Sumset Inequalities for Differential Entropy and Mutual Information

Ioannis Kontoyiannis
Department of Informatics
Athens University of Economics & Business
Patission 76
Athens 10434, Greece.
yiannis@aueb.gr

Mokshay Madiman
Department of Statistics
Yale University
24 Hillhouse Avenue
New Haven, CT 06511, USA.
mokshay.madiman@yale.edu

***Abstract*—The Plünnecke-Ruzsa *sumset theory* gives bounds connecting the cardinality of the sumset $A+B$ defined as $\{a+b\,;\,a\in A, b\in B\}$ with the cardinalities of the original sets A, B. For example, the sum-difference bound states that, $|A+B|\,|A|\,|B| \le |A-B|^3$, where $A-B=\{a-b\,;\,a\in A,\,b\in B\}$. Interpreting the differential entropy $h(X)$ as (the logarithm of) the size of the effective support of X, the main results here are a series of natural information-theoretic analogs for these bounds. For example, the sum-difference bound becomes the new inequality, $h(X+Y)+h(X)+h(Y) \le 3h(X-Y)$, for independent X, Y. Our results include differential-entropy versions of Ruzsa's triangle inequality, the Plünnecke-Ruzsa inequality, and the Balog-Szemerédi-Gowers lemma. 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 several cases requiring substantially new proof strategies. The basic property that naturally replaces functional submodularity is the data processing property of mutual information.**

I. Introduction

The field of *additive combinatorics* [14] provides tools that allow us to count the number of occurrences of particular additive structures in specific subsets of a discrete group. 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.

The Plünnecke-Ruzsa *sumset theory* offers an important collection of tools, consisting primarily of *sumset inequalities* [14]. 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, we have the obvious bounds,

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

as well as much more subtle results, like the Ruzsa triangle inequality [10],

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

or the sum-difference bound [10],

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

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

Recall that Shannon's asymptotic equipartition property (AEP) [2] 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 like, e.g., $|A+B|$, and corresponding bounds for the entropy of sums of discrete random variables, e.g., $H(X+Y)$. First identified by Ruzsa [11], this connection has also been explored in the last few years in different directions by, among others, Tao and Vu [15], Lapidoth and Pete [6], Madiman and Kontoyiannis [8], and Madiman, Marcus and Tetali [9].

Most recently (and most extensively), this connection was developed by Tao in [13]. 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. Thus, for independent discrete random variables X, Y, Z, the simple bounds (1) become [2], respectively,

$$H(X), H(Y) \le H(X+Y) \le H(X)+H(Y),$$

and Ruzsa's bounds (2) and (3) become [13],

$$H(X-Z)+H(Y) \le H(X-Y)+H(Y-Z)$$
$$\text{and } H(X+Y)+H(X)+H(Y) \le 3H(X-Y).$$

[0]I.K. was supported, in part, by a Marie Curie International Outgoing Fellowship, PIOF-GA-2009-235837. M.M. was supported by the NSF CAREER grant DMS-1056996 and by NSF grant CCF-1065494.

Following recent work reported in [7][8], our main motivation is to examine the extent to which this analogy can be carried further: According to the AEP [2], 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 bounds, many of which were proved for the discrete Shannon entropy in [4][15][7][11][13][9].

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 II. The fact that differential entropy is *not* functionally submodular was the source of the main difficulty as well as the excitement for the present development. Instead, the main technical ingredient in our proofs is the *data processing* property of mutual information. Indeed, most of our results 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 these results.

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 the bounds developed here may have applications in core information-theoretic problems. Preliminary connections in this direction can be found in the recent work [1][3][16].

II. Preliminaries

Throughout the paper, $\log$ denotes the natural logarithm $\log_e$, the discrete entropy is defined in nats (in terms of $\log_e$), and the alphabet of any discrete random variable X is assumed to be a (finite or countably infinite) subset A 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 [2],

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

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 [2],

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

gives the upper bound,

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

Our starting point is the recent work of Tao [13], where a series of sumset bounds are established for $H(X)$, beginning with the elementary inequalities (4) and (5). The arguments in [13] are largely based on the following important observation [13][9]:

Lemma 2.1 (Functional submodularity of discrete 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, $H(X_1) + H(X_2) - H(X_{12}) \geq H(X_1) + H(X_2) - H(X_1, X_2) = I(X_1; X_2) \geq I(X_0; X_0) = H(X_0)$. □

Our main goal in this work is to examine the extent to which the bounds in [13] and in earlier work extend to the continuous case. The differential entropy $h(X)$ of a continuous random variable (or random vector) X is defined in nats, and 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 function of X is typically different from that of X itself, even for linear functions [2]: 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,

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

as can be easily seen using standard properties of the entropy and mutual information.

Our overall development 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.

III. Differential Entropy Sumset Bounds

Unless explicitly stated otherwise, all random variables are assumed to be continuous (i.e., taking real values with an absolutely continuous density with respect to Lebesgue measure), 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.

A. Ruzsa distance and the doubling and difference constants

In analogy with the corresponding definition for discrete random variables [13], 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): 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 [13] is based on the discrete entropy analog of the bound,

$$h(X,Y,Z) + h(X-Z) \leq h(X-Y, Y-Z) + h(X,Z), \quad (7)$$

which is proved using functional submodularity. Although in general it fails for differential entropy, we may try to adapt the proof of Lemma 2.1 itself in this particular setting. But 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 the last term, $h(X-Y, Y-Z, X, Z) = -\infty$. 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:

$$I(X; (X-Y, Y-Z)) \geq I(X; X-Z)$$

The proof of Theorem 3.1 based on Lemma 3.2 and the proof of the lemma itself are both given in [8].

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,

$$I(X+Y+Z; X) \leq I(X+Y; X).$$

The proof of Lemma 3.4 is given in [7][8]. Combining the last two lemmas, yields:

Theorem 3.5 (Doubling-difference inequality): If X_1, X_2 are independent and identically distributed (i.i.d.), then:

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

Equivalently:

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

If we define the *doubling constant* and the *difference constant* of a random variable X as,

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

respectively, where X' is an independent copy of X, then Theorem 1 says that:

Corollary 3.6: For any random variable X,

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

equivalently,

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

Note. As mentioned on pp. 64-65 of [14], 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. See Section III-B and the discussion in [12].

The proofs of Theorem 3.5 and Corollary 3.6 are given in [8].

We now come to the first result whose proof in the continuous case is necessarily significantly different than its discrete counterpart. It is also the only major result here for which we give a complete proof. For the proofs of all other results, see [5].

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

$$h(X+Y) \leq 3h(X-Y) - h(X) - h(Y). \quad (8)$$

Equivalently, for any pair X, Y,

$$\mathrm{dist}_R(X,-Y) \leq 3\mathrm{dist}_R(X,Y). \quad (9)$$

The equivalence of (9) and (8) 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 $0 \leq \alpha \leq 1$:

$$\begin{aligned}\alpha I(X+Y;X) + (1-\alpha) I(X+Y;Y) \\ \leq (1+\alpha) I(X-Y;X) + (1+(1-\alpha)) I(X-Y;Y).\end{aligned}$$

Proof. Applying (8) with X, Y in place of X', Y', and subtracting $h(X)$ from both sides, yields,

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

or equivalently,

$$\begin{aligned}h(X+Y) - h(X+Y|Y) \leq & 2[h(X-Y) - h(X-Y|Y)] \\ & + [h(X-Y) - h(X-Y|X)],\end{aligned}$$

which, in terms of mutual information becomes,

$$I(X+Y;Y) \leq 2I(X-Y;Y) + I(X-Y;X). \tag{10}$$

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

$$I(X+Y;X) \leq 2I(X-Y;X) + I(X-Y;Y). \tag{11}$$

Multiplying (10) by α, (11) by $(1-\alpha)$, and adding the two inequalities gives the stated result. □

The main result (8) of Theorem 3.7 is a simple 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:

$$\begin{aligned}h(X_3+Y_3) + & h(X_1) + h(Y_2) \\ & \leq h(X_3 - Y_2) + h(X_1 - Y_3) + h(Z).\end{aligned} \tag{12}$$

The proof of the discrete analog of the bound (12) in [13] contains two important steps, both of which fail for differential entropy. First, functional submodularity is used to deduce the discrete version of,

$$\begin{aligned}h(X_1, X_2, X_3, Y_1, Y_2, Y_3) + h(X_3+Y_3) \\ \leq h(X_3, Y_3) + h(X_3 - Y_2, X_1 - Y_3, X_2, Y_1),\end{aligned} \tag{13}$$

but (13) is trivial because the first term above is equal to $-\infty$. Second, the following simple mutual information identity (implicit in [13]) 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, we have that $h(Z, Y_1, Y_2) + h(Z) - h(Y_1) - h(Y_2)$ equals,

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

as claimed. □

Proof of Proposition 3.9. The most important step of the proof is the realization that the (trivial) result (13) needs to be replaced by the following:

$$\begin{aligned}h(Z, X_3, Y_1, Y_2, Y_3) + h(X_3+Y_3) \\ \leq h(X_3, Y_3) + h(X_3 - Y_2, X_1 - Y_3, X_2, Y_1).\end{aligned} \tag{14}$$

Before establishing (14) we note that it implies,

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

which, combined with Lemma 3.10, gives the required result.

To establish (14) 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,

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

or, equivalently, $h(X_3+Y_3) - h(Y_3)$ equals

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

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 (14) and completes the proof. □

B. The differential entropy Plünnecke-Ruzsa inequality

In additive combinatorics, the Plünnecke-Ruzsa inequality for iterated sumsets is a subtle result with an involved proof based on the theory of commutative directed graphs; see Chapter 6 of [14]. It is interesting that its entropy version can be proved as a simple consequence of the data processing bound in Lemma 3.4; see [5] for details.

Theorem 3.11 (Plünnecke-Ruzsa inequality): 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 $I(X + Y_i; Y_i) \leq \log K_i$ for finite constants $K_1, K_2, \ldots, K_n$. In other words,

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

Then,

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

or, equivalently,

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

By an application of the entropy Plünnecke-Ruzsa inequality we can establish the following bound on iterated sums; see [5] for the proof.

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

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

C. 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 3.13: (Balog-Szemerédi-Gowers lemma) Suppose that X, Y are weakly dependent in the sense that $I(X; Y) \leq \log K$, i.e.,

$$h(X, Y) \geq h(X) + h(Y) - \log K,$$

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; in other words, the sequence X_2, Y, X_1, Y' forms a Markov chain. Then:

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

Following the corresponding development in [13] for discrete random variables, first we establish a weaker result in the following proposition.

Proposition 3.14: (Weak Balog-Szemerédi-Gowers lemma) Under the assumptions of Theorem 3.13, we have:

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

The proofs of the last two results are both significantly different from the proofs of the corresponding discrete versions in [13]. The main step, in each case, is the identification of the "correct" data processing bound that needs to replace the use of functional submodularity. See [5] for details.

REFERENCES

[1] 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.

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

[3] 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.

[4] 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.

[5] I. Kontoyiannis and M. Madiman. Sumset and inverse sumset inequalities for differential entropy and mutual information. *Preprint*, 2012.

[6] 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.

[7] M. Madiman. On the entropy of sums. In *Information Theory Workshop, 2008. ITW '08. IEEE*, pages 303–307, May 2008.

[8] M. Madiman and I. Kontoyiannis. The entropies of the sum and the difference of two IID random variables are not too different. In *Proceedings of the 2010 IEEE International Symposium on Information Theory*, pages 1369 –1372, June 2010.

[9] M. Madiman, A. Marcus, and P. Tetali. Entropy and set cardinality inequalities for partition-determined functions, with applications to sumsets. *Random Structures & Algorithms*, 2011, DOI: 10.1002/rsa.20385. [Online]. Available: `http://arxiv.org/abs/0901.0055`

[10] 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.

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

[12] T. Tao. An entropy Plünnecke-Ruzsa inequality. Blog entry, at `http://terrytao.wordpress.com/`, October 27, 2009.

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

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

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

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