A quasi-stability result for low-degree Boolean functions on $S_n$

David Ellis, Yuval Filmus*, and Ehud Friedgut†

October 2012

Abstract

We prove that Boolean functions on $S_n$, whose Fourier transform is highly concentrated on irreducible representations indexed by partitions of $n$ whose largest part has size at least $n-t$, are close to being unions of cosets of stabilizers of $t$-tuples. We also obtain an edge-isoperimetric inequality for the transposition graph on $S_n$, which is asymptotically sharp for sets of measure $1/poly(n)$, using eigenvalue techniques. We then combine these two results to obtain a best possible edge-isoperimetric inequality for sets of size $(n-t)!$, where $n$ is large compared to $t$, confirming a conjecture of Ben-Efraim in these cases.

1 Introduction

This paper (together with [12] and [13]) is part of a trilogy dealing with stability and ‘quasi-stability’ results concerning Boolean functions on the symmetric group, which are of ‘low complexity’.

Let us begin with some notation and definitions which will enable us to present the Fourier-theoretic context of our results.

Let $S_n$ denote the symmetric group of order $n$, the group of all permutations of $[n] := \{1, 2, \ldots, n\}$. For $i, j \in [n]$, we let

$$T_{ij} = \{ \sigma \in S_n : \sigma(i) = j \}.$$  

We call the $T_{ij}$’s the 1-cosets of $S_n$, since they are the cosets of stabilizers of points.

Similarly, for $t > 1$, and for two ordered $t$-tuples of distinct elements of $[n]$, $I = (i_1, \ldots, i_t)$ and $J = (j_1, \ldots, j_t)$, we let

$$T_{IJ} = \{ \sigma \in S_n : \sigma(I) = J \} = \{ \sigma \in S_n : \sigma(i_k) = j_k \forall k \in [t] \};$$

we call these the $t$-cosets of $S_n$. Abusing notation, we will also use $T_{ij}$ and $T_{IJ}$ to denote their own characteristic functions.

*Supported by the Canadian Friends of the Hebrew University / University of Toronto Permanent Endowment.

†Supported in part by I.S.F. grant 0398246, and BSF grant 2010247.
For any non-negative integer \( t \), we let \( U_t \) be the vector space of real-valued functions on \( S_n \) whose Fourier transform is supported on irreducible representations of \( S_n \) which are indexed by partitions of \( n \), whose largest part has size at least \( n - t \). In [14], it is proved that \( U_t \) is the space spanned by the \( t \)-cosets:

**Theorem 1.** If \( n \) and \( t \) are integers with \( 0 \leq t \leq n \), then

\[
U_t = \text{Span}\{T_{IJ} : I, J \text{ are ordered } t\text{-tuples of distinct elements of } [n]\}.
\]

If \( t \) is fixed and \( n \) is large, the irreducible representations of \( S_n \) which are indexed by partitions of \( n \) with largest part of size \( t \), all have dimension \( \Theta(n^t) \). The functions in \( U_t \) therefore have Fourier transform supported on irreducible representations of dimension \( O(n^t) \), and can be viewed as having low complexity. If \( f \) is a real-valued function on \( S_n \), we define the degree of \( f \) to be the minimum \( t \) such that \( f \in U_t \). This can be seen as a measure of the complexity of \( f \), analogous to the degree of a Boolean function on \( \{0,1\}^n \). Indeed, by Theorem 1, it is precisely the minimum possible total degree of a polynomial in the \( T_{ij} \)'s which is equal to \( f \).

The following theorem from [14] characterizes the Boolean functions in \( U_t \).

**Theorem 2** (Ellis, Friedgut, Pilpel). Let \( f : S_n \to \{0,1\} \) be in \( U_t \). Then \( f \) is the characteristic function of a disjoint union of \( t \)-cosets.

Note that a disjoint union of 1-cosets is necessarily of the form

\[
\bigcup_{j \in S} T_{ij}
\]

for some \( i \in [n] \) and some \( S \subset [n] \), or of the form

\[
\bigcup_{i \in S} T_{ij}
\]

for some \( j \in [n] \) and some \( S \subset [n] \), so is determined by the image or the pre-image of a fixed element. We call such a family a dictatorship. By contrast, a disjoint union of \( t \)-cosets need not be determined by the image or the preimage of a fixed \( t \)-tuple. Consider, for example, the family

\[
T_{(1,2)(1,2)} \cup T_{(1,3)(2,3)} \cup T_{(1,4),(3,4)} \cup T_{(1,5),(4,5)} \cup T_{(1,6),(5,6)} \cup \ldots \cup T_{(1,n),(n-1,n)},
\]

which is a disjoint union of 2-cosets, but is not even determined by the images or pre-images of a bounded number of 2-tuples.

In [12], we proved that a Boolean function of expectation \( O(1/n) \), whose Fourier transform is highly concentrated on irreducible representations corresponding to the partitions \((n)\) and \((n-1,1)\), is close in structure to a union of 1-cosets. Put another way, a Boolean function of expectation \( O(1/n) \), which is close to \( U_1 \) (in Euclidean distance), is close in structure to a union of 1-cosets. This is not true stability, as the Boolean function corresponding to \( T_{11} \cup T_{22} \) is \( O(1/n^2) \) close to \( U_1 \), but is not \( O(1/n) \) close to any dictatorship, whereas a Boolean function in \( U_1 \) must be a dictatorship. We call it a ‘quasi-stability’ result. (In [13], on the other hand, we prove that a Boolean
function of expectation bounded away from 0 and 1, which is close to $U_1$, must be close in structure to a dictatorship; this is ‘genuine’ stability.)

Our aim in this paper is to prove an analogue of our quasi-stability result in [12], for Boolean functions of degree at most $t$. Namely, we show that a Boolean function on $S_n$ with expectation $O(n^{-t})$, whose Fourier transform is highly concentrated on irreducible representations indexed by partitions of $n$ with first row of length at least $n - t$, are close to being unions of cosets of stabilizers of $t$-tuples. Put another way, a Boolean function of expectation $O(n^{-t})$, which is close to $U_t$ (in Euclidean distance), is close in structure to a union of $t$-cosets.

This can be viewed as a non-Abelian analogue of the theorems in [5] and [22], which concern Boolean functions whose Fourier transform is highly concentrated on small sets. In [22], for example, it is shown that a Boolean function on $\{0,1\}^n$ whose Fourier transform is highly concentrated on sets of size at most $t$, must be close in structure to a junta depending upon at most $F(t)$ coordinates, for some function $F$.

Our proof is similar in some respects to the proof in [12], but the representation-theoretic tools are significantly more involved. It also turns out to be easier to deal with fourth moments (rather than third moments, as in [12]); these have the advantage of always being non-negative, although we pay the price of having to substitute approximations for many of the exact expressions in [12].

We use one of our representation-theoretic lemmas to obtain an edge-isoperimetric inequality for the transposition graph on $S_n$, which is asymptotically sharp for sets of measure $1/\text{poly}(n)$. We then combine this with our quasi-stability result to obtain a best possible edge-isoperimetric inequality for sets of size $(n-t)!$, where $n$ is large compared to $t$, confirming a conjecture of Ben Efraim in these cases.

## 2 Background

### Background on general representation theory

In this section, we recall the basic notions and results we need from general representation theory. For more background, the reader may consult [25].

Let $G$ be a finite group. A representation of $G$ over $\mathbb{C}$ is a pair $(\rho, V)$, where $V$ is a finite-dimensional complex vector space, and $\rho: G \to GL(V)$ is a group homomorphism from $G$ to the group of all invertible linear endomorphisms of $V$. The vector space $V$, together with the linear action of $G$ defined by $gv = \rho(g)(v)$, is sometimes called a $\mathbb{C}G$-module. A homomorphism between two representations $(\rho, V)$ and $(\rho', V')$ is a linear map $\phi: V \to V'$ such that such that $\phi(\rho(g)(v)) = \rho'(g)(\phi(v))$ for all $g \in G$ and $v \in V$. If $\phi$ is a linear isomorphism, the two representations are said to be equivalent, or isomorphic, and we write $(\rho, V) \cong (\rho', V')$. If $\dim(V) = n$, we say that $\rho$ has dimension $n$, and we write $\dim(\rho) = n$.

The representation $(\rho, V)$ is said to be irreducible if it has no proper subrepresentation, i.e. there is no proper subspace of $V$ which is $\rho(g)$-invariant for all $g \in G$.

It turns out that for any finite group $G$, there are only finitely many equivalence classes of complex irreducible representations of $G$, and any complex representation of $G$ is isomorphic to a direct sum of irreducible representations of $G$. Hence, we may
choose a set of representatives $\mathcal{R}$ for the equivalence classes of complex irreducible representations of $G$.

If $(\rho, V)$ is a representation of $G$, the character $\chi_\rho$ of $\rho$ is the map defined by

$$\chi_\rho : G \to \mathbb{C};$$
$$\chi_\rho(g) = \text{Tr}(\rho(g)),$$

where $\text{Tr}(\alpha)$ denotes the trace of the linear map $\alpha$ (i.e. the trace of any matrix of $\alpha$). Note that $\chi_\rho(\text{Id}) = \dim(\rho)$, and that $\chi_\rho$ is a class function on $G$ (meaning that it is constant on each conjugacy-class of $G$.)

The usefulness of characters lies in the following

**Fact.** Two complex representations of $G$ are isomorphic if and only if they have the same character. The complex irreducible characters form an orthonormal basis for the vector space of complex-valued class functions on $G$.

Since we will work only with complex representations, from now on, ‘representation’ will mean ‘complex representation’.

We now define the Fourier transform of a function on an arbitrary finite group.

**Definition.** Let $\mathcal{R}$ be a complete set of non-isomorphic, irreducible representations of $G$, i.e. containing one representative from each isomorphism class of irreducible representations of $G$. Let $f : G \to \mathbb{C}$ be a complex-valued function on $G$. We define the Fourier transform of $f$ at an irreducible representation $\rho \in \mathcal{R}$ as

$$\hat{f}(\rho) = \frac{1}{|G|} \sum_{\sigma \in G} f(\sigma) \rho(\sigma);$$

(1)

Note that $\hat{f}(\rho)$ is a linear endomorphism of $V$.

Let $G$ be a finite group. Let $\mathbb{C}[G]$ denote the vector space of all complex-valued functions on $G$. Let $\mathbb{P}$ denote the uniform probability measure on $G$,

$$\mathbb{P}(\mathcal{A}) = |\mathcal{A}|/|G| \quad (\mathcal{A} \subset G),$$

let

$$\langle f, g \rangle = \frac{1}{|G|} \sum_{\sigma \in G} f(\sigma) \overline{g(\sigma)}$$

denote the corresponding inner product, and let

$$||f||_2 = \sqrt{\mathbb{E}[f^2]} = \sqrt{\frac{1}{|G|} \sum_{\sigma \in G} |f(\sigma)|^2}$$

denote the induced Euclidean norm.

For each irreducible representation $\rho \in \mathcal{R}$, let

$$U_\rho := \{ f \in \mathbb{C}[G] : \hat{f}(\pi) = 0 \text{ for all } \pi \in \mathcal{R} \setminus \{\rho\} \}.$$
We refer to this as the subspace of functions whose Fourier transform is supported on the irreducible representation \( \rho \), and we refer to the \( U_\rho \) as the isotypical subspaces. It turns out that the \( U_\rho \) are pairwise orthogonal, and that

\[
\mathbb{C}[G] = \bigoplus_{\rho \in \mathcal{R}} U_\rho. \tag{2}
\]

For each \( \rho \in \mathcal{R} \), let \( f_\rho \) denote the orthogonal projection of \( f \) onto the subspace \( U_\rho \). It follows from the above that

\[
||f||^2 = \sum_{\rho \in \mathcal{R}} ||f_\rho||^2. \tag{3}
\]

The group algebra \( \mathbb{C}G \) denotes the complex vector-space with basis \( G \) and multiplication defined by extending the group multiplication linearly. In other words,

\[
\mathbb{C}G = \left\{ \sum_{g \in G} x_g g : x_g \in F \ \forall g \in G \right\},
\]

and

\[
\left( \sum_{g \in G} x_g g \right) \left( \sum_{h \in G} y_h h \right) = \sum_{g, h \in G} x_g y_h (gh).
\]

As a vector space, \( \mathbb{C}G \) may be identified with \( \mathbb{C}[G] \), by identifying \( \sum_{g \in G} x_g g \) with the function \( g \mapsto x_g \).

**Background on the representation theory of \( S_n \).**

**Definition.** A partition of \( n \) is a non-increasing sequence of integers summing to \( n \), i.e. a sequence \( \lambda = (\lambda_1, \ldots, \lambda_k) \) with \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k \) and \( \sum_{i=1}^k \lambda_i = n \); we write \( \lambda \vdash n \). For example, \( (3,2,2) \vdash 7 \).

The following two orders on partitions of \( n \) will be useful.

**Definition.** (Dominance order) Let \( \lambda = (\lambda_1, \ldots, \lambda_k) \) and \( \mu = (\mu_1, \ldots, \mu_l) \) be partitions of \( n \). We say that \( \lambda \succeq \mu \) (\( \lambda \) dominates \( \mu \)) if \( \sum_{j=1}^k \lambda_i \geq \sum_{j=1}^l \mu_i \ \forall i \) (where we define \( \lambda_i = 0 \ \forall i > k \), \( \mu_i = 0 \ \forall i > l \)).

It is easy to see that this is a partial order.

**Definition.** (Lexicographic order) Let \( \lambda = (\lambda_1, \ldots, \lambda_r) \) and \( \mu = (\mu_1, \ldots, \mu_s) \) be partitions of \( n \). We say that \( \lambda > \mu \) if \( \lambda_j > \mu_j \), where \( j = \min \{i \in [n] : \lambda_i \neq \mu_i \} \).

It is easy to see that this is a total order which extends the dominance order.

It is well-known that there is an explicit 1-1 correspondence between irreducible representations of \( S_n \) (up to isomorphism) and partitions of \( n \). The reader may refer to [26] for a full description of this correspondence, or to the paper [14] for a shorter description.
For each partition $\alpha$ of $n$, we write $[\alpha]$ for the corresponding isomorphism class of irreducible representations of $S_n$, and we write $U_\alpha = U_{[\alpha]}$ for the vector space of complex-valued functions on $S_n$ whose Fourier transform is supported on $[\alpha]$. Similarly, if $f \in \mathbb{C}[S_n]$, we write $f_\alpha$ for the orthogonal projection of $f$ onto $U_\alpha$. By (2), we have

$$\mathbb{C}[S_n] = \bigoplus_{\alpha \vdash n} U_\alpha,$$

and the $U_\alpha$ are pairwise orthogonal. For any $f \in \mathbb{C}[S_n]$, we have

$$f = \sum_{\alpha \vdash n : \alpha_1 \geq n-t} f_\alpha,$$

and

$$\|f\|_2^2 = \sum_{\alpha \vdash n : \alpha_1 \geq n-t} \|f_\alpha\|^2.$$

We write $\chi_\alpha$ for the character of the irreducible representation corresponding to $\alpha$, and we write $\dim[\alpha]$ for the dimension of $[\alpha]$.

**Definition.** If $\lambda = (\lambda_1, \ldots, \lambda_l)$ is a partition of $n$, the Young diagram of shape $\lambda$ is an array of left-justified cells with $\lambda_i$ cells in row $i$, for each $i \in [l]$.

For example, the Young diagram of the partition $(3, 2^2)$ is:

```
  +---+
  |   |
  |   |
  +---+---+
        |   |
  +---+---+---+
```

**Definition.** A $\lambda$-tableau is a Young diagram of shape $\lambda$, each of whose cells contains a number between 1 and $n$. If $\mu$ is a partition of $n$, a Young tableau is said to have content $\mu$ if it contains $\mu_i$ i’s for each $i \in \mathbb{N}$.

**Definition.** A Young tableau is said to be standard if it has content $(1, 1, \ldots, 1)$ and the numbers are strictly increasing down each row and along each column.

**Definition.** A Young tableau is said to be semistandard if the numbers are non-decreasing along each row and strictly increasing down each column.

The relevance of standard Young tableaux stems from the following.

**Theorem 3.** If $\alpha$ is a partition of $n$, then $\dim[\alpha]$ is the number of standard $\alpha$-tableaux.

We say that two $\lambda$-tableaux are row-equivalent if they contain the same numbers in each row. A row-equivalence-class of $\lambda$-tableaux is called a $\lambda$-tabloid. Consider the natural action of $S_n$ on the set of $\lambda$-tabloids, and let $M^\lambda$ denote the induced permutation representation. We write $\xi_\lambda$ for the character of $M^\lambda$; the $\xi_\lambda$ are called the permutation characters of $S_n$.

**Definition.** Let $\lambda$ and $\mu$ be partitions of $n$. The Kostka number $K_{\lambda, \mu}$ is the number of semistandard $\lambda$-tableaux of content $\mu$. 

Young's theorem gives a decomposition of permutation representations into irreducible representations of $S_n$.

**Theorem 4** (Young's theorem). If $\mu$ is a partition of $n$, then

$$M^\mu \cong \bigoplus_{\lambda \vdash n, \lambda \geq \mu} K_{\lambda, \mu}[\lambda].$$

It follows that for each partition $\mu$ of $n$, we have

$$\xi_\mu = \sum_{\lambda \vdash n, \lambda \geq \mu} K_{\lambda, \mu} \chi_\lambda.$$

On the other hand, we can express the irreducible characters in terms of the permutation characters using the determinantal formula: for any partition $\alpha$ of $n$,

$$\chi_\alpha = \sum_{\pi \in S_n} \text{sgn}(\pi) \xi_{\alpha \cdot \text{id} + \pi}. \quad (4)$$

Here, if $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_l)$, $\alpha \cdot \text{id} + \pi$ is defined to be the sequence

$$(\alpha_1 - 1 + \pi(1), \alpha_2 - 2 + \pi(2), \ldots, \alpha_l - l + \pi(l)).$$

If this sequence has all its entries non-negative, we let $\alpha \cdot \text{id} + \pi$ be the partition of $n$ obtained by reordering its entries, and we define $\xi_{\alpha \cdot \text{id} + \pi} = \xi_{\alpha \cdot \text{id} + \pi}$. If the sequence has a negative entry, we define $\xi_{\alpha \cdot \text{id} + \pi} = 0$. Note that if $\xi_\beta$ appears on the right-hand side of (4), then $\beta \geq \alpha$, so the determinantal formula expresses $\chi_\alpha$ in terms of $\{\xi_\beta : \beta \geq \alpha\}$.

For each $t \in \mathbb{N}$, we define

$$U_t = \bigoplus_{\alpha \vdash n : \alpha_1 \geq n - t} U_\alpha,$$

i.e. $U_t$ is the subspace of functions on $S_n$ whose Fourier transform is concentrated on irreducible representations corresponding to partitions with first row of length at least $n - t$. It was proved in [14] that $U_t$ is the linear span of the characteristic functions of the $t$-cosets of $S_n$:

$$U_t = \text{Span}\{T_{IJ} : I, J \text{ are ordered } t\text{-tuples of distinct elements of } [n]\}.$$

For brevity, we shall sometimes write $C(n, t)$ for the set of $t$-cosets of $S_n$.

If $f : S_n \to \mathbb{C}$, we write $f_t$ for the orthogonal projection of $f$ onto $U_t$, for each $t \in [n]$; equivalently,

$$f_t = \sum_{\alpha \vdash n, \alpha_1 \geq n - t} f_\alpha.$$

For each $t \in \mathbb{N}$, we write

$$V_t = \bigoplus_{\alpha \vdash n, \alpha_1 = n - t} U_\alpha.$$

Note that the $V_t$ are pairwise orthogonal, and that

$$U_t = U_{t-1} \oplus V_t$$

for each $t \leq n$. 

7
Background on normal Cayley graphs

Several of our results will involve the analysis of normal Cayley graphs on $S_n$. We recall the definition of a Cayley graph on a finite group.

**Definition.** Let $G$ be a finite group, and let $S \subset G \setminus \{Id\}$ be inverse-closed (meaning that $S^{-1} = S$). The Cayley graph on $G$ with generating set $S$ is the graph with vertex-set $G$, where we join $g$ to $gs$ for every $g \in G$ and $s \in S$; we denote it by $\text{Cay}(G,S)$.

Formally, 

$$V(\text{Cay}(G,S)) = G, \quad E(\text{Cay}(G,S)) = \{\{g,gs\} : g \in G, s \in S\}.$$  

Note that the Cayley graph $\text{Cay}(G,S)$ is $|S|$-regular. If the generating set $S$ is conjugation-invariant, i.e. is a union of conjugacy classes of $G$, then the Cayley graph $\text{Cay}(G,S)$ is said to be a normal Cayley graph.

The connection between normal Cayley graphs and representation theory arises from the following fundamental theorem, which states that for any normal Cayley graph, the eigenspaces of its adjacency matrix are in 1-1 correspondence with the isomorphism classes of irreducible representations of the group.

**Theorem 5** (Frobenius / Schur / Diaconis-Shahshahani). Let $G$ be a finite group, let $S \subset G$ be an inverse-closed, conjugation-invariant subset of $G$, let $\Gamma = \text{Cay}(G,S)$ be the Cayley graph on $G$ with generating set $S$, and let $A$ be the adjacency matrix of $\Gamma$. Let $\mathcal{R}$ be a complete set of non-isomorphic complex irreducible representations of $G$. Then we have

$$\mathbb{C}[G] = \bigoplus_{\rho \in \mathcal{R}} U_\rho,$$

and each $U_\rho$ is an eigenspace of $A$ with dimension $\dim(\rho)^2$ and eigenvalue

$$\lambda_\rho = \frac{1}{\dim(\rho)} \sum_{g \in S} \chi_\rho(g). \quad (5)$$

Note that the adjacency matrix $A$ in the above theorem is given by

$$A_{g,h} = 1_S(gh^{-1}) \quad (g,h \in G).$$

We will need a slight generalization of Theorem 5. Recall that if $G$ is a finite group, a class function on $G$ is a function on $G$ which is constant on each conjugacy-class of $G$. If $S$ is conjugation-invariant, then $1_S$ is clearly a class function. Observe that if $A$ is any matrix in $\mathbb{C}[G \times G]$ defined by

$$A_{g,h} = w(gh^{-1}),$$

where $w$ is a class function on $G$, then $A$ is a linear combination of the adjacency matrices of normal Cayley graphs. Therefore, the isotypical subspaces $U_\rho$ are again eigenspaces of $A$, and the corresponding eigenvalues are given by

$$\lambda_\rho = \frac{1}{\dim(\rho)} \sum_{g \in S} w(g)\chi_\rho(g). \quad (6)$$
Applying Theorem 5 to $S_n$, we see that if $\Gamma = \text{Cay}(S_n, X)$ is a normal Cayley graph on $S_n$, and $A$ is the adjacency matrix of $\Gamma$, then we have

$$\mathbb{C}[S_n] = \bigoplus_{\alpha \vdash n} U_\alpha,$$

and each $U_\alpha$ is an eigenspace of $A$ with dimension $\dim[\alpha]$ and eigenvalue

$$\lambda_\alpha = \frac{1}{\dim[\alpha]} \sum_{\sigma \in X} \chi_\alpha(\sigma). \quad (7)$$

3 Preliminary results

In this section, we prove some preliminary representation-theoretic lemmas.

Preliminary results on the characters of $S_n$

In this section, we prove some easy results on the characters of $S_n$. We will use the following standard notation. For $n, r \in \mathbb{N}$, we will write $(n)_r := n(n-1) \ldots (n-r+1)$ for the $r$th falling factorial moment of $n$, and we will write $([n])_r$ for the set of all $r$-tuples of distinct elements of $[n]$.

If $f = f(n,t)$ and $g = g(n,t)$ are functions of $n$ and $t$, we will write $f = O_t(g)$ to mean that for every $t$, there exists $C_t > 0$ such that $f(n,t) \leq C_t g(n,t)$ for all $n \in \mathbb{N}$. Similarly, we write $f = \Omega_t(g)$ to mean that for every $t$, there exists $c_t > 0$ such that $f(n,t) \geq c_t g(n,t)$ for all $n \in \mathbb{N}$. We write $f = \Theta_t(g)$ if $f = O_t(g)$ and $f = \Omega_t(g)$.

We begin with a crude upper bound on the dimensions of the irreducible characters $\chi_\alpha$.

**Lemma 6.** Let $\alpha$ be a partition of $n$ with $\alpha_1 = n - s$. Then

$$\dim[\alpha] \leq (n)_s.$$

**Proof.** Since $\alpha_1 = n - s$, we have $\alpha \geq (n-s, 1^s)$. Recall (or observe from Theorem 4) that for each partition $\beta \geq (n-s, 1^s)$, $[\beta]$ is an irreducible constituent of $M^{(n-s,1^s)}$, which has dimension $(n)_s$. It follows that $\dim[\alpha] \leq (n)_s$, as required. \qed

We also need crude bounds on the Kostka numbers $K_{\alpha,(n-s,1^s)}$.

**Lemma 7.** Let $\alpha$ be a partition of $n$ with $\alpha_1 = n - s$. Then

$$\left(\begin{array}{c} n-s \\ s \end{array}\right) K_{\alpha,(n-s,1^s)} \leq \dim[\alpha] \leq \left(\begin{array}{c} n \\ s \end{array}\right) K_{\alpha,(n-s,1^s)}.$$

**Proof.** Let us write $\alpha = (n-s, \gamma)$, where $\gamma$ is a partition of $s$. Recall from Theorem 3 that $\dim[\alpha]$ is the number of standard $\alpha$-tableaux. Observe that we may construct $\left(\begin{array}{c} n-s \\ s \end{array}\right)$ distinct standard $\alpha$-tableaux as follows:
1. Place $i$ in cell $(1, i)$ for $i = 1, 2, \ldots, s$.

2. Choose any $s$-set $\{i_1, \ldots, i_s\}$ from $\{s + 1, \ldots, n\}$ $\binom{n-s}{s}$ choices).

3. Place $\{i_1, \ldots, i_s\}$ in the cells below row 1 so that they are strictly increasing along each row and down each column. The number of ways of doing this is precisely the number of standard $\gamma$-tableaux, which is $\dim[\gamma]$.

4. Place the other numbers in the remaining cells of row 1, in increasing order from left to right.

Hence, there are at least $\binom{n-s}{s} \dim[\gamma]$ standard $\alpha$-tableaux. On the other hand, removing the first row from any standard $\alpha$-tableau produces a standard $\gamma$-tableau (filled with some $s$ of the numbers between 1 and $n$, rather than with $\{1, 2, \ldots, s\}$). Hence, the number of standard $\alpha$-tableaux is at most $\binom{n-s}{s} \dim[\gamma]$.

Recall from Theorem 4 that $K_{\alpha, (n-s, 1^s)}$ is the number of semistandard $\alpha$-tableaux of content $(n-s, 1^s)$. A semistandard $\alpha$-tableau of content $(n-s, 1^s)$ must have all $n-s$ of its 1’s in the first row; after deleting the first row, what is left is precisely a standard $\gamma$-tableau, filled with the numbers $\{2, 3, \ldots, s+1\}$, rather than with $\{1, 2, \ldots, s\}$. Hence, the number of semistandard $\alpha$-tableaux of content $(n-s, 1^s)$ is precisely the number of standard $\gamma$-tableaux, so $K_{\alpha, (n-s, 1^s)} = \dim[\gamma]$. Therefore,

\[
\binom{n-s}{s} K_{\alpha, (n-s, 1^s)} \leq \dim[\alpha] \leq \binom{n-s}{s} K_{\alpha, (n-s, 1^s)},
\]

as required.

Finally, we need a crude lower bound on the $L^1$-norm of the characters of the symmetric group.

**Lemma 8.** For each $s \in \mathbb{N}$, there exists $K_s > 0$ such that for any $n \in \mathbb{N}$ and any partition $\alpha$ of $n$ with $\alpha_1 \geq n-s$, we have

\[
\frac{1}{n!} \sum_{\sigma \in S_n} |\chi_\alpha(\sigma)| \geq K_s.
\]

**Proof.** To assist with the proof, we introduce the following notation/definition.

**Definition.** Let $s \in \mathbb{N}$, and let $\beta = (\beta_1, \ldots, \beta_k)$ be a partition of $n$ with $\beta_1 \geq n-s$. We define $X_{\beta, s}$ to be the set of all permutations with cycle-type in the set

\[
\{(\lambda_1, \ldots, \lambda_k, \beta_2, \ldots, \beta_k) : (\lambda_1, \ldots, \lambda_k) \vdash \beta_1, |\lambda_i| > s \; \forall \; i \in [k]\}.
\]

**Claim 1.** Let $\beta = (\beta_1, \ldots, \beta_k)$ be a partition of $n$ with $\beta_1 \geq n-s$. Then

\[
|X_{\beta, s}| \geq L_s n!,
\]

where $L_s > 0$ depends upon $s$ alone.
Proof of Claim. If $s < m$, let $N_{m,s}$ denote the number of permutations in $S_m$ whose cycles all have lengths greater than $s$. Let $\beta_1 = n - r$, where $r \leq s$. Observe that

$$|X_{\beta,s}| \geq \binom{n}{r} N_{n-r,s},$$

since there are $\binom{n}{r}$ choices for the $r$ numbers to go in the cycles of lengths $\beta_2, \ldots, \beta_l$, and then $N_{n-r,s}$ choices for placing the other $n-r$ numbers in the cycles of lengths greater than $s$. It is easy to see that $N_{m,s} \geq c_s m!$, for some $c_s > 0$ depending upon $s$ alone. (For a proof of this, see [14].) It follows that

$$|X_{\beta,s}| \geq \binom{n}{r} c_s (n-r)! = c_s n!/r! \geq c_s n!/s! = L_s n!,$$

where $L_s = c_s/s!$, proving the claim.

Claim 2. If $\alpha, \beta$ are partitions of $n$ with $\alpha_1 \geq \beta_1 = n - s$, where $n > 2s$, then $\chi_\alpha$ is constant on $X_{\beta,s}$.

Proof of Claim. Recall that $\chi_\alpha$ can be expressed as a linear combination of the permutation characters $\{\xi_\gamma : \gamma \geq \alpha\}$. Recall also that $\xi_\gamma(\sigma)$ is simply the number of $\gamma$-tabloids fixed by $\sigma$, which is the number of $\gamma$-tabloids that can be produced by taking each row to be a union of cycles of $\sigma$. This is clearly the same for all $\sigma \in X_{\beta,s}$, since in order to produce a $\gamma$-tabloid from $\sigma \in X_{\beta,s}$ as above, the top row must contain the union of all the cycles of length greater than $s$.

It follows that $\chi_\alpha(X_{\beta,s}) = \chi_\alpha(\sigma_\beta)$, where $\sigma_\beta$ is a permutation of cycle-type $\beta$ — so $\chi_\alpha(X_{\beta,s})$ is simply the $(\alpha, \beta)$-entry of the character table of $S_n$.

We can now prove the lemma. Let $\alpha$ be a partition of $n$ with $\alpha_1 \geq n - s$. Recall that for any partition $\alpha$ of $n$, the top-left minor of the character table of $S_n$ indexed by partitions $\geq \alpha$, is invertible. It follows that it cannot have a row of zeros, so there exists $\beta \geq \alpha$ such that $\chi_\alpha(\sigma_\beta) \neq 0$. Since the irreducible characters of $S_n$ are all integer valued, we have $|\chi_\alpha(\sigma_\beta)| \geq 1$. Hence, we have

$$\frac{1}{n!} \sum_{\sigma \in S_n} |\chi_\alpha(\sigma)| \geq \frac{1}{n!} |X_{\beta,s}| |\chi_\alpha(\sigma_\beta)| \geq L_s.$$

This proves the lemma when $n > 2s$. Since no character vanishes on all of $S_n$, by taking $K_s$ to be sufficiently small, we see that the lemma holds for all $n$.

A preliminary result on the projections of Boolean functions

The following result will be useful both in the proof of our quasi-stability theorem, and in that of our approximate isoperimetric inequality for the transposition graph.

Lemma 9. For each $t \in \mathbb{N}$, there exists a real number $C_t > 0$ such that the following holds. If $A \subset S_n$, and $f = 1_A$ denotes the characteristic function of $A$, then

$$\|f_{t-1}\|_2^2 \leq C_t (n)^{t-1} (|A|/n!)^2.$$
Proof. Let $\alpha$ be a partition of $n$ with $\alpha_1 = n - s$. Our aim is to prove that

$$||f_\alpha||_2^2 \leq C'_s(n)_s(|A|/n!)^2,$$

where $C'_s > 0$ depends upon $s$ alone. If $s = 0$, then $\alpha = (n)$, and we have

$$||f_{(n)}||_2^2 = (|A|/n!)^2,$$

so we may take $C'_s = 1$.

Suppose now that $s \geq 1$. Let $\Gamma = \text{Cay}(S_n, X)$ be any normal Cayley graph on $S_n$, and let $A$ denote its adjacency matrix. If $B \subset S_n$, let $e(B)$ denote the number of edges of $\Gamma$ within $B$, and let $(\lambda_\beta)_{\beta \vdash n}$ denote the eigenvalues of $A$. By Theorem 5, we have

$$Af = \sum_{\beta \vdash n} \lambda_\beta f_\beta,$$

and therefore

$$2e(A) = n!(f, Af) = n! \sum_{\beta \vdash n} \lambda_\beta ||f_\beta||^2.$$

Trivially, we have $e(A) \leq \binom{|A|}{2}$, and therefore

$$|A|^2 \geq |A|(|A| - 1) \geq 2e(A) = n! \sum_{\beta \vdash n} \lambda_\beta ||f_\beta||^2 \geq \lambda_\alpha ||f_\alpha||^2.$$

It follows that

$$||f_\alpha||_2^2 \leq \frac{|A|^2}{n!\lambda_\alpha}.$$

In order to minimize the right-hand side, we will choose $\Gamma$ with $\lambda_\alpha$ as large as possible.

By (7), we have

$$\lambda_\alpha = \frac{1}{\dim(\alpha)} \sum_{\sigma \in X} \chi_\alpha(\sigma).$$

To maximize $\lambda_\alpha$, we simply take $X = \{\sigma \in S_n : \chi_\alpha(\sigma) > 0\}$. The corresponding Cayley graph $\text{Cay}(S_n, X)$ then has

$$\lambda_\alpha = \frac{1}{\dim(\alpha)} \sum_{\substack{\sigma \in S_n \\lambda_\alpha(\sigma) > 0}} \chi_\alpha(\sigma).$$

Since $\alpha \neq (n)$, by the orthogonality of the irreducible characters, we have

$$\langle \chi_\alpha, \chi_{(n)} \rangle = 0,$$

i.e.

$$\sum_{\sigma \in S_n} \chi_\alpha(\sigma) = 0,$$

and therefore

$$\sum_{\substack{\sigma \in S_n \\chi_\alpha(\sigma) > 0}} \chi_\alpha(\sigma) = \frac{1}{2} \sum_{\sigma \in S_n} |\chi_\alpha(\sigma)|.$$
Hence,
\[ \lambda_\alpha = \frac{1}{\dim[\alpha]} \sum_{\sigma \in S_n} |\chi_\alpha(\sigma)|. \]

We therefore obtain
\[ ||f_\alpha||_2^2 \leq \frac{2 \dim[\alpha]|A|^2}{n! \sum_{\sigma \in S_n} |\chi_\alpha(\sigma)|}. \]

By Lemma 6, we have \( \dim[\alpha] \leq (n)_s \), and by Lemma 8, we have
\[ \frac{1}{n!} \sum_{\sigma \in S_n} |\chi_\alpha(\sigma)| \geq K_s. \]

Hence,
\[ ||f_\alpha||_2^2 \leq \frac{2(n)_s}{K_s} (|A|/n!)^2 = C'_s (n)_s (|A|/n!)^2, \]
where \( C'_s = 2/K_s \).

Recall that
\[ ||f_{t-1}||_2^2 = \sum_{\alpha_1 \geq n-t+1} ||f_\alpha||_2^2. \]

Note that for each \( n \geq 2s \), the number of partitions \( \alpha \) of \( n \) with \( \alpha_1 = n - s \) is equal to the number of partitions of \( s \). Hence, the number of terms in the above sum is bounded from above by a function of \( t \) alone, so there exists \( C_t > 0 \) such that
\[ ||f_{t-1}||_2^2 \leq C_t (n)_t (|A|/n!)^2, \]
as required.

Observe that this implies that if \( A \subset S_n \) with \( |A| = o((n-t+1)!) \), then \( ||(1_A)_t||_2^2 = o(|A|/n!) \), i.e. the projection of \( 1_A \) onto \( U_{t-1} \) has small \( L^2 \) norm. In other words, the Fourier transform of \( 1_A \) has very little ‘mass’ on the irreducible representations \( \{[\alpha] : \alpha_1 \geq n-t+1 \} \).

A Cauchy-Schwarz-type inequality

We will need a special case of the following general inequality, which we prove using multiple applications of Cauchy-Schwarz.

**Lemma 10.** Let \( X \) be a finite set, let \( m \geq 2 \), and let \( a_1, \ldots, a_m, L \in \mathbb{N} \). Suppose \( f_j : X^{a_j} \to \mathbb{R} \) for \( j = 1, 2, \ldots, m \). Suppose \( \sigma_j : [a_j] \to [L] \) are injections, such that each \( \ell \in [L] \) lies in the image of at least two of these injections. Then
\[
\left( \sum_{i_1, \ldots, i_L \in X} \prod_{j=1}^m f_j(i_{\sigma_j(1)}, \ldots, i_{\sigma_j(a_j)}) \right)^2 \leq \prod_{j=1}^m \left( \sum_{i_{\sigma_j(1)}, \ldots, i_{\sigma_j(a_j)} \in X} f_j(i_1, \ldots, i_{a_j})^2 \right). 
\]
**Remark.** This contains, as special cases, several simple inequalities that are frequently used in discrete analysis, for example inequalities

\[
\sum_{i,j,k \in X} f(i,j)f(j,k)f(k,i) \leq \left( \sum_{i,j \in X} f(i,j) \right)^{3/2}
\]

and

\[
\sum_{i,j,l \in X} f(i,j)f(i,l)f(k,j)f(k,l) \leq \left( \sum_{i,j \in X} f(i,j)^2 \right)^{2/3}.
\]

The hypothesis that each \( \ell \in [L] \) lies in the image of at least two of the injections is necessary — for example, we do not always have

\[
\left( \sum_{i,j \in X} f(i)f(j) \right)^2 \leq \left( \sum_{i \in X} f(i)^2 \right)^2.
\]

**Proof.** We proceed by induction on \( m \). When \( m = 2 \), each \( \ell \in [L] \) must lie in the image of both \( \sigma_1 \) and \( \sigma_2 \), so \( a_1 = a_2 = L \), and both \( \sigma_1 \) and \( \sigma_2 \) are permutations of \([L]\). We must therefore show that if \( f_1, f_2 : X^L \to \mathbb{R} \), and \( \sigma_1, \sigma_2 \in S_L \), then

\[
\left( \sum_{i_1, \ldots, i_L \in X} f_1(i_{\sigma_1(1)}, \ldots, i_{\sigma_1(L)})f_2(i_{\sigma_2(1)}, \ldots, i_{\sigma_2(L)}) \right)^2 \leq \prod_{j=1}^2 \left( \sum_{i_1, \ldots, i_L \in X} f_j(i_1, \ldots, i_L)^2 \right).
\]

This is an immediate consequence of the Cauchy-Schwarz inequality.

Now let \( m > 2 \), and suppose that the statement of the theorem holds for all smaller values of \( m \). By replacing each \( f_j \) by \( |f_j| \) if necessary, we may assume that \( f_j : X^{a_j} \to \mathbb{R}_{\geq 0} \) for each \( j \in [m] \). Without loss of generality, we may assume that \( \sigma_1 = \text{Id}_{[a_1]} \). We must bound:

\[
Q := \left( \sum_{i_1, \ldots, i_L \in X} \prod_{j=1}^m f_j(i_{\sigma_j(1)}, \ldots, i_{\sigma_j(a_j)}) \right)^2.
\]

Applying the Cauchy-Schwarz inequality, we obtain

\[
Q \leq \sum_{i_1, \ldots, i_{a_1} \in X} f_1(i_1, \ldots, i_{a_1})^2 \sum_{i_1, \ldots, i_{a_1} \in X} \left( \sum_{i_{a_1+1}, \ldots, i_L \in X} \prod_{j=2}^m f_j(i_{\sigma_j(1)}, \ldots, i_{\sigma_j(a_j)}) \right)^2. \tag{8}
\]

Let

\[
Q' = \sum_{i_1, \ldots, i_{a_1} \in X} \left( \sum_{i_{a_1+1}, \ldots, i_L \in X} \prod_{j=2}^m f_j(i_{\sigma_j(1)}, \ldots, i_{\sigma_j(a_j)}) \right)^2.
\]
Observe that some of \{1, 2, \ldots, a_j\} may appear in the image of just one of the injections \(\sigma_2, \ldots, \sigma_m\). Without loss of generality, we may assume that \(i_1, \ldots, i_M\) appear in just one of \(\sigma_2, \ldots, \sigma_m\), and \(i_M+1, \ldots, i_a\) appear in at least two. (Note that we may have \(M = 0\).) Since each \(f_j\) is non-negative, by the convexity of \(y \mapsto y^2\), we have

\[
Q' = \sum_{i_1, \ldots, i_M \in X} \left( \sum_{i_{M+1}, \ldots, i_a \in X} \prod_{j=2}^m f_j(i_{\sigma_j(1)}, \ldots, i_{\sigma_j(a_j)}) \right)^2
\]
\[
\leq \sum_{i_1, \ldots, i_M \in X} \left( \sum_{i_{M+1}, \ldots, i_a \in X} \prod_{j=2}^m f_j(i_{\sigma_j(1)}, \ldots, i_{\sigma_j(a_j)}) \right)^2.
\] (9)

For each fixed \((i_1, \ldots, i_M) \in X^M\), we now apply the induction hypothesis to

\[
\left( \sum_{i_{M+1}, \ldots, i_a \in X} \prod_{j=2}^m f_j(i_{\sigma_j(1)}, \ldots, i_{\sigma_j(a_j)}) \right)^2.
\]

Without loss of generality, we may assume that for each \(j \in \{2, \ldots, m\}\), we have \(\sigma_j(k) \notin [M]\) for \(k \leq c_j\), and \(\sigma_j(k) \in [M]\) for each \(k > c_j\). When we apply the induction hypothesis to the above expression, we think of \(\sigma_j\) as an injection from \([c_j]\) to \([L] \setminus [M]\), suppressing occurrences of \(i_1, \ldots, i_M\). We obtain:

\[
\left( \sum_{i_{M+1}, \ldots, i_a \in X} \prod_{j=2}^m f_j(i_{\sigma_j(1)}, \ldots, i_{\sigma_j(a_j)}) \right)^2
\]
\[
\leq \prod_{j=2}^m \left( \sum_{p_1, \ldots, p_{c_j} \in X} f_j(p_1, \ldots, p_{c_j}, i_{\sigma_j(c_j+1)}, \ldots, i_{\sigma_j(a_j)})^2 \right).
\] (10)

Since the elements

\((i_{\sigma_j(k)} : c_j + 1 \leq k \leq a_j, \ 2 \leq j \leq m)\)

are simply a reordering of the (fixed) sequence \((i_1, \ldots, i_M) \in X^M\), substituting (10) into (9) and relabelling yields

\[
Q' \leq \prod_{j=2}^m \left( \sum_{p_1, \ldots, p_{a_j} \in X} f_j(p_1, \ldots, p_{a_j})^2 \right).
\]

Combining this with (8) yields

\[
Q \leq \prod_{j=1}^m \left( \sum_{p_1, \ldots, p_{a_j} \in X} f_j(p_1, \ldots, p_{a_j})^2 \right),
\]

completing the proof.
4 The quasi-stability result

Our aim in this section is to prove the following ‘quasi-stability’ result.

**Theorem 11.** For each $t \in \mathbb{N}$, there exists $C_t > 0$ such that the following holds. Let $A \subset S_n$ with $|A| = c(n-t)!$, and let $f = 1_A : S_n \to \{0,1\}$ be the characteristic function of $A$, so that $E[f] = c/n$. If $E[(f - f_t)^2] \leq \epsilon c/(n_t)$, where $\epsilon \leq 1/2$, then there exists a Boolean function $F$ such that

$$E[(f - F)^2] \leq (100c^{1/2} + C_t c/\sqrt{n})c/(n_t),$$

and $F$ is the characteristic function of a union of round($c$) $t$-cosets of $S_n$. Moreover, we have $|c - \text{round}(c)| \leq (32c^{1/2} + C_t c/\sqrt{n})c$.

**Proof overview.** The proof proceeds as follows. Instead of working with $f_t$, the projection of $f$ onto $U_t$, it will be more convenient to work with two different functions, which are both close to $f_t$ in $L^2$-norm. Namely, we will let

$$g_t = f_t - f_{t-1};$$

note that $g_t$ is the orthogonal projection of $f$ onto $V_t$, as $V_t$ is the orthogonal complement of $U_{t-1}$ in $U_t$. We will see that if $c = o(n)$, then $\|f_{t-1}\|^2_2 = o(||f||^2_2)$, i.e. the projection of $f$ onto $U_{t-1}$ is small, so $g_t$ is close to $f_t$ in $L^2$-norm.

For each $t$-coset $T$, we define

$$a_T = |A \cap T|/(n - t)!;$$

note that $a_T = (n)_t(f, 1_T) = E[f|_T]$, the expectation of $f$ restricted to $T$. Our aim is to show that there are at approximately $c$ $t$-cosets $T$ such that $a_T$ is close to 1. However, there are no simple relationships between the $a_T$'s and the moments of $f_t$, so instead, we work with the quantities

$$b_T := (n)_t(g_t, 1_T) = E[g_t|_T].$$

We define

$$h_t := \sum_T b_T 1_T.$$ 

Because $h_t$ is in $V_t$ (as opposed to $f_t$, which is in $U_t$ but not in $V_t$), there is a close relationship between $E[h_t^2]$ and $\sum_T b_T^2$, and between $E[h_t^4]$ and $\sum_T b_T^4$. Moreover, we will also see that $h_t$ is close to $g_t$ (and therefore to $f$) in $L^2$-norm, so $E[h_t^2] \approx c/(n_t)$. This, together with the relationship between $E[h_t^4]$ and $\sum_T b_T^4$, will imply that $\sum_T b_T^2 \approx c$.

Using the fact that $E[(h_t - f)^2]$ is small, together with the fact that $f$ is a Boolean function with expectation $c/(n_t)$, we will show that $E[h_t^2] \approx c/(n_t)$. This, together with the relationship between $\sum_T b_T^2$ and $E[h_t^2]$, will imply that $\sum_T b_T^2 \approx c$. As long as $c = o(n)$, we will have $|b_T - a_T| = o(1)$ for each $T$, and therefore $b_T \leq 1 + o(1)$. Hence, the only way we can have $\sum_T b_T^2 \approx c$ and $\sum_T b_T^4 \approx c$ is if approximately $c$ of the $b_T$'s are close to 1. This in turn implies that the corresponding $a_T$'s are close to 1, completing the proof.
Note that this proof relies upon analysing fourth moments, as opposed to our proof of the $t = 1$ case in [12], which relied upon analysing the third moments of a non-negative function. For $t > 1$, there is no simple relationship between $h_t$ and the non-negative function

$$v_t = \sum_T a_T 1_T,$$

which is the natural analogue of non-negative function we used in the $t = 1$ case. The relationship between $\mathbb{E}[h_t^3]$ and $\mathbb{E}[v_t^3]$ seems too hard to analyse for general $t$, and we were unable to find another non-negative function whose third moment was simply related to that of $h_t$. This led us to consider $\mathbb{E}[h_t^4]$ instead; being non-negative, this is easier to bound from below than $\mathbb{E}[h_t^3]$. The price we pay is that the relationship between $\mathbb{E}[h_t^4]$ and $\sum_T b_T^4$ is considerably more complicated than that between $\mathbb{E}[h_t^3]$ and $\sum_T b_T^3$; we make do with an approximate relationship, as opposed to the exact one in [12].

We now begin our formal proof.

**Proof of Theorem 11**

Let $A, f$ satisfy the hypotheses of Theorem 11. For each $t$-coset $T$, define

$$a_T = \frac{|A \cap T|}{(n-t)!}.$$

Our ultimate aim is to show that approximately $c$ of the $a_T$'s are close to 1. Observe that

$$a_T = (n)_t \langle f, 1_T \rangle = \langle f_t, 1_T \rangle \forall T \in \mathcal{C}(n, t),$$

since $1_T \in U_t$.

Define

$$g_t = f_t - f_{t-1};$$

note that $g_t$ is the orthogonal projection of $f$ onto $V_t$. Instead of working with the $a_T$'s, for most of the proof we will work with the quantities

$$b_T := (n)_t(g_t, 1_T) = \mathbb{E}[g_t|_T].$$

By Lemma 9, we have

$$||f_{t-1}\|_2^2 \leq C_t n^{t-1} (c/(n)_t)^2 = O_t(c/n) c/(n)_t = O_t(c/n)\mathbb{E}[f^2],$$

so if $c = o(n)$, $f_t$ is close to $g_t := f_t - f_{t-1}$ in $L^2$-norm, as desired.

It follows that $b_T$ is close to $a_T$, for each $t$-coset $T$. Indeed, we have

$$|a_T - b_T| = (n)_t |\langle f_{t-1}, 1_T \rangle|$$

$$\leq (n)_t ||f_{t-1}||_2 ||1_T||_2$$

$$\leq (n)_t \sqrt{C_t n^{t-1} c/(n)_t} \sqrt{1/(n)_t}$$

$$\leq O_t(c/\sqrt{n}),$$

$$17$$
using the Cauchy-Schwarz inequality and (11).

Since \( a_T \in [0, 1] \), it follows that
\[
-\mathcal{O}_t(c/\sqrt{n}) \leq b_T \leq 1 + \mathcal{O}_t(c/\sqrt{n}) \quad \forall T \in \mathcal{C}(n, t),
\]
so \( b_T \) cannot lie too far outside the interval \([0, 1]\).

We now define the function
\[
h_t = \sum_{T \in \mathcal{C}(n, t)} b_T 1_T = (n)_t \sum_{T \in \mathcal{C}(n, t)} \langle g, 1_T \rangle 1_T;
\]
this function is the crucial one in our proof. Our next step is to show that \( h_t \) must be close to \( g_t \). To this end, we define the linear operator
\[
M : \mathbb{C}[S_n] \to \mathbb{C}[S_n];
\]
\[
g \mapsto (n)_t \sum_{T \in \mathcal{C}(n, t)} \langle g, 1_T \rangle 1_T.
\]
(Note that \( h_t = Mg_t \).) We now prove the following.

**Lemma 12.** The eigenspaces of \( M \) are precisely the \( U_\alpha \). The operator \( M|_{V_t} \) is an invertible endomorphism of \( V_t \), and all its eigenvalues are \( 1 + \mathcal{O}_t(1/n) \).

**Proof.** For each \( \sigma \in S_n \), let us write \( e_\sigma \) for the function on \( S_n \) which is 1 at \( \sigma \) and 0 elsewhere. With slight abuse of notation, we use \( (M(\sigma, \pi))_t \) to denote the matrix of \( M \) with respect to the standard basis \( \{ e_\sigma : \sigma \in S_n \} \) of \( \mathbb{C}[S_n] \). Observe that
\[
M(\sigma, \pi) = M(e_\pi)(\sigma)
\]
\[
= (n)_t \sum_{T \in \mathcal{C}(n, t)} \frac{1}{n!} 1\{ \pi \in T \} 1\{ \sigma \in T \}
\]
\[
= \frac{(n)_t}{n!} |\{ T \in \mathcal{C}(n, t) : \sigma, \pi \in T \}|
\]
\[
= \frac{(n)_t}{n!} |\{ x \in [n]^t : \sigma(i) = \pi(i) \forall i \in x \}|
\]
\[
= \frac{(n)_t}{n!t!} |\{ (i_1, \ldots, i_t) \in [n]^t : i_1, \ldots, i_t \text{ are all distinct, } \sigma(i_k) = \pi(i_k) \forall k \in [t] \}|
\]
\[
= \frac{1}{n!} \binom{n}{t} \xi_{(n-t, 1^t)}(\sigma^{-1})
\]
Define
\[
w_\sigma = \frac{1}{n!} \binom{n}{t} \xi_{(n-t, 1^t)}(\sigma);
\]
observe that \( \sigma \mapsto w_\sigma \) is a class function, and that \( M_{\sigma, \pi} = w(\sigma \pi^{-1}) \) for all \( \sigma, \pi \in S_n \).
By (6), the eigenvalues of $M$ are given by

$$
\lambda_\alpha = \frac{(n)}{n!f_\alpha} \sum_{\sigma \in S_n} \xi_{n-t,1'}(\sigma) \chi_\alpha(\sigma) \\
= \frac{(n)}{f_\alpha} (\xi_{n-t,1'}, \chi_\alpha) \\
= \frac{(n)}{f_\alpha} K_{\alpha,n-t,1'} \ (\alpha \vdash n),
$$

(14)

with the $U_\alpha$ being the corresponding eigenspaces.

In particular, it follows that if $g \in V_\ell = \bigoplus_{\alpha \vdash n: \alpha_1 = n-t} U_\alpha$, then $Mg \in V_\ell$ also, so $M|_{V_\ell}$ is a linear endomorphism of $V_\ell$. Since $K_{\alpha,n-t,1'} \geq 1$ for all $\alpha \geq (n-t,1')$, we have $\lambda_\alpha > 0$ for each $\alpha \vdash n$ with $\alpha_1 = n-t$, so $M|_{V_\ell}$ is invertible. Combining Lemma 7 and (14), we see that for each $\alpha \vdash n$ with $\alpha_1 = n-t$, we have $\lambda_\alpha = 1 + O_t(1/n)$, completing the proof.

Since $g_t \in V_\ell$, it follows that $h_t = Mg_t \in V_\ell$. Moreover,

$$
||h_t - g_t||_2 = ||Mg_t - g_t||_2 \leq O_t(1/n)||g_t||_2 < O_t(1/n)\sqrt{\epsilon/(n)t}.
$$

(15)

Combining (11) and (15), using the triangle inequality, yields

$$
||h_t - f||_2 \leq ||h_t - g_t||_2 + ||g_t - f_t||_2 + ||f_t - f||_2 \\
\leq O_t(1/n)\sqrt{\epsilon/(n)t} + O_t(\sqrt{\epsilon/n})\sqrt{c/(n)t} + \sqrt{c/(n)t} \\
= (O_t(1/n) + O_t(\sqrt{\epsilon/n}) + \sqrt{c/(n)t})\sqrt{c/(n)t} \\
= \psi\sqrt{c/(n)t} \\
= \psi||f||_2,
$$

(16)

where $\psi := \sqrt{\epsilon} + O_t(1/n) + O_t(\sqrt{\epsilon/n})$.

Moreover, by Lemma 12, we have

$$
|||h_t||_2 - ||g_t||_2| \leq ||h_t - g_t||_2 \leq O_t(1/n)||g_t||_2 < O_t(1/n)\sqrt{\epsilon/(n)t},
$$

(17)

and therefore

$$
\mathbb{E}[h_t^2] \leq (1 + O_t(1/n))c/(n)t.
$$

(18)

By (16), we have

$$
||h_t||_2 \geq (1 - \psi)||f||_2,
$$

and therefore

$$
\mathbb{E}[h_t^2] \geq (1 - \psi)^2 c/(n)t \geq (1 - 2\psi)c/(n)t,
$$

(19)

so $\mathbb{E}[h_t^2]$ is close to $c/(n)t$.

We can simplify our expression for $\psi$ by noting that $c$ cannot be too small. Indeed, we have

$$
||f_t||_2^2 \geq \frac{1}{2}c/(n)t,
$$

19
whereas by Lemma 9, we have
\[ ||f_t||_2^2 \leq C_{t+1}(n)_t(c/(n)_t)^2. \]

It follows that
\[ c \geq 1/(2C_{t+1}). \]

Therefore, we can absorb the \( O_t(1/n) \) term in the \( O_t(\sqrt{c/n}) \) term in our expression for \( \psi \), giving \( \psi = \sqrt{\epsilon} + O_t(\sqrt{c/n}) \).

We have proved the following.

**Proposition 13.**
\[ (1 - 2\sqrt{\epsilon} - O_t(\sqrt{c/n})c/(n)_t \leq E[h_t^2] \leq (1 + O_t(1/n))c/(n)_t. \]

Our next step is to obtain a lower bound on \( E[h_t^4] \), using (16). To this end, we prove the following.

**Lemma 14.** Suppose \( F: [0, 1] \to \{0, 1\} \) is a Lebesgue measurable function with \( E[F] = \theta \), and \( h: [0, 1] \to \mathbb{R} \) is a measurable function such \( E[(h - F)^2] \leq \eta \). Then, provided \( \eta \leq \theta(1 - \theta) \), we have
\[ E[h^4] \geq \theta - 8\sqrt{\theta(1 - \theta)}\eta. \]

**Proof.** We will solve the following optimization problem.

**Problem P.** Among all measurable functions \( h: [0, 1] \to \mathbb{R} \) such that \( E[h] = E[F] \) and \( E[(h - F)^2] \leq \eta \), find the minimum value of \( E[h^4] \).

Let \( A = F^{-1}([1]) \), and let \( B = F^{-1}([0]) \); then \( A \) and \( B \) are measurable sets, and \( \lambda(A) = \theta \). Observe that if \( h: [0, 1] \to \mathbb{R}_{\geq 0} \) is feasible for \( P \), then the function
\[ \tilde{h}(x) = \begin{cases} \frac{1}{\theta} \int_A h(x) \, dx & \text{if } x \in A, \\ \frac{1}{1-\theta} \int_B h(x) \, dx & \text{if } \theta \leq x \in B, \end{cases} \]

obtained by averaging \( h \) first over \( A \) and then over \( B \), is also feasible. Indeed, we clearly have \( E[\tilde{h}] = E[h] \), and
\[ E[(h - F)^2] = \int_A (h(x) - 1)^2 \, dx + \int_B h(x)^2 \, dx \]
\[ \geq \frac{1}{\theta} \left( \int_A (h(x) - 1) \, dx \right)^2 + \frac{1}{1-\theta} \left( \int_B h(x) \, dx \right)^2 \]
\[ = \theta \left( \frac{1}{\theta} \int_A h(x) \, dx - 1 \right)^2 + (1 - \theta) \left( \frac{1}{1-\theta} \int_B h(x) \, dx \right)^2 \]
\[ = E[(\tilde{h} - F)^2], \]
by the Cauchy-Schwarz inequality. Moreover, we have

\[
E[h^4] = \int_A h(x)^4 \, dx + \int_B h(x)^4 \, dx
= \theta \cdot \frac{1}{\theta} \int_A h(x)^4 \, dx + (1 - \theta) \cdot \frac{1}{1 - \theta} \int_B h(x)^4 \, dx
\geq \theta \left( \frac{1}{\theta} \int_A h(x) \, dx \right)^4 + (1 - \theta) \left( \frac{1}{1 - \theta} \int_B h(x) \, dx \right)^4
= E[\tilde{h}^4],
\]

by the convexity of \( y \mapsto y^4 \). Hence, replacing \( h \) with \( \tilde{h} \) if necessary, we may assume that \( h \) is constant on \( A \) and on \( B \). In other words, we may assume that \( h \) has the following form:

\[
h(x) = \begin{cases} r & \text{if } x \in A, \\ s & \text{if } x \in B. \end{cases}
\]

Therefore, \( P \) is equivalent to the following problem:

**Problem Q:**

Minimize \( \theta r^4 + (1 - \theta)s^4 \)

subject to \( \theta r + (1 - \theta)s = \theta, \quad \theta(r - 1)^2 + (1 - \theta)s^2 \leq \eta, \quad r, s \geq 0. \)

Or, writing \( r = 1 - (1 - \theta)\delta \) (so that \( s = \theta\delta \), we obtain the following reformulation:

**Problem Q’:**

Minimize \( \theta(1 - (1 - \theta)\delta)^4 + (1 - \theta)(\theta\delta)^4 \)

subject to \( \theta(1 - \theta)\delta^2 \leq \eta, \quad 0 \leq \delta \leq 1/(1 - \theta). \)

When \( \delta = 1 \), the function \( g \) is constant. By the strict convexity of the function \( y \mapsto y^4 \), the objective function is strictly decreasing on \([0, 1]\) as a function of \( \delta \).\(^1\) Hence, provided \( \sqrt{\eta/(\theta(1 - \theta))} \leq 1 \), the minimum is attained at \( \delta = \sqrt{\eta/(\theta(1 - \theta))} \), at which point the value of the objective function is

\[
\theta - 4\sqrt{\theta(1 - \theta)\eta} + 6(1 - \theta)\eta - 4(1 - \theta)(\eta/\theta)\sqrt{\theta(1 - \theta)\eta} + \frac{(1 - \theta)^3 + \theta^3}{\theta(1 - \theta)}\eta^2
\geq \theta - 8\sqrt{\theta(1 - \theta)\eta}.
\]

\(^1\) Alternatively, this can be seen by considering the derivative of the objective function, which is

\[-4\theta(1 - \theta)(1 - (1 - \theta)\delta)^3 + 4\theta^4(1 - \theta)^3 \delta < 0 \forall \delta \in [0, 1].\]
We apply this to functions on the discrete probability space \((S_n, \mathcal{F})\), by considering them as step functions on \([0, 1]\). Applying Lemma 14 with \(h = h_t, F = f, \theta = c/(n)_t\) and \(\eta = \psi^2c/(n)_t\), we obtain
\[
\mathbb{E}[h_t^2] \geq (1 - 8\sqrt{1 - c/(n)_t})(1 - 8\psi)c/(n)_t.
\]
(20)

The next lemma relates \(\mathbb{E}[h_t^2]\) to \(\sum_{T \in \mathcal{C}(n, t)} b_T^2\).

**Lemma 15.** Suppose \(h \in V_t\) with
\[
h = \sum_{T \in \mathcal{C}(n, t)} b_T 1_T.
\]
Then
\[
\mathbb{E}[h^2] = (1 + O_t(1/n)) \frac{1}{(n)_t} \sum_{T \in \mathcal{C}(n, t)} b_T^2.
\]

**Proof.** Let us abbreviate \(\sum_{T \in \mathcal{C}(n, t)}\) to \(\sum_T\). We expand as follows:
\[
\mathbb{E}[h^2] = \sum_T b_T^2 \mathbb{E}[1_T] + \sum_{(S, T) : S \neq T} b_S b_T \mathbb{E}[1_S 1_T] = \frac{1}{(n)_t} \sum_T b_T^2 + \sum_{(S, T) : S \neq T} b_S b_T \mathbb{E}[1_S \cap T].
\]
We must prove that
\[
\sum_{(S, T) : S \neq T} b_S b_T \mathbb{E}[1_S \cap T] = O_t(1/(n)_{t+1}) \sum_T b_T^2.
\]
(21)
We pause to define some more notation. We define
\[
\mathcal{D}(n, t) := \{(x_1, y_1), \ldots, (x_t, y_t)\} : (x_1, \ldots, x_t) \in ([n])_t, (y_1, \ldots, y_t) \in ([n])_t;
\]
note that there is a natural one-to-one correspondence between \(\mathcal{D}(n, t)\) and \(\mathcal{C}(n, t)\), the set of \(t\)-cosets, given by identifying the \(t\)-coset
\[
\{\sigma \in S_n : \sigma(x_i) = y_i \forall i \in [t]\}\end{equation}
with the set of ordered pairs
\[
\{(x_1, y_1), \ldots, (x_t, y_t)\} \in \mathcal{D}(n, t).
\]
We denote this correspondence by \(\leftrightarrow\). If \(T \leftrightarrow I\), we define \(b_T := b_T\).

We say that two sets \(\{(u_1, v_1), \ldots, (u_s, v_s)\} \in \mathcal{D}(n, s), \{(x_1, y_1), \ldots, (x_t, y_t)\} \in \mathcal{D}(n, t)\) are compatible if the corresponding cosets have nonempty intersection, i.e. if \(u_i = x_j \leftrightarrow v_i = y_j\). Otherwise, we say that they are incompatible.

We say that the two sets are independent if \(u_i \neq x_j\) and \(v_i \neq y_j\) for all \(i \in [s], j \in [t]\). Observe that if \(A, B \in \mathcal{D}(n, t)\) are compatible, then we may express
\[
A = I \cup J, B = I \cup K,
\]
where \( I \in \mathcal{D}(n,e) \) for some \( e \leq t \), \( J, K \in \mathcal{D}(n,t-e) \), and \( I, J \) and \( K \) are pairwise independent.

We pause to observe some crucial linear dependencies between the \( b_T \)'s. Recall from Lemma 12 that \( M|_{V_t} \) is invertible. Let \( g = M^{-1} h \); then \( g \in V_t \), and \( b_T = \langle g, 1_T \rangle \) for each \( T \in \mathcal{C}(n,t) \). We claim that

\[
\sum_{k \neq x_1, \ldots, x_{t-1}} b_{\{x_1, y_1\}, \ldots, (x_{t-1}, y_{t-1}), (k, y_t)} = 0 \tag{22}
\]

for any distinct \( x_1, \ldots, x_{t-1} \in [n] \), and any distinct \( y_1, \ldots, y_t \in [n] \). Indeed, the left-hand side is precisely \( \langle g, 1_S \rangle \), where

\[
\mathcal{C}(n,t-1) \ni S = \{ \sigma \in S_n : \sigma(x_i) = y_i \; \forall i \in [t-1] \} \leftrightarrow \{(x_1, y_1), \ldots, (x_{t-1}, y_{t-1})\}.
\]

Since \( g \in V_t \perp U_{t-1} = \text{Span}\{1_S : S \in \mathcal{C}(n,t-1)\} \), the claim follows.

Similarly, we claim that

\[
\sum_{k \neq y_1, \ldots, y_{t-1}} b_{\{x_1, y_1\}, \ldots, (x_{t-1}, y_{t-1}), (x_t, k)} = 0, \tag{23}
\]

for any distinct \( x_1, \ldots, x_t \in [n] \), and any distinct \( y_1, \ldots, y_{t-1} \in [n] \). Indeed, the left-hand side is precisely \( \langle g, 1_S \rangle \), where \( S \) is as above.

We split up the sum in (21) as follows:

\[
\sum_{S \cap T \neq \emptyset} b_S b_T \mathbb{1}[S \supseteq T] = \sum_{e=0}^{t-1} \frac{1}{(n)_{2t-e}} \sum_{J,K \in \mathcal{D}(n,t-e)} \sum_{I, J, K \text{ independent}} b_{I \cup J} b_{I \cup K}.
\]

We now use an argument similar to the indicator-function proof of inclusion-exclusion. Observe that for any \( I \in \mathcal{D}(n,e) \), we have

\[
\sum_{J,K \in \mathcal{D}(n,t-e)} b_{I \cup J} b_{I \cup K} = \sum_{J,K \in \mathcal{D}(n,t-e)} b_{I \cup J} b_{I \cup K} \prod_{p,q \in [t-e], d \in [2]} (1 - 1\{j_p^{(d)} = k_q^{(d)}\}),
\]

where

\[
J = \{(j_1^{(1)}, j_1^{(2)}), (j_2^{(1)}, j_2^{(2)}), \ldots, (j_{t-e}^{(1)}, j_{t-e}^{(2)})\},
\]

\[
K = \{(k_1^{(1)}, k_1^{(2)}), (k_2^{(1)}, k_2^{(2)}), \ldots, (k_{t-e}^{(1)}, k_{t-e}^{(2)})\}.
\]

Consider what happens when we multiply out the product on the right-hand side, à la inclusion-exclusion. We obtain a \((\pm 1)\)-linear combination of products of indicators of the form

\[
\prod_{(p,q,d) \in \mathcal{T}} 1\{j_p^{(d)} = k_q^{(d)}\}.
\]

For each such (non-zero) product of indicators, let us replace \( k_q^{(d)} \) by \( j_p^{(d)} \) in the sum whenever the product contains the indicator \( 1\{j_p^{(d)} = k_q^{(d)}\} \). We obtain a \((\pm 1)\)-linear combination of terms of the form

\[
\sum_{J, K', I, J \text{ indep.}, K \text{ indep.}} b_{I \cup J} b_{I \cup K'},
\]

23
where \( K' \) is obtained from \( K \) by replacing \( k_q^{(d)} \) with \( j_p^{(d)} \) for various \((p,q,d)\). If \( K' \) contains any (unreplaced) \( k_q^{(d)} \), then the corresponding term is zero, by applying (22) or (23) with \( k = k_q^{(d)} \). Hence, the only non-zero terms are where all the \( k_q^{(d)} \)’s have been replaced, i.e. they are of the form

\[
\sum_{j \in D(n,t-e), \ i,j \text{ indep.}} b_{i,j} b_{j,i} \{j_1^{(1)},j_2^{(2)},\ldots,(j_{t-e}^{(1)},j_{t-e}^{(2)})\},
\]

where \( \sigma \in S_{t-e} \). We now sum each such term over \( I \), and apply the Cauchy-Schwarz inequality:

\[
\left| \sum_{i \in D(n,e), \ i,j \text{ indep.}} b_{i,j} h_{j,i} \{j_1^{(1)},j_2^{(2)},\ldots,(j_{t-e}^{(1)},j_{t-e}^{(2)})\} \right| \leq \binom{t}{e} \sum_T b_T^2.
\]

Here, the factor of \( \binom{t}{e} \) comes from the fact that there are \( \binom{t}{e} \) ways of producing an ordered partition \( (I,J) \) of a \( t \)-set into an \( e \)-set \( I \) and a \((t-e)\)-set \( J \).

There are at most \( O_t(1) \) terms of the form (24) for each \( e \in [t-1] \), so we obtain

\[
\sum_{e=0}^{t-1} \frac{1}{(n)^{2t-e}} \sum_{i \in D(n,e), \ i,j \text{ indep.}} b_{i,j} b_{j,i} \{j_1^{(1)},j_2^{(2)},\ldots,(j_{t-e}^{(1)},j_{t-e}^{(2)})\} \leq O_t(1) \frac{1}{(n)^{t+1}} \sum_{T \in C(n,t)} b_T^2,
\]

as required. \( \square \)

The next lemma relates \( \mathbb{E}[h_T^4] \) to \( \sum_{T \in C(n,t)} b_T^4 \).

**Lemma 16.** Suppose \( h \in V_t \) with

\[
h = \sum_{T \in C(n,t)} b_T 1_T.
\]

Then

\[
\mathbb{E}[h^4] = \frac{1}{(n)^t} \left( \sum_{T \in C(n,t)} b_T^4 + O_t(1/n) \left( \sum_{T \in C(n,t)} b_T^2 \right)^2 \right).
\]

**Proof.** We need one more piece of notation. If \( X \) is a finite set, and \( Z \in X^{(2)} \) is a set of ordered pairs of elements of \( X \), we let \( Z(d) \) denote the set of all elements of \( X \) which occur in the \( d \)th coordinate of an ordered pair in \( Z \), for \( d \in \{1,2\} \).

We expand as follows:

\[
\mathbb{E}[h^4] = \sum_T b_T^4 \mathbb{E}[1_T] + \sum_{(T_1,T_2,T_3,T_4) \text{ not all equal}} b_{T_1} b_{T_2} b_{T_3} b_{T_4} \mathbb{E}[1_{T_1} 1_{T_2} 1_{T_3} 1_{T_4}]
\]

\[
= \frac{1}{(n)^t} \sum_T b_T^4 + \sum_{(T_1,T_2,T_3,T_4) \text{ not all equal}} b_{T_1} b_{T_2} b_{T_3} b_{T_4} \mathbb{E}[1_{T_1 \cap T_2 \cap T_3 \cap T_4}],
\]

24
We must prove that
\[
\sum_{(T_1, T_2, T_3, T_4) \text{ not all equal: } T_1 \cap T_2 \cap T_3 \cap T_4 \neq \emptyset} b_{T_1} b_{T_2} b_{T_3} b_{T_4} \mathbb{E}[1_{T_1 \cap T_2 \cap T_3 \cap T_4}] \leq O_t(1/(n)_{t+1}) \left( \sum_T b_T^2 \right)^2 .
\]

We now split up the sum above, by partitioning the set
\[
\{(T_1, T_2, T_3, T_4) \text{ not all equal: } T_1 \cap T_2 \cap T_3 \cap T_4 \neq \emptyset \}
\]
into a bounded number of classes. We define an equivalence relation \( \sim \) on this set by
\[
(T_1', T_2', T_3', T_4') \sim (T_1, T_2, T_3, T_4) \iff \text{there exist } \sigma, \pi \in S_n : \sigma T_i \pi = T_r \forall r \in [4].
\]
Note that the number of equivalence classes is at most \( \binom{4}{t}^4 = O_t(1) \), since we can choose a representative from each equivalence class which is a 4-tuple of \( t \)-cosets, each of which is the point-wise stabiliser of some \( t \)-element subset of \([4t] \).

Note also that \( \mathbb{E}[1_{T_1 \cap T_2 \cap T_3 \cap T_4}] \) depends only upon the equivalence class of \((T_1, T_2, T_3, T_4)\), and that we always have \( \mathbb{E}[1_{T_1 \cap T_2 \cap T_3 \cap T_4}] = 1/(n)_t \), where \( t+1 \leq l \leq 4t \), and therefore we always have \( \mathbb{E}[1_{T_1 \cap T_2 \cap T_3 \cap T_4}] \leq 1/(n)_{t+1} \).

Let \( S \) be an equivalence class; we wish to bound
\[
\left| \sum_{(T_1, T_2, T_3, T_4) \in S} b_{T_1} b_{T_2} b_{T_3} b_{T_4} \mathbb{E}[1_{T_1 \cap T_2 \cap T_3 \cap T_4}] \right|.
\]
Since we always have \( \mathbb{E}[1_{T_1 \cap T_2 \cap T_3 \cap T_4}] \leq 1/(n)_{t+1} \), we have
\[
\left| \sum_{(T_1, T_2, T_3, T_4) \in S} b_{T_1} b_{T_2} b_{T_3} b_{T_4} \mathbb{E}[1_{T_1 \cap T_2 \cap T_3 \cap T_4}] \right| \leq \frac{1}{(n)_{t+1}} \left| \sum_{(T_1, T_2, T_3, T_4) \in S} b_{T_1} b_{T_2} b_{T_3} b_{T_4} \right| .
\]

Take any \((T_1, T_2, T_3, T_4) \in S\), and let \( T_r \leftrightarrow Z_r \in \mathcal{D}(n, t) \) for each \( r \in [4] \). Since \( T_1 \cap T_2 \cap T_3 \cap T_4 \neq \emptyset \), \( Z_1, Z_2, Z_3, Z_4 \) are pairwise compatible, so for any \( \{u_i, v_i\} \in [4] \) \( \in Z_r \), \( \{x_i, y_i\} \in [4] \) \( \in Z_t \), we have \( u_i = x_j \Leftrightarrow v_i = y_j \) for all \( i, j \in [t] \). It follows that \( \bigcup_{r=1}^{4} Z_r(1) = \bigcup_{r=1}^{4} Z_r(2) \); we denote this number by \( N \). Note that \( N \leq 4t \).

Observe that we may write
\[
\sum_{(T_1, T_2, T_3, T_4) \in S} b_{T_1} b_{T_2} b_{T_3} b_{T_4} = \sum_{(i_1, \ldots, i_N) \text{ distinct, } (i_{N+1}, \ldots, i_{2N}) \text{ distinct}} \prod_{r=1}^{4} b_{(i_{p \cdot r, i_{p \cdot r}})}(p, q) \in A_r ,
\]
where \( A_r \subset [N] \times \{N+1, \ldots, 2N\} \) with \( |A_r| = |A_r(1)| = |A_r(2)| = t \), for each \( r \in [4] \).
(For each \( p \in [2N] \), we shall say that \( p \) is ‘good’ if it appears in at least two of the \( A_r \)-s; otherwise, we say that \( p \) is ‘bad’. Our aim is to use indicator functions (à la inclusion-exclusion), together with the linear dependence relations (22) and (23),

\[
(25)
\]
to rewrite (26) as a \((\pm 1)\)-linear combination of at most \(O_t(1)\) sums containing no ‘bad’ indices. Without loss of generality, we may assume that there is a bad index in \([N+1, \ldots, 2N]\), say \(2N\). (The argument for a bad index in \([N]\) is identical, except that (22) is used instead of (23).) Without loss of generality, we may assume that \(2N\) appears in \(A_4\) alone. Let

\[ P = \{ p \in \{N + 1, \ldots, 2N\} : p \text{ appears in } A_4 \}, \quad Q = \{N + 1, \ldots, 2N\} \setminus P. \]

Then

\[
\sum_{\substack{(i_1, \ldots, i_N) \text{ distinct}, \\ (i_{N+1}, \ldots, i_{2N}) \text{ distinct}}} 4 \prod_{r=1}^4 b_{\{(i_p, i_q)\}}(p, q)_{\in A_r} = \sum_{q \in Q} \sum_{\substack{(i_1, \ldots, i_N) \text{ distinct}, \\ (i_{N+1}, \ldots, i_{2N-1}) \text{ distinct}}} 4 \prod_{r=1}^4 b_{\{(i_p, i_q)\}}(p, q)_{\in A_r} \prod_{q \in Q} (1 - 1\{i_{2N} = i_q\}).
\]

Consider what happens when we expand out the product over all \(q \in Q\). We obtain

\[
\sum_{\substack{(i_1, \ldots, i_N) \text{ distinct}, \\ (i_{N+1}, \cdots, i_{2N-1}) \text{ distinct}}} 4 \prod_{r=1}^4 b_{\{(i_p, i_q)\}}(p, q)_{\in A_r} - \sum_{q \in Q} \sum_{\substack{(i_1, \ldots, i_N) \text{ distinct}, \\ (i_{N+1}, \cdots, i_{2N-1}) \text{ distinct}}} b_{\{(i_p, i_q)\}}(p, q)_{\in A_4(2N \to q)} \prod_{r=1}^3 b_{\{(i_p, i_q)\}}(p, q)_{\in A_r},
\]

where \(A_4(2N \to q)\) is produced from \(A_4\) by replacing \(2N\) with \(q\). We have

\[
\sum_{\substack{(i_1, \ldots, i_N) \text{ distinct}, \\ (i_{N+1}, \cdots, i_{2N-1}) \text{ distinct}}} 4 \prod_{r=1}^4 b_{\{(i_p, i_q)\}}(p, q)_{\in A_r} = 0,
\]

by (23), so the first term above is zero; none of the other terms involve the ‘bad’ index \(2N\). Hence, we have

\[
\sum_{\substack{(i_1, \ldots, i_N) \text{ distinct}, \\ (i_{N+1}, \cdots, i_{2N}) \text{ distinct}}} 4 \prod_{r=1}^4 b_{\{(i_p, i_q)\}}(p, q)_{\in A_r} = - \sum_{q \in Q} \sum_{\substack{(i_1, \ldots, i_N) \text{ distinct}, \\ (i_{N+1}, \cdots, i_{2N-1}) \text{ distinct}}} b_{\{(i_p, i_q)\}}(p, q)_{\in A_4(2N \to q)} \prod_{r=1}^3 b_{\{(i_p, i_q)\}}(p, q)_{\in A_r}
\]

We now have \(|Q|\) sums to deal with, but each has one less bad index than the original sum. By repeating this process until there are no more bad indices remaining
in any sum, we can express the original sum (26) as a \((-1\)-linear combination of at most \(O_t(1)\) terms of the form:

\[
\sum_{(i_1, \ldots, i_K) \in \mathcal{R}} \prod_{r=1}^{4} b_{(i_p, i_q)}(p, q) \in \mathcal{A}_r, 
\]

where

- \(K \leq 2N \leq 8t\),
- \(\mathcal{R} \subset [n]^K\) is a subset defined by constraints of the form \(i_k \neq i_l\),
- For each \(r \in [4]\), we have \(\mathcal{A}_r \subset [K]^2\), and each \(k \in [K]\) occurs in at most one ordered pair in \(\mathcal{A}_r\),
- Each index \(k \in [K]\) is ‘good’, meaning that it appears in at least two of the \(\mathcal{A}_r\)’s.

Note that the \(\mathcal{A}_r\)’s vary with the term we are looking at. Crudely, (27) is at most

\[
\sum_{(i_1, \ldots, i_K) \in [n]^K} \prod_{r=1}^{4} |b_{(i_p, i_q)}(p, q) \in \mathcal{A}_r|, 
\]

where we define \(b_{(i_k, j_k)}(k \in [t]) = 0\) if \(\{i_k, j_k\} \notin \mathcal{D}(n, t)\), i.e. if \(i_k = i_l\) or \(j_k = j_l\) for some \(k \neq l\).

To bound (28), we apply Lemma 10. We shall apply it to the function \(F: [n]^{2t} \to \mathbb{R}\) defined by

\[
F(i_1, \ldots, i_t, j_1, \ldots, j_t) = \begin{cases} 
\sum_{|k|} b_{(i_k, j_k)}(k \in [t]) & \text{if } i_1, \ldots, i_k \text{ are all distinct} \\
0 & \text{and } j_1, \ldots, j_t \text{ are all distinct, otherwise.}
\end{cases}
\]

For each \(r \in [4]\), choose any ordering \(((p_1, q_1), \ldots, (p_t, q_t))\) of the ordered pairs in \(\mathcal{A}_r\), and define the injection \(\sigma_r: [2t] \to [K]\) by

\[
\sigma_r(w) = \begin{cases} 
(p_w) & \text{if } 1 \leq w \leq t, \\
(q_{w-t}) & \text{if } t+1 \leq w \leq 2t.
\end{cases}
\]

Applying Lemma 10 with \(m = 4\), \(L = K\), \(X = [n]\), \(a_j = 2t\) for each \(j \in [4]\), \(f_j = F\) for each \(j \in [4]\), and \(\sigma_1, \sigma_2, \sigma_3, \sigma_4\) as above, yields

\[
\left| \sum_{(i_1, \ldots, i_K) \in [n]^K} \prod_{r=1}^{4} |b_{(i_p, i_q)}(p, q) \in \mathcal{A}_r| \right| \leq \left( t! \sum_T b_T^2 \right)^2.
\]

Here, the factor of \(t!\) comes from the fact that we are summing over all possible orderings of the ordered pairs in each \(Z \in \mathcal{D}(n, t)\). Since there are only \(O_t(1)\) terms of the form (27), we have

\[
\sum_{(T_1, T_2, T_3, T_4) \in S} b_{T_1} b_{T_2} b_{T_3} b_{T_4} \leq O_t(1) \left( \sum_T b_T^2 \right)^2.
\]
Using (25), together with the fact that there are only \(O(t)\) different equivalent classes, yields

\[
\sum_{(T_1,T_2,T_3,T_4) \text{ not all equal: } T_1 \cap T_2 \cap T_3 \cap T_4 \neq \emptyset} b_{T_1} b_{T_2} b_{T_3} b_{T_4} E[1_{T_1 \cap T_2 \cap T_3 \cap T_4}] \leq O_t(1/(n+1)) \left( \sum_T b_T^2 \right)^2,
\]

completing the proof of Lemma 16.

Combining Proposition 13 and Lemma 15 shows that \(\sum_T b_T^2\) is close to \(c\):

**Proposition 17.**

\[
(1 - 2\psi - O_t(1/n))c \leq \sum_T b_T^2 \leq (1 + O_t(1/n))c,
\]

where \(\psi = \sqrt{\epsilon} + O_t(\sqrt{c/n})\).

Similarly, combining Lemma 16 and (20) yields

\[
\sum_T b_T^4 + O_t(1/n) \left( \sum_T b_T^2 \right)^2 > (1 - 8\psi)c.
\]

Using the fact that \(\sum_T b_T^2 = \Theta(c)\) yields

\[
\sum_T b_T^4 > (1 - 8\psi)c - O_t(c^2/n) = (1 - 8\psi - O_t(c/n))c = (1 - 8\psi')c,
\]

where

\[
\psi' := \psi + O_t(c/n) = \sqrt{\epsilon} + O_t(\sqrt{c/n}).
\]

Using (13) and Proposition 17, we have

\[
\sum_T b_T^4 \leq (1 + O_t(c/\sqrt{n})) \sum_T b_T^2 \\
\leq (1 + O_t(c/\sqrt{n}))(1 + O_t(1/n))c \\
= (1 + O_t(c/\sqrt{n}))c.
\]

Combining this with (29) shows that \(\sum_T b_T^4\) is close to \(c\):

**Proposition 18.**

\[
(1 - 8\psi')c \leq \sum_T b_T^4 \leq (1 + O_t(c/\sqrt{n}))c,
\]

where \(\psi' = \sqrt{\epsilon} + O_t(\sqrt{c/n})\).
Let \( x_1, \ldots, x_N \) denote the entries \((b^2_T)_{T \in C(n,t)}\) in non-increasing order. By Proposition 17, we have
\[
\sum_{k=1}^{N} x_k \leq (1 + O_t(1/n))c, \tag{30}
\]
and by Proposition 18, we have
\[
\sum_{k=1}^{N} x_k^2 \geq c(1 - 8\psi'). \tag{31}
\]
Subtracting (31) from (30) yields:
\[
\sum_{k=1}^{N} x_k (1 - x_k) \leq c(8 \psi' + O_t(1/n)) = 8c\psi'', \tag{32}
\]
where \( \psi'' := \psi' + O_t(1/n) \). By an appropriate choice of the \( O_t(1/n) \) term in the definition of \( \psi'' \), we may ensure that
\[
\sum_{k=1}^{N} x_k \geq (1 - 2\psi'')c,
\]
by Proposition 17.
Let \( m \) be the largest index \( k \) such that \( x_k \geq 1/2 \) (recall that the \( x_k \) are arranged in non-increasing order). Then
\[
\sum_{k=m+1}^{N} x_k \leq 2 \sum_{k=m+1}^{N} x_k (1 - x_k) \leq 16c\psi'',
\]
by (32). Therefore,
\[
(1 + O_t(c/\sqrt{m}))m \geq \sum_{k=1}^{m} x_k \geq c(1 - 18\psi''). \tag{33}
\]
On the other hand, we have
\[
\sum_{k=1}^{m} (1 - x_k) \leq 2 \sum_{k=1}^{m} x_k (1 - x_k) \leq 16c\psi''.
\]
Rearranging,
\[
\sum_{k=1}^{m} x_k \geq m - 16c\psi''. \tag{34}
\]
Since \( 2x_k - 1 \leq x_k^2 \), we have
\[
(1 + O_t(c/\sqrt{m}))c \geq \sum_{k=1}^{m} x_k^2 \geq 2 \sum_{k=1}^{m} x_k - m \geq m - 32c\psi''. \tag{35}
\]
Combining (33) and (35) yields:

$$|m - c| \leq (32\epsilon^{1/2} + O_t(c/\sqrt{n}))c. \quad (36)$$

Let $$T_1, \ldots, T_m$$ be the $$t$$-cosets corresponding to $$x_1, \ldots, x_m$$, and let

$$C' = \bigcup_{k=1}^{m} T_i$$

denote the corresponding union of $$m$$ $$t$$-cosets of $$S_n$$. We have

$$\sum_{k=1}^{m} |A \cap T_k|/(n-t)! = \sum_{k=1}^{m} a_{T_i} \geq \sum_{k=1}^{m} b_{T_i} - mO_t(c/\sqrt{n})$$

$$\geq \frac{1}{1 + O_t(c/\sqrt{n})} \sum_{k=1}^{m} b_{T_i}^2 - O_t(c^2/\sqrt{n})$$

$$\geq (1 - 18\psi'' - O_t(c/\sqrt{n}))c$$

$$\geq (1 - 18\epsilon^{1/2} - O_t(c/\sqrt{n}))c,$$

using (12), (13), and (33). Since $$|T_i \cap T_j| \leq (n-t-1)!$$ for each $$i \neq j$$, we have

$$|A \cap C'| \geq \sum_{k=1}^{m} |A \cap T_k| - \binom{m}{2} (n-t-1)! \geq (1 - 18\epsilon^{1/2} - O_t(c/\sqrt{n}))(n-t)!,$$

i.e. $$A$$ contains almost all of $$C'$$. Since $$|A| = c(n-t)!$$ and $$|C'| \leq c(n-t)!$$, we must have

$$|A \Delta C'| = |A| + |C'| - 2|A \cap C'| \leq (36\epsilon^{1/2} + O(c/\sqrt{n}))(n-t)!.$$

Crudely, we have

$$|m - \text{round}(c)| \leq 2|m - c| \leq (64\epsilon^{1/2} + O_t(c/\sqrt{n}))c.$$  

By adding or deleting $$|m - \text{round}(c)|$$ $$t$$-cosets to or from $$C'$$, we may produce a family $$C \subset S_n$$ which is a union of $$\text{round}(c)$$ $$t$$-cosets, and satisfies

$$|A \Delta C| \leq |A \Delta C'| + |C' \Delta C|$$

$$\leq (36\epsilon^{1/2} + O(c/\sqrt{n}))c(n-t)! + (64\epsilon^{1/2} + O_t(c/\sqrt{n}))c(n-t)!$$

$$= (100\epsilon^{1/2} + O_t(c/\sqrt{n}))c(n-t)!.$$

Let $$F = 1_C$$ denote the characteristic function of $$C$$; then we have

$$\mathbb{E}[(f - F)^2] = |A \Delta C|/n! \leq (100\epsilon^{1/2} + O_t(c/\sqrt{n}))c/(n).$$

Since $$|c - \text{round}(c)| \leq |m - c| \leq (32\epsilon^{1/2} + O_t(c/\sqrt{n}))c$$, this completes the proof of Theorem 11.
5 An isoperimetric inequality in the transposition graph

In this section, we will apply Theorem 11 to obtain an isoperimetric inequality for $S_n$. We first give some background and notation on discrete isoperimetric inequalities.

Isoperimetric problems are of ancient interest in mathematics. In general, they ask for the smallest possible ‘boundary’ of a set of a certain ‘size’. Discrete isoperimetric inequalities deal with discrete notions of boundary in graphs. There are two different notions of boundary in graphs, the vertex-boundary and the edge-boundary; here, we deal with the latter.

If $G = (V, E)$ is any graph, and $S, T \subseteq V$, we write $E_G(S, T)$ for the set of edges of $G$ between $S$ and $T$, and we write $e_G(S, T) = |E_G(S, T)|$. We write $\partial_G S = E_G(S, S^c)$ for the set of edges of $G$ between $S$ and its complement; this is called the edge-boundary of $S$ in $G$. An edge-isoperimetric inequality for $G$ gives a lower bound on the minimum size of the edge-boundary of a set of size $k$, for each integer $k$.

The transposition graph $T_n$ is the Cayley graph on $S_n$ generated by the transpositions in $S_n$; equivalently, two permutations are joined if, as sequences, one can be obtained from the other by transposing two elements. In this section, we are concerned with the edge-isoperimetric problem for $T_n$.

It would be of great interest to prove an isoperimetric inequality for the transposition graph which is sharp for all set-sizes. Ben Efraim [3] conjectures that initial segments of the lexicographic order on $S_n$ have the smallest edge-boundary of all sets of the same size. (The lexicographic order on $S_n$ is defined as follows: if $\sigma, \pi \in S_n$, we say that $\sigma < \pi$ if $\sigma(j) < \pi(j)$, where $j = \min\{i \in [n] : \sigma(i) \neq \pi(i)\}$. The initial segment of size $k$ of the lexicographic order on $S_n$ simply means the smallest $k$ elements of $S_n$ in the lexicographic order.)

**Conjecture 19** (Ben Efraim). For any $A \subseteq S_n$, $|\partial A| \geq |\partial C|$, where $C$ denotes the initial segment of the lexicographic order on $S_n$ of size $|A|$.

This is a beautiful conjecture; it may be compared to the edge-isoperimetric inequality in $\{0, 1\}^n$, due to Harper [17], Lindsey [24], Bernstein [4] and Hart [18], stating that among all subsets of $\{0, 1\}^n$ of size $k$, the first $k$ elements of the binary ordering on $\{0, 1\}^n$ has the smallest edge boundary. (Recall that if $x, y \in \{0, 1\}^n$, we say that $x < y$ in the binary ordering if $x_j = 0$ and $y_j = 1$, where $j = \min\{i \in [n] : x_i \neq y_i\}$.)

In this section, we use eigenvalue techniques to prove an approximate version of Conjecture 19; this version is asymptotically sharp for sets of size $n!/\text{poly}(n)$. We then combine eigenvalue techniques with Theorem 11 to prove the exact conjecture for sets of size $(n-t)!$, for $n$ sufficiently large depending on $t$.

For sets of size $c(n-1)!$, (where $c \in [n]$), Conjecture 19 follows from calculating the second eigenvalue of the Laplacian of $T_n$. We briefly outline the argument.

If $G = (V, E)$ is a finite graph, and $A$ is the adjacency matrix of $G$, the Laplacian matrix of $G$ may be defined by

$$L = D - A,$$

where $D$ is the diagonal $|V| \times |V|$ matrix with rows and columns indexed by $V$, and
with
\[ D_{u,v} = \begin{cases} \deg(v) & \text{if } u = v \\ 0 & \text{if } u \neq v. \end{cases} \]

The following well-known theorem, due independently to Dodziuk [9] and Alon and Milman [2], provides an edge-isoperimetric inequality for a graph \( G \) in terms of the smallest non-trivial eigenvalue of its Laplacian matrix (the ‘spectral gap’):

**Theorem 20** (Dodziuk / Alon-Milman). If \( G = (V,E) \) is any graph, \( L \) is the Laplacian matrix of \( G \), and \( 0 = \mu_1 \leq \mu_2 \leq \ldots \leq \mu_{|V|} \) are the eigenvalues of \( G \) (repeated with their multiplicities), then for any set \( S \subset V \),
\[ e(S,S^c) \geq \mu_2 \frac{|S||S^c|}{|V|}. \]
If equality holds, then the characteristic vector \( 1_S \) of \( S \) satisfies
\[ 1_S - \frac{|S|}{|G|} f \in \ker(L - \mu_2 I), \quad (37) \]
where \( f \) denotes the all-1’s vector.

If \( G \) is a \( d \)-regular graph (note that \( T_n \) is \( \binom{n}{2} \)-regular), then the Laplacian matrix is given by \( L = dI - A \). Therefore, if the eigenvalues of the adjacency matrix are
\[ d = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{|V|}, \]
then \( \mu_i = d - \lambda_i \) for each \( i \). In particular, \( \mu_2 = d - \lambda_2 \), so for any set \( S \subset V(G) \),
\[ e(S,S^c) \geq (d - \lambda_2) \frac{|S||S^c|}{|V|}. \]

The transposition graph is a normal Cayley graph, and therefore its eigenvalues are given by (7). Frobenius gave the following formula for the value of \( \chi_\alpha \) at a transposition, where \( \alpha = (\alpha_1, \ldots, \alpha_m) \) is a partition of \( n \).
\[ \chi_\alpha((1 2)) = \dim(\rho_\alpha) \frac{1}{2} \sum_{j=1}^{m} ((\alpha_j - j)(\alpha_j - j + 1) - j(j - 1)) \quad (\alpha \vdash n). \]
Combining this with (7) yields the following formula for the eigenvalues of the transposition graph.
\[ \lambda_\alpha = \frac{1}{2} \sum_{j=1}^{m} ((\alpha_j - j)(\alpha_j - j + 1) - j(j - 1)) \quad (\alpha \vdash n). \quad (38) \]
Note that \( \lambda_{(n)} = \binom{n}{2} \), \( \lambda_{(n-1,1)} = \binom{n}{2} - n \) and \( \lambda_{(n-2,2)} = \binom{n}{2} - 2n + 2 \). Diaconis and Shahshahani [8] verify that if \( \alpha \) and \( \alpha' \) are two partitions of \( n \), then
\[ \alpha \triangleright \alpha' \quad \Rightarrow \quad \lambda_\alpha \geq \lambda_{\alpha'}. \quad (39) \]
Hence, if \( n \geq 2 \), then \( \mu_2 = n \), and the \( \mu_2 \)-eigenspace of the Laplacian is precisely \( U_{(n-1,1)} \). Theorem 20 therefore yields the following.
**Theorem 21** (essentially due to Diaconis, Shahshahani). If $\mathcal{A} \subset S_n$, then

$$|\partial \mathcal{A}| \geq \frac{|\mathcal{A}|(n! - |\mathcal{A}|)}{(n-1)!}.$$  

Equality holds only if

$1_{\mathcal{A}} \in U(n) \oplus U(n-1,1)$ \hspace{1cm} (40)

Observe that equality holds in Theorem 21 if and only if $\mathcal{A}$ is a disjoint union of 1-cosets of $S_n$. (The ‘only if’ part follows from (40) and the $t=1$ case of Theorem 1.)

When $|\mathcal{A}| = o(n!)$, Theorem 21 merely implies $|\partial \mathcal{A}| \geq (1-o(1))n|\mathcal{A}|$, whereas Conjecture 19 would imply that $|\partial \mathcal{A}| \geq (1-o(1))n(t+1)|\mathcal{A}|$ whenever $|\mathcal{A}| = o((n-t)!)$. Our aim is to prove the latter when $t$ is small. We prove the following approximate isoperimetric inequality.

**Theorem 22.** Let $\mathcal{A} \subset S_n$ with $|\mathcal{A}| \leq (n-t+1)!$. Then

$$|\partial \mathcal{A}| \geq (1 - C_t|\mathcal{A}|/(n-t+1)!)(t(n-1)|\mathcal{A}|),$$

where $C_t > 0$ depends upon $t$ alone.

**Proof.** Let $f = 1_\mathcal{A}$ and let $L$ be the Laplacian of $T_n$. By Lemma 9, we have

$$\|f_{t-1}\|^2 \leq C_t n^{t-1}(|\mathcal{A}|/n!)^2 \leq O_t(|\mathcal{A}|/(n-t+1)!)|\mathcal{A}|/n!.$$  

Therefore, we have

$$\sum_{\alpha \subset \mathcal{A}} \|f_{\alpha}\|^2 = \|f\|^2 - \|f_{t-1}\|^2 \geq (1 - O_t(|\mathcal{A}|/(n-t+1)!))|\mathcal{A}|/n!.$$  

Note that if $\alpha_1 \leq n - t$, then $\alpha \leq (n-t,t)$, and therefore $\lambda_\alpha \leq \lambda_{(n-t,t)}$, by (39). By (38), we have

$$\lambda_{(n-t,t)} = \binom{n}{2} - tn + t(t-1),$$

so $\lambda_\alpha \leq \binom{n}{2} - tn + t(t-1)$ whenever $\alpha_1 \leq n - t$. In other words, $\mu_\alpha \geq tn - t(t-1)$ whenever $\alpha_1 \leq n - t$. We have

$$|\partial \mathcal{A}| = n! \langle f, Lf \rangle$$

$$= n! \sum_{\alpha \subset \mathcal{A}} \mu_\alpha \|f_{\alpha}\|^2$$

$$\geq n! \sum_{\alpha_1 \leq n-t} \mu_\alpha \|f_{\alpha}\|^2$$

$$\geq n!(tn - t(t-1))(1 - O_t(|\mathcal{A}|/(n-t+1)!))|\mathcal{A}|/n!$$

$$\geq (1 - O_t(|\mathcal{A}|/(n-t+1)!))t(n-1)|\mathcal{A}|,$$

as required. $\square$
Note that when $t$ is fixed and $\mathcal{A} \subset S_n$ with $|\mathcal{A}| = o((n-t+1)!)$, Theorem 22 yields

$$|\partial \mathcal{A}| \geq (1 - o(1))t(n-1)(n-t)!.$$ 

If 

$$\mathcal{A} = \{\sigma \in S_n : \sigma(i) = i \ \forall i \in [t-1], \ \sigma(t) \in \{t, \ldots, t+m-1\}\},$$

then $|\mathcal{A}| = m(n-t)!$ and

$$|\partial \mathcal{A}| = (t(n-1) - m + 1)|\mathcal{A}|.$$ 

If $m = o(n)$, then we obtain

$$|\partial \mathcal{A}| = (1 - o(1))t(n-1)|\mathcal{A}|,$$

showing that Theorem 22 is asymptotically sharp in this case.

We now use a similar argument, together with Theorem 11, to prove Conjecture 19 for sets of size $(n-t)!$, when $n$ is sufficiently large depending on $t$.

**Theorem 23.** Let $\mathcal{A} \subset S_n$ with $|\mathcal{A}| = (n-t)!$. If $n$ is sufficiently large depending on $t$, then

$$|\partial \mathcal{A}| \geq |\partial T_{1-1,2-2,\ldots,t-t}|.$$ 

Equality holds if and only if $\mathcal{A}$ is a $t$-coset of $S_n$.

**Proof.** Let $\mathcal{A} \subset S_n$ with $|\mathcal{A}| = (n-t)!$. Note that

$$|\partial T_{1-1,2-2,\ldots,t-t}| \leq t(n-1)(n-t)!.$$ 

Assume that $|\partial \mathcal{A}| \leq |\partial T_{1-1,2-2,\ldots,t-t}|$; we will show that $\mathcal{A}$ must be a $t$-coset.

Let $f = 1_\mathcal{A}$; then $||f||_2^2 = 1/(n!)t$. Write

$$||f - f_i||_2^2 = \phi||f||_2^2 = \phi|\mathcal{A}|/n!,$$

where $\phi \in [0,1]$. Our first aim is to show that $\phi$ must be small.

By Lemma 9, we have

$$||f_{t-1}||_2^2 \leq Cn^{t-1}(|\mathcal{A}|/n!)^2 = O_t(1/n)|\mathcal{A}|/n!.$$ 

Therefore, we have

$$\sum_{a_1=\ldots=a_{n-t}} ||f_{a_1}||_2^2 = ||f||_2^2 - ||f_{t-1}||_2^2 = (1 - \phi - O_t(1/n)|\mathcal{A}|/n!.$$ 

Using $L$ for the Laplacian of $T_n$, we have

$$|\partial \mathcal{A}| = n!(f,Lf)$$

$$= n! \sum_{a=\ldots=1} \mu_a ||f_{a}||_2^2$$

$$\geq n! \sum_{a_1=\ldots=a_{n-t}} \mu_a ||f_{a_1}||_2^2 + \sum_{a_1=\ldots=a_{n-t-1}} \mu_a ||f_{a_1}||_2^2$$

$$\geq n!(tn - t(t-1))(1 - \phi - O_t(1/n)|\mathcal{A}|/n! + n!(t+1)n - t(t+1))\phi|\mathcal{A}|/n!$$

$$= (tn - t(t-1))(1 - \phi - O_t(1/n)) + ((t+1)n - t(t+1))\phi(n-t)!$$

$$= (t(n-t+1) + \phi n - O_t(1))(n-t)!.$$
Since we are assuming that $|\partial A| \leq t(n-1)(n-t)!$, it follows that $\phi \leq O_t(1/n)$. Hence,
we may apply Theorem 11 with $\epsilon = O_t(1/n)$. We see that there exists a $t$-coset $C \subset S_n$
such that
\[ |A \triangle C| \leq O_t(1/\sqrt{n})(n-t)! |. \]

Our aim is to show that $A = C$. Let us write
\[ |A \setminus C| = \psi(n-t)!, \quad \psi = O_t(1/\sqrt{n}). \]

Let $E = A \setminus C$, and let $M = C \setminus A$; then
\[ |E| = |M| = \psi(n-t)!. \]

Let $X = A \cap C$ and $Y = S_n \setminus (A \cup C)$, so that $A = E \cup X$ and $C = M \cup Y$.

Observe that
\[ |\partial C| + |\partial E| - 2e(C, E) - e(M, S_n \setminus C) + e(M, X) + 2e(E, M) \]
\[ = e(C, E) + e(C, Y) + e(E, \partial Y) + e(\partial Y, \partial Y) - 2e(C, E) - [e(M, E) + e(M, Y)] + e(M, X) + 2e(E, M) \]
\[ = e(C, Y) + e(E, \partial Y) - e(M, Y) + e(\partial Y, M) + e(E, M) \]
\[ = e(A, Y) + e(A, M) = |\partial A|. \]

In particular,
\[ |\partial A| \geq |\partial C| + |\partial E| - 2e(C, E) - e(M, S_n \setminus C) + e(M, X). \] (41)

We first bound $e(C, E)$. By definition, we have $E \cap C = \emptyset$. Without loss of generality,
we may assume that $C = \{ \sigma \in S_n : \sigma(i) = i \text{ for all } i \in [t] \}$. For any $\sigma \in S_n \setminus C$,
choose $i \in [t]$ such that $\sigma(i) \neq i$. If $\sigma(i \sigma^{-1}(i)) \in C$, then it is the unique neighbour of $\sigma$ in $C$;
otherwise, $\sigma$ has no neighbour in $C$. It follows that
\[ e(C, E) \leq |E| = \psi(n-t)!. \]

Next, we bound $|\partial E|$. Since $|E| \leq (n-t)!$, Theorem 22 yields
\[ |\partial E| \geq (1 - O_t(1/n))t(n-1)|E| = (1 - O_t(1/n))t(n-1)\psi(n-t)!. \]

Next, we calculate $e(M, S_n \setminus C)$. Observe that each $\sigma \in C$ has exactly $t(n-1)$
neighbours in $S_n \setminus C$. Indeed, the neighbours of $\sigma$ in $S_n \setminus C$ are precisely $\{ \sigma(i j) : i \in [t], j \neq i \}$. It follows that
\[ e(M, S_n \setminus C) = (t(n-1))|M|. \]

Finally, we bound $e(M, X)$. Since $T_n[C]$ is isomorphic to $T_{n-t}$, applying Theorem 21 to $T_{n-t}$ yields:
\[ e(C \setminus A, A \cap C) \geq \psi(n-t)!/(1 - \psi)(n-t)!/(n-t-1)! = \psi(1 - \psi)(n-t)(n-t)!. \]
Substituting all of these bounds into (41), we obtain:

\[
|\partial A| \geq |\partial C| + (1 - O_t(1/n)) t(n - 1) \psi(n - t)!
- 2\psi(n - t)! - (t(n - 1)) \psi(n - t)! + \psi(1 - \psi)(n - t)(n - t)!
= |\partial C| + [(1 - O_t(1/n)) t(n - 1) \psi
- 2\psi \psi + \psi(1 - \psi)(n - t)](n - t)!
= |\partial C| + \psi((1 - O_t(1/\sqrt{n}))(n - t) - O_t(1))(n - t)!
\]

If \( \psi > 0 \) and \( n \) is sufficiently large depending on \( t \), then the right-hand side is greater than \( |\partial C| \), contradicting our assumption that \( |\partial A| \leq |\partial C| \). It follows that \( \psi = 0 \) and therefore \( A = C \), proving Theorem 23.

6 A note on \( t \)-intersecting families of permutations

In [14], we showed how a \( t = 1 \) version of Theorem 11 could be used to give a natural proof of the Cameron-Ku conjecture on large intersecting families of permutations. Let us briefly give the background to this result. Recall that a family of permutations \( A \subset S_n \) is said to be intersecting if for any two permutations \( \sigma, \pi \in A \), there exists \( i \in [n] \) such that \( \sigma(i) = \pi(i) \). Using a simple partitioning argument, Deza and Frankl proved that if \( A \subset S_n \) is intersecting, then \( |A| \leq (n - 1)! \), i.e. the 1-cosets are intersecting families of maximum size. Cameron and Ku [6] proved that equality holds only if \( A \) is a 1-coset of \( S_n \). They made the following ‘stability’ conjecture.

Conjecture 24 (Cameron-Ku, 2003). There exists \( c > 0 \) such that if \( A \subset S_n \) is an intersecting family with \( |A| \geq c(n - 1)! \), then \( A \) is contained within a 1-coset.

Conjecture 24 was first proved by the first author in [10], but the proof in [12] is, in a sense, more natural.

We say that a family of permutations \( A \subset S_n \) is \( t \)-intersecting if for any \( \sigma, \pi \in A \), there exist \( i_1, \ldots, i_t \in [n] \) such that \( \sigma(i_k) = \pi(i_k) \) for all \( k \in [t] \). In 1977, Deza and Frankl [7] conjectured that if \( n \) is sufficiently large depending on \( t \), then the maximum-sized \( t \)-intersecting families in \( S_n \) are precisely the \( t \)-cosets. This is proved in [14]. In [11], the first author proved the following analogue of Conjecture 24 for \( t \)-intersecting families of permutations.

Theorem 25. For any \( t \in \mathbb{N} \), and any \( c > 1 - 1/e \), there exists \( n_0(t, c) \in \mathbb{N} \) such that if \( n \geq n_0 \), then any \( t \)-intersecting family \( A \subset S_n \) with \( |A| \geq c(n - t)! \) is contained within a \( t \)-coset.

Almost exactly as in the \( t = 1 \) case, Theorem 11 can be used to give a more ‘natural’ proof of Theorem 25. We do not give the details here; it suffices to say that one simply replaces adjacency matrix of the derangement graph (used in the \( t = 1 \) case) with the ‘weighted’ analogue constructed in [14].
7 Conclusion and open problems

As mentioned in [13], we conjecture the following strengthening of Theorem 11.

**Conjecture 26.** Let $A \subset S_n$, and let $t \in \mathbb{N}$. Let $f$ denote the characteristic function of $A$, and let $f_t$ denote the orthogonal projection of $f$ onto $U_t$. If

$$\mathbb{E}[(f - f_t)^2] \leq \epsilon \mathbb{E}[f],$$

then there exists a family $B \subset S_n$ which is a union of $t$-cosets, such that

$$|A \triangle B| \leq C_0 \epsilon |A|,$$

where $C_0$ is an absolute constant.

In section 5, we gave an example of how Fourier-analytic arguments can yield sharp isoperimetric inequalities for relatively small sets, even when the classical ‘eigenvalue gap’ inequality in Theorem 20 is far from sharp. We speculate that this technique of combining eigenvalue arguments, Fourier-analytic arguments and stability arguments may be useful in obtaining other isoperimetric inequalities. Conjecture 19 remains very much open, however; we suspect that more combinatorial techniques are required to prove it in full generality.

**Acknowledgement**

The authors wish to thank Gil Kalai for many useful discussions.

**References**


