Combinatorics, Probability and Computing (2007) 16, 693–712. © 2007 Cambridge University Press
doi:10.1017/S0963548306008340 First published online 23 January 2007 Printed in the United Kingdom

Edge-Isoperimetric Inequalities and Influences

DVIR FALIK[†] and ALEX SAMORODNITSKY[‡]

School of Computer Science and Engineering, Hebrew University, Jerusalem, Israel
(e-mail: dfalik, salex@cs.huji.ac.il)

Received 26 December 2005; revised 22 October 2006

We give a combinatorial proof of the result of Kahn, Kalai and Linial [16], which states that every balanced boolean function on the n-dimensional boolean cube has a variable with influence of at least $\Omega\left(\frac{\log n}{n}\right)$. The methods of the proof are then used to recover additional isoperimetric results for the cube, with improved constants.

We also state some conjectures about optimal constants.

1. Introduction

This paper deals with isoperimetric problems on graphs. Given a graph $G = (V, E)$, the vertex boundary of a subset $S \subseteq V$ contains the vertices of S which have neighbours outside S. The edge boundary of S is the set of edges crossing from S to its complement.

The question of the smallest-possible boundary possessed by a set of given cardinality is an important combinatorial question with obvious connections to classical isoperimetry. Good estimates of the minimal boundary size are also very useful in applications. We briefly mention two. Lower bounds on the vertex boundary show how fast a neighbourhood of a set has to grow when the allowed distance from the set increases, and this leads to concentration of measure results for Lipschitz functions on the graph [21]. Lower bounds on the edge boundary suggest that a simple random walk on the graph does not remain in any subset for too long, and this leads to upper bounds on its mixing time [14].

Early isoperimetric results on graphs include isoperimetric theorems for the boolean cube $\{0,1\}^n$. This is a graph with 2^n vertices indexed by boolean strings of length n. Two vertices are connected by an edge if they differ only in one coordinate. The metric defined by this graph is called the Hamming distance. Two vertices x and y are at distance d if they differ in d coordinates. A Hamming ball is a ball in this metric. A subcube is a subset

[†] Supported by ISF grant number 039-7165.
[‡] Supported by ISF grant number 039-7165 and by GIF grant number I-2052.

https://doi.org/10.1017/S0963548306008340 Published online by Cambridge University Press

of the vertices obtained by fixing the value in some of the coordinates. The number of fixed coordinates is called the co-dimension of the subcube. It turns out [12, 11, 13] that the vertex boundary of a Hamming ball is smallest among all sets of equal size, and the same is true for the edge boundary of a subcube. (It is also possible to interpolate these results for all the intermediate subset sizes.)

While many other exact vertex and edge isoperimetric results are known (see [3] for a survey), in most cases exact results seem hard to obtain. In many of these cases they could be replaced by sufficiently strong approximate isoperimetric results [15, 23, 28].

In this paper we focus on edge-isoperimetric questions in the boolean cube. A major result in this area was obtained by Kahn, Kalai and Linial [16], who showed that any balanced boolean function has a variable with large *influence*. We proceed to describe this result, starting with some background.

For a subset $A \subseteq \{0,1\}^n$ and an index $1 \leqslant i \leqslant n$ let $I_i(A)$ be the fraction of edges in direction i between A and its complement A^c. This means that $2^{n-1} \cdot I_i(A)$ counts the edges with one vertex in A and another in A^c, the vertices disagreeing in the ith coordinate. $\sum_{i=1}^n I_i(A)$ is the total (normalized) cardinality of the edge boundary of A.

The familiar edge-isoperimetric inequality in the cube states that, for any subset $A \subseteq \{0,1\}^n$, we have[1]

$$\sum_{i=1}^{n} I_i(A) \geqslant \frac{2}{\log 2} \cdot \frac{|A|}{2^n} \log \frac{2^n}{|A|}.$$

This is tight if A is a subcube of (arbitrary) co-dimension $0 \leqslant t \leqslant n$.

Let f be the characteristic function of A, with expectation $\mu = \mathbb{E}_{x \in \{0,1\}^n} f(x) = \frac{|A|}{2^n}$. Then the edge-isoperimetric inequality asserts that

$$\sum_{i=1}^{n} I_i(f) \geqslant \frac{2}{\log 2} \cdot \mu \log \frac{1}{\mu}. \tag{1.1}$$

We interchange freely between a set and its characteristic function. Whenever this does not cause confusion we do not mention either, and simply write I_i.

The inequality (1.1) has several easy proofs [11, 13]. The one most relevant to this discussion is by induction on dimension. To illustrate its outlay and its simplicity, here is a sketch (the base $n = 1$ is easy). Assume the inequality holds for dimension $n-1$ and consider the case of dimension n. Write $A = A_0 \cup A_1$, where A_i contains all the elements of A with i in the nth coordinate. Think of the A_i as subsets of the $(n-1)$-dimensional cube, and observe that $I_n(A) \geqslant \left(1/2^{n-1}\right) \cdot ||A_0| - |A_1||$. Taking $a_i = |A_i|$, and slightly simplifying, it remains to check that for any non-negative a_0, a_1 we have

$$a_0 \log \frac{2^{n-1}}{a_0} + a_1 \log \frac{2^{n-1}}{a_1} + \log 2 \cdot |a_0 - a_1| \geqslant (a_0 + a_1) \log \frac{2^n}{a_0 + a_1},$$

which is not hard to verify, using, *e.g.*, concavity of the entropy function $H(x) = -x \log x - (1-x)\log(1-x)$ on $[0,1]$.

[1] We use natural logarithms throughout the paper.

https://doi.org/10.1017/S0963548306008340 Published online by Cambridge University Press

Things become more complicated when we ask for more detailed information. Interpreting the set A as the set of positive outcomes of a game with n players, the number $I_i(A)$ acquires a game-theoretic interpretation as the *influence* of the ith player on the outcome of the game, namely the probability that the outcome of the game remains uncertain if the decisions of other players are chosen at random. Motivated by questions from computational game theory, Ben-Or and Linial [2] conjectured that, for any balanced game (namely $|A| = 2^{n-1}$), there is a player with influence of at least $\Omega\left(\frac{\log n}{n}\right)$.

This conjecture was proved by Kahn, Kalai and Linial in [16].

Theorem 1.1. *Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function with expectation $\mathbb{E}f = \mu$. Then*

$$\sum_{i=1}^{n} I_i^2(f) \geqslant \Omega\left(\frac{\mu^2(1-\mu)^2 \log^2 n}{n}\right). \tag{1.2}$$

In particular, there exists i with $I_i \geqslant \Omega\left(\frac{\mu(1-\mu)\log n}{n}\right)$.

Kahn, Kalai and Linial's paper [16] is one of the first to use Fourier analysis on $\mathbb{Z}_2^n$ in a combinatorial setting. Rather surprisingly, a crucial tool in the proof is an inequality [1, 5, 10] which is easiest to describe in Fourier-analytic terms. Let $\{w_S\}$ be the Walsh–Fourier basis of the vector space of real-valued functions on the cube. For a function $f : \{0,1\}^n \to \mathbb{R}$, $f = \sum_{S\in\{0,1\}^n} \hat{f}(S)w_s$, and a non-negative real ϵ, let $T_\epsilon(f) = \sum_{S\in\{0,1\}^n} \epsilon^{|S|}\hat{f}(S)w_s$. Then

$$\|T_\epsilon f\|_2 \leqslant \|f\|_{1+\epsilon^2}.$$

Following its application in [16], this inequality, known (for historical reasons) as the Bonami–Beckner inequality, became a very important tool in combinatorics and theory of computer science. Still, it is very different from the familiar combinatorial tools, and its appearance in the proof is somewhat mysterious. Thus it seemed of interest to look for a combinatorial proof of Theorem 1.1, possibly along the lines of the above-mentioned proof of (1.1). Let us mention two papers dealing with this problem along very different routes. The first of these papers [9] gives a combinatorial (entropic) proof of the Bonami–Beckner inequality for $\epsilon = \sqrt{3}/3$. This special case is already strong enough to be instrumental in the proof of Theorem 1.1. The second paper [29] presents an inductive proof that the maximal influence of a balanced function is at least $\Omega\left(\frac{\log^\alpha(n)}{n}\right)$ for some $0 < \alpha < 1$.

In this paper we give a fully combinatorial proof of Theorem 1.1. After completing our work, we learned that a very similar proof was recently obtained by Rossignol [25].

We start with a functional form of (1.1). For a non-negative function $f : \{0,1\}^n \to \mathbb{R}$,

$$\mathbb{E}_x \sum_{y\sim x} (f(x) - f(y))^2 \geqslant 2 \cdot \mathbb{E}f^2 \log \frac{\mathbb{E}f^2}{\mathbb{E}^2 f}. \tag{1.3}$$

Functional forms of isoperimetric inequalities are widely used in local theory of Banach spaces [18]. They turn out to be useful in our setting too. We show Theorem 1.1 to be a simple consequence of inequality (1.3).

https://doi.org/10.1017/S0963548306008340 Published online by Cambridge University Press

This inequality can be proved by induction on dimension (see Section 5), as in (1.1), though the proof is somewhat more complicated. The isoperimetric constant $C_{1.3} = 2$ is tight, if we want it to be independent of the dimension. In Section 4 we give examples of functions satisfying (1.3) with equality if the constant 2 is replaced by $2 + o_n(1)$. These are symmetric functions (a function f on the cube is symmetric if $f(x)$ depends only on the distance of x from zero) closely related to a classical family of orthogonal polynomials of discrete variable: the Krawtchouk polynomials.

We also suggest a reason behind the relevance of the Bonami–Beckner inequality. It is well known [10] that this inequality is equivalent to the logarithmic Sobolev inequality

$$\mathbb{E}_x \sum_{y \sim x} (f(x) - f(y))^2 \geqslant 2 \cdot \mathrm{Ent}(f^2) = 2 \cdot (\mathbb{E} f^2 \log f^2 - \mathbb{E} f^2 \log \mathbb{E} f^2) \tag{1.4}$$

in the cube. It turns out that this inequality implies inequality (1.3). In this sense the Bonami–Beckner inequality can be thought of as a refined form of the edge-isoperimetric inequality in the discrete cube.

The actual result we prove seems to be somewhat stronger than Theorem 1.1.

Theorem 1.2. *If f is a boolean function with expectation μ, then*

$$\sum_{i=1}^{n} I_i^2(f) \geqslant 4\mu(1-\mu) \exp\left\{ - \frac{1/2}{\mu(1-\mu)} \sum_{i=1}^{n} I_i(f) \right\}. \tag{1.5}$$

This inequality implies (1.2) with a constant $c_{1.2} = 4$, recovering the estimate of [16].

Theorem 1.2 is a special case of our main result, Theorem 2.2, stated and proved in Section 2. This theorem presents a more general inequality valid for real-valued functions on the discrete cube endowed with an arbitrary measure. The theorem and the approach used in its proof seem to provide convenient tools for dealing with a certain type of isoperimetric statements in the cube. We illustrate this by giving simple proofs of two results from [7] and [8], with better isoperimetric constants.

It should be mentioned that inequality (1.5), in its turn, is implied, *up to a constant in the exponent*, by an inequality of Talagrand [27]. A special case of this inequality asserts that, for a boolean function f with expectation μ,

$$\sum_{i=1}^{n} \frac{I_i}{\log(e/I_i)} \geqslant \Omega\big(\mu(1-\mu)\big). \tag{1.6}$$

It is not hard to see that this gives (1.5) if $1/2$ in the exponent is replaced by a sufficiently large constant.

Next, we focus our attention on the best-possible constants for the above-mentioned inequalities. Specifically, we are interested in the exact constant $C_{1.5}$ that should appear in the exponent on the right-hand side of (1.5).

To be more specific, by Theorem 1.2, $C_{1.5} \leqslant 1/2$. On the other hand, taking f to be a characteristic function of a subcube of large co-dimension shows $C_{1.5} \geqslant \log 2/2$. We believe the lower bound to be the right one.

https://doi.org/10.1017/S0963548306008340 Published online by Cambridge University Press

Conjecture 1.3.

$$C_{1.5} = \frac{\log 2}{2}.$$

In particular, we conjecture small subcubes to be (nearly)-isoperimetric sets for this inequality. If Conjecture 1.3 holds, this would, in particular, give the optimal constant $C_{1.2} = 4/\log^2 2 \approx 8.3$ in inequality (1.2). It is easy to see that[2] $C_{1.2} \geqslant 1/C_{1.5}^2$. Therefore the conjecture would imply $C_{1.2} \geqslant 4/\log^2 2$. It turns out that the best currently known candidate to be an isoperimetric function for inequality (1.2), namely the 'tribes' function of Ben-Or and Linial [2], implies[3] $C_{1.2} \leqslant 4/\log^2 2$. We are grateful to Amites Sarkar [26] for pointing this out to us.

We conclude this section by saying a few words about a possible approach to the proof of Conjecture 1.3. We will say more about this in Section 3. The main step in the proof of Theorem 1.2 is a variant (3.2) of the logarithmic Sobolev inequality for the discrete cube. This inequality applies to general real-valued functions on the cube, and is tight with constant $C = 2$. To prove the conjecture we need to improve this constant to $2/\log 2$, taking into account the specific structure of boolean functions. The familiar approach using tensorization does not seem to be convenient for this. We give a proof of (3.2) which works by induction on the dimension, similar to the proof of (1.1), and seems to be more conducive for this purpose.

This paper is organized as follows. In the next section we prove the main claim, Theorem 2.2. In Section 3 several corollaries are derived from this theorem, and the main technical conjecture is stated. Section 4 constructs non-negative real-valued functions which are almost isoperimetric for inequality (2.1) and hence for several other inequalities in this paper, including (1.3). Inductive proofs of (1.3) and (3.2) are given in Sections 5 and 6, respectively.

2. The main theorem

We start with some definitions and notation.

Let $\mathscr{F}_j$, for $0 \leqslant j \leqslant n$, be the algebra of subsets of $\{0,1\}^n$ generated by the first j bits. More precisely, $\mathscr{F}_j$ is generated by the atoms $\{A_{\epsilon_1 \cdots \epsilon_j} : \epsilon_i \in \{0,1\}\}$ where $A_{\epsilon_1 \cdots \epsilon_j} = \{x : x_1 = \epsilon_1, \ldots, x_j = \epsilon_j\}$. Then $\{\mathscr{F}_j\}$ is an increasing sequence of algebras. In particular, $\mathscr{F}_0 = \{\emptyset, \{0,1\}^n\}$ and $\mathscr{F}_n = 2^{\{0,1\}^n}$.

For a function $f : \{0,1\}^n \to \mathbb{R}$, let $f_i = \mathbb{E}(f \mid \mathscr{F}_i)$ be the conditional expectation of f given the algebra $\mathscr{F}_i$. This means that $f_i(x)$ is the average of f over the points y that coincide with x in the first i coordinates. In particular, $f_0 = \mathbb{E}f$, $f_n = f$. The sequence $f_0, \ldots, f_n$ is a martingale with respect to $\{\mathscr{F}_j\}$.[4] Let d_i, $i = 1 \cdots n$ be the sequence of martingale differences; $d_i = f_i - f_{i-1}$, $1 \leqslant i \leqslant n$.

[2] More precisely, $C_{1.2} \geqslant 1/C_{1.5}^2 - o_n(1)$. Here and in the rest of this paper we ignore negligible factors when comparing constants.

[3] We choose the tribe size appropriately, so that the expectation is small.

[4] Essentially the only martingale property we use is the fact that conditional expectation is an orthogonal projection on a subspace.

https://doi.org/10.1017/S0963548306008340 Published online by Cambridge University Press

Let $\mathscr{E}(f,f) = \mathbb{E}_x \sum_{y\sim x}(f(x)-f(y))^2$. Let us mention that $\mathscr{E}(f,g) = E_x \sum_{y\sim x}(f(x)-f(y))(g(x)-g(y))$ is sometimes called the canonical Dirichlet form on $\{0,1\}^n$ [4].

The following lemma is simple and well known [22]. For completeness, we will give a proof at the end of this section.

Lemma 2.1.

$$\mathscr{E}(f,f) = \sum_{i=1}^{n} \mathscr{E}(d_i, d_i).$$

Let m be a measure on $\{0,1\}^n$. Let C be the best constant in the logarithmic Sobolev inequality for $\{0,1\}^n$ with m. That is to say, C is maximal such that, for any function $f : \{0,1\}^n \to \mathbb{R}$, we have $\mathscr{E}(f,f) \geqslant C \cdot \mathrm{Ent}(f^2)$.

For a function $f : \{0,1\}^n \to \mathbb{R}$, let $\sigma^2(f) = \mathbb{E}f^2 - \mathbb{E}^2 f$.

Our main result is as follows.

Theorem 2.2.

$$\sum_{i=1}^{n} \mathbb{E}^2|d_i| \geqslant \sigma^2(f) \exp\left\{-\frac{\mathscr{E}(f,f)}{C\sigma^2(f)}\right\}. \tag{2.1}$$

Proof. First we need a simple lemma.

Lemma 2.3. *For a non-negative function f on the discrete cube* $\mathrm{Ent}(f^2) \geqslant \mathbb{E}f^2 \log \frac{\mathbb{E}f^2}{\mathbb{E}^2 f}$.

Proof of Lemma 2.3. Since both sides of this inequality are 2-homogeneous, this amounts to showing $\mathbb{E}f^2 \log f^2 + \log \mathbb{E}^2(f) \geqslant 0$, given $\mathbb{E}f^2 = 1$. This is the same as $\log \mathbb{E}f - \mathbb{E}f^2 \log \frac{1}{f} \geqslant 0$; and this is true since logarithm is concave. □

Now we can conclude the proof of Theorem 2.2. Using (1.4) and Lemma 2.3,

$$\mathscr{E}(f,f) = \sum_{i=1}^{n} \mathscr{E}(d_i,d_i) \geqslant C \cdot \sum_{i=1}^{n} \mathrm{Ent}(d_i^2) \geqslant C \cdot \sum_{i=1}^{n} \mathbb{E}d_i^2 \log \frac{\mathbb{E}d_i^2}{\mathbb{E}^2|d_i|}.$$

Observe $\sum_{i=1}^{n} \mathbb{E}d_i^2 = \mathbb{E}f^2 - \mathbb{E}^2 f =: \sigma^2(f)$. By the convexity of $-\log x$, the last sum is

$$C\sigma^2(f) \sum_{i=1}^{n} \frac{\mathbb{E}d_i^2}{\sigma^2(f)} \log \frac{\mathbb{E}d_i^2}{\mathbb{E}^2|d_i|} \geqslant C\sigma^2(f) \log\left(\frac{\sigma^2(f)}{\sum_{i=1}^{n} \mathbb{E}^2|d_i|}\right). \qquad \square$$

We remark that, instead of the logarithmic Sobolev inequality (1.4), it is possible to use isoperimetric inequality (1.3) directly. Hence the logarithmic Sobolev constant C in this statement can be replaced by a potentially bigger isoperimetric constant C'.

Proof of Lemma 2.1. Let $1 \leqslant i \leqslant n$ be an index, and let g and h be two functions on $\{0,1\}^n$ such that $h(x) = (g(x) + g(x \oplus e_i))/2$. We will show

$$\mathscr{E}(g,g) = \mathscr{E}(h,h) + \mathscr{E}(g-h, g-h).$$

https://doi.org/10.1017/S0963548306008340 Published online by Cambridge University Press

Applying this identity successively to the functions $g = f = f_n$, $g = f_{n-1}, \ldots, g = f_1$, we obtain

$$\mathscr{E}(f,f) = \mathscr{E}(f_0,f_0) + \sum_{i=1}^{n} \mathscr{E}(d_i,d_i) = \sum_{i=1}^{n} \mathscr{E}(d_i,d_i),$$

completing the proof.

So it remains to verify the basic identity. For an index $1 \leqslant i \leqslant n$, and a function g, let g^i be a function defined by $g^i(x) = g(x \oplus e_i)$. Then

$$\begin{aligned}\mathscr{E}(g,g) &= \sum_{i=1}^{n} \|g - g^i\|_2^2 = \sum_{i=1}^{n} \|(h - h^i) + ((g - g^i) - (h - h^i))\|_2^2 \\ &= \sum_{i=1}^{n} \big(\|h - h^i\|_2^2 + \|(g - g^i) - (h - h^i)\|_2^2\big) \\ &= \sum_{i=1}^{n} \big(\|h - h^i\|_2^2 + \|(g - h) - (g - h)^i\|_2^2\big) = \mathscr{E}(h,h) + \mathscr{E}(g - h, g - h).\end{aligned}$$

The third equality follows from Pythagoras' theorem, since h and $g - h$ are orthogonal. □

3. Some corollaries for product measures

In this section we derive Theorem 1.1 and theorems of Friedgut and Friedgut–Kalai from Theorem 2.2. We also state our main technical conjecture, Conjecture 3.1.

We will assume the measure m to be a product probability measure, that is, $m = \otimes_{k=1}^{n} m_k$, with $m_k(1) = p_k$, and $m_k(0) = 1 - p_k$. In this case $\mathbb{E}|d_i|$ has a simple upper bound.

Lemma 3.1. *For a product measure m,*

$$\mathbb{E}|d_i| \leqslant 2p_i(1 - p_i) \cdot \mathbb{E}_x|f(x) - f(x \oplus e_i)|.$$

Proof. This lemma is based on an elementary observation that conditional expectation (or, in other words, averaging) decreases ℓ_1-norm.

Consider first the case of the uniform measure. As in the proof of the preceding lemma, let i be an index, and g and h two functions on $\{0,1\}^n$ with $h(x) = (g(x) + g(x \oplus e_i))/2$. Let a function g^i be defined by $g^i(x) = g(x \oplus e_i)$. Then $h = (g + g^i)/2$ and $\|g - h\|_1 = \|(g - g^i)/2\|_1 = 1/2 \cdot \|g - g^i\|_1$. Applying this argument to $g = f_i$ we have $\mathbb{E}|d_i| = \|d_i\|_1 = 1/2 \cdot \|f_i - f_i^i\|_1 \leqslant 1/2 \cdot \|f - f^i\|_1$. The last inequality uses the fact that $f_i - f_i^i$ is a conditional expectation of $f - f^i$.

The argument for the general case is similar. It is convenient to state it in the language of conditional expectations. We have

$$\mathbb{E}|d_i| = \mathbb{E}|f_i - f_{i-1}| = \mathbb{E}|\mathbb{E}(f \mid \mathscr{F}_i) - \mathbb{E}(f \mid \mathscr{F}_{i-1})|.$$

Let $\mathscr{G}_i$ be the algebra of subsets of $\{0,1\}^n$ generated by all the bits but j. That is, for x with $x_i = 0$, we have $\mathbb{E}(f \mid \mathscr{G}_i)(x) = \mathbb{E}(f \mid \mathscr{G}_i)(x \oplus e_i) = (1 - p_i)f(x) + p_i f(x \oplus e_i)$.

https://doi.org/10.1017/S0963548306008340 Published online by Cambridge University Press

Then, for a product measure m, we have $\mathbb{E}(\mathbb{E}(f \mid \mathscr{F}_i) \mid \mathscr{G}_i) = \mathbb{E}(f \mid \mathscr{F}_{i-1})$. Therefore

$$\begin{aligned}\mathbb{E}|\mathbb{E}(f \mid \mathscr{F}_i) - \mathbb{E}(f \mid \mathscr{F}_{i-1})| &= \mathbb{E}|\mathbb{E}\big(f - \mathbb{E}(f \mid \mathscr{G}_i) \mid \mathscr{F}_i\big)| \\ &\leqslant \mathbb{E}|f - \mathbb{E}(f \mid \mathscr{G}_i)| \\ &= 2p_i(1-p_i) \cdot \mathbb{E}_x|f(x) - f(x \oplus e_i)|\end{aligned}$$

□

3.1. Uniform measure

The best logarithmic Sobolev constant C for a cube endowed with the uniform measure is $C = 2$. Hence the theorem gives, for a real-valued function f on the discrete cube,

$$\sum_{i=1}^{n} \mathbb{E}^2|d_i| \geqslant \sigma^2(f) \exp\left\{-\frac{\mathscr{E}(f,f)}{2\sigma^2(f)}\right\}.$$

This may be somewhat simplified for monotone functions, for which $\mathbb{E}|d_i| = \hat{f}(\{i\})$, in which case

$$\sum_{i=1}^{n} \hat{f}^2(\{i\}) \geqslant \sigma^2 \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} I_i(f)\right\}.$$

Section 4 presents a construction of monotone functions for which this inequality is essentially tight.

Our main concerns are boolean functions. We present several easy implications of Theorem 2.2 and a related conjecture.

Proof of Theorem 1.2. For a boolean function f we have $\mathbb{E}_x|f(x) - f(x \oplus e_i)| = I_i$, and using Theorem 2.2 together with Lemma 3.1 yields

$$\sum_{i=1}^{n} I_i^2(f) \geqslant 4\sigma^2(f) \exp\left\{-\frac{1}{2\sigma^2(f)}\sum_{i=1}^{n} I_i(f)\right\},$$

proving (1.5) and Theorem 1.2. □

Example. Let f be the characteristic function of a subcube of dimension $n-t$. Then f has t non-zero influences of size 2^{-t+1}. Assume t is large enough so that $\mu = 2^{-t}$ may be replaced with 1, and $\sigma^2 = \frac{2^t-1}{4^t}$ may be replaced with 2^{-t}. Then (1.5) gives

$$\frac{4t}{4^t} \geqslant \frac{4}{2^t}\exp\left\{-2^{t-1} \cdot \frac{2t}{2^t}\right\} = \frac{1}{2^t}\exp\{-t\} \quad \text{or} \quad t2^{-t} \geqslant e^{-t}.$$

We suggest (Conjecture 1.3) that for boolean functions we have a stronger inequality,

$$\sum_{i=1}^{n} I_i^2(f) \geqslant 4\sigma^2(f) \exp\left\{-\frac{\log 2}{2\sigma^2(f)}\sum_{i=1}^{n} I_i(f)\right\}.$$

Such an inequality would be tight by the example above.

This inequality would follow from the following version of logarithmic Sobolev inequality for boolean functions.

https://doi.org/10.1017/S0963548306008340 Published online by Cambridge University Press

Conjecture 3.2. *For a boolean function f on the discrete cube,*

$$\mathscr{E}(f,f) \geqslant \frac{2}{\log 2} \cdot \sum_{i=1}^{n} \mathrm{Ent}(d_i^2). \tag{3.1}$$

Discussion. This inequality holds with a constant $C = 2$ for real-valued functions on $\{0,1\}^n$, that is,

$$\mathscr{E}(f,f) \geqslant 2 \cdot \sum_{i=1}^{n} \mathrm{Ent}(d_i^2). \tag{3.2}$$

The inequality is proved by applying logarithmic Sobolev inequality to functions d_i, $1 \leqslant i \leqslant n$. It is almost tight, as shown by functions constructed in Section 4. To improve the constant for boolean functions, a different approach seems to be required, one that 'remembers' that $\{d_i\}$ are difference functions of a martingale defined by a boolean function. In particular, the familiar tensorization approach might not be sufficient here since it does not keep track of the combinatorial structure of the functions involved.

We give an inductive proof of inequality (3.2) in Section 6. This proof seems to be better suited for handling functions f with a specific structure, such as boolean functions.

Proof of Theorem 1.1. We show

$$\sum_{i=1}^{n} I_i^2(f) \geqslant \big(4 - o_n(1)\big) \cdot \frac{\mu^2(1-\mu)^2 \log^2 n}{n}.$$

For a boolean function f with expectation μ we have $\sigma^2(f) = \mu(1-\mu)$. Let $\epsilon(n) \gg \frac{\log\log n}{\log n}$. There are two cases to consider. First,

$$\sum_{i=1}^{n} I_i < (2 - \epsilon(n)) \cdot \mu(1-\mu)\frac{\log n}{n}.$$

In this case (1.5) implies $\sum_{i=1}^{n} I_i^2(f) \gg \mu(1-\mu)\frac{\log^2 n}{n}$. The second case is

$$\sum_{i=1}^{n} I_i \geqslant (2 - \epsilon(n)) \cdot \mu(1-\mu)\frac{\log n}{n}.$$

The Cauchy–Schwarz inequality now implies $\sum_{i=1}^{n} I_i^2 \geqslant (2-\epsilon(n))^2 \cdot \mu^2(1-\mu)^2 \frac{\log^2 n}{n}$. □

Proof of a theorem of Friedgut

Theorem 3.3. ([7]) *For a boolean function f and an arbitrary $\epsilon > 0$, there is a function g depending only on*

$$\left(\sum_{i=1}^{n} I_i\right) \cdot \exp\left\{\frac{\sum_{i=1}^{n} I_i}{(2 - o_\epsilon(1))\epsilon}\right\}$$

coordinates[5] *(a junta) such that* $\|f - g\|_2^2 \leqslant \epsilon$.

[5] The original proof in [7] has a somewhat larger estimate for the required number of variables in the junta. The constant in the exponent is 2 rather than 1/2.

https://doi.org/10.1017/S0963548306008340 Published online by Cambridge University Press

Proof. Let $K = \sum_{i=1}^{n} I_i$ and take $\alpha = \exp\{-\frac{K}{(2-o_\epsilon(1))\epsilon}\}$. The error term $o_\epsilon(1)$ will be chosen later.

Without loss of generality, assume the influences I_i to decrease with i, and let r be the maximal index with $I_r \geqslant \alpha$. Clearly $r \leqslant \frac{K}{\alpha}$ and $\sum_{i=r+1}^{n} I_i^2 \leqslant K\alpha$. Take $g = \mathbb{E}(f \mid \mathscr{F}_r)$. The function g depends only on r variables. We will show $\|f-g\|_2^2 \leqslant \epsilon$.

Let $h = f - g$. Take $h_i = \mathbb{E}(h|\mathscr{F}_i)$, $i = 1 \cdots n$, to be the martingale defined by h, and let $d_i(h)$ be its difference functions. Then $d_i(h) = d_i(f)$ for $i > r$ and $d_i(h) = 0$ otherwise. Note that $\mathbb{E}h = 0$ and therefore $\sigma^2(h) = \|h\|_2^2$. By Theorem 2.2,

$$K = \sum_{i=1}^{n} I_i = \mathscr{E}(f,f) \geqslant \mathscr{E}(h,h) \geqslant 2\|h\|_2^2 \log\left(\frac{\|h\|_2^2}{\sum_{i=r+1}^{n} I_i^2}\right) \geqslant 2\|h\|_2^2 \log\left(\frac{\|h\|_2^2}{K\alpha}\right).$$

Recalling the definition of α, it is now easy to choose the error term $o_\epsilon(1)$ appropriately, so that the last inequality implies $\|h\|_2^2 \leqslant \epsilon$. □

3.2. The measure m_p

Let m_p be a product distribution, $m_p = \otimes_{k=1}^{n} m$, with $m(1) = p$, $m(0) = 1-p$. Assume $p \leqslant 1/2$. The best constant C in the logarithmic Sobolev inequality in this case is known [6] to be

$$C(p) = \frac{1-2p}{p(1-p)} \cdot \frac{1}{\log(1-p) - \log p}.$$

Proof of a theorem of Friedgut and Kalai

Theorem 3.4. ([8]) *Let f be a boolean function with expectation μ on $\{0,1\}^n$ endowed with the measure m_p. Assume that $1 \gg p \geqslant n^{-o_n(1)}$. Then there is a variable whose influence on f at least*

$$\Omega\left(\frac{\mu(1-\mu)}{p\log\frac{1}{p}} \cdot \frac{\log n}{n}\right).$$

Proof. For a boolean function f, Theorem 2.2 gives

$$\sum_{i=1}^{n} I_i^2(f) \geqslant \frac{\sigma^2(f)}{4p^2(1-p)^2} \exp\left\{-\frac{1}{C(p)\sigma^2(f)} \sum_{i=1}^{n} I_i(f)\right\}.$$

The expression on the right-hand side is somewhat complicated. It simplifies for $p \ll 1$, for which $C(p) \approx \frac{1}{p\log\frac{1}{p}}$, and we get (ignoring negligible errors)

$$\sum_{i=1}^{n} I_i^2(f) \geqslant \frac{\sigma^2(f)}{4p^2} \exp\left\{-\frac{p\log\frac{1}{p}}{\sigma^2(f)} \sum_{i=1}^{n} I_i(f)\right\}.$$

Proceeding as in the proof of Theorem 1.1, we get that

$$\sum_{i=1}^{n} I_i^2(f) \geqslant \frac{\mu^2(1-\mu)^2}{p^2\log^2\frac{1}{p}} \cdot \frac{\log^2 n}{n}. \tag{3.3}$$

https://doi.org/10.1017/S0963548306008340 Published online by Cambridge University Press

In particular, there is a variable i with influence at least

$$I_i \geqslant \frac{\mu(1-\mu)}{p \log \frac{1}{p}} \cdot \frac{\log n}{n}. \tag{3.4}$$

□

We remark that this proof provides (3.4) with an explicit constant 1, and does not rely on the assumption $p \geqslant n^{-o(1)}$.

4. Construction of almost isoperimetric functions

In this section we construct non-negative functions on the cube endowed with uniform measure, for which inequality (2.1) is almost tight. That is, we show the constant $C = 2$ to be optimal in this case. This directly implies that these functions are almost isoperimetric for inequalities (1.3), (1.5), and (3.2).

These functions were constructed in [24] (for a different purpose). Here we will repeat parts of this construction for completeness. Before this, let us make several observations.

Contrary to the Euclidean case, the edge-isoperimetric set in $\{0,1\}^n$ is a subcube rather than a Hamming ball. This is explained by the 'non-spherical' geometry of $\{0,1\}^n$, which has n distinguished directions. It turns out that if we expand the problem and consider an isoperimetric inequality (1.3) for arbitrary real-valued functions, there is a nearly optimal solution which is spherically invariant, that is, invariant under permutations of coordinates.

Definition. A function f on $\{0,1\}^n$ is *symmetric* if, for any permutation $\tau \in S_n$, and for any $x = (x_1 \cdots x_n) \in \{0,1\}^n$,

$$f(x) = f(x_1 \cdots x_n) = f(x_{\tau(1)\cdots\tau(n)}).$$

The functions we construct are going to be symmetric. Note that the value of a symmetric function at a point depends only on the distance of this point from zero. Therefore there is a natural one-to-one correspondence between symmetric functions on $\{0,1\}^n$ and functions on $\{0 \cdots n\}$. We will abuse notation and speak about both $f(x)$, $x \in \{0,1\}^n$, and $f(k)$, where k is the Hamming weight of x, $0 \leqslant k \leqslant n$.

We choose an index s such that $\log n \ll s \ll n$. The first step is to construct an auxiliary symmetric function k_s, which is fully defined by its values $k_s(0), \ldots, k_s(n)$ at distances $0 \cdots n$. Set (formally) $k_s(-1) = 0$ and $k_s(0) = 1$, and define $k_s(r)$ for $1 \leqslant r \leqslant n$ by the relation $(n-r+1) \cdot k_s(r) = (n-2s) \cdot k_s(r-1) - (r-1) \cdot k_s(r-2)$.

The function $k_s(r)$ coincides on integer points $r = 0 \cdots n$ with a normalized Krawtchouk polynomial

$$K_s(x) = \frac{1}{\binom{n}{s}} \cdot \sum_{k=0}^{s} (-1)^k \binom{x}{k} \binom{n-x}{s-k}.$$

This will be useful, since much is known about these polynomials (see [19]).

In particular, Krawtchouk polynomials $\{K_s\}_{s=0}^n$ are a family of polynomials orthogonal with respect to a measure supported on $0 \cdots n$. Hence their roots are simple and are

https://doi.org/10.1017/S0963548306008340 Published online by Cambridge University Press

located in the interval $(0,n)$ [30]. Let x_s be the first root of K_s. We now define $f = f_s$ to be a symmetric function on $\{0,1\}^n$ defined by $f(x) = k_s(x)$ for points whose distance from zero is at most x_s, and $f_s(x) = 0$ otherwise.

We will need an estimate [19] for location of the first root: $x_s \leqslant \frac{n}{2} - (1-o(1)) \cdot \sqrt{sn}$. Let $m = \lfloor x_s \rfloor$, and let $\mathrm{supp}(f)$ denote the support of f. Standard estimates imply that

$$|\mathrm{supp}(f)| \leqslant \sum_{k=0}^{m} \binom{n}{k} \leqslant 2^{nH(1/2-(1-o(1))\cdot\sqrt{s/n})} \leqslant 2^n \cdot e^{-(1-o(1))\cdot 2s}.$$

By the Cauchy–Schwarz inequality,

$$\frac{\mathbb{E}^2 f}{\mathbb{E} f^2} \leqslant \frac{|\mathrm{supp}(f)|}{2^n} \leqslant e^{-(1-o(1))\cdot 2s} \leqslant o(1).$$

That is, $\sigma^2(f) = \mathbb{E} f^2 - \mathbb{E}^2 f \geqslant (1-o(1)) \cdot \mathbb{E} f^2$.

Let $C = 2$ be the appropriate constant in (2.1).

We start by estimating the right-hand side of (2.1) from below. For $x \in \{0,1\}^n$, let

$$N(x) = \sum_{y:\ y \sim x} f(y).$$

Then it is not hard to see [24, Lemma 3.4] that $N(x) \geqslant (n-2s) f(x)$ and therefore

$$\mathscr{E}(f,f) = \mathbb{E}_x \sum_{y \sim x} (f(x) - f(y))^2 = 2n\mathbb{E} f^2 - 2\langle f, N \rangle \leqslant 4s\mathbb{E} f^2.$$

This means that the right-hand side of (2.1) can be estimated from below by $\mathbb{E} f^2 \cdot e^{-(1+o(1))\cdot 2s}$.

Now to the left-hand side. Here it will be convenient to use simple facts from Fourier analysis on $\{0,1\}^n$. We refer to [16] for more details on the subject. It is easy to see from the recursion (see also [19]) that the function f is monotone. Therefore $\mathbb{E}|d_i| = \hat{f}(\{i\})$. Since f is symmetric, so is its Fourier transform, and these Fourier coefficients have the same value for all indices $1 \leqslant i \leqslant n$. Let I denote this common value. We proceed to upper-bound I. Let $f(k)$ denote the value of f in points at distance k from zero. Then

$$nI = \sum_{i=1}^{n} \hat{f}(\{i\}) = \frac{1}{2^n} \sum_{k=0}^{n} \binom{n}{k} (n-2k) f(k) = \frac{1}{2^n} \sum_{k=0}^{m} \binom{n}{k} (n-2k) f(k) \leqslant \frac{n}{2^n} \sum_{k=0}^{m} \binom{n}{k} f(k).$$

Since $\mathbb{E} f^2 = \frac{1}{2^n} \sum_{k=0}^{n} \binom{n}{k} f^2(k)$, we have

$$f(k) \leqslant \sqrt{\frac{2^n \mathbb{E} f^2}{\binom{n}{k}}},$$

and so

$$I \leqslant \frac{1}{2^n} \sum_{k=0}^{m} \binom{n}{k} f(k) \leqslant \sqrt{\frac{\mathbb{E} f^2}{2^n}} \cdot \sum_{x=0}^{m} \sqrt{\binom{n}{x}} \leqslant n \sqrt{\frac{\binom{n}{m} \mathbb{E} f^2}{2^n}}.$$

Therefore the left-hand side of (2.1) is

$$nI^2 \leqslant \mathbb{E} f^2 \cdot n^3 \frac{\binom{n}{m}}{2^n} \leqslant \mathbb{E} f^2 \cdot n^3 2^{n(H(m/n)-1)} \leqslant \mathbb{E} f^2 \cdot e^{-(1-o(1))\cdot 2s}.$$

https://doi.org/10.1017/S0963548306008340 Published online by Cambridge University Press

Combining the estimates for both sides, we see that $C = 2$ is indeed the best-possible constant in (2.1).

5. An inductive proof of an isoperimetric inequality

In this section we give an inductive proof of the inequality (1.3)

$$\mathscr{E}(f,f) \geqslant 2 \cdot \mathbb{E}f^2 \log \frac{\mathbb{E}f^2}{\mathbb{E}^2 f}$$

for a real non-negative function $f : \{0,1\}^n \to \mathbb{R}$. By homogeneity, we may and will assume $\mathbb{E}f = 1$.

The proof is by induction on the dimension n.

For $n = 1$, let $f(0) = a$, and $f(1) = 2 - a$. Then $\mathbb{E}f^2 = \frac{1}{2}(a^2 + (2-a)^2) = 1 + (1-a)^2$; and $\mathbb{E}_x \sum_{y \sim x}(f(x) - f(y))^2 = (2-2a)^2 = 4(1-a)^2$. Let $x = (1-a)^2$. It remains to verify that $2x \geqslant (1+x)\log(1+x)$ for $0 \leqslant x \leqslant 1$, which is easily seen to be true. In fact a stronger inequality $2x \geqslant (1+x)\log_2(1+x)$ is also valid in this interval, since the right-hand side is a convex function which is 0 at zero and 2 at one.

Assume the inequality to hold for $n-1$. Let f_0 and f_1 be the restrictions of f to $(n-1)$-dimensional half-cubes determined by the value of the nth coordinate. Let μ_i be the expectations of f_i, and v_i the second moments of f_i for $i = 0, 1$.

Then

$$\mathscr{E}(f,f) = \frac{1}{2} \cdot \big(\mathscr{E}(f_0,f_0) + \mathscr{E}(f_1,f_1)\big) + \|f_0 - f_1\|_2^2. \tag{5.1}$$

Note that the expectations and the distance in this formula are computed on $(n-1)$-dimensional cubes.

By the induction hypothesis we can lower-bound the first summand by

$$v_0 \log \frac{v_0}{\mu_0^2} + v_1 \log \frac{v_1}{\mu_1^2}.$$

For the second summand we need a simple lemma.

Lemma 5.1. *Let f_0 and f_1 be two functions with expectations μ_0, μ_1 and variances $\sigma_0^2 = v_0 - \mu_0^2$, $\sigma_1^2 = v_1 - \mu_1^2$. Then*

$$\|f_0 - f_1\|^2 \geqslant (\sigma_0 - \sigma_1)^2 + (\mu_0 - \mu_1)^2.$$

Proof. Let $g_i = f_i - \mu_i$, $i = 0, 1$. Then $\mathbb{E}g_i = 0$ and therefore

$$\begin{aligned} \|f_0 - f_1\|^2 &= \langle g_0 - g_1, g_0 - g_1\rangle + (\mu_0 - \mu_1)^2 \\ &= \|g_0 - g_1\|^2 + (\mu_0 - \mu_1)^2 \\ &\geqslant (\sigma_0 - \sigma_1)^2 + (\mu_0 - \mu_1)^2. \end{aligned}$$

□

Going back, and substituting in (5.1),

$$\mathscr{E}(f,f) \geqslant v_0 \log\left(\frac{v_0}{\mu_0^2}\right) + v_1 \log\left(\frac{v_1}{\mu_1^2}\right) + \left[v_0 + v_1 - 2\sqrt{v_0 - \mu_0^2}\sqrt{v_1 - \mu_1^2} - 2\mu_0\mu_1\right].$$

https://doi.org/10.1017/S0963548306008340 Published online by Cambridge University Press

So it suffices to show that under the assumptions

(1) $\mu_0, \mu_1, v_0, v_1 \geqslant 0$,
(2) $\frac{\mu_0+\mu_1}{2} = 1$, and
(3) $\frac{v_0+v_1}{2} = v := \mathbb{E}f^2$,

we have

$$v_0 \log\left(\frac{v_0}{\mu_0^2}\right) + v_1 \log\left(\frac{v_1}{\mu_1^2}\right) + \left[v_0 + v_1 - 2\sqrt{v_0 - \mu_0^2}\sqrt{v_1 - \mu_1^2} - 2\mu_0\mu_1\right] \geqslant 2v \log v. \quad (5.2)$$

The next few steps swap variables to simplify this expression.

Take $t = \frac{\mu_0-\mu_1}{2}$. Then $\mu_0 = 1 + t$ and $\mu_1 = 1 - t$. Similarly, take $v_0 = v(1 + y)$ and $v_1 = v(1 - y)$. Note that $-1 \leqslant t, y \leqslant 1$.

Substituting in (2.2), and dividing out by $2v$, it needs to be seen that

$$\frac{1+y}{2}\log\left(\frac{1+y}{(1+t)^2}\cdot v\right) + \frac{1-y}{2}\log\left(\frac{1-y}{(1-t)^2}\cdot v\right) + 1$$
$$\geqslant \log v + \frac{1}{v}\cdot\left[\sqrt{v(1+y)-(1+t)^2}\sqrt{v(1-y)-(1-t)^2} + (1-t^2)\right],$$

or

$$\frac{1+y}{2}\log\left(\frac{1+y}{(1+t)^2}\right) + \frac{1-y}{2}\log\left(\frac{1-y}{(1-t)^2}\right) + 1$$
$$\geqslant \frac{1}{v}\cdot\left[\sqrt{v(1+y)-(1+t)^2}\sqrt{v(1-y)-(1-t)^2} + (1-t^2)\right].$$

We first take on the right-hand side and show it to be at most $\sqrt{1-y^2}$. Indeed, it suffices to show

$$\sqrt{v^2(1-y^2) - v(1+y)(1-t)^2 - v(1-y)(1+t)^2 + (1-t^2)^2} \leqslant v\sqrt{1-y^2} - (1-t^2).$$

Note that the right-hand side is non-negative, since going back to the definitions of t and y, this is $\sqrt{v_0 v_1} - \mu_0\mu_1$. Squaring both expressions, and rearranging, we get to

$$(1+y)(1-t)^2 + (1-y)(1+t)^2 \geqslant 2\sqrt{1-y^2}(1-t^2).$$

This inequality is a special case of the arithmetic–geometric inequality.

Now to the left-hand side. Let $H(x) = -x\log x - (1-x)\log(1-x)$ be the (natural) entropy function. For $0 \leqslant p, q \leqslant 1$ let $D(p||q) = p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}$ denote the divergence between two-point distributions $(p, 1-p)$ and $(q, 1-q)$. It is well known (and is a simple consequence of the concavity of the logarithm) that the divergence is non-negative. Now,

$$1 + \left[\frac{1+y}{2}\log\left(\frac{(1+y)^2}{(1+t)^2}\right) + \frac{1-y}{2}\log\left(\frac{(1-y)^2}{(1-t)^2}\right)\right]$$
$$- \left[\frac{1+y}{2}\log(1+y) + \frac{1-y}{2}\log(1-y)\right]$$
$$= 1 + 2\left[\frac{1+y}{2}\log\left(\frac{(1+y)}{(1+t)}\right) + \frac{1-y}{2}\log\left(\frac{(1-y)}{(1-t)}\right)\right] + H\left(\frac{1-y}{2}\right) - \log 2$$
$$= 2D\left(\frac{1-y}{2}\middle\|\frac{1-t}{2}\right) + H\left(\frac{1-y}{2}\right) + (1-\log 2) \geqslant H\left(\frac{1-y}{2}\right) + (1-\log 2).$$

https://doi.org/10.1017/S0963548306008340 Published online by Cambridge University Press

Therefore we need to show that

$$1 - \sqrt{1-y^2} \geqslant \log 2 - H\left(\frac{1-y}{2}\right)$$

for all $-1 \leqslant y \leqslant 1$.

We will need two facts. First, the function $\phi(t) = \frac{1}{2} - \sqrt{t(1-t)}$ is an involution on $[0, \frac{1}{2}]$. Second, we have the following result.

Lemma 5.2. *The function $R(x) = H(\phi(x))$ is convex on $[0, \frac{1}{2}]$.*

Since $R(0) = \log 2$ and $R'(0) = -2$, by convexity $R(x) \geqslant \log 2 - 2x$ for $x \in [0, \frac{1}{2}]$. Rearranging and taking $x = \phi(z)$,

$$2\phi(z) \geqslant \log 2 - H(z).$$

Substituting $y = 1 - 2z$,

$$1 - \sqrt{1-y^2} \geqslant \log 2 - H\left(\frac{1-y}{2}\right),$$

and we are done.

It remains to prove Lemma 5.2.

Proof. We will show $R''(x) \geqslant 0$ on $(0, \frac{1}{2})$. Direct computation of the second derivative shows that this is the same as

$$\phi(1-\phi)\log\frac{1-\phi}{\phi} \geqslant \frac{(\phi')^2}{\phi''}.$$

The function ϕ satisfies

$$\phi' = \frac{\sqrt{1-(1-2\phi)^2}}{1-2\phi},$$

and therefore, after some simplifications,

$$\phi'' = -(2\phi') \cdot \frac{1}{(1-2\phi)^2\sqrt{1-(1-2\phi)^2}}.$$

Hence

$$\frac{(\phi')^2}{\phi''} = (1-(1-2\phi)^2)\left(\frac{1}{2} - \phi\right).$$

Since $1-(1-2\phi)^2 = 4\phi(1-\phi)$, we are left with verifying $\log\frac{1-\phi}{\phi} \geqslant 2(1-2\phi)$, for $\phi \in (0, \frac{1}{2})$.

Substituting $z = 1 - 2\phi$, this is the same as showing

$$\log\frac{1+z}{1-z} \geqslant 2z$$

for $0 < z < 1$. Since both of these functions are 0 at zero, it is enough to compare their derivatives. The derivative of right-hand side is 2, and the derivative of the left-hand side is $\frac{2}{1-z^2} > 2$, completing the proof. □

https://doi.org/10.1017/S0963548306008340 Published online by Cambridge University Press

6. An inductive proof of a logarithmic Sobolev inequality

In this section we give an inductive proof of the inequality (3.2)

$$\mathscr{E}(f,f) \geqslant 2 \cdot \sum_{i=1}^{n} \mathrm{Ent}(d_i^2),$$

for a real function $f : \{0,1\} \to \mathbb{R}$, with d_i the difference functions of f (*cf.* Section 2).

Observe that the right-hand side might depend on the ordering of coordinates.

The proof is by induction on n. For $n = 1$, $d_1 = f - \mathbb{E}f$, and therefore d_1^2 is a constant function with zero entropy. The claim follows. Assume the claim for $n-1$, and consider it for n.

Let the influence $I_i(f)$ of the ith bit on a real-valued function f to be given by $\mathbb{E}_x(f(x) - f(x \oplus e_i))^2$. Thus $\mathscr{E}(f,f) = \sum_{i=1}^{n} I_i$.

Let f_0, f_1 be the restrictions of f to subcubes defined by the value of the nth coordinate. We write $f \leftrightarrow (f_0, f_1)$. These are functions on $n-1$ variables. Let their influences, their conditional expectations, and their difference functions with respect to the natural ordering $1 \cdots n-1$ of the coordinates be denoted by $I_{i,0}, I_{i,1}, f_{i,0}, f_{i,1}, d_{i,0}, d_{i,1}$ correspondingly. Then, by the induction hypothesis,

$$\sum_{i=1}^{n} I_i = \frac{1}{2}\left(\sum_{i=1}^{n-1} I_{i,0} + \sum_{i=1}^{n-1} I_{i,1}\right) + I_n \geqslant \sum_{i=1}^{n-1} \mathrm{Ent}(d_{i,0}^2) + \sum_{i=1}^{n-1} \mathrm{Ent}(d_{i,1}^2) + I_n.$$

Now consider a slightly different ordering of the coordinates for f, which is $n, 1, 2, \ldots, n-1$. Let f_i and d_i be the conditional expectations and the difference functions in this new ordering. We will show

$$\sum_{i=1}^{n-1} \mathrm{Ent}(d_{i,0}^2) + \sum_{i=1}^{n-1} \mathrm{Ent}(d_{i,1}^2) + I_n \geqslant 2 \cdot \sum_{i=1}^{n} \mathrm{Ent}(d_i^2).$$

This will prove the inequality for the ordering $n, 1, 2, \ldots, n-1$ of the coordinates.[6]

Observe $f_i \leftrightarrow (f_{i-1,0},\ f_{i-1,1})$ for $i = 2 \cdots n$, and similarly for difference functions. Note also that d_1^2 is constant, and therefore has zero entropy.

Lemma 6.1. *Let $k \geqslant 0$, $k \leftrightarrow (g,h)$. Then*

$$\mathrm{Ent}(k) = \frac{1}{2}(\mathrm{Ent}(g) + \mathrm{Ent}(h)) + \frac{1}{2}\left(\mathbb{E}g \log \mathbb{E}g + \mathbb{E}h \log \mathbb{E}h - (\mathbb{E}g + \mathbb{E}h) \log \frac{\mathbb{E}g + \mathbb{E}h}{2}\right).$$

Proof. By inspection. □

[6] Alternatively, we could have insisted on the 'natural' order of the coordinates for f, and changed the order of coordinates for f_0, f_1.

https://doi.org/10.1017/S0963548306008340 Published online by Cambridge University Press

Therefore, we need to show

$$\frac{1}{2}\cdot I_n \geqslant \sum_{i=2}^{n}\left(\mathrm{Ent}(d_i^2)-\frac{1}{2}\left(\mathrm{Ent}(d_{i-1,0}^2)+\mathrm{Ent}(d_{i-1,1}^2)\right)\right)$$
$$=\frac{1}{2}\sum_{i=1}^{n-1}\left(\mathbb{E}d_{i,0}^2\log\mathbb{E}d_{i,0}^2+\mathbb{E}d_{i,1}^2\log\mathbb{E}d_{i,1}^2-(\mathbb{E}d_{i,0}^2+\mathbb{E}d_{i,1}^2)\log\frac{\mathbb{E}d_{i,0}^2+\mathbb{E}d_{i,1}^2}{2}\right).$$

For this purpose we consider the joint behaviour of the sequences $\mathbb{E}d_{i,0}^2$ and $\mathbb{E}d_{i,1}^2$.

Applying the Cauchy–Schwarz inequality twice,

$$\begin{aligned}(\mathbb{E}d_{i,0}^2-\mathbb{E}d_{i,1}^2)^2&=\mathbb{E}^2(d_{i,0}-d_{i,1})(d_{i,0}+d_{i,1})\leqslant\mathbb{E}^2|d_{i,0}-d_{i,1}||d_{i,0}+d_{i,1}|\\&\leqslant\mathbb{E}(d_{i,0}-d_{i,1})^2\mathbb{E}(d_{i,0}+d_{i,1})^2\\&\leqslant\left(\mathbb{E}d_{i,0}^2+\mathbb{E}d_{i,1}^2+2\sqrt{\mathbb{E}d_{i,0}^2\mathbb{E}d_{i,1}^2}\right)\cdot\mathbb{E}(d_{i,0}-d_{i,1})^2.\end{aligned}$$

Observe that $d_{i,0}-d_{i,1}$ is a difference sequence for f_0-f_1 and therefore

$$\sum_{i=1}^{n-1}\mathbb{E}(d_{i,0}-d_{i,1})^2\leqslant(f_0-f_1)^2=I_n.$$

Take $a_i:=\mathbb{E}d_{i,0}^2$, $b_i:=\mathbb{E}d_{i,1}^2$, and consider the following optimization problem:

$$\text{maximize}\quad\sum_{i=1}^{n-1}\left(a_i\log a_i+b_i\log b_i-(a_i+b_i)\log\frac{a_i+b_i}{2}\right),$$
$$\text{given}\quad\sum_{i=1}^{n-1}\frac{(a_i-b_i)^2}{a_i+b_i+2\sqrt{a_ib_i}}\leqslant I_n,\quad a_i,b_i\geqslant 0.$$

Let $m_i=\min\{a_i,b_i\}$ and $c_i=|a_i-b_i|$. Then an equivalent formulation is:

$$\text{maximize}\quad\sum_{i=1}^{n-1}\left((m_i+c_i)\log(m_i+c_i)+m_i\log m_i-(2m_i+c_i)\log\frac{2m_i+c_i}{2}\right),$$
$$\text{given}\quad\sum_{i=1}^{n-1}\frac{c_i^2}{2m_i+c_i+2\sqrt{m_i(m_i+c_i)}}\leqslant I_n,\quad m_i,c_i\geqslant 0.$$

Assume $c_i>0$ for all i, since removing coordinates with $c_i=0$ does not affect either the target function or the constraint. Therefore, we are allowed to consider $r_i=\frac{m_i}{c_i}$, leading to the following formulation:

$$\text{maximize}\quad\sum_{i=1}^{n-1}c_i\cdot\left((1+r_i)\log(1+r_i)+r_i\log r_i-(1+2r_i)\log\frac{1+2r_i}{2}\right),$$
$$\text{given}\quad\sum_{i=1}^{n-1}\frac{c_i}{1+2r_i+2\sqrt{r_i(1+r_i)}}\leqslant I_n,\quad r_i,c_i\geqslant 0.$$

The following technical claim completes the analysis.

https://doi.org/10.1017/S0963548306008340 Published online by Cambridge University Press

Lemma 6.2. *For any $r \geqslant 0$,*

$$(1+r)\log(1+r) + r\log r - (1+2r)\log\frac{1+2r}{2} \leqslant \frac{1}{1+2r+2\sqrt{r(1+r)}}.$$

In addition,

$$\left(1+2r+2\sqrt{r(1+r)}\right)\cdot\left((1+r)\log(1+r) + r\log r - (1+2r)\log\frac{1+2r}{2}\right) \to_{r\to\infty} 1.$$

Therefore, the supremum of the above maximization problem is bounded by I_n (and it can easily be seen that it actually equals I_n), completing the proof.

Proof of Lemma 6.2. Let $g(r) = 1+2r+2\sqrt{r(1+r)} = (\sqrt{r}+\sqrt{1+r})^2$, and $h(r) = (1+r)\log(1+r) + r\log r - (1+2r)\log\frac{1+2r}{2}$. We want to show $(gh)(r) \leqslant 1$, for all $r \geqslant 0$. At zero, $(gh)(0) = \log 2 < 1$, at infinity, $g(r) \sim 4r$ and $h(r) \sim \frac{1}{4r}$, and thus $(gh)(r) \to_{r\to\infty} 1$, proving the second part of the lemma. Thus it is sufficient to show gh is increasing, or $\frac{g'h}{g} \geqslant -h'$.

Computing, we have

$$\frac{g'h}{g} = \frac{h}{\sqrt{r(1+r)}} \quad \text{and} \quad -h' = \log\frac{(1+2r)^2}{4r(1+r)}.$$

It remains to show

$$h(r) \geqslant \sqrt{r(1+r)}\cdot\log\frac{(1+2r)^2}{4r(1+r)}.$$

Rewriting $h(r)$ as

$$\log\frac{2+2r}{1+2r} - r\log\frac{(1+2r)^2}{4r(1+r)},$$

this is the same as

$$\log\frac{2+2r}{1+2r} \geqslant \left(\sqrt{r(1+r)}+r\right)\cdot\log\frac{(1+2r)^2}{4r(1+r)} = \frac{r}{\sqrt{r(1+r)}-r}\cdot\log\frac{(1+2r)^2}{4r(1+r)},$$

that is,

$$\left(\sqrt{\frac{1+r}{r}}-1\right)\cdot\log\frac{2+2r}{1+2r} \geqslant \log\frac{(1+2r)^2}{4r(1+r)}.$$

Let $t = \sqrt{\frac{1+r}{r}}$. Then $t \in (1,\infty)$, and, rewriting in terms of t, we want to have

$$(t-1)\log\frac{2t^2}{t^2+1} \geqslant 2\log\frac{t^2+1}{2t}.$$

This is evident. Comparing the derivatives, it suffices to show

$$\log\frac{2t^2}{t^2+1} \geqslant \frac{2t-2}{t^2+1}.$$

https://doi.org/10.1017/S0963548306008340 Published online by Cambridge University Press

Once again, this is evident. Comparing the derivatives for the last time, we have to show

$$\frac{1}{t} \geqslant \frac{-t^2 + 2t + 1}{t^2 + 1},$$

or $t^3 + 1 \geqslant t^2 + t$, which is immediate for $t \geqslant 1$. □

Acknowledgements

We are grateful to Ehud Friedgut for his suggestions, which led to a significant simplification of the proof of Theorem 2.2. We also thank Gil Kalai, Nati Linial and Amites Sarkar for many valuable remarks.

References

[1] Beckner, W. (1975) Inequalities in Fourier analysis. *Annals of Math.* **102** 159–182.
[2] Ben-Or, M. and Linial, N. (1989) Collective coin flipping. In *Randomness and Computation* (S. Micali, ed.), Academic Press, pp. 91–115.
[3] Bezrukov, S. (1994) Isoperimetric problems in discrete spaces. In *Extremal Problems for Finite Sets*, Vol. 3 of *Bolyai Soc. Math. Stud.* (P. Frankl, Z. Füredi, G. Katona and D. Miklos, eds), pp. 59–91.
[4] Bobkov, S. G. and Tetali, P. (2003) Modified logarithmic Sobolev inequalities in discrete settings. In *Proc. STOC 2003*, pp. 287–296.
[5] Bonami, A. (1970) Etude des coefficients Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier* (Grenoble) **20** 335–402.
[6] Diaconis, P. and Saloff-Coste, L. (1996) Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.* **6** 695–750.
[7] Friedgut, E. (1998) Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica* **18** 27–36.
[8] Friedgut, E. and Kalai, G. (1996) Every monotone graph property has a sharp threshold. *Proc. Amer. Math. Soc.* **124** 2993–3002.
[9] Friedgut, E. and Rödl, V. (2001) Proof of a hypercontractive estimate via entropy. *Israel J. Math.* **125** 369–380.
[10] Gross, L. (1975) Logarithmic Sobolev inequalities. *Amer. J. Math.* **97** 1061–1083.
[11] Harper, L. H. (1964) Optimal assignment of numbers to vertices. *J. Soc. Ind. Appl. Math.* **12** 131–135.
[12] Harper, L. H. (1966) Optimal numberings and isoperimetric problems on graphs. *J. Combin. Theory* **1** 385–393.
[13] Hart, S. (1976) A note on the edges of the n-cube. *Discrete Math.* **14** 157–163.
[14] Jerrum, M. (2003) *Counting, Sampling and Integrating: Algorithms and Complexity*, Lectures in Mathematics, ETH Zürich. Birkhäuser, Basel.
[15] Jerrum, M. and Sinclair, A. (1996) The Markov chain Monte Carlo method: An approach to approximate counting and integration. In *Approximation Algorithms for NP-hard Problems* (D. Hochbaum, ed.), PWS, pp. 482–520.
[16] Kahn, J., Kalai, G. and Linial, N. (1988) The influence of variables on boolean functions. In *FOCS 1988*, pp. 68–80.
[17] Kalai, G. Personal communication.
[18] Latala, R. On some inequalities for Gaussian measures. In *Proc. International Congress of Mathematicians*, Vol. II, pp. 813–822.
[19] Levenshtein, V. I. (1995) Krawtchouk polynomials and universal bounds for codes and designs in Hamming spaces. *IEEE Trans. Inform. Theory* **IT-41**, pp. 1303–1321.

https://doi.org/10.1017/S0963548306008340 Published online by Cambridge University Press

[20] van Lint, J. H. (1999) *Introduction to Coding Theory*, Springer, Berlin.
[21] Milman, V. D. and Schechtman, G. (1986) *Asymptotic Theory of Finite Dimensional Normed Spaces*, Springer, New York.
[22] Mossel, E. Personal communication.
[23] Mossel, E., O'Donnell, R. and Oleszkiewicz, K. (2005) Noise stability of functions with low influences: Invariance and optimality. In *FOCS 2005*, pp. 21–30.
[24] Navon, M. and Samorodnitsky, A. (2005) On Delsarte's linear programming bounds for binary codes. In *FOCS 2005*, pp. 327–338.
[25] Rossignol, R. (2005) Threshold for monotone symmetric properties through a logarithmic Sobolev inequality. *Ann. Probab.* **34** 1707–1725. arXiv:math.PR/0511607.
[26] Sarkar, A. Personal communication.
[27] Talagrand, M. (1994) On Russo's approximate $0-1$ law. *Ann. Probab.* **22** 1576–1587.
[28] Talagrand, M. (1995) Concentration of measure and isoperimetric inequalities in product spaces. *Publ. Math. Inst. Hautes Etudes Sci.* **81** 73–203.
[29] Talagrand, M. (1997) On boundaries and influences. *Combinatorica* **17** 275–285.
[30] Szegö, G. (1939) *Orthogonal Polynomials*, AMS.

https://doi.org/10.1017/S0963548306008340 Published online by Cambridge University Press