Sumset and Inverse Sumset Inequalities for Differential Entropy and Mutual Information

Ioannis Kontoyiannis, *Fellow, IEEE* [*†] Mokshay Madiman, *Member, IEEE* [‡§]

June 3, 2012

Abstract

The *sumset* and *inverse sumset* theories of Freiman, Plünnecke and Ruzsa, give bounds connecting the cardinality of the sumset $A+B=\{a+b\,;\,a\in A,\,b\in B\}$ of two discrete sets A,B, to the cardinalities (or the finer structure) of the original sets A,B. For example, the sum-difference bound of Ruzsa states that, $|A+B|\,|A|\,|B|\leq|A-B|^3$, where the difference set $A-B=\{a-b\,;\,a\in A,b\in B\}$. Interpreting the differential entropy $h(X)$ of a continuous random variable X as (the logarithm of) the size of the effective support of X, the main contribution of this paper is a series of natural information-theoretic analogs for these results. For example, the Ruzsa sum-difference bound becomes the new inequality, $h(X+Y)+h(X)+h(Y)\leq 3h(X-Y)$, for any pair of independent continuous random variables X and Y. Our results include differential-entropy versions of Ruzsa's triangle inequality, the Plünnecke-Ruzsa inequality, and the Balog-Szemerédi-Gowers lemma. Also we give a differential entropy version of the Freiman-Green-Ruzsa inverse-sumset theorem, which can be seen as a quantitative converse to the entropy power inequality. Versions of most of these results for the discrete entropy $H(X)$ were recently proved by Tao, relying heavily on a strong, functional form of the submodularity property of $H(X)$. Since differential entropy is *not* functionally submodular, in the continuous case many of the corresponding discrete proofs fail, in many cases requiring substantially new proof strategies. We find that the basic property that naturally replaces the discrete functional submodularity, is the data processing property of mutual information.

Keywords — Shannon entropy, differential entropy, sumset bounds, inequalities, submodularity, data processing, mutual information

[0]Preliminary versions of parts of this work were presented at the 2008 IEEE Information Theory Workshop [9] and at the 2010 IEEE International Symposium on Information Theory [10].

[*]Department of Informatics, Athens University of Economics and Business, Patission 76, Athens 10434, Greece. Email: yiannis@aueb.gr.

[†]I.K. was supported, in part, by a Marie Curie International Outgoing Fellowship, PIOF-GA-2009-235837.

[‡]Department of Statistics, Yale University, 24 Hillhouse Avenue, New Haven, CT 06511, USA. Email: mokshay.madiman@yale.edu.

[§]M.M. was supported by the NSF CAREER grant DMS-1056996 and by NSF grant CCF-1065494.

1 Introduction

1.1 Motivation

Roughly speaking, the field of *additive combinatorics* provides tools that allow us to count the number of occurrences of particular additive structures in specific subsets of a discrete group; see [17] for a broad introduction. The prototypical example is the study of the existence of arithmetic progressions within specific sets of integers – as opposed to the multiplicative structure that underlies prime factorization and much of classical combinatorics and number theory. There have been several major developments and a lot of high-profile mathematical activity in connection with additive combinatorics in recent years, perhaps the most famous example being the celebrated Green-Tao theorem on the existence of arbitrarily long arithmetic progressions within the set of prime numbers.

An important collection of tools in additive combinatorics is a variety of *sumset inequalities*, the so-called Plünnecke-Ruzsa *sumset theory*; see [17] for details. The *sumset* $A+B$ of two discrete sets A and B is defined as, $A+B = \{a+b : a \in A, b \in B\}$, and a *sumset inequality* is an inequality connecting the cardinality $|A+B|$ of $A+B$ with the cardinalities $|A|, |B|$ of A and B, respectively. For example, there are the obvious bounds,

$$\max\{|A|, |B|\} \leq |A+B| \leq |A|\,|B|, \tag{1}$$

as well as much more subtle results, like the Ruzsa triangle inequality [13],

$$|A-C| \leq \frac{|A-B|\,|B-C|}{|B|}, \tag{2}$$

or the sum-difference bound [13],

$$|A+B| \leq \frac{|A-B|^3}{|A|\,|B|}, \tag{3}$$

all of which hold for arbitrary subsets A, B, C of the integers or any other discrete abelian group, and where the *difference set* $A-B$ is defined as, $A-B = \{a-b : a \in A, b \in B\}$.

In the converse direction, the Freiman-Ruzsa *inverse sumset theory* provides information about sets A for which $|A+A|$ is close to being as small as possible; see Section 4 for a brief discussion or the text [17] for details.

In this context, recall that Shannon's asymptotic equipartition property (AEP) [3] says that the entropy $H(X)$ of a discrete random variable X can be thought of as the logarithm of the *effective cardinality* of the alphabet of X. This suggests a correspondence between bounds for the cardinalities of sumsets, e.g., $|A+B|$, and corresponding bounds for the entropy of sums of independent discrete random variables, e.g., $H(X+Y)$. First identified by Ruzsa [14], this connection has also been explored in the last few years in different directions by, among others, Tao and Vu [18], Lapidoth and Pete [8], Madiman and Kontoyiannis [10], and Madiman, Marcus and Tetali [11]; additional pointers to the relevant literature are given below.

This connection was developed most extensively by Tao in [16]. The main idea is to replace sets by (independent, discrete) random variables, and then replace the log-cardinality, $\log|A|$, of each set A by the (discrete, Shannon) entropy of the corresponding random variable (where log denotes the natural logarithm $\log_e$). Thus, for independent discrete random variables X, Y, Z, the simple bounds (1) become,

$$\max\{H(X), H(Y)\} \leq H(X+Y) \leq H(X) + H(Y),$$

which is a trivial exercise in manipulating entropy [3]. On the other hand, again for independent discrete random variables, Ruzsa's influential bounds (2) and (3) become, respectively,

$$\begin{aligned} H(X-Z)+H(Y) &\leq H(X-Y)+H(Y-Z) \\ \text{and } H(X+Y)+H(X)+H(Y) &\leq 3H(X-Y), \end{aligned}$$

which are nontrivial facts proved in [16].

Our main motivation is to examine the extent to which this analogy can be carried further: According to the AEP [3], the *differential entropy* $h(X)$ of a *continuous* random variable X can be thought of as the logarithm of the "size of the effective support" of X. In this work we state and prove natural "differential entropy analogs" of various sumset and inverse-sumset bounds, many of which were proved for the discrete Shannon entropy in the recent work of Tao [16] and in earlier papers by Kaimonovich and Vershik [7], Tao and Vu [18], Madiman [9], Ruzsa [14], and Madiman, Marcus and Tetali [11].

Of particular interest in motivating these results is the fact that the main technical ingredient in the proofs of many of the corresponding discrete bounds was a strong, functional form of the *submodularity* property of the discrete Shannon entropy; see Section 2 for details. The fact that differential entropy is *not* functionally submodular was the source of the main difficulty as well as the main interest for the present development.

1.2 Outline of main results

In Section 2, after briefly reviewing some necessary background and basic definitions, we discuss the functional submodularity of the discrete entropy, and explain how it fails for differential entropy.

Section 3 contains most of our main results, namely, a series of natural differential entropy analogs of the sumset bounds in [16] and in the earlier papers mentioned above. In Theorem 3.1 we prove the following version of the *Ruzsa triangle inequality*: If X, Y, Z are independent, then:

$$h(X-Z) \leq h(X-Y)+h(Y-Z)-h(Y).$$

In Theorem 3.5, we prove the *doubling-difference inequality*: If X_1, X_2 are independent and identically distributed (i.i.d.), then:

$$\frac{1}{2} \leq \frac{h(X_1+X_2)-h(X_1)}{h(X_1-X_2)-h(X_1)} \leq 2.$$

More generally, when X_1, X_2 are independent but not identically distributed, the following *sum-difference inequality* holds, given in Theorem 3.7: If X_1, X_2 are independent, then:

$$h(X_1+X_2) \leq 3h(X_1-X_2)-h(X_1)-h(X_2).$$

A version of the *Plünnecke-Ruzsa inequality* for differential entropy is given in Theorem 3.11: If $X, Y_1, Y_2, \ldots, Y_n$ are independent and there are constants $K_1, K_2, \ldots, K_n$ such that,

$$h(X+Y_i) \leq h(X)+\log K_i, \quad \text{for each } i,$$

then

$$h(X+Y_1+Y_2+\cdots+Y_n) \leq h(X)+\log K_1 K_2 \cdots K_n.$$

An application of this result gives the *iterated sum bound* of Theorem 3.12.

Next we prove a *Balog-Szemerédi-Gowers lemma* for differential entropy. It says that, if X, Y are weakly dependent and $X+Y$ has small entropy, then there exist conditionally independent versions of X, Y that have almost the same entropy, and whose *independent* sum still has small entropy. Specifically, in Theorem 3.14 we prove the following: Suppose X, Y satisfy $I(X;Y) \leq \log K$, for some $K \geq 1$, and suppose also that,

$$h(X+Y) \leq \frac{1}{2}h(X) + \frac{1}{2}h(Y) + \log K.$$

Let X_1, X_2 be conditionally independent versions of X given Y, and let Y' be a conditionally independent version of Y, given X_2 and Y. Then:

$$\begin{aligned} h(X_2|X_1,Y) &\geq h(X) - \log K \\ h(Y'|X_1,Y) &\geq h(Y) - \log K \\ h(X_2+Y'|X_1,Y) &\leq \frac{1}{2}h(X) + \frac{1}{2}h(Y) + 7\log K. \end{aligned}$$

The main interest in the proofs of the above results is that, in most cases, the corresponding discrete-entropy proofs do not generalize in a straightforward manner. There, the main technical tool used is a strong, functional submodularity property of $H(X)$, which does *not* hold for differential entropy. Moreover, for several results it is the overall proof structure that does not carry over to the continuous case; not only the method, but some of the important intermediate steps fail to hold for differential entropy, requiring substantially new proof strategies.

The main technical ingredient in our proofs is the *data processing* property of mutual information. Indeed, most of the bounds in Section 3 can be equivalently stated in terms of mutual information instead of differential entropy. And since data processing is universal in that it holds regardless of the space in which the relevant random variables take values, these proofs offer alternative derivations for the discrete counterparts of the results. The earlier discrete versions are discussed in Section 3.4, where we also describe the entropy version of the *Ruzsa covering lemma* and the fact that its obvious generalization fails in continuous case.

In Section 4 we give a version of the Freiman-Green-Ruzsa inverse-sumset theorem for differential entropy. Roughly speaking, Tao in [16] proves that, if the entropy $H(X+X')$ of the sum of two i.i.d. copies of a discrete random variable X is close to $H(X)$, then X is approximately uniformly distributed on a generalized arithmetic progression.

In the continuous case, the entropy power inequality [3] says that, if X, X' are i.i.d., then,

$$h(X+X') - h(X) \geq \frac{1}{2}\log 2,$$

with equality if and only if X is Gaussian. In Theorem 4.1 we state and prove a quantitative converse to this statement: Under certain regularity conditions on the density of X, we show that if $h(X+X')$ is not much larger than $h(X)$, then X will necessarily be approximately Gaussian, in that the relative entropy between its density and that of a Gaussian with the same mean and variance will be correspondingly small.

Finally we note that, in view of the fact that additive noise is one of the most common modeling assumptions in Shannon theory, it is natural to expect that, likely, some of the bounds developed here may have applications in core information-theoretic problems. Preliminary connections in this direction can be found in the recent work of Cohen and Zamir [2], Etkin and Ordentlich [4], and Wu, Shamai and Verdú [19].

2 Elementary Bounds and Preliminaries

The entropy of a discrete random variable X with probability mass function P on the alphabet A is $H(X) = E[-\log P(X)] = \sum_{x\in A} P(x)\log(1/P(x))$. Throughout the paper, log denotes the natural logarithm $\log_e$, and the support (or alphabet) of any discrete random variable X is assumed to be a (finite or countably infinite) subset of the real line or of an arbitrary discrete abelian group. Perhaps the simplest bound on the entropy $H(X+Y)$ of the sum of two independent random variables X, Y is,

$$H(X+Y) \geq \max\{H(X), H(Y)\},$$

which easily follows from elementary properties [3],

$$\begin{aligned} H(X)+H(Y) = H(X,Y) = H(Y,X+Y) &= H(X+Y)+H(Y|X+Y) \\ &\leq H(X+Y)+H(Y), \end{aligned} \tag{4}$$

and similarly with the roles of X and Y interchanged. The first and third equalities follow from the chain rule and independence, the second equality follows from the "data processing" property that $H(F(Z)) = H(Z)$ if F is a one-to-one function, and the inequality follows from the fact that conditioning reduces entropy.

A similar argument using the nonnegativity of conditional entropy [3],

$$H(X)+H(Y) = H(Y,X+Y) = H(X+Y)+H(Y|X+Y) \geq H(X+Y),$$

gives the upper bound,

$$H(X+Y) \leq H(X)+H(Y). \tag{5}$$

The starting point of our development is the recent work of Tao [16], where a series of sumset bounds are established for $H(X)$, beginning with the elementary inequalities (4) and (5). The arguments in [16] are largely based on the following important observation [16][11]:

Lemma 2.1 (Functional submodularity of Shannon entropy) *If $X_0 = F(X_1) = G(X_2)$ and $X_{12} = R(X_1, X_2)$, then:*

$$H(X_{12}) + H(X_0) \leq H(X_1) + H(X_2).$$

Proof. By data processing for mutual information and entropy, $H(X_1)+H(X_2)-H(X_{12}) \geq H(X_1)+H(X_2)-H(X_1,X_2) = I(X_1;X_2) \geq I(X_0;X_0) = H(X_0)$. □

One of our main goals in this work is to examine the extent to which the bounds in [16] and in earlier work extend to the continuous case. The differential entropy of a continuous random variable X with density f on $\mathbb{R}$ is $h(X) = E[-\log f(X)] = \int_{-\infty}^{\infty} f(x)\log(1/f(x))\,dx$. The differential entropy of any finite-dimensional, continuous random vector $\mathrm{X} = (X_1, X_2, \ldots, X_n)$ is defined analogously, in terms of the joint density of the X_i. In order to avoid uninteresting technicalities, we assume throughout that the differential entropies in the statements of all our results exist and are finite.

The first important difference between $H(X)$ and $h(X)$ is that the differential entropy of a one-to-one function of X is typically different from that of X itself, even for linear functions [3]:

For any continuous random vector X and any nonsingular matrix T, $h(T\mathrm{X}) = h(\mathrm{X}) + \log|\det(T)|$, which is different from $h(X)$ unless T has determinant equal to ± 1.

The upper bound in (5) also fails in general for independent continuous X, Y: Take, e.g., X, Y to be independent Gaussians, one with variance $\sigma^2 > 2\pi e$ and the other with variance $1/\sigma^2$. And the functional submodularity Lemma 2.1 similarly fails for differential entropy. For example, taking $X_1 = X_2$ an arbitrary continuous random variable with finite entropy, $F(x) = G(x) = x$ and $R(x, x') = ax$ for some $a > 1$, the obvious differential-entropy analog of Lemma 2.1 yields $\log a \leq 0$.

On the other hand, the simple lower bound in (5) does generalize,

$$h(X+Y) \geq \max\{h(X), h(Y)\}, \tag{6}$$

and is equivalent to the data processing inequality,

$$\min\{I(X+Y;X),\ I(X+Y;Y)\} \geq 0,$$

since,

$$0 \leq I(X+Y;X) = h(X+Y) - h(X+Y|X) = h(X+Y) - h(Y|X) = h(X+Y) - h(Y),$$

and similarly for $h(X)$ in place of $h(Y)$; here we use the fact that differential entropy is translation-invariant.

In the rest of the paper, all standard properties of $h(X)$ and $H(X)$ will be used without explicit reference; they can all be found, e.g., in [3]. Since it will play a particularly central role in what follows, we recall that the mutual information $I(\mathrm{X};\mathrm{Y})$ between two arbitrary continuous random vectors X, Y can be defined as,

$$I(\mathrm{X};\mathrm{Y}) = h(\mathrm{X}) - h(\mathrm{X}|\mathrm{Y}) = h(\mathrm{Y}) - h(\mathrm{Y}|\mathrm{X}) = h(\mathrm{X}) + h(\mathrm{Y}) - h(\mathrm{X},\mathrm{Y}),$$

and the *data processing inequality* states that, whenever X and Z are conditionally independent given Y, we have,

$$I(\mathrm{X};\mathrm{Y}) \geq I(\mathrm{X},\mathrm{Z}).$$

The development in Section 3 will be largely based on the idea that the use of functional submodularity can be avoided by reducing the inequalities of interest to data-processing inequalities for appropriately defined mutual informations. This reduction is sometimes straightforward, but sometimes far from obvious.

3 Sumset Bounds for Differential Entropy

Throughout the rest of the paper, unless explicitly stated otherwise, all random variables are assumed to be real-valued and continuous (i.e., with distributions absolutely continuous with respect to Lebesgue measure, or in other words, having a probability density function), and the differential entropy of any random variable or random vector appearing in the statement of any of our results is assumed to exist and be finite.

3.1 Ruzsa distance and the doubling and difference constants

In analogy with the corresponding definition for discrete random variables [16], we define the *Ruzsa distance* between any two continuous random variables X and Y as,

$$\mathrm{dist}_R(X,Y) = h(X'-Y') - \frac{1}{2}h(X') - \frac{1}{2}h(Y'),$$

where $X' \sim X$ and $Y' \sim Y$ are independent. It is obvious that dist_R is symmetric, and it is nonnegative because of the lower bound in (6). Our first result states that it also satisfies the triangle inequality:

Theorem 3.1 (Ruzsa triangle inequality for differential entropy) *If X, Y, Z are independent, then:*

$$h(X-Z) \leq h(X-Y) + h(Y-Z) - h(Y).$$

Equivalently, for arbitrary random variables X, Y, Z:

$$\mathrm{dist}_R(X,Z) \leq \mathrm{dist}_R(X,Y) + \mathrm{dist}_R(Y,Z).$$

The proof of the discrete version of this result in [16] is based on the discrete entropy analog of the bound,

$$h(X,Y,Z) + h(X-Z) \leq h(X-Y, Y-Z) + h(X,Z), \tag{7}$$

which is proved using the functional submodularity Lemma 2.1. Although in general Lemma 2.1 fails for differential entropy, we may try to adapt its proof in this particular setting. However the obvious modification of the discrete proof in the continuous case also fails; the analog of the first inequality in the proof of Lemma 2.1, corresponding to $H(X_{12}) \leq H(X_1, X_2)$, is,

$$h(X,Y,Z) \leq h(X-Y, Y-Z, X, Z),$$

which is false, since $(X-Y, Y-Z, X, Z)$ is concentrated on a lower-dimensional subspace of $\mathbb{R}^4$, and so the term on the right side is $-\infty$. Nevertheless, the actual inequality (7) does hold true.

Lemma 3.2 *The inequality (7) holds true for any three independent random variables X, Y, Z, and it is equivalent to the following data processing inequality:*

$$I(X; (X-Y, Y-Z)) \geq I(X; X-Z). \tag{8}$$

Proof. Inequality (8) is an immediate consequence of data processing, since $X - Z = (X - Y) + (Y - Z)$, therefore, X and $X - Z$ are conditionally independent given $(X - Y, Y - Z)$. To see that it is equivalent to (7), note that the right-hand side of (8) is,

$$h(X - Z) - h(X - Z|X) = h(X - Z) - h(Z),$$

while the left-hand side is,

$$\begin{aligned} h(X - Y, Y - Z) + h(X) - h(X, X - Y, Y - Z) &= h(X - Y, Y - Z) + h(X) - h(X, Y, Z) \\ &= h(X - Y, Y - Z) - h(Y) - h(Z), \end{aligned}$$

where the first equality above follows from the fact that the linear map, $(x, x-y, y-z) \mapsto (x, y, z)$ has determinant 1. □

Proof of Theorem 3.1. Rearranging (7) and using independence,

$$h(X - Z) \leq h(X - Y, Y - Z) - h(Y) \leq h(X - Y) + h(Y - Z) - h(Y).$$

This is easily seen to be the same as the claimed inequality upon substituting the definition of the Ruzsa distances in terms of differential entropies. □

Replacing Y by $-Y$, the triangle inequality yields:

Lemma 3.3 *If X, Y, Z are independent, then:*

$$h(X - Z) + h(Y) \leq h(X + Y) + h(Y + Z).$$

In a similar vein we also have:

Lemma 3.4 *If X, Y, Z are independent, then,*

$$h(X + Y + Z) + h(Y) \leq h(X + Y) + h(Y + Z),$$

which is equivalent to the data processing inequality,

$$I(X + Y + Z; X) \leq I(X + Y; X). \tag{9}$$

Proof. The equivalence of the two stated inequalities follows from the observation that

$$\begin{aligned} I(X + Y + Z; X) &= h(X + Y + Z) - h(X + Y + Z|X) \\ &= h(X + Y + Z) - h(Y + Z|X) \\ &= h(X + Y + Z) - h(Y + Z), \end{aligned}$$

and similarly,

$$I(X + Y; X) = h(X + Y) - h(X + Y|X) = h(X + Y) - h(Y|X) = h(X + Y) - h(Y).$$

By the data processing inequality for mutual information, and the assumed independence,

$$I(X + Y + Z; X) \leq I(X + Y, Z; X) = I(X + Y; X) + I(Z; X|X + Y) = I(X + Y; X)$$

which proves (9) and hence the lemma. □

Combining the last two lemmas, yields:

Theorem 3.5 (Doubling-difference inequality) *If X_1, X_2 are i.i.d., then:*

$$\frac{1}{2} \leq \frac{h(X_1 + X_2) - h(X_1)}{h(X_1 - X_2) - h(X_1)} \leq 2.$$

Equivalently:

$$\frac{1}{2} \leq \frac{I(X_1 + X_2; X_2)}{I(X_1 - X_2; X_2)} \leq 2.$$

Proof. For the upper bound, from Lemma 3.4 taking $X, -Y$ and Z i.i.d., we have,

$$h(X + Z) + h(Y) \leq h(X + Y + Z) + h(Y) \leq h(X + Y) + h(Z + Y),$$

so that,

$$h(X + Z) + h(X) \leq 2h(X - Z),$$

or,

$$h(X + Z) - h(X) \leq 2[h(X - Z) - h(X)],$$

as required. For the lower bound, Lemma 3.3 with X, Y, Z i.i.d. yields,

$$h(X - Y) + h(X) \leq 2h(X + Y),$$

i.e.,

$$h(X - Y) - h(X) \leq 2[h(X + Y) - h(X)],$$

which is the stated lower bound. The fact that the entropy bounds are equivalent to the corresponding mutual information bounds can be established easily as in the first part of the proof of Lemma 3.4. □

Theorem 3.5 examines a basic constraint that the differential entropies of the sum and difference of two i.i.d. random variables place on each other. These quantities are, of course, identical when the density under consideration is symmetric, but there does not seem to be an immediate intuitive reason for them to be mutually constraining in the case when the difference $X_1 - X_2$ has a symmetric density but the sum $X_1 + X_2$ does not. Indeed, Lapidoth and Pete [8] showed that the entropies of the sum and difference of two i.i.d. random variables can differ by an arbitrarily large amount: Given any $M > 0$, there exist i.i.d. X_1, X_2 of finite differential entropy, such that,

$$h(X_1 + X_2) - h(X_1 - X_2) > M. \tag{10}$$

If we consider the "entropy-increase" due to addition of subtraction,

$$\begin{aligned} \Delta_+ &= h(Y + Y') - h(Y) \\ \Delta_- &= h(Y - Y') - h(Y); \end{aligned}$$

then (10) states that the *difference* $\Delta_+ - \Delta_-$ can be arbitrarily large, while Theorem 3.5 asserts that the *ratio* Δ_+/Δ_- must always lie between $\frac{1}{2}$ and 2.

In other words, we define the *doubling constant* and the *difference constant* of a random variable X as,

$$\sigma[X] = \exp\{h(X + X') - h(X)\} \quad \text{and} \quad \delta[X] = \exp\{h(X - X') - h(X)\},$$

respectively, where X' is an independent copy of X, then Theorem 3.5 says that:

Corollary 3.6 *For any random variable X,*

$$\frac{1}{2}\text{dist}_R(X,X) \leq \text{dist}_R(X,-X) \leq 2\text{dist}_R(X,X),$$

equivalently,

$$\delta[X]^{1/2} \leq \sigma[X] \leq \delta[X]^2.$$

Note. As mentioned on pp. 64-65 of [17], the analog of the above upper bound, $\sigma[X] \leq \delta[X]^2$, in additive combinatorics is established via an application of the Plünnecke-Ruzsa inequalities. It is interesting to note that the entropy version of this result (both in the discrete and continuous case) can be deduced directly from elementary arguments. Perhaps this is less surprising in view of the fact that strong versions of the Plünnecke-Ruzsa inequality can also be established by elementary methods in the entropy setting, and also because of the (surprising and very recent) work of Petridis [12], where an elementary proof of the Plünnecke-Ruzsa inequality for sumsets is given. See Sections 3.2 and 3.4, and the discussion in [15, 11].

We now come to the first result whose proof in the continuous case is necessarily significantly different than its discrete counterpart.

Theorem 3.7 (Sum-difference inequality for differential entropy) *For any two independent random variables X, Y:*

$$h(X+Y) \leq 3h(X-Y) - h(X) - h(Y). \tag{11}$$

Equivalently, for any pair of random variables X, Y,

$$\text{dist}_R(X,-Y) \leq 3\text{dist}_R(X,Y). \tag{12}$$

The equivalence of (12) and (11) follows simply from the definition of the Ruzsa distance. Before giving the proof, we state and prove the following simple version of the theorem in terms of mutual information:

Corollary 3.8 (Sum-difference inequality for information) *For any pair of independent random variables X, Y, and all $0 \leq \alpha \leq 1$:*

$$\alpha I(X+Y;X) + (1-\alpha)I(X+Y;Y) \leq (1+\alpha)I(X-Y;X) + (1+(1-\alpha))I(X-Y;Y).$$

Proof. Subtracting $h(X)$ from both sides of (11) yields

$$h(X+Y) - h(X) \leq 3h(X-Y) - 2h(X) - h(Y),$$

or equivalently,

$$h(X+Y) - h(X+Y|Y) \leq 2[h(X-Y) - h(X-Y|Y)] + [h(X-Y) - h(X-Y|X)],$$

which, in terms of mutual information becomes,

$$I(X+Y;Y) \leq 2I(X-Y;Y) + I(X-Y;X). \tag{13}$$

Repeating the same argument, this time subtracting $h(Y)$ instead of $h(X)$ from both sides, gives,

$$I(X+Y;X) \leq 2I(X-Y;X) + I(X-Y;Y). \tag{14}$$

Multiplying (13) by α, (14) by $(1-\alpha)$, and adding the two inequalities gives the stated result. □

The inequality (11) of Theorem 3.7 is an immediate consequence of the following proposition.

Proposition 3.9 *Suppose X, Y are independent, let $Z = X - Y$, and let (X_1, Y_1) and (X_2, Y_2) be two conditionally independent versions of (X, Y) given Z. If $(X_3, Y_3) \sim (X, Y)$ are independent of (X_1, Y_1, X_2, Y_2), then:*

$$h(X_3 + Y_3) + h(X_1) + h(Y_2) \leq h(X_3 - Y_2) + h(X_1 - Y_3) + h(Z). \tag{15}$$

The proof of the discrete analog of the bound (15) in [16] contains two important steps, both of which fail for differential entropy. First, functional submodularity is used to deduce the discrete version of,

$$h(X_1, X_2, X_3, Y_1, Y_2, Y_3) + h(X_3 + Y_3) \leq h(X_3, Y_3) + h(X_3 - Y_2, X_1 - Y_3, X_2, Y_1), \tag{16}$$

but (16) is trivial because the first term above is equal to $-\infty$. Second, the following simple mutual information identity (implicit in [16]) fails: If $Z = F(X)$ and X, X' are conditionally independent versions of X given Z, then $I(X; X') = H(Z)$. Instead, for continuous random variables, Z and X are conditionally independent given X', and hence,

$$I(X; X') \geq I(X; Z) = h(Z) - h(Z|X) = +\infty.$$

Instead of this, we will use:

Lemma 3.10 *Under the assumptions of Proposition 3.9:*

$$h(Z, Y_1, Y_2) + h(Z) - h(Y_1) - h(Y_2) = h(X_1) + h(X_2).$$

Proof. Expanding and using elementary properties,

$$\begin{aligned} h(Z, Y_1, Y_2) + h(Z) - h(Y_1) - h(Y_2) &= h(Y_1, Y_2|Z) + 2h(Z) - h(Y_1) - h(Y_2) \\ &= h(Y_1|Z) + h(Y_2|Z) + 2h(Z) - h(Y_1) - h(Y_2) \\ &= h(Y_1, Z) + h(Y_2, Z) - h(Y_1) - h(Y_2) \\ &= h(Z|Y_1) + h(Z|Y_2) \\ &= h(X_1 - Y_1|Y_1) + h(X_2 - Y_2|Y_2) \\ &= h(X_1) + h(X_2), \end{aligned}$$

as claimed. □

Proof of Proposition 3.9. The most important step of the proof is the realization that the (trivial) result (16) needs to be replaced by the following:

$$h(Z, X_3, Y_1, Y_2, Y_3) + h(X_3 + Y_3) \leq h(X_3, Y_3) + h(X_3 - Y_2, X_1 - Y_3, X_2, Y_1). \tag{17}$$

Before establishing (17) we note that it implies,

$$h(X_3 + Y_3) \leq h(X_3 - Y_2) + h(X_1 - Y_3) + h(X_2) + h(Y_1) - h(Z, Y_1, Y_2),$$

using the independence of (X_3, Y_3) and (Y_1, Y_2, Z). Combined with Lemma 3.10, this gives the required result.

To establish (17) we first note that, by construction, $X_1 - Y_1 = X_2 - Y_2 = Z$, therefore,

$$\begin{aligned} X_3 + Y_3 &= X_3 + Y_3 + (X_2 - Y_2) - (X_1 - Y_1) \\ &= (X_3 - Y_2) - (X_1 - Y_3) + X_2 + Y_1, \end{aligned}$$

and hence, by data processing for mutual information,

$$I(X_3; X_3 + Y_3) \le I(X_3; X_3 - Y_2, X_1 - Y_3, X_2, Y_1),$$

or, equivalently,

$$\begin{aligned} h(X_3 + Y_3) - h(Y_3) &= h(X_3 + Y_3) - h(X_3 + Y_3 | X_3) \\ &\le h(X_3) + h(X_3 - Y_2, X_1 - Y_3, X_2, Y_1) - h(X_3 - Y_2, X_1 - Y_3, X_2, Y_1, X_3) \\ &= h(X_3) + h(X_3 - Y_2, X_1 - Y_3, X_2, Y_1) - h(Z, Y_1, Y_2, Y_3, X_3), \end{aligned}$$

where the last equality follows from the fact that the linear map, $(z, y_1, y_2, y_3, x_3) \mapsto (x_3 - y_2, y_1 + z - y_3, y_2 + z, y_1, x_3)$, has determinant 1. Rearranging and using the independence of X_3 and Y_3 gives (17) and completes the proof. □

3.2 The differential entropy Plünnecke-Ruzsa inequality

In additive combinatorics, the Plünnecke-Ruzsa inequality for iterated sumsets is a subtle result that was originally established through an involved proof based on the theory of commutative directed graphs; see Chapter 6 of [17]. It is interesting that its entropy version can be proved as a simple consequence of the data processing bound in Lemma 3.4. See also the remark following Corollary 3.6 above.

Theorem 3.11 (Plünnecke-Ruzsa inequality for differential entropy) *Suppose that the random variables $X, Y_1, Y_2, \ldots, Y_n$ are independent, and that, for each i, Y_i is only weakly dependent on $(X + Y_i)$, in that $I(X + Y_i; Y_i) \le \log K_i$ for finite constants $K_1, K_2, \ldots, K_n$. In other words,*

$$h(X + Y_i) \le h(X) + \log K_i, \quad \text{for each } i.$$

Then,

$$h(X + Y_1 + Y_2 + \cdots + Y_n) \le h(X) + \log K_1 K_2 \cdots K_n,$$

or, equivalently,

$$I(X + Y_1 + Y_2 + \cdots + Y_n; Y_1 + Y_2 + \cdots + Y_n) \le \log K_1 K_2 \cdots K_n.$$

Proof. Using Lemma 3.4:

$$\begin{aligned} h(X + Y_1 + Y_2 + \cdots + Y_n) &\le h(X + Y_1 + Y_2 + \cdots + Y_{n-1}) + h(X + Y_n) - h(X) \\ &\le h(X + Y_1 + Y_2 + \cdots + Y_{n-1}) + \log K_n, \end{aligned}$$

and continuing inductively yields the result. □

By an application of the entropy Plünnecke-Ruzsa inequality we can establish the following bound on iterated sums.

Theorem 3.12 (Iterated sum bound) *Suppose X, Y are independent random variables, let $(X_0, Y_0), (X_1, Y_1), \dots, (X_n, Y_n)$ be i.i.d. copies of (X, Y), and write $S_i = X_i + Y_i$ for the sums of the pairs, $i = 0, 1, \dots, n$. Then:*

$$h(S_0 + S_1 + \cdots + S_n) \leq (2n+1)h(X+Y) - nh(X) - nh(Y).$$

Proof. Suppose the result is true for $n = 1$. Then, for each i,

$$h(S_0 + S_i) \leq 3h(X+Y) - h(X) - h(Y) = h(S_0) + [2h(X+Y) - h(Y) - h(Y)],$$

and the general claim follows from an application of the entropy Plünnecke-Ruzsa inequality (Theorem 3.11). The case $n = 1$ is an immediate consequence of the following lemma (which generalizes Lemma 3.4) with $X \sim Z$ and $Y \sim W$. □

Lemma 3.13 *If X, Y, Z, W are independent, then:*

$$h(X+Y+Z+W) + h(Y) + h(Z) \leq h(X+Y) + h(Y+Z) + h(Z+W).$$

Proof. Applying Lemma 3.4 with $Z + W$ in place of Z,

$$h(X+Y+Z+W) + h(Y) \leq h(X+Y) + h(Y+Z+W),$$

and using Lemma 3.4 again on the last term above,

$$h(X+Y+Z+W) + h(Y) \leq h(X+Y) + h(Y+Z) + h(Z+W) - h(Z).$$

Rearranging, proves the claim. □

Let us briefly comment on the interpretation of Theorem 3.12. The result may be rewritten as

$$(n+1)h(X+Y) - h\Big(\sum_{i=1}^n X_i + \sum_{i=1}^n Y_i\Big) \geq n[h(X) + h(Y) - h(X+Y)],$$

and hence as

$$h(X_0 + Y_0, \dots, X_n + Y_n) - h\Big(\sum_{i=1}^n X_i + \sum_{i=1}^n Y_i\Big) \geq n[h(X,Y) - h(X+Y)]. \tag{18}$$

Thus the "differential entropy loss from summation" of the collection of $n+1$ independent random variables $\{X_i + Y_i : i = 1, \dots, n\}$ is at least n times the "differential entropy loss from summation" of the two independent random variables $\{X, Y\}$. (In the discrete case, one would have a stronger interpretation as the entropy loss would be precisely the *information* lost in addition.)

3.3 The differential entropy Balog-Szemerédi-Gowers lemma

The differential entropy version of the *Balog-Szemerédi-Gowers lemma* stated next says that, if X, Y are weakly dependent and $X + Y$ has small entropy, then there exist conditionally independent versions of X, Y that have almost the same entropy, and whose *independent* sum still has small entropy.

Theorem 3.14 (Balog-Szemerédi-Gowers lemma for differential entropy) *Suppose X, Y are weakly dependent in the sense that $I(X;Y) \leq \log K$, i.e.,*

$$h(X,Y) \geq h(X) + h(Y) - \log K, \tag{19}$$

for some $K \geq 1$, and suppose also that,

$$h(X+Y) \leq \frac{1}{2}h(X) + \frac{1}{2}h(Y) + \log K. \tag{20}$$

Let X_1, X_2 be conditionally independent versions of X given Y, and let Y' be a conditionally independent version of Y, given X_2 and Y; in other words, the sequence X_2, Y, X_1, Y' forms a Markov chain. Then:

$$h(X_2|X_1,Y) \geq h(X) - \log K \tag{21}$$

$$h(Y'|X_1,Y) \geq h(Y) - \log K \tag{22}$$

$$h(X_2+Y'|X_1,Y) \leq \frac{1}{2}h(X) + \frac{1}{2}h(Y) + 7\log K. \tag{23}$$

Following the corresponding development in [16] for discrete random variables, first we establish a weaker result in the following proposition. The main step in the proof – which is also a very significant difference from the proof of the discrete version of the result in [16] – is the identification of the "correct" data processing bound (26) that needs to replace the use of functional submodularity.

Proposition 3.15 (Weak Balog-Szemerédi-Gowers lemma) *Under the assumptions of Theorem 3.14, we have:*

$$h(X_1 - X_2|Y) \leq h(X) + 4\log K.$$

Proof. Let X_1 and X_2 be conditionally independent as above, and let (X_1, Y, X_2) and (X_1, Y'', X_2) be conditionally independent versions of (X_1, Y, X_2), given (X_1, X_2). We claim that,

$$h(X_1, X_2, Y, Y'') + h(X_1 - X_2, Y) \leq h(X_1, X_2, Y) + h(X_1 + Y'', X_2 + Y'', Y), \tag{24}$$

which is equivalent to,

$$h(X_1, X_2, Y''|Y) + h(X_1 - X_2|Y) \leq h(X_1, X_2|Y) + h(X_1 + Y'', X_2 + Y''|Y). \tag{25}$$

This follows from the data processing argument:

$$\begin{aligned}
&h(X_1 - X_2|Y) \\
&= I(X_1 - X_2; X_1|Y) + h(X_1 - X_2|X_1, Y) \\
&\leq I(X_1 + Y'', X_2 + Y''; X_1|Y) + h(X_2|Y) \\
&= h(X_1 + Y'', X_2 + Y''|Y) + h(X_1|Y) - h(X_1 + Y'', X_2 + Y'', X_1|Y) + h(X_2|Y) \\
&= h(X_1, X_2|Y) + h(X_1 + Y'', X_2 + Y''|Y) - h(X_1 + Y'', X_2 + Y'', X_1|Y) \\
&= h(X_1, X_2|Y) + h(X_1 + Y'', X_2 + Y''|Y) - h(X_1, X_2, Y''|Y),
\end{aligned} \tag{26}$$

where the last equality follows from the fact that the linear map $(x_1, x_2, y) \mapsto (x_1 + y, x_2 + y, x_1)$ has determinant -1. This establishes (25) and hence (24).

We now deduce the result from (24). By the independence bound for joint entropy, the second term in the right-hand side of (24) is,

$$h(X_1 + Y'', X_2 + Y'', Y) \leq 2h(X + Y) + h(Y).$$

By conditional independence and the chain rule, the first term in the right-hand side of (24) is,

$$h(X_1, X_2, Y) = h(X_1, X_2|Y) + h(Y) = h(X_1|Y) + h(X_2|Y) + h(Y) = 2h(X, Y) - h(Y).$$

Using this, conditional independence, and the independence bound for joint entropy, the first term in the left-hand of (24) is,

$$\begin{aligned} h(X_1, X_2, Y, Y'') &= h(X_1, X_2) + h(Y, Y''|X_1, X_2) \\ &= h(X_1, X_2) + h(Y|X_1, X_2) + h(Y''|X_1, X_2) \\ &= 2h(X_1, X_2, Y) - h(X_1, X_2) \\ &= 4h(X, Y) - 2h(Y) - h(X_1, X_2) \\ &\geq 4h(X, Y) - 2h(Y) - 2h(X). \end{aligned}$$

And by the chain rule, the second term in the left-hand side of (24) is,

$$h(X_1 - X_2, Y) = h(X_1 - X_2|Y) + h(Y).$$

Finally combining all the above estimates yields,

$$h(X_1 - X_2|Y) + h(Y) + 4h(X, Y) - 2h(Y) - 2h(X) \leq 2h(X + Y) + h(Y) + 2h(X, Y) - h(Y),$$

or,

$$h(X_1 - X_2|Y) \leq 2h(X + Y) + h(Y) - 2h(X, Y) + 2h(X),$$

and the claim then follows from the assumptions in (19), (20). □

The proof of Theorem 3.14, given next, is similar. Again, the key step is an application of the data processing inequality in (29).

Proof of Theorem 3.14. The bound (21) immediately follows from (19) and the definitions:

$$h(X_2|X_1, Y) = h(X_2|Y) = h(X|Y) = h(X, Y) - h(Y) \geq h(X) - \log K.$$

Similarly, (22) follows from (19):

$$h(Y'|X_1, Y) = h(Y'|X_1) = h(Y|X) = h(X, Y) - h(X) \geq h(Y) - \log K.$$

For (23), we first claim that the following holds,

$$h(X_1, X_2, Y, Y') + h(X_2 + Y', Y) \leq h(X_2, Y', Y) + h(X_1 - X_2, X_1 + Y', Y), \tag{27}$$

or, equivalently,

$$h(X_1, X_2, Y'|Y) + h(X_2 + Y'|Y) \leq h(X_2, Y'|Y) + h(X_1 - X_2, X_1 + Y'|Y). \tag{28}$$

As in the previous proof, this follows from a data-processing argument,

$$\begin{aligned} h(X_2+Y'|Y) &= I(X_2+Y';X_2|Y)+h(X_2+Y'|X_2,Y) \\ &\le I(X_1-X_2,X_1+Y';X_2|Y)+h(Y'|X_2,Y) \\ &= h(X_1-X_2,X_1+Y'|Y)+h(X_2|Y)-h(X_1-X_2,X_1+Y',X_2|Y)+h(Y'|Y) \\ &= h(X_1-X_2,X_1+Y'|Y)-h(X_1,X_2,Y'|Y)+h(X_2,Y'|Y), \end{aligned} \tag{29}$$

where the last equality follows from the fact that X_2 and Y' are conditionally independent given Y, and also from the fact that the linear map $(x_1,x_2,y)\mapsto(x_1-x_2,x_1+y,x_2)$ has determinant -1. This proves (28) and hence (27).

As in the previous proof, we bound each of the four terms in (27) as follows. By the chain rule and conditional independence, the first term is,

$$\begin{aligned} h(X_1,X_2,Y,Y') &= h(X_1,X_2,Y)+h(Y'|X_1,X_2,Y) \\ &= h(X_1,X_2,Y)+h(Y'|X_1) \\ &= h(Y)+2h(X|Y)+h(X,Y)-h(X) \\ &= 3h(X,Y)-h(X)-h(Y). \end{aligned}$$

By the chain rule the second term is, $h(X_2+Y',Y)=h(X_2+Y'|Y)+h(Y)$, and by the independence entropy bound for the third term we have,

$$h(X_2,Y',Y)\le h(X_2,Y)+h(Y')=h(X,Y)+h(Y).$$

Finally by the chain rule and the fact that conditioning reduces entropy,

$$\begin{aligned} h(X_1-X_2,X_1+Y',Y) &\le h(Y)+h(X_1-X_2|Y)+h(X_1+Y') \\ &= h(X_1-X_2|Y)+h(Y)+h(X+Y). \end{aligned}$$

Substituting these into (27) gives,

$$h(X_2+Y'|Y)\le h(X_1-X_2|Y)+2h(Y)+h(X)+h(X+Y)-2h(X,Y),$$

and applying the weak Balog-Szemerédi-Gowers lemma of Proposition 3.15 together with (19) and (20), yields,

$$h(X_2+Y'|X_1,Y)=h(X_2+Y'|Y)\le\frac{1}{2}h(X)+\frac{1}{2}h(Y)+7\log K,$$

as claimed. □

Let us comment some more on the interpretation of Theorem 3.14. The conditions assumed may be rewritten as follows:

1. The variables (X_2,Y,X_1,Y') form a Markov chain, with the marginal distributions of the pairs $(X_2,Y),(X_1,Y)$ and (X_1,Y') all being the same as the distribution of (X,Y).
2. $I(X;Y)\le\log K$, i.e., if we represent the Markov chain (X_2,Y,X_1,Y') on a graph, then the information shared across each edge is the same (by 1.) and it is bounded by $\log K$.
3. $I(X+Y;X)+I(X+Y;Y)\le 2\log K$.

The first 2 parts of the conclusion may be rewritten as:

1. $I(X_2; X_1, Y) \leq \log K$;
2. $I(Y'; X_1, Y) \leq \log K$.

These are obvious from looking at the graph structure of the dependence. To rewrite the third part of the conclusion, note that,

$$\begin{aligned} h(X) + h(Y) - 2h(X_2 + Y'|X_1, Y) &= [h(X_2) - h(X_2|X_1, Y)] + [h(Y') - h(Y'|X_1, Y)] \\ &\quad + [h(X_2|X_1, Y) - h(X_2 + Y'|X_1, Y)] \\ &\quad + [h(Y'|X_1, Y) - h(X_2 + Y'|X_1, Y)] \\ &= I(X_2; X_1, Y) + I(Y'; X_1, Y) \\ &\quad + I(X_2 + Y'; Y'|X_1, Y) + I(X_2 + Y'; X_2|X_1, Y), \end{aligned}$$

so that using the first 2 parts of the conclusion, the third part says that

$$I(X_2 + Y'; Y'|X_1, Y) + I(X_2 + Y'; X_2|X_1, Y) \leq 16 \log K.$$

This is not the same as saying that the boundedness of $I(X + Y; X) + I(X + Y; Y)$ for the dependent pair (X, Y) translates to boundedness of the corresponding quantity for independent X and Y with the same marginals (since conditioning will change the marginal distributions), but it does mean that if we embed the dependent pair (X_1, Y) into a Markov chain that has X_2 and Y' at the ends, one has boundedness *on average* of the corresponding *conditional* quantity for the pair (X_2, Y') (which is conditionally independent given X_1 and Y).

3.4 Sumset bounds for discrete entropy

Here we give a brief discussion of the discrete versions of the results presented so far in this section, their origin and the corresponding discrete proofs.

The discrete version of the Ruzsa triangle inequality as in Theorem 3.1 was given in [14] and [16]. The analog of Lemma 3.2 for discrete random variables was established in [16], and of Lemma 3.4 in [11]. The discrete entropy version of the lower bound in the doubling-difference inequality of Theorem 3.5 is implicit in [14] and [16], and the corresponding upper bound is implicitly derived in [11]. The discrete version of the sum-difference inequality of Theorem 3.7 is proved in [16]; the form given in Corollary 3.8 in terms of mutual information is new even in the discrete case, as is Lemma 3.10.

The discrete analog of Proposition 3.9 is implicit in [16]. The Plünnecke-Ruzsa inequality (Theorem 3.11) for discrete random variables in implicitly proved in [7], and explicitly stated and discussed in [15]. The iterated sum bound of Theorem 3.12 in the discrete case is implicit in [16], while the discrete versions of the strong and weak forms of the Balog-Szemerédi-Gowers lemma (Theorem 3.14 and Proposition 3.15) are both given in [16].

Finally, in the unpublished notes of Tao and Vu [18], the following is stated as an exercise:

Proposition 3.16 (Ruzsa covering lemma for Shannon entropy) *Suppose X, Y are independent discrete random variables, and let (X_1, Y_1), (X_2, Y_2) be versions of (X, Y) that are conditionally independent given $X + Y$. Then:*

$$H(X_1, X_2, Y_1|Y_2) = 2H(X) + H(Y) - H(X + Y). \tag{30}$$

We give a proof below for the sake of completeness, but first we note that the result actually fails for differential entropy: By construction we have that $Y_2 = X_1 + Y_1 - X_2$, therefore, the left-hand side of the continuous analog of (30) is,

$$h(X_1, X_2, Y_1|Y_2) = h(X_1, X_2, Y_2|Y_2) = -\infty.$$

Proof. Since, by definition, $X_1 + Y_1 = X_2 + Y_2$, the triplet $(X_1, X_2, X_1 + Y_1)$ determines all four random variables. Therefore, by data processing for the discrete entropy and elementary properties, we have,

$$\begin{aligned}
H(X_1, X_2, Y_1|Y_2) &= H(X_1, X_2, Y_1, Y_2) - H(Y_2) \\
&= H(X_1, X_2, X_1 + Y_1) - H(Y) \\
&= H(X_1 + Y_1) + H(X_1, X_2|X_1 + Y_1) - H(Y) \\
&= H(X + Y) + H(X_1|X_1 + Y_1) + H(X_2|X_1 + Y_1) - H(Y) \\
&= 2H(X, X + Y) - H(X + Y) - H(Y) \\
&= 2H(X, Y) - H(X + Y) - H(Y) \\
&= 2H(X) + H(Y) - H(X + Y),
\end{aligned}$$

as claimed. □

4 A Differential Entropy Inverse Sumset Theorem

The inverse sumset theorem of Freiman-Green-Ruzsa states that, if a discrete set is such that the cardinality of the sumset $A+A$ is close to the cardinality of A itself, then A is "as structured as possible" in that it is close to a generalized arithmetic progression; see [5] or [17] for details. Roughly speaking, the discrete entropy version of this result established by Tao [16] says that, if X, X' are i.i.d. copies of a discrete random variable and $H(X+X')$ is not much larger than $H(X)$, then the distribution of X is close to the uniform distribution on a generalized arithmetic progression. In other words, if the doubling constant $\sigma[X] = \exp\{H(X+X') - H(X)\}$ is small, then X is close to having a maximum entropy distribution.

Here we give a quantitative version of this result for continuous random variables. First we note that the entropy power inequality [3] for i.i.d. summands states that,

$$e^{2h(X+X')} \geq 2e^{2h(X)},$$

or, equivalently, recalling the definition of the doubling constant from Section 3.1,

$$\sigma[X] := \exp\{h(X'+X) - h(X)\} \geq \sqrt{2},$$

with equality iff X is Gaussian. Note that, again, the extreme case where $h(X+X')$ is as close as possible to $h(X)$ is attained by the distribution which has maximum entropy on $\mathbb{R}$, subject to a variance constraint.

Next we give conditions under which the doubling constant $\sigma[X]$ of a continuous random variable is small only if the distribution of X is appropriately close to being Gaussian. Recall that the *Poincaré constant* $R(X)$ of a continuous random variable X is defined as,

$$R(X) = \sup_{g \in H_1(X)} \frac{E[g(X)^2]}{E[g'(X)^2]},$$

where the supremum is over all functions g in the space $H_1(X)$ of absolutely continuous functions with $E[g(X)] = 0$ and $0 < \mathrm{Var}(g(X)) < \infty$. As usual, we write $D(f\|g)$ for the relative entropy $\int f \log(f/g)$ between two densities f and g.

Theorem 4.1 (Freiman-Green-Ruzsa theorem for differential entropy) *Let X be an arbitrary continuous random variable with density f.*

(i) *$\sigma[X] \geq \sqrt{2}$, with equality iff X is Gaussian.*

(ii) *If $\sigma[X] \leq C$ and X has finite Poincaré constant $R = R(X)$, then X is approximately Gaussian in the sense that,*

$$\frac{1}{2}\|f - \phi\|_1^2 \leq D(f\|\phi) \leq \Big(\frac{2R}{\sigma^2} + 1\Big) \log\Big(\frac{C}{\sqrt{2}}\Big),$$

where σ^2 is the variance of X and ϕ denotes the Gaussian density with the same mean and variance as X.

At first sight, the assumption of a finite Poincaré constant may appear unnecessarily restrictive in the above result. Indeed, we conjecture that this assumption may be significantly relaxed. On the other hand, a related counterexample by Bobkov, Chistyakov and Götze [1] suggests that there is good reason for caution.

Theorem 4.2 *Let X be an arbitrary continuous random variable with density f and finite variance. Let ϕ be the Gaussian density with the same mean and variance as f.*

(i) $\sigma[X] \leq \sqrt{2}\,\exp\{D(f\|\phi)\}$, *with equality iff X is Gaussian,*

(ii) [1] *For any $\eta > 0$ there exists a continuous random variable X with,*

$$\sigma[X] > (\sqrt{2} - \eta)\exp\{D(f\|\phi)\}, \tag{31}$$

but with a distribution well separated from the Gaussian, in that,

$$\|f - \phi\|_1 > C, \tag{32}$$

where C is an absolute constant independent of η.

Proof Theorem 4.1. Part (i) follows from the entropy power inequality, as discussed in the beginning of the section, and the first inequality in part (ii) is simply Pinsker's inequality [3].

For the main estimate, assume without loss of generality that X has zero mean, and recall that Theorem 1.3 of [6] says that,

$$D\Big(\frac{X+X'}{\sqrt{2}}\Big) \leq D(X)\Big(\frac{2R}{\sigma^2+2R}\Big),$$

where, for any finite-variance, continuous random variable Y with density g, $D(Y)$ denotes the relative entropy between g and the normal density ϕ_Y with the same mean and variance as Y. Since $D(Y)$ can be expanded to, $D(Y) = h(\phi_Y) - h(Y)$, the above expression simplifies to,

$$h\Big(\frac{X+X'}{\sqrt{2}}\Big) - h(X) \geq \Big(\frac{\sigma^2}{2R+\sigma^2}\Big)[h(\phi) - h(X)], \tag{33}$$

or,

$$\log\Big(\frac{C}{\sqrt{2}}\Big) \geq \log\Big(\frac{\sigma[X]}{\sqrt{2}}\Big) \geq \Big(\frac{\sigma^2}{2R+\sigma^2}\Big)D(f\|\phi),$$

as claimed. □

Proof of Theorem 4.2. As in the last proof, for any finite-variance, continuous random variable Y with density g, write $D(Y)$ for the relative entropy between g and the normal density ϕ_Y with the same mean and variance as Y, so that $D(Y) = h(\phi_Y) - h(Y)$. Then, letting X, X' be two i.i.d. copies of X,

$$\begin{aligned} D\Big(\frac{X+X'}{\sqrt{2}}\Big) &= h(\phi) - h\Big(\frac{X+X'}{\sqrt{2}}\Big) \\ &= h(X) - h(X+X') + h(\phi) - h(X) + \log\sqrt{2} \\ &= \log\Big(\frac{\sqrt{2}\exp\{D(f\|\phi)\}}{\sigma[X]}\Big). \end{aligned} \tag{34}$$

The result of part (i) follows from (34) upon noting that relative entropy is always nonnegative. Part (ii) is a simple restatement of the counterexample in Theorem 1.1 of [1]. Taking in their result $\epsilon = -\log(1-\eta/\sqrt{2})$, we are guaranteed the existence of an absolute constant C and a random variable X such that (32) holds and $D(X+X') < \epsilon$. But, using (34) this translates to,

$$-\log\Big(1-\frac{\eta}{\sqrt{2}}\Big) = \epsilon > D(X+X') = D\Big(\frac{X+X'}{\sqrt{2}}\Big) = \log\Big(\frac{\sqrt{2}\exp\{D(f\|\phi)\}}{\sigma[X]}\Big),$$

and, rearranging, this is exactly condition (31). The fact that X can be chosen to have finite variance is a consequence of the remarks following Theorem 1.1 in [1]. □

We close this section by noting that the two results above actually generalize to the *difference constant*,

$$\delta[X] := \exp\{h(X - X') - h(X)\},$$

for i.i.d. copies X, X' of X, in place of the doubling constant $\sigma[X]$.

Corollary 4.3 *Let X be an arbitrary continuous random variable with density f.*

(i) *$\delta[X] \geq \sqrt{2}$, with equality iff X is Gaussian.*

(ii) *If $\delta[X] \leq C$ and X has finite Poincaré constant $R = R(X)$, then X is approximately Gaussian in the sense that,*

$$\frac{1}{2}\|f - \phi\|_1^2 \leq D(f\|\phi) \leq \Big(\frac{2R}{\sigma^2} + 1\Big) \log\Big(\frac{C^2}{\sqrt{2}}\Big),$$

where σ^2 is the variance of X and ϕ denotes the Gaussian density with the same mean and variance as X.

Proof. Since the entropy power inequality [3] holds for arbitrary independent random variables, the proof of (i) is identical to that in Theorem 4.1, with $-X'$ in place of X'. For (ii) we assume again without loss of generality that X has zero mean and recall that, from Theorem 3.5, we have,

$$h(X + X') \leq 2h(X - X') - h(X).$$

Combining this with the estimate (33) obtained in the proof of Theorem 4.1, yields,

$$2h(X - X') - 2h(X) - \log\sqrt{2} \geq \Big(\frac{\sigma^2}{2R + \sigma^2}\Big)[h(\phi) - h(X)].$$

or,

$$\log\Big(\frac{C^2}{\sqrt{2}}\Big) \geq \log\Big(\frac{\delta[X]}{\sqrt{2}}\Big) \geq \Big(\frac{\sigma^2}{2R + \sigma^2}\Big) D(f\|\phi),$$

as claimed. □

Corollary 4.4 *Let X be an arbitrary continuous random variable with density f and finite variance. Let ϕ be the Gaussian density with the same mean and variance as f.*

(i) *$\delta[X] \leq \sqrt{2}\exp\{D(f\|\phi)\}$, with equality iff X is Gaussian,*

(ii) *For any $\eta > 0$ there exists a continuous random variable X with,*

$$\delta[X] > (\sqrt{2} - \eta)\exp\{D(f\|\phi)\}, \tag{35}$$

but with a distribution well separated from the Gaussian, in that,

$$\|f - \phi\|_1 > C, \tag{36}$$

where C is an absolute constant independent of η.

Proof. As in the proof of Theorem 4.2, with $-X'$ in place of X' we have,

$$\begin{aligned} 0 \leq D\Big(\frac{X-X'}{\sqrt{2}}\Big) &= h(\phi) - h\Big(\frac{X-X'}{\sqrt{2}}\Big) \\ &= h(X) - h(X-X') + h(\phi) - h(X) + \log\sqrt{2} \\ &= \log\Big(\frac{\sqrt{2}\exp\{D(f\|\phi)\}}{\delta[X]}\Big), \end{aligned}$$

giving (i). Part (ii) follows from Theorem 1.1 of [1] exactly as in the proof of the corresponding result in Theorem 4.2, since the distribution of the random variables in the counterexample given in [1] can be taken to be symmetric, so that $X + X'$ has the same distribution as $X - X'$. □

References

[1] S. G. Bobkov, G. P. Chistyakov, and F. Götze. Stability problems in Cramer-type characterization in case of i.i.d. summands. *Preprint*, 2011.

[2] A.S. Cohen and R. Zamir. Entropy amplification property and the loss for writing on dirty paper. *Information Theory, IEEE Transactions on*, 54(4):1477 –1487, April 2008.

[3] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. J. Wiley, New York, 1991.

[4] R.H. Etkin and E. Ordentlich. The degrees-of-freedom of the K-user Gaussian interference channel is discontinuous at rational channel coefficients. *Information Theory, IEEE Transactions on*, 55(11):4932 –4946, Nov. 2009.

[5] B. Green and I.Z. Ruzsa. Freiman's theorem in an arbitrary abelian group. *Journal of the London Mathematical Society*, 75(1):163–175, 2007.

[6] O. Johnson and A.R. Barron. Fisher information inequalities and the central limit theorem. *Probability Theory and Related Fields*, 129(3), 1581-1583 2001.

[7] V.A. Kaimanovich and A.M. Vershik. Random walks on discrete groups: Boundary and entropy. *The Annals of Probability*, 11(3):pp. 457–490, 1983.

[8] A. Lapidoth and G. Pete. On the entropy of the sum and of the difference of two independent random variables. *Proc. IEEEI 2008, Eilat, Israel*, 2008.

[9] M. Madiman. On the entropy of sums. In *Information Theory Workshop, 2008. ITW '08. IEEE*, pages 303–307, May 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 *2010 ISIT Proceedings, IEEE International Symposium on Information Theory*, pages 1369 –1372, June 2010.

[11] M. Madiman, A. Marcus, and P. Tetali. Entropy and set cardinality inequalities for partition-determined functions. *Random Structures & Algorithms*, 40: 399–424, 2012. [Online] `http://arxiv.org/abs/0901.0055`.

[12] G. Petridis. New proofs of Plünnecke-type estimates for product sets in groups. *Preprint, arXiv:1101.3507*, 2011.

[13] I.Z. Ruzsa. Sums of finite sets. In G.V. Chudnovsky D.V. Chudnovsky and M.B. Nathanson, editors, *Number Theory: New York Seminar*. Springer-Verlag, 1996.

[14] I.Z. Ruzsa. Sumsets and entropy. *Random Structures & Algorithms*, 34(1):1–10, 2009.

[15] T. Tao. An entropy Plünnecke-Ruzsa inequality. At `http://terrytao.wordpress.com/`, blog entry, October 27, 2009.

[16] T. Tao. Sumset and inverse sumset theory for Shannon entropy. *Combinatorics, Probability and Computing*, 19:603–639, 2010.

[17] T. Tao and V. Vu. *Additive combinatorics*. Cambridge studies in advanced mathematics. Cambridge University Press, 2006.

[18] T. Tao and V. Vu. Entropy methods. Available at: `www.math.ucla.edu/~tao/` . Unpublished notes, 2006.

[19] Y. Wu, S. Shamai (Shitz), and S. Verdú. A general formula for the degrees of freedom of the interference channel. *Preprint*, 2011.