Entropy bounds on abelian groups and the Ruzsa divergence

Mokshay Madiman, *Member, IEEE* and Ioannis Kontoyiannis, *Fellow, IEEE*

Abstract

Over the past few years, a family of interesting new inequalities for the entropies of sums and differences of random variables has been developed by Ruzsa, Tao and others, motivated by analogous results in additive combinatorics. The present work extends these earlier results to the case of random variables taking values in $\mathbb{R}^n$ or, more generally, in arbitrary locally compact and Polish abelian groups. We isolate and study a key quantity, the *Ruzsa divergence* between two probability distributions, and we show that its properties can be used to extend the earlier inequalities to the present general setting. The new results established include several variations on the theme that the entropies of the sum and the difference of two independent random variables severely constrain each other. Although the setting is quite general, the result are already of interest (and new) for random vectors in $\mathbb{R}^n$. In that special case, quantitative bounds are provided for the stability of the equality conditions in the entropy power inequality; a reverse entropy power inequality for log-concave random vectors is proved; an information-theoretic analog of the Rogers-Shephard inequality for convex bodies is established; and it is observed that some of these results lead to new inequalities for the determinants of positive-definite matrices. Moreover, by considering the multiplicative subgroups of the complex plane, one obtains new inequalities for the differential entropies of products and ratios of nonzero, complex-valued random variables.

Index Terms

Ruzsa divergence; additive combinatorics; Shannon entropy; differential entropy; Haar measure; abelian groups; inequalities; sumset bounds; entropy power inequality; log-concave; Rogers-Shephard inequality.

M. Madiman is with the Department of Mathematical Sciences, University of Delaware, 501 Ewing Hall, Newark DE 19716, USA. Email: madiman@udel.edu

I. Kontoyiannis is with the Department of Informatics, Athens University of Economics and Business, Patission 76, Athens 10434, Greece. Email: yiannis@aueb.gr

Manuscript dates. M. Madiman was supported by NSF CAREER grant DMS-1056996 and NSF grant CCF-1065494. I. Kontoyiannis was supported by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) Research Funding Program “Thales - Investing in knowledge society through the European Social Fund.” Parts of this paper were presented at the IEEE Information Theory Workshop 2013 in Seville, Spain, and at the Eighth Workshop on Information Theoretic Methods in Science and Engineering (WITMSE 2015) in Copenhagen, Denmark.

I. Introduction

A. Motivation

THE properties of the entropy of sums and differences of random variables have attracted a great deal of interest in almost every area of information theory. Classical results were primarily motivated by the study of additive noise channels, and in the past three decades connections with several other fields have emerged, including the foundations of probabilistic limit theorems, functional inequalities and probabilistic bounds.

More recently, it was also observed that inequalities involving the entropies of sums and differences are closely tied to basic questions and results in the area of additive combinatorics, which in turn also have applications in communications. A prominent collection of tools in additive combinatorics are those provided by the Plünnecke-Ruzsa sumset theory; see, e.g., [55] for a broad introduction. A simple example of such a result is the following. Given two discrete sets A and B, the *sumset* $A+B$ is defined as, $A+B=\{a+b\,:\,a\in A, b\in B\}$, and the *difference set* $A-B$ is, $A-B=\{a-b\,:\,a\in A, b\in B\}$. The *Ruzsa triangle inequality* [48] states that, for any three sets A,B,C, we have,

$$|A-C|\cdot|B|\leq|A-B|\cdot|B-C|, \tag{1}$$

where $|E|$ denotes the cardinality of a set E, and A, B and C are subsets of the integers, or any other discrete abelian group. A fascinating connection between such inequalities and corresponding results for the Shannon entropy H was identified initially by Tao and Vu [56] and by Ruzsa [49], and it has been developed quite extensively by several authors over the past 10 years; see, e.g., [34], [54] and the references therein. The main idea is that, interpreting the entropy as the effective log-cardinality of the support of a random variable, then replacing the log-cardinality of every sumset (or difference set) by the entropy of a corresponding sum (respectively, difference) of independent discrete random variables, produces a candidate entropy inequality. For example, (1) becomes,

$$H(X-Z)+H(Y)\leq H(X-Y)+H(Y-Z), \tag{2}$$

for independent X,Y,Z, where H denotes the Shannon entropy.

For discrete random variables, this connection was studied in detail by [34] and Tao [54], who established numerous such entropy inequalities. The main technical tool in Tao's proofs was the submodularity property of the discrete entropy, which, as observed in our subsequent work [30], fails to hold in the case of differential entropy. Therefore, in order to extend Tao's results to continuous random variables, new arguments were necessary, and the key property which replaced submodularity in the proofs of almost all of the corresponding differential entropy inequalities in [30], was the data processing inequality for mutual information. As for the results of [34], some of them can be extended without too much effort to continuous random variables, while others rely too much on the bijection-invariance of discrete entropy and cannot be extended both because of this and because of delicate measure-theoretic issues that only arise in the continuous case.

The starting point of the present work is the desire to explore how this family of inequalities can be extended to random vectors $\mathbb{R}^n$ and, more generally, to random variables taking values in general (locally compact, Polish) abelian groups. Our main results, outlined below, include unified proofs for many of the earlier results in [54], [34] and all the results of [30]; a key ingredient in our approach is the identification of the *Ruzsa divergence* as the central quantity of interest.

We note in passing that strong communication-theoretic motivation for the present work comes from the fact that our results can be used powerfully in the study of the degrees of freedom of interference channels (for which the computation of fundamental limits is a notoriously hard open problem). The results of our prior work [30] played a key role in the works of Wu, Shamai and Verdú [57] and Stotz and Bölcskei [52], [53]; we anticipate that the more general results developed herein will also find applications to communication theory.

B. Outline of main results

The first contribution of this work is to isolate and study, in a general setting, a quantity that plays a key role in the behavior of entropy of sums and differences; we call this the Ruzsa divergence. Let X and Y denote two random variables which can be discrete, continuous, vector-valued, or, more generally, taking values in a locally compact abelian group G. The *Ruzsa divergence* between X and Y is defined as,[1]

$$d_R(X\|Y) := h(X'-Y') - h(X') = I(X'-Y';Y'),$$

where X' and Y' are independent and have the same marginal distributions as X and Y, respectively, and h denotes the entropy on G. As described formally in the following section, h is the usual Shannon entropy if G is discrete, it is the (joint) differential entropy when $G=\mathbb{R}^n$, and in general it is the entropy defined with respect to Haar measure on G. Much of the remainder of this section will summarize how the basic properties of the Ruzsa divergence can be used to provide unified proofs for all existing (discrete and continuous) entropy inequalities in this area, as well as their extensions to general groups, offering an analysis on spaces satisfying essentially minimal assumptions – specifically, on abelian groups equipped with the minimal topological structure necessary to guarantee the existence of a Haar measure so that a natural notion of entropy can be defined.

The second contribution of this work is to highlight some interesting connections of the aforementioned techniques and ideas with problems related to the differential entropies of products of positive random variables, the entropy power inequality, results in convex geometry, and determinantal inequalities.

We begin in Section II by introducing the main definitions and assumptions that will remain in effect throughout the paper. We first formally define the Ruzsa divergence $d_R(X\|Y)$, as well as two related quantities, the conditional Ruzsa divergence and the Ruzsa difference. After some elementary observations, we then state in Theorem 1 the triangle inequality for $d_R(X\|Y)$, which implies the inequality (2), and which is seen to be a simple consequence of a stronger result, Theorem 2. This is stated and proved in Section III, where we also establish a number of the important properties of $d_R(X\|Y)$. In Theorem 3 we show that it is subadditive with respect to convolution, $d_R(X\|Y_1+Y_2)\leq d_R(X\|Y_1)+d_R(X\|Y_2)$, and in Theorem 5 we give a general information-theoretic version of the Balog-Szemeredi-Gowers theorem, a significant inequality from additive combinatorics.

In Section IV we first re-interpret the subadditivity property of Theorem 3 in the context of important inequalities for the cardinalities of sumsets in additive combinatorics, called the Plünnecke-Ruzsa inequalities. Specifically, in Theorem 6 we observe that, if $X, Y_1, Y_2, \ldots, Y_n$ are independent, then,

$$h\left(X-\sum_{i=1}^n Y_i\right) + (n-1)h(X) \leq \sum_{i=1}^n h(X-Y_i).$$

We then examine the question of how different the entropies of $X+X'$ and $X-X'$ can be, when X and X' are independent and identically distributed (i.i.d.). As was pointed out by Lapidoth and Pete [31], the difference between the two can be arbitrarily large, which may be rephrased as saying that $d_R(X\|X')$ and $d_R(X\|-X')$ can differ by an arbitrarily large amount. However, in Corollary 3 we show that the ratio between these two Ruzsa divergences is always bounded between $1/2$ and 2; this generalizes the doubling-difference inequality of [30]. In Theorem 7 we give the general version of the sum-difference inequality [30], relating $h(X+X')$ and $h(X-X')$ [equivalently, relating $d_R(X\|X')$ and $d_R(X\|-X')$] when X and X' are independent but not necessarily identically distributed. We close this section by giving general versions of some recent results by Wu, Shamai and Verdú [57] on discrete random variables, which were used in a study of the degrees of freedom of the M-user interference channel. In Lemma 6 and Theorem 8 we state and prove corresponding results for the entropy of weighted linear combinations of random variables of the form $aX+bY$, where X, Y take values in a general (locally compact and Polish) abelian group, and a, b are integers.

In Section V, we consider the special cases of three subgroups G of the complex plane $\mathbb{C}$, equipped with the multiplication operation: the half-line $(0,\infty)$, the unit circle $T\subseteq\mathbb{C}$, and the nonzero complex numbers $\mathbb{C}\setminus\{0\}$. In each of these cases, the application of our general results lead to new inequalities for the differential entropies of products and ratios of G-valued random variables.

In the last four sections we concentrate on the special case of real random vectors, taking $G=\mathbb{R}^n$ and h to be the usual (joint) differential entropy. In Section VI we look at the difference between $h(X+X')$ and $h(X-X')$

[1] Although a symmetrical variant of this quantity, namely $\frac{1}{2}(d_R(X\|Y)+d_R(Y\|X))$, has been called the "Ruzsa distance" has been studied before in the discrete setting by Tao [54] and for real-valued random variables in [30], we find that focusing on this non-symmetric version makes various developments clearer. Furthermore, the Ruzsa divergence is a particular instance of the Kullback-Leibler divergence or relative entropy, so that it inherits many of its characteristics, but it also has special properties that justify its close study.

from a different perspective, and provide results in the spirit of the Freiman-Green-Ruzsa inverse sumset theorems. In Corollary 7 we show (under certain conditions), based on a recent result from [7], that if $h(X+X')-2h(X)$ is small, then the distribution of X is necessarily close to being Gaussian, in a way that can be precisely quantified in terms of relative entropy. Then, in Theorem 10 we prove a converse result: If the two entropies $h(X-X')$ and $h(X+X')$ are significantly different, then the distribution of X will also be significantly different (in the relative entropy sense) from being Gaussian. These results can be seen as quantitative versions of the condition for equality in the entropy power inequality [50], [51]. Recall that, when applied to i.i.d. random vectors X, X', the entropy power inequality implies that,

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

where, throughout the paper, $\log$ denotes the natural logarithm $\log_e$, so that the entropy and all other familiar information-theoretic quantities are expressed in nats. In Section VII we establish a reverse inequality of this sort: Corollary 8 states that, if X, X' are i.i.d. with a log-concave distribution, then,

$$h(X+X') \leq h(X) + n\log 2.$$

In Section VIII we argue that the Ruzsa divergence is a natural analog of volume-based functionals that arise in the geometry of convex sets. In Corollary 9 we establish the following information-theoretic analog of the Rogers-Shepard inequality: If X and X' are i.i.d. with a log-concave distribution on $\mathbb{R}^n$, then,

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

In fact, we conjecture that the same result holds without the factor of 2 in the last term above. Finally, in Section IX, we briefly indicate how the earlier inequalities for the entropy can be used to develop corresponding inequalities for the determinants of positive-definite matrices. In particular, in Corollary 11 we establish the following variant of an inequality due to Rotfel'd [45]: If $K, K_1, K_2, \ldots, K_n$ are positive-definite matrices, then,

$$\det(K+K_1+\ldots+K_n) \leq [\det(K)]^{-(n-1)}\prod_{j=1}^{n}\det(K+K_j).$$

II. The Ruzsa divergence

We begin by introducing the basic definitions of Haar measure and random variables with values in an abelian group. Readers not interested in the general formulation can simply skip to the two main examples below, and read the rest of this paper keeping only these two key examples in mind.

Let G be an abelian topological group, i.e., a topological space endowed with a commutative, associative and continuous operation (i.e., a continuous function from $G \times G$ to G that takes (x, y) to an element of G denoted $x + y$), which has an identity element 0 (such that $x + 0 = x$ for all x in G) and with every element having an inverse (i.e., for each $x \in G$ there is an element in G denoted $-x$ such that $x + (-x) = 0$). We will always assume that the topology on G is Polish (i.e., it is metrizable so that the resulting metric space is complete and separable), and locally compact (i.e., every point has a compact neighborhood). The Borel σ-algebra $\mathcal{G}$ on G is the σ-algebra generated by all open sets. It is a classical fact (see, e.g., [25], [40], [24]) that under these assumptions, there exists a (countably additive) measure λ defined on $\mathcal{G}$ that is translation-invariant, i.e., such that $\lambda(A + x) = \lambda(A)$ for each $A \in \mathcal{G}$ and each $x \in G$, where $A + x = \{a + x : a \in A\}$. Such a measure is called a Haar measure, and it is unique up to scaling by a positive constant. In any given situation, we will assume that the scaling is chosen at the beginning and fixed; thus we will talk without further comment about "the" Haar measure on G.

For our analysis, the normalization (particular scaling chosen) of the Haar measure does not matter. Nonetheless, it is useful to keep in mind the common normalizations used for the most important examples – namely discrete groups and the additive group $\mathbb{R}^n$. When G is a countable group with the discrete topology, we will always take the Haar measure λ to be counting measure, i.e., $\lambda(\{g\}) = 1$ for every element $g \in G$, and define λ on any subset of G as its (possibly infinite) cardinality. When G is not compact, the Haar measure is infinite, and then it is common to fix the normalization by fixing the measure of some special set; in the case of $\mathbb{R}^n$, as usual, by requiring $\lambda([0, 1]^n) = 1$, we obtain the Lebesgue measure.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and X be a G-valued random variable on it (i.e., a function from Ω to G measurable with respect to $\mathcal{F}$ and $\mathcal{G}$). We say that the random variable X taking values in G has a continuous distribution if its probability distribution, namely the image measure P_X induced by the mapping X on $\mathcal{G}$, is absolutely continuous with respect to the Haar measure λ. In this case, denoting the Radon-Nikodym derivative $\frac{dP_X}{d\lambda}(x)$ by $f(x) = f_X(x)$, we say that X has density f, and write $X \sim f$.

Example 1. *When G is countable, every G-valued random variable X has a continuous distribution, and its density is simply the probability mass function of X, i.e., $f_X(x) = \mathbb{P}\{X = x\}$.*

Example 2. *Let G be the set $\mathbb{R}^n$ equipped with the addition operation, so that λ is the usual Lebesgue measure. Let X be a G-valued random variable. If X is a continuous random variable, then its density is the usual probability density function $f_X : G \to [0, \infty)$ of the random vector X with respect to Lebesgue measure, satisfying,*

$$\mathbb{P}\{X \in B\} = \int_B f_X(x)dx,$$

for each $B \in \mathcal{G}$, where $\mathcal{G}$ is the collection of Borel subsets of $\mathbb{R}^n$.

If X has density f on the group G, the *entropy* of X is defined by,

$$h(X) = -\int_G f(x)\log f(x)\,dx,$$

provided that the integral exists in the Lebesgue sense. As usual, we write $h(X)$ even though the entropy depends only on the density f of X. Clearly, h is precisely the discrete entropy in the setting of Example 1, and the differential entropy in the setting of Example 2.

To summarize, *we assume throughout that G is a Polish, locally compact, abelian group,* equipped with the Haar measure λ on its Borel σ-field $\mathcal{G}$. Then it is easy to check that the same properties are satisfied by the Cartesian product G^n (with coordinate-wise addition defining the group structure, the product topology defining the topological structure, and the product measure λ^n being its Haar measure), for any $n \in \mathbb{N}$. Thus we can define the entropy of any finite collection of jointly distributed random variables $(X_1, \dots, X_n)$, each with values in G, simply by treating $(X_1, \dots, X_n)$ as a measurable function from Ω to the Cartesian product G^n, and computing the entropy of its density. [Generally we will not use the common term "joint entropy," since we prefer to think of the collection of random variables as a single random object.] In particular, we can define the conditional entropy between two G-valued random elements X and Y by the usual chain rule, as $h(Y|X) = h(X, Y) - h(X)$.

Although particular care is needed to see which of the standard properties of discrete entropy and differential entropy carry over to the general case, we note that it is immediate from the definition that some key properties remain true. First, the entropy is always translation-invariant in that, for any constant $a \in G$, $h(X + a) = h(X)$, because of the translation-invariance of the Haar measure. Also, the chain rule holds in general, and, if we define the mutual information as usual as a difference of entropies, the chain rule for mutual information also holds in this general setting. Finally, the property which will play the most central role in our subsequent development, namely the data processing inequality for mutual information, also holds in complete generality.

Definition 1. *Suppose X and Y are G-valued random variables with finite entropy. The* Ruzsa divergence *between X and Y is defined as,*

$$d_R(X\|Y) := h(X' - Y') - h(X'),$$

where X' and Y' are taken to be independent random variables with the same distributions as X and Y, respectively.

Let us note that even though the entropies of X and Y above are assumed to be finite, it is possible that $h(X'-Y')$ and hence $d_R(X\|Y)$ are $+\infty$ (see, e.g., [12] for examples). In order to avoid uninteresting technicalities, in the statements of all subsequent definitions and results, *we will always implicitly assume that the entropies and Ruzsa divergences that appear are well-defined and finite*. The adjustments that need to be made to address possible infinities are left to the reader; see, e.g., the discussion after Lemma 7 where we work out explicitly the precise finiteness conditions in one particular case.

A more precise way of writing the Ruzsa divergence would have been to write it as $d_R(f_1\|f_2)$, where $X \sim f_1$ and $Y \sim f_2$, but we find it convenient to highlight the random vectors in the notation. The term "divergence" is designed to invoke comparison with the relative entropy or Kullback-Leibler divergence (in that d_R also satisfies some properties of a distance but not others, e.g., it is not symmetric); in fact, it is immediately obvious that the Ruzsa divergence is just a special case of the mutual information (and hence of the relative entropy).

Lemma 1. *For any two G-valued random variables X, Y,*

$$d_R(X\|Y) = I(X' - Y'; Y'),$$

where $I(Z; W) = h(Z) + h(W) - h(Z, W)$ denotes the mutual information between Z and W, and $X' \sim X$ and $Y' \sim Y$ are independent. In particular, $d_R(X, Y) \geq 0$.

Observe that $d_R(X\|X) = I(X - X'; X)$, where X' is an independent copy of X, and this is rarely identically zero. In particular, when $G = \mathbb{R}^n$, $d_R(X\|X)$ is never zero, since the entropy power inequality implies a strictly positive lower bound on $d_R(X\|X)$ depending only on n, as discussed in Section VI. Thus even if we ignore the assymmetry of Ruzsa divergence (which can be fixed by averaging $d_R(X\|Y)$ and $d_R(Y\|X)$), one should be careful in interpreting it as a notion of distance.

However, the quantity d_R satisfies a triangle inequality.

Theorem 1 (Triangle inequality for Ruzsa divergence)**.** *If X_1, X_2, X_3 are independent, then,*

$$d_R(X_1\|X_3) \leq d_R(X_1\|X_2) + d_R(X_2\|X_3).$$

Theorem 1 was proved originally (in an equivalent form) for discrete random variables by Ruzsa [49]; see also Tao [54]. Since the discrete arguments used in these proofs rely on the property of submodularity which fails in the continuous setting, a different proof for Theorem 1 was recently provided in [30] for real-valued random variables. The proof we present for the general setting in Section III uses both a re-interpretation of the approach used in [30], and a sufficient condition for bijections in locally compact abelian groups to preserve the entropy, recently obtained in [35] and stated in Lemma 5.

We now define a conditional version of the Ruzsa divergence. Throughout this paper, we say that $X \leftrightarrow Z \leftrightarrow Y$ form a Markov chain if they are defined on a common probability space and the conditional distribution of X given (Z, Y) is the same as that of X given Z alone. The assertion that $X \leftrightarrow Z \leftrightarrow Y$ form a Markov chain is easily seen to be symmetric, i.e., it is equivalent to the statement that $Y \leftrightarrow Z \leftrightarrow X$ form a Markov chain.

Definition 2. *Suppose* X_1, Y, *and* X_2 *are* G*-valued random variables, such that* $X_1 \leftrightarrow Y \leftrightarrow X_2$ *forms a Markov chain. The* conditional Ruzsa divergence between X_1 and X_2 given Y *is,*

$$d_R(X_1 \| X_2 | Y) := h(X_1 - X_2 | Y) - h(X_1 | Y).$$

Lemma 2. *If* $X_1 \leftrightarrow Y \leftrightarrow X_2$ *form a Markov chain, then,*

$$d_R(X_1 \| X_2 | Y) = I(X_1 - X_2; X_2 | Y),$$

where $I(Z; W|V) = h(Z|V) + h(W|V) - h(Z, W|V)$ *denotes the conditional mutual information between* Z *and* W*, given* V*. In particular,* $d_R(X_1 \| X_2 | Y) \geq 0$.

Proof: Observe that,

$$\begin{aligned} d_R(X_1 \| X_2 | Y) &= h(X_1 - X_2 | Y) - h(X_1 | Y) \\ &= h(X_1 - X_2 | Y) - h(X_1 | Y, X_2) \\ &= h(X_1 - X_2 | Y) - h(X_1 - X_2 | Y, X_2) \\ &= I(X_1 - X_2; X_2 | Y) \\ &\geq 0. \end{aligned}$$

The Markov condition was used in an essential way in the second equality of the above display, while the translation-invariance of entropy was used in the third equality. ■

Observe that $d_R(X_1 \| X_2 | Y) \neq d_R(X_2 \| X_1 | Y)$ in general, but that both quantities are non-negative.

Finally we introduce a more general version of the Ruzsa divergence, involving dependent random variables.

Definition 3. *The* Ruzsa difference *of the two* G*-valued random variables* X *and* Y *is,*

$$\tilde{d}_R(X \| Y) := h(X - Y) - h(X).$$

Clearly, $\tilde{d}_R(X \| Y) = d_R(X \| Y)$ when X and Y are independent, but in general $\tilde{d}_R(X \| Y)$ is not a divergence and need not be non-negative. Indeed, it is easy to see that one always has the following identity.

Lemma 3. *For any pair of* X, Y*,*

$$\tilde{d}_R(X \| Y) = I(X - Y; Y) - I(X; Y).$$

III. Properties of Ruzsa divergence

A special case of the Markov chain condition $X_1 \leftrightarrow Y \leftrightarrow X_2$ is when X_1 is independent of (Y, X_2). Then, the conditional Ruzsa divergence can be related to the (unconditional) Ruzsa divergence.

Lemma 4. (Conditioning reduces Ruzsa divergence) *If X_1 is independent of (Y, X_2), then,*

$$d_R(X_1 \| X_2) = d_R(X_1 \| X_2 | Y) + I(Y; X_1 - X_2),$$

and, in particular, $d_R(X_1 \| X_2 | Y) \leq d_R(X_1 \| X_2)$.

Proof: By Lemma 2 and the chain rule for mutual information,

$$\begin{aligned} d_R(X_1 \| X_2 | Y) &= I(X_1 - X_2; X_2 | Y) \\ &= I(X_1 - X_2; (X_2, Y)) - I(X_1 - X_2; Y). \end{aligned}$$

But,

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

by translation-invariance of entropy. The assumed independence now implies that,

$$\begin{aligned} I(X_1 - X_2; (X_2, Y)) &= h(X_1 - X_2) - h(X_1) \\ &= d_R(X_1 \| X_2), \end{aligned}$$

so that,

$$\begin{aligned} d_R(X_1 \| X_2 | Y) &= d_R(X_1 \| X_2) - I(Y; X_1 - X_2) \\ &\leq d_R(X_1 \| X_2). \end{aligned}$$

■

To motivate the next property of Ruzsa divergence we will develop, it is useful to consider the special case $G = \mathbb{R}^n$, equipped with Lebesgue measure. In this case, it is an elementary fact that for any matrix $A \in GL_n(\mathbb{R})$ (i.e., any invertible $n \times n$ matrix), and for any random vector X taking values in $\mathbb{R}^n$, $h(AX) = h(X) + \log \det(A)$, where $\det(\cdot)$ denotes the determinant. This has two useful consequences. Firstly, $d_R(X \| AY) = d_R(A^{-1}X \| Y)$ so that in particular, $d_R(X \| -Y) = d_R(-X \| Y)$. Secondly, for any matrix $A \in SL_n(\mathbb{R})$ (i.e., any invertible matrix with determinant 1), entropy is preserved by the corresponding linear transformation, i.e., $h(AX) = h(X)$.

For a general locally compact abelian group G, the notion of a linear transformation on G^n defined by a matrix A no longer makes sense. However, when the elements of an $n \times n$ matrix $A = (a_{ij})_{1 \leq i,j \leq n}$ are integers, we can talk about the group homomorphism induced by A on G^n. Specifically, for $(x_1, \ldots, x_n) = x \in G^n$, we denote by Ax the element,

$$\left(\sum_{j=1}^{n} a_{1j} x_j, \sum_{j=1}^{n} a_{2j} x_j, \ldots, \sum_{j=1}^{n} a_{nj} x_j \right) \in G^n,$$

where ax denotes the element $x + \cdots + x \in G$, added a times. Even though G^n is not a linear space, we will sometimes call an integer matrix A a "linear transformation," with the understanding that this refers to the group homomorphism induced by it as above.

The general linear group over the integer ring $\mathbb{Z}$ (strictly speaking, of the module $\mathbb{Z}^n$), denoted $GL_n(\mathbb{Z})$, is the set of all $n \times n$ matrices with integer entries and determinant $+1$ or -1. The following result was recently shown in [35].

Lemma 5. *Let X be a random variable taking values in G^n. If $A \in GL_n(\mathbb{Z})$, then,*

$$h(AX) = h(X).$$

This allows us to extend the observation that the Ruzsa divergence behaves nicely when the random vectors involved are linearly transformed.

Corollary 1. *For any $A \in GL_n(\mathbb{Z})$, and any pair of G-valued random variables X, Y,*

$$d_R(X\|AY) = d_R(A^{-1}X\|Y).$$

In particular, $d_R(X\|-Y) = d_R(-X\|Y)$.

Proof: Assume, without loss of generality, that X, Y are independent. By Lemma 5,

$$\begin{aligned} d_R(X\|AY) &= h(X - AY) - h(X) \\ &= h(A^{-1}X - Y) - h(A^{-1}X) \\ &= d_R(A^{-1}X\|Y). \end{aligned}$$

■

We now prove a sharpened version of the triangle inequality in Theorem 1.

Theorem 2. *If X_1, X_2, X_3 are independent, then,*

$$d_R(X_1\|X_3) \leq d_R(X_1\|X_2|X_2 - X_3) + d_R(X_2\|X_3).$$

Proof: By an application of Lemma 1 and the data processing inequality for mutual information,

$$\begin{aligned} d_R(X_1\|X_3) &= I(X_1 - X_3; X_3) \\ &\leq I((X_1 - X_2, X_2 - X_3); X_3). \end{aligned}$$

By the chain rule for mutual information, however,

$$\begin{aligned} & I((X_1 - X_2, X_2 - X_3); X_3) \\ = \; & I(X_2 - X_3; X_3) + I(X_1 - X_2; X_3|X_2 - X_3) \\ = \; & d_R(X_2\|X_3) + I(X_1 - X_2; X_3|X_2 - X_3), \end{aligned}$$

where we used Lemma 1 in the last equality. All that remains is to show that,

$$d_R(X_1\|X_2|X_2 - X_3) = I(X_1 - X_2; X_3|X_2 - X_3),$$

or, in view of Lemma 2, that,

$$I(X_1 - X_2; X_2|X_2 - X_3) = I(X_1 - X_2; X_3|X_2 - X_3). \tag{3}$$

Let us observe the following general fact:

$$I(X; Y, Y - Z) = I(X; Y, Z). \tag{4}$$

To see this, write,

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

where the second identity relied on Lemma 5, and the fact that the mapping of (y, z) to $(y, y - z)$ is represented by the 2×2 matrix,

$$\begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix},$$

which has determinant 1.

Then (4) implies that,

$$I(X_1 - X_2; X_2, X_2 - X_3) = I(X_1 - X_2; X_3, X_2 - X_3),$$

since both these quantities equal $I(X_1 - X_2; X_2, X_3)$. Subtracting $I(X_1 - X_2; X_2 - X_3)$ from both sides, we obtain the inequality in (3), completing the proof. ■

Remark 1. *Using Lemma 4, Theorem 2 can be written in a symmetric form as,*

$$d_R(X_1\|X_3) \quad \leq \quad d_R(X_1\|X_2) + d_R(X_2\|X_3) - I(X_1 - X_2; X_2 - X_3),$$

and Theorem 1 immediately follows.

A useful property of Ruzsa divergence is subadditivity in the second argument, which may be equivalently expressed as a monotonicity property in the first argument.

Theorem 3. *If X, Y_1 and Y_2 are independent, then,*

$$d_R(X\|Y_1 + Y_2) \leq d_R(X\|Y_1) + d_R(X\|Y_2).$$

Equivalently, if X_1, X_2 and Y are independent, then,

$$d_R(X_1 + X_2\|Y) \leq d_R(X_1\|Y).$$

Proof: Observe that,

$$\begin{aligned}
&d_R(X\|Y_1 + Y_2) - d_R(X\|Y_1)\\
&= h(X - Y_1 - Y_2) - h(X) - [h(X - Y_1) - h(X)]\\
&= h(X - Y_1 - Y_2) - h(X - Y_1)\\
&= d_R(X - Y_1; Y_2).
\end{aligned}$$

By relabeling variables, we see that the two formulations are equivalent.

To prove the second formulation (and hence also the first), note that by Lemma 1, and the data processing inequality and the chain rule for mutual information,

$$\begin{aligned}
d_R(X_1 + X_2\|Y) \quad &= \quad I(X_1 + X_2 - Y; Y)\\
&\leq \quad I(X_1 - Y, X_2; Y)\\
&= \quad I(X_1 - Y; Y) + I(X_2; Y|X_1 - Y).
\end{aligned}$$

The second term in the last line is 0 since X_2 is independent of (X_1, Y), so that another application of Lemma 1 gives the desired result. ■

Remark 2. *Written out in terms of entropies, Theorem 3 is equivalent to the assertion that the entropy of a sum of independent group-valued random variables is a submodular set function, i.e., $h(X + Y + Z) + h(Z) \leq h(X + Z) + h(Y + Z)$. For discrete entropy, this assertion is implicit in Kaimanovich and Vershik [27], and explicitly and independently developed in [32], [34]; [34] also contains a generalization from sums to a more general class of so-called partition-determined functions that can make sense on sets with less algebraic structure. For differential entropy, this assertion was first presented in [32], and further explored for the case of $\mathbb{R}$-valued random variables in [30].*

If we do not make assumptions about the nature of the underlying distributions, the Ruzsa divergence and conditional Ruzsa divergence can be unbounded. In Sections VII and VIII, we will make such assumptions and demonstrate a uniform bound on Ruzsa divergence for a log-concave density on $\mathbb{R}^n$. On the other hand, it is possible to obtain a bound on conditional Ruzsa divergence under mild assumptions on the dependence structure.

Theorem 4. *If $X_1 \leftrightarrow Y \leftrightarrow X_2$ form a Markov chain, then,*

$$d_R(X_1\|X_2|Y) \quad \leq 2I(X_1; Y) + I(X_2; Y) + \tilde{d}_R(X_1\|Y) + \tilde{d}_R(Y\|X_2).$$

Proof: 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). By the data processing inequality:

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

where the last equality follows from Lemma 5, and the fact that the linear map $(x_1, x_2, y) \mapsto (x_1 + y, x_2 + y, x_1)$ has determinant -1. Therefore,

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

We have established that,

$$h(X_1, X_2, Y, Y'') + h(X_1 - X_2, Y) \leq h(X_1, X_2, Y) + h(X_1 + Y'', X_2 + Y'', Y). \tag{5}$$

We now deduce the result from (5). First note that by conditional independence of Y and Y'' given X_1, X_2, the first term in the left-hand of (5) is,

$$h(X_1, X_2, Y, Y'') + h(X_1, X_2) = h(X_1, X_2, Y) + h(X_1, X_2, Y'') = 2h(X_1, X_2, Y),$$

so that,

$$\begin{aligned} h(X_1 - X_2, Y) &\leq h(X_1 + Y'', X_2 + Y'', Y) \\ &\quad - h(X_1, X_2, Y) + h(X_1, X_2) \\ &\leq \sum_i h(X_i + Y) + h(Y) \\ &\quad - h(X_1, X_2, Y) + h(X_1, X_2). \end{aligned}$$

By conditional independence and the chain rule,

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

Thus,

$$\begin{aligned} h(X_1 - X_2|Y) + h(Y) \quad &\leq \quad \sum_i h(X_i + Y) + h(Y) + h(X_1, X_2) \\ &\quad\quad -[h(X_1|Y) + h(X_2|Y) + h(Y)] \\ &= \quad h(X_1 + Y) - h(X_1|Y) \\ &\quad\quad +h(X_2 + Y) - h(X_2|Y) + h(X_1, X_2). \end{aligned}$$

So,

$$\begin{aligned} &h(X_1 - X_2|Y) - h(X_1|Y) \\ &\leq I(X_1; Y) + \tilde{d}_R(X_1\|Y) + h(X_2 + Y) - h(X_2|Y) + h(X_1, X_2) - h(X_1, Y). \end{aligned}$$

Since,

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

and since,

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

we are done. ■

Let us note two corollaries of Theorem 4. Firstly, if we assume X_1, X_2 and Y to be independent, we recover the Ruzsa triangle inequality (Theorem 1). Secondly, the case where the joint distribution is symmetric in (X_1, X_2) is of interest.

Corollary 2. *Suppose $X_1 \leftrightarrow Y \leftrightarrow X_2$ form a Markov chain, and X_1 and X_2 have the same conditional distribution given Y. Then,*

$$d_R(X_1 \| X_2|Y) \leq 3I(X;Y) + \tilde{d}_R(X\|Y) + \tilde{d}_R(Y\|X).$$

One may interpret this as follows. For every possible value y of Y, consider the Ruzsa divergence between the conditional distribution of X given $Y = y$, and itself; then the conditional Ruzsa divergence $d_R(X_1\|X_2|Y)$ is the average of these quantities under the distribution of Y. This follows from the fact that X_1, X_2 are conditionally i.i.d. given Y. Thus Corollary 2 says that, for weakly dependent random variables X, Y, having bounds on the two (not particularly well behaved) Ruzsa differences between X and Y, allows one to get a bound on this averaged self-divergence of the conditional distribution of X given Y (which is a well behaved divergence).

Let us recall the Balog-Szemeredi-Gowers theorem, which has become an extremely useful tool in additive combinatorics in the last two decades. There are several formulations, but the one we focus on is stated in terms of the *restricted sumset* $A \overset{E}{+} B$, defined as,

$$A \overset{E}{+} B = \{a + b : a \in A, b \in B, (a, b) \in E\},$$

where E is some subset of the Cartesian product $A \times B$. If A and B are finite nonempty subsets of an abelian group G, and $E \subset A \times B$ satisfies $|E| \geq \frac{1}{K}|A| \cdot |B|$ and $|A \overset{E}{+} B| \leq K\sqrt{|A| \cdot |B|}$ for some $K \geq 1$, then there exist subsets $A_0 \subset A$ and $B_0 \subset B$ such that $|A_0| \geq \frac{1}{K}|A|$, $|B_0| \geq \frac{1}{K}|B|$, and

$$|A_0 + B_0| \leq K^7 \sqrt{|A_0| \cdot |B_0|}.$$

The natural probabilistic analogue of a restricted sumset is a sum of dependent random variables. Theorem 4 may be thought of as an information-theoretic form of the Balog-Szemeredi-Gowers theorem, since bounds for dependent random vectors are used to deduce bounds for (conditionally) independent random vectors. It is not directly analogous to the Balog-Szemeredi-Gowers theorem since the bounds are not in terms of the Ruzsa differences between X_1 and X_2, but rather in terms of the Ruzsa differences between either of them and the auxiliary random variable Y. However, such a direct analogue can be constructed using Theorem 4. This was done in the discrete case by Tao [54], and in the case of the additive group $\mathbb{R}$ by the authors in [30]. We state below the resulting theorem in the general setting, using the notation developed in this paper.

Theorem 5. *Let (X_2, Y_1, X_1, Y_2) form a Markov chain, with the marginal distributions of the pairs $(X_2, Y_1), (X_1, Y_1)$ and (X_1, Y_2) all being the same as the distribution of (X, Y). Then,*

$$d_R(X_2\|Y_2|X_1, Y_1) + d_R(Y_2\|X_2|X_1, Y_1) \leq 3I(X;Y) + \tilde{d}_R(X\|Y) + \tilde{d}_R(Y\|X).$$

Proof: The proof of [30, Theorem 3.14] for real-valued random variables carries over almost exactly in the general case, if one uses Lemma 5 to justify one of the steps. This yields, under the present assumptions, that,

$$I(X_2 + Y_2; Y_2|X_1, Y_1) + I(X_2 + Y_2; X_2|X_1, Y_1) \leq I(X;Y) + I(X + Y;X) + I(X + Y;Y).$$

To obtain the desired result in the stated form, one just needs to replace all occurrences of Y, Y_1 or Y_2 by their respective inverses (i.e., $-Y, -Y_1$ or $-Y_2$), and then make appropriate use of Lemma 1, Lemma 2, and Lemma 5. ■

IV. Entropies of weighted sums and differences

The Plünnecke inequality in additive combinatorics [43], [46], [47] states that, if $|A+B|\leq\alpha|A|$ for finite nonempty subsets A, B of an abelian group, then for every $k>1$, there exists a nonempty subset $A'\subset A$ such that

$$|A'+kB|\leq\alpha^k|A'|, \tag{6}$$

where kB refers to the sumset $B+\cdots+B$ with k summands. A very elegant and considerably simpler proof, obtained by Petridis [41], also shows that the same subset A' can be used for all positive integers k. The inequality (6) can be generalized to different summands: if A and B_i are nonempty finite sets, with $|A+B_i|\leq\alpha_i|A|$ for each i, then there exists a nonempty subset $A'\subset A$ such that,

$$|A'+B_1+\ldots+B_m|\leq\left(\prod_{i=1}^m\alpha_i\right)|A'|.$$

This is usually called the Plünnecke-Ruzsa inequality, since it was proved by Ruzsa [46], [47] using an ingenious combinatorial argument. These inequalities are very influential in additive combinatorics– for example, as expounded in [55], they are sufficient to obtain Freiman-type inverse theorems for groups with bounded torsion. The analogue of the Plünnecke-Ruzsa inequality for the entropy is the following subadditivity property of Ruzsa divergence, which is an immediate consequence of Theorem 3; the same historical remarks made in Remark 2 therefore also apply here.

Theorem 6. *If $X, Y_1, \ldots, Y_k$ are independent, then:*

$$d_R\left(X\middle\|\sum_{i=1}^k Y_i\right)\leq\sum_{i=1}^k d_R(X\|Y_i).$$

To see that this is analogous to the Plünnecke-Ruzsa inequality as stated above, we can trivially rewrite it in the following form: if $d_R(X\|Y_i)\leq\alpha_i$, then $d_R(X\|\sum_{i=1}^k Y_i)\leq\sum_{i=1}^k\alpha_i$. Unlike in the case of sets where one potentially needs to pass to a subset to obtain a valid inequality, the entropy analogue works with the original random variables of interest.

The properties of Ruzsa divergence developed in Section III can also be used to understand how the differential entropy of the sum of two independent random vectors constrains the differential entropy of their difference.

Theorem 7. *For any G-valued random variables X, Y,*

$$d_R(X\|-Y)\leq 2d_R(X\|Y)+d_R(Y\|X).$$

Proof: Let (X_1, Y_1) be independent, with $Z=X_1-Y_1$. Assume (X_2, Y_2) is conditionally independent of (X_1, Y_1) given Z, and has the same conditional distribution given Z as (X_1, Y_1); thus in particular $Z=X_2-Y_2$. Let (X, Y) be independent of (X_1, Y_1, X_2, Y_2), but have the same distribution as either pair (X_i, Y_i).

Since, by construction, $X_1-Y_1=X_2-Y_2=Z$,

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

and hence, by data processing for mutual information,

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

where the last equality follows from the fact that the linear map, $(z,y_1,y_2,y,x)\mapsto(x-y_2,y_1+z-y,y_2+z,y_1,x)$, has determinant 1.

Using the independence of X and Y from each other and all other random variables for the second term on the above right-hand side, we have,

$$\begin{aligned} d_R(Y\|-X) &\leq h(X-Y_2)+h(X_1-Y)+h(X_2)+h(Y_1) \\ &\quad -[h(Z,Y_1,Y_2)+h(Y)] \\ &=[d_R(Y\|X)+h(Y)]+[d_R(X\|Y)+h(X)] \\ &\quad +h(X_2)-h(Z,Y_1,Y_2). \end{aligned} \tag{7}$$

However, observe that, since $I(Y_1;Y_2|Z)=0$,

$$\begin{aligned} h(Z,Y_1,Y_2)+h(Z) &= h(Y_1,Z)+h(Y_2,Z) \\ &= h(X_1,Y_1)+h(X_2,Y_2)=2h(X,Y). \end{aligned} \tag{8}$$

Plugging (8) into (7) gives,

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

which is the desired result. ■

In the case where X and Y are not just independent but also identically distributed, Theorem 7 simply says that $d_R(X\|-X)\leq 3d_R(X\|X)$, while taking X and $-Y$ to have the same distribution gives,

$$d_R(X\|X)\leq d_R(X\|-X)+2d_R(X\|-X)=3d_R(X\|-X).$$

In fact, one can obtain tighter bounds in these special cases.

Corollary 3. *If X,Y are i.i.d., then:*

$$\frac{d_R(X\|-X)}{d_R(X\|X)}\in[\tfrac{1}{2},2].$$

Proof: The desired statement is equivalent, for X_1,X_2 that are i.i.d., to:

$$\frac{1}{2}\leq\frac{h(X_1+X_2)-h(X_1)}{h(X_1-X_2)-h(X_1)}\leq 2. \tag{9}$$

As observed in [30], the upper bound in the inequality (9) follows from Theorem 6, and the lower bound follows from Theorem 1, both of which we have already proved for the general setting. ■

Corollary 3 provides inequalities between $h(X+Y)$ and $h(X-Y)$ when X,Y are i.i.d. and $h(X)$ is known. The requirement to know $h(X)$ to make the comparison cannot be dispensed with in the general setting of locally compact abelian groups. However, this requirement can be dispensed with for discrete groups– as observed by [1], $h(X+Y)/h(X-Y)$ must lie between 3/4 and 4/3 if X and Y are i.i.d. random variables in a discrete group.

Finally, let us examine what can be said about weighted sums and differences, i.e., about random variables of the form $aX+bY$ where a,b are non-zero integers. Discrete entropy inequalities for such random variables play a key role in the recent work of Wu, Shamai and Verdú [57] on the degrees of freedom of the M-user interference channel – specifically, they immediately yield inequalities of similar form for the Rényi information dimension of weighted sums of random variables, which imply, using the single-letter characterization of [57], that for rational channel coefficients the number of degrees of freedom is strictly smaller than $M/2$. In the following theorem, we extend all the inequalities proved by [57] for discrete entropy of weighted sums and differences to the general abelian setting. First we give the generalization of [57, Lemma 18].

Lemma 6. *Let X,X' and Z be independent G-valued random variables, where X' has the same distribution as X. Let a,b be nonzero integers. Then:*

$$h(aX+b)\leq h((a-b)X+bX'+Z)+d_R(X\|X).$$

Furthermore, if a is even, then:

$$h(aX+b) \le h\left(\frac{a}{2}X+Z\right) + h(2X-X') - h(X).$$

Proof: One can simply follow the proof strategy of [57, Lemma 18], which on inspection relies only on the subadditivity of Ruzsa divergence and the Ruzsa triangle inequality, both of which we have already proved in the general setting. ■

Finally we give the generalization of [57, Theorem 14]; its proof is again the same as in the discrete case, using the subadditivity of Ruzsa divergence and Ruzsa triangle inequality established earlier. The result of Theorem 8 can be compared to the inequalities of Bukh [18] for dilated sums of sets.

Theorem 8. *Let X and Y be independent G-valued random variables, and a, b be nonnegative integers. Then,*

$$h(aX+bY) - h(X+Y) \le \tau_{a,b} \left\{ d_R(X\| -Y) + d_R(Y\| -X) \right\},$$

where,

$$\tau_{a,b} = 6\big(\lfloor \log|a| \rfloor + \lfloor \log|b| \rfloor + 2\big).$$

V. Entropies of products and ratios

Since we will need to discuss entropies with respect to two different measures on the same group, we introduce some additional notation to keep things unambiguous. All the examples considered in this section involve subgroups G of the group $\mathbb{C}^\times = \mathbb{C}\setminus\{0\}$ equipped with the multiplication operation. The Haar measure for such multiplicative groups is typically not the same as the familiar Lebesgue measure used to compute differential entropies of real-valued or complex-valued random variables (the one-dimensional and two-dimensional Lebesgue measures are Haar measures for the groups $\mathbb{R}$ and $\mathbb{C}$ respectively, but only when the group structure comes from the addition operation).

A. Positive random variables

Consider the group $\mathbb{R}_{>0} = (0, \infty)$ equipped with the multiplication operation. Its Haar measure is given by,

$$\lambda(dx) = \frac{dx}{x},$$

where dx is Lebesgue measure on $(0, \infty)$. To see this, all we need to do is check the translation-invariance of λ with respect to multiplication, i.e., that for any fixed $c > 0$, we have $\lambda(cA) = \lambda(A)$ when

$$\lambda(A) = \int_A \frac{dx}{x},$$

and dx represents Lebesgue measure on $\mathbb{R}$. [And this in turn is an immediate consequence of the fact that the logarithmic function is an isomorphism between $(\mathbb{R}_{>0}, \times)$ and $(\mathbb{R}, +)$, using the standard translation-invariance of Lebesgue measure for addition.]

We are interested in two entropies of a positive (i.e., $\mathbb{R}_{>0}$-valued) random variable X. To define them, let us assume that X has a density f with respect to Lebesgue measure on $(0, \infty)$. Then:

1) The differential entropy of X is,

$$h_{\mathbb{R}}(X) = -\int_0^\infty f(x) \log f(x) dx.$$

2) The intrinsic entropy $h_\times(X)$ with respect to Haar measure λ on $(\mathbb{R}_{>0}, \times)$ is given by,

$$h_\times(X) = -\int_0^\infty [xf(x)] \log[xf(x)] \lambda(dx) = h_{\mathbb{R}}(X) - \mathbf{E}[\log X], \tag{10}$$

since the density of X with respect to λ is $xf(x)$. We use $h_\times$ to emphasize that this is the intrinsic entropy with respect to the multiplicative structure on $\mathbb{R}_{>0}$ rather than the additive structure on $\mathbb{R}$.

Observe that $\mathbb{R}_{>0}$ is a Polish, locally compact, abelian group to which all of our preceding results apply and yield statements of interest. For illustration, we only write out one consequence: Corollary 3 says that,

$$\frac{1}{2} \leq \frac{h_\times(XY) - h_\times(X)}{h_\times(X/Y) - h_\times(X)} \leq 2,$$

which, using relation (10), translates to the following statement for the usual differential entropy.

Corollary 4. *If X, Y are i.i.d. random variables taking values in $(0, \infty)$, then:*

$$\begin{aligned} h_{\mathbb{R}}(XY) &\leq 2h_{\mathbb{R}}(X/Y) - h_{\mathbb{R}}(X) + 3\mathbf{E}[\log X], \\ h_{\mathbb{R}}(X/Y) &\leq 2h_{\mathbb{R}}(XY) - h_{\mathbb{R}}(X) - 3\mathbf{E}[\log X]. \end{aligned}$$

B. Random variables on the circle group

Consider the unit circle $T = \{z \in \mathbb{C} : |z| = 1\}$ in the complex plane; this is of course a group under multiplication, and is isomorphic to $\mathbb{R}/\mathbb{Z}$ equipped with addition via the isomorphism $t \mapsto e^{2\pi i t}$. Alternatively we can parametrize T using the angle θ subtended by the arc of the circle between the point on T and the real axis (which is just $2\pi t$). With this parametrization, the Haar measure λ is the uniform distribution on the angle or, equivalently, Lebesgue measure on $[0, 2\pi)$. For a T-valued random variable Θ that has a density f with respect to the uniform measure,

$$h(\Theta) = -\int_T f(x) \log f(x) \lambda(dx) = -D(\Theta \| U),$$

where $U \sim \lambda$ is uniformly distributed on T, and $D(\Theta \| U)$ denotes the relative between Θ and a uniformly distributed random variable U on T. Thus, the fact that entropy increases on convolution captures in this setting the fact that convolution brings any distribution closer to the uniform.

In this case, Corollary 3 becomes the following statement.

Corollary 5. *If Θ, Θ' are i.i.d. random variables taking values in T, then:*

$$\frac{1}{2} \leq \frac{D(\Theta \| U) - D(\Theta + \Theta' \| U)}{D(\Theta \| U) - D(\Theta - \Theta' \| U)} \leq 2.$$

C. Non-zero complex random variables

Finally we consider the full group $(\mathbb{C}^\times, \times)$, whose Haar measure is given by,

$$\frac{dz}{|z|^2},$$

where dz is 2-dimensional Lebesgue measure (using the identification of $\mathbb{C}$ with $\mathbb{R}^2$). If f is the density of a $\mathbb{C}^\times$-valued random variable Z with respect to 2-dimensional Lebesgue measure, one has the intrinsic entropy,

$$h_\times(Z) = -\int_{\mathbb{C}^\times} [|z|^2 f(z)] \log[|z|^2 f(z)] \frac{dz}{|z|^2} = h_{\mathbb{R}^2}(Z) - \mathbf{E}[\log(|Z|^2)],$$

where we use $h_{\mathbb{R}^2}(Z)$ to denote the usual differential entropy of Z.

Then Corollary 3 becomes the following statement.

Corollary 6. *If Z_1, Z_2 are i.i.d. random variables taking values in $\mathbb{C}^\times$, then:*

$$\begin{aligned} h_{\mathbb{R}^2}(Z_1 Z_2) &\leq 2h_{\mathbb{R}^2}(Z_1/Z_2) - h_{\mathbb{R}^2}(Z_1) + 6\mathbf{E}[\log |Z_1|], \\ h_{\mathbb{R}^2}(Z_1/Z_2) &\leq 2h_{\mathbb{R}^2}(Z_1 Z_2) - h_{\mathbb{R}^2}(Z_1) - 6\mathbf{E}[\log |Z_1|]. \end{aligned}$$

VI. Freiman-type results for the entropy on $\mathbb{R}^n$

For the rest of the paper, our focus is on the additive group $\mathbb{R}^n$ equipped with Lebesgue measure, so that h denotes the usual differential entropy. Our first observation is a uniform lower bound on the Ruzsa divergence between a distribution and itself. A simple application of the entropy power inequality [50][51] to two i.i.d. random variables easily gives the following result.

Lemma 7. *For any $\mathbb{R}^n$-valued random vector X with finite differential entropy,*

$$d_R(X\|X) \geq \frac{n}{2}\log 2.$$

Furthermore, $d_R(X\|-X) \geq \frac{n}{2}\log 2$.

The assumption of finite differential entropy in Lemma 7 is in fact essential. As shown by Bobkov and Chistyakov [12, Proposition 1], there exists a $\mathbb{R}$-valued random variable X of finite entropy such that if X, X' are i.i.d., the entropy of $X+X'$ does not exist. However, [12] also shows that for any such example, necessarily the entropy of X is $-\infty$, so that it remains true that if the entropy exists and is a real number, then the entropy of the self-convolution also exists (although, thanks to another example constructed in [12], it may then be $+\infty$!). Henceforth, as stated in Section II, if nothing is stated, we will assume that all entropies and Ruzsa divergences exist and are finite.

We find it convenient to restate Lemma 7 in terms of the doubling and difference constants associated with a random vector.

Definition 4. *For an $\mathbb{R}^n$-valued random vector X, the entropy power of X is defined as,*

$$\mathcal{N}(X) = \exp\left\{\frac{2h(X)}{n}\right\}.$$

For an $\mathbb{R}^n$-valued random vector X, the doubling constant is defined by,

$$\sigma_+(X) = \frac{\mathcal{N}(X+X')}{2\mathcal{N}(X)},$$

and the difference constant is defined by,

$$\sigma_-(X) = \frac{\mathcal{N}(X-X')}{2\mathcal{N}(X)},$$

where X' is an independent copy of X.

Then entropy power inequality immediately implies that if X has finite entropy, then $\sigma_+(X) \geq 1$ and $\sigma_-(X) \geq 1$; this is just a restatement of Lemma 7 since,

$$\sigma_-(X) = \tfrac{1}{2}\exp\left\{\frac{2}{n}d_R(X\|X)\right\}, \tag{11}$$

and,

$$\sigma_+(X) = \tfrac{1}{2}\exp\left\{\frac{2}{n}d_R(X\|-X)\right\}. \tag{12}$$

Furthermore, because of the equality conditions of the entropy power inequality, $\sigma_+(X)$ (or $\sigma_-(X)$) is equal to 1 if and only if X is a Gaussian (with non-singular covariance matrix). Note that the definitions of doubling and difference constants of scalar random variables in [54] (for discrete random variables) and in [30] (for $\mathbb{R}$-valued random variables) used a different normalization, but we have chosen the normalization above so that the minimum value achieved at Gaussians for both σ_+ and σ_- is 1.

A natural question is whether the extremality of Gaussians is a stable phenomenon. In other words, if $\sigma_+(X) \leq K$ for some K, does this imply that the distribution of X is necessarily not far from being Gaussian, in a sense that can be quantified in terms of K? It is a perhaps somewhat surprising result due to Bobkov, Chistyakov and Götze [14] that the answer is “no,” even in the one-dimensional setting. Nonetheless, as observed in [30], under the additional assumption that X has a finite Poincaré constant (and using results independently obtained by Johnson and Barron [26] and Artstein, Ball, Barthe and Naor [3] on the rate of convergence in the information-theoretic central limit theorem for $\mathbb{R}$-valued random variables) it can be shown that such a stability bound can indeed be established.

This result cannot be directly extended to the case of $\mathbb{R}^n$-valued random vectors, since non-asymptotic bounds that exhibit convergence rates for the entropic central limit theorem in the multivariate case are not known under just the assumption of a finite Poincaré constant[2]. However, by relying on recent work of Ball and Nguyen [7], one can see that such stability does hold under the stronger assumption of log-concavity.

Recall that a probability density function f defined on $\mathbb{R}^n$ is said to be log-concave if,

$$f(\alpha x + (1-\alpha)y) \geq f(x)^\alpha f(y)^{1-\alpha},$$

for each $x, y \in \mathbb{R}^n$ and each $0 \leq \alpha \leq 1$. If f is log-concave, we will also use the adjective "log-concave" for a random variable X distributed according to f, and for the probability measure induced by it. Note that the class of log-concave probability measures is quite broad, including the uniform distribution on any compact, convex set, the exponential distribution, and of course any Gaussian. On the other hand, log-concavity can also be fairly restricting: For instance, it implies at least exponentially decaying tails, and a finite Poincaré constant.

Now we state the main result of [7] we will need. For a random vector $X \sim f$ we write $D(X)$ for its relative entropy distance from a Gaussian,

$$D(X) = D(f \| f^G) = h(f^G) - h(f),$$

where f^G is the Gaussian density with the same mean and covariance matrix as f, and D is the usual relative entropy.

Theorem 9. *[7] Suppose X is a log-concave random vector in $\mathbb{R}^n$, and that it satisfies a Poincaré inequality with constant c, i.e., if for any smooth function u with $E[u(X)] = 0$,*

$$cE[u(X)^2] \leq E[|\nabla u(X)|^2].$$

Then,

$$h\left(\frac{X_1 + X_2}{\sqrt{2}}\right) - h(X) \geq \frac{c}{4(1+c)} D(X),$$

where X_1 and X_2 denote independent copies of X.

Simply rearranging the conclusion of Theorem 9 gives the following stability result.

Corollary 7. *If X is a log-concave random vector in $\mathbb{R}^n$, with Poincaré constant c, then:*

$$\frac{D(X)}{n} \leq \frac{2(1+c)}{c} \log \sigma_+(X).$$

Remark 3. *It was proved in [10, Proposition V.6], by relying on an important result of Klartag [29], that log-concave distributions are not too far from Gaussianity, in the sense that,*

$$\frac{D(X)}{n} \leq \frac{1}{4} \log n + C,$$

for some absolute constant C. Therefore, the main value of the result in Corollary 7 is in that it explicitly connects "non-Gaussianity" with the doubling constant $\sigma_+(X)$, and especially when $\sigma_+(X)$ is small.

Interestingly, it is not hard to give a much more elementary converse result when we know something about both the doubling and the difference constants. Indeed, we show below that any random vector whose doubling and difference constants differ significantly, must also be significantly far from Gaussianity.

Theorem 10. *If X_1 and X_2 are independent copies of any random vector X in $\mathbb{R}^n$ with finite differential entropy, then,*

$$\frac{D(X)}{n} \geq \frac{1}{4} |\log \sigma_+(X) - \log \sigma_-(X)|.$$

[2] While asymptotic estimates of this sort are known [13], [15], [19], estimates that only hold for a sufficiently large number of summands are not strong enough for our purposes.

Proof: By the invariance of the entropy under linear transformations of determinant 1,

$$\begin{aligned} h(X_1)+h(X_2) &= h(X_1,X_2) \\ &= h\left(\frac{X_1+X_2}{\sqrt{2}},\frac{X_1-X_2}{\sqrt{2}}\right) \\ &\leq h\left(\frac{X_1+X_2}{\sqrt{2}}\right)+h\left(\frac{X_1-X_2}{\sqrt{2}}\right). \end{aligned}$$

Let a be the greater of the quantities $h\big(\frac{X_1+X_2}{\sqrt{2}}\big)$ and $h\big(\frac{X_1-X_2}{\sqrt{2}}\big)$, and b be the lesser of them. The above display implies that,

$$\frac{a+b}{2} \geq h(X). \tag{13}$$

Now, by the scaling property of differential entropy, we have,

$$\begin{aligned} \tfrac{1}{2}|h(X_1-X_2)-h(X_1+X_2)| &= \tfrac{1}{2}\left|h\left(\frac{X_1+X_2}{\sqrt{2}}\right)-h\left(\frac{X_1-X_2}{\sqrt{2}}\right)\right| \\ &= \frac{a-b}{2} = a-\frac{a+b}{2} \\ &\leq a-h(X), \end{aligned}$$

using (13) to obtain the inequality. Since both $\frac{X_1+X_2}{\sqrt{2}}$ and $\frac{X_1-X_2}{\sqrt{2}}$ have the same covariance matrix as X, the maximum entropy property of the Gaussian implies that $h(Z) \geq a$, where Z is a Gaussian random vector with the same covariance matrix as X. Thus we have,

$$\tfrac{1}{2}|h(X_1-X_2)-h(X_1+X_2)| \leq h(Z)-h(X) = D(X),$$

which is equivalent to the desired statement by using the relations (11) and (12). ■

VII. An explicit reverse entropy power inequality

In recent work [9], a reverse entropy power inequality was developed for the class of log-concave distributions. Recall that the entropy power inequality due to Shannon and Stam [50], [51] asserts that $\mathcal{N}(X+Y) \geq \mathcal{N}(X) + \mathcal{N}(Y)$, for any two independent random vectors X and Y in $\mathbb{R}^n$ for which the entropy is defined. The entropy power inequality may be formally strengthened by using the invariance of entropy under affine transformations of determinant ±1, i.e., $\mathcal{N}(u(X)) = \mathcal{N}(X)$ whenever $|\det(u)| = 1$. Specifically,

$$\inf_{u_1,u_2} \mathcal{N}(u_1(X) + u_2(Y)) \geq \mathcal{N}(X) + \mathcal{N}(Y), \tag{14}$$

where the maps $u_i : \mathbb{R}^n \to \mathbb{R}^n$ range over all affine entropy-preserving transformations. What [9] showed was that the inequality (14) can be reversed with a constant independent of dimension if we restrict to log-concave distributions.

Theorem 11 (Reverse EPI, [9])**.** *If X and Y are independent random vectors in $\mathbb{R}^n$ with log-concave densities, there exist linear entropy-preserving maps $u_i : \mathbb{R}^n \to \mathbb{R}^n$ such that*

$$\mathcal{N}\big(\widetilde{X} + \widetilde{Y}\big) \;\leq\; C\,(\mathcal{N}(X) + \mathcal{N}(Y)), \tag{15}$$

where $\widetilde{X} = u_1(X)$, $\widetilde{Y} = u_2(Y)$, and where C is a universal constant.

This reverse entropy power inequality is analogous to Milman's [37] reverse Brunn-Minkowski inequality (see also [38], [39], [42]), which is a celebrated result in convex geometry. In this light, Theorem 11 can be seen as an extension of the analogies between geometry and information theory that were previously observed by Dembo, Cover and Thomas [23], among others. Also, Theorem 11 can be extended to the larger subclass of so-called "convex measures" [11].

Observe that the universal constant provided by the proof of Theorem 11 is not explicit, and it is not easy to even get bounds on it. But in the special case when X and Y have the same distribution, we show below that an explicit constant can be obtained rather simply. To do this, we first note that the following result of Cover and Zhang [21] easily generalizes to higher dimensions: If X and X' are (possibly dependent) random variables with the same log-concave marginal distribution on $\mathbb{R}$, then, $h(X + X') \leq h(2X)$.

Theorem 12. *If X and Y are (possibly dependent) random vectors in $\mathbb{R}^n$, with the same log-concave marginal density, then,*

$$h(X+Y) \leq h(2X).$$

Proof: Suppose the common marginal density of X and Y is f, and let g be the density of $Z = X+Y$. Since f is log-concave, Jensen's inequality implies that,

$$\begin{aligned} E\log f\left(\frac{X+Y}{2}\right) &\geq E\tfrac{1}{2}[\log f(X) + \log f(Y)] \\ &= \tfrac{1}{2}[E\log f(X) + E\log f(Y)] \\ &= -h(f). \end{aligned}$$

Observe that independence is not required here, and all expectations are taken with respect to the joint distribution of (X,Y). In particular, we have that,

$$\begin{aligned} \int g(z)\log\tilde{f}(z)dz &= \int g(z)\log f(\tfrac{z}{2}) - 1 \\ &\geq -h(f) - 1 \\ &= -h(\tilde{f}), \end{aligned}$$

where $\tilde{f}(z) = \frac{1}{2}f(z/2)$ is the density of $Z^* = 2X$. In other words,

$$D(g\|\tilde{f}) + h(g) \leq h(\tilde{f}).$$

Thus $h(g) = h(X+Y)$ is maximized if and only if $g = \tilde{f}$, i.e., when X and Y are identical. ■

Theorem 12 immediately implies that for i.i.d. random vectors with log-concave distribution, the reverse entropy power inequality (Theorem 11) holds with both linear transformations being the identity, and with a universal constant of 2.

Corollary 8. *If X, X' are independent random vectors with the same log-concave distribution, then,*

$$\mathcal{N}(X+X') \leq 2[\mathcal{N}(X)+\mathcal{N}(X')].$$

In other words, for any log-concave random vector X, $\sigma_+(X) \leq 2$.

Proof: From Theorem 12,

$$\mathcal{N}(X+X') \leq \mathcal{N}(2X) = 4\mathcal{N}(X) = 2[\mathcal{N}(X)+\mathcal{N}(X)].$$

■

A version of Corollary 8 was obtained (contemporaneously with this work) by a different method in [16]; however the bound on the doubling constant in that work is $e^4/2 \approx 27.3$, which is significantly worse than the bound of 2 we obtain. Soon after the first version of this paper was released, some related results and a nice conjecture about reverse forms of the entropy power inequality were released by Ball, Nayar and Tkocz [6].

Although it already seems rather restrictive that the doubling constant of any log-concave random vector lies between 1 and 2, we do not believe the upper bound is optimal. However, Corollary 8 represents yet another way in which general log-concave random vectors resemble Gaussian ones; as mentioned in Remark 3, [10] gives a different formulation of this intuition.

Another way to view Corollary 8 is in the context of the central limit theorem. Recall that the central limit theorem in terms of relative entropy ([8], [4], see also [33]) asserts that if $X, X_1, X_2, \ldots$ are i.i.d. random vectors with $h(X) > -\infty$, then, as $n \to \infty$,

$$h\left(\frac{X_1+\ldots+X_n}{\sqrt{n}}\right) \uparrow h(N(0, I_n)).$$

Corollary 8 implies that,

$$\mathcal{N}(X) \leq \mathcal{N}\left(\frac{X_1+X_2}{\sqrt{2}}\right) \leq 2\mathcal{N}(X),$$

and hence constrains the rate at which entropy can increase when doubling sample size in the central limit theorem for i.i.d. log-concave summands.

The above development is also closely related to a very nice observation of K. Ball, dating back to around 2003 but with details only being published much later in [7], relating two important conjectures in convex geometry, namely the Kannan-Lovász-Simonovits conjecture [28] and the hyperplane conjecture or slicing problem of Bourgain [17]. We explain this connection in our language; the reasoning is related to that of K. Ball even if it differs in details. The Kannan-Lovász-Simonovits (KLS) conjecture asserts that the Poincaré constant c is bounded from below for all log-concave densities by a universal constant C independent of dimension. If this is true, then Corollary 7 implies that

$$\frac{D(X)}{n} \leq 2\left(1+\frac{1}{C}\right)\sigma_+(X) \leq 2\left(1+\frac{1}{C}\right)\log 2,$$

using Corollary 8 for the second inequality. In other words, $D(X)/n$ is bounded by a universal constant for any log-concave random vector X in $\mathbb{R}^n$, which by [10, Corollary 5.3], is equivalent to the hyperplane conjecture (whose original formulation in [17] we do not bother to state here). Hence the KLS conjecture implies the hyperplane conjecture.

VIII. Towards a Rogers-Shephard inequality for entropy

The Rogers-Shephard inequality [44] asserts that, if $K \subset \mathbb{R}^n$ is a convex body, then

$$\mathrm{Vol}(K - K) \le \binom{2n}{n} \mathrm{Vol}(K), \tag{16}$$

with equality if and only if K is the n-dimensional simplex. It complements the fact, implied by the Brunn-Minkowski inequality, that,

$$\mathrm{Vol}(K - K) \ge 2^n \mathrm{Vol}(K). \tag{17}$$

Indeed, since by Stirling's formula and some algebraic manipulation,

$$\binom{2n}{n} < 4^n,$$

the inequalities (16) and (17) together imply,

$$2\mathrm{Vol}(K)^{1/n} \le \mathrm{Vol}(K-K)^{1/n} < 4\mathrm{Vol}(K)^{1/n}.$$

As suggested by the analogy between the reverse entropy power inequality and the reverse Brunn-Minkowski inequality discussed in the preceding section, the natural probabilistic analogue of a convex set is a log-concave distribution, and a natural probabilistic analogue of volume is entropy. Therefore, it is natural to ask whether there is a probabilistic analogue of the Rogers-Shephard inequality. Indeed, we show that for X, X' i.i.d. log-concave random vectors, $\mathcal{N}(X - X')$ is bounded by a multiple of $\mathcal{N}(X)$.

Corollary 9. *If X, X' are independent random vectors drawn from the same log-concave distribution, then*

$$\mathcal{N}(X - X') \le 16\mathcal{N}(X).$$

In other words, for any log-concave random vector X, $\sigma_-(X) \le 8$.

Proof: By Corollary 8,

$$\mathcal{N}(X + X') \le 4\mathcal{N}(X),$$

and by Corollary 3,

$$\mathcal{N}(X - X') \le \frac{\mathcal{N}^2(X + X')}{\mathcal{N}(X)} \le 16\mathcal{N}(X).$$

■

Corollary 9 does not provide a tight bound. Indeed, in the contemporaneous work [16], a different approach is used to obtain a bound on the difference constant of $e^2/2 \approx 3.7$, which is better than our bound of 8. We state below a conjecture for the sharp constant in the one-dimensional case.

Conjecture 1. *If X, X' are independent $\mathbb{R}$-valued random variables drawn from the same log-concave distribution, then,*

$$\mathcal{N}(X - X') \le 4\mathcal{N}(X),$$

with equality if and only if X is a translated and scaled version of the (one-sided) exponential distribution. In other words, for any log-concave random variable X, $\sigma_-(X) \le 2$.

Of course, we may also write Corollary 9 and Corollary 8 in terms of the Ruzsa divergence using the identities (11) and (12).

Corollary 10. *If X is a log-concave random vector taking values in $\mathbb{R}^n$, then,*

$$d_R(X\|X) \le 2n\log 2 \quad \textit{and} \quad d_R(X\| - X) \le n\log 2.$$

Let us note that a sharp functional analogue of the Rogers-Shephard inequality has been proved by Colesanti [20] for log-concave functions as opposed to densities (see also [5], [2]).

IX. Determinant inequalities

Differential entropy inequalities have been used to to deduce inequalities for positive-definite matrices since Cover and El Gamal's work in [22]; see also [23] and [36]. However, in most of these cases, the inequalities deduced relate determinants of a positive-definite matrix to those of its square submatrices. We discuss below the use of differential entropy inequalities to prove determinantal inequalities for sums of positive-definite matrices. As in the above papers, the main idea is to use the fact that, for the Gaussian distribution on $\mathbb{R}^n$ with covariance matrix K, written $\gamma_K = N(0, K)$, the differential entropy is given by,

$$h(\gamma_K) = \tfrac{1}{2}\log\left[(2\pi e)^n \det(K)\right].$$

A classical inequality for the determinant of sums is Minkowski's inequality, which asserts that, for $n \times n$ positive-definite matrices,

$$\det(A+B)^{\frac{1}{n}} \geq \det(A)^{\frac{1}{n}} + \det(B)^{\frac{1}{n}}.$$

This may be seen as a consequence of the entropy power inequality (by specializing to Gaussians), but there are also elementary means of deriving it.

On the other hand, upper bounds for the determinant of a sum of positive-definite matrices are not as well known. This is partly due to the fact that the most straightforward inequalities that one might try to check, like subadditivity, are actually false. However, Rotfel'd [45] did obtain such a bound when one of the matrices involved is the identity matrix:

$$\det(I+A+B) \leq \det(I+A)\cdot\det(I+B). \tag{18}$$

Indeed, he obtained this as a special case of a more general inequality for arbitrary square matrices,

$$\det(I+|A+B|)| \leq \det(I+|A|)\cdot\det(I+|B|),$$

where $|A| = \sqrt{A^*A}$ and A^* is the adjoint of A.

Our final observation is that, substituting normals in Theorem 6, provides an extremely simple alternative proof of a generalization of inequality (18), not requiring any of the matrices to be the identity:

Corollary 11. *Let K and K_i be positive-definite matrices of the same dimension. Then:*

$$\det(K+K_1+\ldots+K_n) \leq [\det(K)]^{-(n-1)}\prod_{j=1}^{n}\det(K+K_j).$$

References

[1] E. Abbe, J. Li, and M. Madiman. Entropies of weighted sums in cyclic groups and applications to polar codes. *Preprint*, 2015.

[2] D. Alonso-Gutiérrez, B. González, C. Hugo Jiménez, and R. Villa. Rogers-Shephard inequality for log-concave functions. *Preprint*, arXiv:1410.2556, 2014.

[3] S. Artstein, K. M. Ball, F. Barthe, and A. Naor. On the rate of convergence in the entropic central limit theorem. *Probab. Theory Related Fields*, 129(3):381–390, 2004.

[4] S. Artstein, K. M. Ball, F. Barthe, and A. Naor. Solution of Shannon's problem on the monotonicity of entropy. *J. Amer. Math. Soc.*, 17(4):975–982 (electronic), 2004.

[5] S. Artstein-Avidan, K. Einhorn, D. Y. Florentin, and Y. Ostrover. On Godbersen's conjecture. *Preprint*, arXiv:1408.2135, 2014.

[6] K. Ball, P. Nayar, and T. Tkocz. A reverse entropy power inequality for log-concave random vectors. *Preprint*, arXiv:1509.05926, 2015.

[7] K. Ball and V. H. Nguyen. Entropy jumps for isotropic log-concave random vectors and spectral gap. *Studia Math.*, 213(1):81–96, 2012.

[8] A.R. Barron. Entropy and the central limit theorem. *Ann. Probab.*, 14:336–342, 1986.

[9] S. Bobkov and M. Madiman. Dimensional behaviour of entropy and information. *C. R. Acad. Sci. Paris Sér. I Math.*, 349:201–204, Février 2011.

[10] S. Bobkov and M. Madiman. The entropy per coordinate of a random vector is highly constrained under convexity conditions. *IEEE Trans. Inform. Theory*, 57(8):4940–4954, August 2011.

[11] S. Bobkov and M. Madiman. Reverse Brunn-Minkowski and reverse entropy power inequalities for convex measures. *J. Funct. Anal.*, 262:3309–3339, 2012.

[12] S. G. Bobkov and G. P. Chistyakov. Entropy power inequality for the Rényi entropy. *IEEE Trans. Inform. Theory*, 61(2):708–714, February 2015.

[13] S. G. Bobkov, G. P. Chistyakov, and F. Götze. Rate of convergence and Edgeworth-type expansion in the entropic central limit theorem. *Ann. Probab.*, 41(4):2479–2512, 2013.

[14] S. G. Bobkov, G. P. Chistyakov, and F. Götze. Stability problems in Cramer-type characterization in case of i.i.d. summands. *Theory Probab. Appl.*, 57(4):568–588, 2013.

[15] S. G. Bobkov, G. P. Chistyakov, and F. Götze. Berry-Esseen bounds in the entropic central limit theorem. *Probab. Theory Related Fields*, 159(3-4):435–478, 2014.

[16] S. G. Bobkov and M. M. Madiman. On the problem of reversibility of the entropy power inequality. In P. Eichelsbacher et al., editor, *Limit Theorems in Probability, Statistics, and Number Theory (in honor of Friedrich Götze)*, volume 42 of *Springer Proceedings in Mathematics and Statistics*. Springer-Verlag, 2013. Available online at http://arxiv.org/abs/1111.6807.

[17] J. Bourgain. On high-dimensional maximal functions associated to convex bodies. *Amer. J. Math.*, 108(6):1467–1476, 1986.

[18] B. Bukh. Sums of dilates. *Combin. Probab. Comput.*, 17(5):627–639, 2008.

[19] E. A. Carlen and A. Soffer. Propogation of localization, optimal entropy production, and convergence rates for the central limit theorem. *Preprint*, arXiv:1106.2256, 2011.

[20] A. Colesanti. Functional inequalities related to the Rogers-Shephard inequality. *Mathematika*, 53(1):81–101 (2007), 2006.

[21] T. M. Cover and Z. Zhang. On the maximum entropy of the sum of two dependent random variables. *IEEE Trans. Inform. Theory*, 40(4):1244–1246, 1994.

[22] T.M. Cover and A. El Gamal. An information theoretic proof of Hadamard's inequality. *IEEE Trans. Inform. Theory*, 29(6):930–931, 1983.

[23] A. Dembo, T.M. Cover, and J.A. Thomas. Information-theoretic inequalities. *IEEE Trans. Inform. Theory*, 37(6):1501–1518, 1991.

[24] J. Diestel and A. Spalsbury. *The joys of Haar measure*, volume 150 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2014.

[25] P. R. Halmos. *Measure Theory*. D. Van Nostrand Company, Inc., New York, N. Y., 1950.

[26] O. Johnson and A.R. Barron. Fisher information inequalities and the central limit theorem. *Probab. Theory Related Fields*, 129(3):391–409, 2004.

[27] V. A. Kaĭmanovich and A. M. Vershik. Random walks on discrete groups: boundary and entropy. *Ann. Probab.*, 11(3):457–490, 1983.

[28] R. Kannan, L. Lovász, and M. Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete Comput. Geom.*, 13(3-4):541–559, 1995.

[29] B. Klartag. On convex perturbations with a bounded isotropic constant. *Geom. Funct. Anal.*, 16(6):1274–1290, 2006.

[30] I. Kontoyiannis and M. Madiman. Sumset and inverse sumset inequalities for differential entropy and mutual information. *IEEE Trans. Inform. Theory*, 60(8):4503–4514, August 2014.

[31] 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.

[32] M. Madiman. On the entropy of sums. In *Proc. IEEE Inform. Theory Workshop*, pages 303–307. Porto, Portugal, 2008.

[33] M. Madiman and A.R. Barron. Generalized entropy power inequalities and monotonicity properties of information. *IEEE Trans. Inform. Theory*, 53(7):2317–2329, July 2007.

[34] M. Madiman, A. Marcus, and P. Tetali. Entropy and set cardinality inequalities for partition-determined functions. *Random Struct. Alg.*, 40:399–424, 2012.

[35] M. Madiman and P. Singla. A note on $GL_n(\mathbb{Z})$-actions on locally compact abelian groups. *Preprint*, 2015.

[36] M. Madiman and P. Tetali. Information inequalities for joint distributions, with interpretations and applications. *IEEE Trans. Inform. Theory*, 56(6):2699–2713, June 2010.

[37] V. D. Milman. Inégalité de Brunn-Minkowski inverse et applications à la théorie locale des espaces normés. *C. R. Acad. Sci. Paris Sér. I Math.*, 302(1):25–28, 1986.

[38] V. D. Milman. Entropy point of view on some geometric inequalities. *C. R. Acad. Sci. Paris Sér. I Math.*, 306(14):611–615, 1988.

[39] V. D. Milman. Isomorphic symmetrizations and geometric inequalities. In *Geometric aspects of functional analysis (1986/87)*, volume 1317 of *Lecture Notes in Math.*, pages 107–131. Springer, Berlin, 1988.

[40] L. Nachbin. *The Haar integral*. Robert E. Krieger Publishing Co., Huntington, N.Y., 1976. Translated from the Portuguese by L. Bechtolsheim, Reprint of the 1965 edition.

[41] G. Petridis. New proofs of Plünnecke-type estimates for product sets in groups. *Combinatorica*, 32(6):721–733, 2012.

[42] G. Pisier. *The volume of convex bodies and Banach space geometry*, volume 94 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1989.

[43] H. Plünnecke. Eine zahlentheoretische Anwendung der Graphentheorie. *J. Reine Angew. Math.*, 243:171–183, 1970.

[44] C. A. Rogers and G. C. Shephard. The difference body of a convex body. *Arch. Math. (Basel)*, 8:220–233, 1957.

[45] S. Ju. Rotfel'd. Remarks on the singular values of a sum of completely continuous operators. *Funkcional. Anal. i Priložen*, 1(3):95–96, 1967.

[46] I. Z. Ruzsa. An application of graph theory to additive number theory. *Scientia Ser. A Math. Sci. (N.S.)*, 3:97–109, 1989.

[47] I. Z. Ruzsa. Addendum to: An application of graph theory to additive number theory. *Scientia Ser. A Math. Sci. (N.S.)*, 4:93–94, 1990/91.

[48] I. Z. Ruzsa. Sums of finite sets. In *Number theory (New York, 1991–1995)*, pages 281–293. Springer, New York, 1996.

[49] I. Z. Ruzsa. Entropy and sumsets. *Random Struct. Alg.*, 34:1–10, 2009.

[50] C.E. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423, 623–656, 1948.

[51] A.J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2:101–112, 1959.

[52] D. Stotz and H. Bölcskei. Degrees of freedom in vector interference channels. *Preprint*, `arXiv:1210:2259`, 2014.

[53] D. Stotz and H. Bölcskei. Characterizing degrees of freedom through additive combinatorics. *Preprint*, `arXiv:1506:01866`, 2015.

[54] T. Tao. Sumset and inverse sumset theory for Shannon entropy. *Combin. Probab. Comput.*, 19(4):603–639, 2010.

[55] T. Tao and V. Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.

[56] T. Tao and V. Vu. Entropy methods. Unpublished note available at: `www.math.ucla.edu/˜tao/`, 2006.

[57] Y. Wu, S. Shamai, and S. Verdú. Information dimension and the degrees of freedom of the interference channel. *IEEE Trans. Inform. Theory*, 61(1):256–279, 2015.