On the Entropy of Sums

Mokshay Madiman
Department of Statistics
Yale University
Email: mokshay.madiman@yale.edu

***Abstract*— It is shown that the entropy of a sum of independent random vectors is a submodular set function, and upper bounds on the entropy of sums are obtained as a result in both discrete and continuous settings. These inequalities complement the lower bounds provided by the entropy power inequalities of Madiman and Barron (2007). As applications, new inequalities for the determinants of sums of positive-definite matrices are presented.**

I. Introduction

Entropies of sums are not as well understood as joint entropies. Indeed, for the joint entropy, there is an elaborate history of entropy inequalities starting with the chain rule of Shannon, whose major developments include works of Han, Shearer, Fujishige, Yeung, Matúš, and others. The earlier part of this work, involving so-called Shannon-type inequalities that use the submodularity of the joint entropy, was synthesized and generalized by Madiman and Tetali [1], who give an array of lower as well as upper bounds for joint entropy of a collection of random variables generalizing inequalities of Han [2], Fujishige [3] and Shearer [4]. For a review of the later history, involving so-called non-Shannon inequalities, one may consult, for instance, Matúš [5].

When one considers the entropy of sums of independent random variables instead of joint entropies, the most important inequalities known are the entropy power inequalities, which provide lower bounds on entropy of sums. In the setting of independent summands, the most general entropy power inequalities known to date are described by Madiman and Barron [6].

In this note, we develop a basic submodularity property of the entropy of sums of independent random variables, and formulate a "chain rule for sums". Surprisingly, these results are entirely elementary and rely on the classical properties of joint entropy. We also demonstrate as a consequence upper bounds on entropy of sums that complement entropy power inequalities.

The paper is organized as follows. In Section II, we review the notation and definitions we use. Section III presents the submodularity of the entropy of sums of independent random vectors, which is a central result. Section IV develops general upper bounds on entropies of sums, while Section V reviews known lower bounds and conjectures some additional lower bounds. In Section VI, we apply the entropy inequalities of the previous sections to give information-theoretic proofs of inequalities for the determinants of sums of positive-definite matrices.

II. Preliminaries

Let $X_1, X_2, \ldots, X_n$ be a collection of random variables taking values in some linear space, so that addition is a well-defined operation. We assume that the joint distribution has a density f with respect to some reference measure. This allows the definition of the entropy of various random variables depending on $X_1, \ldots, X_n$. For instance, $H(X_1, X_2, \ldots, X_n) = -E[\log f(X_1, X_2, \ldots, X_n)]$. There are the familiar two canonical cases: (a) the random variables are real-valued and possess a probability density function, or (b) they are discrete. In the former case, H represents the differential entropy, and in the latter case, H represents the discrete entropy. We simply call H the entropy in all cases, and where the distinction matters, the relevant assumption will be made explicit.

Since we wish to consider sums of various subsets of random variables, the following notational conventions and definitions will be useful. Let $[n]$ be the index set $\{1, 2, \ldots, n\}$. We are interested in a collection $\mathcal{C}$ of subsets of $[n]$. For any set $s \subset [n]$, X_s stands for the random variable $(X_i : i \in s)$, with the indices taken in their increasing order, while

$$Y_s = \sum_{i \in s} X_i.$$

For any index i in $[n]$, define the *degree* of i in $\mathcal{C}$ as $r(i) = |\{t \in \mathcal{C} : i \in t\}|$. A function $\alpha : \mathcal{C} \to \mathbb{R}_+$, is called a *fractional covering*, if for each $i \in [n]$, we have $\sum_{s \in \mathcal{C}: i \in s} \alpha_s \geq 1$. A function $\beta : \mathcal{C} \to \mathbb{R}_+$, is called a *fractional packing*, if for each $i \in [n]$, we have $\sum_{s \in \mathcal{C}: i \in s} \beta_s \leq 1$. If α is both a fractional packing and a fractional covering, it is called a *fractional partition*.

The mutual information between two jointly distributed random variables X and Y is

$$I(X; Y) = H(X) + H(Y) - H(X, Y),$$

and is a measure of the dependence between X and Y. In particular, $I(X; Y) = 0$ if and only if X and Y are independent. The following three simple properties of mutual information can be found in elementary texts on information theory such as Cover and Thomas [7]:

- When $Y = X + Z$, where X and Z are independent, then

$$I(X; Y) = H(Y) - H(Z). \tag{1}$$

Indeed, $I(X; X + Z) = H(X + Z) - H(X + Z|X) = H(X + Z) - H(Z|X) = H(X + Z) - H(Z)$.

- The mutual information cannot increase when one looks at functions of the random variables (the "data processing inequality"):

$$I(f(X);Y) \le I(X;Y).$$

- When Z is independent of (X,Y), then

$$I(X;Y) = I(X,Z;Y). \tag{2}$$

Indeed,

$$\begin{aligned}
&I(X,Z;Y) - I(X;Y)\\
&= H(X,Z) - H(X,Z|Y) - [H(X) - H(X|Y)]\\
&= H(X,Z) - H(X) - [H(X,Z|Y) - H(X|Y)]\\
&= H(Z|X) - H(Z|X,Y)\\
&= 0.
\end{aligned}$$

III. Submodularity

Although the inequality below is a simple consequence of these well known facts, it does not seem to have been noticed before.

Theorem I:[Submodularity for Sums] If X_i are independent $\mathbb{R}^d$-valued random vectors, then

$$H(X_1+X_2) + H(X_2+X_3) \ge H(X_1+X_2+X_3) + H(X_2).$$

Proof: First note that

$$\begin{aligned}
&H(X_1+X_2) + H(X_2+X_3) - H(X_1+X_2+X_3) - H(X_2)\\
&= H(X_1+X_2) - H(X_2) -\\
&\quad \big[H(X_1+X_2+X_3) - H(X_2+X_3)\big]\\
&= I(X_1+X_2;X_1) - I(X_1+X_2+X_3;X_1),
\end{aligned}$$

using (1). Thus we simply need to show that

$$I(X_1+X_2;X_1) \ge I(X_1+X_2+X_3;X_1).$$

Now

$$\begin{aligned}
I(X_1+X_2+X_3;X_1) &\overset{(a)}{\le} I(X_1+X_2,X_3;X_1)\\
&\overset{(b)}{=} I(X_1+X_2;X_1),
\end{aligned}$$

where (a) follows from the data processing inequality, and (b) follows from (2); so the proof is complete. ■

Some remarks about this result are appropriate. First, this result implies that the "rate region" of vectors $(R_1,\ldots,R_n)$ satisfying

$$\sum_{i\in s} R_i \le H(Y_s)$$

for each $s \subset [n]$ is polymatroidal (in particular, it has a non-empty dominant face on which $\sum_{i\in[n]} R_i = H(Y_{[n]})$). It is not clear how this fact is to be interpreted since this is not the rate region of any real information theoretic problem to the author's knowledge. Second, a consequence of the submodularity of entropy of sums is an entropy inequality that obeys the partial order constructed using compressions, as introduced by Bollobas and Leader [8]. Since this inequality requires the development of some involved notation, we defer the details to the full version [9] of this paper. Third, an equivalent statement of Theorem I is that for independent random vectors,

$$I(X;X+Y) \ge I(X;X+Y+Z), \tag{3}$$

as is clear from the proof of Theorem I.

IV. Upper Bounds

Remarkably, if we wish to use Theorem I to obtain upper bounds for the entropy of a sum, there is a dramatic difference between the discrete and continuous cases, even though this distinction does not appear in Theorem I.

First consider the case where all the random variables are *discrete*. Shannon's chain rule for joint entropy says that

$$H(X_1,X_2) = H(X_1) + H(X_2|X_1) \tag{4}$$

where $H(X_2|X_1)$ is the conditional entropy of X_2 given X_1. This rule extends to the consideration of n variables; indeed,

$$H(X_1,\ldots,X_n) = \sum_{i=1}^{n} H(X_i|X_{<i}) \tag{5}$$

where $X_{<i}$ is used to denote $(X_j : j < i)$. Similarly, one may also talk about a "chain rule for sums" of independent random variables, and this is nothing but the well known identity (1) that is ubiquitous in the analysis of communication channels. Indeed, one may write

$$H(X_1+X_2) = H(X_1) + I(X_1+X_2;X_2)$$

and iterate this to get

$$\begin{aligned}
H\Bigg(\sum_{i\in[n]} X_i\Bigg) &= H(Y_{[n-1]}) + I(Y_{[n]};Y_n)\\
&= H(Y_{[n-2]}) + I(Y_{[n-1]};Y_{n-1}) + I(Y_{[n]};Y_n)\\
&= \ldots\\
&= \sum_{i\in[n]} I(Y_{[i]};Y_i),
\end{aligned}$$

which has the form of a "chain rule". Here we have used Y_s to denote the sum of the components of X_s, and in particular $Y_i = Y_{\{i\}} = X_i$. Below we also use $\le i$ for the set of indices less than or equal to i, $Y_{\le i}$ for the corresponding subset sum of X_i's etc.

Theorem II:[Upper Bound for Discrete Entropy] If X_i are independent discrete random variables, then

$$H(X_1+\ldots+X_n) \le \sum_{s\in\mathcal{C}} \alpha_s H\Bigg(\sum_{i\in s} X_i\Bigg),$$

for any fractional covering α using any collection $\mathcal{C}$ of subsets of $[n]$.

Theorem II can easily be derived by combining the chain rule for sums described above with Theorem I. Alternatively, it follows from Theorem I because of the general fact that submodularity implies "fractional subadditivity" (see, e.g., [1]).

For any collection $\mathcal{C}$ of subsets, [1] introduced the degree covering, given by

$$\alpha_s = \frac{1}{r_-(s)},$$

where $r_-(s) = \min_{i\in s} r(i)$. Specializing Theorem II to this particular fractional covering, we obtain

$$H(X_1 + \ldots + X_n) \leq \sum_{s\in\mathcal{C}} \frac{1}{r_-(s)} H\bigg(\sum_{i\in s} X_i\bigg).$$

A simple example is the case of the collection $\mathcal{C}_m$, consisting of all subsets of $[n]$ with m elements. For this collection, we obtain

$$H(X_1 + \ldots + X_n) \leq \frac{1}{\binom{n-1}{m-1}} \sum_{s\in\mathcal{C}} H\bigg(\sum_{i\in s} X_i\bigg).$$

since the degree of each index with respect to $\mathcal{C}_m$ is $\binom{n-1}{m-1}$.

Now consider the case where all the random variables are *continuous*. At first sight, everything discussed above for the discrete case should also go through in this case, since Theorem I holds in general. However this is not true; indeed, the differential entropy of a sum does not even satisfy simple subadditivity! To see why we cannot obtain subadditivity from Theorem I, note that setting $X_2 = 0$ makes Theorem I trivial because the differential entropy of a constant is $-\infty$ (unlike the discrete entropy of a constant which is 0). Similarly, the chain rule for sums described above is only partly true: while

$$H\bigg(\sum_{i\in[n]} X_i\bigg) = H(X_1) + \sum_{i=2}^{n} I(Y_{[i]}; Y_i), \tag{6}$$

we cannot equate $H(X_1) = H(Y_1)$ to $I(Y_1; Y_1)$, since the latter is ∞.

Still one finds that (6) is nonetheless useful; it yields the following upper bounds for differential entropy of a sum.

Theorem III:[Upper Bound for Differential Entropy] Let X and $X_i, i \in [n]$ be independent $\mathbb{R}^d$-valued random vectors with densities. Suppose α is a fractional covering for the collection $\mathcal{C}$ of subsets of $[n]$. Then

$$H\bigg(X + \sum_{i\in[n]} X_i\bigg) \leq \sum_{s\in\mathcal{C}} \alpha_s H\bigg(X + \sum_{i\in s} X_i\bigg) - \bigg(\sum_{s\in\mathcal{C}} \alpha_s - 1\bigg) H(X).$$

Proof: The modified chain rule (6) for sums implies that

$$H(X + Y_s) = H(X) + \sum_{i\in s} I(Y_i; Y_{s\cap\leq i}).$$

Thus

$$\begin{aligned}
\sum_{s\in\mathcal{C}} \alpha_s H(X + Y_s) &= \sum_{s\in\mathcal{C}} \alpha_s [H(X) + \sum_{i\in s} I(Y_i; Y_{s\cap\leq i})] \\
&\overset{(a)}{\geq} \sum_{s\in\mathcal{C}} \alpha_s H(X) + \sum_{s\in\mathcal{C}} \alpha_s \sum_{i\in s} I(Y_i; Y_{\leq i}) \\
&\overset{(b)}{=} \sum_{s\in\mathcal{C}} \alpha_s H(X) + \sum_{i\in[n]} I(Y_i; Y_{\leq i}) \sum_{s\ni i} \alpha_s \\
&\overset{(c)}{\geq} \sum_{s\in\mathcal{C}} \alpha_s H(X) + \sum_{i\in[n]} I(Y_i; Y_{\leq i}) \\
&= \bigg(\sum_{s\in\mathcal{C}} \alpha_s - 1\bigg) H(X) + H(Y_{[n]}),
\end{aligned}$$

where (a) follows from (3), (b) follows from an interchange of sums, and (c) follows from the definition of a fractional covering, and the last identity in the display again uses the chain rule (6) for sums. ■

Note in particular that if $H(X) \geq 0$, then Theorem III implies

$$H\bigg(X + \sum_{i\in[n]} X_i\bigg) \quad \leq \textstyle\sum_{s\in\mathcal{C}} \alpha_s H\bigg(X + \sum_{i\in s} X_i\bigg),$$

which is identical to Theorem II except that the random variable X is an additional summand in every sum.

V. Lower bounds

In this section, we only consider continuous random vectors. An entropy power inequality relates the entropy power of a sum to that of summands. We make the following conjecture that generalizes the known entropy power inequalities.

Conjecture I:[Fractional Superadditivity of Entropy Power] Let $X_1, \ldots, X_n$ be independent $\mathbb{R}^d$-valued random vectors with densities and finite covariance matrices. For any collection $\mathcal{C}$ of subsets of $[n]$, let β be a fractional packing. Then

$$e^{\frac{2H(X_1+\ldots+X_n)}{d}} \geq \sum_{s\in\mathcal{C}} \beta_s e^{\frac{2H(\sum_{j\in s} X_j)}{d}}. \tag{7}$$

Equality holds if and only if all the X_i are normal with proportional covariance matrices, and β is a fractional partition.

Remark I: A canonical example of a fractional packing is given by the coefficients

$$\beta_s = \frac{1}{r_+(s)}, \tag{8}$$

where $r_+(s) = \max_{i\in s} r(i)$ and $r(i)$ is the degree of i. Thus Conjecture I implies that

$$e^{\frac{2H(X_1+\ldots+X_n)}{d}} \geq \sum_{s\in\mathcal{C}} \frac{1}{r_+(s)} e^{\frac{2H(\sum_{j\in s} X_j)}{d}}.$$

The following weaker statement is the main result of Madiman and Barron [6]: if r_+ is the maximum degree (over all $i \in [n]$), then

$$e^{\frac{2H(X_1+\ldots+X_n)}{d}} \geq \frac{1}{r_+} \sum_{s \in \mathcal{C}} e^{\frac{2H(\sum_{j \in s} X_j)}{d}}. \tag{9}$$

It is easy to see that this generalizes the entropy power inequalities of Shannon-Stam [10], [11] and Artstein, Ball, Barthe and Naor [12].

Of course, one may further conjecture that the entropy power is also supermodular for sums, which would be stronger than Conjecture I. In general, it is an interesting question to characterize the class of all real numbers that can arise as entropy powers of sums of a collection of underlying *independent* random variables. Then the question of supermodularity of entropy power is just the question of whether this class, which one may call the *Stam class* to honour Stam's role in the study of entropy power, is contained in the class of polymatroidal vectors.

There has been much progress in recent decades on characterizing the class of all entropy inequalities for the joint distributions of a collection of (dependent) random variables. Major contributions were made by Han and Fujishige in the 1970's, and by Yeung and collaborators in the 1990's; we refer to [13] for a history of such investigations; more recent developments and applications are contained in [5] and [14]. The study of the Stam class above is a direct analogue of the study of the entropic vectors by these authors. Let us point out some particularly interesting aspects of this analogy, using X_s to denote $(X_i : i \in s)$ and H to denote (with some abuse of notation) the discrete entropy, where $(X_i : i \in [n])$ is a collection of dependent random variables taking values in some discrete space. Fujishige [3] proved that $h(s) = H(X_s)$ is a submodular set function (analogous to the conjecture on the supermodularity of entropy power), and Madiman and Tetali [1] proved that

$$\sum_{s \in \mathcal{C}} \beta_s H(X_s | X_{s^c}) \leq h([n]) \leq \sum_{s \in \mathcal{C}} \alpha_s h(s), \tag{10}$$

the upper bound of which is analogous to the entropy power inequalities for sums suggested in Conjecture I. However, it does not seem to be easy to derive entropy lower bounds for sums and inequalities for joint distributions using a common framework; even the recent innovative treatment of entropy power inequalities by Rioul [15], which partially addresses this issue, requires some delicate analysis in the end to justify the vanishing of higher-order terms in a Taylor expansion. On the other hand, inequalities for joint distributions tend to rely entirely on elementary facts such as chain rules. Nonetheless the formal similarity begs the question of how far this parallelism extends; in particular, are there "non-Shannon-type" inequalities for entropy powers of sums that impose restrictions beyond those imposed by the conjectured supermodularity?

VI. Application to determinants

It has been long known that the probabilistic representation of a matrix determinant using multivariate normal probability density functions can be used to prove useful matrix inequalities (see, for instance, Bellman's classic book [16]). A new variation on the theme of using properties of normal distributions to deduce inequalities for positive-definite matrices was begun by Cover and El Gamal [17], and continued by Dembo, Cover and Thomas [18] and Madiman and Tetali [1]. Their idea was to use the representation of the determinant in terms of the entropy of normal distributions, rather than the integral representation of the determinant in terms of the normal density. They showed that the classical Hadamard, Szasz and more general inequalities relating the determinants of square submatrices followed from properties of the joint entropy function. Our development below also uses the representation of the determinant in terms of entropy, namely, the fact that the entropy of the the d-variate normal distribution with covariance matrix K, written $N(0,K)$, is given by

$$H(N(0,K)) = \tfrac{1}{2} \log \left[(2\pi e)^d |K| \right].$$

The ground is prepared to present an upper bound on the determinant of a sum of positive matrices.

Theorem IV:[Upper Bound on Determinant of Sum] Let K and K_i be positive matrices of dimension d. Then the function $D(s) = \log |\sum_{i \in s} K_i|$, defined on the power set of $[n]$, is submodular. Furthermore, for any fractional covering α using any collection of subsets $\mathcal{C}$ of $[n]$,

$$|K + K_1 + \ldots + K_n| \leq |K|^{-a} \prod_{s \in \mathcal{C}} \left| K + \sum_{j \in s} K_j \right|^{\alpha_s},$$

where $a = \sum_{s \in \mathcal{C}} \alpha_s - 1$.

Proof: Substituting normals in Theorems I and III gives the result. ∎

As a corollary, one obtains for instance

$$|K + K_1 + \ldots + K_n|^{n-1} |K| \leq \prod_{i \in [n]} \left| K + \sum_{j \neq i} K_j \right|.$$

We now present lower bounds for the determinant of a sum.

Proposition I:[Generalized Minkowski determinant inequality] Let $K_1, \ldots, K_n$ be $d \times d$ positive matrices. Let $\mathcal{C}$ be a a collection of subsets of $[n]$ maximum degree r_+. Then

$$|K_1 + \ldots + K_n|^{\frac{1}{d}} \geq \frac{1}{r_+} \sum_{s \in \mathcal{C}} \left| \sum_{j \in s} K_j \right|^{\frac{1}{d}}.$$

Equality holds if and only if all the matrices K_i are proportional to each other, and each index appears in exactly r_+ subsets.

In particular, when $\mathcal{C}$ is the collection of singleton sets, $r_+ = 1$, and one recovers the classical inequality. Specifically, if K_1 and K_2 are $d \times d$ symmetric, positive-definite matrices, then

$$|K_1 + K_2|^{\frac{1}{d}} \geq |K_1|^{\frac{1}{d}} + |K_2|^{\frac{1}{d}}.$$

Many proofs of this inequality exist, see, e.g., [19]. Although the bound in Proposition I is better in general than the classical Minkowski bound of $|K_1|^{\frac{1}{d}} + \ldots + |K_n|^{\frac{1}{d}}$, mathematically the former can be seen as a consequence of the latter. Indeed, below we prove a more general version of Proposition I.

Theorem V: Let $K_1, \ldots, K_n$ be $d \times d$ positive matrices. Then for any fractional packing β using the collection $\mathcal{C}$ of subsets of $[n]$,

$$|K_1 + \ldots + K_n|^{\frac{1}{d}} \geq \sum_{s \in \mathcal{C}} \beta_s \left| \sum_{j \in s} K_j \right|^{\frac{1}{d}}.$$

Equality holds iff the matrices $\{\sum_{j \in s} K_j \, , \, s \in \mathcal{C}\}$ are proportional.

Proof: First note that for any fractional partition,

$$K_1 + \ldots + K_n = \sum_{s \in \mathcal{C}} \beta_s \sum_{j \in s} K_j.$$

Applying the usual Minkowski inequality gives

$$|K_1 + \ldots + K_n|^{\frac{1}{d}} \geq \sum_{s \in \mathcal{C}} \left| \beta_s \sum_{j \in s} K_j \right|^{\frac{1}{d}} = \sum_{s \in \mathcal{C}} \beta_s \left| \sum_{j \in s} K_j \right|^{\frac{1}{d}}.$$

For equality, we clearly need the matrices

$$\beta_s \sum_{j \in s} K_j \, , \, s \in \mathcal{C}$$

to be proportional. ■

It is not a priori clear that the bounds obtained using Proposition I (or Theorem V) are better than the direct Minkowski lower bound of

$$|K_1|^{\frac{1}{d}} + \ldots + |K_n|^{\frac{1}{d}},$$

but we observe that they are in general better. Consider the collection $\mathcal{C}_{n-1}$ of leave-one-out sets, i.e., $\mathcal{C} = \{s \subset [n] : |s| = n-1\}$. The maximum degree here is $r_+ = n - 1$, so

$$|K_1 + \ldots + K_n|^{\frac{1}{d}} \geq \frac{1}{n-1} \sum_{s \in \mathcal{C}_{n-1}} \left| \sum_{j \in s} K_j \right|^{\frac{1}{d}}. \tag{11}$$

Lower bounding each term on the right by iterating this inequality, and using $\mathcal{C}_k$ to denote the collection of all sets of size k, one obtains

$$|K_1 + \ldots + K_n|^{\frac{1}{d}} \geq \frac{2}{(n-1)(n-2)} \sum_{i \in [n]} \left| \sum_{j \in s} K_j \right|^{\frac{1}{d}}.$$

Repeating this procedure yields the hierarchy of inequalities

$$|K_1 + \ldots + K_n|^{\frac{1}{d}} \geq \mathrm{LB}_{n-1} \geq \mathrm{LB}_{n-2} \geq \ldots \geq \mathrm{LB}_1,$$

where LB_k is the lower bound given by Proposition I using $\mathcal{C}_k$. Note that the direct Minkowski lower bound, namely LB_1, is the worst of the hierarchy.

Note that Theorem V gives evidence towards Conjecture I, since it is simply the special case of Conjecture I for multivariate normals! (Conversely, the inequality (9) may be used to give an information-theoretic proof of Proposition I, which is a special case of Theorem V.)

There are other interesting related entropy and matrix inequalities that we do not have space to cover in this note; these can be found in the full paper [9].

Acknowledgment

The author would like to thank Andrew Barron and Prasad Tetali, who have collaborated with him on related work.

References

[1] M. Madiman and P. Tetali, “Information inequalities for joint distributions, with interpretations and applications,” *Submitted*, 2007.

[2] T. S. Han, “Nonnegative entropy measures of multivariate symmetric correlations,” *Information and Control*, vol. 36, no. 2, pp. 133–156, 1978.

[3] S. Fujishige, “Polymatroidal dependence structure of a set of random variables,” *Information and Control*, vol. 39, pp. 55–72, 1978.

[4] F. Chung, R. Graham, P. Frankl, and J. Shearer, “Some intersection theorems for ordered sets and graphs,” *J. Combinatorial Theory, Ser. A*, vol. 43, pp. 23–37, 1986.

[5] F. Matúš, “Two constructions on limits of entropy functions,” *IEEE Trans. Inform. Theory*, vol. 53, no. 1, pp. 320–330, 2007.

[6] M. Madiman and A. Barron, “Generalized entropy power inequalities and monotonicity properties of information,” *IEEE Trans. Inform. Theory*, vol. 53, no. 7, pp. 2317–2329, July 2007.

[7] T. Cover and J. Thomas, *Elements of Information Theory*. New York: J. Wiley, 1991.

[8] B. Bollobás and I. Leader, “Compressions and isoperimetric inequalities,” *J. Combinatorial Theory Ser. A*, vol. 56, no. 1, pp. 47–62, 1991.

[9] M. Madiman, “The entropy of sums of independent random elements,” *Preprint*, 2008.

[10] C. Shannon, “A mathematical theory of communication,” *Bell System Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.

[11] A. Stam, “Some inequalities satisfied by the quantities of information of Fisher and Shannon,” *Information and Control*, vol. 2, pp. 101–112, 1959.

[12] S. Artstein, K. M. Ball, F. Barthe, and A. Naor, “Solution of Shannon’s problem on the monotonicity of entropy,” *J. Amer. Math. Soc.*, vol. 17, no. 4, pp. 975–982 (electronic), 2004.

[13] J. Zhang and R. Yeung, “On characterization of entropy function via information inequalities,” *IEEE Trans. Inform. Theory*, vol. 44, pp. 1440–1452, 1998.

[14] R. Dougherty, C. Freiling, and K. Zeger, “Networks, matroids and non-Shannon information inequalities,” *IEEE Trans. Inform. Theory*, vol. 53, no. 6, pp. 1949–1969, June 2007.

[15] O. Rioul, “A simple proof of the entropy-power inequality via properties of mutual information,” *Proc. IEEE Intl. Symp. Inform. Theory, Nice, France*, pp. 46–50, 2007.

[16] R. Bellman, *Introduction to Matrix Analysis*. McGraw-Hill, 1960.

[17] T. Cover and A. El Gamal, “An information theoretic proof of Hadamard’s inequality,” *IEEE Trans. Inform. Theory*, vol. 29, no. 6, pp. 930–931, 1983.

[18] A. Dembo, T. Cover, and J. Thomas, “Information-theoretic inequalities,” *IEEE Trans. Inform. Theory*, vol. 37, no. 6, pp. 1501–1518, 1991.

[19] R. Bhatia, *Matrix Analysis*. Springer, 1996.