Hypergraphs, Entropy, and Inequalities

Ehud Friedgut

1. INTRODUCTION. Hypergraphs. Information Theory. Cauchy-Schwarz. It seems reasonable to assume that most mathematicians would be puzzled to find these three terms as, say, key words for the same mathematical paper. (Just in case this puzzlement is a result of being unfamiliar with the term "hypergraph": a hypergraph is nothing other than a family of sets, and will be defined formally later.) To further pique the curiosity of the reader we consider a simple "triangle inequality" that we will later associate with a very simple (hyper)graph—namely, the triangle K_3. Let X, Y, and Z be three independent probability spaces, and let

$$f : X \times Y \to \mathbb{R}, \quad g : Y \times Z \to \mathbb{R}, \quad h : Z \times X \to \mathbb{R}$$

be functions that are square-integrable with respect to the relevant product measures. Then

$$\int f(x, y)g(y, z)h(z, x)\, dx\, dy\, dz \\ \leq \sqrt{\int f^2(x, y)\, dx\, dy \int g^2(y, z)\, dy\, dz \int h^2(z, x)\, dz\, dx}.$$

The astute reader will notice that this inequality is slightly different from the generalized Hölder inequality that would apply had all three functions been defined on the same measure space. We will see in sections 4 and 5 that this inequality and many others are a consequence of an information-theoretic principle.

When studying the multitude of proofs of various famous inequalities, such as the arithmetic-geometric mean inequality, one often finds that they follow from the convexity of some underlying function and can be deduced from Jensen's inequality. Another prime example of an area where results implied by convexity play a central role is information theory: many of the properties of the entropy of a random variable have clear intuitive meaning, yet their proofs depend on the convexity of the function $x \mapsto x \log x$.

In this paper we wish to point out a common underlying information-theoretic theme that can be found in many well-known inequalities. The formulation of this principle is combinatorial and follows from a generalization of a result about hypergraphs commonly known as Shearer's Entropy Lemma.

Shearer's Entropy Lemma [**9**] is a combinatorial result comparing the size of the edge set of a hypergraph with the sizes of certain projections of this set. In a recent paper of the author with V. Rödl [**7**], a generalization of this lemma for weighted hypergraphs was established. This is Lemma 3.2, which we present here with its proof. The main tool we use is a more general version of this lemma, which we present in Lemma 3.3.

By applying Lemma 3.3 to certain specific hypergraphs, we recover the Cauchy-Schwarz inequality, Hölder's inequality, the monotonicity of the Euclidean norm, the monotonicity of weighted means, and other lesser known inequalities.

Finally we present a continuous version of this method. A nice aspect of the latter is that it offers a concise way of representing inequalities by weighted hypergraphs, and vice versa. After presenting the content of this paper in a workshop the author learned from Andrew Thomason that the continuous version had already been proved by Helmut Finner in [**6**] (albeit without using the language of hypergraphs or the notion of entropy).

2. DEFINITIONS AND A SIMPLE EXAMPLE.

Entropy. We start by giving a simple example of how the Cauchy-Schwarz inequality can be interpreted as an information-theoretic result. First we recall some basic definitions. (For background on entropy and proofs of the results stated here see, for example, [**5**].) In what follows X, Y, and so forth denote discrete random variables taking values in finite sets. Also, log signifies the logarithm base 2.

The *entropy* $H(X)$ of a random variable X is defined by

$$H(X) = \sum_x p(x) \log \frac{1}{p(x)},$$

where we write $p(x)$ for $\Pr(X = x)$ and extend this notation in the natural way to other contexts, as they arise. Note that the entropy of X does not depend on the values that X assumes, but only on the probabilities with which they are assumed. The intuitive meaning of $H(X)$ is that it expresses the expected number of bits of information one might need in order to communicate the value of X, or equivalently, the number of bits of information conveyed by X. It is always true that

$$H(X) \leq \log \left|\text{Support}(X)\right|, \tag{1}$$

where Support(X) is the set of values that X assumes and $|A|$ denotes the cardinality of a set A. Equality occurs if and only if X is uniformly distributed on its support (i.e., assumes all possible values with the same probability). The *conditional entropy* $H(X \mid Y)$ of X given Y is given by

$$H(X \mid Y) = \mathbf{E} H(X \mid Y = y) = \sum_y p(y) \sum_x p(x \mid y) \log \frac{1}{p(x \mid y)},$$

where the expectation $\mathbf{E}$ is taken over the values of Y. Intuitively $H(X \mid Y)$ measures the expected amount of information X conveys to an observer who knows the value of Y. It is therefore not surprising that

$$H(X \mid Y) \leq H(X), \tag{2}$$

and

$$H(X \mid Y) = H(X) \text{ if and only if } X \text{ and } Y \text{ are independent.} \tag{3}$$

Given a collection of random variables $\{Y_i : i \in I\}$, let $Y_I = (Y_i : i \in I)$ be the corresponding vector of random variables. Note that Y_I is itself a random variable. We use the notation $H(X \mid Y_i : i \in I)$ to denote $H(X \mid Y_I)$. With this notation inequality (2) generalizes as follows: for any subset J of I

$$H(X \mid Y_i : i \in I) \leq H(X \mid Y_i : i \in J). \tag{4}$$

This inequality, which is central to the proof of the main lemma in this paper, has the intuitive meaning that the more information one knows, the less information is conveyed by X. For a vector of random variables $X = (X_1, \dots, X_n)$ we have the following chain rule:

$$H(X) = H(X_1) + H(X_2 \mid X_1) + \cdots + H(X_n \mid X_1, \dots, X_{n-1}). \tag{5}$$

A simple example: Cauchy-Schwarz. Here, and in most of the rest of the paper, we prefer the discrete setting and will be proving inequalities involving variables with integer values. The case of real values can be deduced by approximating reals with rationals and deriving the rational case from the integral one. The continuous analogs follow as limiting cases of the discrete.

Let $A_1, A_2, \dots, A_n$ be pairwise disjoint disjoint finite subsets of $\mathbb{Z}$ with $|A_i| = a_i$ for $i = 1, \dots, n$. Likewise let $B_1, B_2, \dots, B_n$ be pairwise disjoint sets of integers with $|B_i| = b_i$. Let $R_i = A_i \times B_i \subset \mathbb{Z}^2$. Choose an index r at random from $\{1, \dots, n\}$ with

$$\Pr(r = i) = \frac{a_i b_i}{\sum_k a_k b_k}.$$

Now choose two random points independently, uniformly from the points of R_r. Denote these points by (x_1, y_1) and (x_2, y_2). Note that both points are uniformly distributed among all $\sum_k a_k b_k$ points of $\bigcup R_k$.

Next, write the four numbers x_1, y_1, x_2, y_2 on four cards and hand them to Alice and Bob, two of the mythical heroes of information theory. Assume that Alice gets the cards with x_1 and y_1, and Bob gets those with x_2 and y_2. Our two characters now set out to sell their information on the free market, where each bit of information sells for, say, one dollar. Interpreting the entropy of a random variable as the number of bits of information it carries they should hope to gain together

$$H(x_1, y_1) + H(x_2, y_2) = 2 \log \left(\sum_k a_k b_k \right)$$

dollars. (Recall that (x_i, y_i) is uniformly distributed over all relevant values.)

Now imagine that, before selling their information, Alice and Bob meet for lunch and agree to redistribute their prospective wealth, Alice taking the cards with x_1 and x_2, and Bob taking the y-cards. It seems obvious that by doing so they have not enlarged their combined wealth; indeed,

$$\begin{aligned} H\big((x_1, y_1)\big) + H\big((x_2, y_2)\big) &= H(R_r) + H(x_1 \mid R_r) + H(y_1 \mid R_r) \\ &\quad + H(R_r) + H(x_2 \mid R_r) + H(y_2 \mid R_r) \\ &= H\big((x_1, x_2)\big) + H\big((y_1, y_2)\big). \end{aligned} \tag{6}$$

But Alice's new random variable (x_1, x_2) is distributed (although not necessarily uniformly) over $\sum a_k^2$ different values, whereas Bob's random variable (y_1, y_2) can assume $\sum b_k^2$ different values. Hence together they can hope to earn at the very most $\log \sum a_k^2 + \log \sum b_k^2$ dollars. The conclusion is that

$$2 \log \left(\sum_k a_k b_k \right) \le \log \sum_k a_k^2 + \log \sum_k b_k^2,$$

i.e.,

$$\left(\sum_k a_k b_k\right)^2 \le \left(\sum_k a_k^2\right)\left(\sum_k b_k^2\right).$$

This, of course, is the Cauchy-Schwarz inequality.

3. ENTROPY LEMMAS. Before stating the lemmas of this section we recall the definition of a hypergraph. A *hypergraph* $H = (V, E)$ consists of a set V of vertices and a set E of edges, where each member e of E is a subset of V. A graph is a hypergraph such that all edges have cardinality 2.

Shearer's Entropy Lemma is a lemma relating the number of edges in a hypergraph to the numbers of edges in certain projections of the hypergraph. In a recent paper [**7**] the author and V. Rödl used a weighted version of this lemma to prove a special case of a hypercontractive estimate of Bonami and Beckner [**7**], [**1**], [**4**].

We start by quoting the result from [**9**]:

Lemma 3.1 (Shearer's Entropy Lemma). *Let t be a positive integer, let $H = (V, E)$ be a hypergraph, and let $F_1, F_2, \dots, F_r$ be subsets of V such that every vertex in V belongs to at least t of the sets F_i. Let H_i $(i = 1, 2, , \dots, r)$ be the projection hypergraphs: $H_i = (V, E_i)$, where $E_i = \{e \cap F_i : e \in E\}$. Then*

$$|E|^t \le \prod |E_i|.$$

The original proof of this lemma uses induction on t. However, there exists a more intuitive proof that takes an information-theoretic approach. This proof, probably first discovered by Jaikumar Radhakrishnan [**8**], has existed only as an item of folklore. The proof of Lemma 3.2 found in [**7**] is a generalization of the folklore proof.

Lemma 3.2 (Weighted Entropy Lemma). *Let H, E, V, t, and F_i be as in Lemma 3.1, and for each e in E let the edge $e_i = e \cap F_i$ of E_i be endowed with a nonnegative real weight $w_i(e_i)$. Then*

$$\left(\sum_{e\in E}\prod_{i=1}^{r} w_i(e_i)\right)^t \le \prod_i \sum_{e_i\in E_i} w_i(e_i)^t. \tag{7}$$

Furthermore, if for for each e in E

$$w(e) = \prod_{i=1}^{r} w_i(e_i),$$

then a necessary condition for equality to hold in (7) is that for $i = 1, \dots, r$ and for each e^* in E_i

$$\frac{w_i(e^*)^t}{\sum_{e_i\in E_i} w_i(e_i)^t} = \frac{\sum\{w(e') : e' \in E, e'_i = e^*\}}{\sum_{e\in E} w(e)}. \tag{8}$$

Of course, setting all weights equal to 1 in Lemma 3.2 gives Shearer's lemma. Moreover, it turns out that in all the applications that we intend to study the necessary condition for equality in (8) is also sufficient.

We include the proof of Lemma 3.2 here both for the sake of completeness and because it sheds some light on the connection between information theory and the inequalities we prove. Essentially what follows can be interpreted as a generalization of the story of Alice and Bob from the previous section, where the pair is replaced by a set of r information merchants.

Proof. Clearly we may assume all weights are positive integers. For simplicity of notation we assume that $V = \{1, \ldots, n\}$. We will now work with *multihypergraphs*, a simple generalization of hypergraphs: the set of edges of a multihypergraph is a multiset (i.e., some edges may appear with multiplicity). Define a new multihypergraph $H' = (V, E')$ by creating $\prod_i w_i(e_i)$ copies $e^{(c_1,\ldots,c_r)}$ of each edge e, where $1 \le c_i \le w_i(e_i)$ for $1 \le i \le r$. Similarly for $1 \le i \le r$ we define H_i' by creating $w_i(e_i)$ copies of every edge e_i.

Consider an edge e of H' chosen uniformly at random from the collection of all edges. Set

$$Y = Y(e) = (X, C) = (x_1, \ldots, x_n, c_1, \ldots, c_r),$$

where $X = (x_1, \ldots, x_n)$ is the characteristic vector of e (i.e., $x_k = 1$ if k belongs to e and $x_k = 0$ otherwise) and $C = (c_1, \ldots, c_r)$ gives the index of the copy of the given edge.

This defines a random variable Y that is uniformly distributed over a set of $\sum_{e\in E} \prod_i w_i(e_i)$ values, whence

$$H(Y) = \log\left(\sum_{e\in E}\prod_i w_i(e_i)\right). \tag{9}$$

We now define r random variables Y^i ($1 \le i \le r$): Y^i corresponds to picking an edge e uniformly from H', observing its projection e_i, and then choosing with replacement t independent copies of e_i from the $w_i(e_i)$ possible copies. For this, let $X^i = (x_1^i, \ldots, x_n^i)$ be the characteristic vector of $e_i = e \cap F_i$. Note that this vector is derived from X by setting the coordinates not in F_i to 0, hence the variables x_k^i (with k in F_i) have the same joint distribution as the corresponding x_k. Next, let $c_i^1, \ldots, c_i^t$ $(1 \le i \le r)$ be t independent random variables such that the joint distribution of (X, c_i^k) is the same as that of (X, c_i). Observe that $1 \le c_i^k \le w_i(e_i)$ for $1 \le k \le t$. Finally, define

$$Y^i = (X^i, C^i) = (X^i, c_i^1, \ldots, c_i^t).$$

Then Y^i can take on at most $\sum_{e_i\in E_i} w_i(e_i)^t$ different values. It follows from (1) that

$$H(Y^i) \le \log\left(\sum_{e_i\in E_i} w_i(e_i)^t\right). \tag{10}$$

To complete the proof of the lemma we must show that

$$tH(Y) \le \sum H(Y^i).$$

By (5),

$$H(Y) = \sum_{m=1}^{n} H(x_m \mid x_l : l < m) + \sum_{i=1}^{r} H(c_i \mid x_1, \ldots, x_n).$$

Similarly,

$$H(Y^i) = \sum_{m \in F_i} H(x_m \mid x_l : l < m, l \in F_i) + tH(c_i \mid x_l : l \in F_i).$$

Here we have used the fact $(c_i^k \mid X)$ has the distribution of $(c_i \mid X)$ and depends only on $\{x_l : l \in F_i\}$. Accordingly,

$$\begin{aligned}\sum H(Y^i) - tH(Y) \\ &= \sum_{m=1}^{n} \left(\left[\sum_{i : m \in F_i} H(x_m \mid x_l : l < m, l \in F_i) \right] - tH(x_m \mid x_l : l < m) \right) \\ &\quad + t \left(\sum_{i=1}^{r} H(c_i \mid x_l, l \in F_i) - H(c_i \mid x_1, \dots, x_n) \right).\end{aligned}$$

Finally, using the fact that every m belongs to at least t of the F_i and appealing to (4), we observe that all terms in the last expression are positive. Therefore, $tH(Y) \leq \sum H(Y^i)$ as required. A necessary condition for $tH(Y) = \sum H(Y^i)$ to hold is that equality hold in (10) for each i. This implies that all the Y^i are uniformly distributed on their respective supports, which is exactly equivalent to (8). ■

In the setting of the previous lemma the fact that every vertex is covered by at least t of the F_i can be restated as follows: if each F_i is endowed with the weight $1/t$, then every vertex is "covered" by weight at least 1. It turns out that it is useful to generalize this to allow the sets F_i to receive different weights in a manner known as a *fractional covering* of the hypergraph they describe. This is formulated precisely in the next lemma.

Lemma 3.3 (Generalized Weighted Entropy Lemma). *Let H, E, V, F_i, w_i, and w be as in Lemma 3.2. If $\alpha = (\alpha_1, \dots, \alpha_r)$ is a vector of weights such that*

$$\sum_{v \in F_i} \alpha_i \geq 1$$

for each v in V (α is a "fractional cover" of the hypergraph whose edges are the F_i), then

$$\sum_{e \in E} \prod_{i=1}^{r} w_i(e_i) \leq \prod_i \left(\sum_{e_i \in E_i} w_i(e_i)^{1/\alpha_i} \right)^{\alpha_i}.$$

Equality holds only if for $i = 1, \dots, r$ and for each e^* in E_i

$$\frac{w_i(e^*)^{1/\alpha_i}}{\sum_{e_i \in E_i} w_i(e_i)^{1/\alpha_i}} = \frac{\sum\{w(e') : e' \in E, e'_i = e^*\}}{\sum_{e \in E} w(e)}. \tag{11}$$

This lemma can be proved directly or deduced from Lemma 3.2 by approximating real numbers with rationals and constructing an appropriate multihypergraph for which the number of copies of each edge is determined by its weight. We omit the details.

4. APPLICATIONS. We now see how using Lemma 3.3 on certain hypergraphs yields interesting inequalities. Our first three examples involve a very simple hypergraph with disjoint edges.

- Let $V = \{1, 2, \ldots, n\}$ and $F_1 = F_2 = V$, let $E = V$ (each edge consists of one vertex), and let $\alpha = (1/2, 1/2)$. For each edge k in E set $w_1(k_1) = a_k$ and $w_2(k_2) = b_k$, where the a_k and b_k are real numbers. (We retain the notation k_1 and k_2 even though $k_1 = k_2 = k$). Then Lemma 3.3 gives

$$\left(\sum a_k b_k\right) \leq \sqrt{\sum a_k^2}\sqrt{\sum b_k^2},$$

  which is the Cauchy-Schwarz inequality. By condition (11) equality occurs only if

$$(a_1^2, \ldots, a_n^2) \sim (b_1^2, \ldots, b_n^2) \sim (a_1 b_1, \ldots, a_n b_n),$$

  which implies that the vectors $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ are proportional. As promised, this is also a sufficient condition for equality.
- Take the same hypergraph as before, but now use the fractional cover

$$\alpha = (\lambda, 1 - \lambda).$$

  This leads to

$$\sum a_k b_k \leq \left(\sum a_k^{1/\lambda}\right)^{\lambda} \left(\sum a_k^{1/(1-\lambda)}\right)^{(1-\lambda)}$$

  (i.e., to Hölder's inequality). Once again, using (11) one may recover the condition for equality.
- It is easy to generalize the previous examples, still using the same hypergraph but now with k sets $F_1, \ldots, F_k$, to get the generalized Hölder's inequality: for any positive real numbers $r_1, \ldots, r_k$ such that $\sum 1/r_i = 1$ and any kn nonnegative numbers $a_{11}, a_{12}, \ldots, a_{1k}, \ldots, a_{nk}$ it is true that

$$\sum_{i=1}^{n} \prod_{j=1}^{k} a_{ij} \leq \prod_{j=1}^{k} \left(\sum_{i=1}^{n} a_{ij}^{1/r_j}\right)^{r_j}.$$

- Here is an example of a slightly different kind. We take $t < r$, $V = \{1, \ldots, r\} \times \{1, \ldots, n\}$, $F_i = \{i, i+1, \ldots, i+t-1\} (\mathrm{mod}\ r) \times \{1, \ldots, n\}$, $\alpha = (1/t, 1/t, \ldots, 1/t)$, and $E = \{e^{(k)} = \{1, \ldots, r\} \times \{k\} : k = 1, \ldots, n\}$. Assign weights as follows: for $e^{(k)}$ in E

$$w_1(e_1^{(k)}) = w_2(e_2^{(k)}) = \cdots = w_r(e_r^{(k)}) = a_k,$$

  where a_k is a nonnegative real number. Applying Lemma 3.3 we get

$$\sum_k a_k^r \leq \left(\sum_k a_k^t\right)^{r/t}$$

  or, equivalently,

$$\left(\sum_k a_k^r\right)^{1/r} \leq \left(\sum_k a_k^t\right)^{1/t}. \tag{12}$$

Recalling that $r > t$ and noting the homogeneity of (12) we recover the monotonicity of the ℓ^p norm: if $p > q > 0$ and w belongs to $\mathbf{R}^n$, then $\|w\|_p \leq \|w\|_q$. Condition (11) now implies that equality holds if and only if

$$(a_1^r, \ldots, a_n^r) \sim (a_1^t, \ldots, a_n^t)$$

(i.e., all the a_k are equal).

- A variation on this theme is the following: we now use $V = \{1, \ldots, n\}$, $F_1 = \cdots = F_r = V$, $E = \{1, \ldots, n\}$, and $\alpha = (1/r, 1/r, \ldots, 1/r)$. Fix an integer s less than r. Assign weights by the rule

$$w_1(k_1) = \cdots = w_s(k_s) = a_k,$$

and

$$w_{s+1}(k_{s+1}) = \cdots = w_r(k_r) = 1$$

for k in E. We invoke Lemma 3.3 to obtain

$$\sum a_k^s \leq \left(\sum a_k^r\right)^{s/r} n^{(r-s)/r},$$

or rearranging

$$\left(\frac{\sum a_k^s}{n}\right)^{1/s} \leq \left(\frac{\sum a_k^r}{n}\right)^{1/r}$$

for any nonnegative real a_k. Recalling that $r > s$ and, as in the previous example, taking the homogeneity into account we find that the p-average of $(a_1, \ldots, a_n)$,

$$\left(\frac{\sum a_k^p}{n}\right)^{1/p},$$

is monotone increasing in p and is strictly increasing unless all the a_k are equal. This result contains the arithmetic-geometric mean inequality, since the arithmetic mean is the p-average with $p = 1$ and the geometric mean is the limit of the p-averages as p tends to 0.

- Having now warmed up, we move to hypergraphs that have more interesting combinatorial structures. We begin with a complete tripartite hypergraph. Consider three arbitrary disjoint sets I, J, and K. As vertices of the hypergraph we take the elements of $I \cup J \cup K$, and let the edge set consist of all sets $\{i, j, k\}$ with i in I, j in J, and k in K. Next define $F_1 = I \cup J$, $F_2 = J \cup K$, $F_3 = K \cup I$, and take the fractional cover $\alpha = (1/2, 1/2, 1/2)$. For an edge $e = \{i, j, k\}$ set $w_1(e_1) = a_{ij}$, $w_2(e_2) = b_{jk}$, $w_3(e_3) = c_{ki}$. Lemma 3.3 tells us that

$$\sum a_{ij} b_{jk} c_{ki} \leq \sqrt{\sum a_{ij}^2 \sum b_{ij}^2 \sum c_{ij}^2}, \tag{13}$$

where a_{ij}, b_{jk}, c_{ki} are real numbers. This inequality can be written more efficiently in matrix form:

$$\operatorname{Tr}(ABC) \leq \sqrt{\operatorname{Tr}(AA^t) \operatorname{Tr}(BB^t) \operatorname{Tr}(CC^t)}. \tag{14}$$

(Here M^t signifies the transpose and $\mathrm{Tr}(M)$ the trace of a matrix M.) Equality holds in (14) if and only if

$$\frac{A^t}{\mathrm{Tr}(AA^t)} = \frac{BC}{\mathrm{Tr}(ABC)},$$
$$\frac{B^t}{\mathrm{Tr}(BB^t)} = \frac{CA}{\mathrm{Tr}(ABC)},$$

and

$$\frac{C^t}{\mathrm{Tr}(CC^t)} = \frac{AB}{\mathrm{Tr}(ABC)}.$$

The continuous version of this inequality is also aesthetically pleasing and will be discussed in the next section.

- The previous example can be generalized in several directions, of which we mention only one here: if $A_1, \dots, A_k$ are matrices such that the product $A_i A_{i+1}$ is defined for $i = 1, \dots, k-1$ and the product $A_1 A_2 \cdots A_k$ is square, then

$$\mathrm{Tr}\left(\prod A_i\right) \le \sqrt{\prod \mathrm{Tr}(A_i A_i^t)}.$$

5. CONTINUOUS INEQUALITIES. Lemma 3.3 has a continuous analog that we state without proof. The continuous version was discovered by Finner (see [**6**], where it is proved without the use of entropy). The continuous version allows us to give many inequalities a nice graphic (or perhaps we should say hypergraphic) representation. If the reader believes (as we do) that a picture is worth 10^3 words, then he or she is urged to skip the statement of the theorem and proceed immediately to the figures that follow, only afterwards returning to the theorem for precise details. It is important to note, however, that *the functions in question are defined on different spaces.*

Lemma 5.1. *Let $H = (V, \mathcal{F})$ be a hypergraph, where $\mathcal{F} = \{F_1, \dots, F_r\}$, and let $\alpha = (\alpha_1, \dots, \alpha_r)$ be a nonnegative vector such that*

$$\sum_{v \in F_i} \alpha_i \ge 1$$

holds for each v in V. Assume that with each v in V is associated a measure space X_v with measure μ_v and with each F_i in $\mathcal{F}$ a nonnegative function $w_i : \prod_{v \in F_i} X_v \to \mathbf{R}$. Then, under the assumption that all functions involved are integrable,

$$\int \prod_i w_i \prod_{v \in V} d\mu_v \le \prod_i \left(\int w_i^{1/\alpha_i} \prod_{v \in F_i} d\mu_v \right)^{\alpha_i},$$

or

$$\left\| \prod w_i \right\|_1 \le \prod \|w_i\|_{1/\alpha_i}.$$

Equality holds only if for each i

$$w_i^{1/\alpha_i} \sim \int \prod w_i \prod_{v \notin F_i} d\mu_v.$$

The shift in notation in Lemma 5.1 from that in Lemma 3.3 is not accidental: the edges F_i in this lemma arise from the sets F_i in Lemma 3.3 when translating the discrete to the continuous.

We now present a series of figures that exemplify instances of Lemma 5.1. We start with the simplest, namely, a hypergraph with one vertex and two edges of size one. This is Figure 1, which represents Hölder's inequality.

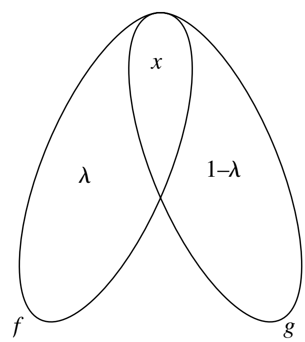

Figure 1. Hölder's inequality: $\|fg\|_1 \leq \|f\|_\lambda \|g\|_{1-\lambda}$.

The next example, depicted in Figure 2, is an unusual cyclical version of Hölder's inequality. It is instructive because it involves a more subtle combinatorial structure.

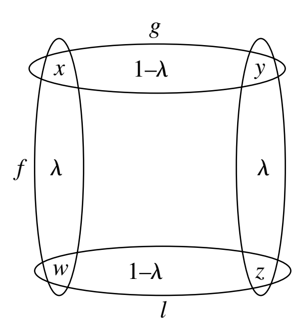

Figure 2. The cyclic Hölder inequality: $\|fghl\|_1 \leq \|f\|_\lambda \|g\|_{1-\lambda} \|h\|_\lambda \|l\|_{1-\lambda}$.

Observe two things about this example:

- Each of the four different functions is defined on the product of a different pair of spaces. Written out in long form the inequality becomes

$$\int \left| f(w,x)g(x,y)h(y,z)l(z,w) \right| dw\,dx\,dy\,dz \leq$$
$$\left(\int |f|^\lambda \,dw\,dx \right)^{1/\lambda} \times \left(\int |g|^{1-\lambda} \,dx\,dy \right)^{1/(1-\lambda)}$$
$$\times \left(\int |h|^\lambda \,dy\,dz \right)^{1/\lambda} \times \left(\int |l|^{1-\lambda} \,dz\,dw \right)^{1/(1-\lambda)}.$$

- The inequality is not symmetric in the functions! The combinatorial structure of the four-cycle actually plays a role. For example, if the roles of f and g are switched, the resulting inequality is false.

Returning to a more symmetrical setting, Figure 3 represents a "triangle" inequality:

$$\int f(x, y)g(y, z)h(z, x)\, dx\, dy\, dz \leq \sqrt{\int f^2(x, y)\, dx\, dy \int g^2(y, z)\, dy\, dz \int h^2(z, x)\, dz\, dx}.$$

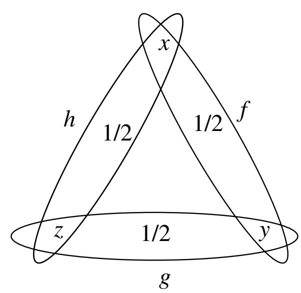

Figure 3. "Triangle" Inequality: $\|fgh\|_1 \leq \|f\|_2\|g\|_2\|h\|_2$.

This is the continuous version of inequality (13). It can also be proved by repeated application of the Cauchy-Schwarz inequality. Note the difference between the "triangle" inequality and the Hölder inequality

$$\int f(x)g(x)h(x)\, dx \leq \sqrt[3]{\int |f(x)|^3\, dx \int |g(x)|^3\, dx \int |h(x)|^3\, dx}.$$

The triangular version asserts a stronger statement, replacing the 3-norms in the right hand side of Hölder's inequality with 2-norms. This holds only because each of the functions in question is defined on the product of a different pair of spaces.

This last inequality has, of course, infinitely many generalizations that may be neatly represented by simple graphs (e.g., the inequalities that arise from cycles of length more than three). Some of these generalizations, which can be established by clever repeated application of Hölder's inequality, are found in [**2**], a 1979 paper by Ron Blei, but they were probably known even earlier to Dvoretzky [**3**].

A different direction of generalization results from moving to hypergraphs with edges of size larger than two. We give one example arising from the complete 3-uniform hypergraph on four vertices. This is Figure 4, which represents a "tetrahedron" inequality:

$$\int \left(\prod_{i=0}^{3} f_i\right) dx_0\, dx_1\, dx_2\, dx_3 \leq \prod_{i=0}^{3} \sqrt[3]{\int \left(|f_i|^3\right)\, dx_i\, dx_{i+1}\, dx_{i+2}},$$

where all indices are taken modulo 4.

Finally we would like to remark that the discovery of Lemma 3.2 in [**7**] was motivated by applying it to the hypergraph of Figure 5. It turns out that, with the appropriate

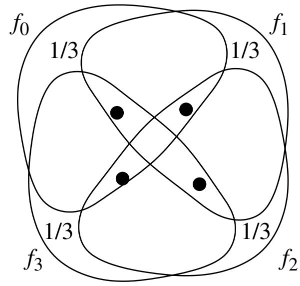

Figure 4. The tetrahedron inequality: $\| \prod f_i \|_1 \leq \prod \| f_i \|_3$.

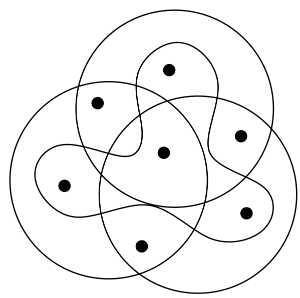

Figure 5. The even-Venn hypergraph.

fractional covering, this hypergraph gives rise to an inequality that can be used to deduce valuable information concerning Boolean functions on product spaces (see [**7**]).

ACKNOWLEDGMENTS. I would like to thank Ron Blei, Vojtech Rödl, Leonard Schulman, Andrew Thomason, and Avi Wigderson for useful remarks, Jeff Kahn for pointing out a subtlety concerning the conditions for equality, and Shlomo Huri for technical issues.

REFERENCES

1. W. Beckner, Inequalities in Fourier analysis, *Ann. of Math.* **102** (1975) 139–182.
2. R. Blei, Fractional Cartesian products of sets, *Ann. Inst. Fourier (Grenoble)* **29** (2) (1979) 79–105.
3. ———, private communication.
4. A. Bonami, Etude des coefficients Fourier des fonctiones de $L^p(G)$, *Ann. Inst. Fourier (Grenoble)* **20** (2) (1970) 335–402.
5. I. Csiszár and J. Körner, *Information Theory. Coding Theorems for Discrete Memoryless Systems,* Akadémiai Kiadó, Budapest, 1981.
6. H. Finner, A generalization of Hölder's inequality and some probability inequalities, *Ann. Probab.* **20** (1992) 1893–1901.
7. E. Friedgut and V. Rödl, Proof of a hypercontractive estimate via entropy, *Israel J. Math.* **125** (2001) 369–380.
8. J. Radhakrishnan, private communication.
9. F. R. K. Chung, P. Frankl, R. L. Graham, and J. B. Shearer, Some intersection theorems for ordered sets and graphs, *J. Comb. Theory Ser. A* **43** (1986) 23–37.

EHUD FRIEDGUT graduated from the Hebrew University of Jerusalem in 1997 and returned to the mathimatics department there in 2000 as a faculty member. His research interests are mainly probabilistic combinatorics and discrete harmonic analysis. In his spare time he practices yoga, which he considers to be an antipodal point to mathematics on the sphere of human affairs.
Institute of Mathematics, Hebrew University of Jerusalem, Givat Ram, Jerusalem, 91904, Israel
ehudf@math.huji.ac.il