Combinatorics, Probability and Computing (2011) 20, 363–380.
doi:10.1017/S0963548311000083

Almost Isoperimetric Subsets of the Discrete Cube

DAVID ELLIS

St John's College, Cambridge, UK
(e-mail: dce27@cam.ac.uk)

Received 18 May 2010; revised 25 January 2011; first published online 17 February 2011

We show that a set $A \subset \{0,1\}^n$ with edge-boundary of size at most

$$|A|(\log_2(2^n/|A|) + \epsilon)$$

can be made into a subcube by at most $(2\epsilon/\log_2(1/\epsilon))|A|$ additions and deletions, provided ϵ is less than an absolute constant.

We deduce that if $A \subset \{0,1\}^n$ has size 2^t for some $t \in \mathbb{N}$, and cannot be made into a subcube by fewer than $\delta|A|$ additions and deletions, then its edge-boundary has size at least

$$|A|\log_2(2^n/|A|) + |A|\delta\log_2(1/\delta) = 2^t(n - t + \delta\log_2(1/\delta)),$$

provided δ is less than an absolute constant. This is sharp whenever $\delta = 1/2^j$ for some $j \in \{1, 2, \ldots, t\}$.

1. Introduction

We work in the n-dimensional discrete cube $\{0,1\}^n$, the set of all 0–1 vectors of length n. This may be identified with $\mathcal{P}([n])$, the set of all subsets of $[n] = \{1, 2, \ldots, n\}$, by identifying a set $x \subset [n]$ with its characteristic vector χ_x in the usual way. A *d-dimensional subcube* of $\{0,1\}^n$ is a set of the form

$$\{x \in \{0,1\}^n : \ x_{i_1} = a_1, x_{i_2} = a_2, \ldots, x_{i_{n-d}} = a_{n-d}\},$$

where $i_1 < i_2 < \cdots < i_{n-d}$ are coordinates, and $a_1, a_2, \ldots$ and a_{n-d} are fixed elements of $\{0,1\}$. The coordinates $i_1, i_2, \ldots, i_{n-d}$ are called the *fixed* coordinates; the other coordinates are called the *moving* coordinates, and $n - d$ is called the *codimension* of the subcube.

Consider the graph Q_n with vertex-set $\{0,1\}^n$, where we join two 0–1 vectors if they differ in exactly one coordinate; this graph is called the *n-dimensional hypercube*. Given a set $A \subset \{0,1\}^n$, the *edge-boundary* of A is defined to be the set of all edges of Q_n joining a point in A to a point not in A. We write ∂A for the edge-boundary of A.

For $1 \leqslant k \leqslant 2^n$, let $C_{n,k}$ be the first k elements of the *binary ordering* on $\mathcal{P}([n])$, defined by

$$x < y \Leftrightarrow \max(x\Delta y) \in y.$$

The edge-isoperimetric inequality of Harper [6], Lindsey [12], Bernstein [2] and Hart [7] states that among all subsets of $\{0,1\}^n$ of size k, $C_{n,k}$ has the smallest possible edge-boundary.

A slightly weaker form is as follows:

$$|\partial A| \geqslant |A| \log_2(2^n/|A|) \quad \forall A \subset \{0,1\}^n; \tag{1.1}$$

equality holds if and only if A is a subcube. We call $|\partial A|/|A|$ the *average out-degree* of A; (1.1) says that the average out-degree of A is at least $\log_2(2^n/|A|)$ (which is the average out-degree of a subcube of size $|A|$, when $|A|$ is a power of 2). Writing $p = |A|/2^n$ for the measure of the set A, we may rewrite (1.1) as

$$|\partial A| \geqslant 2^n p \log_2(1/p) \quad \forall A \subset \{0,1\}^n.$$

Hence, if $|A| = 2^{n-1}$, $|\partial A| \geqslant 2^{n-1}$, and equality holds only if A is a codimension-1 subcube, in which case the edge-boundary consists of all the edges in one direction.

It is natural to ask whether it is always possible to find a direction in which there are many boundary edges. For $i \in [n]$, we write

$$A_i^+ = \{x \setminus \{i\} : x \in A,\ i \in x\} \subset \mathcal{P}([n] \setminus \{i\}),$$

and

$$A_i^- = \{x \in A : i \notin x\} \subset \mathcal{P}([n] \setminus \{i\});$$

A_i^+ and A_i^- are called the *upper* and *lower i-sections* of A, respectively. We write

$$\partial_i A = |A_i^+ \Delta A_i^-|$$

for the number of edges of the boundary of A in direction i. The *influence* of the coordinate i on the set A is defined to be

$$\beta_i = |A_i^+ \Delta A_i^-|/2^{n-1},$$

i.e., the fraction of direction-i edges of Q_n which belong to ∂A. This is simply the probability that if $S \subset \mathcal{P}([n])$ is chosen uniformly at random, A contains exactly one of S and $S\Delta\{i\}$. Clearly,

$$|\partial A| = 2^{n-1} \sum_{i=1}^{n} \beta_i.$$

Ben-Or and Linial [1] conjectured that for any set $A \subset \{0,1\}^n$ with $|A| = 2^{n-1}$, there exists a coordinate with influence at least $\Omega\left(\frac{\log_2 n}{n}\right)$. This was proved by Kahn, Kalai and Linial; it follows from the celebrated KKL theorem.

Theorem 1.1 (Kahn, Kalai and Linial [9]). *If $A \subset \{0,1\}^n$ with measure p, then*

$$\sum_{i=1}^{n} \beta_i^2 \geqslant Cp^2(1-p)^2(\ln n)^2/n,$$

where $C > 0$ is an absolute constant.

Corollary 1.2. *If $A \subset \{0,1\}^n$ with measure p, then there exists a coordinate $i \in [n]$ with*

$$\beta_i \geqslant C'p(1-p)(\ln n)/n,$$

where $C' > 0$ is an absolute constant.

Corollary 1.2 is sharp up to the value of the absolute constant C', as can be seen from the 'tribes' construction of Ben-Or and Linial [1]. Let $n = kl$, and split $[n]$ into l 'tribes' of size k. Let A be the set of all 0–1 vectors which are identically 0 on at least one tribe. Observe that

$$|A| = (1-(1-2^{-k})^l)2^n,$$
$$|\partial A| = n2^{n-k}(1-2^{-k})^{l-1},$$

and

$$\beta_i = 2^{-(k-1)}(1-2^{-k})^{l-1} \quad \forall i \in [n].$$

Let $k = 2^j$ for some $j \in \mathbb{N}$, and let $l = 2^k/k$, so that $n = 2^k = 2^{2^j}$; then

$$1-p = (1-2^{-k})^l = (1-2^{-k})^{2^k/k} = 1 - 1/k + O(1/k^2),$$

and

$$\beta_i = \frac{2(1-p)}{n(1-2^{-k})} = \frac{2(1-1/k+O(1/k^2))}{n} \quad \forall i \in [n],$$

so

$$\frac{\beta_i}{p(1-p)\ln(n)/n} = \frac{2(1-1/k+O(1/k^2))}{(1/k - O(1/k^2)(1-O(1/k))k\ln 2} = \frac{2}{\ln 2}(1+O(1/k)).$$

The best possible values of the constants C and C' (in Theorem 1.1 and Corollary 1.2 respectively) remain unknown. Falik and Samorodnitsky [3] have shown that one can take $C = 4$, and therefore $C' = 2$.

Kahn, Kalai and Linial's proof of Theorem 1.1 is one of the first instances of Fourier analysis on $\{0,1\}^n$ being used to prove a purely combinatorial result; Fourier analysis has since become a very important tool in both probabilistic and extremal combinatorics. More recently, Falik and Samorodnitsky [3] gave an entirely combinatorial proof of Theorem 1.1; a similar proof was found independently by Rossignol [13]. In [3], Falik and Samorodnitsky use influence-based methods to obtain several other results on subsets of $\{0,1\}^n$ with small edge-boundary.

What happens if the edge-boundary of A has size *close* to $|A|\log_2(2^n/|A|)$? How close must A be to a subcube? Using the techniques of Fourier analysis, Friedgut, Kalai and

https://doi.org/10.1017/S0963548311000083 Published online by Cambridge University Press

Naor [5] proved that if $A \subset \{0,1\}^n$ with $|A| = 2^{n-1}$ and $|\partial A| \leqslant 2^{n-1}(1+\epsilon)$, then A can be made into a codimension-1 subcube by at most $K\epsilon 2^{n-1}$ additions and deletions, where K is an absolute constant. Bollobás, Leader and Riordan [11] conjectured that for any $N \in \mathbb{N}$, there exists a constant K_N depending on N such that any $A \subset \{0,1\}^n$ with $|A| = 2^{n-N}$ and

$$|\partial A| \leqslant (1+\epsilon)|A| \log_2(2^n/|A|)$$

can be made into a codimension-N subcube by at most $K_N \epsilon 2^{n-N}$ additions and deletions. They proved this for $N = 2$ and $N = 3$, also using the techniques of Fourier analysis. We remark that K_N must necessarily depend on N. Indeed, as was observed by Samorodnitsky [14], a variant of the 'tribes' construction of Ben-Or and Linial provides an example of a (small) set A satisfying

$$|\partial A| \leqslant (1+\epsilon)|A| \log_2(2^n/|A|),$$

and yet requiring at least $(1-o(1))|A|$ additions and deletions to make it into a subcube. As above, let $n = kl$, split $[n]$ into l 'tribes' of size k, and let A be the set of all 0–1 vectors which are identically 0 on at least one tribe. Fix an integer s. Let $k = 2^j$, and let $l = 2^{k/2^s}/k = 2^{2^{j-s}-j}$, so that $n = 2^{k/2^s} = 2^{2^{j-s}}$. Let $j \to \infty$. Then

$$1 - p = (1-2^{-k})^l = 1 - l2^{-k} + O((l2^{-k})^2) \geqslant 1 - l2^{-k},$$

so

$$p \leqslant l2^{-k},$$

and therefore

$$\log_2(1/p) \geqslant k - \log_2 l = (1-2^{-s})k + \log_2 k.$$

Note that

$$|\partial A| = n2^{n-k}(1-2^{-k})^{l-1} = \frac{n2^{n-k}(1-p)}{1-2^{-k}} = n2^{n-k}(1+O(l2^{-k})).$$

Hence,

$$\begin{aligned}\frac{|\partial A|}{|A|\log_2(2^n/|A|)} &\leqslant \frac{n2^{n-k}(1+O(l2^{-k}))}{(l2^{-k}(1-O(l2^{-k})))((1-2^{-s})k+\log_2 k)2^n}\\ &= \frac{kl(1+O(l2^{-k}))}{l((1-2^{-s})k+\log_2 k)}\\ &= \frac{1+O(l2^{-k})}{1-2^{-s}+(\log_2 k)/k}\\ &< \frac{1}{1-2^{-s}},\end{aligned}$$

provided j is sufficiently large depending on s. For any $\epsilon > 0$, this can clearly be made $\leqslant 1+\epsilon$ by choosing s to be sufficiently large depending on ϵ. However, A is a union of l codimension-k subcubes with disjoint sets of fixed coordinates, and therefore requires at least $(1-o(1))|A|$ additions and deletions to make it into a subcube.

https://doi.org/10.1017/S0963548311000083 Published online by Cambridge University Press

Samorodnitsky [14] conjectured that given any $\delta > 0$, there exists an $a > 0$ such that any $A \subset \{0,1\}^n$ with

$$|\partial A| \leqslant (1 + a/n)|A| \log_2(2^n/|A|)$$

can be made into a subcube by at most $\delta|A|$ additions and deletions. Making use of a result of Keevash [10] on the structure of r-uniform hypergraphs with small shadows, he proved that any $A \subset \{0,1\}^n$ with

$$|\partial A| \leqslant (1 + n^{-4})|A| \log_2(2^n/|A|)$$

can be made into a subcube by at most $o(|A|)$ additions and deletions.

It turns out that the correct condition to ensure that A is close to a subcube is that $|\partial A|/|A|$, the average out-degree of A, is close to $\log_2(2^n/|A|)$. Our first main result (Theorem 2.4) implies that if $A \subset \{0,1\}^n$ has edge-boundary of size at most

$$|A|(\log_2(2^n/|A|) + \epsilon), \tag{1.2}$$

where ϵ is less than an absolute constant, then it can be made into a subcube by at most

$$(1 + O(1/\log_2(1/\epsilon)))\frac{\epsilon}{\log_2(1/\epsilon)}|A| \leqslant \frac{2\epsilon}{\log_2(1/\epsilon)}|A|$$

additions and deletions. This proves the above conjecture of Bollobás, Leader and Riordan, and also that of Samorodnitsky.

We then prove Theorem 2.5, which states that if $A \subset \{0,1\}^n$ has size 2^t for some $t \in \mathbb{N}$, and edge-boundary of size at most

$$|A|(\log_2(2^n/|A|) + \epsilon) = 2^t(n - t + \epsilon),$$

where ϵ is less than an absolute constant, then it can be made into a t-dimensional subcube by at most $\delta_1(\epsilon)|A|$ additions and deletions, where $\delta_1(\epsilon)$ is the unique root of

$$x \log_2(1/x) = \epsilon$$

in $(0, 1/e)$. It follows that if $A \subset \{0,1\}^n$ has size 2^t for some $t \in \mathbb{N}$, and cannot be made into a subcube by fewer than $\delta|A|$ additions and deletions, then

$$|\partial A| \geqslant |A| \log_2(2^n/|A|) + |A|\delta \log_2(1/\delta) = 2^t(n - t + \delta \log_2(1/\delta)),$$

provided δ is less than an absolute constant. This is sharp whenever $\delta = 1/2^j$ for some $j \in \{1, 2, \ldots, t\}$.

Our first aim is to prove a 'rough' stability result (Theorem 2.3), stating that if A is 'almost isoperimetric', in the sense that the average out-degree of ∂A is not too far above $\log_2(2^n/|A|)$, then A can be made into a subcube by a small number of additions and deletions. Influence-based methods play a crucial role in our proof. Indeed, it will turn out that a set $A \subset \{0,1\}^n$ satisfying (1.2) must have each influence either very small or very large. We will use the following theorem of Talagrand [16].

Theorem 1.3 (Talagrand). *Suppose $A \subset \{0,1\}^n$ with measure*

$$\frac{|A|}{2^n} = p.$$

https://doi.org/10.1017/S0963548311000083 Published online by Cambridge University Press

Then its influences satisfy

$$\sum_{i=1}^{n} \beta_i / \log_2(1/\beta_i) \geqslant K p(1-p),$$

where $K > 0$ *is an absolute constant.*

This implies that if all the influences are small, the edge-boundary must be very large. This will help to show that there must be a coordinate, i say, of very large influence. It will follow that one of the i-sections of A is very small. An inductive argument will enable us to complete the proof.

2. Main results

We first prove a sequence of results on the rough structure of subsets of $\{0,1\}^n$ with small edge-boundary. If $A \subset \{0,1\}^n$, and $i \in [n]$, we define

$$\gamma_i = \frac{\min\{|A_i^+|, |A_i^-|\}}{|A|}.$$

(Observe that we always have $\gamma_i \leqslant 1/2$.) We first show that if $A \subset \{0,1\}^n$ has small edge-boundary, then for each $i \in [n]$, either one of the i-sections of A is very small, or else the upper and lower i-sections of A have very similar sizes.

Lemma 2.1. *Let* $A \subset \{0,1\}^n$ *with*

$$|\partial A| = |A|(\log_2(2^n/|A|) + \epsilon_0). \tag{2.1}$$

Then, for each $i \in [n]$*, either*

(1) $\gamma_i \leqslant \epsilon_0/(5(\log_2 5 - 2))$*, or*
(2) $1/2 - \epsilon_0 < \gamma_i \leqslant 1/2$.

Proof. Let $A \subset \{0,1\}^n$, satisfying the hypothesis of the lemma. Write

$$p = \frac{|A|}{2^n}$$

for the measure of A; then

$$|\partial A| = 2^n p(\log_2(1/p) + \epsilon_0).$$

Fix $i \in [n]$. Without loss of generality, we may assume that $|A_i^+| \leqslant |A_i^-|$, so

$$\gamma_i = \frac{|A_i^+|}{|A|}.$$

Write $\gamma = \gamma_i$. Let

$$p^+ = \frac{|A_i^+|}{2^{n-1}}, \; p^- = \frac{|A_i^-|}{2^{n-1}}.$$

https://doi.org/10.1017/S0963548311000083 Published online by Cambridge University Press

Note that

$$p^+ = 2\gamma p,\ \ p^- = 2(1-\gamma)p.$$

Define ϵ^+, ϵ^- by

$$|\partial A_i^+| = |A_i^+|(\log_2(2^{n-1}/|A_i^+|) + \epsilon^+), \quad |\partial A_i^-| = |A_i^-|(\log_2(2^{n-1}/|A_i^-|) + \epsilon^-).$$

Observe that

$$\begin{aligned}
|\partial A| &= |\partial A_i^+| + |\partial A_i^-| + |A_i^+ \Delta A_i^-| \\
&= |A_i^+|(\log_2(2^{n-1}/|A_i^+|) + \epsilon^+) + |A_i^-|(\log_2(2^{n-1}/|A^-|) + \epsilon^-) + |A_i^+ \Delta A_i^-| \\
&= \gamma|A|\log_2(2^n/(2\gamma|A|)) + (1-\gamma)|A|(\log_2(2^n/(2(1-\gamma)|A|) + \epsilon^+|A_i^+| + \epsilon^-|A_i^-| \\
&\qquad + |A_i^+ \Delta A_i^-| \\
&= |A|\log_2(2^n/|A|) - (1 - H_2(\gamma))|A| + \epsilon^+|A_i^+| + \epsilon^-|A_i^-| + |A_i^+ \Delta A_i^-| \qquad (2.2)\\
&\geqslant |A|\log_2(2^n/|A|) - (1 - H_2(\gamma))|A| + \epsilon^+|A_i^+| + \epsilon^-|A_i^-| + ||A_i^+| - |A_i^-|| \\
&= |A|\log_2(2^n/|A|) - (1 - H_2(\gamma))|A| + \epsilon^+|A_i^+| + \epsilon^-|A_i^-| + (1-2\gamma)|A| \\
&= |A|\log_2(2^n/|A|) + (H_2(\gamma) - 2\gamma)|A| + \epsilon^+|A_i^+| + \epsilon^-|A_i^-| \\
&= |A|\log_2(2^n/|A|) + F(\gamma)|A| + \epsilon^+|A_i^+| + \epsilon^-|A_i^-|,
\end{aligned}$$

where $H_2 : [0,1] \to \mathbb{R}$ denotes the *binary entropy* function,

$$H_2(\gamma) := \gamma\log_2(1/\gamma) + (1-\gamma)\log_2(1/(1-\gamma)),$$

and

$$F(\gamma) := H_2(\gamma) - 2\gamma.$$

Hence, (2.1) implies that

$$\gamma\epsilon^+ + (1-\gamma)\epsilon^- + F(\gamma) \leqslant \epsilon_0. \qquad (2.3)$$

Therefore, crudely,

$$F(\gamma) \leqslant \epsilon_0.$$

The function F is concave on $[0,1/2]$, and attains its maximum at $\gamma = 1/5$, where it takes the value $\log_2 5 - 2$. Hence, for $\gamma \leqslant 1/5$,

$$F(\gamma) \geqslant 5(\log_2 5 - 2)\gamma,$$

whereas for $1/5 \leqslant \gamma \leqslant 1/2$,

$$F(1/2 - \eta) \geqslant \tfrac{10}{3}(\log_2 5 - 2)\eta > \eta.$$

Hence, for each $i \in [n]$, either

(1) $\gamma_i \leqslant \epsilon_0/(5(\log_2 5 - 2))$, or
(2) $1/2 - \epsilon_0 < \gamma_i \leqslant 1/2$,

proving the lemma. □

https://doi.org/10.1017/S0963548311000083 Published online by Cambridge University Press

Remark 1. We can of course rephrase the conclusion of Lemma 2.1 in terms of influences. Let $A \subset \{0,1\}^n$ satisfying (2.5). Observe that if case (1) occurs for $i \in [n]$, then

$$\beta_i \geqslant (1-2\gamma_i)|A|/2^{n-1} = 2(1-2\gamma_i)p \geqslant 2\left(1 - 2\frac{\epsilon_0}{5(\log_2 5 - 2)}\right)p, \tag{2.4}$$

(the ith influence is 'large').

If, on the other hand, case (2) occurs, then by (2.2), we have

$$|A_i^+ \Delta A_i^-| \leqslant |\partial A| - |A|\log_2(2^n/|A|) + (1 - H_2(\gamma_i))|A| = (\epsilon_0 + 1 - H_2(\gamma_i))|A|.$$

Since H_2 is concave, with $H_2(1/2) = 1$, we have

$$1 - H_2(1/2 - \eta) \leqslant 2\eta \ (0 \leqslant \eta \leqslant 1/2),$$

and therefore

$$|A_i^+ \Delta A_i^-| < 3\epsilon_0 |A|,$$

i.e.,

$$\beta_i < 6\epsilon_0 p,$$

(the ith influence is 'small').

We now show that if the edge-boundary of A is sufficiently small, then case (1) in Lemma 2.1 must occur for some $i \in [n]$.

Lemma 2.2. *There exists an absolute constant $c > 0$ such that the following holds. If $\epsilon \leqslant c$, and $A \subset \{0,1\}^n$ with measure*

$$\frac{|A|}{2^n} \leqslant 1 - \epsilon,$$

and

$$|\partial A| \leqslant |A|(\log_2(2^n/|A|) + \epsilon), \tag{2.5}$$

then case (1) *must occur for some $i \in [n]$, i.e., $\gamma_i \leqslant \epsilon/(5(\log_2 5 - 2))$ for some $i \in [n]$.*

Proof. We can easily prove the lemma for sets with measure $p \in [1/2, 7/8]$. Suppose $A \subset \{0,1\}^n$ has measure $p \in [1/2, 7/8]$ and satisfies (2.5). Suppose for a contradiction that case (2) occurs for every $i \in [n]$. Then, by Remark 1, $\beta_i < 6\epsilon p$ for every $i \in [n]$, and therefore, by Theorem 1.3,

$$\sum_{i=1}^{n} \beta_i > Kp(1-p)\log_2\left(\frac{1}{6\epsilon p}\right).$$

The right-hand side is at least

$$2p(\log_2(1/p)+\epsilon)$$

provided

$$\tfrac{K}{8}\log_2\left(\frac{1}{6\epsilon}\right) \geqslant 2(1+\epsilon),$$

which holds for all $\epsilon \leqslant c := 2^{-32K}/6$. This contradicts (2.5), proving the lemma for $p \in [1/2, 7/8]$.

Now observe that *any* set $A \subset \{0,1\}^n$ with measure $p \in [7/8, 1-\epsilon]$ has

$$|\partial A| > |A|(\log_2(2^n/|A|)+\epsilon), \tag{2.6}$$

To see this, just apply the edge-isoperimetric inequality (1.1) to A^c:

$$|\partial A| = |\partial(A^c)| \geqslant 2^n(1-p)\log_2(1/(1-p)).$$

It is easily checked that

$$2^n(1-p)\log_2(1/(1-p)) > 2^n p(\log_2(1/p)+1-p) \quad \forall p \geqslant 7/8,$$

so (2.6) holds for all $p \in [7/8, 1-\epsilon]$. Hence, any set $A \subset \{0,1\}^n$ satisfying (2.5) must have measure $p \leqslant 7/8$.

It remains to prove the lemma for all sets of measure $p \leqslant 1/2$. Suppose A has measure $p \leqslant 1/2$ and satisfies (2.5). Suppose for a contradiction that case (2) occurs for every $i \in [n]$.

Fix any $i \in [n]$. Without loss of generality, we may assume that $|A_i^+| \leqslant |A_i^-|$, so that

$$\gamma_i = \frac{|A_i^+|}{|A|}.$$

Write $\gamma = \gamma_i$. Define ϵ^+ and ϵ^- as in the proof of Lemma 2.1. By (2.3), we have

$$\gamma\epsilon^+ + (1-\gamma)\epsilon^- + F(\gamma) \leqslant \epsilon.$$

Hence, crudely,

$$\gamma\epsilon^+ + (1-\gamma)\epsilon^- \leqslant \epsilon,$$

so either $\epsilon^+ \leqslant \epsilon$ or $\epsilon^- \leqslant \epsilon$.

If $\epsilon^+ \leqslant \epsilon$, then let $A' = A_i^+$. The set A' is a subset of $\mathcal{P}([n]\setminus\{i\})$ of measure $p' := 2\gamma p \in ((1-2\epsilon)p, p) \subset [0,1/2]$, satisfying the conditions of the lemma.

If $\epsilon^- \leqslant \epsilon$, then let $A' = A_i^-$. The set A' is a subset of $\mathcal{P}([n]\setminus\{i\})$ of measure $p' := 2(1-\gamma)p < 2(1/2+\epsilon)p \leqslant 1/2+\epsilon < 7/8$, satisfying the conditions of the lemma.

If A' has case (1) occurring for some j, then by (2.4),

$$\begin{aligned}
\beta'_j &\geqslant 2\left(1 - 2\frac{\epsilon}{5(\log_2 5 - 2)}\right)p' \\
&\geqslant 2\left(1 - 2\frac{\epsilon}{5(\log_2 5 - 2)}\right)(1-2\epsilon)p \\
&> 2(1-2\epsilon)^2 p,
\end{aligned}$$

and therefore

$$\beta_j > (1-2\epsilon)^2 p > 6\epsilon p,$$

contradicting our assumption that A has case (2) occurring for every $i \in [n]$. Therefore, A' also has case (2) occurring for every coordinate. Hence, it must have measure $p' < 1/2$, by the above argument for sets of measure in $[1/2, 7/8]$. Repeat the same argument for A', and continue; we obtain a sequence of set systems $(A^{(l)})$ on ground sets of sizes $n-l$, all with measure $< 1/2$, satisfying the conditions of the lemma, and with case (2) occurring for every coordinate. Stop at the minimum M such that $A^{(M)} = \emptyset$; clearly, $M \leqslant n-1$. Then $A^{(M-1)}$ has one of its j-sections empty for some j, so case (1) must occur for this j, a contradiction. This proves the lemma. □

We can now prove a rough stability result for subsets of $\{0,1\}^n$ with small edge-boundary.

Theorem 2.3. *There exists an absolute constant $c > 0$ such that if $A \subset \{0,1\}^n$ with*

$$|\partial A| \leqslant |A| \log_2(2^n/|A|) + \epsilon|A|$$

for some $\epsilon \leqslant c$, then

$$|A\Delta C|/|A| < 3\epsilon$$

for some subcube C.

Proof. Let c be the constant in Lemma 2.2. Let $A \subset \{0,1\}^n$ be such that

$$|\partial A| \leqslant |A| \log_2(2^n/|A|) + \epsilon|A|,$$

for some $\epsilon \leqslant c$. Let $\epsilon_0 \leqslant \epsilon$ be such that

$$|\partial A| = |A|(\log_2(2^n/|A|) + \epsilon_0).$$

By Lemma 2.2, there exists $i \in [n]$ with case (1) occurring, *i.e.*, with

$$\gamma_i \leqslant \epsilon/(5(\log_2 5 - 2)).$$

Without loss of generality, we may assume that $i = n$, and that $|A_n^+| \leqslant |A_n^-|$. In keeping with our earlier notation, we write $\gamma = \gamma_n = |A_n^+|/|A|$.

To avoid confusion, we now write $B^{(0)} = A$, $p^{(0)} = p$, $\epsilon^{(0)} = \epsilon_0$, and $\gamma^{(0)} = \gamma$. Let $B^{(1)} = A_n^- \subset \mathcal{P}([n-1])$, let $p^{(1)} = p_n^-$, and let $\epsilon^{(1)} = \epsilon_n^-$.

By (2.3), we have

$$(1-\gamma^{(0)})\epsilon^{(1)} + F(\gamma^{(0)}) \leqslant \epsilon^{(0)}.$$

Since $F(\gamma^{(0)}) \geqslant 5(\log_2 5 - 2)\gamma^{(0)}$, we have

$$(1-\gamma^{(0)})\epsilon^{(1)} + 5(\log_2 5 - 2)\gamma^{(0)} \leqslant \epsilon^{(0)};$$

https://doi.org/10.1017/S0963548311000083 Published online by Cambridge University Press

it follows that $\epsilon^{(1)} \leqslant \epsilon \leqslant c$. Hence, $B^{(1)} \subset \mathcal{P}([n-1])$ also satisfies the hypothesis of Theorem 2.3 (with n replaced by $n-1$). Its measure $p^{(1)}$ satisfies

$$\begin{aligned} p^{(1)} &= 2(1-\gamma^{(0)})p^{(0)} \\ &\geqslant 2\left(1 - \frac{\epsilon^{(0)}}{5(\log_2 5 - 2)}\right)p^{(0)} \\ &> 2(1-\epsilon^{(0)})p^{(0)} \\ &\geqslant 2(1-c)p^{(0)}. \end{aligned}$$

Repeat the same argument for $B^{(1)}$. We obtain a sequence of set systems $(B^{(k)})$ on ground sets of sizes $n-k$, satisfying the hypotheses of Theorem 2.3 with ϵ replaced by $\epsilon^{(k)} \leqslant \epsilon_0 \leqslant c$, with measures $p^{(k)}$ satisfying

$$p^{(k+1)} > 2(1-\epsilon^{(k)})p^{(k)} \quad \forall k \geqslant 0,$$

and with

$$(1-\gamma^{(k)})\epsilon^{(k+1)} + F(\gamma^{(k)}) \leqslant \epsilon^{(k)} \quad \forall k \geqslant 0. \tag{2.7}$$

Without loss of generality, we may assume that $B^{(k)} \subset \mathcal{P}([n-k])$.

We may continue this process until we produce a set system $B^{(N)}$ at stage N, for which $p^{(N)} > 1-\epsilon_0$, at which point we can no longer apply Lemma 2.2. We must now show that A is close to $\mathcal{P}([n-N])$. Observe that

$$\begin{aligned} |A \setminus B^{(N)}| &= \sum_{k=0}^{N-1} \gamma^{(k)} p^{(k)} 2^{n-k} \\ &= \sum_{k=0}^{N-1} 2^k \left(\prod_{j<k}(1-\gamma^{(j)})\right)\gamma^{(k)} p_0 2^{n-k} \\ &= \sum_{k=0}^{N-1} \left(\prod_{j<k}(1-\gamma^{(j)})\right)\gamma^{(k)} p_0 2^{n} \\ &= \sum_{k=0}^{N-1} \left(\prod_{j<k}(1-\gamma^{(j)})\right)\gamma^{(k)} |A|. \end{aligned}$$

By repeatedly applying the inequality (2.7), we obtain

$$\sum_{k=0}^{N-1}\left(\prod_{j<k}(1-\gamma^{(j)})\right)F(\gamma^{(k)}) + \left(\prod_{j=0}^{N-1}(1-\gamma^{(j)})\right)\epsilon_N \leqslant \epsilon_0,$$

so, certainly,

$$\sum_{k=0}^{N-1}\left(\prod_{j<k}(1-\gamma^{(j)})\right)F(\gamma^{(k)}) \leqslant \epsilon_0.$$

https://doi.org/10.1017/S0963548311000083 Published online by Cambridge University Press

Since $F(\gamma^{(k)}) \geqslant 5(\log_2 5 - 2)\gamma^{(k)}$ $(0 \leqslant k \leqslant N-1)$, it follows that

$$\sum_{k=0}^{N-1} \left(\prod_{j<k} (1-\gamma^{(j)}) \right) \gamma^{(k)} \leqslant \frac{\epsilon_0}{5(\log_2 5 - 2)}.$$

Hence,

$$|A \setminus B^{(N)}| \leqslant \frac{\epsilon_0}{5(\log_2 5 - 2)} |A| < \epsilon_0 |A|.$$

Let $C = \mathcal{P}([n-N])$, a codimension-N subcube. Then

$$|A \setminus C| = |A \setminus B^{(N)}| < \epsilon_0 |A|. \tag{2.8}$$

Since $p^{(N)} > 1 - \epsilon_0$, we have

$$|C \setminus A| < \epsilon_0 |C|. \tag{2.9}$$

Hence,

$$|C| < \frac{1}{1-\epsilon_0} |A|,$$

and therefore

$$|C \setminus A| < \frac{\epsilon_0}{1-\epsilon_0} |A| < 2\epsilon_0 |A|.$$

Combining this with (2.8) yields

$$|A \Delta C| < 3\epsilon_0 |A|, \tag{2.10}$$

proving Theorem 2.3. □

We may use this rough stability result to obtain a more precise one.

Theorem 2.4. *There exists an absolute constant $c > 0$ such that if $A \subset \{0,1\}^n$ with*

$$|\partial A| \leqslant |A| \log_2(2^n/|A|) + \epsilon |A|,$$

for some $\epsilon \leqslant c$, then

$$|A \Delta C| < \delta_0(\epsilon) |A|$$

for some subcube C, where $\delta_0(\epsilon)$ is the smallest positive solution of

$$x \log_2(1/x) - 3x = \epsilon.$$

Proof. Write

$$|\partial A| = |A| (\log_2(2^n/|A|) + \epsilon_0), \tag{2.11}$$

where $0 \leqslant \epsilon_0 \leqslant \epsilon$. Choose a subcube C such that $|A \Delta C|$ is minimal, and let $\delta = |A \Delta C|/|A|$. By Theorem 2.3, $\delta < 3\epsilon_0 \leqslant 3c < 1/2$.

https://doi.org/10.1017/S0963548311000083 Published online by Cambridge University Press

Without loss of generality, we may assume that $C = \mathcal{P}([n-N])$. Let $B = C \setminus A$ and let $D = A \setminus C$; then

$$|B| + |D| < 3\epsilon_0|A|.$$

Since every point of D is adjacent to at most one point of C, the number of edges in ∂A between points of $A \cap C$ and points of $\{0,1\}^n \setminus C$ is at least

$$N(2^{n-N} - |B|) - |D|.$$

The number of edges in ∂A between points of C is at least

$$|B| \log_2(2^{n-N}/|B|).$$

Finally, the number of edges of the cube in ∂D is at least

$$|D| \log_2(2^n/|D|),$$

and the number of edges of the cube between points in D and points in C is at most $|D|$, so the number of edges of the cube between points of D and points of $(\{0,1\}^n \setminus C) \setminus A$ is at least

$$|D|(\log_2(2^n/|D|) - 1).$$

It follows that

$$\begin{aligned}
|\partial A| &\geqslant N(2^{n-N} - |B|) - |D| + |B|\log_2(2^{n-N}/|B|) + |D|(\log_2(2^n/|D|) - 1)\\
&= N2^{n-N} + (\log_2(2^{n-N}/|B|) - N)|B| + (\log_2(2^n/|D|) - 2)|D|\\
&= N(|A| - |D| + |B|) + (\log_2(2^{n-N}/|B|) - N)|B|\\
&\quad + (\log_2(2^n/|D|) - 2)|D|\\
&= N|A| + |B|(\log_2(2^n/|B|) - N) + |D|(\log_2(2^n/|D|) - N - 2).
\end{aligned} \tag{2.12}$$

Write $|B| = \phi|A|$ and $|D| = \psi|A|$. Then $\delta = \psi + \phi$. Note that

$$N = \log_2\left(\frac{2^n}{|A| - |D| + |B|}\right) = \log_2\left(\frac{2^n}{|A|}\right) - \log_2(1 - \psi + \phi).$$

Hence, we obtain

$$\begin{aligned}
|\partial A| &\geqslant |A|\log_2(2^n/|A|) - |A|\log_2(1 - \psi + \phi)\\
&\quad + \phi|A|(\log_2(1/\phi) + \log_2(1 - \psi + \phi))\\
&\quad + \psi|A|(\log_2(1/\psi) - 2 + \log_2(1 - \psi + \phi))\\
&= |A|\log_2(2^n/|A|)\\
&\quad + |A|(\phi\log_2(1/\phi) + \psi\log_2(1/\psi) - 2\psi + (\psi + \phi - 1)\log_2(1 - \psi + \phi))\\
&> |A|\log_2(2^n/|A|) + |A|(\psi\log_2(1/\psi) + \phi\log_2(1/\phi) - 3\psi - 3\phi),
\end{aligned}$$

where the last inequality follows from the fact that $\psi, \phi < 1/2$. Observe that the function

$$\begin{aligned}
h : (0,1] &\to \mathbb{R},\\
x &\mapsto x\log_2(1/x)
\end{aligned}$$

is concave, and therefore

$$\psi \log_2(1/\psi) + \phi \log_2(1/\phi) \geqslant (\psi + \phi) \log_2(1/(\psi + \phi)).$$

We obtain

$$|\partial A| > |A| \log_2(2^n/|A|) + |A|((\psi + \phi) \log_2(1/(\psi + \phi)) - 3(\psi + \phi)).$$

Hence, by (2.11),

$$(\psi + \phi) \log_2(1/(\psi + \phi)) - 3(\psi + \phi) < \epsilon_0,$$

i.e.,

$$\delta(\log_2(1/\delta) - 3) < \epsilon_0.$$

It is easy to check that the function

$$g : (0, 1] \to \mathbb{R},$$
$$x \mapsto x \log_2(1/x) - 3x$$

is strictly increasing between 0 and $2^{-(3+1/\ln(2))}$; provided $3c \leqslant 2^{-(3+1/\ln(2))}$, it follows that $\delta < \delta_0(\epsilon)$, where $\delta_0(\epsilon)$ is the smallest positive solution of

$$x \log_2(1/x) - 3x = \epsilon,$$

proving Theorem 2.4. □

Remark 2. Observe that

$$\delta_0(\epsilon) = (1 + O(1/\log_2(1/\epsilon))) \frac{\epsilon}{\log_2(1/\epsilon)} \leqslant \frac{2\epsilon}{\log_2(1/\epsilon)}.$$

Similarly, we may obtain an exact stability result for set systems whose size is a power of 2.

Theorem 2.5. *There exists an absolute constant $c > 0$ such that if $A \subset \{0, 1\}^n$ with size $|A| = 2^{n-N}$ for some $N \in \mathbb{N}$, and with edge-boundary*

$$|\partial A| \leqslant |A| \log_2(2^n/|A|) + \epsilon|A|,$$

where $\epsilon \leqslant c$, then there exists a codimension-N subcube C such that

$$|A \Delta C| \leqslant \delta_1(\epsilon)|A|,$$

where $\delta_1(\epsilon)$ is the unique root of the equation

$$x \log_2(1/x) = \epsilon$$

in $(0, 1/e)$.

Proof. Write

$$|\partial A| = |A|(\log_2(2^n/|A|) + \epsilon_0), \tag{2.13}$$

https://doi.org/10.1017/S0963548311000083 Published online by Cambridge University Press

where $0 \leqslant \epsilon_0 \leqslant \epsilon$. Choose a subcube C such that $|A\Delta C|$ is minimal, and let $\delta = |A\Delta C|/|A|$. By Theorem 2.3, $\delta < 3\epsilon_0 \leqslant 3c < 1/2$.

Suppose C has codimension N'. Note that if $N \neq N'$, then $|A|$ and $|C|$ would differ by a factor of at least 2, so

$$|A\Delta C|/|A| \geqslant ||A| - |C||/|A| \geqslant 1/2,$$

a contradiction. Hence, $N' = N$, *i.e.*, $|C| = |A|$.

Let $B = C \setminus A$; then $|A \setminus C| = |C \setminus A| = |B|$. From (2.12), we have

$$\begin{aligned}|\partial A| &\geqslant |A|\log_2(2^n/|A|) + |B|(\log_2(2^n/|B|) - N) + |B|(\log_2(2^n/|B|) - N - 2)\\ &= |A|\log_2(2^n/|A|) + 2|B|\log_2(2^n/|B|) - 2|B|\log_2(2^n/|A|) - 2|B|\\ &= |A|\log_2(2^n/|A|) + |A|\delta\log_2(1/\delta).\end{aligned}$$

It follows that

$$\delta\log_2(1/\delta) \leqslant \epsilon.$$

Observe that the function

$$\begin{aligned}h : (0,1] &\to \mathbb{R},\\ x &\mapsto x\log_2(1/x)\end{aligned}$$

has

$$h'(x) = -\frac{1}{\ln 2}(1 + \ln x),$$

and is therefore strictly increasing between 0 and $1/e$, where it attains its maximum of $1/(e\ln 2)$, and strictly decreasing between $1/e$ and 1. Since $\delta < 3\epsilon \leqslant 3c < 1/e$, it follows that $\delta \leqslant \delta_1(\epsilon)$, where $\delta_1(\epsilon)$ is the unique root of the equation

$$x\log_2(1/x) = \epsilon$$

in $(0, 1/e)$, proving the theorem. □

The following is an immediate consequence of Theorem 2.5.

Corollary 2.6. *If $A \subset \{0,1\}^n$ has size 2^t for some $t \in \mathbb{N}$, and cannot be made into a subcube by fewer than $\delta|A|$ additions and deletions, then its edge-boundary satisfies*

$$|\partial A| \geqslant |A|\log_2(2^n/|A|) + |A|\max\{\delta\log_2(1/\delta), c\} = 2^t(n - t + \max\{\delta\log_2(1/\delta), c\}),$$

where $c > 0$ is an absolute constant. There exists an absolute constant $c' > 0$ such that if $\delta \leqslant c'$, then

$$|\partial A| \geqslant |A|\log_2(2^n/|A|) + |A|\delta\log_2(1/\delta) = 2^t(n - t + \delta\log_2(1/\delta)).$$

Remark 3. Observe that all we need from Theorem 2.3 to prove Theorem 2.5 is that

$$\delta = |A\Delta C|/|A| < 1/e.$$

https://doi.org/10.1017/S0963548311000083 Published online by Cambridge University Press

If we just knew that $\delta < 1/2$, we could still deduce from the above argument that $\delta \log_2(1/\delta) \leqslant \epsilon$.

Remark 4. Observe that Theorem 2.5 is best possible, apart from the restriction $\epsilon \leqslant c$. To see this, let $C = \mathcal{P}([n-N])$, a codimension-N-subcube, where $1 \leqslant N \leqslant n-1$. Let $2 \leqslant M \leqslant n-N$, and delete from C the codimension-$(N+M)$ subcube

$$B = \{x \cup \{n-N\} \,:\, x \in \mathcal{P}([n-N-M])\}.$$

Now add on the codimension-$(N+M)$ subcube

$$D = \{x \cup \{n\} \,:\, x \in \mathcal{P}([n-N-M])\}.$$

The resulting family $A = (C \setminus B) \cup D$ has

$$|A\Delta C|/|A| = 2^{-(M-1)} \leqslant 1/2;$$

it is easy to check that all other subcubes $C' \neq C$ have

$$|A\Delta C'| > |A\Delta C|.$$

Hence,

$$\delta := \min\{|A\Delta C'| \,:\, C' \text{ is a subcube}\}/|A| = |A\Delta C|/|A| = 2^{-(M-1)}.$$

Observe that we have equality in (2.12) for A, and therefore

$$|\partial A| = |A| \log_2(2^n/|A|) + |A|\delta \log_2(1/\delta).$$

3. Conclusion and open problems

Consider the function

$$f(\delta) = \inf\left\{ \frac{|\partial A| - |A|\log_2(2^n/|A|)}{|A|} \,:\, n \in \mathbb{N},\ A \subset \{0,1\}^n, \right.$$
$$\left. |A| \text{ is a power of 2, } |A\Delta C| \geqslant \delta|A| \text{ for all subcubes } C \right\}.$$

We have shown that $f(\delta) = \max(\delta \log_2(1/\delta), c)$ when $\delta = 1/2^j$ for some $j \in \mathbb{N}$, where $c > 0$ is an absolute constant, implying that $f(2^{-j}) = j2^{-j}$ for $j \in \mathbb{N}$ sufficiently large. We conjecture that the restriction on j could be removed.

Conjecture 3.1. *For any $j \in \mathbb{N}$,*

$$f(2^{-j}) = j2^{-j}.$$

As observed above, the function

$$h : (0,1] \to \mathbb{R},$$
$$x \mapsto x \log_2(1/x)$$

https://doi.org/10.1017/S0963548311000083 Published online by Cambridge University Press

is strictly decreasing between $1/e$ and 1, whereas f is clearly a non-decreasing function of δ. It would be interesting to determine the behaviour of $f(\delta)$ for $1/2 < \delta \leqslant 1$.

We also conjecture that Talagrand's theorem (Theorem 1.3) holds with $K = 2$. This was independently conjectured by Samorodnitsky [14]. It would be best possible, as can be seen by taking A to be a t-dimensional subcube; then $n-t$ influences are $2^{-(n-t-1)}$, and the rest are zero, so

$$\sum_{i=0}^{n} \beta_i / \log_2(1/\beta_i) = \frac{(n-t)2^{-(n-t-1)}}{n-t-1}.$$

Hence,

$$\frac{1}{p(1-p)} \sum_{i=0}^{n} \beta_i / \log_2(1/\beta_i) = \frac{2(n-t)}{(n-t-1)(1-2^{-(n-t)})} \to 2 \quad \text{as } n \to \infty.$$

Knowing this would obviously weaken the upper bound on ϵ required to prove Theorem 2.3, though it would not result in a proof of Conjecture 3.1.

It would be interesting to determine the structure of subsets $A \subset \{0,1\}^n$ satisfying

$$|\partial A| \leqslant L \log_2(2^n/|A|)$$

for L a fixed positive constant. Kahn and Kalai [8] conjecture the following.

Conjecture 3.2 (Kahn and Kalai). *For any $L > 0$, there exist $L' > 0$ and $\delta > 0$ such that, for any monotone increasing $A \subset \{0,1\}^n$ with measure*

$$p = \frac{|A|}{2^n} \leqslant 1/2,$$

there exists a subcube C with codimension at most $L' \log_2(1/p)$ and all fixed coordinates equal to 1, such that

$$\frac{|A \cap C|}{|C|} \geqslant (1+\delta)p. \tag{3.1}$$

We believe Conjecture 3.2 to be true for non-monotone sets as well, if one allows the subcube C to have fixed zeros as well as fixed ones.

Conjecture 3.3. *For any $L > 0$, there exist $L' > 0$ and $\delta > 0$ such that, for any $A \subset \{0,1\}^n$ with measure*

$$p = \frac{|A|}{2^n} \leqslant 1/2,$$

there exists a subcube C with codimension at most $L' \log_2(1/p)$, such that

$$\frac{|A \cap C|}{|C|} \geqslant (1+\delta)p.$$

Acknowledgements

The author would like to thank Alex Samorodnitsky for much helpful advice, and also Ehud Friedgut, Imre Leader, and Benny Sudakov for valuable discussions.

References

[1] Ben-Or, M. and Linial, N. (1985) Collective coin flipping, robust voting games, and minima of Banzhaf value. In *Proc. 26th IEEE Symposium on the Foundations of Computer Science*, pp. 408–416.
[2] Bernstein, A. J. (1967) Maximally connected arrays on the n-cube. *SIAM J. Appl. Math.* **15** 1485–1489.
[3] Falik, D. and Samorodnitsky, A. (2007) Edge-isoperimetric inequalities and influences. *Combin. Probab. Comput.* **16** 693–712.
[4] Friedgut, E. (1998) Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica* **18** 27–36.
[5] Friedgut, E., Kalai, G. and Naor, A. (2002) Boolean functions whose Fourier transform is concentrated on the first two levels. *Adv. Appl. Math.* **29** 427–437.
[6] Harper, L. H. (1964) Optimal assignments of numbers to vertices. *SIAM J. Appl. Math.* **12** 131–135.
[7] Hart, S. (1976) A note on the edges of the n-cube. *Discrete Math.* **14** 157–163.
[8] Kahn, J. and Kalai, G. (2007) Thresholds and expectation thresholds. *Combin. Probab. Comput.* **16** 495–502.
[9] Kahn, J., Kalai, G. and Linial, N. (1988) The influence of variables on boolean functions. In *FOCS* 1988, pp. 68–80.
[10] Keevash, P. (2008) Shadows and intersections: Stability and new proofs. *Adv. Math.* **218** 1685–1703.
[11] Leader, I. Personal communication.
[12] Lindsey, J. H., II (1964) Assignment of numbers to vertices. *Amer. Math. Monthly* **71** 508–516.
[13] Rossignol, R. (2005) Threshold for monotone symmetric properties through a logarithmic Sobolev inequality. *Ann. Probab.* **34** 1707–1725.
[14] Samorodnitsky, A. (2009) Edge isoperimetric inequalities in the Hamming cube. Talk given at the IPAM Long Program in Combinatorics, Workshop IV: *Analytical Methods in Combinatorics, Additive Number Theory and Computer Science*, November 2009.
[15] Samorodnitsky, A. Personal communication.
[16] Talagrand, M. (1994) On Russo's approximate 0–1 law. *Ann. Probab.* **22** 1576–1587.