Journal of Combinatorial Theory, Series A 155 (2018) 360–375

Contents lists available at ScienceDirect

Journal of Combinatorial Theory, Series A

www.elsevier.com/locate/jcta

A stability result for the cube edge isoperimetric inequality

Peter Keevash [1], Eoin Long

Mathematical Institute, University of Oxford, Oxford, UK

ARTICLE INFO

Article history:
Received 7 April 2017
Available online 24 November 2017

Keywords:
Isoperimetric inequality
Hypercube
Boolean analysis
Stability

ABSTRACT

We prove the following stability version of the edge isoperimetric inequality for the cube: any subset of the cube with average boundary degree within K of the minimum possible is ε-close to a union of L disjoint cubes, where $L \leq L(K, \varepsilon)$ is independent of the dimension. This extends a stability result of Ellis, and can viewed as a dimension-free version of Friedgut's junta theorem.

1. Introduction

The edge isoperimetric inequality is a fundamental result in Extremal Combinatorics concerning the distribution of edges in the cube. The *n-cube* Q_n is the graph on vertex set $\{0,1\}^n$ in which vertices are adjacent if they differ in a single coordinate. The *edge boundary* of a set $\mathcal{A} \subset V(Q_n)$ is the set of edges $\partial_e(\mathcal{A}) \subset E(Q_n)$ that leave $\mathcal{A}$, i.e. $\partial_e(\mathcal{A}) = \{xy \in E(Q_n) : x \in \mathcal{A}, y \notin \mathcal{A}\}$. A tight lower bound on $|\partial_e(\mathcal{A})|$ was given by Bernstein [4], Harper [16], Hart [17] and Lindsey [22], who proved that the extremal sets are initial segments of the 'binary ordering' on Q_n (see also Chapter 16 of [5]). In

E-mail addresses: keevash@maths.ox.ac.uk (P. Keevash), long@maths.ox.ac.uk (E. Long).
[1] Research supported in part by ERC Consolidator Grant 647678.

https://doi.org/10.1016/j.jcta.2017.11.005

particular, the following bound is tight when $|\mathcal{A}| = 2^d$ for some $d \in \mathbb{N}$ (take $\mathcal{A}$ to be the vertices of a d-dimensional subcube).

Theorem 1. *Every $\mathcal{A} \subset V(Q_n)$ satisfies $|\partial_e(\mathcal{A})| \geq |\mathcal{A}| \cdot \log_2 \left(2^n/|\mathcal{A}|\right)$.*

The next natural question is to understand the structure of subsets in the cube for which the inequality in Theorem 1 is close to an equality: must they be close to an extremal example? Indeed, for any problem in Extremal Combinatorics, the study of this 'stability' question often leads to a deeper understanding of the original question. The following stability version of Theorem 1 was obtained by Ellis [10], solving a conjecture of Bollobás, Leader and Riordan.

Theorem 2. *There is $\varepsilon_0 > 0$ so that for $0 \leq \varepsilon \leq \varepsilon_0$, the following holds. Suppose $\mathcal{A} \subset V(Q_n)$ with $|\partial_e(\mathcal{A})| \leq |\mathcal{A}|\left(\log_2(2^n/|\mathcal{A}|) + \varepsilon\right)$. Then there is a subcube C of Q_n with $|\mathcal{A} \triangle C| \leq \frac{3\varepsilon}{\log_2(\varepsilon^{-1})}|\mathcal{A}|$.*

This result has recently been refined by Ellis, Keller and Lifshitz [11] in the regime of extremely close approximation: they proved that a set $\mathcal{A}$ whose edge boundary is within an additive constant d of the minimum possible is $O(d)$-close to (an isomorphic copy of) the unique isoperimetric set.

It is generally more challenging to obtain any structural information in an extremal problem as the distance from the extremum increases. Kahn and Kalai [19] made a series of compelling conjectures around the theme of thresholds of monotone properties, some of which explore a potential connection with the stability problem for the edge-isoperimetric inequality.

One such conjecture, in a strengthened form proposed by Ellis in [10, Conjecture 3.3], suggests that small edge-boundary should imply some correlation with a large subcube. Concretely, for any $K > 0$ there are K', δ so that if $\mathcal{A} \subset \{0,1\}^n$ with $|\mathcal{A}| = \alpha 2^n$ and $|\partial_e(\mathcal{A})| \leq K|\mathcal{A}| \log_2 \alpha^{-1}$ then there should be a subcube $\mathcal{C}$ of $\{0,1\}^n$ of codimension at most $K' \log_2 \alpha^{-1}$ with $|\mathcal{A} \cap \mathcal{C}| \geq (1+\delta)\alpha|\mathcal{C}|$. Kahn and Kalai proposed this conjecture in the special case of monotone properties, but in the more general setting of biased measures on the cube. Kahn and Kalai further conjecture (see [19, Conjecture 4.1(b)], again for monotone properties) that such $\mathcal{A}$ must be close to a union of at most $\alpha^{-K'}$ cubes.

A weaker form of the latter conjecture follows from a result of Friedgut [13] in the 'dense' regime.

Theorem 3. *Let $K, \varepsilon > 0$. Suppose that $\mathcal{A} \subset \{0,1\}^n$ with $|\partial_e(\mathcal{A})| \leq K2^n$. Then there are disjoint cubes $\mathcal{C}_1, \ldots, \mathcal{C}_L$ with $|\mathcal{A} \triangle (\mathcal{C}_1 \cup \cdots \cup \mathcal{C}_L)| \leq \varepsilon 2^n$, where $L \leq L(K,\varepsilon) = 2^{2^{C(K/\varepsilon)}}$ for some constant $C > 0$.*

Remark. Friedgut actually proved that given such $\mathcal{A}$ there is set $S \subset [n]$ with $|S| \leq D := 2^{C(K/\varepsilon)}$ so that $\mathcal{A}$ is ε-approximated by disjoint cubes, all of whose fixed coordinate sets lie in S (often stated as $\mathcal{A}$ is ε-close to a D-Junta). Theorem 3 is an immediate consequence of this.

Our main result gives an analogue of Friedgut's theorem that also applies in the sparse regime.

Theorem 4. *Let $K, \varepsilon > 0$. Suppose that $\mathcal{A} \subset V(Q_n)$ with $|\partial_e(\mathcal{A})| \le |\mathcal{A}|\big(\log_2(2^n/|\mathcal{A}|) + K\big)$. Then there are disjoint cubes $\mathcal{C}_1, \dots, \mathcal{C}_L$ with $|\mathcal{A} \triangle (\mathcal{C}_1 \cup \dots \cup \mathcal{C}_L)| \le \varepsilon|\mathcal{A}|$, where $L \le L(K, \varepsilon) = 2^{2^{C(K/\varepsilon)^2}}$ for some constant $C > 0$.*

Remark. Letting $E(\mathcal{A})$ denote the set of edges in $\mathcal{A}$, i.e. $E(\mathcal{A}) = \{xx' \in E(Q_n) : x, x' \in \mathcal{A}\}$, Theorem 1 is equivalent to $|E(\mathcal{A})| \le |\mathcal{A}|(\log_2 |\mathcal{A}|)/2$. In this setting, Theorem 4 says that if $\mathcal{A} \subset \{0,1\}^n$ with $|E(\mathcal{A})| \ge |\mathcal{A}|(\log_2 |\mathcal{A}| - K)/2$ then $\mathcal{A}$ can be ε-approximated by at most $L(K, \varepsilon)$ subcubes. In this sense, Theorem 4 gives a 'dimension free' stability theorem.

We conclude this introduction with a brief outline of our argument, and how the paper will be organised to implement this. Most of the proof is geared towards showing that $\mathcal{A}$ has a coordinate of significant influence. This exploits the connection between edge-boundary and the influences of Boolean functions, which we will discuss in the next section, together with two inequalities (due to Talagrand and to Polyanskiy) that we will use in our proof. The starting point of our strategy is to choose an appropriate partition of the coordinate set, such that we maintain control on two important quantities: the constant K appearing in Theorem 4 (which we call the *isoperimetric excess* of $\mathcal{A}$) and a certain 'mutual information' quantity (in the sense of information theory). In section 3 we prove a partitioning lemma that will enable us to control both these quantities. The mutual information is then used in section 4 to show that $\mathcal{A}$ is 'product-like' in a certain sense. The control on the isoperimetric excess will be such that we can apply Ellis's theorem to approximate certain sections of $\mathcal{A}$ by cubes, provided that they are not too dense. To address the latter point (density of sections), in section 5 we apply Polyanskiy's hypercontractive inequality to show that $\mathcal{A}$ is typically not too dense in random subcubes (this result can be viewed as a sparse variant of the "It Ain't Over Till It's Over" conjecture, proved by Mossel, O'Donnell and Oleszkiewicz [23]). The results of the previous sections are combined in section 6 in finding a coordinate of significant influence. This is the main ingredient of an inductive proof of our main theorem, given in the final section.

2. Influences of Boolean functions

Edge boundary has a natural reformulation in terms of the analysis of Boolean functions, which is an active area in its own right, with many applications to other fields, including Social Choice and Computational Complexity; we refer the reader to the book [24] for an introduction. While our approach in this paper will be generally combinatorial rather than analytical, we will require some auxiliary results obtained by these analytic means.

To discuss this connection we require some notation and terminology. Given $f : \{0,1\}^n \to \mathbb{R}$, let $\mathbb{E}(f) = 2^{-n}\sum_{x\in\{0,1\}^n} f(x)$ and $\mathbb{V}ar(f) = \mathbb{E}(f-\mathbb{E}(f))^2$, the expectation and variance of f respectively. The function f is said to be Boolean if $f : \{0,1\}^n \to \{0,1\}$. Subsets of $V(Q_n)$ are naturally identified with Boolean functions, where a set $\mathcal{A} \subset \{0,1\}^n$ corresponds to the indicator function $1_{\mathcal{A}}$, with $1_{\mathcal{A}}(x) = 1$ if $x \in \mathcal{A}$ and 0 otherwise. Given $x \in \{0,1\}^n$ and $i \in [n]$ let $x \oplus e_i$ denote the element of $V(Q_n)$ obtained by changing the ith coordinate of x. The *influence of f in direction i* is defined as $I_i(f) := |\{x \in \{0,1\}^n : f(x) \neq f(x \oplus e_i)\}|/2^n$. The *influence of f*, denoted $I(f)$, is simply the sum of the individual influences, i.e. $I(f) = \sum_{i\in[n]} I_i(f)$. Thus $I_i(f)$ denotes the proportion of edges in direction i whose vertices disagree under f, and so $I(1_{\mathcal{A}}) = |\partial_e(\mathcal{A})|/2^{n-1}$ for all $\mathcal{A}$. Thus any statement regarding the edge boundary of $\mathcal{A}$ is equivalent to a statement on the influence of $1_{\mathcal{A}}$.

The notion of influence was first introduced by Ben-Or and Linial [3] in the context of social choice theory. They conjectured that any Boolean function $f : \{0,1\}^n \to \{0,1\}$ with $\mathbb{E}(f) = 1/2$ satisfies $\max_{i\in[n]} I_i(f) = \Omega\big((\log n)/n\big)$. This was later established by the fundamental KKL theorem of Kahn, Kalai and Linial [20], who proved that such f satisfy $\sum_{i\in[n]} I_i(f)^2 = \Omega\big((\log^2 n)/n\big)$. The following related inequality, that we will use in this paper, was given by Talagrand [28].

Theorem 5. *Any Boolean function $f : \{0,1\}^n \to \{0,1\}$ satisfies $\sum_{i\in[n]} \frac{I_i(f)}{1-\log_2 I_i(f)} \geq c \cdot \mathbb{V}ar(f)$, where $c > 0$ is a constant.*

An important tool in the proof of the KKL theorem (and many results in this area) is hypercontractivity of the noise operator, due to Bonami [6] and Beckner [2] (see also [24, Chapter 9]). (An alternative approach based on martingales and the log-Sobolev inequality for the cube was given by Falik and Samorodnitsky [12] and Rossignol [26].) Hypercontractivity will also be important for us in this paper, via the following estimate for spherical averages due to Polyanskiy [25].

For $p \in [1,\infty]$ let $L_p(\{0,1\}^n)$ denote the set of functions $f : \{0,1\}^n \to \mathbb{R}$ equipped with the norm $\|\cdot\|_p$ where $\|f\|_p = (2^{-n}\sum_{x\in\{0,1\}^n} |f(x)|^p)^{1/p}$. For $p = 2$, the space $L_2(\{0,1\}^n)$ also forms a Hilbert space equipped with the usual inner product, given by $\langle f, g\rangle = \frac{1}{2^n}\sum_{x\in\{0,1\}^n} f(x)g(x)$. Writing $d_H(x,\widetilde{x})$ for the Hamming distance between x and $\widetilde{x}$, let $S_\ell : L_2(\{0,1\}^n) \to L_2(\{0,1\}^n)$ denote the linear operator acting on $f \in L_2(\{0,1\}^n)$ pointwise by

$$S_\ell(f)(x) = \frac{1}{\binom{n}{\ell}} \sum_{\widetilde{x}:d_H(x,\widetilde{x})=\ell} f(\widetilde{x}). \tag{1}$$

This operator can be seen as a variant of the standard noise operator. In [25] Polyanskiy gave a hypercontractive estimate for S_ℓ, proving the following result (see Theorem 1, together with the remark following it).

Theorem 6. *Let $\ell \in [0, 0.15n] \cap \mathbb{N}$. Then for any $f : \{0,1\}^n \to \mathbb{R}$*

$$\|S_\ell(f)\|_2 \le 2^{1/2}\|f\|_{1+(1-2\ell/n)^2}.$$

While we do not need it for this paper, we should also remark that the threshold conjectures of Kahn and Kalai are intimately connected via Russo's lemma [27] to the large literature on influences under p-biased measures, which can be viewed as a weighted edge boundary (see e.g. [7,14,15,18,28]).

3. A partitioning lemma

In this section we establish some notation for partitions of the coordinate set and the corresponding sections of $\mathcal{A}$ that will be used throughout the paper. We also prove a lemma which shows that lower-dimensional sections of $\mathcal{A}$ tend to have smaller isoperimetric excess than $\mathcal{A}$, and also bounds a certain 'mutual information' that will be used in the next section to show that $\mathcal{A}$ has an approximate product structure.

Given $x = (x_i)_{i\in[n]} \in \{0,1\}^n$ and a set $I \subset [n]$ the *I-restriction* of x is the vector $x_I = (x_i)_{i\in I} \in \{0,1\}^I$. Given a partition $[n] = I_1 \cup \cdots \cup I_M$ and vectors $x^{(m)} \in \{0,1\}^{I_m}$ for all $m \in [M]$, let $x^{(1)} \circ \cdots \circ x^{(M)} \in \{0,1\}^n$ denote the *concatenation* of $x^{(1)}, \dots, x^{(M)}$, the unique vector $y \in \{0,1\}^n$ with $y_{I_m} = x^{(m)}$ for all $m \in [M]$. Note that $x = x_{I_1} \circ \cdots \circ x_{I_M}$ for all $x \in \{0,1\}^n$.

Let $\mathrm{Dir} : E(Q_n) \to [n]$ denote the function with $\mathrm{Dir}(xx') = i$ if $x' = x \oplus e_i$. For $I \subset [n]$ and $\mathcal{A} \subset \{0,1\}^n$ the set $\partial_e^I(\mathcal{A}) := \{xx' \in \partial_e(\mathcal{A}) : \mathrm{Dir}(xx') \in I\}$ is the *I-edge boundary* of $\mathcal{A}$. Clearly $\partial_e^{[n]}(\mathcal{A}) = \partial_e(\mathcal{A})$. Given a partition $I \cup J = [n]$ and $y \in \{0,1\}^J$ the *y-section of* $\mathcal{A}$ is the set

$$\mathcal{A}_y^I := \{z \in \{0,1\}^I : y \circ z \in \mathcal{A}\} \subset \{0,1\}^I.$$

The contributions from different sections give $\partial_e^I(\mathcal{A}) = \bigcup_{y\in\{0,1\}^J} \partial_e(\mathcal{A}_y^I)$.

Given a set $I \subset [n]$, with complement $J = [n] \setminus I$, let:

- $\alpha^I = (\alpha_y^I)$ be the probability distribution on $\{0,1\}^J$, with $\alpha_y^I = |\mathcal{A}_y^I|/|\mathcal{A}|$ for all $y \in \{0,1\}^J$;
- $|\partial_e^I(\mathcal{A}_y^I)| = |\mathcal{A}_y^I|\big(\log_2(2^{|I|}/|\mathcal{A}_y^I|) + K_y^I\big)$ for all $y \in \{0,1\}^I$, and set $K^I = \sum_y \alpha_y^I K_y^I$.

Note in particular that $\alpha^\emptyset$ is uniformly distributed on $\mathcal{A}$, i.e. $\alpha^\emptyset(x)$ is $1/|\mathcal{A}|$ if $x \in \mathcal{A}$ or 0 otherwise.

These section sizes can be naturally reformulated in terms of the following random variables. Consider selecting $x \in \mathcal{A}$ uniformly at random. Let $\mathbf{X}_i$ for $i \in [n]$ denote the random variable $\mathbf{X}_i(x) = x_i$. Write $\mathbf{X}_I = (\mathbf{X}_i)_{i \in I}$ for $I \subset [n]$. Then $\mathbf{X}_J$ satisfies $\mathbb{P}(\mathbf{X}_J = y) = \alpha_y^I$.

We will see that the entropy of these random variables appears naturally in the edge-isoperimetric problem. First we recall some standard definitions (for an introduction to information theory see the book [9]). Given a probability distribution $\mathbf{p} = (p_\omega)_{\omega \in \Omega}$ on a finite set Ω, the binary entropy of $\mathbf{p}$ is given by $H(\mathbf{p}) = -\sum_{\omega \in \Omega} p_\omega \log_2 p_\omega$. Given $\gamma \in [0,1]$ we will also sometimes write $H(\gamma)$ for the binary entropy of the probability distribution $\{\gamma, 1-\gamma\}$, i.e. $H(\gamma) = -\gamma \log_2 \gamma - (1-\gamma)\log_2(1-\gamma)$. The entropy of a random variable $\mathbf{X}$, denoted $H(\mathbf{X})$, taking values in Ω is the entropy of its probability mass function, i.e. $H(\mathbf{X}) = H(\mathbf{p})$ where $\mathbf{p} = (p_\omega)_{\omega \in \Omega}$ and $p_\omega = \mathbb{P}(\mathbf{X} = \omega)$.

We will use the following entropy inequality of Shearer (see [8] or Chapter 15 of [1]). We say that a family of sets $\mathcal{S} = \{S_m\}_{m \in [M]}$ forms a D-cover of $[n]$ if every $j \in [n]$ appears in at least D sets from $\mathcal{S}$.

Theorem 7. *Let $\mathbf{X} = (\mathbf{X}_i)_{i \in [n]}$ be a random variable taking values in a finite set Ω and let $\mathbf{X}_S$ denote the random variable $\mathbf{X}_S = (X_i)_{i \in S}$ for all $S \subset [n]$. Then given a D-cover $\mathcal{S}$ of $[n]$, we have $\sum_{S \in \mathcal{S}} H(\mathbf{X}_S) \geq D \cdot H(\mathbf{X})$.*

With this notation in place, we can state the partitioning lemma.

Lemma 8. *Let $\mathcal{A} \subset \{0,1\}^n$ with $|\partial_e(\mathcal{A})| \leq |\mathcal{A}|\big(\log_2(2^n/|\mathcal{A}|) + K\big)$. Suppose $I_1 \cup \cdots \cup I_M = [n]$ is a partition. Then*

(i) $\sum_{m \in [M]} H(\alpha^{I_m}) - (M-1)H(\alpha^{\emptyset}) \leq K$;
(ii) $\sum_{m \in [M]} K^{I_m} \leq K$.

Proof. As $\mathcal{A}_y^I \subset \{0,1\}^I$, by Theorem 1 we have $|\partial_e^I(\mathcal{A}_y^I)| = |\mathcal{A}_y^I|\big(\log\left(2^{|I|}/|\mathcal{A}_y^I|\right) + K_y^I\big)$ with $K_y^I \geq 0$. Expanding this expression, we find

$$|\partial_e^I(\mathcal{A}_y^I)| = \alpha_y^I |\mathcal{A}| \log_2\left(2^{|I|}/|\mathcal{A}|\right) - \alpha_y^I |\mathcal{A}| \log_2(\alpha_y^I) + \alpha_y^I K_y^I |\mathcal{A}|.$$

Summing over $y \in \{0,1\}^J$, as $\sum_{y \in \{0,1\}^J} \alpha_y^I = 1$ we obtain

$$|\partial_e^I(\mathcal{A})| = \sum_{y \in \{0,1\}^J} |\partial_e^I(\mathcal{A}_y^I)| = |\mathcal{A}| \log_2\left(2^{|I|}/|\mathcal{A}|\right) + |\mathcal{A}|\big(H(\boldsymbol{\alpha}^I) + K^I\big).$$

Apply this equality for $I_1, \ldots, I_M$. Using $|\partial_e(\mathcal{A})| = \sum_{m \in [M]} |\partial_e^{I_m}(\mathcal{A})|$, we obtain

$$|\partial_e(\mathcal{A})| = |\mathcal{A}| \log_2\left(2^n/|\mathcal{A}|\right) + |\mathcal{A}|\Big(\sum_{m \in [M]} H(\boldsymbol{\alpha}^{I_m}) - (M-1)\log_2|\mathcal{A}| + \sum_{m \in [M]} K^{I_m}\Big).$$

Since $\log_2|\mathcal{A}| = H(\alpha^\emptyset)$ and $|\partial_e(\mathcal{A})| \le |\mathcal{A}|\big(\log_2(2^n/|\mathcal{A}|) + K\big)$ this gives

$$\sum_{m\in[M]} H(\boldsymbol{\alpha}^{I_m}) - (M-1)H(\alpha^\emptyset) + \sum_{m\in[M]} K^{I_m} \le K. \tag{2}$$

Both (i) and (ii) now follow from (2). Indeed, (i) holds since $K^{I_m} \ge 0$ for all $m \in [M]$. To see (ii), by (2) it suffices to show $\sum_{m\in[M]} H(\alpha^{I_m}) \ge (M-1)H(\alpha^\emptyset)$. To see this, consider selecting $x \in \mathcal{A}$ uniformly at random, and for all $i \in [n]$ let $\mathbf{X}_i$ denote the random variable given by $\mathbf{X}_i(x) = x_i$. For all $I' \subset [n]$ also let $\mathbf{X}_{I'} = (\mathbf{X}_i)_{i\in I'}$. Then $\mathbf{X}_{J_m}$ satisfies $\mathbb{P}(\mathbf{X}_{J_m} = y_m) = \alpha^{I_m}_{y_m}$ for all $m \in [M]$, giving $H(\mathbf{X}_{J_m}) = H(\alpha^{I_m})$. Furthermore $H(\mathbf{X}_{[n]}) = \log_2|\mathcal{A}|$. However $\{J_m\}_{m\in[M]}$ forms a $(M-1)$-cover for $[n]$. Theorem 7 therefore gives $\sum_{m\in[M]} H(\alpha^{I_m}) = \sum_{m\in[M]} H(\mathbf{X}_m) \ge (M-1)H(\mathbf{X}_{[n]}) = (M-1)H(\alpha^\emptyset)$. This completes the proof of the lemma. □

4. Approximate product structure

In this section we will use Lemma 8 (i) with $M = 2$ to show that if $\mathcal{A}$ has small isoperimetric excess then it has an approximate product structure with respect to any partition $[n] = I \cup J$, in the sense that for most elements $x \in A$ the product of 'orthogonal sections' $|\mathcal{A}^J_{x_I}||\mathcal{A}^I_{x_J}|$ is comparable with $|\mathcal{A}|$.

Recall that for $y \in \{0,1\}^J$ we let $\alpha^I_y = |\mathcal{A}^I_y|/|\mathcal{A}|$, for $z \in \{0,1\}^I$ we let $\alpha^J_z = |\mathcal{A}^J_z|/|\mathcal{A}|$, and $\alpha^\emptyset_x = 1/|\mathcal{A}|$ for all $x \in \mathcal{A}$ and 0 otherwise. We also let α^I, α^J and $\alpha^\emptyset$ denote the corresponding probability distributions. The quantity $H(\alpha^I) + H(\alpha^J) - H(\alpha^\emptyset)$ can be viewed as the mutual information of the random variables $\mathbf{X}_I$ and $\mathbf{X}_J$ considered in the previous section. If the mutual information were zero, then the variables would be independent, and $\mathcal{A}$ would have a product structure. The following lemma can be viewed as a stability version of this observation.

Lemma 9. *Let $K, \varepsilon > 0$ and suppose $H(\alpha^I) + H(\alpha^J) - H(\alpha^\emptyset) \le K$. Then for at least $(1-\varepsilon)|\mathcal{A}|$ elements $x \in \mathcal{A}$ we have $|\mathcal{A}^J_{x_I}||\mathcal{A}^I_{x_J}| \ge |\mathcal{A}|/(e \cdot 2^{K/\varepsilon})$.*

Proof. Write $b_x = \alpha^\emptyset_x/(\alpha^I_{x_J}\alpha^J_{x_I})$ and let $f(t) := t\log_e t + 1 - t$. We claim that

$$\begin{aligned}\log_e 2 \times \big(H(\alpha^I) + H(\alpha^J) - H(\alpha^\emptyset)\big) &= \sum_{x\in\{0,1\}^{[n]}} \alpha^I_{x_J}\alpha^J_{x_I}\big(b_x \log_e b_x\big) \\ &= \sum_{x\in\{0,1\}^{[n]}} \alpha^I_{x_J}\alpha^J_{x_I} f(b_x).\end{aligned}$$

To see this, first note that

$$H(\alpha^J) = -\sum_{y\in\{0,1\}^J} \alpha^I_y \log_2\big(\alpha^I_y\big) = -\sum_{x\in\{0,1\}^\emptyset} \alpha^\emptyset_x \log_2\big(\alpha^I_{x_J}\big).$$

Using the analogous expressions for $H(\alpha^I)$ and $H(\alpha^\emptyset)$, we obtain

$$H(\alpha^I) + H(\alpha^J) - H(\alpha^\emptyset) = \sum_{x\in\{0,1\}^{[n]}} \alpha_x^\emptyset \log_2(b_x) = \sum_{x\in\{0,1\}^{[n]}} \alpha^I_{x_J}\alpha^J_{x_I} b_x \log_2 b_x.$$

This gives the first equality of the claim. The second follows as

$$\sum_{x\in\{0,1\}^{[n]}} \alpha^I_{x_J}\alpha^J_{x_I} = \Big(\sum_{z\in\{0,1\}^I} \alpha^J_z\Big)\Big(\sum_{y\in\{0,1\}^J} \alpha^I_y\Big) = 1, \text{ and}$$

$$\sum_{x\in\{0,1\}^{[n]}} (\alpha^I_{x_J}\alpha^J_{x_I}) b_x = \sum_{x\in\{0,1\}^{[n]}} \alpha^\emptyset_x = 1.$$

Now consider $\mathcal{A}_D := \{x \in \mathcal{A} : b_x \geq D\} \subset \mathcal{A}$ for $D > 1$. We have

$$\begin{aligned}\frac{|\mathcal{A}_D|}{|\mathcal{A}|} &= \sum_{x\in\mathcal{A}_D} \alpha^\emptyset_x = \sum_{x\in\mathcal{A}:b_x\geq D} \alpha^I_{x_J}\alpha^J_{x_I} b_x = \sum_{x\in\mathcal{A}:b_x\geq D} \alpha^I_x \alpha^J_{x_I} f(b_x)\Big(\frac{b_x}{f(b_x)}\Big)\\ &\leq \Big(\frac{D}{f(D)}\Big) \times \sum_{x\in\mathcal{A}} \alpha^I_x\alpha^J_{x_I} f(b_x) = \Big(\frac{D}{f(D)}\Big) \times \log_e 2 \times \big(H(\alpha^I) + H(\alpha^J) - H(\alpha^\emptyset)\big)\\ &\leq \frac{(\log_e 2)K}{\log_e D - 1}.\end{aligned}$$

The first inequality holds as $f(t) \geq 0$ for $t > 0$ and $g(t) := f(t)/t$ satisfies $g'(t) = t^{-1} - t^{-2} \geq 0$ for $t \geq 1$. The following equality holds by the claim, and then the final inequality holds since $D/f(D) \leq 1/(\log_e D - 1)$ and $H(\alpha^I) + H(\alpha^J) - H(\alpha^\emptyset) \leq K$ by Lemma 8(i). Setting $D = e2^{K/\varepsilon}$ gives $|\mathcal{A}_D| \leq \varepsilon|\mathcal{A}|$. Since $|\mathcal{A}|/(|\mathcal{A}^J_{x_I}||\mathcal{A}^I_{x_J}|) = b_x \leq D$ for all $x \notin \mathcal{A}_D$, this completes the proof. □

5. Sparse sections

In this section we prove the following result, which shows that if $\mathcal{A} \subset V(Q_n)$ is sparse, then this is also true of typical sections of $\mathcal{A}$. Another way to interpret the result (which is also convenient for the proof) is to consider a random element x of $\mathcal{A}$, reveal all but d of its coordinates, then sample a new element of $\widetilde{x} \in \{0,1\}^n$ that agrees with the revealed coordinates. Then there is typically still some uncertainty as to whether $\widetilde{x}$ is in $\mathcal{A}$ (It Ain't Over Till It's Over).

Lemma 10. *Let $\mathcal{A} \subset V(Q_n)$ with $|\mathcal{A}| = \alpha 2^n$ and $d \in \mathbb{N}$ with $d \leq 0.15n$. Independently select:*

- *$x \in \mathcal{A}$ uniformly at random,*
- *$I \subset [n]$ with $|I| = d$, uniformly at random.*

Then $\mathbb{E}_{x,I}(|\mathcal{A}^I_{x_J}|) \leq 2\alpha^{d/8n}2^d$, where $J = [n] \setminus I$.

Note that the exponent of α is tight up to a constant factor (for example, when $\mathcal{A}$ is a subcube).

Proof. Given x and I, we also select $\widetilde{x} \in \{0,1\}^n$ uniformly at random subject to $\widetilde{x}_J = x_J$. Note that

$$\mathbb{E}_{x,I}\big(|\mathcal{A}^I_{x_J}|\big) = \mathbb{E}_{x,I}\big(\mathbb{P}(\widetilde{x} \in \mathcal{A}|x,I)\cdot 2^d\big) = \mathbb{P}(\widetilde{x} \in \mathcal{A})\cdot 2^d.$$

The lemma is thus equivalent to showing that $\mathbb{P}(\widetilde{x} \in \mathcal{A}) \leq 2\alpha^{d/8n}$.

To see this, we note that given $w \in \mathcal{A}$ and $\widetilde{w} \in \{0,1\}^n$ with $d_H(w,\widetilde{w}) = \ell$, we have

$$\begin{aligned}\mathbb{P}\big(\widetilde{x} = \widetilde{w}|x = w\big) &= \mathbb{P}\big(d_H(x,\widetilde{x}) = \ell|x = w\big)\cdot\mathbb{P}\big(\widetilde{x} = \widetilde{w}|x = w, d_H(\widetilde{x},x) = \ell\big)\\ &= \begin{cases} 2^{-d}\binom{d}{\ell}.\binom{n}{\ell}^{-1} & \text{if } \ell \leq d;\\ 0 & \text{otherwise.}\end{cases}\end{aligned}$$

Let S denote the linear operator $S = 2^{-d}\sum_{\ell=0}^{d}\binom{d}{\ell}S_\ell$, with S_ℓ as in (1). Let $1_{\mathcal{A}}$ denote the indicator function of $\mathcal{A}$. Then for $w \in \mathcal{A}$ we have

$$S\big(1_{\mathcal{A}}\big)(w) = \left(\frac{1}{2^d}\sum_{\ell=0}^{d}\binom{d}{\ell}S_\ell\right)1_{\mathcal{A}}(w) = \mathbb{P}\big(\widetilde{x} \in \mathcal{A}|x = w\big).$$

We deduce that $\mathbb{P}(\widetilde{x} \in \mathcal{A}) = \sum_{w\in\{0,1\}^n}\mathbb{P}(x = w)\mathbb{P}\big(\widetilde{x} = \widetilde{w}|x = w\big) = \langle\alpha^{-1}1_{\mathcal{A}}, S1_{\mathcal{A}}\rangle$. Separating S and using the Cauchy–Schwarz inequality gives

$$\mathbb{P}(\widetilde{x} \in \mathcal{A}) \leq \alpha^{-1}\sum_{\ell=0}^{d}\binom{d}{\ell}2^{-d}\|1_{\mathcal{A}}\|_2\|S_\ell 1_{\mathcal{A}}\|_2. \tag{3}$$

However, by Theorem 6 we have $\|S_\ell(1_{\mathcal{A}})\|_2 \leq 2\|\mathbf{1}_{\mathcal{A}}\|_{1+(1-2\ell/n)^2}$ for $\ell \leq d$. As

$$\begin{aligned}(1+(1-2\ell/n)^2)^{-1} - 1/2 = (2\ell/n - 2\ell^2/n^2)(1+(1-2\ell/n)^2)^{-1} &\geq \ell/n - \ell^2/n^2\\ &\geq \ell/2n,\end{aligned}$$

since $\ell \leq n/2$, this gives

$$\|1_{\mathcal{A}}\|_2\|S_\ell 1_{\mathcal{A}}\|_2 \leq 2\|1_{\mathcal{A}}\|_2\|1_{\mathcal{A}}\|_{1+(1-2\ell/n)^2} = 2\alpha^{\frac{1}{2}+(1+(1-2\ell/n)^2)^{-1}} \leq 2\alpha^{1+\ell/2n}.$$

Combined with (3) this gives

$$\mathbb{P}(\widetilde{x} \in \mathcal{A}) \leq \frac{2}{2^d}\sum_{\ell=0}^{d}\binom{d}{\ell}\alpha^{\ell/2n} = 2\left(\frac{1+\alpha^{1/2n}}{2}\right)^d.$$

To simplify, let $\alpha = e^{-L}$. As $e^\gamma \le 1 + \gamma/2$ for $\gamma \in [-1, 0]$ and $1 + \gamma \le e^\gamma$ for all $\gamma \in \mathbb{R}$, we find

$$\left(\frac{1+\alpha^{1/2n}}{2}\right)^d = \left(\frac{1+e^{-L/2n}}{2}\right)^d \le \left(\frac{1+(1-L/4n)}{2}\right)^d$$
$$= \left(1 - \frac{L}{8n}\right)^d \le e^{-dL/8n} = \alpha^{d/8n}.$$

Therefore $\mathbb{P}(\widetilde{x} \in \mathcal{A}) \le 2\alpha^{d/8n}$, completing the proof of the lemma. □

Remark. Lemma 10 may be seen as a variant of 'small set expansion in the noisy hypercube' (see Section 9, [24]).

6. Finding a coordinate of large influence

In this section we prove that if $\mathcal{A} \subset \{0, 1\}^n$ has small isoperimetric excess and is not close to being the whole cube then there is a coordinate of large influence.

Theorem 11. *Let $\mathcal{A} \subset V(Q_n)$ with $|\mathcal{A}| \le \left(\frac{7}{8}\right)2^n$ and $|\partial_e(\mathcal{A})| \le |\mathcal{A}|\big(\log_2(2^n/|\mathcal{A}|) + K\big)$. Then $\max_{i\in[n]} I_i(\mathbf{1}_{\mathcal{A}}) \ge 2^{-C(K+1)^2}|\mathcal{A}|/2^n$, for some constant $C > 0$.*

Proof. Let $|\mathcal{A}| = \alpha 2^n$, where $\alpha \le 7/8$, and let $\beta_i = I_i(\mathbf{1}_{\mathcal{A}})$ for all $i \in [n]$ and $\beta = \max_{i\in[n]} \beta_i$. We will also fix a number of parameters to be used in the proof. Set $c_0 = \min\{\varepsilon_0, 1/8\}$, where ε_0 is as in Theorem 2. Also set $C_1 = 2^{12}/c_0$, $\delta = c_0/(32(K + 1))$ and $M = \lceil 1/4\delta \rceil$. Lastly set $C = 32C_1/c$, with c as in Theorem 5.

We first consider the case when $\alpha \ge 2^{-C_1(K+1)^2}$, where the result follows from Talagrand's inequality. Indeed, in this case $|\partial_e(\mathcal{A})| \le |\mathcal{A}|\big(C_1(K+1)^2+K\big) \le 2C_1(K+1)^2|\mathcal{A}|$. As $\alpha \le 7/8$, we have $\mathbb{V}ar(\mathbf{1}_{\mathcal{A}}) \ge |\mathcal{A}|/2^{n+3}$, so Theorem 5 gives

$$\frac{c|\mathcal{A}|}{2^{n+3}} \le c \cdot \mathbb{V}ar(\mathbf{1}_{\mathcal{A}}) \le \sum_{i\in[n]} \frac{\beta_i}{1 - \log_2 \beta_i} \le \frac{I(\mathbf{1}_{\mathcal{A}})}{\log_2(\beta^{-1})} = \frac{|\partial_e(\mathcal{A})|}{2^{n-1}\log_2(\beta^{-1})}$$
$$\le \frac{2C_1(K+1)^2|\mathcal{A}|}{2^{n-1}\log_2(\beta^{-1})}.$$

Rearranging $\beta \ge 2^{-32C_1(K+1)^2/c} = 2^{-C(K+1)^2}$, as required.

It remains to consider the case $\alpha \le 2^{-C_1(K+1)^2}$. We start by giving an overview of the argument in this case. We will find a partition $I \cup J$ of $[n]$ so that for many elements x of $\mathcal{A}$ the I-section is sparse and has small isoperimetric excess, and the product of orthogonal sections through x is comparable with $\mathcal{A}$. We can then apply Ellis' theorem to find a subcube $\mathcal{C} \subset \{0, 1\}^I$ such that many elements of $\mathcal{A}$ have an I-restriction in $\mathcal{C}$. Finally, we show that one of the coordinates that is influential for $\mathcal{C}$ must also be influential for $\mathcal{A}$.

To begin, select a partition $[n] = I_0 \cup I_1 \cup \ldots \cup I_M$ uniformly at random, with $|I_m| = d = \lceil \delta n \rceil$ for all $m \in [M]$ and $|I_0| = n - Md$. (This is possible as $n \geq C_1(K+1)^2$, $\delta n \geq C_1 c_0 (K+1)/32 \geq 1$ and $Md \leq (2\delta)^{-1}(2\delta n) \leq n$.) Write $J_m = [n] \setminus I_m$ for all $m \in [M]$. We say I_m is *controlled* if

$$\left|\left\{x \in \mathcal{A} : |\mathcal{A}^{I_m}_{x_{J_m}}| \leq 16\alpha^{d/8n}2^d\right\}\right| \geq 3|\mathcal{A}|/4.$$

By Lemma 10 we have

$$\frac{1}{4} \times \mathbb{P}(I_m \text{ is not controlled}) \times 16\alpha^{d/8n}2^d \leq \mathbb{E}_{x,I_m}(|\mathcal{A}^{I_m}_{x_{J_m}}|) \leq 2\alpha^{d/8n}2^d.$$

This gives $\mathbb{P}(I_m \text{ is controlled}) \geq 1/2$. Letting $S_1 = \{m \in [M] : I_m \text{ is controlled}\}$, we find that $\mathbb{E}(|S_1|) \geq M/2$. Fix a choice of $I_1, \ldots, I_M$ such that $|S_1| \geq M/2$.

Now set $S_2 = \{m \in [M] : K^{I_m} \leq 4K/M\}$. As $\sum_{m\in[M]} K^{I_m} \leq K$ by Lemma 8 (*ii*), Markov's inequality gives $|S_2| \geq 3M/4$. Combined with the previous paragraph, this gives $S_1 \cap S_2 \neq \emptyset$. Fix $m \in S_1 \cap S_2$ and take $I = I_m$ and $J = J_m$.

To proceed we now consider the following subsets of $\mathcal{A}$:

(i) $\mathcal{A}_1 = \left\{x \in \mathcal{A} : |\mathcal{A}^I_{x_J}| \leq 16\alpha^{d/8n}2^d\right\}$;
(ii) $\mathcal{A}_2 = \left\{x \in \mathcal{A} : K^I_{x_J} \leq 16K/M\right\}$;
(iii) $\mathcal{A}_3 = \left\{x \in \mathcal{A} : |\mathcal{A}^I_{x_J}||\mathcal{A}^J_{x_I}| \geq |\mathcal{A}|/(e2^{4K})\right\}$.

Further let $\mathcal{B} = \mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3$. We claim that $|\mathcal{B}| \geq |\mathcal{A}|/4$. To see this, first note that $|\mathcal{A}_1| \geq 3|\mathcal{A}|/4$ as $m \in S_1$. Also, since $\mathbb{E}_{x\sim\mathcal{A}} K^I_{x_J} = K^I \leq 4K/M$ as $m \in S_2$, by Markov's inequality $|\mathcal{A}_2| \geq 3|\mathcal{A}|/4$. Lastly, since $H(\alpha^I) + H(\alpha^J) - H(\alpha^\emptyset) \leq K$ from Lemma 8 (i), applying Lemma 9 with $\varepsilon = 1/4$ gives $|\mathcal{A}_3| \geq 3|\mathcal{A}|/4$. Therefore $|\mathcal{B}| \geq |\mathcal{A}|/4$ as claimed.

We will now show that for some $x \in \mathcal{B}$ the set $\mathcal{A}^I_{x_J} \in \{0,1\}^I$ has a coordinate of large influence which also gives large influence for $\mathcal{A}$. To see this, note that partitioning $\mathcal{B}$ over the I-sections gives

$$\sum_{y\in\{0,1\}^J} |\mathcal{B}^I_y| = |\mathcal{B}| \geq \frac{|\mathcal{A}|}{4} = \sum_{y\in\{0,1\}^J} \frac{|\mathcal{A}^I_y|}{4}.$$

Therefore $|\mathcal{B}^I_{y_0}| \geq |\mathcal{A}^I_{y_0}|/4 > 0$ for some $y_0 \in \{0,1\}^J$. For any $z \in \mathcal{B}^I_{y_0}$ we have $x = y_0 \circ z \in \mathcal{B} \subset \mathcal{A}_2$ which gives $K^I_{y_0} = K^I_{x_J} \leq 16K/M \leq c_0$, i.e.

$$|\partial^I_e(\mathcal{A}^I_{y_0})| \leq |\mathcal{A}^I_{y_0}|\left(\log_2(2^{|I|}/|\mathcal{A}^I_{y_0}|) + c_0\right).$$

Theorem 2 therefore gives $|\mathcal{A}^I_{y_0} \triangle \mathcal{C}| \leq 3c_0|\mathcal{A}^I_{y_0}|/\log_2(c_0^{-1}) \leq |\mathcal{A}^I_{y_0}|/8$ for some subcube $\mathcal{C} \subset \{0,1\}^I$. As $\mathcal{B}^I_{y_0} \subset \mathcal{A}^I_{y_0}$ and $|\mathcal{B}^I_{y_0}| \geq |\mathcal{A}^I_{y_0}|/4$, this gives $|\mathcal{B}^I_{y_0} \cap \mathcal{C}| \geq |\mathcal{A}^I_{y_0}|/8$.

Set $\mathcal{D} = \{x \in \mathcal{A} : x_I \in \mathcal{C}\}$. Note that $\mathcal{D}$ is insensitive to coordinates in J, in the sense that if $x \in \mathcal{D}$ and $\widetilde{x} \in \mathcal{A}$ with $x_I = \widetilde{x}_I$ then $\widetilde{x} \in \mathcal{D}$. Therefore $\mathcal{D} \supset \bigcup_{z\in\mathcal{C}}\{y \circ z : y \in \mathcal{A}^J_z\}$

and in particular $|\mathcal{D}| \geq \sum_{z\in\mathcal{C}} |\mathcal{A}_z^J|$. However for each $z \in \mathcal{B}_{y_0}^I \cap \mathcal{C} \subset \mathcal{A}_{y_0}^I$ we have $x := y_0 \circ z \in \mathcal{A}_3$ and so $|\mathcal{A}_z^I||\mathcal{A}_{y_0}^J| = |\mathcal{A}_{x_J}^I||\mathcal{A}_{x_I}^J| \geq |\mathcal{A}|/(e2^{4K})$. Combined, this gives

$$|\mathcal{D}| \geq \sum_{z\in\mathcal{B}_{y_0}^I\cap\mathcal{C}} |\mathcal{A}_z^J| \geq \sum_{z\in\mathcal{B}_{y_0}^I\cap\mathcal{C}} \frac{|\mathcal{A}|}{e2^{4K}|\mathcal{A}_{y_0}^I|} = |\mathcal{B}_{y_0}^I \cap \mathcal{C}| \cdot \frac{|\mathcal{A}|}{e2^{4K}|\mathcal{A}_{y_0}^I|}$$
$$\geq \frac{|\mathcal{A}_{y_0}^I|}{8} \cdot \frac{|\mathcal{A}|}{e2^{4K}|\mathcal{A}_{y_0}^I|} \geq \frac{|\mathcal{A}|}{2^{4K+5}}. \qquad (4)$$

Thus a large proportion of elements $x \in \mathcal{A}$ satisfy $x_I \in \mathcal{C}$.

We will now show that one of the coordinates that are influential for $\mathcal{C} \subset \{0,1\}^I$ must also be influential for $\mathcal{A}$. To see this, as $\mathcal{C} \subset \{0,1\}^I$ is a subcube there is $T = \{i_1,\dots,i_t\} \subset I$ and $z_0 \in \{0,1\}^T$ with $\mathcal{C} = \{z \in \{0,1\}^I : z_T = z_0\}$ and $\log_2|\mathcal{C}| = d - t = |I \setminus T|$. As $|\mathcal{C}| \leq |\mathcal{A}_{y_0}^I| + |\mathcal{A}_{y_0}^I \triangle \mathcal{C}| \leq 2|\mathcal{A}_{y_0}^I| \leq 2^5\alpha^{d/8n}2^d$, we find

$$t = d - \log_2|\mathcal{C}| \geq \frac{d}{8n}\log_2(\alpha^{-1}) - 5 > \frac{\delta}{8}\log_2(\alpha^{-1}) - 5$$
$$\geq \frac{c_0C_1(K+1)}{2^8} - 5 \geq 4K + 7. \qquad (5)$$

Here we used $\alpha \leq 2^{-C_1(K+1)^2}$, $\delta = c_0/(32(K+1))$ and $C_1 = 2^{12}/c_0$.

Finally, suppose for a contradiction that $\beta_i \leq |\mathcal{A}|/(2^{4K+6}2^n)$ for all $i \in T$. For $0 \leq \ell \leq t$ let

$$\mathcal{A}_\ell = \mathcal{A} \cap \left\{x \in \{0,1\}^n : x_{\{i_{\ell+1},\dots,i_t\}} = z_{\{i_{\ell+1},\dots,i_t\}}\right\}.$$

Clearly $\mathcal{D} = \mathcal{A}_0 \subset \mathcal{A}_1 \subset \cdots \subset \mathcal{A}_t = \mathcal{A}$. As $\beta_{i_\ell} \leq |\mathcal{A}|/2^{4K+6}2^n$, we have $|\mathcal{A}_\ell| \geq 2|\mathcal{A}_{\ell-1}| - |\mathcal{A}|/2^{4K+6}$. Equivalently $|\mathcal{A}_\ell| - |\mathcal{A}|/2^{4K+6} \geq 2\big(|\mathcal{A}_{\ell-1}| - |\mathcal{A}|/2^{4K+6}\big)$. Taking $\ell = t$, we find

$$|\mathcal{A}| > |\mathcal{A}_t| - |\mathcal{A}|/2^{4K+6} \geq 2^t(|\mathcal{A}_0| - |\mathcal{A}|/2^{4K+6}) \geq 2^{t-4K-6}|\mathcal{A}|.$$

The final inequality holds since $|\mathcal{A}_0| = |\mathcal{D}| \geq |\mathcal{A}|/2^{4K+5}$ by (4). However, as $t \geq 4K + 7$ from (5), this is a contradiction, and so, as $C_1 \geq 6$ we have $\beta \geq \max_{i\in T}\beta_i \geq 2^{-4K-6}|\mathcal{A}|/2^n \geq 2^{-C_1(K+1)^2}|\mathcal{A}|/2^n$, as claimed. □

7. Almost isoperimetric sets are close to a union of cubes

With Theorem 11 in hand, we can now prove Theorem 4.

Proof of Theorem 4. To begin, let $C_1 \geq 1$ be the constant given by Theorem 3 and let $C_2 \geq 1$ be the constant given in Theorem 11. Set $C = \max\{6C_1, 8C_2\}$ and let $g : \mathbb{R} \to \mathbb{R}$ be the function $g(x) = 2^{2^{C(x+1)^2}}$. We will show that for all $K \geq 0$ and $\varepsilon > 0$, given a set

$\mathcal{A} \subset \{0,1\}^n$ with $|\partial_e(\mathcal{A})| \le |\mathcal{A}|\big(\log_2(2^n/|\mathcal{A}|) + K\big)$ there are disjoint cubes $\mathcal{C}_1, \dots, \mathcal{C}_L$ with $|\mathcal{A} \triangle (\mathcal{C}_1 \cup \cdots \cup \mathcal{C}_L)| \le \varepsilon|\mathcal{A}|$ and $L \le g(K/\varepsilon)$.

Before beginning the proof, we note that this seemingly weaker bound on L implies the bound stated in Theorem 4. Indeed, if $\varepsilon > 1$ then $\mathcal{A}$ can be ε-approximated 0 subcubes. For $\varepsilon \le 1$, if $K \le K/\varepsilon < c_0 := \max\{\varepsilon_0, 1/8\}$ then $\mathcal{A}$ can ε-approximated by $1 \le L(K/\varepsilon)$ subcube by Theorem 2. Otherwise, $1 \le c_0^{-1} K/\varepsilon$ and $g(K/\varepsilon) \le L(K/\varepsilon) := 2^{2^{4Cc_0^{-2}(K/\varepsilon)^2}}$.

We will prove the result by induction on $|\mathcal{A}| + n$. Clearly it holds when $|\mathcal{A}| = 1$ for all n. We also claim that the result holds when $|\mathcal{A}| \ge \big(\frac{7}{8}\big)2^n$. Indeed, in this case we consider $\mathcal{A}^c = \{0,1\}^n \setminus \mathcal{A}$, and write $|\mathcal{A}^c| = \alpha 2^n$ so that $\alpha \le \frac{1}{8}$. Using $1 - x \ge 2^{-2x}$ for $x \in [0, 1/8]$ and applying Theorem 1 to $\mathcal{A}^c$ we find

$$\begin{aligned} 2^n\big(\log_2(1-\alpha)^{-1} + K\big) \ge |\partial_e(\mathcal{A})| = |\partial_e(\mathcal{A}^c)| &\ge \alpha 2^n \log_2(\alpha^{-1}) \ge 3\alpha 2^n \\ &\ge \Big(\frac{3}{2}\Big) 2^n \log_2(1-\alpha)^{-1}. \end{aligned}$$

Thus $K \ge \frac{1}{2}\log_2(2^n/|\mathcal{A}|)$ and $|\partial_e(\mathcal{A})| \le 3K2^n$. By Theorem 3 there are disjoint subcubes $\mathcal{C}_1, \dots, \mathcal{C}_L$ such that $|\mathcal{A} \triangle (\mathcal{C}_1 \cup \cdots \cup \mathcal{C}_L)| \le \big(\frac{\varepsilon}{2}\big)2^n \le \varepsilon|\mathcal{A}|$ with $L \le 2^{2^{C_1(3K/(\varepsilon/2))}} \le g(K/\varepsilon)$, as desired.

Let $\mathcal{A} \subset \{0,1\}^n$ and assume that the result holds for smaller values of $|\mathcal{A}| + n$, and that $|\mathcal{A}| \le \big(\frac{7}{8}\big)2^n$. We can apply Theorem 11 to $\mathcal{A}$ find a coordinate $j \in [n]$ with $I_j(\mathbf{1}_{\mathcal{A}}) = b_j|\mathcal{A}|/2^n$ where $b_j \ge c(K) := 2^{-C_2(K+1)^2}$. Without loss of generality $j = n$. Set $\mathcal{A}^- = \mathcal{A}_0^{[n-1]}$ and $\mathcal{A}^+ = \mathcal{A}_1^{[n-1]}$ and let $\gamma \in [0,1]$ with $|\mathcal{A}^-| = \gamma|\mathcal{A}|$ and $|\mathcal{A}^+| = (1-\gamma)|\mathcal{A}|$. By symmetry we may assume $\gamma \le 1/2$. By Theorem 1, there are $K^-, K^+ \ge 0$ with

$$\begin{aligned} |\partial_e^{[n-1]}(\mathcal{A}^-)| &= |\mathcal{A}^-|\big(\log_2(2^{n-1}/|\mathcal{A}^-|) + K^-\big), \\ |\partial_e^{[n-1]}(\mathcal{A}^+)| &= |\mathcal{A}^+|\big(\log_2(2^{n-1}/|\mathcal{A}^+|) + K^+\big). \end{aligned}$$

Expanding these expressions gives

$$\begin{aligned} |\partial_e^{[n-1]}(\mathcal{A}^-)| &= \gamma|\mathcal{A}|\log_2(2^n/|\mathcal{A}|) + \gamma|\mathcal{A}|(\log_2\gamma^{-1} - 1 + K^-), \\ |\partial_e^{[n-1]}(\mathcal{A}^+)| &= (1-\gamma)|\mathcal{A}|\log_2(2^n/|\mathcal{A}|) + (1-\gamma)|\mathcal{A}|(\log_2(1-\gamma)^{-1} - 1 + K^+). \end{aligned}$$

Combining these identities together with the contribution $b_n|\mathcal{A}|$ from the edges in direction n gives

$$\begin{aligned} |\partial_e(\mathcal{A})| &= |\partial_e^{[n-1]}(\mathcal{A}^-)| + |\partial_e^{[n-1]}(\mathcal{A}^+)| + |\mathcal{A}^- \triangle \mathcal{A}^+| \\ &= |\mathcal{A}|\log_2 2^n/|\mathcal{A}| + |\mathcal{A}|\big(H(\gamma) - 1 + b_n + \gamma K^- + (1-\gamma)K^+\big). \end{aligned}$$

By possibly decreasing K we can assume that $|\partial_e(\mathcal{A})| = |\mathcal{A}|\big(\log_2(2^n/|\mathcal{A}|) + K\big)$. Then

$$\gamma K^- + (1-\gamma)K^+ = K - (H(\gamma) - 2\gamma) - (b_n - (1-2\gamma)) := \widetilde{K}. \tag{6}$$

Note that both bracketed terms here are non-negative. Indeed, $H(\gamma)$ is concave on $[0,1/2]$ as $H'(\gamma) = \log_2(\gamma/(1-\gamma)) \geq 0$ and so $H(\gamma) \geq 2\gamma$ for $\gamma \in [0,1/2]$ as $H(0) = 0$ and $H(1/2) = 1$. The second term is also non-negative as $b_n|\mathcal{A}| = |\mathcal{A}^+ \triangle \mathcal{A}^-| \geq |\mathcal{A}^+| - |\mathcal{A}^-| = (1-2\gamma)|\mathcal{A}|$. By partitioning the contribution of (6) we find $\delta \in [0,1]$ with $\gamma K^- = \delta \widetilde{K}$ and $(1-\gamma)K^+ = (1-\delta)\widetilde{K}$. Also fix $E := 2^{-2C_2(K/\varepsilon+1)^2}$.

First suppose that $\widetilde{K} \leq K - E$. In this case, we will approximate both $\mathcal{A}^-$ and $\mathcal{A}^+$ by appropriate collections of cubes. Set $\varepsilon^- = \delta\varepsilon/\gamma$ and $\varepsilon^+ = (1-\delta)\varepsilon/(1-\gamma)$. By the inductive hypothesis, there are disjoint subcubes $\mathscr{C}^- = \{\mathcal{C}_j^-\}_{j\in[L^-]}$ with $L^- \leq L(K^-, \varepsilon^-)$ and $\mathscr{C}^+ = \{\mathcal{C}_i^+\}_{i\in[L^+]}$ with $L^+ \leq L(K^+, \varepsilon^+)$ so that

$$|\mathcal{A}^- \triangle (\cup_{\mathcal{C}\in\mathscr{C}^-}\mathcal{C})| \leq \varepsilon^-|\mathcal{A}^-| = \delta\varepsilon|\mathcal{A}| \qquad \text{and}$$
$$|\mathcal{A}^+ \triangle (\cup_{\mathcal{C}\in\mathscr{C}^+}\mathcal{C})| \leq \varepsilon^+|\mathcal{A}^+| = (1-\delta)\varepsilon|\mathcal{A}|.$$

We can naturally identify cubes in $\mathscr{C}^-$, $\mathscr{C}^+$ with subcubes of Q_n in which the nth coordinate is 0, 1, respectively. Taking $\mathscr{C} = \mathscr{C}^- \cup \mathscr{C}^+$ we find $|\mathcal{A} \triangle (\cup_{\mathcal{C}\in\mathscr{C}}\mathcal{C})| \leq \varepsilon|\mathcal{A}|$. Therefore

$$|\mathscr{C}| \leq L(K^-, \varepsilon^-) + L(K^+, \varepsilon^+) \leq 2g(\widetilde{K}/\varepsilon) \leq 2g\big((K-E)/\varepsilon\big) \leq 2g\big(K/\varepsilon - E\big). \tag{7}$$

Note that the function $h(x) = \log_2 g(x) = 2^{C(x+1)^2}$ satisfies $h'(x) \geq 2(\log_e 2)C(x+1)h(x) \geq h(x)$ and so $h'(x)$ is increasing. By the mean value theorem, using $E \leq (K/\varepsilon + 1)/2$, this gives

$$\begin{aligned} 1 + h(K/\varepsilon - E) &\leq 1 + h(K/\varepsilon) - Eh'(K/\varepsilon - E) \leq 1 + h(K/\varepsilon) - E2^{C(K/\varepsilon+1)^2/4} \\ &\leq h(K/\varepsilon). \end{aligned}$$

Here $E2^{C(K/\varepsilon+1)^2/4} = 2^{(C/4-2C_2)(K/\varepsilon+1)^2} \geq 1$ as $C \geq 8C_2$. Exponentiating, and combining with (7) we find $|\mathscr{C}| \leq 2g(K/\varepsilon - E) \leq g(K/\varepsilon)$, completing the proof of this case.

It remains to deal with the case $\widetilde{K} \geq K - E$. We claim that this is only possible if $\gamma \leq E$. To see this, note that by (6) in this case we have (i) $b_n - (1-2\gamma) \leq E$ and (ii) $H(\gamma) - 2\gamma \leq E$. Since $b_n \geq c(K) \geq 2E$, by (i) we have $\gamma \leq 1/2 - c(K)/4$. Also $H(\gamma) - 2\gamma \geq 2\gamma - 4\gamma^2 = 2\gamma(1-2\gamma) \geq \min(\gamma, 1-2\gamma)$ for $\gamma \in [0,1/2]$. Therefore $H(\gamma) - 2\gamma > E$ for $\gamma \in (E, 1/2 - c(K)/4]$, which by (ii) forces $\gamma \leq E$, as claimed.

As $\mathcal{A}^-$ is small, we can approximate $\mathcal{A}$ by deleting $\mathcal{A}^-$ and approximating $\mathcal{A}^+$ by subcubes with accuracy $\varepsilon' = (\varepsilon - \gamma)/(1-\gamma)$. By induction, there are disjoint cubes $\mathscr{C} = \{\mathcal{C}_i\}_{i\in[L]}$ with $|\mathcal{A}^+ \triangle (\cup_{\mathcal{C}\in\mathscr{C}}\mathcal{C})| \leq \varepsilon'|\mathcal{A}^+|$ and $L \leq g(K^+/\varepsilon')$. But then $|\mathcal{A} \triangle (\cup_{\mathcal{C}\in\mathscr{C}}\mathcal{C})| \leq \varepsilon'|\mathcal{A}^+| + |\mathcal{A}^-| \leq \varepsilon|\mathcal{A}|$. Lastly, using $\gamma \leq E$, $C_2 \geq 1$ and $K^+ \leq \big(K - (H(\gamma) - 2\gamma)\big)/(1-\gamma)$ we have

$$\frac{K^+}{\varepsilon'} \leq \frac{K - \gamma\log_2\gamma^{-1}/3}{\varepsilon - \gamma} \leq \frac{K - 2\gamma C_2(K/\varepsilon+1)^2/3}{\varepsilon - \gamma} \leq \frac{K - \gamma K/\varepsilon}{\varepsilon - \gamma} = \frac{K}{\varepsilon}.$$

Therefore $L \leq g(K^+/\varepsilon') \leq g(K/\varepsilon)$, completing the proof of this case and the theorem. $\square$

Note. Keller and Lifshitz [21] have independently and simultaneously proved a stronger version of our main theorem, with an essentially optimal bound of $L(K,\varepsilon) \leq 2^{2^{C(K/\varepsilon)}}$. Although our bounds are weaker, as our approach is significantly different we feel that the methods may be useful for similar problems in the future, particularly if they are not amenable to the compression arguments used in [21].

Acknowledgments

We would like to thank David Ellis for bringing the independent work of Keller and Lifshitz [21] to our attention. We would also like to thank the referees for their helpful comments.

References

[1] N. Alon, J. Spencer, The Probabilistic Method, Wiley, 2008.
[2] W. Beckner, Inequalities in Fourier analysis, Ann. of Math. 102 (1975) 159–182.
[3] M. Ben-Or, N. Linial, Collective coin flipping, robust voting schemes and minima of Banzhaf values, in: FOCS, IEEE Computer Society, 1985, pp. 408–416.
[4] A.J. Bernstein, Maximally connected arrays on the n-cube, SIAM J. Appl. Math. 15 (1967) 1485–1489.
[5] B. Bollobás, Combinatorics, Cambridge University Press, 1986.
[6] A. Bonami, Étude des coefficients de Fourier des fonctions de $L_p(G)$, Ann. Inst. Fourier 20 (1970) 335–402.
[7] J. Bourgain, appendix to [14].
[8] F.R.K. Chung, P. Frankl, R.L. Graham, J.B. Shearer, Some intersection theorems for ordered sets and graphs, J. Combin. Theory Ser. A 43 (1986) 23–37.
[9] T.M. Cover, J.A. Thomas, Elements of Information Theory, Wiley, 2006.
[10] D. Ellis, Almost isoperimetric subsets of the discrete cube, Combin. Probab. Comput. 20 (2011) 363–380.
[11] D. Ellis, N. Keller, N. Lifshitz, On the structure of subsets of the discrete cube with small edge boundary, arXiv:1612.06680.
[12] D. Falik, A. Samorodnitsky, Edge-isoperimetric inequalities and influences, Combin. Probab. Comput. 16 (2007) 693–712.
[13] E. Friedgut, Boolean functions with low average sensitivity depend on few coordinates, Combinatorica 18 (1998) 27–35.
[14] E. Friedgut, Sharp thresholds of graph properties, and the k-sat problem, J. Amer. Math. Soc. 12 (1999) 1017–1054.
[15] E. Friedgut, G. Kalai, Every monotone graph property has a sharp threshold, Proc. Amer. Math. Soc. 124 (1996) 2993–3002.
[16] L.H. Harper, Optimal assignments of numbers to vertices, SIAM J. Appl. Math. 12 (1964) 131–135.
[17] S. Hart, A note on the edges of the n-cube, Discrete Math. 14 (1976) 157–163.
[18] H. Hatami, A structure theorem for Boolean functions with small total influences, Ann. of Math. 176 (2012) 509–533.
[19] J. Kahn, G. Kalai, Thresholds and expectation thresholds, Combin. Probab. Comput. 16 (2007) 495–502.
[20] J. Kahn, G. Kalai, N. Linial, The influence of variables on Boolean functions, in: Proceedings of the 29th Annual IEEE Symposium on Foundations of Computer Science, 1988, pp. 68–80.
[21] N. Keller, N. Lifshitz, Approximation of biased Boolean functions of small total influence by DNF's, preprint.

[22] J.H. Lindsey II, Assignment of numbers to vertices, Amer. Math. Monthly 71 (1964) 508–516.
[23] E. Mossel, R. O'Donnell, K. Oleszkiewicz, Noise stability of functions with low influences: invariance and optimality, Ann. of Math. 171 (2010) 295–341.
[24] R. O'Donnell, Analysis of Boolean Functions, Cambridge University Press, 2014.
[25] Y. Polyanskiy, Hypercontractivity of spherical averages in Hamming space, arXiv:1309.3014.
[26] R. Rossignol, Threshold for monotone symmetric properties through a logarithmic Sobolev inequality, Ann. Probab. 34 (2005) 1707–1725.
[27] L. Russo, An approximate zero–one law, Z. Wahrsch. Verw. Gebiete 61 (1982) 129–139.
[28] M. Talagrand, On Russo's approximate zero–one law, Ann. Probab. 22 (1994) 1576–1587.