} |\phi_X(\omega)| \frac{\omega^2}{2} d\omega, \end{aligned}$$

which is finite by assumption. Back to Equation (24),

$$\begin{aligned} & \frac{h(X + U_{\tau+t}) - h(X + U_{\tau})}{t} \\ & = - \int_{\mathbb{R}} \ln p_{X+U_{\tau+h(t)}}(y) \mathcal{F}^{-1} \left\{ \phi_{X+U_{\tau+h(t)}}(\omega) \ln \phi_U(\omega) \right\} (-y) dy \\ & = -\sigma_U^2 \int_{\mathbb{R}} \int_{\mathbb{R}} \ln p_{X+U_{\tau+h(t)}}(y) \left(p_{X+U_{\tau+h(t)}}(y-x) - p_{X+U_{\tau+h(t)}}(y) - xp'_{X+U_{\tau+h(t)}}(y) \right) \frac{1}{x^2} dH_U(x) dy \\ & = -\sigma_U^2 \int_{\mathbb{R}} \int_{\mathbb{R}} \ln p_{X+U_{\tau+h(t)}}(y) \left(p_{X+U_{\tau+h(t)}}(y-x) - p_{X+U_{\tau+h(t)}}(y) - xp'_{X+U_{\tau+h(t)}}(y) \right) dy \frac{1}{x^2} dH_U(x), \end{aligned} \tag{27}$$

where the interchange in the order of integration in Equation (27) will be validated next by Fubini. Considering the inner integral,

$$\begin{aligned} & - \int_{\mathbb{R}} \ln p_{X+U_{\tau+h(t)}}(y) \left(p_{X+U_{\tau+h(t)}}(y-x) - p_{X+U_{\tau+h(t)}}(y) - xp'_{X+U_{\tau+h(t)}}(y) \right) dy \\ & = D(p_{X+U_{\tau+h(t)}}(y-x) \| p_{X+U_{\tau+h(t)}}(y)) + x \int_{\mathbb{R}} p'_{X+U_{\tau+h(t)}}(y) dy \\ & = D(p_{X+U_{\tau+h(t)}}(u-x) \| p_{X+U_{\tau+h(t)}}(u)). \end{aligned}$$

Finally, Equation (27) gives

$$\frac{h(X + U_{\tau+t}) - h(X + U_{\tau})}{t} = \sigma_U^2 \int_{\mathbb{R}} D(p_{X+U_{\tau+h(t)}}(u-x) \| p_{X+U_{\tau+h(t)}}(u)) \frac{1}{x^2} dH_U(x) \tag{28}$$

$$\leq \sigma_U^2 C_{X+U_{\tau+h(t)}} \tag{29}$$

$$\leq \sigma_U^2 C_X, \tag{30}$$

which is finite. Equation (29) is due to the definition Equation (19) and Equation (30) is due to Equation (20). The finiteness of the end result justifies the interchange of the order of integration in Equation (27) by Fubini. □

We point out that the result of this lemma is sufficient for the purpose of the main result of this paper. Nevertheless, it is worth noting that Equation (21) could be strengthened to:

$$\frac{d}{dt} h(X + U_t) \Big|_{t=\tau} = \sigma_U^2 \int_{\mathbb{R}} D(p_{X+U_{\tau}}(u-x) \| p_{X+U_{\tau}}(u)) \frac{1}{x^2} dH_U(x). \tag{31}$$

In fact, the set of points where the left and right derivative of the concave function $h(X + U_t)$ differ is of zero measure. One can therefore state that the derivative exists almost-everywhere and it is upperbounded almost-everywhere by Equation (30). Furthermore, considering Equation (28), one can see that taking the limit as $t \rightarrow 0^+$ will yield

$$\frac{d^+}{dt} h(X + U_t) \Big|_{t=\tau} = \sigma_U^2 \int_{\mathbb{R}} \lim_{t \rightarrow 0} D(p_{X+U_{\tau+h(t)}}(u-x) \| p_{X+U_{\tau+h(t)}}(u)) \frac{1}{x^2} dH_U(x), \tag{32}$$

by the Monotone Convergence Theorem. The continuity of relative entropy may be established using techniques similar to those in [23] when appropriate conditions on p_X hold.

Finally, When U is purely Gaussian, $U_t \sim \sqrt{t}U$, $H_U(x)$ is the unit step function and Equation (31) boils down to de Bruijn’s identity for Gaussian perturbations (Equation (4)).

3.3. Proof of Theorem 1

Proof. We assume without loss of generality that X and Z have a zero mean. Otherwise, define $Y' = Y - (\mu_X + \mu_1 + \mu_2)$ and $X' = X - \mu_X$ for which $h(Y') = h(Y)$, $h(X') = h(X)$ and $J(X') = J(X)$ since differential entropy and Fisher information are translation invariant. We divide the proof into two steps and we start by proving the theorem when Z is purely Gaussian.

Z is purely Gaussian:

We decompose Z as follows: Let $n \in \mathbb{N}^* \setminus \{1\}$ and $\epsilon = \frac{1}{n}$, then

$$Z = \sqrt{\epsilon} \sum_{i=1}^n Z^i,$$

where the $\{Z_i\}_{0 \leq i \leq n}$ are IID with the same law as Z . We write Equation (7) in an incremental formulation as follows:

$$\begin{aligned} Y^0 &= X \\ Y^l &= Y^{l-1} + \sqrt{\epsilon} Z^l = X + \sqrt{\epsilon} \sum_{i=1}^l Z^i \end{aligned}$$

for $l \in \{1, \dots, n\}$. Note that Y^n has the same statistics as Y . Using the de Bruijn's identity (Equation (4)), we write

$$h(Y^l) = h(Y^{l-1}) + \frac{\epsilon \sigma_1^2}{2} J(Y^{l-1}) + o(\epsilon),$$

and

$$h(Y^l) \leq h(Y^{l-1}) + \frac{\epsilon \sigma_1^2}{2} J(Y^{l-1}), \quad l \in \{1, \dots, n\},$$

by virtue of the fact that $h(Y^l)$ is concave in $\epsilon \geq 0$ (Section VII, p. 51, [5]). Using the FII Equation (3) on Y^l we obtain

$$\begin{aligned} \frac{1}{J(Y^l)} &\geq \frac{1}{J(Y^{l-1})} + \frac{1}{J(\sqrt{\epsilon} Z^l)} = \frac{1}{J(Y^{l-1})} + \frac{\epsilon}{J(Z^l)} \\ &\geq \frac{1}{J(X)} + l \frac{\epsilon}{J(Z^1)} \quad l \in \{1, \dots, n\}. \end{aligned} \tag{33}$$

Examining now $h(Y) = h(Y^n)$,

$$\begin{aligned} h(Y^n) &\leq h(Y^{n-1}) + \frac{\epsilon \sigma_1^2}{2} J(Y^{n-1}) \\ &\leq h(X) + \frac{\epsilon \sigma_1^2}{2} \sum_{l=0}^{n-1} J(Y^l) \\ &\leq h(X) + \frac{\epsilon \sigma_1^2}{2} \left[J(X) + \sum_{l=1}^{n-1} \frac{J(X)J(Z^1)}{J(Z^1) + l\epsilon J(X)} \right] \end{aligned} \tag{34}$$

$$\begin{aligned} &\leq h(X) + \frac{\epsilon \sigma_1^2}{2} J(X) \left[1 + \int_0^{n-1} \frac{J(Z^1)}{J(Z^1) + u\epsilon J(X)} du \right] \\ &= h(X) + \frac{\epsilon \sigma_1^2}{2} J(X) + \frac{\sigma_1^2}{2} J(Z^1) \ln \left[1 + (n-1)\epsilon \frac{J(X)}{J(Z^1)} \right] \\ &= h(X) + \frac{\epsilon \sigma_1^2}{2} J(X) + \frac{1}{2} \ln \left[1 + (1-\epsilon) \sigma_1^2 J(X) \right], \end{aligned} \tag{35}$$

where, in order to write the last equality, we used the fact that $J(Z^1) = \frac{1}{\sigma_1^2}$ since $Z_1 \sim \mathcal{N}(0, \sigma_1^2)$. Equation (34) is due to the bounds in Equation (33) and Equation (35) is justified since the function $\frac{J(Z^1)}{J(Z^1) + u\epsilon J(X)}$ is decreasing in u . Since the upper bound is true for any small-enough ϵ , necessarily

$$h(Y) \leq h(X) + \frac{1}{2} \ln \left(1 + \sigma_1^2 J(X) \right),$$

which completes the proof of the second part of the theorem. When X and Z are both Gaussian, evaluating the quantities shows that equality holds. In addition, equality only holds whenever the FII is satisfied with equality that is whenever X and Z are Gaussian.

Z is a sum of a Gaussian variable with a non-Gaussian infinitely divisible one:

Let $\{U_t\}_{t \geq 0}$ be a family of RVs associated with the non-Gaussian infinitely divisible RV Z_2 and distributed according to Equation (15). Using concavity Equation (18) in t ,

$$\begin{aligned} h(Y) &= h(X + Z_1 + Z_2) = h(X + Z_1 + U_1) \\ &\leq h(X + Z_1) + \left. \frac{dh(X + Z_1 + U_t)}{dt} \right|_{t=0^+} \\ &\leq h(X + Z_1) + \sigma_2^2 C_{X+Z_1} \end{aligned} \tag{36}$$

$$\leq h(X) + \frac{1}{2} \ln \left(1 + \sigma_1^2 J(X) \right) + \sigma_2^2 C_{X+Z_1} \tag{37}$$

$$\leq h(X) + \frac{1}{2} \ln \left(1 + \sigma_1^2 J(X) \right) + \sigma_2^2 \min \{C_X; C_Z\} \tag{38}$$

$$= h(X) + \frac{1}{2} \ln \left(1 + \sigma_1^2 J(X) \right) + \sigma_2^2 \min \left\{ \sup_x \frac{D(p_X(u-x) \| p_X(u))}{x^2}, \frac{1}{2\sigma_1^2} \right\}.$$

Equation (36) is an application of Lemma 2 since $(X + Z_1)$ has a positive PDF and satisfies all the required technical conditions. The upperbound proven in the previous paragraph gives Equation (37) and (38) is due to Equation (20). □

On a final note, using Lemmas 1 and 2 one could have applied the Plünnecke-Ruzsa inequality (Theorem 3.11, [10]) which yields

$$h(Y) \leq h(X) + \frac{\sigma_1^2}{2} J(X) + \sigma_2^2 C_{X+Z_1},$$

which is looser than Equation (37).

4. Extension

The bound in Equation (10) may be extended to the n -dimensional vector Gaussian case, $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$, where \mathbf{Z} is an n -dimensional Gaussian vector. In this case, if $\mathbf{J}(\mathbf{X})$ denotes the Fisher information matrix, the Fisher information and the entropy power are defined as

$$\mathbf{J}(\mathbf{X}) = \text{Tr}(\mathbf{J}(\mathbf{X})) \quad N(\mathbf{X}) = \frac{1}{2\pi e} e^{\frac{2}{n} h(\mathbf{X})}.$$

- When \mathbf{Z} has n IID Gaussian components –i.e., with covariance matrix $\Lambda_Z = \sigma^2 \mathbf{I}$, following similar steps lead to:

$$N(\mathbf{X} + \mathbf{Z}) \leq N(\mathbf{X}) + N(\mathbf{Z}) \frac{N(\mathbf{X})\mathbf{J}(\mathbf{X})}{n}, \tag{39}$$

with equality if and only if \mathbf{X} is Gaussian with IID components.

- In general, for any positive-definite matrix Λ_Z with a singular value decomposition UDU^T , if we denote by $B = UD^{-\frac{1}{2}}U^T$ then

$$BY = B(\mathbf{X} + \mathbf{Z}) = B\mathbf{X} + B\mathbf{Z} = B\mathbf{X} + \mathbf{Z}'$$

where \mathbf{Z}' is Gaussian distributed with an identity covariance matrix. Equation (39) gives

$$\begin{aligned} N(B(\mathbf{X} + \mathbf{Z})) &\leq N(B\mathbf{X}) + N(\mathbf{Z}') \frac{N(B\mathbf{X})J(B\mathbf{X})}{n} \\ \iff N(\mathbf{X} + \mathbf{Z}) &\leq N(\mathbf{X}) + \frac{N(\mathbf{X})J(\mathbf{X})}{n} \\ \iff N(\mathbf{X} + \mathbf{Z}) &\leq N(\mathbf{X}) + \frac{N(\mathbf{X})\text{Tr}(J(\mathbf{X})\Lambda_Z)}{n}, \end{aligned}$$

where we used $N(B\mathbf{X}) = |\det(B)|^{\frac{2}{n}} N(\mathbf{X})$.

5. Conclusions

We have derived a novel tight upper bound on the entropy of the sum of two independent random variables where one of them is infinitely divisible with a Gaussian component. The bound is shown to be tighter than previously known ones and holds for variables with possibly infinite second moment. With the isoperimetric inequality in mind, the “symmetry” that this bound provides with the known lower bounds is remarkable and hints to possible generalizations to scenarios where no Gaussian component is present.

Acknowledgments: The authors would like to thank the Assistant Editor for his patience and the anonymous reviewers for their helpful comments. This work was supported by AUB’s University Research Board and the Lebanese National Council for Scientific Research (CNRS-L).

Author Contributions: Both authors contributed equally on all stages of this work: formulated and solved the problem and prepared the manuscript. Both authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shannon, C.E. A mathematical theory of communication, part I. *Bell Syst. Tech. J.* **1948**, *27*, 379–423.
2. Bobkov, S.G.; Chistyakov, G.P. Entropy power inequality for the Renyi entropy. *IEEE Trans. Inf. Theory* **2015**, *61*, 708–714.
3. Stam, A.J. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Inf. Control* **1959**, *2*, 101–112.
4. Blachman, N.M. The convolution inequality for entropy powers. *IEEE Trans. Inf. Theory* **1965**, *11*, 267–271.
5. Rioul, O. Information theoretic proofs of entropy power inequality. *IEEE Trans. Inf. Theory* **2011**, *57*, 33–55.
6. Dembo, A.; Cover, T.M.; Thomas, J.A. Information Theoretic Inequalities. *IEEE Trans. Inf. Theory* **1991**, *37*, 1501–1518.
7. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: New York, NY, USA, 2006.
8. Ruzsa, I.Z. Sumsets and entropy. *Random Struct. Algorithms* **2009**, *34*, 1–10.
9. Tao, T. Sumset and inverse sumset theory for Shannon entropy. *Comb. Probab. Comput.* **2010**, *19*, 603–639.
10. Kontoyiannis, I.; Madiman, M. Sumset and inverse sumset inequalities for differential entropy and mutual information. *IEEE Trans. Inf. Theory* **2014**, *60*, 4503–4514.
11. Madiman, M. On the entropy of sums. In Proceedings of the 2008 IEEE Information Theory Workshop, Oporto, Portugal, 5–9 May 2008.
12. Cover, T.M.; Zhang, Z. On the maximum entropy of the sum of two dependent random variables. *IEEE Trans. Inf. Theory* **1994**, *40*, 1244–1246.
13. Ordentlich, E. Maximizing the entropy of a sum of independent bounded random variables. *IEEE Trans. Inf. Theory* **2006**, *52*, 2176–2181.

14. Bobkov, S.; Madiman, M. On the Problem of Reversibility of the Entropy Power Inequality. In *Limit Theorems in Probability, Statistics and Number Theory*; Springer-Verlag: Berlin/Heidelberg, Germany, 2013; pp. 61–74.
15. Miclo, L. Notes on the speed of entropic convergence in the central limit theorem. *Progr. Probab.* **2003**, *56*, 129–156.
16. Luisier, F.; Blu, T.; Unser, M. Image denoising in mixed Poisson-Gaussian noise. *IEEE Trans. Image Process.* **2011**, *20*, 696–708.
17. Fahn, J.; Abou-Faycal, I. Using Hermite bases in studying capacity-achieving distributions over AWGN channels. *IEEE Trans. Inf. Theory* **2012**, *58*, 5302–5322.
18. Heyer, H. *Structural Aspects in the Theory of Probability: A Primer in Probabilities on Algebraic-Topological Structures*; World Scientific: Singapore, Singapore, 2004; Volume 7.
19. Costa, M.H.M. A new entropy power inequality. *IEEE Trans. Inf. Theory* **1985**, *31*, 751–760.
20. Verdú, S. On channel capacity per unit cost. *IEEE Trans. Inf. Theory* **1990**, *36*, 1019–1030.
21. Kullback, S. *Information Theory and Statistics*; Dover Publications: Mineola, NY, USA, 1968.
22. Steutel, F.W.; Harn, K.V. *Infinite Divisibility of Probability Distributions on the Real Line*; Marcel Dekker Inc.: New York, NY, USA, 2006.
23. Fahn, J.; Abou-Faycal, I. On the finiteness of the capacity of continuous channels. *IEEE Trans. Commun.* **2015**, doi:10.1109/TCOMM.2015.2503403.



© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).