A Log-Sobolev Inequality for the Multislice, with Applications

Yuval Filmus[1]

Department of Computer Science, Technion – Israel Institute of Technology, Haifa, Israel
yuvalfi@cs.technion.ac.il

Ryan O'Donnell[2]

Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, USA
odonnell@cs.cmu.edu

Xinyu Wu

Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, USA
xinyuw1@andrew.cmu.edu

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

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

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.

2012 ACM Subject Classification Mathematics of computing → Markov processes

Keywords and phrases log-Sobolev inequality, small-set expansion, conductance, hypercontractivity, Fourier analysis, representation theory, Markov chains, combinatorics

Digital Object Identifier 10.4230/LIPIcs.ITCS.2019.34

Related Version Full version at `https://arxiv.org/abs/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*.

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

[1] Taub Fellow – supported by the Taub Foundations. The research was funded by ISF grant 1337/16.
[2] Supported by NSF grant CCF-1717606. This material is based upon work supported by the National Science Foundation under grant numbers listed above. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation (NSF).

10th Innovations in Theoretical Computer Science (ITCS 2019).
Editor: Avrim Blum; Article No. 34; pp. 34:1–34:12
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

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 $\kappa = (\#_1 u, \ldots, \#_\ell u) \in \mathbb{N}^\ell$ 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 $\kappa = (1, 1, \ldots, 1)$, the set $\mathcal{U}_\kappa$ is the set of all permutations of $[n]$.

1.1 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.

1.2 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., [25, 11]). For a subset $A \subseteq \mathcal{U}_\kappa$, define its *conductance* (or *expansion*) to be

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

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*,

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

is bounded below by a constant. The "small-set expansion" phenomenon [28, 37, 47] (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

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

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

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

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

We remark that Inequality (1) 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) 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) to nonnegative functions holds; namely,

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

(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 $\mathrm{KL}(\phi\pi \parallel \pi)$ denotes the *KL divergence* between distributions $\phi\pi$ and π.) Inequality (2) includes Inequality (1) by taking $\phi = 1_A/\mathrm{vol}(A)$.

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

▶ **Theorem 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* (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).

1.3 On the sharpness of Theorem 1

When ℓ is considered to be a constant, Theorem 1 is sharp up to constant factors (which we did not attempt to optimize); i.e.,

$$\varrho_\kappa^{-1} = \Theta(n) \cdot \log\left(\frac{n}{\min_i\{\kappa_i\}}\right) \quad \text{for } \ell = O(1). \tag{3}$$

To see the upper bound on ϱ_κ, assume without loss of generality that $\ell = \mathrm{argmin}_i\{\kappa_i\}$, 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) shows the claimed upper bound on ϱ_κ.

At the opposite extreme, when $\ell = n$ and $\kappa = (1, 1, \dots, 1)$, we have the random transposition walk on the symmetric group S_n. In this case, Theorem 1 as stated gives the poor bound of $\varrho_\kappa \geq \Omega(1/n^2 \log n)$, whereas the optimal bound is $\varrho_\kappa = \Theta(1/n \log n)$ [11, 35]. In fact, our proof of Theorem 1 (which generalizes that of [35]) 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 (3) always holds, up to universal constants, without the assumption of $\ell = O(1)$; we leave investigation of this for future work.

2 Applications

There are many known applications of log-Sobolev and hypercontractive inequalities in combinatorics and theoretical computer science (see, e.g., [42, Ch. 9, 10]). In this paper we present four particular consequences of Theorem 1 for analysis/combinatorics of Boolean functions on the multislice. We anticipate the possibility of several more.

2.1 KKL and Kruskal–Katona for multislices

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 and Inequality (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. [28] and Talagrand's strengthening of it [51]), follows immediately from previous work [43, 44]. It says that for any set with $\Phi[A] \le O(1/n)$, there must exist some transposition $\tau \in \mathrm{Trans}(n)$ with at least constant *influence* on A, where the influence of the transposition τ on A is defined to be

$$\mathrm{Inf}_\tau[A] = \mathop{\mathbf{Pr}}_{\boldsymbol{u}\sim\pi}\left[1_A(\boldsymbol{u}) \ne 1_A(\boldsymbol{u}^\tau)\right].$$

Formally, Talagrand's strengthening of the KKL theorem in this setting is:

▶ **Theorem 2.** *Let $f\colon \mathcal{U}_\kappa \to \{0,1\}$. Then*

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

Substituting our lower bound on ρ_κ from Theorem 1 yields concrete new results. For example, consider our model scenario of $n \longrightarrow \infty$ with $\ell = O(1)$ and $\kappa_i/n \ge \Omega(1)$ for each i; suppose further that f is "roughly balanced", meaning $\Omega(1) \le \mathbf{Var}[f] \le 1 - \Omega(1)$. Then

$$\mathop{\mathrm{avg}}_{\tau\in\mathrm{Trans}(n)}\left\{\frac{\mathrm{Inf}_\tau[f]}{\lg(2/\mathrm{Inf}_\tau[f])}\right\} \gtrsim \frac{1}{n},$$

and hence the maximum influence $\mathcal{M}[f]$ satisfies

$$\mathcal{M}[f] \gtrsim \frac{\log n}{n}.$$

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

Let us record here one more concrete corollary of Theorem 2. In our model scenario, that theorem (roughly speaking) says that the energy $\mathcal{E}[1_A] = \mathrm{avg}_{\tau\in\mathrm{Trans}(n)}\,\mathrm{Inf}_\tau[1_A]$ is at least $\Omega\big(\frac{\log n}{n}\big)$ unless some transposition $(i\ j)$ has a rather large influence, like $1/n^{.01}$, on 1_A.

▶ **Corollary 3.** *Let $A \subseteq \mathcal{U}_\kappa$. Assume $\kappa_i \ge pn$ for all $i \in [\ell]$ and that $\epsilon \le \mathrm{vol}(A) \le 1-\epsilon$. Then*

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

It is the hallmark of a junta A that every transposition τ has either $\mathrm{Inf}_\tau[A] = 0$ or $\mathrm{Inf}_\tau[A] \ge \Omega(1)$. Mirroring the original KKL Theorem, our work shows that: (i) if $\Phi[A] \le c/n$ then there exists τ with $\mathrm{Inf}_\tau[A] \ge \exp(-O(c))$; (ii) for *any* $A \subseteq \mathcal{U}_\kappa$ with $\Omega(1) \le \mathrm{vol}(A) \le 1-\Omega(1)$, there exists τ with $\mathrm{Inf}_\tau[A] \ge \Omega\big(\frac{\log n}{n}\big)$.

From our KKL Theorem, we obtain various versions of the Kruskal–Katona Theorem for multislices. The classical Kruskal–Katona Theorem [50, 33, 29] 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

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

It is not hard to show that $\mathrm{vol}(\partial A) \ge \mathrm{vol}(A)$ 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 [43, 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.** 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 $\kappa' \rhd \kappa$.

▶ **Definition 5.** 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 $\kappa' \lhd \kappa$.

▶ **Definition 6.** Given a histogram $\kappa \in \mathbb{N}_+^\ell$, we define a natural probability distribution $\mathbf{lower}(\kappa)$ on the lower neighbors of κ as follows. To draw $\kappa' \sim \mathbf{lower}(\kappa)$: take an arbitrary $u \in \mathcal{U}_\kappa$; choose $\boldsymbol{j}, \boldsymbol{j}' \sim [n]$ independently and randomly, conditioned on $u_{\boldsymbol{j}} \neq u_{\boldsymbol{j}'}$; let $\boldsymbol{c}, \boldsymbol{d}$ denote the two colors $u_{\boldsymbol{j}}, u_{\boldsymbol{j}'}$, with the convention $\boldsymbol{c} \prec \boldsymbol{d}$; finally, let κ' be the lower neighbor of κ with $\kappa'_{\boldsymbol{c}} = \kappa_{\boldsymbol{c}} + 1$ and $\kappa'_{\boldsymbol{d}} = \kappa_{\boldsymbol{d}} - 1$.

We similarly define a probability distribution $\mathbf{upper}(\kappa)$ on the upper neighbors of κ by interchanging the roles of $\boldsymbol{c}$ and $\boldsymbol{d}$.

▶ **Theorem 7.** *For $A \subseteq \mathcal{U}_\kappa$ we have*

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

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 $\mathrm{vol}(A) + \Omega(1/n)$. Using our KKL Theorem (Corollary 3) 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 8.** *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*

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

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 [43], we give a conceptual improvement to the "or else" clause in Theorem 8. 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:

▶ **Theorem 9.** *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*

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

or else there exists $j \in [n]$ and colors $c \prec d \in [\ell]$ with

$$\mathop{\mathbf{Pr}}_{\boldsymbol{u} \sim \pi_\kappa}[\boldsymbol{u} \in A \mid \boldsymbol{u}_j = d] - \mathop{\mathbf{Pr}}_{\boldsymbol{u} \sim \pi_\kappa}[\boldsymbol{u} \in A \mid \boldsymbol{u}_j = c] \geq 1/n^{.01}.$$

2.2 Friedgut Junta Theorem for multislices

A closely related consequence of our work is a **Friedgut Junta Theorem** for the multislice (cf. [24]), which follows (using a small amount of representation theory) from work of Wimmer [53] (see also [18] 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.

▶ **Theorem 10.** *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 $\mathbf{Pr}_{\boldsymbol{u} \sim \pi_\kappa}[f(\boldsymbol{u}) \neq h(\boldsymbol{u})] \leq \epsilon$.*

2.3 Nisan–Szegedy Theorem for multislices

Finally, with a little more representation theory effort, we are able to derive from Theorem 1 a **Nisan–Szegedy Theorem** for the multislice (cf. [41]), which is (roughly) an $\epsilon = 0$ version of the Friedgut Junta Theorem; this generalizes previous work on the Hamming slice [20]. 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 [21].)

More formally, 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 $3 \cdot 2^{k-1} - 2 \leq \gamma_2(k) < 22 \cdot 2^k$.) 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 [42, Def. 8.32]). We can obtain the following Nisan–Szegedy Theorem:

▶ **Theorem 11.** *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.*

3 Context and prior work

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

3.1 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 [25] 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 [40]); 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 [15].

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., [34] or [42, 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 [26], Lindsey [36], Bernstein [2], and Hart [27], 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 [16]. The same set of authors also recently proved a stability version of the edge isoperimetric inequality in the 99% regime [17].

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 $n = 1$.

3.2 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 [11] showed that for the $n = 1$ chain, the log-Sobolev constant is

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

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 (3) may hold with absolute constants for any ℓ.

3.3 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 [38], to child psychology [45], to computational learning theory [43]. An asymptotically exact analysis of the time to stationarity of this Markov chain was given by Diaconis and Shahshahani [12], 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 [11] before finally being determined (up to constants) by Lee and Yau in 1998 [35]. 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 [43, 8, 22], for computational learning theory [52, 43], for property testing [39], and for generalizing classic "analysis of Boolean functions" results [43, 44, 18, 19, 23, 22, 5].

3.4 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 [12, 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 [31, 14, 13, 1, 32] (with the analogous problems on the Johnson graphs serving as an important warmup [30]). Still, it seems fair to say that the mixing properties of the Grassmann graph are far from being fully understood.

3.5 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 [12], [9, p. 59], [20]. 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) [46].

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 [48] giving the asymptotic mixing time for the transposition walk on balanced multislices $\mathcal{U}_{(n/\ell,\ldots,n/\ell)}$ (see also [10, 49]). 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.

3.6 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.

References

1 Boaz Barak, Pravesh Kothari, and David Steurer. Small-Set Expansion in Shortcode Graph and the 2-to-2 Conjecture. Technical Report 1804.08662, arXiv, 2018.

2 Arthur Jay Bernstein. Maximally connected arrays on the n-cube. *SIAM J. Appl. Math.*, 15:1485–1489, 1967. doi:10.1137/0115129.

3 Aline Bonami. Étude des coefficients Fourier des fonctions de $L^p(G)$. *Annales de l'Institut Fourier*, 20(2):335–402, 1970.

4 Raphaël Bouyrie. An unified approach to the Junta theorem for discrete and continuous models. Technical report, arXiv, 2017. arXiv:1702.00753.

5 Raphaël Bouyrie. On quantitative noise stability and influences for discrete and continuous models. *Combin. Probab. Comput.*, 27(3):334–357, 2018. doi:10.1017/S0963548318000044.

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.*, 8:197–223, 2011.

7 John Chiarelli, Pooya Hatami, and Michael Saks. Tight Bound on the Number of Relevant Variables in a Bounded degree Boolean function. Technical report, arXiv, 2018. arXiv:1801.08564.

8 Pat Devlin and Jeff Kahn. On "stability" in the Erdős-Ko-Rado theorem. *SIAM J. Discrete Math.*, 30(2):1283–1289, 2016. doi:10.1137/15M1012992.

9 Persi Diaconis. *Group representations in probability and statistics*, volume 11 of *Institute of Mathematical Statistics Lecture Notes—Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 1988.

10 Persi Diaconis and Susan Holmes. Random walks on trees and matchings. *Electron. J. Probab.*, 7:no. 6, 17, 2002. doi:10.1214/EJP.v7-105.

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

12 Persi Diaconis and Mehrdad Shahshahani. Time to reach stationarity in the Bernoulli–Laplace diffusion model. *SIAM Journal on Mathematical Analysis*, 18(1):208–218, 1987.

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

14 Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. Towards a Proof of the 2-to-1 Games Conjecture? In *Proceedings of the 50th Annual ACM Symposium on Theory of Computing*, pages 376–389, 2018.

15 Paul Ehrenfest and Tatiana Ehrenfest. Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem. *Physikalische Zeitschrift*, 8(9):311–314, 1907.

16 David Ellis, Nathan Keller, and Noam Lifshitz. On a biased edge isoperimetric inequality for the discrete cube. Technical report, arXiv, 2017. `arXiv:1702.01675`.

17 David Ellis, Nathan Keller, and Noam Lifshitz. On the structure of subsets of the discrete cube with small edge boundary. *Discrete Analysis*, 9:1–29, 2018. `doi:10.19086/da.3668`.

18 Yuval Filmus. An orthogonal basis for functions over a slice of the Boolean hypercube. *Electron. J. Combin.*, 23(1):Paper 1.23, 27, 2016.

19 Yuval Filmus. Friedgut–Kalai–Naor theorem for slices of the Boolean cube. *Chic. J. Theoret. Comput. Sci.*, pages Art. 14, 17, 2016.

20 Yuval Filmus and Ferdinand Ihringer. Boolean constant degree functions on the slice are juntas. Technical report, arXiv, 2018. `arXiv:1801.06338`.

21 Yuval Filmus and Ferdinand Ihringer. Boolean degree 1 functions on some classical association schemes. Technical report, arXiv, 2018. `arXiv:1801.06034`.

22 Yuval Filmus, Guy Kindler, Elchanan Mossel, and Karl Wimmer. Invariance principle on the slice. *Transactions on Computation Theory*, 10(3):11, 2018.

23 Yuval Filmus and Elchanan Mossel. Harmonicity and invariance on slices of the Boolean cube. In *Proceedings of the 31st Annual Computational Complexity Conference*, pages Art. No. 16, 13, 2016.

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

25 Leonard Gross. Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975.

26 Lawrence H. Harper. Optimal assignments of numbers to vertices. *J. Soc. Indust. Appl. Math.*, 12:131–135, 1964.

27 Sergiu Hart. A note on the edges of the n-cube. *Discrete Math.*, 14(2):157–163, 1976. `doi:10.1016/0012-365X(76)90058-3`.

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

29 Gyula Katona. A theorem of finite sets. In *Theory of graphs (Proc. Colloq., Tihany, 1966)*, pages 187–207. Academic Press, New York, 1968.

30 Subhash Khot, Dor Minzer, Dana Moshkovitz, and Muli Safra. Small Set Expansion in The Johnson Graph. Technical Report TR18-078, ECCC, 2018.

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

32 Subhash Khot, Dor Minzer, and Muli Safra. Pseudorandom sets in Grassmann graph have near-perfect expansion. Technical Report TR18-006, ECCC, 2018.

33 Joseph B. Kruskal. The number of simplices in a complex. In *Mathematical optimization techniques*, pages 251–278. Univ. of California Press, Berkeley, Calif., 1963.

34 Michel Ledoux. Concentration of measure and logarithmic Sobolev inequalities. In *Séminaire de Probabilités XXXIII*, pages 120–216. Springer, 1999.

35 Tzong-Yau Lee and Horng-Tzer Yau. Logarithmic Sobolev inequality for some models of random walks. *Annals of Probability*, 26(4):1855–1873, 1998.

36 John H. Lindsey II. Assignment of numbers to vertices. *Amer. Math. Monthly*, 71:508–516, 1964. doi:10.2307/2312587.

37 Lászlo Lovász and Ravi Kannan. Faster mixing via average conductance. In *Proceedings of the 31st Annual ACM Symposium on Theory of Computing*, pages 282–287, 1999.

38 Patrick Moran. Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society*, 54(1):60–71, 1958.

39 Dana Moshkovitz. Direct Product Testing With Nearly Identical Sets. Technical Report TR14-182, ECCC, 2014.

40 Edward Nelson. The free Markoff field. *Journal of Functional Analysis*, 12:211–227, 1973.

41 Noam Nisan and Mario Szegedy. On the Degree of Boolean Functions as Real Polynomials. *Computational Complexity*, 4(4):301–313, 1994.

42 Ryan O'Donnell. *Analysis of Boolean Functions.* Cambridge University Press, 2014.

43 Ryan O'Donnell and Karl Wimmer. KKL, Kruskal-Katona, and monotone nets. *SIAM Journal on Computing*, 42(6):2375–2399, 2013.

44 Ryan O'Donnell and Karl Wimmer. Sharpness of KKL on Schreier graphs. *Electronic Communications in Probability*, 18:1–12, 2013.

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

46 D. H. J. Polymath. A new proof of the density Hales–Jewett theorem. *Annals of Mathematics*, 175(3):1283–1327, 2012.

47 Prasad Raghavendra and David Steurer. Graph expansion and the Unique Games Conjecture. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing*, pages 755–764, 2010.

48 Fabio Scarabotti. Time to reach stationarity in the Bernoulli–Laplace diffusion model with many urns. *Adv. in Appl. Math.*, 18(3):351–371, 1997. doi:10.1006/aama.1996.0514.

49 Fabio Scarabotti and Filippo Tolli. Harmonic analysis on a finite homogeneous space II: the Gelfand–Tsetlin decomposition. *Forum Math.*, 22(5):879–911, 2010. doi:10.1515/FORUM.2010.047.

50 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:117–118, 1959.

51 Michel Talagrand. On Russo's approximate zero-one law. *Annals of Probability*, 22(3):1576–1587, 1994.

52 Karl Wimmer. *Fourier methods and combinatorics in learning theory.* PhD thesis, Carnegie Mellon University, 2009.

53 Karl Wimmer. Low influence functions over slices of the Boolean hypercube depend on few coordinates. In *Proceedings of the 29th Annual Computational Complexity Conference*, pages 120–131, 2014.