Log-Sobolev inequality for the multislice, with applications[*]

Yuval Filmus[†]   Ryan O'Donnell[‡]   Xinyu Wu[§]

Abstract

Let $\kappa \in \mathbb{N}_+^\ell$ satisfy $\kappa_1 + \cdots + \kappa_\ell = n$, and let $\mathcal{U}_\kappa$ denote the *multislice* of all strings $u \in [\ell]^n$ having exactly κ_i coordinates equal to i, for all $i \in [\ell]$. Consider the Markov chain on $\mathcal{U}_\kappa$ where a step is a random transposition of two coordinates of u. We show that the log-Sobolev constant ϱ_κ for the chain satisfies

$$\varrho_\kappa^{-1} \le n \cdot \sum_{i=1}^\ell \tfrac{1}{2} \log_2(4n/\kappa_i),$$

which is sharp up to constants whenever ℓ is constant. From this, we derive some consequences for small-set expansion and isoperimetry in the multislice, including a KKL Theorem, a Kruskal–Katona Theorem for the multislice, a Friedgut Junta Theorem, and a Nisan–Szegedy Theorem.

Keywords: log-Sobolev inequality; small-set expansion; conductance; hypercontractivity; Fourier analysis; representation theory; Markov chains; combinatorics.
MSC2020 subject classifications: 60J10; 05E18; 68R05.
Submitted to EJP on September 10, 2018, final version accepted on January 30, 2022.
Supersedes arXiv:1809.03546.

1 Introduction

Suppose we have a deck of n cards, with κ_1 of them colored red, κ_2 of them colored blue, and κ_3 of them colored green. If we "shuffle" the cards by repeatedly transposing random pairs of cards, how long does it take for the deck to get to a well-mixed configuration? This question is asking about the mixing time and expansion in a Markov chain known variously as the *multi-urn Bernoulli–Laplace diffusion process* or the *multislice*.

[†]Technion Computer Science Department. E-mail: yuvalfi@cs.technion.ac.il
[‡]Computer Science Department, Carnegie Mellon University. E-mail: odonnell@cs.cmu.edu
[§]Computer Science Department, Carnegie Mellon University. E-mail: xinyuwu@cmu.edu

Let $\ell \in \mathbb{N}_+$ denote a number of *colors* and let $n \in \mathbb{N}_+$ denote a number of *coordinates* (or *positions*). Following computer science terminology, we refer to elements $u \in [\ell]^n$ as *strings*. Given a color $i \in [\ell]$, we write $\#_i u$ for the number of coordinates $j \in [n]$ for which $u_j = i$. The vector $\kappa = (\#_1 u, \dots, \#_\ell u) \in \mathbb{N}^\ell$ is referred to as the *histogram* of u. In general, if $\kappa \in \mathbb{N}_+^\ell$ satisfies $\kappa_1 + \cdots + \kappa_\ell = n$ (so κ is a *composition* of n), we define the associated *multislice* to be

$$\mathcal{U}_\kappa = \{u \in [\ell]^n : \#_i u = \kappa_i \text{ for all } i \in [\ell]\}.$$

The terminology here is inspired by the well-studied case when $\ell = 2$, in which case $\mathcal{U}_\kappa$ is a Hamming *slice* of the Boolean cube. We also remark that when $\ell = n$ and $\kappa = (1, 1, \dots, 1)$, the set $\mathcal{U}_\kappa$ is the set of all permutations of $[n]$.

The random transposition Markov chain. The symmetric group S_n acts on strings $u \in [\ell]^n$ in the natural way, by permuting coordinates: $(u^\sigma)_j = u_{\sigma(j)}$ for $\sigma \in S_n$. This action preserves each multislice $\mathcal{U}_\kappa$. This paper is concerned with the Markov chain on $\mathcal{U}_\kappa$ generated by *random transpositions*. Let $\mathrm{Trans}(n) \subseteq S_n$ denote the set of transpositions on n coordinates. We will specifically be interested in the reversible, discrete-time Markov chain on state space $\mathcal{U}_\kappa$ in which a step from $u \in \mathcal{U}_\kappa$ consists of moving to $u^{\boldsymbol{\tau}}$, where $\boldsymbol{\tau} \sim \mathrm{Trans}(n)$ is chosen uniformly at random. (We always use **boldface** to denote random variables.) One also has the associated *Schreier graph*, with vertex set $\mathcal{U}_\kappa$ and edges $\{u, u^\tau\}$ for all $u \in \mathcal{U}_\kappa$ and $\tau \in \mathrm{Trans}(n)$. Since this graph is regular, it follows that the invariant distribution for the Markov chain is the uniform distribution on $\mathcal{U}_\kappa$. We will denote this distribution by π_κ, or just π if κ is clear from context.

Log-Sobolev inequalities. One of the most powerful ways to study mixing time and "small-set expansion" in Markov chains is through *log-Sobolev* inequalities (see, e.g., [32, 14]). For a subset $A \subseteq \mathcal{U}_\kappa$, define its *conductance* (or *expansion*) to be

$$\Phi[A] = \Pr_{\substack{\boldsymbol{u} \sim A \\ \boldsymbol{\tau} \sim \mathrm{Trans}(n)}} [\boldsymbol{u}^{\boldsymbol{\tau}} \notin A].$$

Sets A with small conductance are natural bottlenecks for mixing in the Markov chain. An example when $\ell = 2$ and $\kappa = (n/2, n/2)$ is the "dictator" set $A = \{u : u_1 = 1\}$. It has expansion $\Phi[A] = \frac{1}{n-1}$, and indeed, if we start the random walk from a string u with $u_1 = 1$, it will take about $n/2$ steps on average before there's even a chance that u_1 will change from 1.

One feature of this example is that the set A is "large"; its *(fractional) volume*,

$$\mathrm{vol}(A) = |A|/|\mathcal{U}_\kappa| = \Pr_{\boldsymbol{u} \sim \pi} [\boldsymbol{u} \in A],$$

is bounded below by a constant. The "small-set expansion" phenomenon [35, 44, 54] (occurring most famously in the standard random walk on the Boolean cube $\{0,1\}^n$) refers to the possibility that all "small" sets have high conductance. Intuitively, if small-set expansion holds for a Markov chain, then a random walk with a deterministic starting point should mix rapidly in its early stages, with the possibility for slowdown occurring only when the chain is somewhat close to mixed.

A *log-Sobolev inequality* for the Markov chain is one way that such a phenomenon may be captured. In particular, if the *log-Sobolev constant* for the transposition chain on $\mathcal{U}_\kappa$ is ϱ_κ, it follows that

$$\Phi[A] \geq \tfrac{1}{2}\varrho_\kappa \cdot \ln(1/\mathrm{vol}(A)) \quad \text{for all nonempty subsets } A \subseteq \mathcal{U}_\kappa. \tag{1.1}$$

Log-Sobolev inequality for the multislice, with applications

So sets of constant volume must have conductance $\Omega(\varrho_\kappa)$, but sets of volume $2^{-\Theta(n)}$ (for example) must have conductance $\Omega(n\varrho_\kappa)$. A known further consequence of a log-Sobolev inequality is a *hypercontractive inequality*, which concerns expansion in the continuous-time version of the Markov chain. It implies that if $\boldsymbol{\sigma}$ is the random permutation generated by performing the continuous-time chain for $t = \frac{\ln c}{2\varrho_\kappa}$ time — i.e.,

$$\boldsymbol{\sigma} \text{ is the product of Poisson}\left(\frac{\ln c}{2\varrho_\kappa}\right) \text{ random transpositions, } c \geq 1$$

— then

$$\mathop{\mathbf{Pr}}_{\substack{\boldsymbol{u} \sim A \\ \boldsymbol{\sigma} \sim \text{Trans}(n)}} [\boldsymbol{u}^{\boldsymbol{\sigma}} \notin A] \geq 1 - \text{vol}(A)^{(c-1)/(c+1)} \quad \text{for all nonempty subsets } A \subseteq \mathcal{U}_\kappa.$$

Thus again, if $\text{vol}(A)$ is small, then the Markov chain will almost surely exit A after running for $\Theta(\varrho_\kappa^{-1})$ steps.

We remark that Inequality (1.1) is merely a *consequence* of the log-Sobolev constant being ϱ_κ. It is not the case that ϱ_κ is defined to be the largest constant for which Inequality (1.1) holds (for all A) — though this is a reasonable intuition. Instead, ϱ_κ is defined to be the largest constant for which a certain generalization of Inequality (1.1) to nonnegative functions holds; namely,

$$\mathop{\mathbf{E}}_{\substack{\boldsymbol{u} \sim \pi \\ \boldsymbol{\tau} \sim \text{Trans}(n)}} \left(\sqrt{\phi(\boldsymbol{u})} - \sqrt{\phi(\boldsymbol{u}^{\boldsymbol{\tau}})}\right)^2 \geq \varrho_\kappa \cdot \text{KL}(\phi\pi \parallel \pi) \quad \text{for all probability densities } \phi. \quad (1.2)$$

(Here a *probability density function* is a function $\phi\colon \mathcal{U}_\kappa \to \mathbb{R}^{\geq 0}$ satisfying $\mathbf{E}_\pi[\phi] = 1$, and $\text{KL}(\phi\pi \parallel \pi)$ denotes the *KL divergence* between distributions $\phi\pi$ and π.) Inequality (1.2) includes Inequality (1.1) by taking $\phi = 1_A/\text{vol}(A)$. For more details, see Section 2.

Our main theorem in this work is a lower bound on the log-Sobolev constant for $\mathcal{U}_\kappa$:

Theorem 1.1. *Let $\kappa \in \mathbb{N}_+^\ell$ satisfy $\kappa_1 + \cdots + \kappa_\ell = n$, and let ϱ_κ denote the log-Sobolev constant for the transposition chain on the multislice $\mathcal{U}_\kappa$ (i.e., the largest constant for which Inequality (1.2) holds). Then*

$$\varrho_\kappa^{-1} \leq n \cdot \sum_{i=1}^{\ell} \tfrac{1}{2} \log_2(4n/\kappa_i).$$

The main case of interest for us is $n \longrightarrow \infty$ with $\ell = O(1)$ and $\kappa_i/n \geq \Omega(1)$ for each i; in other words, when we are at a "middling" histogram of a high-dimensional multicube $[\ell]^n$. In this case our bound is $\varrho_\kappa \geq \Omega(1/n)$, which is the same bound that holds for the standard random walk on the Boolean cube. Thus for this parameter setting, the random transposition chain on the multislice enjoys all of the same small-set expansion properties as the Boolean cube (up to constants).

On the sharpness of Theorem 1.1. When ℓ is considered to be a constant, Theorem 1.1 is sharp up to constant factors (which we did not attempt to optimize); i.e.,

$$\varrho_\kappa^{-1} = \Theta(n) \cdot \log\left(\frac{n}{\min_i\{\kappa_i\}}\right) \quad \text{for } \ell = O(1). \quad (1.3)$$

To see the upper bound on ϱ_κ, assume without loss of generality that $\ell = \text{argmin}_i\{\kappa_i\}$, and take

$$A = \{u \in \mathcal{U}_\kappa : u_j = \ell \text{ for all } j \in [\kappa_\ell]\}.$$

It is easy to compute that $\Phi[A] = \Theta(\kappa_\ell/n)$ and $\mathrm{vol}(A) = \binom{n}{\kappa_\ell}^{-1}$ (hence $\ln(1/\mathrm{vol}(A)) = \Theta(\kappa_\ell \log(n/\kappa_\ell))$). Putting this into Inequality (1.1) shows the claimed upper bound on ϱ_κ.

At the opposite extreme, when $\ell = n$ and $\kappa = (1, 1, \ldots, 1)$, we have the random transposition walk on the symmetric group S_n. In this case, Theorem 1.1 as stated gives the poor bound of $\varrho_\kappa \geq \Omega(1/n^2 \log n)$, whereas the optimal bound is $\varrho_\kappa = \Theta(1/n \log n)$ [14, 42]. In fact, our proof of Theorem 1.1 (which generalizes that of [42]) can actually achieve the tight lower bound of $\varrho_\kappa \geq \Omega(1/n \log n)$ in this case. However, we tailored our general bound for the case of $\ell = O(1)$, and did not try to optimize for the most general scenario of ℓ varying with n. A reasonable prediction might be that Equation (1.3) always holds, up to universal constants, without the assumption of $\ell = O(1)$; we leave investigation of this for future work.[1]

1.1 Applications

There are many known applications of log-Sobolev and hypercontractive inequalities in combinatorics and theoretical computer science (see, e.g., [49, Ch. 9, 10]). In this paper we present four particular consequences of Theorem 1.1 for analysis/combinatorics of Boolean functions on the multislice. We anticipate the possibility of several more. Full details of these applications appear in Section 4; here we describe them informally.

Throughout the remainder of this section, let us think of n as large, of ℓ as constant, and let us fix a histogram κ (with $\kappa_1 + \cdots + \kappa_\ell = n$) satisfying $\kappa_i/n \geq \Omega(1)$ for all i. For example, we might think of $\ell = 3$ and $\kappa = (n/3, n/3, n/3)$, so that $\mathcal{U}_\kappa$ consists of all ternary strings with an equal number of 1's, 2's, and 3's. The *isoperimetric problem* for $\mathcal{U}_\kappa$ would ask: for a given fixed $0 < \alpha < 1$, which subset $A \subseteq \mathcal{U}_\kappa$ with $\mathrm{vol}(A) = \alpha$ has minimal "edge boundary", i.e., minimal $\Phi[A]$? (Here "edge boundary" is with respect to performing a single transposition, although in our Kruskal–Katona application we will relate this to the size of A's "shadows" at neighboring multislices.)

We typically think of α as "constant", bounded away from 0 and 1. In our example with $\kappa = (n/3, n/3, n/3)$, when $\alpha = 1/3$ the isoperimetric minimizer is a "dictator" set like $A = \{u : u_1 = 1\}$; it has $\Phi[A] = \frac{4/3}{n-1}$. The "99% regime" version of the isoperimetric question would be: if $\Phi[A]$ is within a factor $1 + o(1)$ of minimal, must A be "$o(1)$-close" to a minimizer? This question will be considered in a companion paper. We will instead consider the "1% regime" version of the isoperimetric question: if $\Phi[A]$ is at most $O(1)$ times the minimum, must A at least "slightly resemble" a minimizer?

To orient ourselves, first note that for constant α (bounded away from 0 and 1), the minimum possible value of $\Phi[A]$ among A with $\mathrm{vol}(A) = \alpha$ is $\Theta(1/n)$; indeed, this follows from our Theorem 1.1 and Inequality (1.1). From this fact, we will derive a multislice variant of the **Kruskal–Katona Theorem**. Up to $O(1)$ factors, this minimum is achieved not just by "dictator" sets like $\{u \in \mathcal{U}_{(n/3, n/3, n/3)} : u_1 = 1\}$, but also by any "junta" set, meaning a set A for which absence or presence of $u \in A$ depends only on the colors $(u_j : j \in J)$ for a set $J \subseteq [n]$ of cardinality $c = O(1)$. It is not hard to see that if $A \subseteq \mathcal{U}_\kappa$ is such a *c-junta*, then $\Phi[A] \leq O(c/n)$. We may now ask: if $\Phi[A] \leq O(1/n)$, must A at least slightly "resemble" a junta?

We give two closely related positive answers to this question, as a consequence of our log-Sobolev inequality. The first answer, a **KKL Theorem** for the multislice (cf. [35]), follows immediately from previous work [50, 51]. It says that for any set with $\Phi[A] \leq O(1/n)$, there must exist some pair of coordinates $j, j' \in [n]$ with at least constant *influence* on A, where the influence of the transposition $(j \, j')$ on A is defined to be

$$\mathrm{Inf}_{(j \, j')}[A] = \Pr_{\boldsymbol{u} \sim \pi}\left[1_A(\boldsymbol{u}) \neq 1_A(\boldsymbol{u}^{(j \, j')})\right]. \tag{1.4}$$

---

[1]Salez [55] showed in a subsequent paper that this is indeed true.

It is the hallmark of a junta A that every transposition $(j\ j')$ has either $\mathrm{Inf}_{(j\ j')}[A] = 0$ or $\mathrm{Inf}_{(j\ j')}[A] \geq \Omega(1)$. In fact, mirroring the original KKL Theorem, our work shows that: (i) if $\Phi[A] \leq c/n$ then there exists $(j\ j')$ with $\mathrm{Inf}_{(j\ j')}[A] \geq \exp(-O(c))$; (ii) for *any* $A \subseteq \mathcal{U}_\kappa$ with $\Omega(1) \leq \mathrm{vol}(A) \leq 1 - \Omega(1)$, there exists $(j\ j')$ with $\mathrm{Inf}_{(j\ j')}[A] \geq \Omega\left(\frac{\log n}{n}\right)$. From this, we can also derive a "robust" version of our Kruskal–Katona theorem (a là [50]).

A closely related consequence of our work is a **Friedgut Junta Theorem** for the multislice (cf. [29]), which follows (using a small amount of representation theory) from work of Wimmer [61] (see also [24] for a different account). It states that for any A with $\Phi[A] \leq c/n$, and any $\epsilon > 0$, there is a genuine $\exp(O(c/\epsilon))$-junta $A' \subseteq \mathcal{U}_\kappa$ that is ϵ-close to A, meaning $\mathrm{vol}(A \triangle A') \leq \epsilon$. The junta theorem can also be generalized to real-valued functions, following the work of Bouyrie [4], with a worse dependence on ϵ in the exponent.

Finally, with a little more representation theory effort, we are able to derive from Theorem 1.1 a **Nisan–Szegedy Theorem** for the multislice (cf. [48]), which is (roughly) an $\epsilon = 0$ version of the Friedgut Junta Theorem; this generalizes previous work on the Hamming slice [25]. It says that if $A \subseteq \mathcal{U}_\kappa$ is of "degree k" — meaning that its indicator function can be written as a linear combination of k-junta functions — then A must be an $\exp(O(k))$-junta itself. (The $k = 1$ case of this theorem, with the conclusion that A is a 1-junta, was proven recently in [26].)

1.2 Context and prior work

In this section we review similar contexts where log-Sobolev inequalities and small-set expansion have been studied.

The Boolean cube. The simplest and best-known setting for these kinds of results is the Boolean cube $\{0,1\}^n$ with the nearest-neighbour random walk. The optimal hypercontractive inequality in this setting was proven by Bonami [3]. Later, Gross [32] introduced log-Sobolev inequalities, showed that they were equivalent to hypercontractive inequalities in this setting, and determined the exact log-Sobolev constant for the Boolean cube, namely $\varrho = 2/n$. Gross also observed that all the same results also hold for Gaussian space in any dimension (recovering prior work of Nelson [47]); Gaussian space is in fact a "special case" of the Boolean cube, by virtue of the Central Limit Theorem. The Boolean cube also generalizes the well-studied *Ehrenfest model* of diffusion [19].

These inequalities for the Boolean cube, as well as the associated small-set expansion corollaries, have had innumerable applications in analysis, combinatorics, and theoretical computer science, in topics ranging from communication complexity to inapproximability; see, e.g., [41] or [49, Chapters 9–11].

A different line of work sought to determine the exact minimum value of $\Phi[A]$ in terms of the size of A. This challenge, known as the *edge isoperimetric problem*, has been solved by Harper [33], Lindsey [43], Bernstein [2], and Hart [34], who have shown that the optimal sets are initial segments of a lexicographic ordering of the vertices of the Boolean cube. Recently Ellis, Keller and Lifshitz gave a new proof of the edge isoperimetric inequality using the Kruskal–Katona Theorem [21]. The same set of authors also recently proved a stability version of the edge isoperimetric inequality in the 99% regime [22].

Returning to log-Sobolev inequalities, an extraordinarily helpful feature of the random walk on the Boolean cube is that it is a *product Markov chain*, with a stationary distribution that is *independent* across the n coordinates. Because of this, a simple induction lets one immediately reduce the log-Sobolev (and hypercontractivity) analysis to the base case of $n = 1$.

Other product chains. For *any* product Markov chain, one can similarly reduce the analysis to the $n = 1$ case. In general, let ν be a probability distribution of full support on $[\ell]$, and consider the Markov chain on $[\ell]^n$ in which a step from $u \in [\ell]^n$ consists of choosing a random coordinate $\boldsymbol{j} \sim [n]$ and replacing $u_{\boldsymbol{j}}$ with a random draw from ν. The invariant distribution for this chain is the product distribution $\nu^{\otimes n}$. Though the $n = 1$ case of this chain is, in a sense, trivial — it mixes perfectly in one step — it is not especially easy to work out the optimal log-Sobolev constant. Nevertheless, Diaconis and Saloffe-Coste [14] showed that for the $n = 1$ chain, the log-Sobolev constant is

$$\varrho_\nu^{\mathrm{triv}} = 2\frac{q - p}{\ln q - \ln p}, \quad \text{where } p = \min_{i \in [\ell]}\{\nu(i)\},\ q = 1 - p. \tag{1.5}$$

It follows immediately that the log-Sobolev constant in the general-n case is $\varrho_\nu^{\mathrm{triv}}/n$. In particular, if $\kappa_1 + \cdots + \kappa_\ell = n$ and $\nu(i) = \kappa_i/n$, then $\nu^{\otimes n}$ resembles the uniform distribution π_κ on $\mathcal{U}_\kappa$, and the product chain on $[\ell]^n$ somewhat resembles the random transposition chain on $\mathcal{U}_\kappa$. This gives credence to the possibility that Equation (1.3) may hold with absolute constants for any ℓ.

The Boolean slice / Bernoulli–Laplace model / Johnson graph. Significant difficulties arise when one moves away from product Markov chains. One of the simplest steps forward is to the Boolean slice. This is the $\ell = 2$ case of the Markov chains studied in this paper, with the "balanced" case of $\kappa = (n/2, n/2)$ being the most traditionally studied. This Markov chain is also equivalent to the Bernoulli–Laplace model for diffusion between two incompressible liquids, and to the standard random walk on *Johnson graphs*; taking multiple steps in the chain is similar to the random walk in *generalized Johnson graphs*. The chain has been studied in wide-ranging contexts, from genetics [45], to child psychology [52], to computational learning theory [50]. An asymptotically exact analysis of the time to stationarity of this Markov chain was given by Diaconis and Shahshahani [16], using representation theory. However, the log-Sobolev constant for the chain took a rather long time to be determined; it was left open in Diaconis and Saloff-Coste's 1996 survey [14] before finally being determined (up to constants) by Lee and Yau in 1998 [42]. This sharp log-Sobolev inequality, and its attendant hypercontractivity and small-set expansion inequalities, have subsequently been used in numerous applications — for the Kruskal–Katona and Erdős–Ko–Rado theorems in combinatorics [50, 11, 27], for computational learning theory [60, 50], for property testing [46], and for generalizing classic "analysis of Boolean functions" results [50, 51, 24, 23, 28, 27, 5].

The Grassmann graph. One direction of generalization for the Johnson graphs are their "q-analogues", the *Grassmann graphs*; understanding this Markov chain was posed as an open problem even in the early work of Diaconis and Shahshahani [16, Example 2]. For a finite field $\mathbb{F}$ and integer parameters $n \geq k \geq 1$, the associated Grassmann graph has as its vertices all k-dimensional subspaces of $\mathbb{F}^n$, with two subspaces connected by an edge if their intersection has dimension $k - 1$. Understanding small-set expansion (and lack thereof) in the Grassmann graphs was central to the very recent line of work that positively resolved the 2-to-2 Conjecture [38, 18, 17, 1, 39] (with the analogous problems on the Johnson graphs serving as an important warmup [37]). Still, it seems fair to say that the mixing properties of the Grassmann graph are far from being fully understood.

The multislice. We now come to the multislice, the other natural direction of generalization for the Johnson graphs, and the subject of the present paper. One can see

the multislice as a generalization of the Bernoulli–Laplace model, modeling diffusion between three or more liquids. As well, the space of functions $f : \mathcal{U}_\kappa \to \mathbb{R}$, together with the action of S_n on $\mathcal{U}_\kappa$, is precisely the *Young permutation module* M^κ arising in the representation theory of the symmetric group. Understanding the mixing properties of the $\mathcal{U}_\kappa$ Markov chain with random transpositions was suggested as an open problem several times [16], [12, p. 59], [25]. The multislice has also played a key combinatorial role in problems in combinatorics, such as the Density Hales–Jewett problem (where $\ell = 3$ was the main case under consideration) [53].

Although it might at first appear to be a simple generalization of the Boolean slice, there are several fundamental impediments that arise when moving from $\ell = 2$ even to $\ell = 3$. These include: the fact that a Hamming slice disconnects the nearest-neighbour graph in $[2]^\ell$ but not in $[3]^\ell$; the fact that one can introduce just *one* variable per coordinate when representing functions $[2]^\ell \to \mathbb{R}$ as multilinear polynomials; the fact that 2-row irreps of S_n (Young diagrams) are completely defined by the number of boxes not in the first row; and, the fact that when $\ell \geq 3$, the decomposition of the permutation module M^κ into irreps has multiplicities. The last of these was the main difficulty to be overcome in Scarabotti's work [56] giving the asymptotic mixing time for the transposition walk on balanced multislices $\mathcal{U}_{(n/\ell,\dots,n/\ell)}$ (see also [13, 57]). It also prevents the multislice from forming an association scheme.

For the purposes of this paper, the main difficulty that arises when analyzing the log-Sobolev inequality is the following: when $\ell = 2$, any nontrivial step in the Markov chain (switching a 1 and a 2) has the property that the histogram within $[\ell]^{n-2}$ of the unswitched colors is always the same: $(\kappa_1 - 1, \kappa_2 - 1)$. By contrast, once $\ell \geq 3$, the multiple "kinds" of transpositions (switching a 1 and a 2, or a 1 and a 3, or a 2 and a 3, etc.) lead to differing histograms within $[\ell]^{n-2}$ for the unswitched colors. This significantly complicates inductive arguments.

The symmetric group and beyond. Finally, we mention that analysis of the multislice can also be motivated simply as a necessary first step in a full understanding of spectral analysis on the symmetric group and other algebraic structures, an opinion also espoused in, e.g., [6]. Such structures include classical association schemes such as polar spaces and bilinear forms, matrix groups such as the general linear group, and the q-analog of the multislice.

2 Preliminaries

2.1 Definitions relevant for the log-Sobolev inequality

Given parameters $\ell \in \mathbb{N}_+$ (number of colors) and $n \in \mathbb{N}_+$ (number of coordinates), our objects of study in this paper are *multislices*, parametrized by a histogram $\kappa \in \mathbb{N}_+^\ell$ satisfying $\kappa_1 + \cdots + \kappa_\ell = n$:

$$\mathcal{U}_\kappa = \{u \in [\ell]^n : \#_i u = \kappa_i \text{ for all } i \in [\ell]\}.$$

We will only consider multislices with at least two colors; in other words, $\ell \geq 2$.

We introduce the inner product space of functions on $\mathcal{U}_\kappa$,

$$M^\kappa = \{f : \mathcal{U}_\kappa \to \mathbb{R}\}, \quad \text{with } \langle f, g \rangle = \mathop{\mathbf{E}}_{\boldsymbol{u} \sim \pi}[f(\boldsymbol{u})g(\boldsymbol{u})],$$

where $\pi = \pi_\kappa$ denotes the uniform distribution on the multislice $\mathcal{U}_\kappa$.

Let K denote the *transition / Markov operator* on M^κ associated to the transposition random walk, defined by

$$\mathrm{K}f(u) = \mathop{\mathbf{E}}_{\boldsymbol{\tau} \sim \mathrm{Trans}(n)}[f(u^{\boldsymbol{\tau}})],$$

where $\mathrm{Trans}(n)$ consists of all $\binom{n}{2}$ transpositions. Let L denote the *Laplacian operator* $\mathbb{1}-\mathrm{K}$ (where $\mathbb{1}$ is the identity operator). Then the *energy* (or *Dirichlet form*) of $f\colon \mathcal{U}_\kappa \to \mathbb{R}$ is

$$\mathcal{E}[f] = \langle f, \mathrm{L}f \rangle = \tfrac{1}{2} \mathop{\mathbf{E}}_{\boldsymbol{u} \sim \boldsymbol{v}}\Big[(f(\boldsymbol{u}) - f(\boldsymbol{v}))^2 \Big],$$

where we have introduced the notation $\boldsymbol{u} \sim \boldsymbol{v}$ to denote that $(\boldsymbol{u}, \boldsymbol{v})$ is a random edge in the Schreier graph; equivalently, $\boldsymbol{u} \sim \pi$ and $\boldsymbol{v} = \boldsymbol{u}^{\boldsymbol{\tau}}$ for $\boldsymbol{\tau} \sim \mathrm{Trans}(n)$. One may check that if $A \subseteq \mathcal{U}_\kappa$, then

$$\mathcal{E}[1_A] = \mathrm{vol}(A) \cdot \Phi[A],$$

where $1_A \in M^\kappa$ denotes the 0/1-indicator of A. (Recall that $\Phi[A]$ is the probability, over $\boldsymbol{u} \sim A$ and $\boldsymbol{\tau} \sim \mathrm{Trans}(n)$, that $\boldsymbol{u}^{\boldsymbol{\tau}} \notin A$.)

Before formally defining the log-Sobolev inequality for the transposition chain on $\mathcal{U}_\kappa$, we recall first its simpler counterpart, the *Poincaré inequality*. For $f \in M^\kappa$, this is

$$\mathcal{E}[f] \geq \lambda_1 \cdot \mathop{\mathbf{Var}}_{\pi}[f],$$

where λ_1 is the *spectral gap*; i.e., the lowest eigenvalue of L other than the trivial $\lambda_0 = 0$. For the transposition chain on $\mathcal{U}_\kappa$ it is known that $\lambda_1 = \frac{2}{n-1}$, with "dictator" functions and other "degree-1" functions providing the tight examples; see Theorem 4.15 in Section 4.2.

As for the log-Sobolev inequality, it is typically defined as

$$\mathcal{E}[f] \geq \tfrac{1}{2}\varrho \cdot \mathbf{Ent}[f^2];$$

the largest constant ϱ that is acceptable for all $f \in M^\kappa$ being termed the log-Sobolev constant for $\mathcal{U}_\kappa$. Here $\mathbf{Ent}[g] = \mathbf{E}_\pi[g \ln g] - \mathbf{E}_\pi[g] \cdot \ln \mathbf{E}_\pi[g]$. We remark that Diaconis and Saloff-Coste [14] showed that $\varrho \leq \lambda_1$ always holds.

In this work we will prefer a slightly different (equivalent) definition for the log-Sobolev inequality, used in [42]. It's easy to see that replacing f with $|f|$ does not change $\mathbf{Ent}[f^2]$ but can only decrease $\mathcal{E}[f]$. Thus in the log-Sobolev inequality it suffices to consider nonnegative f. Also, the inequality is 2-homogeneous (both sides are multiplied by c^2 when f is multiplied by c); thus it suffices to consider nonnegative f with $\mathbf{E}_\pi[f^2] = 1$. We write a general such f as $\sqrt{\phi}$, where $\phi\colon \mathcal{U}_\kappa \to \mathbb{R}^{\geq 0}$ satisfies $\mathbf{E}_\pi[\phi] = 1$. We call such a ϕ a *probability density function*, thinking of it as a relative probability density with respect to the uniform distribution $\pi = \pi_\kappa$. In other words, we associate ϕ to the probability distribution in which $u \in \mathcal{U}_\kappa$ has probability mass $\phi(u)\pi(u)$. Now when $\phi = f^2$, we have

$$\mathbf{Ent}[f^2] = \mathbf{Ent}[\phi] = \mathop{\mathbf{E}}_{\pi}[\phi \ln \phi] - \mathop{\mathbf{E}}_{\pi}[\phi] \cdot \ln \mathop{\mathbf{E}}_{\pi}[\phi] = \mathop{\mathbf{E}}_{\pi}[\phi \ln \phi],$$

where we used $\ln \mathbf{E}_\pi[\phi] = \ln 1 = 0$. This quantity is precisely the *Kullback–Leibler divergence* (or *relative entropy*) between the distribution $\phi\pi$ and the distribution π, denoted $\mathrm{KL}(\phi\pi \parallel \pi)$. Thus we have shown that the usual formulation of the log-Sobolev inequality is equivalent to

$$\mathcal{E}\left[\sqrt{\phi}\right] \geq \tfrac{1}{2}\varrho_\kappa \cdot \mathrm{KL}(\phi\pi \parallel \phi) \tag{2.1}$$

for all probability densities ϕ on $\mathcal{U}_\kappa$, as stated in Inequality (1.2).

2.2 Hypercontractivity, influences, and other preliminaries for our applications

In this section we make some further definitions, which will be useful for our applications of the log-Sobolev inequality. We first decompose L into its components on each transposition $\tau \in \mathrm{Trans}(n)$, introducing the operator L_τ defined by

$$\mathrm{L}_\tau f(x) = f(u) - f(u^\tau)$$

Note that

$$\mathrm{L} = \underset{\tau \in \mathrm{Trans}(n)}{\mathrm{avg}} \mathrm{L}_\tau.$$

Now for $f : \mathcal{U}_\kappa \to \mathbb{R}$ we define the *influence* of transposition $\tau \in \mathrm{Trans}(n)$ on f to be

$$\mathrm{Inf}_\tau[f] = \langle f, \mathrm{L}_\tau f \rangle = \langle \mathrm{L}_\tau f, f \rangle = \tfrac{1}{2}\|\mathrm{L}_\tau f\|_2^2$$

(where we are using norm notation $\|f\|_p = \mathbf{E}[|f|^p]^{1/p}$). Note that if $A \subseteq \mathcal{U}_\kappa$ then we have the following combinatorial interpretation, agreeing with our notation from Equation (1.4):

$$\mathrm{Inf}_\tau[A] = \mathrm{Inf}_\tau[1_A] = \underset{\boldsymbol{u} \sim \pi_\kappa}{\mathbf{Pr}}\left[\boldsymbol{u} \in A,\ \boldsymbol{u}^\tau \notin A\right] = \underset{\boldsymbol{u} \sim \pi_\kappa}{\mathbf{Pr}}\left[\boldsymbol{u} \notin A,\ \boldsymbol{u}^\tau \in A\right].$$

For general $f : \mathcal{U}_\kappa \to \mathbb{R}$ we introduce the following additional notation, for the *average influence* (equivalent to *energy*), *total influence*, and *maximum influence* of f:

$$\mathcal{E}[f] = \underset{\tau \in \mathrm{Trans}(n)}{\mathrm{avg}} \mathrm{Inf}_\tau[f], \quad \mathrm{Inf}[f] = \sum_{\tau \in \mathrm{Trans}(n)} \mathrm{Inf}_\tau[f] = \binom{n}{2}\mathcal{E}[f] \quad \mathcal{M}[f] = \max_{\tau \in \mathrm{Trans}(n)} \mathrm{Inf}_\tau[f].$$

As recounted in the important survey of Diaconis and Saloff-Coste [14], there is an equivalence between log-Sobolev inequalities and hypercontractivity for reversible Markov chains. To explain what hypercontractivity means in this abstract setting, we first define the continuous-time analog of the random transposition walk. This is the continuous-time Markov chain, running from time $t = 0$ to $t = \infty$, in which (informally) in any interval of infinitesimal length dt, one performs step of the random transposition chain with probability dt.

More formally, we can provide the following alternative description: if we initialize the continuous-time Markov chain at state u, then its state $\boldsymbol{u}_t$ at time $t \geq 0$ is obtained by performing $\mathrm{Poisson}(t)$ random transpositions on u. From this definition we may define *noise operator* (or *heat kernel*) H_t on M^κ:

$$\mathrm{H}_t f(u) = \mathbf{E}[f(\boldsymbol{u}_t)].$$

It's well known that we can also express H_t in terms of the Laplacian operator, $\mathrm{H}_t = e^{-t\mathrm{L}}$.

Diaconis and Saloff-Coste [14, Theorem 3.5(ii)] show that the log-Sobolev inequality implies hypercontractivity:

Theorem 2.1. *For $q \geq 2$ and $t = \frac{\ln(q-1)}{2\varrho_\kappa}$, the following hold for all $f \in M^\kappa$:*

$$\|\mathrm{H}_t f\|_q \leq \|f\|_2, \qquad \|\mathrm{H}_t f\|_2 \leq \|f\|_{q'},$$

where q' is the Hölder conjugate of q (meaning $1/q + 1/q' = 1$).

(The first inequality directly appears in [14]; the second statement appears only implicitly. It is a consequence of $\mathrm{H}_t f$ being a self-adjoint operator; see, e.g., [49, Prop. 9.19].)

In all our applications, we actually use hypercontractivity, rather than log-Sobolev directly.

2.3 Two numerical lemmas

Here we give some two elementary numerical lemmas we'll need for our proofs. We start by computing and bounding the inverse moment of a hypergeometric random variable. The following simple fact may well be in the literature; the most relevant citation we found was [31]:

Lemma 2.2. *Let* $X \sim \mathrm{Hypergeometric}(N, K, n)$.[2] *Then, writing* $p = 1 - \binom{N-K}{n+1}/\binom{N+1}{n+1}$, *it holds that*

$$\mathbf{E}\left[\frac{1}{X+1}\right] = p\frac{N+1}{(n+1)(K+1)} \leq \frac{N+1}{(n+1)(K+1)}.$$

Proof. Take an urn with $N + 1$ balls, $K + 1$ of them white, one of the white balls being "special". Consider an experiment in which we draw $n + 1$ balls without replacement, and then choose a random white ball among the ones drawn (if any). A "success" occurs if the randomly chosen white ball is the special one.

A necessary condition for the experiment to succeed is that the special white ball was chosen at all, which happens with probability $\frac{n+1}{N+1}$. Assuming that this happened, the number of remaining white balls drawn is distributed as $X \sim \mathrm{Hypergeometric}(N, K, n)$. Thus the probability that we finally choose the special ball is $\frac{n+1}{N+1}\mathbf{E}[\frac{1}{X+1}]$.

Let us now think of the experiment in a different way. A necessary condition for success is that at least one white ball is drawn, which happens with probability p. Given that this occurs, the finally chosen white ball is just a random white ball (among all $K + 1$ white balls), so the probability that it is the special one is exactly $\frac{1}{K+1}$. Thus

$$\frac{n+1}{N+1}\mathbf{E}\left[\frac{1}{X+1}\right] = \frac{p}{K+1},$$

completing the proof of the lemma. $\qquad\square$

Next, we give a bound on the log-Sobolev constant from Equation (1.5) that is more tractable.

Lemma 2.3. *Let* ν *be a probability distribution of full support on* $[\ell]$. *Then the log-Sobolev constant* $\varrho_\nu^{\mathrm{triv}}$ *for the associated trivial Markov chain, given in Equation* (1.5), *satisfies the bound*

$$(\varrho_\nu^{\mathrm{triv}})^{-1} \leq \frac{1}{2}\sum_{i=1}^{\ell} \lg\left(\frac{1}{\nu_i}\right),$$

where $\lg$ *denotes* $\log_2$.

Proof. Let $p = \min_{i\in[\ell]}\{\nu(i)\}$, and assume without loss of generality that this minimum is achieved by $i = \ell$. Recall the formula of Diaconis and Saloffe-Coste [14],

$$\varrho_\nu^{\mathrm{triv}} = 2\frac{q-p}{\ln q - \ln p}, \quad \text{where } q = 1 - p.$$

What we need to show is

$$\frac{\ln(1/p) - \ln(1/(1-p))}{1 - 2p} \leq \lg(1/p) + \sum_{i=1}^{\ell-1} \lg(1/\nu(i)).$$

By convexity of $t \mapsto \lg(1/t)$ for $t \in (0, 1]$, the right-hand side above is at least

$$\lg(1/p) + (\ell - 1)\lg\left(\frac{\ell-1}{1-p}\right) \geq \lg(1/p) + \lg(1/(1-p)),$$

where the second inequality used $\ell \geq 2$. Thus it suffices to show

$$\ln(1/p) - \ln(1/(1-p)) \leq (1-2p)(\lg(1/p) + \lg(1/(1-p))), \quad 0 < p \leq 1/2.$$

(This inequality is simply the lemma we are trying to prove, restricted to the case $\ell = 2$.)

---

[2]In other words, X is the number of white balls among n random balls drawn (without replacement) from an urn containing N balls in total, K of which are white.

Write $p = 1/2 - \delta/2$, where $\delta \in [0, 1)$. Both sides of the above inequality are zero for $\delta = 0$; thus to establish the inequality it suffices to show the right-hand side's derivative is at least the left-hand side's. Taking derivatives, we need to show

$$\frac{2}{1 - \delta^2} \leq \lg\left(\frac{4}{1 - \delta^2}\right) + \frac{2\delta^2}{(\ln 2)(1 - \delta^2)}, \quad 0 \leq \delta < 1.$$

Multiplying this by $\frac{1 - \delta^2}{2} > 0$ gives

$$1 \leq 1 - \delta^2 + \frac{1}{2}(1 - \delta^2) \lg\left(\frac{1}{1 - \delta^2}\right) + \frac{\delta^2}{\ln 2} \iff 0 \leq \left(\frac{1}{\ln 2} - 1\right)\delta^2 + \frac{1}{2}(1 - \delta^2) \lg\left(\frac{1}{1 - \delta^2}\right),$$

which is evidently true as $\frac{1}{\ln 2} - 1 > 0$ and $0 < 1 - \delta^2 \leq 1$. $\qquad\square$

3 Bounding the log-Sobolev constant

In this section we will frequently identify a histogram $\kappa \in \mathbb{N}_+^\ell$ with the associated *multiset* of colors, namely the multiset with κ_i copies of i for each $i \in [\ell]$. Thus we may write $n = |\kappa| = \kappa_1 + \cdots + \kappa_\ell$. We will also use the notation $\boldsymbol{i} \sim \kappa$ to mean that $\boldsymbol{i}$ is chosen uniformly at random from the multiset κ; i.e., according to the probability distribution on $[\ell]$ in which i has probability κ_i/n. We will write $\underline{\kappa}$ for this probability distribution, and will need to refer to the log-Sobolev constant $\varrho_{\underline{\kappa}}^{\mathrm{triv}}$ from Equation (1.5).

Let us introduce one more piece of notation: if ϱ_κ denotes the optimal log-Sobolev constant for the transposition Markov chain on the multislice $\mathcal{U}_\kappa$, we will write

$$R_\kappa = \varrho_\kappa^{-1}.$$

Thus the goal of our Theorem 1.1 is to upper-bound R_κ. The midpoint of the proof will be establishing the following inductive bound:

Lemma 3.1. *Let $\kappa \in \mathbb{N}_+^\ell$ with $n = |\kappa| > 2$. Then*

$$R_\kappa \leq \frac{n-1}{n} \cdot (\varrho_{\underline{\kappa}}^{\mathrm{triv}})^{-1} + \max_{\substack{i_1, i_2 \in [\ell] \\ \mathrm{distinct}}} \left\{ \underset{\boldsymbol{i} \sim \kappa \setminus \{i_1, i_2\}}{\mathbf{E}} R_{\kappa \setminus \boldsymbol{i}} \right\}. \tag{3.1}$$

Given Theorem 3.1, the deduction of Theorem 1.1 will be elementary, though not completely straightforward; this is in Section 3.2. As for the deduction of Theorem 3.1 itself, it will mostly follow the proof Lee and Yau used [42] to analyze ϱ_κ in the cases $\ell = 2$ (the Hamming slice) and $\ell = n$ (the symmetric group S_n). Notice, however, that in both of these cases the "max" appearing in Inequality (3.1) becomes superfluous. In the $\ell = 2$ case, the only possibility for $\{i_1, i_2\}$ is $\{1, 2\}$, so we have the much simpler recursion

$$R_\kappa \leq \frac{n-1}{n} \cdot (\varrho_{\underline{\kappa}}^{\mathrm{triv}})^{-1} + \underset{\boldsymbol{i} \sim \kappa \setminus \{1, 2\}}{\mathbf{E}} R_{\kappa \setminus \boldsymbol{i}}.$$

In the $\ell = n$ case (meaning $\kappa_i = 1$ for all $i \in [\ell]$), we see by symmetry that every choice of i_1, i_2 leads to an isomorphic subproblem, that of bounding the log-Sobolev constant for S_{n-1}. That is, we have the even simpler recursion

$$R_{(1^n)} \leq \frac{n-1}{n} \cdot (\varrho_{(1/n, \ldots, 1/n)}^{\mathrm{triv}})^{-1} + R_{(1^{n-1})} \lesssim \tfrac{1}{2} \ln n + R_{(1^{n-1})},$$

where the asymptotic inequality used Equation (1.5). This recursion straightforwardly yields the known bound for the symmetric group, $R_{(1^N)} \lesssim \frac{1}{2} n \ln n$. A key point of our work is recognizing that one can use the Lee–Yau methodology to obtain Inequality (3.1), and that despite its somewhat complicated form, this recursion can be solved to yield a good bound.

3.1 Proving Theorem 3.1

Although much of the proof of Theorem 3.1 is from [42], we recapitulate it here for completeness and clarity. Fix $\kappa \in \mathbb{N}_+^\ell$ with $n = |\kappa| > 2$. Recall from Inequality (2.1) that R_κ is the smallest constant such that

$$\mathrm{KL}(\phi\pi \parallel \pi) \leq 2R_\kappa \cdot \mathcal{E}\left[\sqrt{\phi}\right] \tag{3.2}$$

holds for all probability densities on $\mathcal{U}_\kappa$.

The way we recursively bound R_κ involves applying the chain rule for KL divergence to $\mathrm{KL}(\phi\pi \parallel \pi)$. We will set up the notation for invoking the chain rule with respect to the nth coordinate. In fact, we will eventually apply it for *each* coordinate $k \in [n]$, and then take expectations over a uniformly random k. However, it will be notationally convenient to focus just on the $k = n$ case.

To this end, given a probability distribution ξ on $\mathcal{U}_\kappa$ (which will be either $\phi\pi$ or π), we will write ξ_n to denote its marginal on the nth coordinate (a probability distribution on $[\ell]$). Also, given a particular $a \in [\ell]$, we will write $\xi_{|a}$ to denote ξ's distribution on $\mathcal{U}_\kappa$ conditioned on the last coordinate having color a.

To begin the analysis of Inequality (3.2), let us write

$$\psi = \phi\pi \quad \text{(a probability distribution on } \mathcal{U}_\kappa\text{)}$$

and then apply the chain rule for KL divergence with respect to the nth coordinate:

$$\mathrm{KL}(\psi \parallel \pi) = \underbrace{\mathrm{KL}(\psi_n \parallel \pi_n)}_{\text{MARGINAL}_n} + \underbrace{\underset{a \sim \psi_n}{\mathbf{E}}\left[\mathrm{KL}(\psi_{|a} \parallel \pi_{|a})\right]}_{\text{CONDITIONAL}_n}.$$

We will now bound $\mathrm{MARGINAL}_n$ and $\mathrm{CONDITIONAL}_n$. In each case we will prove a bound that has a certain "dependence on the nth coordinate". We will then remark that we could have equally well proved an analogous bound involving the kth coordinate, for any $k \in [n]$. Finally, we will average this analogous bound over all $k \in [n]$. Adding the two averaged bounds from the MARGINAL and CONDITIONAL cases yields a valid upper bound on $\mathrm{KL}(\psi \parallel \pi)$,

$$\mathrm{KL}(\psi \parallel \pi) \leq \underset{k \sim [n]}{\mathrm{avg}}\left\{\text{bound on MARGINAL}_k\right\} + \underset{k \sim [n]}{\mathrm{avg}}\left\{\text{bound on CONDITIONAL}_k\right\}. \tag{3.3}$$

From this we will derive a recursive upper bound on R_κ via Inequality (3.2).

3.1.1 Bounding MARGINAL$_n$

Our bound on $\mathrm{MARGINAL}_n$ is from [42]. We think of the trivial Markov chain on $[\ell]$ with invariant distribution π_n, noting that π_n is nothing more than $\underline{\kappa}$. Applying the associated log-Sobolev inequality, we get

$$\mathrm{KL}(\psi_n \parallel \pi_n) \leq 2 \cdot (\varrho_{\underline{\kappa}}^{\mathrm{triv}})^{-1} \cdot \mathcal{E}_{\pi_n}\left[\sqrt{\psi_n/\pi_n}\right], \tag{3.4}$$

where we wrote $\mathcal{E}_{\pi_n}$ to denote the energy functional for the trivial Markov chain. It is simple to check that the function ψ_n/π_n on $[\ell]$ is just $a \mapsto \mathbf{E}_{\boldsymbol{u} \sim \pi_{|a}}[\phi(\boldsymbol{u})]$. Thus

$$\mathcal{E}_{\pi_n}\left[\sqrt{\psi_n/\pi_n}\right] = \frac{1}{2} \underset{\boldsymbol{a},\boldsymbol{b} \sim \pi_n}{\mathbf{E}}\left(\sqrt{\underset{\boldsymbol{u} \sim \pi_{|\boldsymbol{a}}}{\mathbf{E}}[\phi(\boldsymbol{u})]} - \sqrt{\underset{\boldsymbol{v} \sim \pi_{|\boldsymbol{b}}}{\mathbf{E}}[\phi(\boldsymbol{v})]}\right)^2. \tag{3.5}$$

We would prefer if the two inner expectations here were over the same probability space. To that end, observe that for any $a, b \in [\ell]$ (not necessarily distinct), we can make a

draw $\boldsymbol{v} \sim \pi_{|b}$ in the following unusual way. First, draw $\boldsymbol{u} \sim \pi_{|a}$. Next, draw $\boldsymbol{j} \sim \boldsymbol{u}^{-1}(b)$, where we have introduced the notation $u^{-1}(b)$ for the set of all coordinates j with $u_j = b$. Finally, form $\boldsymbol{v} = \boldsymbol{u}^{(\boldsymbol{j}\ n)}$. (Here we are abusing notation by allowing the possibility of $\boldsymbol{j} = n$, so that the "transposition" $(\boldsymbol{j}\ n)$ may be the identity.) Thus we have

$$\mathop{\mathbf{E}}_{\boldsymbol{v} \sim \pi_{|b}}[\phi(\boldsymbol{v})] = \mathop{\mathbf{E}}_{\substack{\boldsymbol{u} \sim \pi_{|a} \\ \boldsymbol{j} \sim \boldsymbol{u}^{-1}(b)}}[\phi^{(\boldsymbol{j}\ n)}(\boldsymbol{u})],$$

where use the notation $\phi^\tau(u) := \phi(u^\tau)$. Putting this into Equation (3.5) (and also "pointlessly" choosing $\boldsymbol{j} \sim \boldsymbol{u}^{-1}(\boldsymbol{b})$ in the first inner expectation) we get

$$\mathcal{E}_{\pi_n}\left[\sqrt{\psi_n/\pi_n}\right] = \frac{1}{2}\mathop{\mathbf{E}}_{\boldsymbol{a},\boldsymbol{b} \sim \pi_n}\left|\sqrt{\mathop{\mathbf{E}}_{\substack{\boldsymbol{u} \sim \pi_{|a} \\ \boldsymbol{j} \sim \boldsymbol{u}^{-1}(\boldsymbol{b})}}[\phi(\boldsymbol{u})]} - \sqrt{\mathop{\mathbf{E}}_{\substack{\boldsymbol{u} \sim \pi_{|a} \\ \boldsymbol{j} \sim \boldsymbol{u}^{-1}(\boldsymbol{b})}}[\phi^{(\boldsymbol{j}\ n)}(\boldsymbol{u})]}\right|^2. \qquad (3.6)$$

Considering the quantity inside the outer expectation, we further use

$$\left|\sqrt{\mathop{\mathbf{E}}_{\substack{\boldsymbol{u} \sim \pi_{|a} \\ \boldsymbol{j} \sim \boldsymbol{u}^{-1}(\boldsymbol{b})}}[\phi(\boldsymbol{u})]} - \sqrt{\mathop{\mathbf{E}}_{\substack{\boldsymbol{u} \sim \pi_{|a} \\ \boldsymbol{j} \sim \boldsymbol{u}^{-1}(\boldsymbol{b})}}[\phi^{(\boldsymbol{j}\ n)}(\boldsymbol{u})]}\right| = \left|\left\|\sqrt{\phi}\right\|_2 - \left\|\sqrt{\phi^{(\cdot\ n)}}\right\|_2\right| \leq \left\|\sqrt{\phi} - \sqrt{\phi^{(\cdot\ n)}}\right\|_2,$$

where $\|\cdot\|_2$ is the 2-norm defined by the distribution on $(\boldsymbol{u}, \boldsymbol{j})$, and we used the triangle inequality. Putting this back into Equation (3.6) yields

$$\mathcal{E}\left[\sqrt{\mu_n/\pi_n}\right] \leq \frac{1}{2}\mathop{\mathbf{E}}_{\boldsymbol{a},\boldsymbol{b} \sim \pi_n}\left[\left\|\sqrt{\phi} - \sqrt{\phi^{(\cdot\ n)}}\right\|_2^2\right] = \mathop{\mathbf{E}}_{\boldsymbol{a},\boldsymbol{b} \sim \pi_n}\mathop{\mathbf{E}}_{\substack{\boldsymbol{u} \sim \pi_{|a} \\ \boldsymbol{j} \sim \boldsymbol{u}^{-1}(\boldsymbol{b})}}\left[e\left(\sqrt{\phi};\boldsymbol{u}, \boldsymbol{u}^{(\boldsymbol{j}\ n)}\right)\right], \quad (3.7)$$

where we have introduced the shorthand

$$e(f; u, v) = \tfrac{1}{2}(f(u) - f(v))^2.$$

In the right-hand expectation in Inequality (3.7), $\boldsymbol{u}$ is simply distributed according to π. Furthermore, suppose we fix any outcome $\boldsymbol{u} = u$. Let us consider the joint distribution of $\boldsymbol{b} \sim \pi_n$ and $\boldsymbol{j} \sim u^{-1}(\boldsymbol{b})$. Since $u \in \mathcal{U}_\kappa$, we could equivalently form $\boldsymbol{b}$ by choosing a random color within $\boldsymbol{u}$. But in this case, $\boldsymbol{j}$ is formed by first choosing a random color within u, and then taking a random coordinate where u has that same color. It is clear that the resulting distribution on $\boldsymbol{j}$ is simply uniformly random on $[n]$. Thus the right-hand side of Inequality (3.7) is simply

$$\mathop{\mathbf{E}}_{\boldsymbol{u} \sim \pi}\mathop{\mathbf{E}}_{\boldsymbol{j} \sim [n]}\left[e\left(\sqrt{\phi}; \boldsymbol{u}, \boldsymbol{u}^{(\boldsymbol{j}\ n)}\right)\right].$$

Putting this into Inequality (3.7) and then into Inequality (3.4), we conclude

$$\mathrm{MARGINAL}_n \leq 2 \cdot (\varrho_{\underline{\kappa}}^{\mathrm{triv}})^{-1} \cdot \mathop{\mathbf{E}}_{\boldsymbol{u} \sim \pi}\mathop{\mathbf{E}}_{\boldsymbol{j} \sim [n]}\left[e\left(\sqrt{\phi}; \boldsymbol{u}, \boldsymbol{u}^{(\boldsymbol{j}\ n)}\right)\right].$$

This bound we have derived depends on the nth coordinate only through the transposition of $\boldsymbol{j}$ with n. If we repeat this derivation for a general coordinate $k \in [n]$, and then average over k, we will get

$$\mathop{\mathrm{avg}}_{\boldsymbol{k} \sim [n]}\left\{\text{bound on } \mathrm{MARGINAL}_{\boldsymbol{k}}\right\} = 2 \cdot \frac{n-1}{n} \cdot (\varrho_{\underline{\kappa}}^{\mathrm{triv}})^{-1} \cdot \mathop{\mathbf{E}}_{\substack{\boldsymbol{u} \sim \pi \\ \boldsymbol{\tau} \sim \mathrm{Trans}(n)}}\left[e\left(\sqrt{\phi}; \boldsymbol{u}, \boldsymbol{u}^{\boldsymbol{\tau}}\right)\right]$$

$$= 2 \cdot \frac{n-1}{n} \cdot (\varrho_{\underline{\kappa}}^{\mathrm{triv}})^{-1} \cdot \mathcal{E}\left[\sqrt{\phi}\right], \qquad (3.8)$$

where the $\frac{n-1}{n}$ factor accounts for the fact that when $\boldsymbol{j}$ and $\boldsymbol{k}$ are uniformly random, there is a $\frac{1}{n}$ chance that they are equal — in which case the "transposition" $(\boldsymbol{j}\ \boldsymbol{k})$ is the identity and we get a contribution of $e(\sqrt{\phi}; \boldsymbol{u}, \boldsymbol{u}) = 0$.

3.1.2 Bounding CONDITIONAL$_n$

To bound CONDITIONAL$_n$, we again follow [42] to a certain point. By definition of R we have

$$\mathop{\mathbf{E}}_{\boldsymbol{a}\sim\psi_n}\left[\mathrm{KL}(\psi_{|\boldsymbol{a}}\parallel\pi_{|\boldsymbol{a}})\right] \leq \mathop{\mathbf{E}}_{\boldsymbol{a}\sim\psi_n}\left[2R_{\kappa\backslash\boldsymbol{a}}\cdot\mathcal{E}_{\kappa\backslash\boldsymbol{a}}\left[\sqrt{\frac{\psi_{|\boldsymbol{a}}}{\pi_{|\boldsymbol{a}}}}\right]\right].$$

Note that $\psi_{|\boldsymbol{a}}=\psi(\cdot,\boldsymbol{a})/\psi_n(\boldsymbol{a})$, similarly for π, and also that $\mathcal{E}[c\cdot f]=c^2\cdot\mathcal{E}[f]$. Thus

$$\mathop{\mathbf{E}}_{\boldsymbol{a}\sim\psi_n}\left[2R_{\kappa\backslash\boldsymbol{a}}\cdot\mathcal{E}_{\kappa\backslash\boldsymbol{a}}\left[\sqrt{\frac{\psi_{|\boldsymbol{a}}}{\pi_{|\boldsymbol{a}}}}\right]\right] = 2\mathop{\mathbf{E}}_{\boldsymbol{a}\sim\psi_n}\left[R_{\kappa\backslash\boldsymbol{a}}\cdot\frac{\pi_n(\boldsymbol{a})}{\psi_n(\boldsymbol{a})}\cdot\mathcal{E}_{\kappa\backslash\boldsymbol{a}}\left[\sqrt{\frac{\psi(\cdot,\boldsymbol{a})}{\pi(\cdot,\boldsymbol{a})}}\right]\right]$$

$$= 2\mathop{\mathbf{E}}_{\boldsymbol{a}\sim\pi_n}\left[R_{\kappa\backslash\boldsymbol{a}}\cdot\mathcal{E}_{\kappa\backslash\boldsymbol{a}}\left[\sqrt{\frac{\psi(\cdot,\boldsymbol{a})}{\pi(\cdot,\boldsymbol{a})}}\right]\right].$$

We can further write

$$\mathcal{E}_{\kappa\backslash\boldsymbol{a}}\left[\sqrt{\frac{\psi(\cdot,\boldsymbol{a})}{\pi(\cdot,\boldsymbol{a})}}\right] = \mathcal{E}_{\kappa\backslash\boldsymbol{a}}\left[\sqrt{\phi(\cdot,\boldsymbol{a})}\right] = \mathop{\mathbf{E}}_{\substack{\overline{\boldsymbol{u}}\sim\pi_{\kappa\backslash\boldsymbol{a}}\\\boldsymbol{\tau}\sim\mathrm{Trans}(n-1)}}\left[e\left(\sqrt{\phi};(\overline{\boldsymbol{u}},\boldsymbol{a}),(\overline{\boldsymbol{u}}^{\boldsymbol{\tau}},\boldsymbol{a})\right)\right].$$

Combining all previous deductions, we get

$$\mathop{\mathbf{E}}_{\boldsymbol{a}\sim\psi_n}\left[\mathrm{KL}(\psi_{|\boldsymbol{a}}\parallel\pi_{|\boldsymbol{a}})\right] \leq 2\mathop{\mathbf{E}}_{\boldsymbol{a}\sim\pi_n}R_{\kappa\backslash\boldsymbol{a}}\cdot\mathop{\mathbf{E}}_{\substack{\overline{\boldsymbol{u}}\sim\pi_{\kappa\backslash\boldsymbol{a}}\\\boldsymbol{\tau}\sim\mathrm{Trans}(n-1)}}\left[e\left(\sqrt{\phi};(\overline{\boldsymbol{u}},\boldsymbol{a}),(\overline{\boldsymbol{u}}^{\boldsymbol{\tau}},\boldsymbol{a})\right)\right]$$

$$= 2\mathop{\mathbf{E}}_{\substack{\boldsymbol{u}\sim\pi\\\boldsymbol{\tau}\sim\mathrm{Trans}(n-1)}}\left[e\left(\sqrt{\phi};\boldsymbol{u},\boldsymbol{u}^{\boldsymbol{\tau}}\right)\cdot R_{\kappa\backslash\boldsymbol{u}_n}\right].$$

This is our desired bound for CONDITIONAL$_n$, except we make a slight adjustment so that the expectation is over all $\boldsymbol{\tau}$ in $\mathrm{Trans}(n)$, obtaining the equivalent bound

$$\mathrm{CONDITIONAL}_n \leq 2\mathop{\mathbf{E}}_{\substack{\boldsymbol{u}\sim\pi\\\boldsymbol{\tau}\sim\mathrm{Trans}(n)}}\left[e\left(\sqrt{\phi};\boldsymbol{u},\boldsymbol{u}^{\boldsymbol{\tau}}\right)\cdot\frac{\mathbb{1}[n\text{ is fixed by }\boldsymbol{\tau}]}{1-2/n}\cdot R_{\kappa\backslash\boldsymbol{u}_n}\right].$$

Again, had we repeated this derivation for an arbitrary coordinate k in place of the nth, and then averaged over k, we would get

$$\mathop{\mathrm{avg}}_{\boldsymbol{k}\sim[n]}\left\{\text{bound on CONDITIONAL}_k\right\} = 2\mathop{\mathbf{E}}_{\substack{\boldsymbol{u}\sim\pi\\\boldsymbol{\tau}\sim\mathrm{Trans}(n)}}\left[e\left(\sqrt{\phi};\boldsymbol{u},\boldsymbol{u}^{\boldsymbol{\tau}}\right)\cdot\mathop{\mathbf{E}}_{\boldsymbol{k}\sim\mathrm{Fix}(\boldsymbol{\tau})}R_{\kappa\backslash\boldsymbol{u}_k}\right], \quad (3.9)$$

where $\mathrm{Fix}(\boldsymbol{\tau})$ denotes the fixed points of transposition $\boldsymbol{\tau}$.

This is the point at which, by necessity, we depart from [42]. To proceed, we simply take a worst-case upper bound on the two colors swapped by $\boldsymbol{\tau}$; no matter what $\boldsymbol{u}$ and $\boldsymbol{\tau}$ are, we have

$$\mathop{\mathbf{E}}_{\boldsymbol{k}\sim\mathrm{Fix}(\boldsymbol{\tau})}R_{\kappa\backslash\boldsymbol{u}_k} \leq \max_{i_1,i_2\in[\ell]}\left\{\mathop{\mathbf{E}}_{\boldsymbol{i}\sim\kappa\backslash\{i_1,i_2\}}R_{\kappa\backslash\boldsymbol{i}}\right\}.$$

In fact, when inserting this into Inequality (3.9), we can do slightly better. Notice that if $\boldsymbol{\tau}$ swaps two colors of $\boldsymbol{u}$ that are the same, then $e\left(\sqrt{\phi};\boldsymbol{u},\boldsymbol{u}^{\boldsymbol{\tau}}\right)=0$ anyway. Thus we may insert the indicator random variable $\mathbb{1}[\boldsymbol{\tau}\text{ swaps distinct colors in }\boldsymbol{u}]$ into the expectation in Inequality (3.9), and then use

$$\mathbb{1}[\boldsymbol{\tau}\text{ swaps distinct colors in }\boldsymbol{u}]\cdot\mathop{\mathbf{E}}_{\boldsymbol{k}\sim\mathrm{Fix}(\boldsymbol{\tau})}R_{\kappa\backslash\boldsymbol{u}_k} \leq \max_{\substack{i_1,i_2\in[\ell]\\\text{distinct}}}\left\{\mathop{\mathbf{E}}_{\boldsymbol{i}\sim\kappa\backslash\{i_1,i_2\}}R_{\kappa\backslash\boldsymbol{i}}\right\}.$$

Putting this inequality into Inequality (3.9) yields

$$\mathop{\mathrm{avg}}_{\boldsymbol{k}\sim[n]}\left\{\text{bound on CONDITIONAL}_k\right\} \leq 2\max_{\substack{i_1,i_2\in[\ell]\\\text{distinct}}}\left\{\mathop{\mathbf{E}}_{\boldsymbol{i}\sim\kappa\backslash\{i_1,i_2\}}R_{\kappa\backslash\boldsymbol{i}}\right\}\cdot\mathcal{E}\left[\sqrt{\phi}\right]. \quad (3.10)$$

3.1.3 Completing the proof of Theorem 3.1

Putting together Inequalities (3.3), (3.8) and (3.10) yields

$$\mathrm{KL}(\phi\pi \parallel \pi) \le 2\left(\frac{n-1}{n} \cdot (\varrho_{\underline{\kappa}}^{\mathrm{triv}})^{-1} + \max_{\substack{i_1,i_2 \in [\ell] \\ \mathrm{distinct}}} \left\{ \mathop{\mathbf{E}}_{\boldsymbol{i} \sim \kappa \setminus \{i_1,i_2\}} R_{\kappa \setminus \boldsymbol{i}} \right\} \right) \cdot \mathcal{E}\left[\sqrt{\phi} \right],$$

which immediately implies Theorem 3.1.

3.2 Deducing Theorem 1.1 from Theorem 3.1

In this section we upper-bound R_κ using the recursion given by Theorem 3.1. Let us make a few simplifications to the statement of Theorem 3.1. First, let us drop the factor $\frac{n-1}{n}$ for simplicity. Second, let us use the upper bound on $(\varrho_{\underline{\kappa}}^{\mathrm{triv}})^{-1}$ from Theorem 2.3. Finally, let us drop the condition that i_1, i_2 be distinct in the $\max$; this condition greatly simplifies the recursion when $\ell = 2$ (as in [42]), but doesn't help us much when $\ell > 2$. Thus we will finally use

$$R_\kappa \le c(\kappa) + \max_{\{i_1,i_2\} \subseteq \kappa} \left\{ \mathop{\mathbf{E}}_{\boldsymbol{i} \sim \kappa \setminus \{i_1,i_2\}} R_{\kappa \setminus \boldsymbol{i}} \right\}, \quad c(\kappa) := \frac{1}{2} \sum_{i\,:\,\kappa_i > 0} \lg\left(\frac{|\kappa|}{\kappa_i} \right) \quad \text{for } |\kappa| > 2. \quad (3.11)$$

We remind the reader that in the above, $\{i_1, i_2\}$ and κ are considered to be multisets of $[\ell]$. In fact, we will always consider κ to merely be a multiset of the colors on which it is supported. That is to say, whenever some κ_i becomes 0 in the above recursion (through the removal of a color inside the expectation), we will treat the color i as no longer existing (rather than allowing $\kappa_i = 0$). This is acceptable, since the definition of R_κ is not affected by removing colors that don't appear in κ. This is why we dropped the hypothesis $\kappa \in \mathbb{N}_+^\ell$ in writing Inequality (3.11), and why we wrote the sum in $c(\kappa)$ as being over $\{i : \kappa_i > 0\}$, rather than over all $i \in [\ell]$. It might seem that dropping the hypothesis $\kappa \in \mathbb{N}_+^\ell$ (and hence $\kappa_i > 0$ for all i) in Inequality (3.11) could cause a problem for the case when the number of colors drops to just one, meaning $\kappa = \{i, i, \dots, i\}$ for some i. However, in this degenerate case, the correct value of R_κ is 0, and we also have $c(\kappa) = 0$.

Regarding the base cases of $|\kappa| = 2$ for our recursion Inequality (3.11), we have the true values

$$R_{\{a,b\}} = \left\{ \begin{array}{ll} 0 & \text{if } a = b \\ 1/2 & \text{if } a \ne b \end{array} \right\} \le 1/2. \quad (3.12)$$

Indeed, $\varrho_{\underline{\{1,2\}}}^{\mathrm{triv}} = 1$ according to Equation (1.5), and the energy of this trivial chain is half that of the transposition chain on $\mathcal{U}_{\{1,2\}}$.

"Strategies". Given κ, let's define a *strategy for* κ to be a mapping p that takes in an arbitrary nonempty $\mu \subseteq \kappa$ and outputs a pair $\{i_1, i_2\} \subseteq \mu$. We'll say that $p(\mu) = \{i_1, i_2\}$ is the pair *protected* by p. We'll write $\mathrm{Strat}(\kappa)$ for the set of all strategies for κ. Then from Inequality (3.11) we have that

$$R_\kappa \le \max_{p \in \mathrm{Strat}(\kappa)} S_p(\kappa), \quad (3.13)$$

where $S_p(\kappa)$ is defined to be the solution of the recursion

$$S_p(\kappa) = c(\kappa) + \mathop{\mathbf{E}}_{\boldsymbol{i} \sim \kappa \setminus p(\kappa)} S_p(\kappa \setminus \boldsymbol{i}), \quad (3.14)$$

with the base case $S_p(\text{pair}) = 1/2$ from Inequality (3.12). Let us write

$$c(\kappa) = \sum_{a \in [\ell]} c^{(a)}(\kappa), \quad \text{where } c^{(a)}(\kappa) := \mathbb{1}[\kappa_a \neq 0] \cdot \frac{1}{2} \lg\left(\frac{|\kappa|}{\kappa_a}\right).$$

Then it follows from Equation (3.14) that we have

$$S_p(\kappa) = \sum_{a=1}^{\ell} S_p^{(a)},$$

where

$$S_p^{(a)}(\kappa) = c^{(a)}(\kappa) + \mathop{\mathbf{E}}_{\boldsymbol{i} \sim \kappa \backslash p(\kappa)} S_p^{(a)}(\kappa \setminus \boldsymbol{i}). \tag{3.15}$$

Thus returning to Inequality (3.13), we have

$$R_\kappa \leq \max_{p \in \text{Strat}(\kappa)} \left\{ \sum_{a=1}^{\ell} S_p^{(a)}(\kappa) \right\} \leq \sum_{a=1}^{\ell} \max_{p \in \text{Strat}(\kappa)} S_p^{(a)}(\kappa). \tag{3.16}$$

Next, it will be convenient if $c^{(a)}(\kappa)$ is a decreasing function of κ_a (considering $|\kappa|$ fixed), *even for $\kappa_a = 0$*. So let us

$$\text{redefine } c^{(a)}(\kappa) := \frac{1}{2} \lg\left(\frac{2|\kappa|}{1 + \kappa_a}\right). \tag{3.17}$$

This redefinition only increases $c^{(a)}(\kappa)$; it increases it from zero to a nonzero value when $\kappa_a = 0$; and, when $\kappa_a > 0$, the increase from κ_a to $\kappa_a + 1$ in the denominator is at most a factor of 2, and this is compensated for by the new factor of 2 in the numerator. Thus Inequality (3.16) is still valid under our redefinition.

The advantage of the redefinition is, as mentioned, that $c^{(a)}(\kappa)$ always goes up when κ_a drops by one, even when dropping from 1 down to 0. Thus for each fixed $a \in [\ell]$, it is now clear from from Equation (3.15) that *the optimal strategies $p \in \text{Strat}(\kappa)$ for maximizing $S_p^{(a)}(\kappa)$ are precisely the "greedy" ones*. Here the "greedy" strategies for color a mean the ones that "always protect non-a colors" (to the extent this is possible — if μ has only $m < 2$ non-a colors then $p(\mu)$ will be obliged to contain $2 - m$ a's). It is also not hard to see that every such greedy strategy g is equally effective; for the purposes of computing $S_g^{(a)}(\kappa)$, it doesn't matter what non-a colors appear in the $\mu \subseteq \kappa$ that arise — only how many of them there are. (This observation relies in part on using the same upper bound, $1/2$, for both cases in Inequality (3.12).)

Greedy strategies. Let $S_g^{(a)}(\kappa)$ denote the solution for a greedy protection strategy, which, as we have argued, equals $\max_{p \in \text{Strat}(\kappa)}\{S_p^{(a)}(\kappa)\}$. Our final goal will be to establish

$$S_g^{(a)}(\kappa) \leq \frac{1}{2} n \lg \frac{4n}{\kappa_a}. \tag{3.18}$$

Putting this bound into Inequality (3.16) will yield Theorem 1.1.

We first dispense with the edge case when κ contains fewer than 2 non-a's. In this case, under the greedy strategy $\boldsymbol{i}$ is always a in Equation (3.15), and so "solving the recursion" just amounts to computing a sum. When $\kappa_a = |\kappa|$, the result is

$$S_g^{(a)} = \frac{1}{2} \lg \frac{2n}{1+n} + \cdots + \frac{1}{2} \lg \frac{2 \cdot 3}{1+3} + \frac{1}{2},$$

and when $\kappa_a = |\kappa| - 1$, the result is

$$S_g^{(a)} = \frac{1}{2} \lg \frac{2n}{n} + \cdots + \frac{1}{2} \lg \frac{2 \cdot 3}{3} + \frac{1}{2}.$$

In both cases, each of the $n-1$ summands is at most $\frac{1}{2}\lg 2$, from which Inequality (3.18) immediately follows.

We now come to the main case, when κ contains at least 2 non-a's. In this case, the greedy strategy involves always protecting two non-a's, and as argued, the quantity $S_g^{(a)}(\kappa)$ only depends on κ_a. Thus for analysis purposes, we may henceforth assume $\ell = 2$ and $a = 1$. Now from Inequality (3.16) and Equation (3.17) we conclude that $S_g^{(a)}(\kappa)$ is the solution of

$$S_g^{(1)}(\kappa) = \frac{1}{2}\lg\left(\frac{2|\kappa|}{1+\kappa_1}\right) + \frac{\kappa_1}{\kappa_1 + \kappa_2 - 2}S_g^{(1)}(\kappa \setminus 1) + \frac{\kappa_2 - 2}{\kappa_1 + \kappa_2 - 2}S_g^{(1)}(\kappa \setminus 2),$$

with base case $S_g^{(1)}(\{2,2\}) = 1/2$. It is easier to analyze this recursion in terms of $\lambda := \kappa \setminus \{2,2\}$. Writing $G(\lambda) = S_g^{(1)}(\kappa)$, we have

$$G(\lambda) = \frac{1}{2}\lg\left(\frac{2(\lambda_1 + \lambda_2 + 2)}{1+\lambda_1}\right) + \frac{\lambda_1}{\lambda_1 + \lambda_2}G(\lambda \setminus 1) + \frac{\lambda_2}{\lambda_1 + \lambda_2}G(\lambda \setminus 2), \qquad (3.19)$$

with base case $G(\emptyset) = 1/2$. In fact, it will be convenient to overpay for the base case, taking $G(\emptyset) = \frac{1}{2}\lg\left(\frac{2(0+0+2)}{1+0}\right) = 1$.

Finally, we can solve the recursion in Equation (3.19) by giving it a probabilistic interpretation. Suppose we choose a random string $\boldsymbol{u}$ in $\mathcal{U}_\lambda$. Then we begin a deterministic process with "stages" numbered $|\lambda|, |\lambda| - 1, |\lambda| - 2, \ldots, 1, 0$. In each stage, we "pay" $\frac{1}{2}\lg\left(\frac{2(\#_1\boldsymbol{u}+\#_2\boldsymbol{u}+2)}{1+\#_1\boldsymbol{u}}\right)$, and then we delete the last character in $\boldsymbol{u}$. It is easy to see that the solution of Equation (3.19) is equal to the expectation (over the initial choice of $\boldsymbol{u}$) of the total payment in this process. By linearity of expectation, this total payment is the sum of the expected payment in each stage, and in the mth stage it is clear that the random variable $\#_1\boldsymbol{u}$ is distributed as Hypergeometric$(\lambda_1 + \lambda_2, \lambda_1, m)$. Thus the expected payment in the mth stage is

$$\mathbf{E}\left[\frac{1}{2}\lg\left(\frac{2(m+2)}{1+\boldsymbol{X}}\right)\right] \leq \frac{1}{2}\lg\left(\mathbf{E}\left[\frac{2(m+2)}{1+\boldsymbol{X}}\right]\right), \quad \text{for } \boldsymbol{X} \sim \text{Hypergeometric}(\lambda_1 + \lambda_2, \lambda_1, m).$$

By Theorem 2.2, this is at most

$$\frac{1}{2}\lg\left(\frac{2(m+2)(\lambda_1 + \lambda_2 + 1)}{(\lambda_1 + 1)(m+1)}\right) \leq 1 + \frac{1}{2}\lg\frac{\lambda_1 + \lambda_2 + 1}{\lambda_1 + 1}.$$

(The bound here is a little loose, but we valued simplicity over optimization of lower-order terms.) When this is summed over $0 \leq m \leq |\lambda|$, we get an upper bound of

$$|\lambda| + 1 + \frac{|\lambda|}{2}\lg\frac{|\lambda| + 1}{\lambda_1 + 1}.$$

Recalling $\lambda = \kappa \setminus \{2,2\}$, and using $|\kappa| - 1 = n - 1 \leq n$, the above upper bound is at most

$$n + \frac{1}{2}n\lg\frac{n}{\kappa_1} = \frac{1}{2}n\lg\frac{4n}{\kappa_1},$$

confirming Inequality (3.18).

4 Applications

4.1 KKL and Kruskal–Katona for multislices

By viewing the multislice as a Schreier graph, we can apply the results of [50, 51] to obtain the KKL Theorem in this setting (in fact, Talagrand's strengthening [59] of it):

Theorem 4.1. *Let $f : \mathcal{U}_\kappa \to \{0, 1\}$. Then*

$$\operatorname*{avg}_{\tau \in \mathrm{Trans}(n)} \left\{ \frac{\mathrm{Inf}_\tau[f]}{\lg(2/\mathrm{Inf}_\tau[f])} \right\} \gtrsim \rho_\kappa \cdot \mathbf{Var}_{\pi_\kappa}[f].$$

(Since we have written $\gtrsim$, hiding a universal constant, it doesn't matter if we take f's range to be $\{0, 1\}$ or $\{-1, 1\}$.) Substituting our lower bound on ρ_κ from Theorem 1.1 yields concrete new results. For example, consider our model scenario of $n \longrightarrow \infty$ with $\ell = O(1)$ and $\kappa_i/n \geq \Omega(1)$ for each i; suppose further that f is "roughly balanced", meaning $\Omega(1) \leq \mathbf{Var}[f] \leq 1 - \Omega(1)$. Then

$$\operatorname*{avg}_{\tau \in \mathrm{Trans}(n)} \left\{ \frac{\mathrm{Inf}_\tau[f]}{\lg(2/\mathrm{Inf}_\tau[f])} \right\} \gtrsim \frac{1}{n}, \quad \text{and hence} \quad \mathcal{M}[f] \gtrsim \frac{\log n}{n}.$$

The latter statement here is the traditional conclusion of the KKL Theorem.

Let us record here one more concrete corollary of Theorem 4.1. In our model scenario, that theorem (roughly speaking) says that the energy $\mathcal{E}[1_A] = \operatorname{avg}_{\tau \in \mathrm{Trans}(n)} \mathrm{Inf}_\tau[1_A]$ is at least $\Omega\left(\frac{\log n}{n}\right)$ unless some transposition $(i \ j)$ has a rather large influence, like $1/n^{.01}$, on 1_A.

Corollary 4.2. *Let $A \subseteq \mathcal{U}_\kappa$. Assume $\kappa_i \geq pn$ for all $i \in [\ell]$ and that $\epsilon \leq \mathrm{vol}(A) \leq 1 - \epsilon$. Then*

$$\mathcal{E}[1_A] \geq \Omega\left(\frac{\epsilon}{\ell \log(1/p)}\right) \cdot \frac{\log(1/\mathcal{M}[1_A])}{n}.$$

Proof. This is immediate from Theorem 4.1 and Theorem 1.1, using $\lg(2/\mathrm{Inf}_\tau[1_A]) \geq \lg(2/\mathcal{M}[1_A])$, $\rho_\kappa^{-1} \lesssim n \cdot \ell \cdot \log(1/p)$, and $\mathbf{Var}[1_A] = \mathrm{vol}(A)(1 - \mathrm{vol}(A)) \geq \epsilon/2$. $\square$

Following [50], we will use this to show a variant of the Kruskal–Katona Theorem for multislices.[3]

The classical Kruskal–Katona Theorem [58, 40, 36] concerns subsets of Hamming slices of the Boolean cube. To recall it, let us write a 2-color histogram $\kappa \in \mathbb{N}_+^2$ as (κ_0, κ_1), with $n = \kappa_0 + \kappa_1$. If $A \subseteq \mathcal{U}_\kappa$, then the *(lower) shadow* of A is defined to be

$$\partial A = \left\{ v \in \mathcal{U}_{(\kappa_0+1, \kappa_1-1)} : v \leq u \text{ for some } u \in A \right\}.$$

It is not hard to show that $\mathrm{vol}(\partial A) \geq \mathrm{vol}(A)$ always (here the fractional volume $\mathrm{vol}(\partial A)$ is vis-à-vis the containing slice $\mathcal{U}_{(\kappa_0+1, \kappa_1-1)}$). The Kruskal–Katona Theorem improves this by giving an exactly sharp lower bound on $\mathrm{vol}(\partial A)$ as a function of $\mathrm{vol}(A)$. The precise function is somewhat cumbersome to state, but the qualitative consequence, assuming that $\mathrm{vol}(A)$ and κ_0/n are bounded away from 0 and 1, is that $\mathrm{vol}(\partial A) \geq \mathrm{vol}(A) + \Omega(1/n)$. This is sharp, up to the constant in the $\Omega(\cdot)$, as witnessed by the "dictator set" $A = \{u : u_1 = 0\}$. See [50, Sec. 1.2] for more discussion.

To extend the Kruskal–Katona Theorem to multislices, we first need to extend the notion of neighboring slices and shadows. Fix an ordering on the colors, $1 \prec 2 \prec \cdots \prec \ell$. This total order extends to a partial order on strings in $[\ell]^n$ in the natural way.

Definition 4.3. *Let $\kappa \in \mathbb{N}_+^\ell$ be a histogram. We say that histogram κ' is a* lower neighbor *of κ, and write $\kappa' \lhd \kappa$, if there exists some $c \prec d \in [\ell]$ such that $\kappa'_c = \kappa_d + 1$, $\kappa'_d = \kappa_c - 1$, and $\kappa'_i = \kappa_i$ for all other colors i. In the opposite case, when $c \succ d$, we say κ' is an* upper neighbor *of κ, and write $\kappa' \rhd \kappa$.*

The main difference between the Boolean case and the multicolored case is that each multislice now has multiple upper and lower neighbors.

---

[3]Our variants are unrelated to those of Clements [8, 9, 10].

Definition 4.4. *Let $A \subseteq \mathcal{U}_\kappa$, and let $\kappa' \lhd \kappa$. The* lower shadow *of A at κ' is*

$$\partial_{\kappa'} A = \{u \in \mathcal{U}_{\kappa'} : u \prec v \text{ for some } v \in A\}.$$

We similarly define upper shadows. We may use the same notation $\partial_{\kappa'} A$ for both kinds of shadows, since whether a shadow is upper or lower is determined by whether $\kappa' \rhd \kappa$ or $\kappa' \lhd \kappa$.

Towards proving Kruskal–Katona theorems, we relate the volume of A's lower shadows to $\mathcal{E}[1_A]$. Recalling that A now has multiple lower shadows, we show that a certain weighted average of their volumes is noticeably larger than the volume of A. We first define the appropriate weighted average.

Definition 4.5. *Given a histogram $\kappa \in \mathbb{N}_+^\ell$, we define a natural probability distribution* **lower**(κ) *on the lower neighbors of κ as follows. To draw $\kappa' \sim$ **lower**(κ): take an arbitrary $u \in \mathcal{U}_\kappa$; choose $j, j' \sim [n]$ independently and randomly, conditioned on $u_j \neq u_{j'}$; let c, d denote the two colors $u_j, u_{j'}$, with the convention $c \prec d$; finally, let κ' be the lower neighbor of κ with $\kappa'_c = \kappa_c + 1$ and $\kappa'_d = \kappa_d - 1$.*

We similarly define a probability distribution **upper**(κ) *on the upper neighbors of κ by interchanging the roles of c and d.*

Proposition 4.6. *Given a histogram $\kappa \in \mathbb{N}_+^\ell$, let $h(\kappa) = 1 - \sum_{i=1}^\ell \frac{\kappa_i(\kappa_i - 1)}{n(n-1)} \leq 1$. (This is the probability that applying a random transposition to a string in $\mathcal{U}_\kappa$ actually changes it.) Then for any $A \subseteq \mathcal{U}_\kappa$,*

$$\underset{\kappa' \sim \mathbf{lower}(\kappa)}{\mathbf{E}}[\text{vol}(\partial_{\kappa'} A)] \geq \text{vol}(A) + \mathcal{E}[1_A]/h(\kappa).$$

In particular, at least one lower neighbor of A has volume at least $\text{vol}(A) + \mathcal{E}[1_A]/h(\kappa)$.

Remark 4.7. The same proposition also holds if we consider upper neighbors, $\kappa' \sim$ **upper**(κ).

Proof. By definition,

$$\mathcal{E}[1_A] = \underset{\substack{u \sim \pi_\kappa \\ \tau \sim \text{Trans}(n)}}{\mathbf{Pr}}[u \in A, \ u^\tau \notin A] = h(\kappa) \cdot \mathbf{Pr}[u \in A, \ v \notin A], \tag{4.1}$$

where the random string $v \in \mathcal{U}_\kappa$ is defined to be u^τ conditioned on $u \neq u^\tau$. In other words, the pair (u, v) is distributed as a random pair of strings differing by a "nontrivial" color-swap. Let $c, d \in [\ell]$ denote the two colors swapped, with the convention $c \prec d$. Then if we define $\kappa_{cd} \lhd \kappa$ to be the lower neighbor of κ having one fewer d and one more c, it holds that κ_{cd} is distributed according to **lower**(κ). Finally, let $w \in \mathcal{U}_{\kappa_{cd}}$ to be the string that agrees with u, v on the unswapped coordinates, and has color c on the swapped coordinates.

It is easy to see the following: v is uniformly distributed on $\mathcal{U}_\kappa$; conditioned on c and d, the string w is uniformly distributed on $\mathcal{U}_{\kappa_{cd}}$; and, $w \prec u, w \prec v$. In light of the last of these, we may make the following deductions: If $v \in A$, then $w \in \partial_{\kappa_{cd}} A$. Furthermore, even when $v \notin A$, if $u \in A$ then $w \in \partial_{\kappa_{cd}} A$. Thus

$$\mathbf{Pr}[w \in \partial_{\kappa_{cd}} A] \geq \mathbf{Pr}[v \in A] + \mathbf{Pr}[u \in A, \ v \notin A]$$
$$\implies \quad \underset{c, d}{\mathbf{E}}[\text{vol}(\partial_{\kappa_{cd}} A)] \geq \text{vol}(A) + \mathcal{E}[1_A]/h(\kappa).$$

In this deduction, on the right we used that v is uniformly distributed on $\mathcal{U}_\kappa$ and we used Equation (4.1). On the left we used that — conditioned on c and d — the string w is uniform on $\mathcal{U}_{\kappa_{cd}}$. The proof is completed by recalling that κ_{cd} is distributed according to **lower**(κ). $\square$

We can now immediately deduce our first Kruskal–Katona Theorem, using just the log-Sobolev inequality Theorem 1.1, and Inequality (1.1):

Theorem 4.8. *For $A \subseteq \mathcal{U}_\kappa$ we have*

$$\mathbf{E}_{\kappa' \sim \mathbf{lower}(\kappa)}[\mathrm{vol}(\partial_{\kappa'} A)] \geq \mathrm{vol}(A) + \frac{1}{n} \cdot \mathrm{vol}(A) \ln(1/\mathrm{vol}(A)) \cdot \left(\sum_{i=1}^{n} \log_2(4n/\kappa_i) \right)^{-1}.$$

In particular, at least one lower shadow of A has volume at least the right-hand side. The analogous statement for upper shadows also holds.

Thus in the model case when $\mathrm{vol}(A)$ and each κ_i/n is bounded away from 0 and 1, and $\ell = O(1)$, we get that the average lower shadow of A has volume at least $\mathrm{vol}(A) + \Omega(1/n)$. Now using our KKL Theorem (Theorem 4.2) we can get a "robust" version of this statement; the volume increase is in fact on the order of $(\log n)/n$ unless there is a highly influential transposition for A:

Theorem 4.9. *Let $A \subseteq \mathcal{U}_\kappa$. Assume $\kappa_i \geq pn$ for all $i \in [\ell]$ and that $\epsilon \leq \mathrm{vol}(A) \leq 1 - \epsilon$. Then for any $\delta > 0$ we have*

$$\mathbf{E}_{\kappa' \sim \mathbf{lower}(\kappa)}[\mathrm{vol}(\partial_{\kappa'} A)] \geq \mathrm{vol}(A) + \frac{\log n}{n} \cdot \Omega\left(\frac{\epsilon \delta}{\ell \log(1/p)} \right),$$

or else there exists $\tau \in \mathrm{Trans}(n)$ with $\mathrm{Inf}_\tau[A] \geq 1/n^\delta$. The analogous statement for upper shadows also holds.

As in [50], we now give a conceptual improvement to the "or else" clause in Theorem 4.9. Let us work with upper shadows rather than lower shadows going forward. The natural example for sets A with upper-shadow expansion "only" $\Omega(1/n)$ are "dictator" sets such as $A = \{u : u_1 = \ell\}$. For such sets, all transpositions of the form $(1\ j)$ indeed have huge influence. However, it's not so natural to single out one such $(1\ j)$ as the "reason" for the small expansion; instead, we would prefer to say the reason is that A is highly "correlated" with coordinate 1. To this end, let us make a definition.

Definition 4.10. *Let $A \subseteq \mathcal{U}_\kappa$, let $j \in [n]$, and let $c \prec d$ be colors in $[\ell]$. The* correlation *of A with respect to coordinate j and colors c, d is*

$$\mathrm{corr}_{j,c,d}[A] = \mathbf{Pr}_{\boldsymbol{u} \sim \pi_\kappa}[\boldsymbol{u} \in A \mid \boldsymbol{u}_j = d] - \mathbf{Pr}_{\boldsymbol{u} \sim \pi_\kappa}[\boldsymbol{u} \in A \mid \boldsymbol{u}_j = c].$$

For simplicity, we present the following theorem without stating the most general possible settings for parameters:

Theorem 4.11. *For $n \longrightarrow \infty$, let $A \subseteq \mathcal{U}_\kappa$, with $\ell = O(1)$, $\kappa_i/n \geq \Omega(1)$ for all $i \in [\ell]$ and $\Omega(1) \leq \mathrm{vol}(A) \leq 1 - \Omega(1)$. Then*

$$\mathbf{E}_{\kappa' \sim \mathbf{upper}(\kappa)}[\mathrm{vol}(\partial_{\kappa'} A)] \geq \mathrm{vol}(A) + \Omega\left(\frac{\log n}{n} \right), \tag{4.2}$$

or else there exists $j \in [n]$ and colors $c \prec d \in [\ell]$ with $\mathrm{corr}_{j,c,d}[A] \geq 1/n^{.01}$.

Suppose that Inequality (4.2) doesn't hold. Then there must exist $\tau \in \mathrm{Trans}(n)$ such that $\mathrm{Inf}_\tau[A] \geq 1/n^{.01}$. Without loss of generality, we can assume that $\tau = (1\ 2)$. We then deduce two consequences of this, relating the volume of A and its upper shadows. Finally we will combine these to get that A is correlated to a single color change on one coordinate. The proof here has similar ideas to [50, Lemma A.6], but we reproduce it for completeness. We also introduce the notation $\mathrm{vol}_c(A) := \mathbf{Pr}_{\boldsymbol{u} \sim \pi_\kappa}[\boldsymbol{u} \in A \mid \boldsymbol{u}_1 = c]$.

Lemma 4.12. *Let $A \subseteq \mathcal{U}_\kappa$, with $\ell = O(1)$ as $n \to \infty$. Let $\epsilon \leq \kappa_c/n \leq 1 - \epsilon$ for all $c \in [\ell]$. Further suppose that*

$$\mathbf{E}_{\kappa' \sim \mathbf{upper}(\kappa)}[\mathrm{vol}(\partial_{\kappa'} A)] - \mathrm{vol}(A) \leq \eta.$$

For every $c, d \in [\ell]$ with $c \prec d$, let $\kappa_{dc} \rhd \kappa$ be the upper neighbor with one more d and one less c than κ. Then we have

$$\sum_{e \in \{c,d\}} \left(\Pr_{\boldsymbol{v} \sim \pi_{\kappa_{dc}}} [\boldsymbol{v} \in \partial_{\kappa_{dc}} A \wedge \boldsymbol{v}_1 = e] - \Pr_{\boldsymbol{u} \sim \pi_\kappa} [\boldsymbol{u} \in A \wedge \boldsymbol{u}_1 = e] \right) \le \frac{1}{\epsilon^2} \eta. \qquad (4.3)$$

Proof. Let $P_{\boldsymbol{c},\boldsymbol{d}}$ be the probability that $\kappa' \sim \mathbf{upper}(\kappa)$ is such that κ' has one more $\boldsymbol{d}$ and one less $\boldsymbol{c}$ than κ. We use the convention that $\boldsymbol{u} \sim \mathcal{U}_\kappa$ and $\boldsymbol{v}$ is sampled from an upper neighbor of $\boldsymbol{u}$. By definition,

$$\mathop{\mathbf{E}}_{\kappa' \sim \mathbf{upper}(\kappa)} [\mathrm{vol}(\partial_{\kappa'} A)] - \mathrm{vol}(A) = \sum_{\boldsymbol{c},\boldsymbol{d},e \in [\ell]} \mathrm{vol}_e(\partial_{\kappa_{\boldsymbol{dc}}} A) \Pr[\boldsymbol{v}_1 = e] P_{\boldsymbol{c},\boldsymbol{d}}$$
$$- \sum_{\boldsymbol{c},\boldsymbol{d},e \in [\ell]} \mathrm{vol}_e(A) \Pr[\boldsymbol{u}_1 = e] P_{\boldsymbol{c},\boldsymbol{d}}.$$

If we sample $\boldsymbol{u} \sim \mathcal{U}_\kappa$ conditioned on $\boldsymbol{u}_1 = e$, and change a random $\boldsymbol{c}$ to a $\boldsymbol{d}$, this distribution is uniform on $\mathcal{U}_{\kappa_{\boldsymbol{dc}}}$ conditioned on the first coordinate being e. Therefore, conditioned on $\boldsymbol{c}$ and $\boldsymbol{d}$, for $e \in [\ell]$, $\mathrm{vol}_e(A) \le \mathrm{vol}_e(\partial_{\kappa_{\boldsymbol{dc}}} A)$. We also have that for $e \ne \boldsymbol{c}, \boldsymbol{d}$, $\Pr[\boldsymbol{v}_1 = e] = \Pr[\boldsymbol{u}_1 = e]$. Putting these together,

$$\sum_{\boldsymbol{c},\boldsymbol{d} \in [\ell]} \sum_{e \in \{\boldsymbol{c},\boldsymbol{d}\}} (\mathrm{vol}_e(\partial_{\kappa_{\boldsymbol{dc}}} A) \Pr[\boldsymbol{v}_1 = e] P_{\boldsymbol{c},\boldsymbol{d}} - \mathrm{vol}_e(A) \Pr[\boldsymbol{u}_1 = e] P_{\boldsymbol{c},\boldsymbol{d}}) \le \eta.$$

In particular, every pair $c, d \in [\ell]$ satisfies

$$\sum_{e \in \{c,d\}} \left(\Pr_{\boldsymbol{v} \sim \pi_{\kappa_{dc}}} [\boldsymbol{v} \in \partial_{\kappa_{dc}} A \wedge \boldsymbol{v}_1 = e] - \Pr_{\boldsymbol{u} \sim \pi_\kappa} [\boldsymbol{u} \in A \wedge \boldsymbol{u}_1 = e] \right) P_{c,d} \le \eta.$$

Finally, we bound $P_{c,d}$. This is the probability that, for any $u \in \mathcal{U}_\kappa$ and $\boldsymbol{i}, \boldsymbol{j} \in [n]$ chosen uniformly and independently, $\{u_{\boldsymbol{i}}, u_{\boldsymbol{j}}\} = \{c, d\}$ conditioned on $u_{\boldsymbol{i}} \ne u_{\boldsymbol{j}}$. We can calculate this probability explicitly:

$$\Pr_{\boldsymbol{i},\boldsymbol{j} \sim [n]} [u_{\boldsymbol{i}} = c, u_{\boldsymbol{j}} = d \mid u_{\boldsymbol{i}} \ne u_{\boldsymbol{j}}] = \frac{\kappa_c}{n} \cdot \frac{\kappa_d}{n - \kappa_c} \cdot h(\kappa)^{-1} \ge \epsilon^2.$$

Therefore,

$$\sum_{e \in \{c,d\}} \left(\Pr_{\boldsymbol{v} \sim \pi_{\kappa_{dc}}} [\boldsymbol{v} \in \partial_{\kappa_{dc}} A \wedge \boldsymbol{v}_1 = e] - \Pr_{\boldsymbol{u} \sim \pi_\kappa} [\boldsymbol{u} \in A \wedge \boldsymbol{u}_1 = d] \right) \le \frac{1}{\epsilon^2} \eta. \qquad \square$$

Lemma 4.13. *Let $A \subseteq \mathcal{U}_\kappa$, with $\ell = O(1)$ as $n \longrightarrow \infty$. Let $\epsilon \le \kappa_c/n \le 1 - \epsilon$ for all $c \in [\ell]$. Further suppose that*

$$\mathrm{Inf}_{(1\ 2)}[A] \ge \ell^2 \gamma.$$

Then there exist $c, d \in [\ell]$ with $c \prec d$ such that the upper neighbor $\kappa_{dc} \rhd \kappa$ satisfies

$$\Pr_{\boldsymbol{v} \sim \pi_{\kappa_{dc}}} [\boldsymbol{v} \in \partial_{\kappa_{dc}} A \mid \boldsymbol{v}_i = d] - \Pr_{\boldsymbol{u} \sim \pi_\kappa} [\boldsymbol{u} \in A \mid \boldsymbol{u}_i = c] \ge \gamma \qquad (4.4)$$

for $i = 1$ or 2.

Proof. Draw $\boldsymbol{u} \sim \pi_\kappa$, and write $\boldsymbol{u} = (\boldsymbol{u}_1, \boldsymbol{u}') = (\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{w})$ with $\boldsymbol{w} \in [\ell]^{n-2}$. We have

$$\mathrm{Inf}_{(1\ 2)}[A] \ge \Pr[(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{w}) \notin A \wedge (\boldsymbol{u}_2, \boldsymbol{u}_1, \boldsymbol{w}) \in A] \ge \ell^2 \gamma.$$

Since there are ℓ^2 choices of colors for $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, it must be true that for some choice of $c, d \in [\ell]$,

$$\Pr[\boldsymbol{u}_1 = c \wedge \boldsymbol{u}_2 = d \wedge (c, d, \boldsymbol{w}) \notin A \wedge (d, c, \boldsymbol{w}) \in A] \ge \gamma.$$

We consider the case that $c \prec d$ and reach the case of $i = 1$ in the conclusion; the other case is similar.

If $(d, c, \boldsymbol{w}) \in A$ then $(d, d, \boldsymbol{w}) \in \partial_{\kappa_{dc}} A$. Therefore

$$\mathbf{Pr}[\boldsymbol{u}_1 = c \wedge \boldsymbol{u}_2 = d \wedge (c, d, \boldsymbol{w}) \notin A \wedge (d, d, \boldsymbol{w}) \in \partial_{\kappa_{dc}} A] \geq \gamma.$$

Every $\boldsymbol{u}$ in the event above also satisfies $\boldsymbol{u} \notin A \wedge (d, \boldsymbol{u}') \in \partial_{\kappa_{dc}} A$, so

$$\mathbf{Pr}[\boldsymbol{u}_1 = c \wedge \boldsymbol{u} \notin A \wedge (d, \boldsymbol{u}') \in \partial_{\kappa_{dc}} A] \geq \gamma,$$

and thus

$$\mathbf{Pr}[\boldsymbol{u} \notin A \wedge (d, \boldsymbol{u}') \in \partial_{\kappa_{dc}} A \mid \boldsymbol{u}_1 = c] \geq \frac{\gamma}{\kappa_c/n} \geq \gamma.$$

Finally, let $\boldsymbol{u}' \sim \pi_{\kappa'}$, where κ' is obtained from κ by removing one c. If $(c, \boldsymbol{u}') \in A$ then $(d, \boldsymbol{u}') \in \partial_{\kappa_{dc}} A$, and so

$$\mathop{\mathbf{Pr}}_{\boldsymbol{v}' \sim \pi_{\kappa'}} [(d, \boldsymbol{v}') \in \partial_{\kappa_{dc}} A] - \mathop{\mathbf{Pr}}_{\boldsymbol{u}' \sim \pi_{\kappa'}} [(c, \boldsymbol{u}') \in A] = \mathop{\mathbf{Pr}}_{\boldsymbol{u}' \sim \pi_{\kappa'}} [(d, \boldsymbol{u}') \in \partial_{\kappa_{dc}} A \wedge (c, \boldsymbol{u}') \notin A].$$

Now $(c, \boldsymbol{u}')$ is uniformly distributed on $\mathcal{U}_\kappa$ conditioned on $\boldsymbol{u}_1 = c$, so

$$\mathop{\mathbf{Pr}}_{\boldsymbol{u}' \sim \pi_{\kappa'}} [(d, \boldsymbol{u}') \in \partial_{\kappa_{dc}} A \wedge (c, \boldsymbol{u}') \notin A] = \mathop{\mathbf{Pr}}_{\boldsymbol{u} \sim \pi_\kappa} [\boldsymbol{u} \notin A \wedge (d, \boldsymbol{u}') \in \partial_{\kappa_{dc}} A \mid \boldsymbol{u}_1 = c] \geq \gamma,$$

and also

$$\mathop{\mathbf{Pr}}_{\boldsymbol{u}' \sim \pi_{\kappa'}} [(c, \boldsymbol{u}') \in A] = \mathop{\mathbf{Pr}}_{\boldsymbol{u} \sim \pi_\kappa} [\boldsymbol{u} \in A \mid \boldsymbol{u}_i = c].$$

Meanwhile, $(d, \boldsymbol{u}')$ is uniformly distributed on $\mathcal{U}_{\kappa_{dc}}$ conditioned on $\boldsymbol{u}_1 = d$, and so

$$\mathop{\mathbf{Pr}}_{\boldsymbol{v}' \sim \pi_{\kappa'}} [(d, \boldsymbol{v}') \in \partial_{\kappa_{dc}} A] = \mathop{\mathbf{Pr}}_{\boldsymbol{v} \sim \pi_{\kappa_{dc}}} [\boldsymbol{v} \in \partial_{\kappa_{dc}} A \mid \boldsymbol{v}_i = d]. \qquad \square$$

Proof of Theorem 4.11. Apply Theorem 4.9 with $\delta = .02$, and assume that Inequality (4.2) does not hold. The theorem shows that $\mathrm{Inf}_\tau[A] \geq 1/n^{.02}$ for some transposition τ, which without loss of generality is $\tau = (1\ 2)$. We can therefore apply Theorem 4.13 (with $\gamma = 1/(\ell^2 n^{.02})$), obtaining two colors $c \prec d$, which without loss of generality satisfy the conclusion of the lemma for $i = 1$:

$$\mathrm{vol}_d(\partial_{\kappa_{dc}} A) \geq \frac{1}{\ell^2 n^{.02}} + \mathrm{vol}_c(A). \tag{4.5}$$

We will show that A is correlated to the first coordinate and the colors c and d.

Since Inequality (4.2) does not hold, we can apply Theorem 4.12 (with $\eta = O(\log n/n)$) to obtain

$$\mathbf{Pr}[\boldsymbol{v}_1 = d]\mathrm{vol}_d(\partial_{\kappa_{dc}} A) + \mathbf{Pr}[\boldsymbol{v}_1 = c]\mathrm{vol}_c(\partial_{\kappa_{dc}} A)$$
$$- \mathbf{Pr}[\boldsymbol{u}_1 = d]\mathrm{vol}_d(A) - \mathbf{Pr}[\boldsymbol{u}_1 = c]\mathrm{vol}_c(A) \leq \frac{1}{\epsilon^2}\eta.$$

Combining Inequality (4.5) with $\mathbf{Pr}[\boldsymbol{v}_1 = d] \geq \epsilon$ and $\mathrm{vol}_c(\partial_{\kappa_{dc}} A) \geq \mathrm{vol}_c(A)$, we deduce

$$\mathbf{Pr}[\boldsymbol{u}_1 = d]\mathrm{vol}_d(A) \geq \mathbf{Pr}[\boldsymbol{v}_1 = d]\mathrm{vol}_d(\partial_{\kappa_{dc}} A)$$
$$+ \mathbf{Pr}[\boldsymbol{v}_1 = c]\mathrm{vol}_c(\partial_{\kappa_{dc}} A) - \mathbf{Pr}[\boldsymbol{u}_1 = c]\mathrm{vol}_c(A) - \frac{1}{\epsilon^2}\eta$$
$$\geq \frac{\epsilon}{\ell^2 n^{.02}} + (\mathbf{Pr}[\boldsymbol{v}_1 = d] + \mathbf{Pr}[\boldsymbol{v}_1 = c] - \mathbf{Pr}[\boldsymbol{u}_1 = c])\mathrm{vol}_c(A) - \frac{1}{\epsilon^2}\eta$$
$$\geq \mathbf{Pr}[\boldsymbol{u}_1 = d]\mathrm{vol}_c(A) + \frac{\epsilon}{\ell^2 n^{.02}} - \frac{1}{\epsilon^2}\eta.$$

Dividing by $\mathbf{Pr}[\boldsymbol{u}_1 = d]$, we conclude that

$$\mathrm{corr}_{1,c,d}[A] = \mathrm{vol}_d(A) - \mathrm{vol}_c(A) \geq \frac{\epsilon}{\ell^2 n^{.02}} - \frac{1}{\epsilon^2}\eta.$$

Since $\eta = O(\log n/n)$, $\ell = O(1)$, and $\epsilon = \Omega(1)$, for large enough n, $\mathrm{corr}_{1,c,d}[A] \geq \frac{1}{n^{.01}}$. $\qquad \square$

4.2 Harmonic analysis on the symmetric group, and Friedgut on the multislice

In this section we will recap some aspects of harmonic analysis on the symmetric group and on the multislice, paying particular attention to the notion of the "low-degree" components of a function. For more details, see e.g. [12].

First, we briefly discuss partitions. A *partition* λ of n is a nonincreasing sequence of positive integers summing to n. (Equivalently, it is a sorted histogram κ; i.e., one with $\kappa_1 \geq \kappa_2 \geq \cdots \geq \kappa_\ell$.) We write $\lambda \vdash n$. Sometimes we extend λ into an infinite sequence, by padding it with infinitely many zeroes. We say that λ *dominates* or *majorizes* μ, written $\lambda \trianglerighteq \mu$, if for all $i \geq 1$ the inequality $\lambda_1 + \cdots + \lambda_i \geq \mu_1 + \cdots + \mu_i$ holds.

Though we will eventually be interested in functions on multislices, we begin by studying the larger vector space V of functions $f \colon S_n \to \mathbb{R}$ on the symmetric group. Note that we can naturally extend the operators K, L, H_t to this space V. The partitions λ of n index the *irreducible representations* of the symmetric group S_n. In particular this means that V has an orthogonal decomposition

$$V = \{f \mid f \colon S_n \to \mathbb{R}\} = \bigoplus_{\lambda \vdash n} V^\lambda,$$

where the *isotypic component* V^λ corresponds to the irreducible representations λ (counted with multiplicity). In analogy with the level/degree decomposition on the Boolean cube, we denote the orthogonal projection of f onto V^λ by $f^{=\lambda}$.

One utility of this decomposition is that V^λ is an eigenspace for the operator K, with eigenvalue equal to $\widehat{\chi}_\lambda(\tau)$, the normalized character evaluated at a(ny) transposition $\tau \in \mathrm{Trans}(n)$. Frobenius [30] determined an explicit formula for these character values:

$$\mathrm{K}f = \sum_{\lambda \vdash n} c_\lambda f^{=\lambda}, \text{ where } c_\lambda = \frac{1}{n(n-1)} \sum_{i=1}^{\ell} [\lambda_i^2 - (2i-1)\lambda_i]. \tag{4.6}$$

See [15, Cor. 1 & Lem. 7] for an explicit proof of the above. Immediate consequences of the above formula are the following:

$$\mathrm{L}f = \sum_{\lambda} d_\lambda f^{=\lambda}, \text{ where } d_\lambda = 1 - c_\lambda; \tag{4.7}$$

$$\mathrm{H}_t f = e^{-t\mathrm{L}} f = \sum_{\lambda} e^{-td_\lambda} f^{=\lambda}. \tag{4.8}$$

From Equation (4.8) we can see that H_t is an invertible operator for all $t \geq 0$, and that it is natural to write $\mathrm{H}_t^{-1} = \mathrm{H}_{-t}$.

An important feature of the formula for c_λ (and hence d_λ) is its relation to majorization order. The following simple calculation was observed in, e.g., [15, Lem. 10]:

Lemma 4.14. *If* $\lambda \trianglerighteq \mu$ *then* $c_\lambda > c_\mu$ *and hence* $d_\lambda < d_\mu$.

From this we may immediately determine the spectral gap[4] of the transposition chain on S_n, which is achieved at $\lambda = (n-1, 1)$.

Corollary 4.15. *The minimal nontrivial eigenvalue of operator* L *on* V *is* $\frac{2}{n-1}$.

As we explain shortly, given λ we will be particularly interested in the parameter $k = n - \lambda_1$. An immediate consequence of Theorem 4.14 is that we can determine the minimal and maximal value of d_λ in terms of this parameter k. We skip the straightforward calculations (most of which appear in [12, Ch. 3D, Lem. 2]):

---

[4]Although the spectral gap is usually notated λ_1, we avoid this notation here due to confusion with the standard notation λ for partitions.

Corollary 4.16. *For $\lambda \vdash n$ and $\lambda_1 = n - k$, we have*

$$\frac{k}{n-1} \leq d_\lambda \leq \frac{2k}{n-1}.$$

The upper bound has equality if $\lambda = (n-k, 1, \ldots, 1)$. Further, if $k \leq n/2$ we have

$$d_\lambda \geq \left(1 - \frac{k-1}{n}\right)\frac{2k}{n-1},$$

with equality if $\lambda = (n-k, k)$.

Why consider the parameter $k = n - \lambda_1$? It turns out that this parameter is very much analogous to "Fourier degree" for functions on the Boolean cube, as the following result (proved in, e.g., [20, Thm. 7]) shows:

Theorem 4.17. *Let $f\colon S_n \to \mathbb{R}$ be a nonzero function. The* degree *of f is the least $k \in \mathbb{N}$ such that f can be represented as a linear combination of "k-juntas" (meaning functions g such that $g(\pi)$ depends only on some k values $\pi(j_1), \ldots, \pi(j_k)$). It is also equal to the least k such that $f^{=\lambda} = 0$ for all λ with $n - \lambda_1 > k$.*

We now provide two simple applications of Theorem 4.16 concerning functions of bounded degree:

Lemma 4.18. *If $f\colon S_n \to \mathbb{R}$ has degree at most k, then $\mathrm{Inf}[f] \leq kn\|f\|_2^2$.*

Proof. Using Equation (4.7) and Theorem 4.16,

$$\mathrm{Inf}[f] = \binom{n}{2}\langle f, \mathrm{L}f\rangle = \binom{n}{2}\sum_{\lambda \trianglerighteq (n-k,1,\ldots,1)} d_\lambda \|f^{=\lambda}\|_2^2 \leq \sum_{\lambda \trianglerighteq \kappa} kn\|f^{=\lambda}\|_2^2 = kn\|f\|_2^2. \qquad \square$$

Lemma 4.19. *If $f\colon S_n \to \mathbb{R}$ has degree at most k, then for all $t \geq 0$,*

$$\|\mathrm{H}_t f\|_2 \geq e^{-2kt/(n-1)}\|f\|_2,$$

and for all $t \leq 0$,

$$\|\mathrm{H}_t f\|_2 \leq e^{-2kt/(n-1)}\|f\|_2.$$

Proof. Theorem 4.16 shows that for $t \geq 0$,

$$\|\mathrm{H}_t f\|_2^2 = \sum_{\lambda \trianglerighteq (n-k,1,\ldots,1)} e^{-2td_\lambda}\|f^{=\lambda}\|_2^2 \geq e^{-4kt/(n-1)}\|f\|_2^2.$$

The bound for $t \leq 0$ follows in a similar fashion. $\qquad \square$

We now move on to discussing functions on the multislice. Let $\kappa \in \mathbb{N}_+^\ell$ be a histogram of size n. By relabeling the colors we may assume $\kappa_1 \geq \kappa_2 \geq \cdots \geq \kappa_\ell$ and hence that $\kappa \vdash n$. Let us denote by $u_0 \in [\ell]^n$ the following *canonical string*:

$$u_0 = \underbrace{11\cdots1}_{\substack{\kappa_1 \\ \text{times}}}\underbrace{22\cdots2}_{\substack{\kappa_2 \\ \text{times}}}\cdots\underbrace{\ell\ell\cdots\ell}_{\substack{\kappa_\ell \\ \text{times}}}.$$

Note that as π runs over all permutations in S_n, the string u_0^π runs over all strings in $\mathcal{U}_\kappa$, with equal multiplicity $\kappa_1!\kappa_2!\cdots\kappa_\ell!$. In this way, each function f in the permutation module M^κ (i.e., the multislice $\mathcal{U}_\kappa$ considered as a representation of S_n) can be naturally identified with a "pullback" function $\overline{f} \in V$, via $\overline{f}(\pi) = f(u_0^\pi)$. Conversely, the functions $g \in V$ that correspond to functions on the multislice $\mathcal{U}_\kappa$ are precisely those that are invariant to the action of the *Young subgroup* $S_{\kappa_1} \times \cdots \times S_{\kappa_\ell}$. Classical results in the

representation theory of the symmetric group show that this subspace has the following isotypic decomposition:

$$M^\kappa = \bigoplus_{\lambda \trianglerighteq \kappa} V_\kappa^\lambda, \tag{4.9}$$

where V_κ^λ is (isomorphic to) a nonzero subspace of V^λ (specifically, V_κ^λ consists of $K_{\lambda\kappa}$ copies of the irrep associated to λ, where $K_{\lambda\kappa}$ is the *Kostka number*. Since this decomposition always includes $V_\kappa^{(n-1,1)} \leq V^{(n-1,1)}$ (unless $\kappa = (n)$), we conclude:

Corollary 4.20. *The minimal nontrivial eigenvalue of the operator* L *on any* M^κ *(for* $\kappa \neq (n)$*) is also* $\frac{2}{n-1}$*.*

We now define the notion of "degree" for functions on multislices:

Definition 4.21. *Let* $f : \mathcal{U}_\kappa \to \mathbb{R}$ *be a nonzero function. The* degree *of* f *is the least* $k \in \mathbb{N}$ *such that* f *can be represented as a linear combination of "k-juntas" (functions* g *such that* $g(u)$ *depends only on some* k *values* $u_{j_1}, \ldots, u_{j_k}$*). It is also equal to the least* k *such that* $f^{=\lambda} = 0$ *for all* λ *with* $n - \lambda_1 > k$ *(in* f*'s decomposition as in Equation* (4.9)*).*

Claim 4.22. *The two definitions of "degree" above are indeed the same.*

Proof. If $g \in M^\kappa$ is a k-junta, it's easy to see that its pullback $\overline{g} : S_n \to \mathbb{R}$ is a k-junta. Thus if $f \in M^\kappa$ is a linear combination of k-juntas, so too is its pullback $\overline{f} : S_n \to \mathbb{R}$. From Theorem 4.17 we get that $\overline{f}^{=\lambda} = 0$ for all λ with $n - \lambda_1 > k$ and so the same is true of $f^{=\lambda}$.

In the other direction, if $f^{=\lambda} = 0$ for all λ with $n - \lambda_1 > k$, the same is true of $\overline{f}^{=\lambda}$, and hence $\overline{f}$ is a linear combination of k-juntas (by Theorem 4.17 again). We need to show that f is also a linear combination of k-juntas. By linearity, it suffices to assume that $\overline{f}$ is itself a k-junta; indeed, it further suffices to assume $\overline{f}$ is of the form $\overline{f}(\pi) = \mathbb{1}[\pi(i_1) = j_1, \ldots, \pi(i_k) = j_k]$ for some coordinates $i_1, \ldots, i_k, j_1, \ldots, j_k \in [n]$. By definition we have $f(v) = \overline{f}(\pi)$ for any $\pi \in S_n$ such that $u_0^\pi = v$. In particular, it equals the average of $\overline{f}(\pi)$ over all such π; i.e.,

$$f(v) = \operatorname*{\mathbf{E}}_{\substack{\boldsymbol{\pi} \in S_n \\ u_0^\pi = v}} [\overline{f}(\boldsymbol{\pi})] = \operatorname*{\mathbf{Pr}}_{\substack{\boldsymbol{\pi} \in S_n \\ u_0^\pi = v}} [\pi(i_1) = j_1, \ldots, \pi(i_k) = j_k] = \mathbb{1}[v_{i_1} = j_1, \ldots, v_{i_k} = j_k].$$

This means that f is indeed a k-junta on $\mathcal{U}_\kappa$. $\qquad\square$

An immediate consequence is the following:

Corollary 4.23. *Theorems 4.18 and 4.19 hold equally well for functions* $f \in M^\kappa$ *of degree at most* k*.*

Finally, we relate the main theorem in our paper to the comparison of norms for low-degree functions on the multislice:

Lemma 4.24. *Fix a histogram* $\kappa \in \mathbb{N}_+^\ell$ *and let* $p = \min_i \kappa_i / n$*. Suppose that* $f \in M^\kappa$ *has degree* k*. Then for all finite* $q \geq 2$*:*

$$\|f\|_q \leq (q-1)^{\Theta(k \log(1/p))} \|f\|_2,$$
$$\|f\|_2 \leq (q-1)^{\Theta(k \log(1/p))} \|f\|_{q'},$$

where q' *is given by* $1/q + 1/q' = 1$*.*

Proof. Theorem 1.1 shows that $\varrho_\kappa^{-1} = \Theta(n \log(1/p))$. Theorem 2.1 therefore shows that $\|\mathrm{H}_t g\|_q \leq \|g\|_2$ and $\|\mathrm{H}_t g\|_2 \leq \|g\|_{q'}$ for all $g \in M^\kappa$, where

$$t = \frac{\ln(q-1)}{2\varrho_\kappa} = \Theta(\ln(q-1) \cdot n \log(1/p)).$$

Applying this to $g = \mathrm{H}_t^{-1} f$ (which has the same degree as f) and using Theorem 4.19 (and Theorem 4.23), we deduce

$$\|f\|_q \leq \|\mathrm{H}_{-t} f\|_2 \leq e^{2tk/(n-1)} \|f\|_2 = (q-1)^{\Theta(k \log(1/p))} \|f\|_2,$$

and similarly for the second claimed inequality. $\qquad\square$

We end this section by providing an analogue of Friedgut's Junta Theorem [29] for functions on multislices:

Theorem 4.25. *Let* $f: \mathcal{U}_\kappa \to \{0,1\}$ *be such that* $\mathrm{Inf}[f] \leq Kn$. *Write* $p_i = \kappa_i / n$. *Then for every* $\epsilon > 0$ *there exists* $h: \mathcal{U}_\kappa \to \{0,1\}$ *depending on at most* $\left(\frac{1}{p_1 p_2 \cdots p_\ell} \right)^{O(K/\epsilon)}$ *coordinates such that* $\mathbf{Pr}_{\boldsymbol{u} \sim \pi_\kappa}[f(\boldsymbol{u}) \neq h(\boldsymbol{u})] \leq \epsilon$.

The proof is essentially identical to Wimmer's proof [61, Sec. VI] of the analogous theorem for functions on the Boolean slice (i.e., the $\ell = 2$ case of the above). After replacing Wimmer's pullback function (notated f^g therein) with our generalization $\overline{f}$, it only remains to substitute in our main log-Sobolev inequality for the multislice $\mathcal{U}_\kappa$.

4.3 Nisan–Szegedy Theorem on the multislice

The Nisan–Szegedy Theorem says that a degree-k Boolean-valued function on the Hamming cube is a $k2^k$-junta. (We remark that the smallest quantity $\gamma_2(k)$ that can replace $k2^k$ here is now known [7] to satisfy $3 \cdot 2^{k-1} - 2 \leq \gamma_2(k) < 22 \cdot 2^k$.) In [25], an analogous result for functions on Hamming slices was shown; they conjectured a similar result for functions on multislices. We resolve this conjecture, following the structure of their proof. This proof structure involves proving three successively stronger versions of the desired theorem.

The first version pertains only to functions on balanced multislices; it was originally established for Hamming slices in [27]:

Theorem 4.26. *Fix* $\ell \geq 2$, *assume* n *is a multiple of* ℓ, *and let* $\kappa = (n/\ell, \ldots, n/\ell) \in \mathbb{N}_+^\ell$. *If* $f: \mathcal{U}_\kappa \to \{0,1\}$ *has degree at most* k, *then* f *is an* $\ell^{O(k)}$*-junta.*

To prove Theorem 4.26, we first use our hypercontractivity result to establish the following analogue of [25, Lem. 3.1]:

Lemma 4.27. *In the setting of Theorem 4.26, every nonzero influence* $\mathrm{Inf}_\tau[f]$ *is at least* $\ell^{-O(k)}$ *(where the* $O(\cdot)$ *hides a universal constant).*

Proof. Since f has degree at most k, the same is true of $\mathrm{L}_\tau f$. Thus Theorem 4.24 shows (taking, say, $q = 4$) that

$$\mathrm{Inf}_\tau[f] = \tfrac{1}{2} \|\mathrm{L}_\tau f\|_2^2 \leq \ell^{O(k)} \cdot \|\mathrm{L}_\tau f\|_{4/3}^2.$$

Since f is Boolean-valued, $\mathrm{L}_\tau f$ takes values in $\{0, \pm 1\}$, and so

$$\|\mathrm{L}_\tau f\|_{4/3}^2 = (\|\mathrm{L}_\tau f\|_2^2)^{3/2} = (2\mathrm{Inf}_\tau[f])^{3/2}.$$

Combining these yields that either $\mathrm{Inf}_\tau[f] = 0$ or else $\mathrm{Inf}_\tau[f]^{-1/2} \leq \ell^{O(k)}$, as needed. $\quad\square$

With Theorem 4.27 in hand (as well as Theorem 4.18) follows exactly as in [25, Sec. 3.1].

The same argument works as long as $\min_i\{\kappa_i\} = \Omega(n)$. The second, stronger version of Nisan–Szegedy for multislices shows that in fact, it suffices to assume only that $\min_i\{\kappa_i\} \geq \ell^{O(k)}$.

Theorem 4.28. *There are universal constants* $C \geq C'$ *such that the following holds. For all* $k \in \mathbb{N}_+$ *and all* $\kappa \in \mathbb{N}_+^\ell$ *with* $\min_i\{\kappa_i\} \geq \ell^{Ck}$, *if* $f: \mathcal{U}_\kappa \to \{0,1\}$ *has degree at most* k, *then* f *is an* $\ell^{C'k}$*-junta.*

Roughly speaking, $C = 3C'$, where C' is the constant hidden in the $O(\cdot)$ of Theorem 4.26. The proof of Theorem 4.28 exactly follows the argumentation in [25, Sec. 3.2]. Essentially, starting from Theorem 4.26, they show that the truth of the statement is preserved whenever one of the quantities κ_i is incremented.

Before stating our third Nisan–Szegedy variant, let us extend the definition of $\gamma_2(k)$; we'll define $\gamma_\ell(k)$ to be the least integer such that the following statement is true:

> Every degree-k Boolean-valued function $f : [\ell]^n \to \{0, 1\}$ on the "ℓ-multicube"
> is a $\gamma_\ell(k)$-junta.

Here we say that $f : [\ell]^n \to \mathbb{R}$ has degree at most k if it is a linear combination of k-juntas (as usual for functions on product spaces, see [49, Def. 8.32]). As remarked at the end of [25], it's easy to show that $\ell^{k-1} \leq \gamma_\ell(k) \leq \gamma_2(\lceil \log_2 \ell \rceil k)$. When ℓ is a power of 2, this upper bound is at most $22 \cdot \ell^k$, very close to the lower bound; in general we have $\gamma_\ell(k) < 22 \cdot \ell^{2k}$.

Our third and final Nisan–Szegedy Theorem for the multislice improves the junta size in Theorem 4.28 to $\gamma_\ell(k)$, which is optimal (since the analogue of [25, Lem. 3.10] equally holds in our setting). We do not know, however, the weakest lower bound we can assume on $\min_i\{\kappa_i\}$.

Theorem 4.29. *There is a universal constant C such that the following holds. For all $k \in \mathbb{N}_+$ and all $\kappa \in \mathbb{N}_+^\ell$ with $\min_i\{\kappa_i\} \geq \ell^{Ck}$, if $f : \mathcal{U}_\kappa \to \{0, 1\}$ has degree at most k, then f is an $\gamma_\ell(k)$-junta.*

The way we prove this departs somewhat from the polynomials-based proof in [25, Lem. 3.9].

Proof. From Theorem 4.28 we know that f is an $\ell^{C'k}$-junta, where $C' \leq C$. Without loss of generality, say that $f(u)$ depends only on coordinates $u_1, \ldots, u_J$, where $J \leq \ell^{C'k}$. Note that as we vary $u \in \mathcal{U}_\kappa$, we see all ℓ^J possibilities for the substring $(u_1, \ldots, u_J)$; this is because $\min_i\{\kappa_i\} \geq J$. As a consequence, we can define a function $g : [\ell]^J \to \{0, 1\}$ by

$$g(u_1, \ldots, u_J) = \mathop{\mathbf{E}}_{\substack{u_{J+1}, \ldots, u_n \in [\ell] : \\ u \in \mathcal{U}_\kappa}} [f(u)].$$

Since f has degree k, it is a linear combination of k-juntas, and in particular, a linear combination of functions of the form $\mathbb{1}[u_{i_1} = c_1, \ldots, u_{i_k} = c_k]$. We show below that

$$h_{i,c} := \mathop{\mathbf{E}}_{\substack{u_{J+1}, \ldots, u_n \in [\ell] : \\ u \in \mathcal{U}_\kappa}} \left[\mathbb{1}[u_{i_1} = c_1, \ldots, u_{i_k} = c_k] \right]$$

is a degree k function, and so g has degree k. Thus g is (by definition) a $\gamma_\ell(k)$-junta, and hence so is f.

It remains to show that $h_{i,c}$ has degree k. Suppose first that $i_1, \ldots, i_k > L$. Let $d = (\#_1 c, \ldots, \#_\ell c)$ be the histogram of $c_1, \ldots, c_k$, and let $w = (\#_1 u_{\leq L}, \ldots, \#_\ell u_{\leq L})$ be the histogram of $u_1, \ldots, u_L$. The reader can verify that

$$h_{i,c} = \frac{(\kappa_1 - w_1)^{\underline{d_1}} \ldots (\kappa_\ell - w_\ell)^{\underline{d_\ell}}}{(n - L)^{\underline{k}}},$$

where $a^{\underline{b}} = a(a-1) \ldots (a - b + 1)$. Since $d_1 + \cdots + d_\ell = k$, this is a degree k function.

When some of the indices $i_1, \ldots, i_k$ are in $[L]$, we have to modify the argument slightly. Suppose that $i_1, \ldots, i_r \in [L]$ and $i_{r+1}, \ldots, i_k \notin [L]$. Redefine d to capture the histogram of $c_{r+1}, \ldots, c_k$. The reader can verify that

$$h_{i,c} = \mathbb{1}[u_{i_1} = c_1, \ldots, u_{i_r} = c_r] \times \frac{(\kappa_1 - w_1)^{\underline{d_1}} \ldots (\kappa_\ell - w_\ell)^{\underline{d_\ell}}}{(n - L)^{\underline{k-r}}},$$

which has degree k since $d_1 + \cdots + d_\ell = k - r$. $\qquad\square$

References

[1] Boaz Barak, Pravesh Kothari, and David Steurer, *Small-set expansion in shortcode graph and the 2-to-2 conjecture*, Tech. Report arXiv:1804.08662, arXiv, 2018. MR3899803

[2] Arthur Jay Bernstein, *Maximally connected arrays on the n-cube*, SIAM J. Appl. Math. **15** (1967), 1485–1489. MR0223260

[3] Aline Bonami, *Étude des coefficients Fourier des fonctions de $L^p(G)$*, Annales de l'Institut Fourier **20** (1970), no. 2, 335–402. MR0283496

[4] Raphaël Bouyrie, *An unified approach to the Junta theorem for discrete and continuous models*, Tech. Report arXiv:1702.00753, arXiv, 2017.

[5] Raphaël Bouyrie, *On quantitative noise stability and influences for discrete and continuous models*, Combin. Probab. Comput. **27** (2018), no. 3, 334–357. MR3788164

[6] Sourav Chatterjee, Jason Fulman, and Adrian Röllin, *Exponential approximation by Stein's method and spectral graph theory*, ALEA Lat. Am. J. Probab. Math. Stat. **8** (2011), 197–223. MR2802856

[7] John Chiarelli, Pooya Hatami, and Michael Saks, *Tight bound on the number of relevant variables in a bounded degree Boolean function*, Tech. Report arXiv:1801.08564, arXiv, 2018.

[8] George F. Clements, *A generalization of the Kruskal-Katona theorem*, J. Combin. Theory Ser. A **37** (1984), no. 1, 91–97. MR749518

[9] George F. Clements, *Another generalization of the Kruskal–Katona theorem*, J. Combin. Theory Ser. A **68** (1994), no. 1, 239–245. MR1295796

[10] George F. Clements, *Yet another generalization of the Kruskal–Katona theorem*, Discrete Math. **184** (1998), no. 1-3, 61–70. MR1609351

[11] Pat Devlin and Jeff Kahn, *On "stability" in the Erdős-Ko-Rado theorem*, SIAM J. Discrete Math. **30** (2016), no. 2, 1283–1289. MR3513857

[12] Persi Diaconis, *Group representations in probability and statistics*, Institute of Mathematical Statistics Lecture Notes—Monograph Series, vol. 11, Institute of Mathematical Statistics, Hayward, CA, 1988. MR964069

[13] Persi Diaconis and Susan Holmes, *Random walks on trees and matchings*, Electron. J. Probab. **7** (2002), no. 6, 17. MR1887626

[14] Persi Diaconis and Laurent Saloff-Coste, *Logarithmic Sobolev inequalities for finite Markov chains*, Annals of Applied Probability **6** (1996), no. 3, 695–750. MR1410112

[15] Persi Diaconis and Mehrdad Shahshahani, *Generating a random permutation with random transpositions*, Z. Wahrsch. Verw. Gebiete **57** (1981), no. 2, 159–179. MR0626813

[16] Persi Diaconis and Mehrdad Shahshahani, *Time to reach stationarity in the Bernoulli–Laplace diffusion model*, SIAM Journal on Mathematical Analysis **18** (1987), no. 1, 208–218. MR0871832

[17] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra, *On non-optimally expanding sets in Grassmann graphs*, Proceedings of the 50th Annual ACM Symposium on Theory of Computing, 2018, pp. 940–951. MR3826307

[18] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra, *Towards a proof of the 2-to-1 Games Conjecture?*, Proceedings of the 50th Annual ACM Symposium on Theory of Computing, 2018, pp. 376–389. MR3826261

[19] Paul Ehrenfest and Tatiana Ehrenfest, *Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem*, Physikalische Zeitschrift **8** (1907), no. 9, 311–314.

[20] David Ellis, Ehud Friedgut, and Haran Pilpel, *Intersecting families of permutations*, J. Amer. Math. Soc. **24** (2011), no. 3, 649–682. MR2784326

[21] David Ellis, Nathan Keller, and Noam Lifshitz, *On a biased edge isoperimetric inequality for the discrete cube*, Tech. Report arXiv:1702.01675, arXiv, 2017. MR3884706

[22] David Ellis, Nathan Keller, and Noam Lifshitz, *On the structure of subsets of the discrete cube with small edge boundary*, Discrete Analysis **9** (2018), 1–29. MR3819050

[23] Yuval Filmus, *Friedgut–Kalai–Naor theorem for slices of the Boolean cube*, Chic. J. Theoret. Comput. Sci. (2016), Art. 14, 17. MR3578425

[24] Yuval Filmus, *An orthogonal basis for functions over a slice of the Boolean hypercube*, Electron. J. Combin. **23** (2016), no. 1, Paper 1.23, 27. MR3484728

[25] Yuval Filmus and Ferdinand Ihringer, *Boolean constant degree functions on the slice are juntas*, Tech. Report arXiv:1801.06338, arXiv, 2018. MR3990024

[26] Yuval Filmus and Ferdinand Ihringer, *Boolean degree 1 functions on some classical association schemes*, Tech. Report arXiv:1801.06034, arXiv, 2018. MR3874601

[27] Yuval Filmus, Guy Kindler, Elchanan Mossel, and Karl Wimmer, *Invariance principle on the slice*, Transactions on Computation Theory **10** (2018), no. 3, 11. MR3831003

[28] Yuval Filmus and Elchanan Mossel, *Harmonicity and invariance on slices of the Boolean cube*, Proceedings of the 31st Annual Computational Complexity Conference, 2016, pp. Art. No. 16, 13. MR3540817

[29] Ehud Friedgut, *Boolean functions with low average sensitivity depend on few coordinates*, Combinatorica **18** (1998), no. 1, 27–36. MR1645642

[30] Ferdinand Georg Frobenius, *Über die Charaktere der symmetrischen Gruppe*, Sitzungsberichte der Königlich preussischen Akademie der Wissenschaften zu Berlin (1900), 516–534.

[31] Zakkula Govindarajulu, *The first two moments of the reciprocal of the positive hypergeometric variable*, Sankhyā Ser. B **26** (1964), 217–236. MR0184315

[32] Leonard Gross, *Logarithmic Sobolev inequalities*, American Journal of Mathematics **97** (1975), no. 4, 1061–1083. MR0420249

[33] Lawrence H. Harper, *Optimal assignments of numbers to vertices*, J. Soc. Indust. Appl. Math. **12** (1964), 131–135. MR0162737

[34] Sergiu Hart, *A note on the edges of the n-cube*, Discrete Math. **14** (1976), no. 2, 157–163. MR0396293

[35] Jeff Kahn, Gil Kalai, and Nathan Linial, *The influence of variables on Boolean functions*, Proceedings of the 29th Annual IEEE Symposium on Foundations of Computer Science, 1988, pp. 68–80.

[36] Gyula Katona, *A theorem of finite sets*, (1968), 187–207. MR0290982

[37] Subhash Khot, Dor Minzer, Dana Moshkovitz, and Muli Safra, *Small set expansion in the Johnson graph*, Tech. Report TR18-078, Electronic Colloquium on Computational Complexity, 2018.

[38] Subhash Khot, Dor Minzer, and Muli Safra, *On independent sets, 2-to-2 games, and Grassmann graphs*, Proceedings of the 49th Annual ACM Symposium on Theory of Computing, 2017, pp. 576–589. MR3678212

[39] Subhash Khot, Dor Minzer, and Muli Safra, *Pseudorandom sets in Grassmann graph have near-perfect expansion*, Tech. Report TR18-006, Electronic Colloquium on Computational Complexity, 2018. MR3899624

[40] Joseph B. Kruskal, *The number of simplices in a complex*, Mathematical optimization techniques, Univ. of California Press, Berkeley, Calif., 1963, pp. 251–278. MR0154827

[41] Michel Ledoux, *Concentration of measure and logarithmic Sobolev inequalities.*, Séminaire de Probabilités XXXIII, Springer, 1999, pp. 120–216. MR1767995

[42] Tzong-Yau Lee and Horng-Tzer Yau, *Logarithmic Sobolev inequality for some models of random walks*, Annals of Probability **26** (1998), no. 4, 1855–1873. MR1675008

[43] John H. Lindsey II, *Assignment of numbers to vertices*, Amer. Math. Monthly **71** (1964), 508–516. MR0168489

[44] Lászlo Lovász and Ravi Kannan, *Faster mixing via average conductance*, Proceedings of the 31st Annual ACM Symposium on Theory of Computing, 1999, pp. 282–287. MR1798047

[45] Patrick Moran, *Random processes in genetics*, Mathematical Proceedings of the Cambridge Philosophical Society **54** (1958), no. 1, 60–71. MR0127989

[46] Dana Moshkovitz, *Direct product testing with nearly identical sets*, Tech. Report TR14-182, Electronic Colloquium on Computational Complexity, 2014.

[47] Edward Nelson, *The free Markoff field*, Journal of Functional Analysis **12** (1973), 211–227. MR0343816

[48] Noam Nisan and Mario Szegedy, *On the degree of Boolean functions as real polynomials*, Computational Complexity **4** (1994), no. 4, 301–313. MR1313531

[49] Ryan O'Donnell, *Analysis of boolean functions*, Cambridge University Press, 2014. MR3443800

[50] Ryan O'Donnell and Karl Wimmer, *KKL, Kruskal-Katona, and monotone nets*, SIAM Journal on Computing **42** (2013), no. 6, 2375–2399. MR3143850

[51] Ryan O'Donnell and Karl Wimmer, *Sharpness of KKL on Schreier graphs*, Electronic Communications in Probability **18** (2013), 1–12. MR3037216

[52] Jean Piaget and Barbel Inhelder, *The origin of the idea of chance in children*, The Norton Library, 1976.

[53] D. H. J. Polymath, *A new proof of the density Hales-Jewett theorem*, Annals of Mathematics **175** (2012), no. 3, 1283–1327. MR2912706

[54] Prasad Raghavendra and David Steurer, *Graph expansion and the Unique Games Conjecture*, Proceedings of the 42nd Annual ACM Symposium on Theory of Computing, 2010, pp. 755–764. MR2743325

[55] Justin Salez, *A sharp log-Sobolev inequality for the multislice*, Ann. H. Lebesgue **4** (2021), 1143–1161. MR4353960

[56] Fabio Scarabotti, *Time to reach stationarity in the Bernoulli–Laplace diffusion model with many urns*, Adv. in Appl. Math. **18** (1997), no. 3, 351–371. MR1436486

[57] Fabio Scarabotti and Filippo Tolli, *Harmonic analysis on a finite homogeneous space II: the Gelfand–Tsetlin decomposition*, Forum Math. **22** (2010), no. 5, 879–911. MR2719761

[58] Marcel-Paul Schützenberger, *A characteristic property of certain polynomials of E. F. moore and C. E. shannon*, Quarterly Progress Report, Research Laboratory of Electronics (RLE) **055.IX** (1959), 117–118.

[59] Michel Talagrand, *On Russo's approximate zero-one law*, Annals of Probability **22** (1994), no. 3, 1576–1587. MR1303654

[60] Karl Wimmer, *Fourier methods and combinatorics in learning theory*, Ph.D. thesis, Carnegie Mellon University, 2009. MR2713255

[61] Karl Wimmer, *Low influence functions over slices of the Boolean hypercube depend on few coordinates*, Proceedings of the 29th Annual Computational Complexity Conference, 2014, pp. 120–131. MR3281002