KKL, Kruskal-Katona, and monotone nets

Ryan O'Donnell
Carnegie Mellon University
odonnell@cs.cmu.edu

Karl Wimmer
Carnegie Mellon University
kwimmer@andrew.cmu.edu

August 15, 2009

Abstract

We generalize the Kahn-Kalai-Linial (KKL) Theorem to random walks on Cayley and Schreier graphs, making progress on an open problem of Hoory, Linial, and Wigderson. In our generalization, the underlying group need not be abelian so long as the generating set is a union of conjugacy classes. An example corollary is that for every $f : \binom{[n]}{k} \to \{0,1\}$ with $\mathbf{E}[f]$ and k/n bounded away from 0 and 1, there is a pair $1 \le i < j \le n$ such that $\mathcal{I}_{ij}(f) \ge \Omega(\frac{\log n}{n})$. Here $\mathcal{I}_{ij}(f)$ denotes the "influence" on f of swapping the ith and jth coordinates. Using this corollary we obtain a "robust" version of the Kruskal-Katona Theorem: Given a constant-density subset A of a middle slice of the Hamming n-cube, the density of ∂A is greater by at least $\Omega(\frac{\log n}{n})$, unless A is noticeably correlated with a single coordinate.

As an application of these results, we show that the set of functions $\{0, 1, x_1, \ldots, x_n, \mathrm{Maj}\}$ is a $(1/2 - \gamma)$-net for the set of all n-bit monotone boolean functions, where $\gamma = \Omega(\frac{\log n}{\sqrt{n}})$. This distance is optimal for polynomial-size nets and gives an optimal weak-learning algorithm for monotone functions under the uniform distribution, solving a problem of Blum, Burch and Langford.

1 Introduction

In this paper we:

- Generalize the celebrated Kahn-Kalai-Linial (KKL) Theorem [KKL88] to non-product-distribution settings; specifically, to random walks on Cayley and Schreier graphs.
- Use this to give a "robust" version of the classical Krukal-Katona Theorem [Kru63, Kat68], showing that a set's shadow is noticeably larger than Kruskal-Katona promises, unless the set has significant correlation with a single coordinate.
- Deduce that every monotone boolean function has correlation $\Omega(\frac{\log n}{\sqrt{n}})$ with one of the functions 0, 1, x_1, x_2, ..., x_n, or Majority. From this we derive an optimal weak-learning algorithm for monotone functions under the uniform distribution (which is also highly efficient).

We proceed to discuss each of these topics (the KKL Theorem, the Kruskal-Katona Theorem, approximating monotone functions) in depth.

1.1 The Kahn-Kalai-Linial Theorem

The paper of Kahn, Kalai, and Linial [KKL88] has been one of the most influential works applying Fourier analysis to theoretical computer science. The KKL Theorem, along with variants from works such as [BKK+92, Tal94, Fri98], has proved enormously useful in a wide variety of areas, from distributed computing [BOL90], to random k-SAT [Fri99] and random graphs [FK96], communication complexity [Raz95], hardness of approximation [DS05, CKK+05, KR08], metric embeddings [KR06, DKSV06], weak random sources [KZ06], and learning theory [OS08].

The usual statement of the Kahn-Kalai-Linial Theorem is:

KKL Theorem: For any $f : \{0,1\}^n \to \{0,1\}$, $\mathcal{M}[f] \geq \Omega\left(\frac{\log n}{n}\right) \cdot \mathbf{Var}[f]$.

This statement uses the following definitions:[1]

$$\mathbf{Var}[f] = \mathbf{E}[f(\boldsymbol{x})^2] - \mathbf{E}[f(\boldsymbol{x})]^2 = \mathbf{Pr}[f(\boldsymbol{x}) = 0] \cdot \mathbf{Pr}[f(\boldsymbol{x}) = 1].$$

$$\mathcal{I}_i[f] = \tfrac{1}{2}\,\mathbf{E}[(f(\boldsymbol{x}) - f(\boldsymbol{x}^{(i)}))^2] = \tfrac{1}{2}\,\mathbf{Pr}[f(\boldsymbol{x}) \neq f(\boldsymbol{x}^{(i)})], \qquad \mathcal{M}[f] = \max_{i \in [n]}\{\mathcal{I}_i[f]\}.$$

Here $x^{(i)}$ denotes the string x with the ith coordinate flipped. The quantity $\mathcal{I}_i[f]$ is called the *influence* of coordinate $i \in [n]$ on f.[2] One often focuses on "balanced" functions, meaning $\mathbf{Pr}[f(\boldsymbol{x}) = 1] = 1/2$, or "roughly balanced" functions, meaning $\Omega(1) \leq \mathbf{Pr}[f(\boldsymbol{x}) = 1] \leq 1 - \Omega(1)$. In either case, $\mathbf{Var}[f] \geq \Omega(1)$ and the KKL Theorem says that there exists at least one coordinate with influence at least $\Omega(\frac{\log n}{n})$.

The KKL Theorem is tight up to the constant, by the "Tribes" example of Ben-Or and Linial [BOL90]. It improves over the elementary lower bound of $\frac{2}{n}\mathbf{Var}[f]$, which follows immediately from the *Poincaré Inequality* for the discrete cube:

$$\mathcal{E}[f] \geq \frac{2}{n}\mathbf{Var}[f], \quad \text{where } \mathcal{E}[f] = \operatorname*{avg}_{i \in [n]}\{\mathcal{I}_i[f]\} \text{ is the } \textit{average} \text{ influence.} \tag{1}$$

[1] Throughout this paper we use boldface to denote random variables, and these are assumed to have the uniform distribution on their domain unless otherwise specified.

[2] The factor of 1/2 in its definition is often omitted; we take it for technical consistency with the later results in the paper.

Although the gain from $\Omega(\frac{1}{n})$ to $\Omega(\frac{\log n}{n})$ might at first seem small, it is often the fact that the gained factor $\log n$ goes to infinity that makes all the difference in applications. For example, $\log n \to \infty$ is the reason why a $o(1)$ fraction of voters can control any two-party election [KKL88], why one has *sharp* thresholds for graph properties [FK96], why the Sparsest-Cut semidefinite program has a superconstant integrality gap [DKSV06], etc.

We will actually be interested in the following strengthening of the KKL Theorem, first stated and proved by Talagrand [Tal94] though following easily from the proof method of KKL:

KKL Theorem 2: For any $f : \{0,1\}^n \to \{0,1\}$, $\mathcal{E}[f] \geq \Omega(1) \cdot \frac{\log(1/\mathcal{M}[f])}{n} \cdot \mathbf{Var}[f]$.

This is a strengthening because we can of course replace the left-hand side by $\mathcal{M}[f]$. Generalizations of the KKL Theorem(s) are known to hold under the p-biased distribution on $\{0,1\}^n$ [Tal94, FK96] and under the uniform distribution on $[0,1]^n$ [BKK+92]. However these generalizations seem to depend heavily on having a *product* probability distribution; this is because all known proofs (even recent alternate ones [Ros06, FS07]) are essentially Fourier-analytic, and Fourier analysis seems to work best with product distributions.[3]

1.1.1 The KKL Theorem on Schreier graphs

In their survey on expander graphs [HLW06], Hoory, Linial, and Wigderson connected the KKL Theorem to *expansion* in the Cayley graph of the group $\mathbb{Z}_2^n$ with the standard set of generators $(e_i)_{i \in [n]}$; this connection is the key to the metric embedding results in [KR06, DKSV06]. Hoory, Linial, and Wigderson asked if this phenomenon could be found in Cayley graphs for other groups.

In this paper we prove such a result. To state it we need some more definitions. Let G be a finite group acting transitively on a finite set X; we write x^g for the action of $g \in G$ on $x \in X$. Let U denote a generating set for G which is *symmetric*: $U = U^{-1}$. The *Schreier graph* $\mathrm{Sch}(G, X, U)$ has vertex set X and an edge (x, y) whenever $x^u = y$ for some $u \in U$. A special case is when $X = G$ and $x^u = xu$; in this case we recover the *Cayley graph* $C(G, U)$. These are connected, regular, undirected graphs. In fact, it is known [Gro77] that *every* regular graph of even degree is a Schreier graph for some group action. Because of the regularity, it is natural to endow X with the uniform distribution (which is not in general a product distribution). Then, for a function $f : X \to \{0,1\}$ we define $\mathbf{Var}[f]$ as before, and we define

$$\mathcal{I}_u[f] = \tfrac{1}{2} \mathop{\mathbf{E}}_{\boldsymbol{x} \sim X}[(f(\boldsymbol{x}) - f(\boldsymbol{x}^u))^2] = \tfrac{1}{2} \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim X}[f(\boldsymbol{x}) \neq f(\boldsymbol{x}^u)],$$

the influence of $u \in U$ on f. We again define $\mathcal{E}[f]$ as the average influence, $\frac{1}{|U|} \sum_{u \in U} \mathcal{I}_u[f]$. Finally, for the natural random walk on $\mathrm{Sch}(G, X, U)$ (where we move from x to $x^{\boldsymbol{u}}$ for a random $\boldsymbol{u} \in U$), there is an associated *log-Sobolev constant*, denoted ρ (see [Gro75, DSC96]). It is defined as the largest constant such that the following holds for all nonconstant $f : X \to \mathbb{R}$:

Log-Sobolev Inequality: $\mathcal{E}[f] \geq \frac{1}{2}\rho \mathbf{Ent}[f^2]$, where $\mathbf{Ent}[g] = \mathbf{E}[g(\boldsymbol{x}) \log g(\boldsymbol{x})] - \mathbf{E}[g(\boldsymbol{x})] \log \mathbf{E}[g(\boldsymbol{x})]$.

We can now state our generalization of the KKL Theorem 2:

[3] The one exception to this rule is the generalization by Graham and Grimmett [GG06] to distributions on $\{0,1\}^n$ which are "monotonic" (which is equivalent to satisfying the FKG lattice condition). Still, the Graham-Grimmett proof is a reduction to the product-distribution case.

Theorem 1.1. *In the Schreier graph* $\mathrm{Sch}(G, X, U)$*, with log-Sobolev constant* ρ*, suppose that the generating set* U *is a union of conjugacy classes. Then for any* $f : X \to \{0,1\}$*,*

$$\mathcal{E}[f] \geq \Omega(1) \cdot \log(1/\mathcal{M}[f]) \cdot \rho \cdot \mathbf{Var}[f]; \qquad \textit{hence also } \exists u \in U \textit{ s.t. } \mathcal{I}_u[f] \geq \Omega(1) \cdot \rho \log(1/\rho) \cdot \mathbf{Var}[f].$$

This recovers the KKL Theorem(s) by taking the Cayley graph on $G = X = \mathbb{Z}_2^n$ with the standard generating set (which is of course a union of conjugacy classes, since every group element is its own conjugacy classes in an abelian group like $\mathbb{Z}_2^n$). The log-Sobolev constant for the associated random walk is known to be $\frac{2}{n}$ [Gro75].

The main application in this paper of Theorem 1.1 takes place in the *nonabelian* setting of the Schreier graph $\mathrm{Sch}(S_n, \binom{[n]}{k}, U)$, where $\binom{[n]}{k}$ denotes the set of n-bit strings of Hamming weight k, S_n is the symmetric group acting on $\binom{[n]}{k}$ in the natural way, and U is the set of all transpositions.[4] Here we have $\binom{n}{2}$ generators (ij), and we write $\mathcal{I}_{ij}[f]$ for the influence of switching the ith and jth coordinates on f. Using the log-Sobolev constant for this graph determined by Lee and Yau [LY98], we are able to conclude:

Corollary 1.2. *For any* $f : \binom{[n]}{k} \to \{0,1\}$*, with* $\Omega(1) \leq k/n \leq 1 - \Omega(1)$*,*

$$\mathcal{E}[f] \geq \Omega(1) \cdot \frac{\log(1/\mathcal{M}[f])}{n} \cdot \mathbf{Var}[f]; \qquad \textit{hence also } \exists (ij) \textit{ s.t. } \mathcal{I}_{ij}[f] \geq \Omega\left(\frac{\log n}{n}\right) \cdot \mathbf{Var}[f].$$

We discuss the Schreier graph framework in greater detail in Section 2.1, and prove a more detailed version of Theorem 1.1 and Corollary 1.2 in Section 2.2.

1.2 The Kruskal-Katona Theorem

The Kruskal-Katona Theorem [Kru63, Kat68] is a widely-used, classical result in combinatorics. We state it in terms of subsets of slices of the boolean hypercube. Given a set of strings $A \subseteq \binom{[n]}{k}$, its *(lower) shadow* and *upper shadow* are

$$\partial A = \{x \in \binom{[n]}{k-1} : x \prec y \text{ for some } y \in A\}, \quad \partial^u A = \{x \in \binom{[n]}{k+1} : x \succ y \text{ for some } y \in A\},$$

where $x \prec y$ denotes that $x \leq y$ component-wise. The Kruskal-Katona Theorem gives an exact lower bound on the size of $|\partial A|$ (respectively, $|\partial^u A|$) as a function of $|A|$; the minimizing A consists of the first (respectively, last) $|A|$ strings in colexicographic order. The precise formulas here are cumbersome to deal with, so slightly weaker results are often stated. The most common one is due to Lovász [Lov79], but we will state an even handier corollary due to Bollobás and Thomason [BT87]. For a set $B \subseteq \binom{[n]}{\ell}$, we use the notation $\mu_\ell(B) = |B|/\binom{n}{\ell}$:

Kruskal-Katona Corollary: Let $A \subseteq \binom{[n]}{k}$. Then[5]

$$\mu_{k+1}(\partial^u A) \quad \geq \quad \mu_k(A)^{1-1/(n-k)} \quad \geq \quad \mu_k(A) + \frac{\mu_k(A) \ln(1/\mu_k(A))}{n-k}.$$

The parameter range of greatest interest to us is when both $\mu_k(A)$ and k/n are bounded away from 0 and 1 by a constant. In this case, the Kruskal-Katona Corollary above implies that

[4] Note that the uniform distribution on $\binom{[n]}{k}$ is not monotonic in the sense of Graham and Grimmett, so their result does not apply. In fact the distribution is *anti*-monotonic.

[5] We can always make an analogous statement for lower shadows, in this case $\mu_{k-1}(\partial A) \geq \mu_k(A)^{1-1/k}$, obtained trivially by boolean duality. For the remainder of the paper we mainly discuss upper shadows.

$\mu_{k+1}(\partial^u A) \geq \mu_k(A) + \Omega(1/n)$. This amount of "density increase" cannot, in general, be improved. (Since we know the original Kruskal-Katona Theorem is precisely sharp, this is to say that the corollary does not lose much.) To see this, consider for example the *dictator* sets $A = \{x : x_i = 1\}$ (at various slices). Clearly $\mu_k(A) = k/n$ and $\mu_{k+1}(\partial^u A) = (k+1)/n$, so the shadow density increased by only $1/n$. This is not the only such example; one gets upper shadow density increases of only $\Omega(1/n)$ for any set A of the form $\{x : f(x_I) = 1\}$, where $I \subseteq [n]$ is of bounded cardinality and $f : \{0,1\}^{|I|} \to \{0,1\}$ is monotone. On the other hand, one may wonder if these are the only such examples.

Our "robust" version of the Kruskal-Katona Theorem states that this is the case: given a set $A \subseteq \binom{[n]}{k}$ in the parameter range of interest, if A is not noticeably correlated with a single coordinate, then the shadow densities increase by the much larger amount $\Omega(\frac{\log n}{n})$.

Theorem 1.3. *For all $\epsilon > 0$ there exists $\delta > 0$ such that the following holds: If $A \subseteq \binom{[n]}{k}$, $\epsilon \leq k/n, \mu_k(A) \leq 1 - \epsilon$, then the following holds:*

$$\mu_{k+1}(\partial^u A) \geq \mu_k(A) + \delta \cdot \frac{\log n}{n}$$

unless there exists $i \in [n]$ with

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \binom{[n]}{k}}[\boldsymbol{x} \in A \mid \boldsymbol{x}_i = 1] - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \binom{[n]}{k}}[\boldsymbol{x} \in A \mid \boldsymbol{x}_i = 0] \geq 1/n^{\epsilon}.$$

We remark that a similar-in-spirit "stability" result for Kruskal-Katona was recently proved via combinatorial means by Keevash [Kee08]. In the terminology of Tao [Tao07], Keevash's is a "99%-structured" result, whereas ours is a "1%-structured" result.

We prove Theorem 1.3 in Appendix A.

1.3 Approximating and learning monotone functions

We use our results to solve optimally the problem of weak-learning monotone functions under the uniform distribution. The problem is an old one, introduced in the first paper on weak-learning [KV89]. An algorithm is given access to uniformly distributed random examples $(\boldsymbol{x}, f(\boldsymbol{x}))$ from an unknown *monotone* function $f : \{0,1\}^n \to \{0,1\}$. The algorithm's task is to construct an approximation to f, called a *hypothesis*. The quality of a hypothesis $h : \{0,1\}^n \to \{0,1\}$ is measured by its *accuracy* with respect to the uniform distribution: $\mathbf{Pr}_{\boldsymbol{x}}[f(\boldsymbol{x}) = h(\boldsymbol{x})]$.

Since the class of all monotone functions is quite rich, one does not expect a polynomial-time algorithm to be able to construct a highly accurate hypothesis. Indeed, the original algorithm of Kearns and Valiant achieves accuracy $1/2 + \Omega(1/n)$ (with high probability[6]); this is termed *weak-learning with advantage* $\Omega(1/n)$ (the "advantage" is over random guessing). The Kearns-Valiant algorithm is very simple: it draws $O(n^2 \log n)$ examples and then chooses one of the hypotheses $\{0, 1, x_1, x_2, \ldots, x_n\}$, whichever has the highest empirical accuracy on the sample. It is easy to show that the true accuracy of the hypothesis x_i for a monotone function f is exactly $1/2 + \mathcal{I}_i[f]$. Thus the basic inequality (1) implies that either 0 or 1 has accuracy $\Omega(1)$ or at least one x_i has accuracy at least $\Omega(1/n)$. A simple Chernoff bound implies that the empirical accuracies of all potential hypotheses are close to the true accuracies, and this proves the correctness of the Kearns-Valiant algorithm.

We introduce the terminology "net-based" for algorithms of the Kearns-Valiant type. This means that they work by showing a information-theoretic result: a *net* for the class of all monotone functions.

[6] At least 2/3, say, which can be boosted by standard means.

Definition 1.4. *Let $\mathcal{C}$ be a class of boolean functions, $\{0,1\}^n \to \{0,1\}$. An* α-net *for $\mathcal{C}$ is a collection $\mathcal{H}$ of n-bit boolean functions such that for all $f \in \mathcal{C}$ there is an $h \in \mathcal{H}$ with $\mathbf{Pr}[f(\boldsymbol{x}) \neq h(\boldsymbol{x})] \leq \alpha$.*

As we've seen, the collection $\{0, 1, x_1, \ldots, x_n\}$ is a $(1/2 - \Omega(\frac{1}{n}))$-net for the class of monotone functions. And given a polynomial-size $(1/2-\gamma)$-net for a class $\mathcal{C}$, a Chernoff argument easily implies a weak-learning algorithm for $\mathcal{C}$ under the uniform distribution using $O(1/\gamma^2) \cdot \log n$ examples and polynomial time.[7]

Remark 1.5. *The definition of a net for the class of monotone functions does not require that the functions in the net themselves be monotone. However a simple shifting argument [Kle66] shows that if one replaces each net function with a monotone shifted version, the net's distance parameter can only decrease. Thus it suffices to look for nets of monotone functions.*

Bshouty and Tamon [BT96] later improved the Kearns-Valiant advantage to $\Omega(\frac{\log n}{n})$, using the KKL Theorem, but the main development came in the work of Blum, Burch, and Langford [BBL98]. They used the Kruskal-Katona Corollary to show that the tiny net $\{0, 1, \mathrm{Maj}\}$ is a $(1/2-\Omega(\frac{1}{\sqrt{n}}))$-net for the class of monotone functions, where Maj denotes the Majority function. They also showed that any learning algorithm which sees f's value on only polynomially many strings can achieve advantage at most $O(\frac{\log n}{\sqrt{n}})$. This is an information-theoretic result, and in particular it implies:

Theorem 1.6. *(Blum-Burch-Langford) If $\mathcal{H}$ is a $(1/2 - \gamma)$-net of polynomial size for the class of monotone n-bit functions, then $\gamma \leq O(\frac{\log n}{\sqrt{n}})$.*

This shows that the net $\{0, 1, \mathrm{Maj}\}$ is close to optimal among polynomial-size nets, but leaves a gap factor of $\log n$. Blum, Burch, and Langford conjectured that efficiently weak-learning monotone functions with advantage $\Omega(\frac{\log n}{\sqrt{n}})$ is possible. Using a net approach to weak-learning monotone functions, we can assume that each function in the net is monotone by shifting

A subsequent work of Amano and Maruoka [AM06] suggested directions for proving this; they conjectured that $\{0, 1, x_1, \ldots, x_n, \mathrm{Maj}\}$ is a $(1/2 - \Omega(\frac{\log n}{\sqrt{n}}))$-net for monotone functions. Indeed, Amano and Maruoka made the strictly stronger conjecture that for balanced monotone f, the hypothesis Maj has accuracy $1/2 + \Omega(\mathcal{E}[f] \cdot \sqrt{n})$. This implies their net conjecture using the KKL Theorem 2. In fact, unbeknownst to the authors, Benjamini, Kalai, and Schramm [BKS99, Theorem 3.1] had shown that for balanced monotone f, the hypothesis Maj has accuracy

$$\frac{1}{2} + \Omega\left(\frac{\mathcal{E}[f] \cdot \sqrt{n}}{\sqrt{\log(1/(\mathcal{E}[f]\sqrt{n}))}}\right). \tag{2}$$

If one combines this with the KKL Theorem 2, one can show that $\{0, 1, x_1, \ldots, x_n, \mathrm{Maj}\}$ is a $(1/2 - \Omega(\frac{\sqrt{\log n}}{\sqrt{n}}))$-net for monotone functions. Unfortunately, in Appendix C we provide a counterexample showing that the stronger conjecture of Amano and Maruoka is false.

Nevertheless, in this paper we use our robust Kruskal-Katona Theorem to show that the first Amano-Maruoka conjecture is true, and hence so is the Blum-Burch-Langford conjecture. Indeed, in Appendix B we show the following result:

Theorem 1.7. *For all $0 < \epsilon < 1/2$ there exists $0 < \delta < 1$ such that the following holds: Let $f : \{0,1\}^n \to \{0,1\}$ be monotone. Then at least one of the below statements about accuracy with*

[7] Assuming the functions in the net are easy to evaluate.

respect to f must hold:

$$\begin{aligned} 0 \text{ or } 1 \text{ has accuracy} &\geq 1-\epsilon; \\ x_1, x_2, \ldots, \text{ or } x_n \text{ has accuracy} &\geq 1/2 + 1/n^{\epsilon}; \\ \text{Maj has accuracy} &\geq 1/2 + \delta \cdot \frac{\log n}{\sqrt{n}}. \end{aligned}$$

This of course gives a $(1/2 - \Omega(\frac{\log n}{\sqrt{n}}))$-net of size $n+3$ for monotone functions, a distance which is optimal (up to the $\Omega(\cdot)$ constant) for polynomial-size nets. In fact, this stronger result yields an unusually efficient weak-learning algorithm. For any constant $\epsilon > 0$, an algorithm can draw $O(n^{2\epsilon} \log n)$ random examples to check whether 0, 1, x_1, ..., or x_n has empirical accuracy at least $1/2+1/n^{\epsilon}$. If so, it outputs that hypothesis. If not, Theorem 1.7 implies that (with high probability, using Chernoff bounds) Maj has accuracy $1/2 + \delta \cdot \frac{\log n}{\sqrt{n}}$. The algorithm *need not verify this*; it can simply immediately output Maj. Hence we get a nearly-linear time optimal weak-learner with subpolynomial sample complexity:

Theorem 1.8. *For any positive constant $\epsilon > 0$, there is an algorithm for weak-learning the class of n-bit monotone functions under the uniform distribution which achieves advantage $\Omega(\frac{\log n}{\sqrt{n}})$ while using $O(n^{\epsilon})$ random examples and $O(n^{1+\epsilon})$ time.*

Further, we give a $(1/2 - \Omega(\frac{\log n}{\sqrt{n}}))$-net of size $O(n/\log n)$ for monotone functions, using slightly more complicated functions.

We prove Theorem 1.7 in Appendix B.

2 KKL

In this section we set up and then prove our generalized KKL Theorem.

2.1 Preliminaries

Recall the setting described in Section 1.1.1, a random walk on the Schreier graph $\mathrm{Sch}(G, X, U)$, where G is a group acting on X with symmetric generating set U. Examples to keep in mind include the standard Cayley graphs of $\mathbb{Z}_2^n$ and $\mathbb{Z}_m^n$, and the Cayley graph of S_n with transpositions. Our main application is the setting where $X = \binom{[n]}{k}$, $G = S_n$, and U is the set of transpositions, acting on strings in X in the natural way. We write $L^2(X)$ for the inner product space of all functions $f : X \to \mathbb{R}$, with inner product $\langle f, g \rangle = \mathbf{E}_{\boldsymbol{x} \sim X}[f(\boldsymbol{x})g(\boldsymbol{x})]$. Here $\boldsymbol{x} \sim X$ denotes that $\boldsymbol{x}$ is drawn from the uniform distribution on X. We will consider some basic operators on this space, associated with the random walk on $\mathrm{Sch}(G, X, U)$. Define the operators L (the *normalized Laplacian*) and L_u (for $u \in U$) as follows:

$$L_u f(x) = f(x) - f(x^u), \qquad L = \frac{1}{|U|} \sum_{u \in U} L_u = id - K,$$

where K is the *Markov operator* or *transition matrix* for the random walk. Next, for $t \in \mathbb{R}^{\geq 0}$ we define the *continuous time Markov semigroup* H_t,

$$H_t = e^{-tL} = e^{-t} e^{tK} = \sum_{m=0}^{\infty} \frac{e^{-t} t^m}{m!} K^m.$$

In other words, $H_t f(x) = \mathbf{E}_{\boldsymbol{y}}[f(\boldsymbol{y})]$, where $\boldsymbol{y}$ is generated from x by taking $\boldsymbol{m}$ steps in the random walk, with $\boldsymbol{m} \sim \text{Poisson}(t)$. The semigroup property is that $H_s H_t = H_{s+t}$.

It is easy to check that when $\boldsymbol{x} \sim X$ and $\boldsymbol{u} \in U$ are chosen uniformly and independently, the pairs $(\boldsymbol{x}, \boldsymbol{x}^{\boldsymbol{u}})$ and $(\boldsymbol{x}^{\boldsymbol{u}}, \boldsymbol{x})$ have the same distribution. From this one concludes that K is a self-adjoint operator, and hence so are K and H_t. For a fixed $u \in U$, the L_u operator is not in general self-adjoint (its adjoint is $L_{u^{-1}}$). However, it does have the property $\langle f, L_u f\rangle = \frac{1}{2}\langle L_u f, L_u f\rangle$, which follows from the fact that $\boldsymbol{x}^u$ is uniformly distributed when $\boldsymbol{x} \sim X$ is.

We next recall the basic functionals on $L^2(X)$ described in Section 1.1.1:

Definition 2.1. *For $u \in U$ we define the* influence of u *on $f \in L^2(X)$ to be*

$$\mathcal{I}_u[f] = \tfrac{1}{2}\|L_u f\|_2^2 = \langle f, L_u f\rangle = \langle L_u f, f\rangle. \tag{3}$$

(The first equation here is the definition; the second two are an easy consequence of the fact that $\boldsymbol{x}^{\boldsymbol{u}}$ is uniformly distributed when $\boldsymbol{x} \in X$, $\boldsymbol{u} \in U$ are uniformly random.)

Definition 2.2. *The* average influence *of $f \in L^2(X)$ is*

$$\mathcal{E}[f] = \frac{1}{|U|}\sum_{u \in U} \mathcal{I}_u[f] = \langle f, Lf\rangle = \langle Lf, f\rangle.$$

Next, we consider the eigenvalue/eigenfunction decomposition of the normalized Laplacian L. It is well-known that L has nonnegative real eigenvalues denoted

$$0 = \lambda_0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_{|X|-1}.$$

(Positivity of λ_1 follows because G acts transitively on X so the random walk is irreducible.) We write $(\psi_i)_{i=0}^{|X|-1}$ for corresponding eigenfunctions forming an orthonormal basis of $L^2(X)$, with $\psi_0 \equiv 1$. Note that the ψ_i's are also eigenfunctions for the operator H_t, with associated eigenvalues $e^{-t\lambda_i}$. For a given $f \in L^2(X)$ we write f^i for its projection into the ith eigenspace, $f^i = \langle f, \psi_i\rangle \psi_i$. We have

$$\mathcal{E}[f] = \sum_{i=0}^{|X|-1} \lambda_i \|f^i\|_2^2 = \sum_{i \ge 1} \lambda_i \|f^i\|_2^2, \tag{4}$$

$$\mathbf{Var}[f] = \mathop{\mathbf{E}}_{\boldsymbol{x} \sim X}[f(\boldsymbol{x})^2] - \mathop{\mathbf{E}}_{\boldsymbol{x} \sim X}[f(\boldsymbol{x})]^2 = \sum_{i \ge 1} \|f^i\|_2^2, \tag{5}$$

the latter of these using $\psi^0 \equiv 1$. From this we immediately deduce the *Poincaré Inequality*, $\mathcal{E}[f] \ge \lambda_1 \mathbf{Var}[f]$. The quantity λ_1 is called the *spectral gap* of the random walk on $\text{Sch}(G, X, U)$. Related to this is the Log-Sobolev Inequality, discussed in Section 1.1.1. We will use the following result of Diaconis and Saloff-Coste [DSC96]:

Theorem 2.3. *$\rho \le \lambda_1$ always.*

A lower bound on ρ implies *hypercontractivity* for the operator H_t. The following is essentially due to L. Gross [Gro75]; see also [DSC96, Theorem 3.5(ii)] (wherein $\alpha = \rho/2$):

Theorem 2.4. *If ρ is the log-Sobolev constant for the random walk, and $1 \le p \le q \le \infty$ satisfy $\frac{q-1}{p-1} \le \exp(2\rho t)$, then for all $f \in L^2(X)$, we have $\|H_t f\|_q \le \|f\|_p$.*

Here $\|g\|_p$ denotes $\mathbf{E}_{\boldsymbol{x} \sim X}[|g(\boldsymbol{x})|^p]^{1/p}$.

Examples: For the Cayley graph on $\mathbb{Z}_2^n$, the spectral gap is $\lambda_1 = \frac{2}{n}$ and the log-Sobolev constant ρ is also $\frac{2}{n}$ [Gro75]. For the standard Cayley graph on $\mathbb{Z}_m^n$, $m \geq 3$, the spectral gap is well known to be $\lambda_1 = \frac{1-\cos(2\pi/m)}{n} \sim \frac{2\pi^2}{m^2 n}$ for large m. The log-Sobolev constant ρ is known to equal λ_1 for even m [CS03]; for odd m it is known [DSC96, CS03] that $\rho \geq \frac{4\pi^2/5}{m^2 n}$. For the Cayley graph on S_n generated by transpositions, Diaconis and Shahshahani [DS81] have shown the spectral gap is $\lambda_1 = \frac{2}{n-1}$. The log-Sobolev constant is known [DSC96, LY98] to satisfy $\rho = \Theta(\frac{1}{n \log n})$.

The main case of interest for our aplications is the Schreier graph on $\binom{[n]}{k}$ generated by transpositions, $0 < k < n$. Here the spectral gap is again $\lambda_1 = \frac{2}{n-1}$, independent of k [DS87]. As for the log-Sobolev constant, Lee and Yau [LY98] have shown that

$$\rho = \Theta\left(\frac{1}{n \log(1/\nu(k))}\right),$$

where we have introduced the notation

$$\nu(k) = \nu(n-k) = k(n-k)/\binom{n}{2}.$$

This quantity $\nu(k)$ is the probability the random walk takes a non-self-loop step; for the parameter range of main interest to us, when k/n is bounded away from 0 and 1, we have $\nu(k) = \Omega(1)$ and hence $\rho = \Theta(\frac{1}{n})$.

2.2 Proof of our KKL Theorem on Schreier graphs

In this section we prove Theorem 1.1, our generalization of KKL to random walks on Schreier graphs. A key hypothesis of that theorem is that the generating set U is a union of conjugacy classes. For generating sets, this condition is equivalent to closure under conjugation. The condition holds in all the example cases we described: $\mathbb{Z}_2^n$ and $\mathbb{Z}_m^n$ are abelian, so it is immediate there; for the $G = S_n$ examples it holds because the set of transpositions is a conjugacy class in G. The utility of generating sets which are a union of conjugacy classes is the following:

Proposition 2.5. *Suppose the generating set U for a Schreier graph is a union of conjugacy classes. Then L and L_u commute for every $u \in U$, and so do H_t and L_u.*

Proof. Regarding L and L_u, since subtracting from the identity operator does not affect commutativity, it suffices to show that the operators K and $A_u = id - L_u$ commute. Since $A_u f(x) = f(x^u)$, we have

$$KA_u f(x) = \mathop{\mathbf{E}}_{\boldsymbol{v} \in U}[A_u f(x^{\boldsymbol{v}})] = \mathop{\mathbf{E}}_{\boldsymbol{v} \in U}[f(x^{\boldsymbol{v}u})] = \mathop{\mathbf{E}}_{\boldsymbol{v} \in U}[f(x^{uu^{-1}\boldsymbol{v}u})].$$

It is immediate that $v \mapsto u^{-1}vu$ is an injection on U, and since U is finite it is also a bijection. Thus

$$\mathop{\mathbf{E}}_{\boldsymbol{v} \in U}[f(x^{uu^{-1}\boldsymbol{v}u})] = \mathop{\mathbf{E}}_{\boldsymbol{v} \in U}[f(x^{u\boldsymbol{v}})] = Kf(x^u) = A_u K f(x).$$

The commutativity of H_t and L_u now follows because L_u commutes with every power of L and hence with $H_t = \exp(-tL)$. □

We our now ready to prove our KKL generalization, which we restate here with explicit constants (which we have not tried to optimize):

Theorem 2.6. *In the Schreier graph* $\mathrm{Sch}(G, X, U)$*, suppose that the generating set* U *is a union of conjugacy classes. Let* ρ *denote the log-Sobolev constant for the standard random walk on the graph. Then for any* $f : X \to \{0, 1\}$*,*

$$\mathcal{E}[f] \geq \frac{1}{4} \cdot \log_3 \left(\frac{1}{2\mathcal{M}[f]} \right) \cdot \rho \cdot \mathbf{Var}[f]. \tag{6}$$

Proof. Let us fix the following parameters:

$$t = \frac{\ln 3}{2} \cdot \frac{1}{\rho}, \qquad \Lambda = \frac{2\mathcal{E}[f]}{\mathbf{Var}[f]}$$

(we clearly may assume $\rho, \mathbf{Var}[f] \neq 0$).

The proof will proceed by lower- and upper-bounding the quantity $\mathcal{E}[H_t f]$. For the lower bound, from equation (4) we have

$$\mathcal{E}[H_t f] = \sum_{i \geq 1} \lambda_i \|H_t f^i\|_2^2 = \sum_{i \geq 1} \lambda_i \exp(-2t\lambda_i) \|f^i\|_2^2 = \sum_{i \geq 1} \lambda_i 3^{-\lambda_i/\rho} \|f^i\|_2^2,$$

Here the third equality used our choice of t, and the second equality used

$$\|H_t f^i\|_2^2 = \langle H_t f^i, H_t f^i \rangle = \langle f^i, H_t H_t f^i \rangle = \langle f^i, H_{2t} f^i \rangle,$$

by the self-adjointness and the semigroup properties of H_t, along with the eigenvalue/eigenfunction decomposition of H_{2t}. We write $w(\lambda) = \lambda 3^{-\lambda/\rho}$, and drop the terms with $\lambda_i \geq \Lambda$ to conclude

$$\mathcal{E}[H_t f] \geq \sum_{i : \lambda_1 \leq \lambda_i < \Lambda} w(\lambda_i) \|f^i\|_2^2 \geq \min_{\lambda_1 \leq \lambda_i < \Lambda} \{w(\lambda_i)\} \cdot \sum_{i : \lambda_1 \leq \lambda_i < \Lambda} \|f^i\|_2^2.$$

From equations (4), (5), and and our choice of Λ, we straightforwardly deduce

$$\sum_{i : \lambda_1 \leq \lambda_i < \Lambda} \|f^i\|_2^2 \geq \frac{1}{2} \mathbf{Var}[f]. \tag{7}$$

As for the other factor, elementary calculus implies that $w(\lambda)$ is decreasing for $\lambda \geq \rho/(\ln 3)$. Since $\lambda_1 \geq \rho \geq \rho/(\ln 3)$ (using Theorem 2.3), we conclude

$$\mathcal{E}[H_t f] \geq \frac{1}{2} \cdot w(\Lambda) \cdot \mathbf{Var}[f] = \mathcal{E}[f] \cdot 3^{-\Lambda/\rho}, \tag{8}$$

using the definitions of w and Λ.

As for upper-bounding $\mathcal{E}[H_t f]$, by definition we have

$$\mathcal{E}[H_t f] = \frac{1}{2|U|} \sum_{u \in U} \|L_u H_t f\|_2^2. \tag{9}$$

Since U is a union of conjugacy classes, Proposition 2.5 implies

$$\|L_u H_t f\|_2^2 = \|H_t L_u f\|_2^2.$$

We now use the hypercontractivity Theorem 2.4, selecting $q = 2$, $p = 1 + \exp(-2\rho t) = 4/3$ (by our choice of t). This gives

$$\|H_t L_u f\|_2^2 \leq \|L_u f\|_{4/3}^2 = \mathbf{E}[|L_u f|^{4/3}]^{3/2}.$$

Since f is $\{0,1\}$-valued, so too is $|L_u f|$. Thus $|L_u f|^{4/3} = |L_u f|^2$ and so we have

$$\mathbf{E}[|L_u f|^{4/3}]^{3/2} = (\|L_u f\|_2^2)^{3/2} = (2\mathcal{I}_u[f])^{3/2} \leq 2\mathcal{I}_u[f] \cdot (2\mathcal{M}[f])^{1/2}.$$

Thus we have shown

$$\|L_u H_t f\|_2^2 \leq 2\mathcal{I}_u[f] \cdot (2\mathcal{M}[f])^{1/2}.$$

Substituting this into (9) yields

$$\mathcal{E}[H_t f] \leq \frac{1}{|U|} \sum_{u \in U} \mathcal{I}_u[f] \cdot (2\mathcal{M}[f])^{1/2} = \mathcal{E}[f] \cdot (2\mathcal{M}[f])^{1/2}. \tag{10}$$

Combining the lower and upper bounds (8) and (10) yields

$$3^{-\Lambda/\rho} \leq (2\mathcal{M}[f])^{1/2},$$

and taking logs gives us

$$\frac{\Lambda}{\rho} \geq \frac{1}{2} \log_3 \left(\frac{1}{2\mathcal{M}[f]} \right),$$

which is (6) by definition of Λ. □

2.3 Corollaries of our KKL Theorem

We now give some corollaries of Theorem 2.6. First, since $\mathcal{M}[f] \geq \mathcal{E}[f]$, we have

$$\frac{\mathcal{M}[f]}{\log(1/\mathcal{M}[f])} \geq \Omega(\rho \mathbf{Var}[f]),$$

and hence

$$\mathcal{M}[f] \geq \Omega(1) \cdot \log \left(\frac{1}{\rho \mathbf{Var}[f]} \right) \cdot \rho \mathbf{Var}[f].$$

There is no point in saving the $\mathbf{Var}[f]$ quantity inside the log, since the Log-Sobolev Inequality already gives us

$$\mathcal{E}[f] \geq \Omega(\rho \log(1/\mathbf{Var}[f]) \mathbf{Var}[f])$$

for functions with range $\{0,1\}$. Hence our corollary in the spirit of the original KKL Theorem is:

Corollary 2.7. *In the setting of Theorem 2.6, there exists at least one generator $u \in U$ with*

$$\mathcal{I}_u[f] \geq \Omega(\rho \log(1/\rho)) \cdot \mathbf{Var}[f].$$

We can now specialize our theorem and corollary to the cases discussed in the four examples. As mentioned, these all have generating sets that are a union of conjugacy classes. In the case of the Cayley graph on $\mathbb{Z}_2^n$, since $\rho = \frac{2}{n}$ we recover the classic KKL Theorem(s). In the case of the Cayley graph on $\mathbb{Z}_m^n$, we get:

Theorem 2.8. *If $f : \mathbb{Z}_m^n \to \{0,1\}$ and we denote $\mathcal{I}_i[f] = \mathbf{Pr}_{\boldsymbol{x} \sim \mathbb{Z}_m^n}[f(\boldsymbol{x}) \neq f(\boldsymbol{x} + e_i)]$, $\mathcal{M}[f] = \max_i \mathcal{I}_i[f]$, $\mathcal{E}[f] = \operatorname{avg}_i \mathcal{I}_i[f]$, then*

$$\mathcal{E}[f] \geq \Omega \left(\frac{\log(1/\mathcal{M}[f])}{m^2 n} \right) \cdot \mathbf{Var}[f], \qquad \mathcal{M}[f] \geq \Omega \left(\frac{\log n + \log m}{m^2 n} \right) \cdot \mathbf{Var}[f].$$

In the case of the Cayley graph on S_n, since $\lambda_1 = \frac{2}{n-1}$ but $\rho = \Theta(\frac{1}{n \log n})$, one can check that unfortunately neither Theorem 2.6 nor Corollary 2.7 yields new information beyond the Poincaré and Log-Sobolev Inequalities.

Our main motivation is the case of the Schreier graph on $\binom{[n]}{k}$ under transpositions, a setting where the uniform distribution is not a product distribution. Recalling $\rho = \Theta(\frac{1}{n \log(1/\nu(k))})$ in this setting, we have:

Theorem 2.9. *If* $f : \binom{[n]}{k} \to \{0,1\}$, *then*

$$\mathcal{E}[f] \geq \Omega(1) \cdot \frac{\log(1/\mathcal{M}[f])}{n \log(1/\nu(k))} \cdot \mathbf{Var}[f].$$

From this we also get Corollary 1.2 from Section 1.1.

3 Remaining results and future directions

For space considerations, the remaining discussions and proofs — how our generalized KKL Theorem implies the robust Kruskal-Katona Theorem, and how this in turn implies our monotone net (and hence optimal weak-learning) theorems — are given in the Appendix. These proofs use more "elementary" combinatorial/probabilistic arguments.

For future work, probably the most interesting direction is to extend the KKL Theorem to even more general settings. It would be especially appealing if one could weaken the the requirement that the Schreier graph's generating set be a union of conjugacy classes.

4 Acknowledgments

We thank Tom Bohman, James Lee, Alex Samorodnitsky, and Avi Wigderson for their comments and suggestions.

References

[AM06] Kazuyuki Amano and Akira Maruoka. On learning monotone Boolean functions under the uniform distribution. *Theoretical Computer Science*, 350(1):3–12, 2006.

[BBL98] Avrim Blum, Carl Burch, and John Langford. On learning monotone Boolean functions. In *Proc. 39th IEEE Symp. on Foundations of Comp. Sci.*, pages 408–415, 1998.

[BKK+92] Jean Bourgain, Jeff Kahn, Gil Kalai, Yitzhak Katznelson, and Nathan Linial. The influence of variables in product spaces. *Israel J. of Math.*, 77(1):55–64, 1992.

[BKS99] Itai Benjamini, Gil Kalai, and Oded Schramm. Noise sensitivity of Boolean functions and applications to percolation. *Inst. Hautes Études Sci. Publ. Math.*, 90:5–43, 1999.

[BOL90] Michael Ben-Or and Nathan Linial. Collective coin flipping. In Silvio Micali, editor, *Randomness and Computation*. Academic Press, New York, 1990.

[BT87] Béla Bollobás and Andrew Thomason. Threshold functions. *Combinatorica*, 7(1):35–38, 1987.

[BT96] Nader Bshouty and Christino Tamon. On the Fourier spectrum of monotone functions. *Journal of the ACM*, 43(4):747–770, 1996.

[CKK+05] Shuchi Chawla, Robert Krauthgamer, Ravi Kumar, Yuval Rabani, and D. Sivakumar. On the hardness of approximating Multicut and Sparsest-Cut. In *Proc. 20th IEEE Conference on Computational Complexity*, pages 144–153, 2005.

[CS03] Guan-Yu Chen and Yuan-Chung Sheu. On the log-Sobolev constant for the simple random walk on the n-cycle: the even cases. *J. Func. Analysis*, 202(2):473–485, 2003.

[DKSV06] Nikhil Devanur, Subhash Khot, Rishi Saket, and Nisheeth Vishnoi. Integrality gaps for sparsest cut and minimum linear arrangement problems. In *Proc. 38th ACM Symp. on the Theory of Computing*, pages 537–546, 2006.

[DS81] Persi Diaconis and Mehrdad Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, 57(2):159–179, 1981.

[DS87] Persi Diaconis and Mehrdad Shahshahani. Time to reach stationarity in the Bernoulli-Laplace diffusion model. *SIAM J. Math. Anal.*, 18(1):208–218, 1987.

[DS05] Irit Dinur and Samuel Safra. On the hardness of approximating minimum vertex cover. *Annals of Mathematics*, 162(1):439–485, 2005.

[DSC96] Persi Diaconis and Laurent Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Annals of Applied Probability*, 6(3):695–750, 1996.

[Fel68] William Feller. *An introduction to probability theory and its applications, vol. 1.* Wiley, New York, 3rd edition, 1968.

[FK96] Ehud Friedgut and Gil Kalai. Every monotone graph property has a sharp threshold. *Proc. AMS*, 124(10):2993–3002, 1996.

[Fri98] Ehud Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–36, 1998.

[Fri99] Ehud Friedgut. Sharp thresholds of graph properties, and the k-SAT problem. *J. AMS*, 12:1017–1054, 1999.

[FS07] Dvir Falik and Alex Samorodnitsky. Edge-isoperimetric inequalities and influences. *Combinatorics, Probability and Computing*, 16(5):693–712, 2007.

[GG06] Ben Graham and Geoffrey Grimmett. Influence and sharp-threshold theorems for monotonic measures. *Ann. Prob.*, 34(5):1726–1745, 2006.

[Gro75] Leonard Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97:1061–1083, 1975.

[Gro77] Jonathan Gross. Every connected regular graph of even degree is a Schreier coset graph. *J. Comb. Theory B*, 22(3):227–232, 1977.

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. of the American Mathematical Society*, 43(4):439–561, 2006.

[Kat68] Gyula O. H. Katona. A theorem of finite sets. In *Theory of Graphs: Proceedings*, pages 187–207. Academic Press, 1968.

[Kee08] Peter Keevash. Shadows and intersections: stability and new proofs. *Adv. Math.*, 218:1685–1703, 2008.

[KKL88] Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on boolean functions. In *Proc. 29th IEEE Symp. on Foundations of Comp. Sci.*, pages 68–80, 1988.

[Kle66] Daniel Kleitman. Families of non-disjoint subsets. *Journal of Combinatorial Theory*, 1:153–155, 1966.

[KR06] Robert Krauthgamer and Yuval Rabani. Improved lower bounds for embeddings into l_1. In *Proc. 17th ACM-SIAM Symp. on Discrete Algorithms*, pages 1010–1017, 2006.

[KR08] Subhash Khot and Oded Regev. Vertex Cover might be hard to approximate to within $2-\epsilon$. *Journal of Computing and Sys. Sci.*, 74(3):335–349, 2008.

[Kru63] Joseph Kruskal. The optimal number of simplices in a complex. *Math. Opt. Techniques*, pages 251–268, 1963.

[KV89] Michael Kearns and Leslie Valiant. Crytographic limitations on learning Boolean formulae and finite automata. In *Proc. 21st ACM Symp. on the Theory of Computing*, pages 433–444, 1989.

[KZ06] Jesse Kamp and David Zuckerman. Deterministic extractors for bit-fixing sources and exposure-resilient cryptography. *SIAM Journal on Computing*, 36(5):1231–1247, 2006.

[Lov79] László Lovász. *Combinatorial problems and exercises.* North-Holland Publ., Amsterdam, 1979.

[LY98] Tzong-Yau Lee and Horng-Tzer Yau. Logarithmic Sobolev inequality for some models of random walks. *Ann. Prob.*, 26(4):1855–1873, 1998.

[Mar74] Grigory Margulis. Probabilistic characteristics of graphs with large connectivity. *Problemy Peredachi Informatsii*, 10(2):101–108, 1974.

[OS08] Ryan O'Donnell and Rocco Servedio. Learning monotone decision trees in polynomial time. *SIAM Journal on Computing*, 37(3):827–844, 2008.

[Raz95] Ran Raz. Fourier analysis for probabilistic communication complexity. *Computational Complexity*, 5(3):205–221, 1995.

[Ros06] Raphaël Rossignol. Threshold for monotone symmetric properties through a logarithmic Sobolev inequality. *Ann. Prob.*, 34(5):1707–1725, 2006.

[Rus82] Lucio Russo. An approximate zero-one law. *Z. Wahrsch. Verw. Gebiete*, 61(1):129–139, 1982.

[Tal94] Michel Talagrand. On Russo's approximate zero-one law. *Ann. Prob.*, 22:1576–1587, 1994.

[Tao07] Terence Tao. Structure and randomness in combinatorics. In *Proc. 48th IEEE Symp. on Foundations of Comp. Sci.*, pages 3–15, 2007.

A Kruskal-Katona

In this section we use our KKL generalization in the setting of functions on $\binom{[n]}{k}$ to prove a robust version of the Kruskal-Katona Theorem; specifically, we will be using Theorem 2.9.

A.1 Density increase and average influence

The connection between the KKL Theorem and influences on one hand, and Kruskal-Katona and shadow densities on the other hand, is the following (recall the notation $\mu_\ell(B) = |B|/\binom{n}{\ell}$, $\nu(k) = k(n-k)/\binom{n}{2}$):

Proposition A.1. *Let $A \subseteq \binom{[n]}{k}$. Then*

$$\begin{aligned} \mu_{k-1}(\partial A) &\geq \mu_k(A) + \mathcal{E}[\mathbf{1}_A]/\nu(k), \\ \mu_{k+1}(\partial^u A) &\geq \mu_k(A) + \mathcal{E}[\mathbf{1}_A]/\nu(k). \end{aligned}$$

Proof. The statements are equivalent by Boolean duality; we prove the second. Let $\boldsymbol{x} \sim \binom{[n]}{k}$ be uniformly random. Let $\boldsymbol{i} \in [n]$ be a uniformly random index where $\boldsymbol{x_i} = 0$ and $\boldsymbol{j} \in [n]$ a uniformly random index where $\boldsymbol{x_j} = 1$. Let $\boldsymbol{y} = \boldsymbol{x}^{\boldsymbol{ij}} \in \binom{[n]}{k}$, and let $\boldsymbol{z} \in \binom{[n]}{k+1}$ be $\boldsymbol{x}$ with its $\boldsymbol{i}$th coordinate changed to 1. It's clear that $\boldsymbol{y}$ is uniformly distributed on $\binom{[n]}{k}$ and $\boldsymbol{z}$ is uniformly distributed on $\binom{[n]}{k+1}$. We also have $\boldsymbol{x}, \boldsymbol{y} \prec \boldsymbol{z}$ with certainty, so if *either* $\boldsymbol{x} \in A$ or $\boldsymbol{y} \in A$ then $\boldsymbol{z} \in \partial^u A$. Hence

$$\mu_{k+1}(\partial^u A) = \mathbf{Pr}[\boldsymbol{z} \in \partial^u A] \geq \mathbf{Pr}[\boldsymbol{x} \in A \vee \boldsymbol{y} \in A] = \mathbf{Pr}[\boldsymbol{x} \in A] + \mathbf{Pr}[\boldsymbol{x} \notin A \wedge \boldsymbol{y} \in A].$$

We have $\mathbf{Pr}[\boldsymbol{x} \in A] = \mu_k(A)$ so it remains to show that $\mathbf{Pr}[C] = \mathcal{E}[\mathbf{1}_A]/\nu(k)$, where C is the event "$\boldsymbol{x} \notin A \wedge \boldsymbol{y} \in A$". The distribution on $(\boldsymbol{x}, \boldsymbol{y})$ is clearly the same as the distribution on $(\boldsymbol{y}, \boldsymbol{x})$, so

$$\mathbf{Pr}[C] = \mathbf{Pr}[\boldsymbol{x} \in A \wedge \boldsymbol{y} \notin A] = \tfrac{1}{2}\mathbf{Pr}[\mathbf{1}_A(\boldsymbol{x}) \neq \mathbf{1}_A(\boldsymbol{y})].$$

Consider now a slightly different experiment: We choose $\boldsymbol{x'} \in \binom{[n]}{k}$ uniformly, choose $1 \leq \boldsymbol{i'} < \boldsymbol{j'} \leq n$ uniformly from among the $\binom{n}{2}$ possibilities, and then form $\boldsymbol{y'} = (\boldsymbol{x'})^{(\boldsymbol{i'j'})}$. We have $\boldsymbol{x'} \neq \boldsymbol{y'}$ iff $\boldsymbol{x'_{i'}} \neq \boldsymbol{y'_{j'}}$ which occurs with probability exactly $k(n-k)/\binom{n}{2} = \nu(k)$, and conditioned on this occurring it's easy to see that $(\boldsymbol{x'}, \boldsymbol{y'})$ has the same distribution as $(\boldsymbol{x}, \boldsymbol{y})$. Thus

$$\tfrac{1}{2}\mathbf{Pr}[\mathbf{1}_A(\boldsymbol{x}) \neq \mathbf{1}_A(\boldsymbol{y})] = \tfrac{1}{2}\mathbf{Pr}[\mathbf{1}_A(\boldsymbol{x'}) \neq \mathbf{1}_A(\boldsymbol{y'}) \mid \boldsymbol{x'} \neq \boldsymbol{y'}] = \tfrac{1}{2}(1/\nu(k)) \cdot \mathbf{Pr}[\mathbf{1}_A(\boldsymbol{x'}) \neq \mathbf{1}_A(\boldsymbol{y'})],$$

since the event $\mathbf{1}_A(\boldsymbol{x'}) \neq \mathbf{1}_A(\boldsymbol{y'})$ obviously implies $\boldsymbol{x'} \neq \boldsymbol{y'}$. But

$$\tfrac{1}{2}\mathbf{Pr}[\mathbf{1}_A(\boldsymbol{x'}) \neq \mathbf{1}_A(\boldsymbol{y'})] = \mathop{\mathbf{E}}_{\boldsymbol{i'},\boldsymbol{j'}}\left[\tfrac{1}{2}\mathop{\mathbf{Pr}}_{\boldsymbol{x'}}[\mathbf{1}_A(\boldsymbol{x'}) \neq \mathbf{1}_A((\boldsymbol{x'})^{(\boldsymbol{i'j'})})]\right] = \mathop{\mathrm{avg}}_{\boldsymbol{i'},\boldsymbol{j'}} \mathcal{I}_{\boldsymbol{i'j'}}[\mathbf{1}_A] = \mathcal{E}[\mathbf{1}_A],$$

completing the proof. □

From this result, we see that results akin to the Kruskal-Katona Corollary follow from lower bounds on the average influence of a set A. Indeed, the simplest such lower bound would use the Poincaré Inequality. As stated in Section 2.1, Diaconis and Shahshahani [DS87] established that the spectral gap λ_1 for the transposition-based random walk on $\binom{[n]}{k}$ is $\frac{2}{n-1}$, independent of k. Thus the Poincaré Inequality for $A \subseteq \binom{[n]}{k}$ is:

$$\mathcal{E}[\mathbf{1}_A] \geq \frac{2}{n-1} \cdot \mathbf{Var}[\mathbf{1}_A] = \frac{2}{n-1} \cdot \mu_k(A)(1-\mu_k(A)). \tag{11}$$

Combining this with Proposition (A.1) yields:

Theorem A.2. *Let $A \subseteq \binom{[n]}{k}$. Then*

$$\begin{aligned}\mu_{k-1}(\partial A) &\geq \mu_k(A) + \mu_k(A)(1-\mu_k(A))\frac{n}{k(n-k)},\\ \mu_{k+1}(\partial^u A) &\geq \mu_k(A) + \mu_k(A)(1-\mu_k(A))\frac{n}{k(n-k)}.\end{aligned}$$

This deduction seems to be new; we could not find it in the literature. It should be compared to the Kruskal-Katona Corollary stated in Section 1.2. The two results are in general incomparable, but are the same up to constants in the parameter range of interest to us, when k/n and $\mu_k(A)$ are bounded away from 0 and 1.

Note that (11), Theorem A.2, Proposition A.1, and the Kruskal-Katona Theorem are all made tight by "dictator sets". As described in Section 1.2, dictator sets — as well as other sets depending on a bounded number of coordinates — show that the density increase from $\mu_k(A)$ to $\mu_{k+1}(\partial^u A)$ can be as low as $\frac{1}{n}$. However our KKL generalization Theorem 2.9 implies that these "junta-type" obstructions are the only thing keeping $\mu_{k+1}(\partial^u A)$ from exceeding $\mu_k(A)$ by $\Omega(\frac{\log n}{n})$:

Theorem A.3. *Let $A \subseteq \binom{[n]}{k}$. Then*

$$\begin{aligned}\mu_{k-1}(\partial A) &\geq \mu_k(A) + \Delta,\\ \mu_{k+1}(\partial^u A) &\geq \mu_k(A) + \Delta,\end{aligned}$$

where

$$\Delta = \Omega\left(\frac{1}{\nu(k)\log(1/\nu(k))}\right)\cdot\frac{\log(1/\mathcal{M}[\mathbf{1}_A])}{n}\cdot\mu_k(A)(1-\mu_k(A)).$$

The proof follows immediately from combining Theorem 2.9 with Proposition A.1. In the parameter setting of interest to us, when k/n and $\mu_k(A)$ are bounded away from 0 and 1, we have

$$\Delta = \Omega\left(\frac{\log(1/\mathcal{M}[\mathbf{1}_A])}{n}\right).$$

Hence we have the following corollary:

Corollary A.4. *For all $\epsilon > 0$ there exists $\delta > 0$ such that the following holds: If $A \subseteq \binom{[n]}{k}$, $\epsilon \leq k/n \leq 1-\epsilon$, and $\epsilon \leq \mu_k(A) \leq 1-\epsilon$, then the following holds:*

$$\mu_{k-1}(\partial A) \geq \mu_k(A) + \delta\cdot\frac{\log n}{n}, \tag{12}$$

$$\mu_{k+1}(\partial^u A) \geq \mu_k(A) + \delta\cdot\frac{\log n}{n}, \tag{13}$$

unless there exists $1 \leq i < j \leq n$ such that $\mathcal{I}_{ij}[\mathbf{1}_A] \geq 1/n^\epsilon$.

A.2 Our robust Kruskal-Katona Theorem

The conclusion of Corollary A.4 is not completely natural: the canonical sets whose upper-shadow densities only change by $\Theta(1/n)$ are "dictator sets" such as $A_i = \{x : x_i = 1\}$. For these sets the corollary's conclusion is certainly true: $\mathcal{I}_{ij}[\mathbf{1}_{A_i}] \geq \Omega(1)$ for *any* j. But there is no canonical choice of j. We would prefer a conclusion saying that A must be noticeably "correlated" with a *single* coordinate.

Definition A.5. *Given $A \subseteq \binom{[n]}{k}$ and $i \in [n]$, we define the* correlation *of A with coordinate i to be*

$$\mathrm{corr}_i[A] = \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \binom{[n]}{k}}[\boldsymbol{x} \in A \mid \boldsymbol{x}_i = 1] - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \binom{[n]}{k}}[\boldsymbol{x} \in A \mid \boldsymbol{x}_i = 0].$$

We observe that it is *not* true that $\mathcal{I}_{ij}[\mathbf{1}_A]$ being large implies either $\mathrm{corr}_i[A]$ or $\mathrm{corr}_j[A]$ is large. For example, if

$$A = \{x : x_1 \oplus x_2 \oplus \cdots \oplus x_{n/2} = 1\} \subseteq \binom{[n]}{n/2},$$

(here $\oplus$ denotes exclusive-or) then, e.g., $\mathcal{I}_{1n}[A] \approx 1/4$ but $\mathrm{corr}_1[A]$ and $\mathrm{corr}_n[A]$ are negligible. However, we can show that if $\mathcal{I}_{ij}[\mathbf{1}_A]$ is large *because* one of (12), (13) fails, then in fact at least one of $\mathrm{corr}_i[A]$ and $\mathrm{corr}_j[A]$ is large:

Proposition A.6. *Suppose $A \subseteq \binom{[n]}{k}$, with*

$$\mu_{k+1}(\partial^u A) - \mu_k(A) \leq \eta. \tag{14}$$

Suppose further that

$$\mathcal{I}_{ij}[\mathbf{1}_A] \geq 2\gamma.$$

Then either $\mathrm{corr}_i[A]$ or $\mathrm{corr}_j[A]$ is at least

$$\frac{\gamma}{\mathbf{Pr}[\boldsymbol{x}_1 = 0]} - \frac{\eta}{\mathbf{Pr}[\boldsymbol{x}_1 = 1]} = \frac{n}{n-k} \cdot \gamma - \frac{n}{k} \cdot \eta.$$

In particular, if k/n is bounded away from 0 and $\gamma \geq C\eta$ for a sufficiently large constant C, then one of the correlations is at least $\Omega(\gamma)$. Observe also that the conclusion cannot be that *both* $\mathrm{corr}_i[A]$ and $\mathrm{corr}_j[A]$ must be large, by the example of dictator sets.

Proof. Assume without loss of generality that $i = 1$, $j = 2$. We write a uniformly random $\boldsymbol{x} \sim \binom{[n]}{k}$ as $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{z})$, where $\boldsymbol{z} \in \{0,1\}^{n-2}$. The assumption $\mathcal{I}_{12}[\mathbf{1}_A] \geq 2\gamma$ is equivalent to $\mathbf{Pr}[C] \geq 4\gamma$, where B is the event

$$B = \text{“}\boldsymbol{x}_1 \neq \boldsymbol{x}_2 \ \wedge \ \mathbf{1}_A(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{z}) \neq \mathbf{1}_A(\boldsymbol{x}_2, \boldsymbol{x}_1, \boldsymbol{z})\text{”}.$$

The event B can be partitioned into the following two mutually exclusive events:

$$\text{“}\boldsymbol{x}_1 \neq \boldsymbol{x}_2 \ \wedge \ (0,1,\boldsymbol{z}) \notin A \ \wedge \ (1,0,\boldsymbol{z}) \in A\text{”} \qquad \text{or} \qquad \text{“}\boldsymbol{x}_1 \neq \boldsymbol{x}_2 \ \wedge \ (0,1,\boldsymbol{z}) \in A \ \wedge \ (1,0,\boldsymbol{z}) \notin A\text{”}.$$

Hence one of these occurs with probability at least 2γ; without loss of generality, say it is the first, in which case we will show the lower bound on $\mathrm{corr}_1[A]$. So we have

$$\mathbf{Pr}[\boldsymbol{x}_1 \neq \boldsymbol{x}_2 \ \wedge \ (0,1,\boldsymbol{z}) \notin A \ \wedge \ (1,0,\boldsymbol{z}) \in A] \geq 2\gamma.$$

Whenever $(1,0,\boldsymbol{z}) \in A$, certainly $(1,1,\boldsymbol{z}) \in \partial^u A$. Hence,

$$\mathbf{Pr}[\boldsymbol{x}_1 \neq \boldsymbol{x}_2 \ \wedge \ (0,1,\boldsymbol{z}) \notin A \ \wedge \ (1,1,\boldsymbol{z}) \in \partial^u A] \geq 2\gamma.$$

By symmetry, $(\boldsymbol{x}_1, \boldsymbol{x}_2) = (0,1)$, $(\boldsymbol{x}_1, \boldsymbol{x}_2) = (1,0)$ occur with equal probability conditioned on $\boldsymbol{z}$, so

$$\begin{aligned} \mathbf{Pr}[\boldsymbol{x}_1 = 0 \ \wedge \ \boldsymbol{x}_2 = 1 \ \wedge \ (0,1,\boldsymbol{z}) \notin A \ \wedge \ (1,1,\boldsymbol{z}) \in \partial^u A] &\geq \gamma \\ \Rightarrow \quad \mathbf{Pr}[\boldsymbol{x}_1 = 0 \ \wedge \ \boldsymbol{x} \notin A \ \wedge \ (1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_n) \in \partial^u A] &\geq \gamma. \end{aligned}$$

Thus we have established

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\binom{[n]}{k}}[\boldsymbol{x} \notin A \ \wedge \ (1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n) \in \partial^u A \mid \boldsymbol{x}_1 = 0] \geq \frac{\gamma}{\mathbf{Pr}[\boldsymbol{x}_1 = 0]}. \tag{15}$$

This begins to look like the claim that $\mathrm{corr}_1[A]$ is lower-bounded. We will use condition (14) to complete the proof. The key to what remains is studying four quantities:

$$\mu^0 := \mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\binom{[n]}{k}}[\boldsymbol{x} \in A \mid \boldsymbol{x}_1 = 0], \qquad \mu^1 := \mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\binom{[n]}{k}}[\boldsymbol{x} \in A \mid \boldsymbol{x}_1 = 1]$$

$$\mu^0_+ := \mathop{\mathbf{Pr}}_{\boldsymbol{y}\sim\binom{[n]}{k+1}}[\boldsymbol{y} \in \partial^u A \mid \boldsymbol{y}_1 = 0], \qquad \mu^1_+ := \mathop{\mathbf{Pr}}_{\boldsymbol{y}\sim\binom{[n]}{k+1}}[\boldsymbol{y} \in \partial^u A \mid \boldsymbol{y}_1 = 1].$$

We will take these parameters to be 0 if the event being conditioned on occurs with probability 0.

By definition, $\mathrm{corr}_1[A] = \mu^1 - \mu^0$ (and also $\mathrm{corr}_1(\partial^u A) = \mu^1_+ - \mu^0_-$). Next, observe that the distribution $(\boldsymbol{y} \mid \boldsymbol{y}_1 = 0)$ can be gotten by choosing $(\boldsymbol{x} \mid \boldsymbol{x}_1 = 0)$ and then changing a randomly chosen 0 from $\boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ into a 1. Under this coupling, $\boldsymbol{x} \in A \Rightarrow \boldsymbol{y} \in \partial^u A$. Hence we may conclude

$$\mu^0 \leq \mu^0_+, \qquad \text{and similarly,} \qquad \mu^1 \leq \mu^1_+. \tag{16}$$

It's also not hard to show that $\mu^0 \leq \mu^1_+$, but we will improve this using (15). Let $\boldsymbol{x}' \sim \binom{[n-1]}{k}$ be uniform. Since the event "$(0, \boldsymbol{x}') \in A$" implies the event "$(1, \boldsymbol{x}') \in \partial^u A$", we have

$$\mathbf{Pr}[(1, \boldsymbol{x}') \in \partial^u A] - \mathbf{Pr}[(0, \boldsymbol{x}') \in A] = \mathop{\mathbf{Pr}}_{\boldsymbol{x}'}[(0, \boldsymbol{x}') \notin A \ \wedge \ (1, \boldsymbol{x}') \in \partial^u A].$$

But $(0, \boldsymbol{x}')$ is distributed as $(\boldsymbol{x} \mid \boldsymbol{x}_1 = 0)$ and $(1, \boldsymbol{x}')$ is distributed as $(\boldsymbol{y} \mid \boldsymbol{y}_1 = 1)$. Thus (15) implies that the right-hand side of the above is at least $\gamma / \mathbf{Pr}[\boldsymbol{x}_1 = 0]$, and we conclude

$$\mu^1_+ - \mu^0 \geq \frac{\gamma}{\mathbf{Pr}[\boldsymbol{x}_1 = 0]}. \tag{17}$$

Finally, by hypothesis (14) we have

$$\mathbf{Pr}[\boldsymbol{y}_1 = 1]\mu^1_+ + \mathbf{Pr}[\boldsymbol{y}_1 = 0]\mu^0_+ - \mathbf{Pr}[\boldsymbol{x}_1 = 1]\mu^1 - \mathbf{Pr}[\boldsymbol{x}_1 = 0]\mu^0 \leq \eta,$$

which implies (using (16), (17) in the second step below),

$$\begin{aligned}
\mathbf{Pr}[\boldsymbol{x}_1 = 1]\mu^1 &\geq \mathbf{Pr}[\boldsymbol{y}_1 = 1]\mu^1_+ + \mathbf{Pr}[\boldsymbol{y}_1 = 0]\mu^0_+ - \mathbf{Pr}[\boldsymbol{x}_1 = 0]\mu^0 - \eta \\
&\geq \mathbf{Pr}[\boldsymbol{y}_1 = 1]\left(\mu^0 + \frac{\gamma}{\mathbf{Pr}[\boldsymbol{x}_1 = 0]}\right) + \mathbf{Pr}[\boldsymbol{y}_1 = 0]\mu^0 - \mathbf{Pr}[\boldsymbol{x}_1 = 0]\mu^0 - \eta \\
&= \left(\mathbf{Pr}[\boldsymbol{y}_1 = 1] + \mathbf{Pr}[\boldsymbol{y}_1 = 0] - \mathbf{Pr}[\boldsymbol{x}_1 = 0]\right)\mu^0 + \frac{\mathbf{Pr}[\boldsymbol{y}_1 = 1]}{\mathbf{Pr}[\boldsymbol{x}_1 = 0]} \cdot \gamma - \eta \\
\Rightarrow \mu^1 - \mu^0 &\geq \frac{\mathbf{Pr}[\boldsymbol{y}_1 = 1]}{\mathbf{Pr}[\boldsymbol{x}_1 = 0]\mathbf{Pr}[\boldsymbol{x}_1 = 1]} \cdot \gamma - \frac{\eta}{\mathbf{Pr}[\boldsymbol{x}_1 = 1]} \\
\Rightarrow \mathrm{corr}_1[A] &\geq \frac{\gamma}{\mathbf{Pr}[\boldsymbol{x}_1 = 0]} - \frac{\eta}{\mathbf{Pr}[\boldsymbol{x}_1 = 1]}.
\end{aligned}$$

□

Combining Proposition A.6 with Theorem A.3 yields a natural robust Kruskal-Katona Theorem. Or, if we use the simpler Corollary A.4, we get the robust Kruskal-Katona Theorem stated in Section 1.2, namely Theorem 1.3.

B Nets for monotone functions

In this section we use our robust Kruskal-Katona theorem to prove that $\{0, 1, x_1, \ldots, x_n, \mathrm{Maj}\}$ is a $(1/2 - \Omega(\frac{\log n}{\sqrt{n}}))$-net for the class of monotone n-bit functions. Indeed, we will prove the stronger statement Theorem 1.7. Recall that Theorem 1.7 also immediately implies our optimal weak-learning algorithm for monotone functions under the uniform distribution, Theorem 1.8.

Starting in this section, we will make a notational switch which greatly simplifies various expressions: we will write -1 and 1 instead of 0 and 1. So henceforth $\binom{[n]}{k}$ will denote strings x in $\{-1,1\}^n$ with exactly k 1's (which we continue to denote by $|x| = k$), and boolean functions will map $\{-1,1\}^n \to \{-1,1\}$. Note that with this notation,

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \{-1,1\}^n}[f(\boldsymbol{x}) = h(\boldsymbol{x})] = \frac{1}{2} + \frac{1}{2} \cdot \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \{-1,1\}^n}[f(\boldsymbol{x})h(\boldsymbol{x})].$$

We define the *correlation* of $f, h : \{-1,1\}^n \to \{-1,1\}$ to be $\mathbf{E}_{\boldsymbol{x}}[f(\boldsymbol{x})h(\boldsymbol{x})]$. This is also twice the "advantage of hypothesis h for function f". With this new notation, our goal Theorem 1.7 becomes equivalent (up to constants) to:

Theorem B.1. *For all $0 < \epsilon < 1/2$ there exists $0 < \delta < 1$ such that the following holds: Let $f : \{-1,1\}^n \to \{-1,1\}$ be monotone. Then at least one of the below statements about correlation with f must hold:*

$$\textit{a constant } \pm 1 \textit{ has correlation} \geq 1 - \epsilon; \tag{18}$$

$$\textit{a dictator } x_1, x_2, \ldots, \textit{ or } x_n \textit{ has correlation} \geq \frac{1}{n^\epsilon}; \tag{19}$$

$$\textit{Majority has correlation} \geq \delta \cdot \frac{\log n}{\sqrt{n}}. \tag{20}$$

As a small note, we need to decide how Maj is defined on strings x with $|x| = n/2$ when n is even. In fact, since $\mathbf{Pr}_{\boldsymbol{x}}[|\boldsymbol{x}| = n/2] \leq O(1/\sqrt{n})$, for any fixed ϵ and hence δ, one can define Maj arbitrarily on the middle layer and it will not change Theorem B.1 for sufficiently large n. For convenience, then, we will define $\mathrm{Maj}(x) = 0$ for $|x| = n/2$, n even; this is well-defined given our notion of correlation.

Proof. The strategy will be to show that if (18) and (20) fail, then (19) must hold. We begin with a straightforward calculation: for any $0 \leq k < n/2$, the correlation of Maj with f is

$$\begin{aligned}
\mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x})\mathrm{Maj}(\boldsymbol{x})] &= \mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x}) \mid |\boldsymbol{x}| \geq n-k] \cdot \mathbf{Pr}[|\boldsymbol{x}| \geq n-k] && (21)\\
&- \mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x}) \mid |\boldsymbol{x}| \leq k] \cdot \mathbf{Pr}[|\boldsymbol{x}| \leq k] && (22)\\
&+ \mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x}) \mid n/2 < |\boldsymbol{x}| < n-k] \cdot \mathbf{Pr}[n/2 < |\boldsymbol{x}| < n-k] && (23)\\
&- \mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x}) \mid k < |\boldsymbol{x}| < n/2] \cdot \mathbf{Pr}[k < |\boldsymbol{x}| < n/2]. && (24)
\end{aligned}$$

(Recall our convention that $\mathrm{Maj}(x) = 0$ if $|x| = n/2$, and interpret (23), (24) as 0 if k is close enough to $n/2$ that the event therein has probability 0.) By symmetry, the probabilities in (23), (24) are identical. Furthermore, the expectation in (23) is at least that in (24), by monotonicity. Hence the contribution from (23), (24) is nonnegative. As well, the probabilities in (21), (22) are identical. Hence we conclude

$$\mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x})\mathrm{Maj}(\boldsymbol{x})] \geq \mathbf{Pr}[|\boldsymbol{x}| \leq k] \cdot \left(\mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x}) \mid |\boldsymbol{x}| \geq n-k] - \mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x}) \mid |\boldsymbol{x}| \leq k] \right). \tag{25}$$

As a small point, this fact allows us to freely assume that n is sufficiently large as a function of the constant ϵ; for otherwise, by taking δ small enough the theorem reduces to showing that $\mathbf{E}[f \cdot \mathrm{Maj}] > 0$ assuming f is not constant, and this is implied by taking $k = 0$ in (25). It also shows the (well-known) fact that $\mathbf{E}[f \cdot \mathrm{Maj}] \geq 0$ for monotone f.

We proceed with some additional straightforward calculations. Let

$$m = \lceil n/2 - 1 \rceil, \qquad , m' = n - m = \lfloor n/2 + 1 \rfloor,$$

indexing the two slices just outside $n/2$. If we take $k = m$ in (25) we get

$$\mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x})\mathrm{Maj}(\boldsymbol{x})] = \mathbf{Pr}[|\boldsymbol{x}| \geq m'] \cdot \mu_+ - \mathbf{Pr}[|\boldsymbol{x}| \leq m] \cdot \mu_-,$$

where

$$\mu_+ = \mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x}) \mid |\boldsymbol{x}| \geq m'], \qquad \mu_- = \mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x}) \mid |\boldsymbol{x}| \leq m].$$

But $|\boldsymbol{x}| \geq m'$ iff $|\boldsymbol{x}| > n/2$, $|\boldsymbol{x}| \leq m$ iff $|\boldsymbol{x}| < n/2$. Since $\mathbf{Pr}[|\boldsymbol{x}| = n/2] \leq O(1/\sqrt{n})$ we have $\mathbf{Pr}[|\boldsymbol{x}| > n/2] = \mathbf{Pr}[|\boldsymbol{x}| < n/2] = 1/2 \pm O(1/\sqrt{n})$. So

$$\mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x})\mathrm{Maj}(\boldsymbol{x})] = \tfrac{1}{2}(\mu_+ - \mu_-) \pm O(1/\sqrt{n}).$$

Similarly, and even simpler,

$$\mathbf{E}[f] = \tfrac{1}{2}(\mu_+ + \mu_-) \pm O(1/\sqrt{n}).$$

We now come to the main part of the proof. Suppose we assume that (18), (20) fail (where we will show how to choose $\delta = \delta(\epsilon)$ later). Then we derive

$$0 \leq \mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x})\mathrm{Maj}(\boldsymbol{x})] = \tfrac{1}{2}(\mu_+ - \mu_-) \pm O(1/\sqrt{n}) \leq \frac{\log n}{\sqrt{n}},$$

i.e., $\mu_- = \mu_+ \pm O(1/\sqrt{n})$, and

$$|\tfrac{1}{2}(\mu_+ + \mu_-) \pm O(1/\sqrt{n})| \leq 1 - \epsilon,$$

hence

$$|\mu_+|, |\mu_-| \leq 1 - \epsilon/2$$

(where we've used $\epsilon/2 \geq O(1/\sqrt{n})$). By monotonicity, $\mu_+ \geq \mu_{m'}$ and $\mu_- \leq \mu_m$, where

$$\mu_m = \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \binom{[n]}{m}}[f(\boldsymbol{x})], \qquad \mu_{m'} = \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \binom{[n]}{m'}}[f(\boldsymbol{x})]. \tag{26}$$

Thus we conclude

$$|\mu_{m'}|, |\mu_m| \leq 1 - \epsilon/2. \tag{27}$$

We next derive an additional conclusion from (20) failing. Define b ("bottom") and t ("top") by

$$b = \lfloor n/2 - \sqrt{n}/2 \rfloor, \qquad t = n - \ell = \lceil n/2 + \sqrt{n}/2 \rfloor$$

(where we have $0 < b < t < n$ if n is sufficiently large), and define μ_b, μ_t analogously with (26). Because (20) fails, we may assume

$$\mu_t - \mu_b \leq 10\delta \cdot \frac{\log n}{\sqrt{n}}. \tag{28}$$

For otherwise, since $\mathbf{E}_{\boldsymbol{x}}[f(\boldsymbol{x}) \mid |\boldsymbol{x}| \geq t] \geq \mu_h$ and $\mathbf{E}_{\boldsymbol{x}}[f(\boldsymbol{x}) \mid |\boldsymbol{x}| \leq b] \leq \mu_b$ by monotonicity, inequality (25) implies

$$\mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x})\mathrm{Maj}(\boldsymbol{x})] \geq \mathbf{Pr}[|\boldsymbol{x}| \leq b] \cdot 10\delta \cdot \frac{\log n}{\sqrt{n}} \geq \delta \cdot \frac{\log n}{\sqrt{n}},$$

where we used $\mathbf{Pr}[|\boldsymbol{x}| \leq \lfloor n/2 - \sqrt{n}/2 \rfloor] \geq .1$ by the Central Limit Theorem (for sufficiently large n). We also have $\mu_b \leq \mu_m < \mu_{m'} \leq \mu_t$ by monotonicity; combining this with (27) and (28) yields

$$|\mu_t|, |\mu_b| \leq 1 - \epsilon/2 + 10\delta \cdot \frac{\log n}{\sqrt{n}} \leq 1 - \epsilon/4 \tag{29}$$

(for n sufficiently large).

Finally, since $t - b \geq \sqrt{n}$, inequality (28) certainly implies that there exists k satisfying $b \leq k < t$ and

$$\mu_{k+1} - \mu_k \leq 10\delta \cdot \frac{\log n}{n}. \tag{30}$$

By monotonicity we have $\mu_b \leq \mu_k \leq \mu_t$ and hence from (29) we get $|\mu_k| \leq 1 - \epsilon/4$. Since we also have $k/n = 1/2 \pm O(1/\sqrt{n})$, we now apply our robust Kruskal-Katona Theorem, specifically Theorem 1.3, taking $A = f^{-1}(1) \cap \binom{[n]}{k}$. (Throughout, recall that n may be assumed sufficiently large.) By monotonicity, f is 1 on all strings in $\partial^u A$. Thus, taking δ suitably small as a function of ϵ, we deduce from (30) that there exists some $i \in [n]$ with

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \binom{[n]}{k}}[f(\boldsymbol{x}) = 1 \mid \boldsymbol{x}_i = 1] - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \binom{[n]}{k}}[f(\boldsymbol{x}) = 1 \mid \boldsymbol{x}_i = -1] \geq \frac{12}{n^\epsilon}.$$

One can easily check that the quantity on the left here is at most

$$2 \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \binom{[n]}{k}}[f(\boldsymbol{x})\boldsymbol{x}_i] + 2\left| \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \binom{[n]}{k}}[\boldsymbol{x}_i = 1] - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \binom{[n]}{k}}[\boldsymbol{x}_i = -1] \right|;$$

hence we get

$$\mathop{\mathbf{E}}_{\boldsymbol{x} \sim \binom{[n]}{k}}[f(\boldsymbol{x})\boldsymbol{x}_i] \geq \frac{5}{n^\epsilon}, \tag{31}$$

using $|\mathbf{Pr}_{\boldsymbol{x} \sim \binom{[n]}{k}}[\boldsymbol{x}_i = 1] - \mathbf{Pr}_{\boldsymbol{x} \sim \binom{[n]}{k}}[\boldsymbol{x}_i = -1]| \leq O(1/\sqrt{n}) \leq 1/n^\epsilon$.

In fact, (31) together with (28) implies

$$\mathop{\mathbf{E}}_{\boldsymbol{x} \sim \binom{[n]}{j}}[f(\boldsymbol{x})\boldsymbol{x}_i] \geq \frac{4}{n^\epsilon} \qquad \forall\ b \leq j \leq t. \tag{32}$$

We illustrate this for the case where $j > k$; the $j < k$ case is similar (using the other inequality in the lemma). When $j > k$, we use Lemma B.2 to yield

$$\frac{1}{j}\mu_j - \frac{1}{k}\mu_k \geq \frac{1}{k} \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \binom{[n]}{k}}[f(\boldsymbol{x})\boldsymbol{x}_i] - \frac{1}{j} \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \binom{[n]}{k}}[f(\boldsymbol{x})\boldsymbol{x}_i].$$

As $j/k = 1 + O(1/\sqrt{n})$, we can deduce that

$$\mu_j - \mu_k \geq \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \binom{[n]}{k}}[f(\boldsymbol{x})\boldsymbol{x}_i] - \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \binom{[n]}{k}}[f(\boldsymbol{x})\boldsymbol{x}_i] + O(1/\sqrt{n}).$$

We know from monotonicity and (28) that $\mu_j - \mu_k \leq 10\delta \cdot \frac{\log n}{\sqrt{n}}$. Because $O(1/\sqrt{n})$ and $10 \cdot \delta \frac{\log n}{\sqrt{n}}$ are negligible compared to $1/n^\epsilon$, we get the result we want.

We now conclude:

$$\begin{aligned} \mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x})\boldsymbol{x}_i] &= \mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x})\boldsymbol{x}_i \mid b \leq |\boldsymbol{x}| \leq t] \cdot \mathbf{Pr}[b \leq |\boldsymbol{x}| \leq t] && (33) \\ &+ \mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x})\boldsymbol{x}_i \mid |\boldsymbol{x}| < b \text{ or } |\boldsymbol{x}| > t] \cdot \mathbf{Pr}[|\boldsymbol{x}| < b \text{ or } |\boldsymbol{x}| > t]. && (34) \end{aligned}$$

The expectation in (33) is at least $4/n^\epsilon$, by (32). The probability therein is at least $1/2$, by the Central Limit Theorem (assuming n large enough). Hence (33) is at least $2/n^\epsilon$. On the other hand, by Lemma B.3 below, (34) is at least $-O(1/\sqrt{n})$. Hence we have

$$\mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x})\boldsymbol{x}_i] \geq \frac{2}{n^\epsilon} - O(1/\sqrt{n}) \geq \frac{1}{n^\epsilon},$$

establishing (19) and completing the proof. □

Below are the two lemmas we needed in the preceding proof.

Lemma B.2. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be monotone and let $i \in [n]$. Write $\mu_k = \mathbf{E}_{\boldsymbol{x} \sim \binom{[n]}{k}}[f(\boldsymbol{x})]$ and $\theta_k(i) = \mathbf{E}_{\boldsymbol{x} \sim \binom{[n]}{k}}[f(\boldsymbol{x})\boldsymbol{x}_i]$. Then for any $0 < k \leq \ell < n$,*

$$\begin{aligned} \frac{\mu_\ell}{\ell} - \frac{\mu_k}{k} &\geq \frac{\theta_k(i)}{k} - \frac{\theta_\ell(i)}{\ell}, \\ \frac{\mu_\ell}{n-\ell} - \frac{\mu_k}{n-k} &\geq \frac{\theta_\ell(i)}{n-\ell} - \frac{\theta_k(i)}{n-k}. \end{aligned}$$

Proof. The two statements imply each other, replacing $f(x)$ with $-f(-x)$; hence it suffices to prove the first one. Write

$$\mu_k^- = \mathop{\mathbf{E}}_{\boldsymbol{x} \in \binom{[n]}{k}}[f(\boldsymbol{x}) \mid \boldsymbol{x}_i = -1],$$

and similarly define μ_k^+, μ_ℓ^-, μ_ℓ^+. We have

$$\begin{aligned} \mu_k &= \frac{k}{n} \cdot \mu_k^+ + \left(1 - \frac{k}{n}\right) \cdot \mu_k^-, \\ \theta_k(i) &= \frac{k}{n} \cdot \mu_k^+ - \left(1 - \frac{k}{n}\right) \cdot \mu_k^-, \end{aligned}$$

and similarly for ℓ. Thus the desired inequality holds iff

$$\frac{1}{\ell}(\mu_\ell + \theta_\ell(i)) \geq \frac{1}{k}(\mu_k + \theta_k(i)) \quad \Leftrightarrow \quad \frac{2}{n}\mu_\ell^+ \geq \frac{2}{n}\mu_k^+.$$

But this last inequality is true; it follows from the fact that restricting f by fixing the ith bit to 1 gives a monotone function. □

Lemma B.3. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be monotone and let $i \in [n]$. For integer $n/2 \leq k \leq n/2 + \sqrt{n}$,*

$$\mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x})\boldsymbol{x}_i \mid |\boldsymbol{x}| > k] \geq -O(1/\sqrt{n}).$$

Similarly, for integer $n/2 - \sqrt{n} \leq k \leq n/2$,

$$\mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x})\boldsymbol{x}_i \mid |\boldsymbol{x}| < k] \geq -O(1/\sqrt{n}).$$

Proof. The two statements imply each other, replacing $f(x)$ with $-f(-x)$, so it suffices to prove the first. We have

$$\begin{aligned}\mathbf{E}[f(\boldsymbol{x})\boldsymbol{x}_i \mid |\boldsymbol{x}| > k] &= \mathbf{E}[f(\boldsymbol{x}) \mid |\boldsymbol{x}| > k+1, \boldsymbol{x}_i = 1] \cdot \mathbf{Pr}[\boldsymbol{x}_i = 1, |\boldsymbol{x}| > k+1 \mid |\boldsymbol{x}| > k] \\ &+ \mathbf{E}[f(\boldsymbol{x}) \mid |\boldsymbol{x}| = k+1, \boldsymbol{x} = 1] \cdot \mathbf{Pr}[\boldsymbol{x}_i = 1, |\boldsymbol{x}| = k+1 \mid |\boldsymbol{x}| > k] \\ &- \mathbf{E}[f(\boldsymbol{x}) \mid |\boldsymbol{x}| > k, \boldsymbol{x}_i = -1] \cdot \mathbf{Pr}[\boldsymbol{x}_i = -1 \mid |\boldsymbol{x}| > k].\end{aligned}$$

The second quantity above is at most $\mathbf{Pr}[|\boldsymbol{x}| = k+1 \mid |\boldsymbol{x}| > k]$ in absolute value, which is at most $O(1/\sqrt{n})$, using the fact that $k \le n/2 + \sqrt{n}$. And in the first and third quantities, the probability factors are $1/2 \pm O(1/\sqrt{n})$. Hence up to an additive $\pm O(1/\sqrt{n})$, the whole expression is equal to

$$\frac{1}{2}\left(\mathbf{E}[f(\boldsymbol{x}) \mid |\boldsymbol{x}| > k+1, \boldsymbol{x}_i = 1] - \mathbf{E}[f(\boldsymbol{x}) \mid |\boldsymbol{x}| > k, \boldsymbol{x}_i = -1]\right).$$

But we can prove this is nonnegative, completing the proof, by a coupling argument: Simply note that we can draw $(\boldsymbol{x} \mid |\boldsymbol{x}| > k+1, \boldsymbol{x}_i = 1)$ by first drawing $(\boldsymbol{x} \mid |\boldsymbol{x}| > k, \boldsymbol{x}_i = -1)$ and then changing the ith coordinate from -1 to 1; and, note that this can only increase f's value, by monotonicity. □

Notice that this $(1/2 - \Omega(\log n/\sqrt{n}))$-net for monotone functions contains $n+3$ functions. An interesting question is how small such a net can be. We now give an explicit $(1/2 - \Omega(\log n/\sqrt{n}))$-net containing $O(n/\log n)$ many functions.

Let t be an integer, and partition the variables into blocks $\{B_1, B_2, \ldots, B_{\lceil \frac{n}{t} \rceil}\}$ where $|B_i| \le t$ for all i. Letting Maj_{B_i} be the majority function where only the bits in B_i are relevant, we prove the following:

Theorem B.4. *For all $0 < \epsilon < 1/2$ there exists $0 < \delta < 1$ such that the following holds: Let $f : \{-1,1\}^n \to \{-1,1\}$ be monotone. Partition the variables into blocks B_i, each of size at most t. Then at least one of the below statements about correlation with f must hold:*

$$\text{a constant } \pm 1 \text{ has correlation} \ge 1 - \epsilon; \tag{35}$$

$$\mathrm{Maj}_{B_i} \text{ has correlation} \ge 2^{-t+1} \cdot \frac{1}{n^\epsilon}; \tag{36}$$

$$\text{Majority has correlation} \ge \delta \cdot \frac{\log n}{\sqrt{n}}. \tag{37}$$

Proof. We will directly use Theorem B.1. Suppose that the first and third case do not hold. (The first and third cases of Theorem B.1 and the theorem to prove are exactly the same.) Then Theorem B.1 says that some variable has $1/n^\epsilon$ correlation with f. Without loss of generality, assume this variable is x_1, that x_1 is in B_1, and $|B_1| = t$.

We first prove the following lemma:

Lemma B.5. *Let $f : \{-1,1\}^t \to \{-1,1\}$ be a monotone function, and let* Maj *be the majority function on all t bits. Then*

$$\mathop{\mathbf{E}}_{\boldsymbol{x}}[\mathrm{Maj}(\boldsymbol{x})f(\boldsymbol{x})] \ge 2^{-t+1} \mathop{\mathbf{E}}_{\boldsymbol{x}}[\boldsymbol{x}_1 f(\boldsymbol{x})].$$

Proof. Because Maj and f are both monotone functions, the left hand side is nonnegative. If $\mathbf{E}_{\boldsymbol{x}}[\mathrm{Maj}(\boldsymbol{x})f(\boldsymbol{x})] > 0$, then clearly $\mathbf{E}_{\boldsymbol{x}}[\mathrm{Maj}(\boldsymbol{x})f(\boldsymbol{x})] \ge 2^{-t+1}$, and because $\mathbf{E}_{\boldsymbol{x}}[\boldsymbol{x}_1 f(\boldsymbol{x})] \le 1$ we are done. Suppose $\mathbf{E}_{\boldsymbol{x}}[\mathrm{Maj}(\boldsymbol{x})f(\boldsymbol{x})] = 0$. We have shown in our proof of Theorem B.1 (specifically, using equation 25) that f must be constant. In this case, both sides of the inequality are 0, completing the proof of the lemma. □

We partition the boolean cube into 2^{n-t} subcubes for each setting of the bits not in B_1. We will partition a string $x \in \{-1,1\}^n$ into $(x_{B_1}, x_{\overline{B_1}})$ in the natural way. Then

$$\mathop{\mathbf{E}}_{\boldsymbol{x}}[\mathrm{Maj}_{B_1}(\boldsymbol{x})f(\boldsymbol{x})] = \mathop{\mathbf{E}}_{\boldsymbol{x}_{\overline{B_1}}}[\mathop{\mathbf{E}}_{\boldsymbol{x}_{B_1}}[\mathrm{Maj}_{B_1}((\boldsymbol{x}_{B_1}, x_{\overline{B_1}}))f((\boldsymbol{x}_{B_1}, x_{\overline{B_1}}))|\boldsymbol{x}_{\overline{B_1}} = x_{\overline{B_1}}]]$$

In the inner expectation, the bits not in B_1 are fixed. Under this restriction, f is still a monotone function, and Maj_{B_1} is the majority of the t unset bits. Thus we can apply the lemma to the inner expectation, yielding:

$$\begin{aligned}
\mathop{\mathbf{E}}_{\boldsymbol{x}_{\overline{B_1}}}[\mathop{\mathbf{E}}_{\boldsymbol{x}_{B_1}}[\mathrm{Maj}_{B_1}((\boldsymbol{x}_{B_1}, x_{\overline{B_1}}))f((\boldsymbol{x}_{B_1}, x_{\overline{B_1}}))|\boldsymbol{x}_{\overline{B_1}} = x_{\overline{B_1}}]] &\geq \mathop{\mathbf{E}}_{\boldsymbol{x}_{\overline{B_1}}}[2^{-t+1}\mathop{\mathbf{E}}_{\boldsymbol{x}_{B_1}}[\boldsymbol{x}_1 f((\boldsymbol{x}_{B_1}, x_{\overline{B_1}}))|\boldsymbol{x}_{\overline{B_1}} = x_{\overline{B_1}}]] \\
&= 2^{-t+1}\mathop{\mathbf{E}}_{\boldsymbol{x}_{\overline{B_1}}}[\mathop{\mathbf{E}}_{\boldsymbol{x}_{B_1}}[\boldsymbol{x}_1 f((\boldsymbol{x}_{B_1}, x_{\overline{B_1}}))|\boldsymbol{x}_{\overline{B_1}} = x_{\overline{B_1}}]] \\
&= 2^{-t+1}\mathop{\mathbf{E}}_{\boldsymbol{x}}[\boldsymbol{x}_1 f(\boldsymbol{x})] \\
&\geq 2^{-t+1}/n^{\epsilon},
\end{aligned}$$

which is what we wanted to prove.

□

The following corollary comes from setting $t = \log(n^{\frac{1}{2}-\epsilon}/(\delta \log n))$ and setting the sizes of the blocks as equal as possible:

Corollary B.6. *For all $0 < \epsilon < 1/2$ there exists $0 < \delta < 1$ such that the following holds: Let $f : \{-1,1\}^n \to \{-1,1\}$ be monotone. Partition the variables into $\lceil \frac{n}{t} \rceil$ many blocks B_i, each of size at most $t = \log(n^{\frac{1}{2}-\epsilon}/(\delta \log n))$. Then at least one of the below statements about correlation with f must hold:*

$$\begin{aligned}
\textit{a constant } \pm 1 \textit{ has correlation} &\geq 1-\epsilon; && (38)\\
\mathrm{Maj}_{B_i} \textit{ has correlation} &\geq \delta \cdot \frac{\log n}{\sqrt{n}}; && (39)\\
\textit{Majority has correlation} &\geq \delta \cdot \frac{\log n}{\sqrt{n}}. && (40)
\end{aligned}$$

The resulting net from this corollary is $\{-1, 1, \mathrm{Maj}_{[n]}, \mathrm{Maj}_{B_1}, \mathrm{Maj}_{B_2}, \ldots, \mathrm{Maj}_{B_{\lceil n/t \rceil}}\}$; the size of this net is $\lceil \frac{n}{t} \rceil + 3 = O(n/\log n)$.

C A counterexample function

In this section we continue to describe boolean functions as maps $\{-1,1\}^n \to \{-1,1\}$. Additionally, we change the notation for Hamming weight, defining:

$$\text{for } x \in \{-1,1\}^n, \qquad |x| = \sum_{i=1}^{n} x_i \in [-n, n].$$

Recall that Amano and Maruoka [AM06] made the following conjecture, which if true would immediately imply our theorem that $\{-1, 1, x_1, \ldots, x_n, \mathrm{Maj}\}$ is $(1/2 - \Omega(\frac{\log n}{\sqrt{n}}))$-net for monotone functions, using the KKL Theorem 2.

Conjecture C.1. *Suppose $f : \{-1,1\} \to \{-1,1\}$ is a monotone function with $\mathbf{E}_{\boldsymbol{x}}[f(\boldsymbol{x})] = 0$. Then*

$$\mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x})\mathrm{Maj}(\boldsymbol{x})] \geq \Omega\left(\mathcal{E}[f] \cdot \sqrt{n}\right).$$

In this section, we give a counterexample showing that this conjecture is, unfortunately, false.

Let $\frac{1}{100}\sqrt{n \log n} < k < n^{3/5}$, where k and n are both odd integers, and write $t = k/\sqrt{n}$, so $\sqrt{\log n}/100 < t < n^{1/10}$ is a real parameter. Our counterexample is based around the non-boolean monotone function $f_t : \{-1,1\}^n \to \{-1,0,1\}$ given by

$$f_t(x) = \begin{cases} 1 & \text{if } |x| > k, \\ 0 & \text{if } -k \leq |x| \leq k, \\ -1 & \text{if } |x| < -k. \end{cases}$$

In what follows, $a(n,t) \sim b(n,t)$ means $a(n,t)/b(n,t) \to 1$ as $n \to \infty$ (and hence $t \to \infty$). Also, we write ϕ for the pdf of a standard Gaussian, $\phi(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)$.

Proposition C.2. *We have*

$$\begin{aligned} \mathcal{E}[f_t] &\sim 2\phi(t)/\sqrt{n}, \\ \mathop{\mathbf{E}}_{\boldsymbol{x}}[f_t(\boldsymbol{x})\mathrm{Maj}(\boldsymbol{x})] &\sim 2\phi(t)/t. \end{aligned}$$

Proof. Clearly $\mathbf{E}[f_t(\boldsymbol{x})\mathrm{Maj}(\boldsymbol{x})] = 2\mathbf{Pr}[|\boldsymbol{x}| > k]$. Further,

$$\begin{aligned} \mathcal{E}[f_t] &= \frac{1}{2n}\left(\mathop{\mathbf{Pr}}_{\boldsymbol{x}}[|\boldsymbol{x}| = k]\left(\frac{n}{2} - \frac{k}{2}\right) + \mathop{\mathbf{Pr}}_{\boldsymbol{x}}[|\boldsymbol{x}| = k+2]\left(\frac{n}{2} + \frac{k+2}{2}\right)\right) \\ &+ \frac{1}{2n}\left(\mathop{\mathbf{Pr}}_{\boldsymbol{x}}[|\boldsymbol{x}| = -k]\left(\frac{n}{2} - \frac{k}{2}\right) + \mathop{\mathbf{Pr}}_{\boldsymbol{x}}[|\boldsymbol{x}| = -k-2]\left(\frac{n}{2} + \frac{k+2}{2}\right)\right) \\ &\sim \frac{1}{2}\left(\mathop{\mathbf{Pr}}_{\boldsymbol{x}}[|\boldsymbol{x}| = k] + \mathop{\mathbf{Pr}}_{\boldsymbol{x}}[|\boldsymbol{x}| = k+2]\right). \end{aligned}$$

By well known error estimates for the Central Limit Theorem [Fel68, Chap. VII], which apply in our situation because $t < n^{1/10} = o(n^{1/6})$, we have

$$\begin{aligned} \mathbf{Pr}[|\boldsymbol{x}| = k] &\sim 2\phi(t)/\sqrt{n}, \\ \mathbf{Pr}[|\boldsymbol{x}| > k] &\sim \phi(t)/t, \end{aligned}$$

and we also have $\mathbf{Pr}[|\boldsymbol{x}| = k+2] \sim 2\phi(t + 2/\sqrt{n})/\sqrt{n} \sim 2\phi(t)/\sqrt{n}$. The result follows. □

With this in hand, we now define our counterexample function f (we remind the reader that we have not picked a value for t yet) via

$$f(x) = \begin{cases} 1 & \text{if } |x| > k, \\ x_1 & \text{if } -k \leq |x| \leq k, \\ -1 & \text{if } |x| < -k. \end{cases}$$

The function $f : \{-1,1\}^n \to \{-1,1\}$ is indeed a monotone function. It is also easily seen to be balanced, $\mathbf{E}[f] = 0$. Next, we have $\mathbf{E}[f \cdot \mathrm{Maj}] = \mathbf{E}[f_t \cdot \mathrm{Maj}] + \mathbf{E}[g_t \cdot \mathrm{Maj}]$, where

$$g_t(x) = \begin{cases} 0 & \text{if } |x| > k, \\ x_1 & \text{if } -k \leq |x| \leq k, \\ 0 & \text{if } |x| < -k. \end{cases}$$

Clearly $0 \le \mathbf{E}[g_t \cdot \mathrm{Maj}] \le \mathbf{E}[\boldsymbol{x}_1 \cdot \mathrm{Maj}] < 1/\sqrt{n}$. Thus using Proposition C.2 we get

$$\frac{2\phi(t)}{t}(1 - o(1)) \le \mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x})\mathrm{Maj}(\boldsymbol{x})] \le \frac{2\phi(t)}{t}(1 + o(1)) + \frac{1}{\sqrt{n}}. \tag{41}$$

Furthermore, it is not hard to calculate that $\mathcal{E}[f] = (2 \pm o(1))\mathcal{E}[f_t] + O(1/n)$. As a sketch of this calculation, we state that roughly half of the strings counted 1 in $\mathcal{E}[f_t]$ now count 0, and roughly half now count 4, hence the $(2 \pm o(1))$. Further, the $O(1/n)$ term is the contribution to $\mathcal{E}[f]$ from the x_1 portion of f.

Hence from Proposition C.2 that

$$\frac{4\phi(t)}{\sqrt{n}}(1 - o(1)) - O(1/n) \le \mathcal{E}[f] \le \frac{4\phi(t)}{\sqrt{n}}(1 + o(1)) + O(1/n). \tag{42}$$

We now get a counterexample to Amano and Muruoka's conjecture by selecting k in such a way that $\phi(t) \sim \frac{\sqrt{\log n}}{\sqrt{n}}$. If we could freely choose t to be any real parameter then we could achieve $\phi(t) = \frac{\sqrt{\log n}}{\sqrt{n}}$ exactly, taking $t = \sqrt{\log n - \log\log n - \log\log\sqrt{2\pi}}$. Because $t = k/\sqrt{n}$ and k is an odd integer, we can not necessarily use this value of t. We can pick the closest value to t that is of the required form, which differs by at most $2/\sqrt{n}$. This still allows for $\phi(t) \sim \frac{\sqrt{\log n}}{\sqrt{n}}$; since we also get $t \sim \sqrt{\log n}$, plugging into (41), (42) gives

$$\mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x})\mathrm{Maj}(\boldsymbol{x})] \sim \frac{2}{\sqrt{n}}, \qquad \mathcal{E}[f] \sim \frac{4\sqrt{\log n}}{n},$$

contradicting the conjecture. Indeed, this parameter setting shows that the Benjamini-Kalai-Schramm relationship (2) between $\mathcal{E}[f]$ and $\mathbf{E}[f \cdot \mathrm{Maj}]$ cannot be improved up to constants.

C.1 On the Russo-Margulis Lemma

Regarding our robust Kruskal-Katona theorem, we remark that in some sense Theorem 1.3 is the "expected" result, given the KKL Theorem 2 and the Russo-Margulis Lemma [Mar74, Rus82]. The Russo-Margulis Lemma implies that for monotone $f : \{-1,1\}^n \to \{-1,1\}$,

$$\left[\frac{d}{dp} \mathop{\mathbf{E}}_{\boldsymbol{x} \sim_p \{-1,1\}^n}[f(\boldsymbol{x})]\right]_{p=1/2} = 2n \cdot \mathcal{E}[f],$$

where $\boldsymbol{x} \sim_p \{-1,1\}^n$ denotes drawing $\boldsymbol{x}$ from the p-biased product distribution. Thus if f is roughly balanced and has all of its influences at most $1/n^\epsilon$, then KKL implies that its p-biased density has derivative at least $\Omega(\log n)$ at $p = 1/2$. (The paper of Friedgut and Kalai [FK96] makes this argument for general p.) Thus we might guess that in moving monotonically from, say, $\binom{[n]}{n/2}$ to $\binom{[n]}{n/2+1}$ a monotone function's density should increase by $\Omega(\frac{\log n}{n})$. But of course, p-biased density makes no sense in the context of the Kruskal-Katona Theorem where one doesn't have a function on all of $\{-1,1\}^n$, just a subset A of some slice $\binom{[n]}{k}$.

Further, the counterexample function f from this section shows that the Russo-Margulis Lemma is not helpful for establishing our monotone net result. To see this, note that Russo-Margulis tells us that

$$\left[\frac{d}{dp} \mathop{\mathbf{E}}_{\boldsymbol{x} \sim_p \{-1,1\}^n}[f(\boldsymbol{x})]\right]_{p=1/2} = \Theta(\sqrt{\log n}).$$

This might "suggest" that f's density on slices $\binom{[n]}{k}$ for k very near $n/2$ should be increasing in jumps of $\Omega(\frac{\sqrt{\log n}}{n})$ per slice. However since $f(x) = x_1$ in this regime, the actual density jumps are only $\Omega(\frac{1}{n})$. Thus we see that Russo-Margulis is insufficiently sensitive to understand density changes on adjacent slices; this is ultimately what necessitates our generalized KKL and Kruskal-Katona Theorems, which are "local" to slices.