

# A stability result for balanced dictatorships in $S_n$

David Ellis\*, Yuval Filmus† and Ehud Friedgut‡

May 2013

## Abstract

We prove that a balanced Boolean function on  $S_n$  whose Fourier transform is highly concentrated on the first two irreducible representations of  $S_n$ , is close in structure to a dictatorship, a function which is determined by the image or pre-image of a single element. As a corollary, we obtain a stability result concerning extremal isoperimetric sets in the Cayley graph on  $S_n$  generated by the transpositions.

Our proof works in the case where the expectation of the function is bounded away from 0 and 1. In contrast, [7] deals with Boolean functions of expectation  $O(1/n)$  whose Fourier transform is highly concentrated on the first two irreducible representations of  $S_n$ . These need not be close to dictatorships; rather, they must be close to a union of a constant number of cosets of point-stabilizers.

**Keywords:** Fourier transform, stability, symmetric group.

## 1 Introduction

### 1.1 Background

This paper (together with [7] and [8]) is part of a trilogy dealing with stability and ‘quasi-stability’ results concerning Boolean functions on the symmetric group, which are of ‘low complexity’, in a Fourier-theoretic sense.

Let us begin with some notation and definitions that will enable us to present the Fourier-theoretic context of our results. Following this, the paper will be essentially Fourier-free, since Lemmas 1 and 2 will translate the relevant Fourier notion into a more combinatorial one.

We write  $[n] := \{1, 2, \dots, n\}$ , and we let  $S_n$  denote the symmetric group on  $[n]$ . If  $i, j \in [n]$ , we write  $T_{ij} := \{\pi \in S_n : \pi(i) = j\}$ . We call the  $T_{ij}$ ’s the *1-cosets* of  $S_n$ , since they are cosets of point-stabilizers. Similarly, for  $t > 1$ , and for two ordered  $t$ -tuples of distinct elements of  $[n]$ ,  $I = (i_1, \dots, i_t)$  and  $J = (j_1, \dots, j_t)$ , we write  $T_{IJ} := \{\pi \in S_n : \pi(I) = J\}$ . We call the  $T_{IJ}$ ’s the *t-cosets* of  $S_n$ . If  $\mathcal{F} \subset S_n$ , we write  $\chi_{\mathcal{F}}$  for its characteristic function, i.e. the Boolean function on  $S_n$  with  $\chi_{\mathcal{F}}(\pi) = 1$  iff  $\pi \in \mathcal{F}$ . Abusing notation slightly, we will often use  $T_{ij}$  and  $T_{IJ}$  to denote their own characteristic functions.

We say that a Boolean function  $f: S_n \rightarrow \{0, 1\}$  is a *dictatorship* if there exists  $i \in [n]$  and  $X \subset [n]$  such that  $f(\pi) = \chi_{\{\pi(i) \in X\}}$  for all  $\pi \in S_n$ , or  $f(\pi) = \chi_{\{\pi^{-1}(i) \in X\}}$  for all  $\pi \in S_n$ , i.e. iff  $f$  is determined by the image or the preimage of just one element of  $[n]$ . It is easy to see that a Boolean function  $f$  on  $S_n$  is a dictatorship if and only if it is the characteristic function of a disjoint union of 1-cosets, i.e. a sum of  $T_{i,j}$ ’s.

If  $f: S_n \rightarrow \mathbb{R}$ , the *Fourier transform* of  $f$  at an irreducible representation  $\rho$  of  $S_n$  is defined by

$$\hat{f}(\rho) = \frac{1}{n!} \sum_{\pi \in S_n} f(\pi) \rho(\pi).$$

---

\*Queen Mary, University of London, UK.

†University of Toronto, Canada. Research supported by the Canadian Friends of the Hebrew University / University of Toronto Permanent Endowment.

‡Weizmann Institute of Science, Israel. Research supported in part by I.S.F. grant 0398246, and BSF grant 2010247.

Recall that the equivalence classes of irreducible representations of  $S_n$  are indexed by partitions of  $n$ . We refer the reader to [18] for background on the representation theory of  $S_n$ , and to [19] for background on the Fourier transform on non-Abelian groups.

For any non-negative integer  $t$ , we let  $U_t$  denote the vector space of real-valued functions on  $S_n$  whose Fourier transform is supported on irreducible representations indexed by partitions of  $n$ , whose largest part has size at least  $n - t$ . (Note that a partition  $\lambda$  of  $n$  has largest part of size at least  $n - t$  if and only if  $\lambda \succeq (n - t, 1^t)$ , where  $\succeq$  denotes the lexicographical order on partitions of  $n$ .) 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 is a measure of the complexity of  $f$ , analogous to the degree of a Boolean function on  $\{0, 1\}^n$ . Indeed, it is precisely the minimum possible total degree of a polynomial in the  $T_{ij}$ 's which is equal to  $f$ .

Note that  $U_0$  is the space of functions whose Fourier transform is supported on the trivial representation — i.e., the space of constant functions. The space  $U_1$ , which is the main subject of this paper, is the space of functions whose Fourier transform is supported on the two irreducible constituents of the permutation representation. As promised, we now de-Fourierize this definition.

First, an easy fact:

**Lemma 1.** *For any  $t \in \mathbb{N}$ , the indicators of  $t$ -cosets (the functions  $T_{IJ}$ ), are in  $U_t$ .*

Next, a slightly more intricate fact, observed and proved in [9]:

**Lemma 2.** *For any  $t \in \mathbb{N}$ , the  $T_{IJ}$ 's span  $U_t$ .*

In this paper, which deals only with  $U_1$ , we will henceforth use the definition  $U_1 = \text{Span}\{T_{ij} : i, j \in [n]\}$ .

Finally, we recall a theorem from [9], which characterizes the Boolean functions in  $U_1$ .

**Theorem 1** (Ellis, Friedgut, Pilpel). *Let  $f : S_n \rightarrow \{0, 1\}$  be in  $U_1$ . Then  $f$  is a dictatorship. (Equivalently,  $f$  is the characteristic function of a disjoint union of 1-cosets.)*

The goal of the current paper, together with [7], is to provide stability versions of Theorem 1. This is in the spirit of similar projects in the Abelian case, which have proved extremely useful and applicable, see e.g. [3], [13], [12], [14], [15] and [17]. The general idea in applications is to prove results in extremal combinatorics using Fourier analysis, and then use the Fourier stability results in order to deduce combinatorial stability results. A good example of this is Theorem 6 in this paper, where we characterize the almost-extremal sets for the edge-isoperimetric inequality in the transposition graph on  $S_n$  (the Cayley graph on  $S_n$  generated by the transpositions). See [7] for more about applications in the symmetric group setting.

We remark that if  $t \geq 2$ , then a Boolean function in  $U_t$  is not necessarily the characteristic function of a union of  $t$ -cosets. Theorem 27 in [9] states that a Boolean function in  $U_t$  is the characteristic function of a disjoint union of  $t$ -cosets, but this is false; a counterexample, and the error in the proof, is pointed out by the second author in [11]. A counterexample when  $t = 2$  is as follows. Let  $n \geq 8$ . For any permutation  $\pi \in S_n$ , define  $x = x(\pi) \in \{0, 1\}^4$  by  $x_i = \chi_{\{\pi(i) \in [4]\}}$ , and consider the function

$$f : S_n \rightarrow \{0, 1\}; \quad \pi \mapsto \chi_{\{x_1 \geq x_2 \geq x_3 \geq x_4 \text{ or } x_1 \leq x_2 \leq x_3 \leq x_4\}}.$$

It can be checked that  $f \in U_2$ , but the value of  $f$  clearly cannot be determined by fixing the images of at most two elements of  $[n]$ , so neither  $f$  nor  $1 - f$  is a union of 2-cosets. It is easy to use  $f$  to construct a counterexample for each  $t \geq 3$ , by considering a product of  $f$  with the characteristic function of the pointwise stabilizer of a  $(t - 2)$ -set. We note that the main application of Theorem 27 in [9] was to characterize (for large  $n$ ) the  $t$ -intersecting families in  $S_n$  of maximum size (i.e., to characterize the cases of equality in the Deza-Frankl conjecture); fortunately, this characterization follows immediately e.g. from the Hilton-Milner type result of the first author in [6], where the proof does not depend on Theorem 27 in [9] (and indeed predates the latter).

The division between the three papers in our trilogy is as follows: in [7], we deal with Boolean functions which are close to  $U_1$ , and have expectation  $O(1/n)$ . We prove that such a functions must be close to a sum of dictatorships — equivalently, close to the characteristic function of a union of 1-cosets. In the current paper, we prove that Boolean functions that are close to  $U_1$ , and whose expectation is bounded away from 0 and 1, must be close to a single dictatorship. Finally, in [8], we deal with Boolean functions close to  $U_t$ , with expectation  $O(n^{-t})$ ; we prove that they must be close to the characteristic

function of a union of  $t$ -cosets. (The latter is perhaps especially interesting, in view of the fact that a Boolean function in  $U_t$  with expectation  $O(n^{-t})$  is not necessarily the characteristic function of a union of  $t$ -cosets.) The term ‘quasi-stability’ in the titles of the other two papers refers to the fact that a Boolean function such as  $T_{11} + T_{22} - T_{11} \cdot T_{22}$  is  $O(1/n^2)$  close to  $U_1$ , and is indeed  $O(1/n^2)$  close to  $T_{11} + T_{22}$ , which is a sum of two dictatorships, but is not  $O(1/n^2)$  close to any single dictatorship. In the case studied in this paper, we have *bona fide* stability, as the functions in question turn out to be close to a dictator. For both ranges of expectation we have studied, however, the Boolean functions which are close to  $U_1$  are close to the characteristic function of a union of 1-cosets. Interestingly, the proof in this paper is quite different from the proof in [7]. It would be interesting to find a common proof that covers the complete spectrum of possible values of  $\mathbb{E}[f]$ .

The outline of the paper is as follows. In the rest of this section, we set up our notation, state our main result, and outline the proof. In section 2, we prove the main theorem for the case of functions with expectation = 1/2. Next, in section 3, we adapt the proof to deal with functions with expectation bounded away from 0 and 1. Finally, in section 4, we give an application of our main theorem: a characterization of the almost-extremal sets for the edge-isoperimetric inequality for the transposition graph on  $S_n$ .

## 1.2 Notation

We now outline our notation systematically. If  $\mathcal{F} \subset S_n$ , we let  $\chi_{\mathcal{F}}$  denote the *characteristic function* of  $\mathcal{F}$ , i.e.  $\chi_{\mathcal{F}} : S_n \rightarrow \{0, 1\}$  with  $\chi_{\mathcal{F}}(\pi) = 1$  iff  $\pi \in \mathcal{F}$ . If  $B$  is a statement, its *indicator*  $\mathbf{1}_B$  is equal to 1 if  $B$  is true and 0 if  $B$  is false.

We write  $[n] := \{1, 2, \dots, n\}$ . Let  $S_n$  denote the symmetric group on  $[n]$ , i.e. the group of all permutations of  $[n]$ . For each  $i, j \in [n]$ , we define

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

to be the set of all permutations sending  $i$  to  $j$ ; we call these the *1-cosets* of  $S_n$ , as they are the cosets of stabilisers of points. Abusing notation slightly, we will also use  $T_{ij}$  to denote its own characteristic vector, more properly written as  $\chi_{T_{ij}}$ .

We define  $U_1(n)$  to be the subspace of  $\mathbb{R}^{S_n}$  spanned by  $\{T_{ij} : i, j \in [n]\}$ . When  $n$  is understood, we abbreviate this to  $U_1$ .

We equip  $\mathbb{R}^{S_n}$  with the inner product induced by the uniform probability measure on  $S_n$ :

$$\langle f, g \rangle = \frac{1}{n!} \sum_{\pi \in S_n} f(\pi)g(\pi).$$

The *expectation* of a real-valued function on  $S_n$  will mean the expectation with respect to the uniform probability measure, i.e.

$$\mathbb{E}[f] = \frac{1}{n!} \sum_{\pi \in S_n} f(\pi).$$

We let  $\|\cdot\|_2$  denote the induced Euclidean norm; i.e.

$$\|f\|_2 = \sqrt{\mathbb{E}[f^2]} = \sqrt{\frac{1}{n!} \sum_{\pi \in S_n} f(\pi)^2}.$$

The distance between functions, or between a function and a subspace, will mean the Euclidean distance as defined by this norm.

Throughout, if  $u$  and  $v$  are functions of several variables, the notation  $u = O(v)$  will mean that there exists an absolute constant  $C$  (not depending upon any of the variables) such that  $|u| \leq C|v|$  pointwise.

The notation  $x \pm \epsilon$  is shorthand for the closed interval  $[x - \epsilon, x + \epsilon]$ . If  $y \in x \pm \epsilon$ , then we say that  $y$  is  $\epsilon$ -close to  $x$ . If  $y \notin x \pm \epsilon$ , then we say that  $y$  is  $\epsilon$ -far from  $x$ . For a set  $S$ , we say that  $x$  is  $\epsilon$ -close to  $S$  if  $|x - y| \leq \epsilon$  for some  $y \in S$ . Otherwise, we say that  $x$  is  $\epsilon$ -far from  $S$ .

We will be dealing throughout with functions on finite probability spaces (i.e., with random variables); we will frequently refer to these simply as ‘functions’ (rather than as ‘random variables’), when the underlying probability space is understood.

### 1.3 Main result

Our main goal in this paper is to prove the following theorem.

**Theorem 2.** *Let  $\mathcal{F} \subset S_n$  be a family of permutations with size  $|\mathcal{F}| = c \cdot n!$ , satisfying*

$$\mathbb{E}[(f - f_1)^2] = \epsilon_1,$$

where  $f = 2\chi_{\mathcal{F}} - 1$ , and  $f_1$  is the orthogonal projection of  $f$  onto  $U_1$ . Then there exists a family  $\mathcal{G} \subset S_n$  which is a union of  $dn$  disjoint 1-cosets, such that

$$|d - c| = O\left(\epsilon_1^{1/7} + \frac{1}{n^{1/3}}\right) \quad \text{and} \quad \frac{|\mathcal{G} \Delta \mathcal{F}|}{n!} = O\left(\frac{1}{\eta} \left(\epsilon_1^{1/7} + \frac{1}{n^{1/3}}\right)\right),$$

where  $\eta = \min\{c, 1 - c\}$ .

**Remark.** *It is convenient to work with the  $\pm 1$ -valued function  $f$ , rather than the  $0/1$ -valued function  $\chi_{\mathcal{F}}$ . Note that if  $(\chi_{\mathcal{F}})_1$  denotes the orthogonal projection of  $\chi_{\mathcal{F}}$  onto  $U_1$ , then the square of the Euclidean distance of  $\chi_{\mathcal{F}}$  from  $U_1$  is*

$$\mathbb{E}[(\chi_{\mathcal{F}} - (\chi_{\mathcal{F}})_1)^2] = \frac{1}{4} \mathbb{E}[(f - f_1)^2].$$

Throughout the proof, a Boolean function will mean a function taking values in  $\{\pm 1\}$ , rather than  $\{0, 1\}$ .

For the entire proof, we will make the assumptions

$$n \geq 4, \tag{1}$$

$$\frac{1}{n^{7/3}} \leq \epsilon_1 < \epsilon_0, \tag{2}$$

where  $\epsilon_0 > 0$  depends only upon  $c$ . Later, we will show how to get rid of these assumptions. During the proof, we will use the phrase *since  $\epsilon_1$  is small enough,  $P$  holds* to mean that for some  $\epsilon_0 > 0$ , the statement  $P$  follows from  $\epsilon_1 < \epsilon_0$ .

For pedagogical reasons, we will first assume that  $c = 1/2$ . This assumption does not affect the proof very much, but it simplifies the expressions appearing therein. After completing the proof in this case, we will show how to extend it to general  $c$ , carefully noting the relation between  $\epsilon_0$  and  $c$ .

### 1.4 Proof overview

We adopt a simple canonical way to express  $f_1$  as a linear combination of the form

$$f_1 = \sum_{i,j} a_{ij} T_{ij};$$

we then study the matrix of coefficients  $(a_{ij})$ . This offers a nice visualization of the function, due to the observation that

$$f_1(\pi) = \sum_{i=1}^n a_{i\pi(i)},$$

i.e.  $f_1(\pi)$  is equal to the sum of the entries on a *generalised diagonal* of the matrix  $(a_{ij})$ . (A generalised diagonal of an  $n \times n$  matrix is a set of  $n$  entries with one entry from each row and one from each column, so corresponds to a permutation of  $\{1, 2, \dots, n\}$ .)

Note that the  $T_{ij}$ 's are linearly dependent (the dimension of  $U_1$  is only  $(n-1)^2 + 1$ , whereas there are  $n^2$  different  $T_{ij}$ 's), so there are many possible ways to represent  $f_1$  in such a manner. It turns out that a particularly useful choice (when  $c = 1/2$ ) is

$$a_{ij} = (n-1) \langle f, T_{ij} \rangle. \tag{3}$$

To illustrate this, here is the matrix corresponding to the dictatorship  $\mathcal{F} = \{\pi \in S_n : 1 \leq \pi(1) \leq n/2\}$  (where  $n$  is even):

$$\begin{pmatrix} \overbrace{1 - \frac{1}{n} \quad \cdots \quad 1 - \frac{1}{n}}^{\frac{n}{2}} & \frac{1}{n} - 1 & \cdots & \frac{1}{n} - 1 \\ -\frac{1}{n} & \cdots & -\frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ -\frac{1}{n} & \cdots & -\frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & & \vdots & \vdots & & \vdots \\ -\frac{1}{n} & \cdots & -\frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}$$

This matrix exemplifies the usefulness of our choice  $a_{ij} = (n-1)\langle f, T_{ij} \rangle$ : the entries that are significant for our dictatorship (row 1, which depends on the image of 1) are all close in absolute value to 1, whereas all other entries are close to 0. The idea of the proof is to discover some properties of the matrix  $(a_{ij})$ , and then show that they imply that it looks roughly like the matrix above — namely, that it has precisely one row or column in which almost half the entries are very close to 1 and almost half the entries are very close to  $-1$ , and that almost all the other entries in the matrix are very close to 0.

The proof breaks down into two main parts. In the first part, we show that for almost all  $\pi \in S_n$ , the generalised diagonal defined by  $\pi$ , namely  $\{a_{i\pi(i)} : 1 \leq i \leq n\}$ , has precisely one entry which is ‘large’ (close to 1 or  $-1$ ), and all the rest of its entries are small. In the second part, we deduce that  $(a_{ij})$  must have either a row or a column, almost all of whose entries are large. This will enable us to complete the proof.

### Part 1

**Step 1 (§ 2.2, § 2.3).** Consider any two sets  $X, Y \subset \{1, \dots, n\}$  with  $|X| = |Y|$ , and the corresponding set of permutations

$$T_{X,Y} = \{\pi \in S_n : \pi(X) = Y\}.$$

When calculating  $f_1$  on  $T_{X,Y}$ , we only need to look at the submatrices of  $(a_{ij})$  defined by  $X \times Y$  and by  $\overline{X} \times \overline{Y}$ . So it is natural to define, for  $\pi_1$  a bijection from  $X$  to  $Y$  and  $\pi_2$  a bijection from  $\overline{X}$  to  $\overline{Y}$ ,

$$g_1(\pi_1) = \sum_{i \in X} a_{i\pi_1(i)}, \quad g_2(\pi_2) = \sum_{i \in \overline{X}} a_{i\pi_2(i)}, \quad g(\pi_1, \pi_2) = g_1(\pi_1) + g_2(\pi_2),$$

where  $(\pi_1, \pi_2)$  denotes the permutation of  $S_n$  whose restrictions to  $X$  and  $\overline{X}$  are  $\pi_1$  and  $\pi_2$  respectively. Notice that  $g$  is just the restriction of  $f_1$  to  $T_{X,Y}$ . For most choices of  $X, Y$ , these functions are ‘well-behaved’, meaning that all of the following hold (§ 2.2):

- For most permutations  $\pi \in T_{X,Y}$ ,  $g(\pi)$  is close to  $\pm 1$ .
- Furthermore, the function  $g$  is close to  $\pm 1$  in an  $L^2$  sense.
- Both  $\mathbb{E} g_1$  and  $\mathbb{E} g_2$  are close to their expected value, 0.

Next, we note that, crucially, the function  $g$  (which can be viewed as a random variable) is the sum of two independent random variables  $g_1, g_2$ , and yet is concentrated near 1 and  $-1$ . How can that happen? We show (§ 2.3) that it must be the case that one of the  $g_i$ ’s (say  $g_1$ ) is concentrated around a constant  $C$ , and that the other (say  $g_2$ ) is concentrated around two values,  $-C - 1$  and  $-C + 1$ . Using the observations above, it follows that  $C$  is very close to 0.

**Step 2 (§ 2.4).** For any permutation  $\pi \in S_n$ , we consider all pairs  $(X, Y)$  compatible with it, i.e. all pairs  $(X, \pi(X))$ . For most choices of  $\pi$  and for most choices of compatible  $(X, Y)$ , it will be true that one of  $g_1(\pi_1), g_2(\pi_2)$  is close to 0, and the other is close to  $\pm 1$ . Note that

$$g_1(\pi_1) = \sum_{i \in X} a_{i\pi(i)}, \quad g_2(\pi_2) = \sum_{i \in \overline{X}} a_{i\pi(i)}.$$

Put differently, for most permutations  $\pi \in S_n$  it is true that for most ways of splitting the generalised diagonal  $D = \{a_{i\pi(i)} : 1 \leq i \leq n\}$  into two parts, one part sums to roughly 0, and the other to roughly  $\pm 1$ . That can only happen if almost all entries in  $D$  are small, and one is close in magnitude to 1.

## Part 2 (§ 2.5)

This part uses induction on  $n$  to prove the following claim. If an  $n \times n$  matrix satisfies *property*  $Q(\delta)$ , namely a  $(1 - \delta)$ -fraction of its generalised diagonals have a single entry which is large in magnitude, then the matrix has a *strong line* — either a row or a column, a  $(1 - C\delta)$ -fraction of whose entries are large. (Here,  $C$  does not depend on  $n$ .)

**Base case.** When  $n$  is small compared to  $1/\delta$ , we can prove directly that there is a line where *all* of the entries are large.

**Induction step.** Given an  $n \times n$  matrix  $M$  satisfying  $Q(\delta)$  and a set  $X$  of  $n/2$  rows, we can always find a set  $Y$  of  $n/2$  columns such that either  $X \times Y$  or  $\overline{X} \times \overline{Y}$  also satisfy  $Q(\delta)$ . The induction hypothesis shows that the relevant submatrix has a strong line. The strong lines for different choices of  $X$  must be the same (on the same row or column of  $M$ ), since otherwise the probability that a generalized diagonal passes through two large entries would be too big. Altogether, these strong lines constitute a line  $\ell$  which is almost as strong as required. A small bootstrapping argument shows that  $\ell$  must indeed have the required number of large entries.

## Culmination (§ 2.6)

At this stage of the proof, we know that the matrix  $(a_{ij})$  has a line, say row  $i$ , almost all of whose entries are close either to  $-1$  or to  $1$ . It follows that for most  $j$ , it holds that  $(n-1)\langle f, T_{ij} \rangle$  is close to  $0$  or to  $1$ . The disjoint union of the 1-cosets corresponding to those entries close to  $1$  form a good approximation to  $\mathcal{F}$ .

Part 2 is largely independent of Part 1. Part 1 shows that most generalized diagonals of the matrix  $(a_{ij})$  are composed of one large entry and  $n-1$  small entries. Part 2 abstracts this situation, and deduces the existence of a strong line. The results of Part 2 work for any definition of which entries are large and which are small, and so are of independent interest.

**Glossary of terminology** *Restrictions* are defined in the beginning of § 2.2. *Typical restrictions* are defined in the end of § 2.2. *Good restrictions* are defined in the beginning of § 2.5. Functions which are *almost Boolean* or *almost close to  $C$*  are defined in § 2.2, just before Lemma 6. *Partitions*, *good partitions* and *good permutations* are defined in the beginning of § 2.4. *Small* and *large* entries are defined in the beginning of § 2.5. *Strong lines* (as well as strong rows and columns) are defined in § 2.5, just after Lemma 13.

## 2 Proof when $c = 1/2$

### 2.1 Matrix representation

Let  $\mathcal{F}$  and  $f$  be as in the statement of the theorem. Since  $f_1 \in U_1$ , it can be represented as a linear combination of 1-cosets  $T_{ij}$ . We single out one such representation:

$$a_{ij} = (n-1)\langle f, T_{ij} \rangle. \quad (3)$$

We start by showing that the  $a_{ij}$  do indeed represent  $f_1$ .

**Lemma 3.** *We have*

$$f_1 = \sum_{i,j} a_{ij} T_{ij}.$$

*Furthermore, each row and each column of the matrix  $(a_{ij})$  sums to zero:*

$$\sum_j a_{ij} = 0 \quad \forall i \in [n], \quad \sum_i a_{ij} = 0 \quad \forall j \in [n].$$

*For each permutation  $\pi \in S_n$ , we have*

$$f_1(\pi) = \sum_i a_{i\pi(i)}.$$

*Proof.* The ‘real’ proof of this fact uses the Fourier inversion formula, and the characters of the first two irreducible representations of  $S_n$ . This is how we derived the formula. However, to avoid digression, we offer a simpler, *ad hoc* argument.

The second statement follows from a simple calculation. For each  $i$ ,

$$\sum_j \langle f, T_{ij} \rangle = \langle f, 1 \rangle = 2\langle \chi_{\mathcal{F}}, 1 \rangle - \langle 1, 1 \rangle = 0,$$

so

$$\sum_j a_{ij} = 0.$$

Similarly, for each  $j$ ,

$$\sum_i \langle f, T_{ij} \rangle = \langle f, 1 \rangle = 0,$$

so

$$\sum_i a_{ij} = 0.$$

For the first statement, both sides of the equation are in  $U_1$ , so it is enough to show that both sides have the same inner product with each  $T_{ij}$ . Note that  $\langle T_{ij}, T_{ij} \rangle = 1/n$ ,  $\langle T_{ij}, T_{kl} \rangle = \frac{1}{n(n-1)}$  if  $i \neq k$  and  $j \neq l$ , and  $\langle T_{ij}, T_{kl} \rangle = 0$  if  $i \neq k$  or  $j \neq l$ . Therefore,

$$\left\langle \sum_{k,l} a_{kl} T_{kl}, T_{ij} \right\rangle = \frac{a_{ij}}{n} + \frac{1}{n(n-1)} \sum_{\substack{k \neq i \\ l \neq j}} a_{kl} = \frac{a_{ij}}{n} + \frac{a_{ij}}{n(n-1)} = \frac{a_{ij}}{n-1} = \langle f, T_{ij} \rangle = \langle f_1, T_{ij} \rangle,$$

using

$$\sum_{\substack{k \neq i \\ l \neq j}} a_{kl} = \sum_{k,l} a_{kl} - \sum_l a_{il} - \sum_k a_{kj} + a_{ij}. \quad (4)$$

Finally, the formula for  $f_1(\pi)$  follows immediately from the first statement.  $\square$

The preceding lemma shows that each value of  $f_1$  is equal to the sum of a generalised diagonal in the matrix  $(a_{ij})$ .

Next, we calculate the  $\ell^2$  norm of the vector formed by the entries  $a_{ij}$ .

**Lemma 4.** *We have*

$$\sum_{i,j} a_{ij}^2 = (n-1)(1 - \epsilon_1).$$

*Proof.* Since  $f_1$  is an orthogonal projection of  $f$ , we have

$$1 = \|f\|_2^2 = \|f_1\|_2^2 + \|f - f_1\|_2^2 = \|f_1\|_2^2 + \epsilon_1.$$

Therefore,  $\|f_1\|_2^2 = 1 - \epsilon_1$ . On the other hand, we have

$$\begin{aligned} \|f_1\|_2^2 &= \sum_{i,j,k,l} a_{ij} a_{kl} \langle T_{ij}, T_{kl} \rangle \\ &= \frac{1}{n} \sum_{i,j} a_{ij}^2 + \frac{1}{n(n-1)} \sum_{i,j} \sum_{\substack{k \neq i \\ l \neq j}} a_{ij} a_{kl} \\ &= \frac{1}{n} \sum_{i,j} a_{ij}^2 + \frac{1}{n(n-1)} \sum_{i,j} a_{ij}^2 = \frac{1}{n-1} \sum_{i,j} a_{ij}^2, \end{aligned}$$

using (4).  $\square$

## 2.2 Random restrictions

For  $X, Y \subset [n]$  of equal size, let  $T_{X,Y}$  denote the set of all permutations sending  $X$  to  $Y$ :

$$T_{X,Y} = \{\pi \in S_n : \pi(X) = Y\}.$$

We call such a pair  $(X, Y)$  a *restriction*. Let  $g(X, Y)$  denote the subvector of  $f_1$  supported on  $T_{X,Y}$ .

The final part of Lemma 3 shows that every value of  $f_1$  is the sum of a generalized diagonal of  $(a_{ij})$ . It is natural to decompose  $g(X, Y)$  into two functions, one depending on the submatrix supported by  $X \times Y$ , the other depending on the submatrix supported by  $\overline{X} \times \overline{Y}$ :

$$g_1(X, Y) = \sum_{i \in X, j \in Y} a_{ij} T_{ij}, \quad g_2(X, Y) = g_1(\overline{X}, \overline{Y}).$$

(In the definition of  $g_1$ ,  $T_{ij}$  is, strictly speaking, the restriction of  $\chi_{T_{ij}}$  to  $T_{X,Y}$ .) Lemma 3 immediately implies that  $g(X, Y) = g_1(X, Y) + g_2(X, Y)$ . Note that  $g_1(X, Y)$  and  $g_2(X, Y)$  are both supported on  $T_{X,Y}$ .

We now define a probability distribution  $\mathcal{R}$  over the set of all restrictions, as follows. Each  $i \in [n]$  is included in  $X$  independently at random with probability  $1/2$ . Then,  $Y$  is chosen uniformly at random from all sets of size  $|X|$ . Note that this definition is symmetric between  $X$  and  $Y$ , and furthermore,  $(X, Y)$  has the same distribution as  $(\overline{X}, \overline{Y})$ .

Most of this subsection will be devoted to the study of properties of typical restrictions. We start by calculating the mean and variance of  $\mathbb{E}[g_1(X, Y)]$  when  $(X, Y) \sim \mathcal{R}$ . This will enable us to show that  $\mathbb{E}[g_1(X, Y)]$  and  $\mathbb{E}[g_2(X, Y)]$  are typically small in magnitude.

**Lemma 5.** *Let  $X, Y \subset [n]$  with  $|X| = |Y|$ , and define*

$$m(X, Y) = \mathbb{E}_{T_{X,Y}} [g_1(X, Y)],$$

where the expectation is with respect to the uniform probability measure on  $T_{X,Y}$ .

If  $(X, Y) \sim \mathcal{R}$ , then the mean and variance of  $m(X, Y)$  with respect to  $\mathcal{R}$  satisfy

$$\mathbb{E}_{\mathcal{R}}[m] = 0$$

and

$$\mathbb{V}_{\mathcal{R}}[m] \leq \frac{1}{2n}.$$

*Proof.* We start with a formula for  $m(X, Y)$ :

$$m(X, Y) = \sum_{i \in X, j \in Y} a_{ij} \mathbb{E}_{T_{X,Y}} [T_{ij}] = \frac{1}{|X|} \sum_{i \in X, j \in Y} a_{ij}.$$

Conditioned upon  $|X|$ , we have

$$\mathbb{E}_{\mathcal{R}}[m(X, Y) | |X|] = \frac{1}{|X|} \frac{|X|^2}{n^2} \sum_{i,j} a_{ij} = 0.$$

Hence,  $\mathbb{E}_{\mathcal{R}}[m] = 0$ .

The next step is to calculate  $\mathbb{E}_{\mathcal{R}}[m^2]$ . Expanding the formula, we get

$$\begin{aligned} |X|^2 m(X, Y)^2 &= \sum_{i,j} \mathbf{1}_{\{i \in X, j \in Y\}} a_{ij}^2 + \sum_i \sum_{j \neq l} \mathbf{1}_{\{i \in X, j, l \in Y\}} a_{ij} a_{il} \\ &\quad + \sum_j \sum_{i \neq k} \mathbf{1}_{\{i, k \in X, j \in Y\}} a_{ij} a_{kj} + \sum_{\substack{k \neq i, \\ l \neq j}} \mathbf{1}_{\{i, k \in X, j, l \in Y\}} a_{ij} a_{kl}. \end{aligned}$$

Taking expectations, we get

$$\begin{aligned} |X|^2 \mathbb{E}_{\mathcal{R}}[m(X, Y)^2 | |X|] &= \Pr[i \in X \wedge j \in Y] \sum_{i,j} a_{ij}^2 + \Pr[i \in X \wedge j, l \in Y] \sum_i \sum_{l \neq j} a_{ij} a_{il} \\ &\quad + \Pr[i, k \in X \wedge j \in Y] \sum_{k \neq i} \sum_j a_{ij} a_{kj} + \Pr[i, k \in X \wedge j, l \in Y] \sum_{\substack{k \neq i, \\ l \neq j}} a_{ij} a_{kl}. \end{aligned}$$

Using (4), together with

$$\sum_{l \neq j} a_{il} = \sum_{k \neq i} a_{kj} = -a_{ij}$$

(from Lemma 3) we obtain

$$\begin{aligned} n^2 \mathbb{E}_{\mathcal{R}}[m(X, Y)^2 | |X|] &= \sum_{i,j} a_{ij}^2 - 2 \frac{|X| - 1}{n - 1} \sum_{i,j} a_{ij}^2 + \frac{(|X| - 1)^2}{(n - 1)^2} \sum_{i,j} a_{ij}^2 \\ &= \left(1 - \frac{|X| - 1}{n - 1}\right)^2 \sum_{i,j} a_{ij}^2. \end{aligned}$$

Taking expectations and using the estimate  $\sum_{i,j} a_{ij}^2 \leq n - 1$  provided by Lemma 4, we conclude that

$$\begin{aligned} \mathbb{E}_{\mathcal{R}}[m(X, Y)^2] &= \frac{n + 1}{4n(n - 1)^2} \sum_{i,j} a_{ij}^2 \\ &\leq \frac{n + 1}{4n(n - 1)} \\ &\leq \frac{1}{2n}. \end{aligned}$$

□

The following lemma states some properties that a random restriction enjoys with probability close to 1. The lemma uses the following nomenclature for functions on a probability space (a.k.a. random variables):

- A function  $\phi$  is  $(\delta, \epsilon)$ -almost Boolean if with probability at least  $1 - \delta$ ,  $\phi$  is  $\epsilon$ -close to  $\pm 1$ . In symbols,

$$\Pr[|\phi| \in 1 \pm \epsilon] \geq 1 - \delta.$$

- A function  $\phi$  is  $(\delta, \epsilon)$ -almost close to  $C$  if with probability at least  $1 - \delta$ ,  $\phi$  is  $\epsilon$ -close to  $C$ . In symbols,

$$\Pr[|\phi - C| \leq \epsilon] \geq 1 - \delta.$$

**Lemma 6.** *Let  $(X, Y) \sim \mathcal{R}$ . With probability at least  $1 - 3\epsilon_1^{1/7}$ ,  $(X, Y)$  satisfies the following properties:*

(a)  $g(X, Y)$  is  $(\epsilon_1^{4/7}, \epsilon_1^{1/7})$ -almost Boolean.

(b)  $\mathbb{E}[g_1(X, Y)]$  and  $\mathbb{E}[g_2(X, Y)]$  are  $\epsilon_1^{1/7}$ -close to zero.

(c)  $\mathbb{E}[ (|g(X, Y)| - 1)^2 ] \leq \epsilon_1^{6/7}$ .

*Proof.* We claim that each of the different parts holds with probability at least  $1 - \epsilon_1^{1/7}$ . The lemma follows using a union bound. We will use the fact that selecting a random partition  $(X, Y) \sim \mathcal{R}$  and then selecting a uniform random element in  $T_{X,Y}$  is the same as choosing a uniform random permutation. This holds because for any  $X$ , the sets  $(T_{X,Y} : |Y| = |X|)$  partition  $S_n$ .

We first deal with part (a). Suppose for a contradiction that

$$\Pr_{\mathcal{R}} \left[ \Pr_{T_{X,Y}} [ |g(X, Y)| \notin 1 \pm \epsilon_1^{1/7} ] \geq \epsilon_1^{4/7} \right] > \epsilon_1^{1/7}.$$

It follows that  $\Pr[|f_1| \notin 1 \pm \epsilon_1^{1/7}] > \epsilon_1^{5/7}$ . This implies that

$$\mathbb{E}[(f_1 - f)^2] \geq \mathbb{E}[ (|f_1| - 1)^2 ] > \epsilon_1^{5/7} (\epsilon_1^{1/7})^2 = \epsilon_1.$$

This contradicts  $\mathbb{E}[(f_1 - f)^2] = \epsilon_1$ , proving the claim for part (a).

For part (b), we use Lemma 5, which gives the mean and variance (with respect to  $\mathcal{R}$ ) of  $\mathbb{E}_{T_{X,Y}}[g_1(X, Y)]$ . We have

$$\Pr_{\mathcal{R}} \left[ \left| \mathbb{E}_{T_{X,Y}} [g_1(X, Y)] \right| \geq \frac{1}{\epsilon_1^{1/14} \sqrt{n}} \right] \leq \frac{\epsilon_1^{1/7} n}{2n} = \frac{1}{2} \epsilon_1^{1/7},$$

using Chebyshev's inequality. Assumption (2) states that  $\epsilon_1 \geq 1/n^{7/3}$ . Hence  $\epsilon_1^{3/14} \geq 1/\sqrt{n}$ , and so  $\epsilon_1^{1/7} \geq \epsilon_1^{-1/14}/\sqrt{n}$ . Therefore,

$$\Pr_{\mathcal{R}} \left[ \left| \mathbb{E}_{T_{X,Y}} [g_1(X, Y)] \right| \geq \epsilon_1^{1/7} \right] \leq \frac{1}{2} \epsilon_1^{1/7}.$$

Since  $(\bar{X}, \bar{Y}) \sim \mathcal{R}$ , the same holds for  $g_2(X, Y)$ . The claim for part (b) follows, using a union bound.

For part (c), the starting point is

$$\mathbb{E}[ (|f_1| - 1)^2 ] \leq \mathbb{E}[ (f_1 - f)^2 ] = \epsilon_1.$$

Therefore,

$$\mathbb{E}_{\mathcal{R}} \left[ \mathbb{E}_{T_{X,Y}} [ (|g(X, Y)| - 1)^2 ] \right] = \mathbb{E}[ (|f_1| - 1)^2 ] \leq \epsilon_1.$$

The claim now follows from Markov's inequality.  $\square$

We call a restriction *typical* if it satisfies the properties (a), (b) and (c) in Lemma 6.

### 2.3 Decomposition under a typical restriction

In this subsection, we show that if  $(X, Y)$  is a typical restriction (meaning a restriction satisfying the properties listed in Lemma 6), then the functions  $g_1(X, Y)$  and  $g_2(X, Y)$  have a particularly simple structure: up to translation, one of them is almost constant, and the other is almost Boolean.

Many of the lemmas in this subsection start by assuming that a particular restriction  $(X, Y)$  is typical. In these lemmas, we will write  $g, g_1, g_2$  for  $g(X, Y), g_1(X, Y), g_2(X, Y)$ .

The following technical lemma tackles the following situation. Suppose that some function  $\phi$  is almost close to  $C_0$ . Can we deduce that  $C_0 \approx \mathbb{E}\phi$ ? The lemma gives a sufficient condition (in the case  $C_0 = 0$ ).

**Lemma 7.** *Suppose that a function  $\phi$  on a probability space satisfies the following properties:*

- (a) *The function  $\phi$  is  $(p, \epsilon)$ -almost close to 0.*
- (b) *There exists  $C \in \mathbb{R}$  such that  $\mathbb{E}[ (|\phi + C| - 1)^2 ] \leq \delta$ .*

Then

$$|\mathbb{E}[\phi]| \leq 3\epsilon + 3p + 6\sqrt{\frac{\delta}{1-p}}.$$

**Remark.** *Condition (b) says that  $\phi + C$  is almost Boolean in the  $L^2$  sense. As we shall see, conditions (a) and (b) together imply that  $C$  must be close to 1 or close to  $-1$ .*

*Proof.* Without loss of generality, we may assume that  $C \geq 0$ . We start by establishing the bound

$$|1 - C| \leq \epsilon + \sqrt{\frac{\delta}{1-p}}. \tag{5}$$

We distinguish between three cases:  $C < 1 - \epsilon$ ,  $C > 1 + \epsilon$  and  $|1 - C| \leq \epsilon$ . In the latter case, we already have the desired bound.

Suppose  $C < 1 - \epsilon$ . Whenever  $|\phi| \leq \epsilon$ , we have

$$1 - |\phi + C| \geq 1 - C - \epsilon > 0.$$

Since this happens with probability at least  $1 - p$ , we deduce that  $(1 - p)(1 - C - \epsilon)^2 \leq \delta$ , verifying (5).

Suppose next that  $C > 1 + \epsilon$ . Whenever  $|\phi| \leq \epsilon$ , we have

$$|\phi + C| - 1 \geq C - \epsilon - 1 > 0.$$

Since this happens with probability at least  $1 - p$ , we deduce that  $(1 - p)(C - \epsilon - 1)^2 \leq \delta$ , again verifying (5). This completes the proof of (5).

When  $|t| \geq 1$ , we have  $|t| \leq t^2$ , and so

$$\mathbb{E}[|\phi + C| - 1 | \chi_{\{\phi + C \geq 2\}}] \leq \mathbb{E}[(|\phi + C| - 1)^2 \chi_{\{\phi + C \geq 2\}}] \leq \delta.$$

The triangle inequality implies that

$$|\mathbb{E}[(\phi + C - 1)\chi_{\{\phi + C \geq 2\}}]| \leq \mathbb{E}[|\phi + C - 1| \chi_{\{\phi + C \geq 2\}}] = \mathbb{E}[|\phi + C| - 1 | \chi_{\{\phi + C \geq 2\}}] \leq \delta.$$

When  $t \leq -2$ , we have  $|t - 1| = 1 - t \leq 3(-t - 1) = 3(|t| - 1) = 3||t| - 1|$ , and so, as before,

$$\mathbb{E}[|\phi + C - 1| \chi_{\{\phi + C \leq -2\}}] \leq 3\mathbb{E}[|\phi + C| - 1 | \chi_{\{\phi + C \leq -2\}}] \leq 3\delta.$$

The triangle inequality implies that

$$|\mathbb{E}[(\phi + C - 1)\chi_{\{\phi + C \leq -2\}}]| \leq 3\delta.$$

Combining the two together, we get

$$|\mathbb{E}[(\phi + C - 1)\chi_{\{|\phi + C| \geq 2\}}]| \leq 4\delta.$$

Define  $\psi = \phi + C - 1$ . Rewriting the last inequality in terms of  $\psi$ , we have

$$|\mathbb{E}[\psi \chi_{\{|\psi + 1| \geq 2\}}]| \leq 4\delta.$$

When  $|\psi + 1| \leq 2$ ,  $|\psi| \leq 3$ . When  $|\phi| \leq \epsilon$ , we have  $|\psi| \leq |1 - C| + \epsilon \leq 2\epsilon + \sqrt{\delta/(1 - p)}$ . Therefore

$$\begin{aligned} |\mathbb{E}[\psi]| &\leq |\mathbb{E}[\psi \chi_{\{|\phi| \leq \epsilon\}}]| + |\mathbb{E}[\psi \chi_{\{|\phi| > \epsilon \text{ and } |\psi + 1| \leq 2\}}]| + |\mathbb{E}[\psi \chi_{\{|\psi + 1| \geq 2\}}]| \\ &\leq 2\epsilon + \sqrt{\frac{\delta}{1 - p}} + 3p + 4\delta. \end{aligned}$$

We conclude that

$$|\mathbb{E}[\phi]| \leq |\mathbb{E}[\psi]| + |1 - C| \leq 3\epsilon + 2\sqrt{\frac{\delta}{1 - p}} + 3p + 4\delta. \quad \square$$

Our first key step is the following lemma, which uses the fact that  $g_1(X, Y)$  and  $g_2(X, Y)$  are independent pieces of  $g(X, Y)$  to deduce that, up to translation, both are close to Boolean. Moreover, at least one of them is close to being constant.

We will use the following notation, when a restriction  $(X, Y)$  is understood. For a permutation  $\pi$ ,  $\pi_1 = \pi|_X$  denotes its restriction to  $X$ , and  $\pi_2 = \pi|_{\overline{X}}$  denotes its restriction to  $\overline{X}$ . Thus,  $g_1$  depends only upon  $\pi_1$ , and  $g_2$  depends only upon  $\pi_2$ .

**Lemma 8.** *Suppose  $(X, Y)$  is a typical restriction. Choose  $\alpha, \beta$  uniformly at random from  $T_{X, Y}$ . Then with probability at least  $1 - 8\epsilon_1^{2/7}$ , one of the following three cases holds:*

- (a)  $|g_1(\alpha_1) - g_1(\beta_1)|$  and  $|g_2(\alpha_2) - g_2(\beta_2)| \leq 2\epsilon_1^{1/7}$ .
- (b)  $|g_1(\alpha_1) - g_1(\beta_1)| \leq 2\epsilon_1^{1/7}$  and  $|g_2(\alpha_2) - g_2(\beta_2)| \in 2 \pm 2\epsilon_1^{1/7}$ .
- (c)  $|g_1(\alpha_1) - g_1(\beta_1)| \in 2 \pm 2\epsilon_1^{1/7}$  and  $|g_2(\alpha_2) - g_2(\beta_2)| \leq 2\epsilon_1^{1/7}$ .

*Proof.* Typicality implies that  $\Pr[|g| - 1| > \epsilon_1^{1/7}] < \epsilon_1^{4/7}$ . This implies that with probability at least  $1 - \epsilon_1^{2/7}$  over the choice of  $\pi_1$ , it is true that  $\Pr_{\pi_2}[|g(\pi) - 1| > \epsilon_1^{1/7}] < \epsilon_1^{2/7}$ . So with probability at least  $1 - 2\epsilon_1^{2/7}$  over the choice of  $(\pi_1, \pi_2)$ , it is true that  $|g(\pi)|$  is  $\epsilon_1^{1/7}$ -close to 1. Thus with probability at least  $1 - 8\epsilon_1^{2/7}$  over the choice of  $\alpha, \beta$ , all of the following are  $\epsilon_1^{1/7}$ -close in magnitude to 1:

$$\begin{aligned} x &= g_1(\alpha_1) + g_2(\alpha_2), & y &= g_1(\alpha_1) + g_2(\beta_2), \\ z &= g_1(\beta_1) + g_2(\alpha_2), & w &= g_1(\beta_1) + g_2(\beta_2). \end{aligned}$$

Since  $\epsilon_1$  is small enough, each of these four values is unambiguously close to either 1 or  $-1$ .

If  $x, y, z$  are all close to the same value, then

$$|g_1(\alpha_1) - g_1(\beta_1)|, |g_2(\alpha_2) - g_2(\beta_2)| \leq 2\epsilon_1^{1/7}.$$

If  $x$  and  $y$  are close to different values, then

$$|g_2(\alpha_2) - g_2(\beta_2)| \in 2 \pm 2\epsilon_1^{1/7}.$$

Without loss of generality, we may assume that  $x$  is close to 1 and  $y$  is close to  $-1$ . Then  $g_2(\alpha_2) - g_2(\beta_2) \in 2 \pm 2\epsilon_1^{1/7}$ . If  $z$  is close to  $-1$ , then  $g_1(\alpha_1) - g_1(\beta_1) \in 2 \pm 2\epsilon_1^{1/7}$ . But this implies that

$$w = x - (g_1(\alpha_1) - g_1(\beta_1)) - (g_2(\alpha_2) - g_2(\beta_2)) \in -3 \pm 5\epsilon_1^{1/7}.$$

Since  $\epsilon_1$  is small enough,  $-3 + 5\epsilon_1^{1/7} < -1 - \epsilon_1^{1/7}$ , and we reach a contradiction. So when  $x$  and  $y$  are close to different values,  $x$  and  $z$  must be close to the same value. This implies that

$$|g_1(\alpha_1) - g_1(\beta_1)| \leq 2\epsilon_1^{1/7}.$$

If  $x$  and  $z$  are close to different values, then we similarly obtain

$$|g_1(\alpha_1) - g_1(\beta_1)| \in 2 \pm 2\epsilon_1^{1/7}, \quad |g_2(\alpha_2) - g_2(\beta_2)| \leq 2\epsilon_1^{1/7}.$$

These cases are exhaustive.  $\square$

The preceding lemma shows that for most choices of  $\alpha, \beta$ , either both  $g_1$  and  $g_2$  act as if they were constant, or one acts as if it were constant, and the other acts as if it were Boolean, up to translation. The following lemma, which is the main result of this section, shows that in fact, one is almost zero, and the other is almost Boolean.

**Lemma 9.** *Suppose  $(X, Y)$  are typical restrictions. Then either  $g_1$  is  $(3\epsilon_1^{1/7}, 19\epsilon_1^{1/7})$ -almost close to zero and  $g_2$  is  $(4\epsilon_1^{1/7}, 24\epsilon_1^{1/7})$ -almost Boolean, or the same is true with the roles of  $g_1$  and  $g_2$  reversed.*

*Proof.* Define

$$\begin{aligned} p_1 &= \Pr_{\alpha, \beta \in T_{X, Y}} [|g_1(\alpha_1) - g_1(\beta_1)| > 2\epsilon_1^{1/7}], \\ p_2 &= \Pr_{\alpha, \beta \in T_{X, Y}} [|g_2(\alpha_2) - g_2(\beta_2)| > 2\epsilon_1^{1/7}]. \end{aligned}$$

Lemma 8 implies that  $p_1 p_2 \leq 8\epsilon_1^{2/7}$ . Thus, either  $p_1 \leq 3\epsilon_1^{1/7}$  or  $p_2 \leq 3\epsilon_1^{1/7}$ . Without loss of generality, we may assume that  $p_1 \leq 3\epsilon_1^{1/7}$ .

A simple averaging argument shows that for some choice of  $\alpha$ , we have

$$\Pr_{\beta \in T_{X, Y}} [|g_1(\alpha_1) - g_1(\beta_1)| \leq 2\epsilon_1^{1/7}] \geq 1 - 3\epsilon_1^{1/7}.$$

Therefore, putting  $C_1 = g_1(\alpha_1)$ , we deduce that  $g_1$  is  $(3\epsilon_1^{1/7}, 2\epsilon_1^{1/7})$ -almost close to  $C_1$ .

Typicality implies that  $g = g_1 + g_2$  satisfies  $\mathbb{E}[|g| - 1]^2 \leq \epsilon_1^{6/7}$ . This must be true for some value  $C_2$  of  $g_2$ . The function  $g_1 - C_1$  is  $(3\epsilon_1^{1/7}, 2\epsilon_1^{1/7})$ -almost close to zero, and so we can apply Lemma 7, with the following parameters:

$$\phi \triangleq g_1 - C_1, \quad p \triangleq 3\epsilon_1^{1/7}, \quad \epsilon \triangleq 2\epsilon_1^{1/7}, \quad \delta \triangleq \epsilon_1^{6/7}, \quad C \triangleq C_1 + C_2.$$

Since  $\epsilon_1$  is small enough, Lemma 7 implies that

$$|\mathbb{E}[g_1] - C_1| \leq 6\epsilon_1^{1/7} + 9\epsilon_1^{1/7} + 6\sqrt{\frac{\epsilon_1^{6/7}}{1 - 3\epsilon_1^{1/7}}} = 15\epsilon_1^{1/7} + O(\epsilon_1^{3/7}) \leq 16\epsilon_1^{1/7}.$$

On the other hand, by typicality,  $|\mathbb{E}[g_1]| \leq \epsilon_1^{1/7}$ . Therefore

$$|C_1| \leq 17\epsilon_1^{1/7}.$$

We conclude that  $g_1$  is  $(3\epsilon_1^{1/7}, 19\epsilon_1^{1/7})$ -almost close to zero.

We now turn our gaze to  $g_2$ . Lemma 8 implies that with probability at least  $1 - 8\epsilon_1^{2/7}$  over the choice of  $\alpha, \beta$ , it holds that  $|g_2(\alpha_2) - g_2(\beta_2)| \in \{0, 2\} \pm 2\epsilon_1^{1/7}$ . A simple averaging argument shows that for some choice of  $\alpha$ , it holds that

$$\Pr_{\beta \in T_{X,Y}} [|g_2(\alpha_2) - g_2(\beta_2)| \notin \{0, 2\} \pm 2\epsilon_1^{1/7}] \leq 8\epsilon_1^{2/7}.$$

Let  $C_3 = g_2(\alpha_2)$ . Then  $g_2$  is concentrated on the three values  $\{C_3 - 2, C_3, C_3 + 2\}$ . Another application of Lemma 8 will show that it is actually concentrated either on  $\{C_3 - 2, C_3\}$  or on  $\{C_3, C_3 + 2\}$ . Define

$$q_1 = \Pr_{\beta \in T_{X,Y}} [g_2(\beta_2) \in C_3 + 2 \pm 2\epsilon_1^{1/7}],$$

$$q_2 = \Pr_{\gamma \in T_{X,Y}} [g_2(\gamma_2) \in C_3 - 2 \pm 2\epsilon_1^{1/7}].$$

When  $g_2(\beta_2) \in C_3 + 2 \pm 2\epsilon_1^{1/7}$  and  $g_2(\gamma_2) \in C_3 - 2 \pm 2\epsilon_1^{1/7}$ , we have  $|g_2(\beta_2) - g_2(\gamma_2)| \in 4 \pm 4\epsilon_1^{1/7}$ . In particular, since  $\epsilon_1$  is small enough, in this case  $\beta_2, \gamma_2$  satisfy none of the options presented by Lemma 8. Hence, we must have  $q_1 q_2 \leq 8\epsilon_1^{2/7}$ . Therefore, either  $q_1 \leq 3\epsilon_1^{1/7}$  or  $q_2 \leq 3\epsilon_1^{1/7}$ . Without loss of generality, we may assume that  $q_1 \leq 3\epsilon_1^{1/7}$ . Putting  $C_4 = C_3 - 1$ , we conclude that  $g_2 - C_4$  is  $(4\epsilon_1^{1/7}, 2\epsilon_1^{1/7})$ -almost Boolean. (Here, we used the estimate  $3\epsilon_1^{1/7} + 8\epsilon_1^{2/7} \leq 4\epsilon_1^{1/7}$ , true since  $\epsilon_1$  is small enough.) Our task is now to show that  $C_4$  is close to zero.

Since  $g_1$  is  $(3\epsilon_1^{1/7}, 19\epsilon_1^{1/7})$ -almost close to zero, it follows that  $g - C_4$  is  $(7\epsilon_1^{1/7}, 21\epsilon_1^{1/7})$ -almost Boolean. By typicality,  $g$  is  $(\epsilon_1^{4/7}, \epsilon_1^{1/7})$ -almost Boolean. Therefore, with probability at least  $1 - 7\epsilon_1^{1/7} - \epsilon_1^{4/7} \geq 1 - 8\epsilon_1^{1/7}$  over the choice of  $\pi \in T_{X,Y}$ ,

$$g(\pi) \in \{C_4 \pm 1\} \pm 21\epsilon_1^{1/7} \text{ and } g(\pi) \in \{\pm 1\} \pm \epsilon_1^{1/7}. \quad (6)$$

Suppose that  $\pi_+ \in T_{X,Y}$  satisfies (6) with  $g(\pi_+) \epsilon_1^{1/7}$ -close to 1. Then either  $C_4$  is  $22\epsilon_1^{1/7}$ -close to zero, or it is  $22\epsilon_1^{1/7}$ -close to 2. Similarly, if  $\pi_- \in T_{X,Y}$  satisfies (6) with  $g(\pi_-) \epsilon_1^{1/7}$ -close to  $-1$ , then either  $C_4$  is  $22\epsilon_1^{1/7}$ -close to zero, or it is  $22\epsilon_1^{1/7}$ -close to  $-2$ . Since  $\epsilon_1$  is small enough, if such permutations  $\pi_+, \pi_-$  exist, then we can conclude that  $|C_4| \leq 22\epsilon_1^{1/7}$ . That would complete the proof of the lemma.

It remains to rule out the case that for all permutations satisfying (6),  $g(\pi)$  has the same sign. That would imply that  $g$  is  $(8\epsilon_1^{1/7}, \epsilon_1^{1/7})$ -almost close to  $L$ , where  $L \in \{\pm 1\}$ . We apply Lemma 7, with the following parameters:

$$\phi \triangleq g - C, \quad p \triangleq 8\epsilon_1^{1/7}, \quad \epsilon \triangleq \epsilon_1^{1/7}, \quad \delta \triangleq \epsilon_1^{6/7}, \quad C \triangleq L.$$

Since  $\epsilon_1$  is small enough, the lemma implies that

$$|\mathbb{E}[g] - L| \leq 3\epsilon_1^{1/7} + 24\epsilon_1^{1/7} + 6\sqrt{\frac{\epsilon_1^{6/7}}{1 - 9\epsilon_1^{1/7}}} \leq 28\epsilon_1^{1/7}.$$

Therefore  $\mathbb{E}[g]$  is  $28\epsilon_1^{1/7}$ -close to  $L$ . On the other hand, typicality implies that  $\mathbb{E}[g]$  is  $2\epsilon_1^{1/7}$ -close to zero. We deduce that  $L$  is  $30\epsilon_1^{1/7}$ -close to zero. Since  $L \in \{\pm 1\}$  and  $\epsilon_1$  is small enough, this is a contradiction.  $\square$

## 2.4 Random partitions

Subsection 2.3 deals with random restrictions  $(X, Y)$ . The main result, Lemma 9, shows that with large probability, in the decomposition  $g(X, Y) = g_1(X, Y) + g_2(X, Y)$ , one of the functions is almost constant, and the other is almost Boolean. In this subsection, we switch the order of the random choices, and deduce a property of random permutations.

We will need the following classical theorem due to Esseen [10]. For a modern proof, see [16, 4.1.b]. Note we require Esseen's version, rather than Berry's slightly weaker result [2].

**Theorem 3** (Berry-Esseen). *Let  $X_1, \dots, X_n$  be independent random variables with finite third moments, and let  $S$  be their sum. Define*

$$\psi = \frac{\sum_{i=1}^n \mathbb{E}[|X_i - \mathbb{E} X_i|^3]}{(\sum_{i=1}^n \mathbb{V}[X_i])^{3/2}}.$$

*Let  $N$  be a normal random variable with the same mean and variance as  $S$ . Then  $S$  and  $N$  are  $C_0\psi$ -close in distribution, where  $C_0 < 1$  is an absolute constant. In other words, for every  $t \in \mathbb{R}$ ,*

$$|\Pr[S < t] - \Pr[N < t]| \leq C_0\psi.$$

Before stating the results, we need some definitions. If  $(X, Y) \sim \mathcal{R}$ , then the marginal distribution of  $X$  is  $U(2^{[n]})$ , the uniform distribution on the power set of  $[n]$ . With slight abuse of terminology, we call the subset  $X \subset [n]$  a *partition*, as it will correspond to the genuine partition  $(X, X^c)$ .

For a permutation  $\pi \in S_n$  and a partition  $X \subset [n]$ , define

$$P_1 := \sum_{i \in X} a_{i\pi(i)}, \quad P_2 := \sum_{i \in X^c} a_{i\pi(i)}.$$

We say that a partition  $X$  is *good* for  $\pi$  if either  $P_1$  is  $25\epsilon_1^{1/7}$ -close to zero and  $P_2$  is  $25\epsilon_1^{1/7}$ -close to  $\pm 1$ , or the same is true with the roles of  $P_1$  and  $P_2$  reversed. Otherwise, we say that  $X$  is *bad* for  $\pi$ . We say that the permutation  $\pi \in S_n$  is *good* if with probability at least  $4/5$ , a random partition  $X$  is good for  $\pi$ . Otherwise, we say that  $\pi$  is *bad*.

The following lemma shows that most permutations are good.

**Lemma 10.** *With probability at least  $1 - 50\epsilon_1^{1/7}$ , a random permutation  $\pi \in S_n$  is good.*

*Proof.* By Lemma 6, a restriction  $(X, Y) \sim \mathcal{R}$  is typical with probability at least  $1 - 3\epsilon_1^{1/7}$ . Suppose  $(X, Y)$  is typical. Choose a uniform random permutation  $\pi \in T_{X,Y}$ . Lemma 9 shows that with probability at least  $1 - 7\epsilon_1^{1/7}$ ,  $X$  is good for  $\pi$ . Hence, if we choose a restriction  $(X, Y) \sim \mathcal{R}$  and a permutation  $\pi \in T_{X,Y}$  uniformly at random, then  $X$  is good for  $\pi$  with probability at least  $1 - 10\epsilon_1^{1/7}$ .

Given  $X$ , the sets  $T_{X,Y}$  partition  $S_n$ . Therefore, the permutation  $\pi$  chosen in the process above is chosen uniformly at random from  $S_n$ . Furthermore, by definition, the marginal distribution of  $X$  is  $U(2^{[n]})$ . Therefore, if we first choose a permutation  $\pi \in S_n$  uniformly at random, and then we choose  $X \sim U(2^{[n]})$ , then  $X$  is good for  $\pi$  with probability at least  $1 - 10\epsilon_1^{1/7}$ . Thus, the average probability (over  $\pi \in S_n$ ) that a random partition is bad is at most  $10\epsilon_1^{1/7}$ :

$$\mathbb{E}_{\pi \in S_n} \left[ \Pr_{X \sim U(2^{[n]})} [X \text{ is bad for } \pi] \right] \leq 10\epsilon_1^{1/7}.$$

Markov's inequality now implies that the probability that  $\pi$  is bad is at most  $50\epsilon_1^{1/7}$ :

$$\Pr_{\pi \in S_n} \left[ \Pr_{X \sim U(2^{[n]})} [X \text{ is bad for } \pi] > 1/5 \right] < \frac{10\epsilon_1^{1/7}}{1/5} = 50\epsilon_1^{1/7}.$$

□

The next lemma shows that if  $\pi$  is a good permutation, then the generalized diagonal  $a_{i\pi(i)}$  corresponding to  $\pi$  has a special structure: one of its elements is 'large', and the rest are 'small'. This is, essentially, a consequence of the main statement of [13], but, for the sake of being self-contained, we give a full proof.

**Lemma 11.** *Suppose  $\pi \in S_n$  is a good permutation. Then for some  $m \in [n]$ ,  $|a_{m\pi(m)}|$  is  $50\epsilon_1^{1/7}$ -close to  $\pm 1$ , and for  $i \neq m$ ,  $|a_{i\pi(i)}| \leq 50\epsilon_1^{1/7}$ .*

*Proof.* The proof is inspired by one of the proofs in [13]. Considering what happens when an element ‘switches sides’ allows us to group the elements  $a_{i\pi(i)}$  into two groups: ‘small’ elements (close to zero) and ‘large’ elements (close to  $\pm 1$ ). Similar considerations show that there can be at most one large element. The crucial part is showing that not all elements can be small. Indeed, in this case, the sum  $P_1$  in the definition of goodness is approximately normal, and so it cannot be concentrated on the two values  $\{0, 1\}$  or  $\{0, -1\}$ . The formal proof is as follows.

Define  $s_i = a_{i\pi(i)}$ . Since  $\pi$  is good,  $S_0 := \sum_{i=1}^n s_i$  is  $50\epsilon_1^{1/7}$ -close to either 1 or  $-1$ . It cannot be close to both, since  $\epsilon_1$  is small enough. Choose  $K \in \{\pm 1\}$  so that  $S_0$  is  $50\epsilon_1^{1/7}$ -close to  $K$ .

Define  $T(X) = \sum_{i \in X} s_i$ . Let  $X \sim U(2^{[n]})$ , and put  $T = T(X)$ , so that  $T$  is also a random variable. Note that  $T = S_0/2 + \sum_{i=1}^n W_i$ , where  $W_i = s_i(\chi_{i \in X} - 1/2)$ . Clearly,  $X$  is a good partition for  $\pi$  if and only if  $\bar{X}$  is a good partition for  $\pi$ . Since  $X$  and  $\bar{X}$  are equidistributed, given that  $X$  is good for  $\pi$ ,  $T$  is  $25\epsilon_1^{1/7}$ -close to zero with probability  $1/2$ , and  $25\epsilon_1^{1/7}$ -close to  $K$  with probability  $1/2$ . We conclude that with probability at least  $2/5$ ,  $T$  is  $25\epsilon_1^{1/7}$ -close to zero, and with probability at least  $2/5$ ,  $T$  is  $25\epsilon_1^{1/7}$ -close to  $K$ .

Consider any  $s_i$ . Since  $2 \cdot 1/5 < 1$  (here,  $1/5$  is an upper bound on the probability that a random partition is bad for  $\pi$ ), there is some choice of  $Y \subset [n] \setminus \{i\}$  such that both  $Y$  and  $Y \cup \{i\}$  are good for  $\pi$ . Since  $|T(Y) - T(Y \cup \{i\})| = |s_i|$ , necessarily either  $|s_i| \leq 50\epsilon_1^{1/7}$  ( $s_i$  is *small*) or  $|s_i|$  is  $50\epsilon_1^{1/7}$ -close to  $|K| = 1$  ( $s_i$  is *large*).

We claim that not all the  $s_i$  can be small. Assume, for the sake of contradiction, that  $|s_i| \leq 50\epsilon_1^{1/7}$  for all  $i$ . Applying Berry-Esseen with  $X_i = W_i$  shows that  $T$  is  $\psi$ -close in distribution to a normal distribution  $N \sim \mathcal{N}(S_0/2, \sigma^2)$ , where

$$\sigma^2 = \frac{1}{4} \sum_{i=1}^n s_i^2, \quad \psi = \frac{\sum_{i=1}^n |s_i|^3}{(\sum_{i=1}^n s_i^2)^{3/2}}.$$

The upper bound on  $|s_i|$  implies that  $|s_i|^3 \leq 50\epsilon_1^{1/7} s_i^2$ , and so

$$\psi \leq \frac{50\epsilon_1^{1/7} \sum_{i=1}^n s_i^2}{(\sum_{i=1}^n s_i^2)^{3/2}} = \frac{50\epsilon_1^{1/7}}{\sqrt{\sum_{i=1}^n s_i^2}} = \frac{25\epsilon_1^{1/7}}{\sigma}.$$

We now obtain a lower bound on  $\sigma$ . With probability at least  $4/5$ ,  $T$  is  $25\epsilon_1^{1/7}$ -close to zero or to  $K$ , and so its distance from its mean  $S_0/2$  is at least  $1/2 - 50\epsilon_1^{1/7}$ . Hence

$$\sigma^2 = \mathbb{V}[T] \geq \frac{4}{5} \left( \frac{1}{2} - 50\epsilon_1^{1/7} \right)^2 = \Omega(1).$$

This shows that  $\psi = O(1)$ . Concretely, when  $\epsilon_1^{1/7}$  is small enough,  $\psi \leq 1/10$ .

For every interval  $I$ , Berry-Esseen shows that  $|\Pr[T \in I] - \Pr[N \in I]| < 2\psi$ . We consider three intervals:  $I_1 = 0 \pm 25\epsilon_1^{1/7}$ ,  $I_2 = K \pm 25\epsilon_1^{1/7}$ , and  $I_3$  is the interval ‘in between’: when  $K = 1$ ,  $I_3 = (25\epsilon_1^{1/7}, 1 - 25\epsilon_1^{1/7})$ , and when  $K = -1$ ,  $I_3 = (-1 + 25\epsilon_1^{1/7}, -25\epsilon_1^{1/7})$ . By assumption,

$$\Pr[T \in I_1] \geq 2/5, \quad \Pr[T \in I_2] \geq 2/5, \quad \Pr[T \in I_3] \leq 1/5.$$

Since  $2\psi \leq 1/5$ , we deduce that

$$\Pr[N \in I_1] \geq 1/5, \quad \Pr[N \in I_2] \geq 1/5, \quad \Pr[N \in I_3] \leq 2/5.$$

The density of a normal distribution is bitonic (increasing and then decreasing), and so

$$\frac{\Pr[N \in I_3]}{|I_3|} \geq \min \left( \frac{\Pr[N \in I_1]}{|I_1|}, \frac{\Pr[N \in I_2]}{|I_2|} \right). \quad (7)$$

We have  $|I_1| = |I_2| = 50\epsilon_1^{1/7}$  and  $|I_3| = 1 - 50\epsilon_1^{1/7}$ . Therefore, the left-hand side of (7) is at most (say)  $4/5$ , and both terms on the right-hand side are at least (say)  $1$  (since  $\epsilon_1$  is small enough), a contradiction.

Concluding, there must be some  $m$  such that  $|s_m|$  is  $25\epsilon_1^{1/7}$ -close to  $1$ . We claim that there cannot be two such indices  $m, l$ . Suppose, for the sake of contradiction, that both  $|s_m|$  and  $|s_l|$  are  $25\epsilon_1^{1/7}$ -close to  $1$ . Since  $4 \cdot 1/5 < 1$ , there is some choice of  $Y \subset [n] \setminus \{m, l\}$  such that all of  $Y, Y \cup \{m\}, Y \cup \{l\}, Y \cup \{m, l\}$  are good for  $\pi$ . Since  $T(Y \cup \{m\}) = T(Y) + s_m$  and  $T(Y \cup \{l\}) = T(Y) + s_l$ , we see that  $s_m, s_l$  must have the same sign. Since  $T(Y \cup \{m, l\}) = T(Y) + s_m + s_l$ , this implies that  $Y$  and  $Y \cup \{m, l\}$  cannot both be good. This contradiction shows that there can be at most one large  $s_m$ .  $\square$

## 2.5 Strong lines

The previous section showed that if we pick a generalized diagonal at random in the matrix  $(a_{ij})$ , then with probability close to  $1$ , we can designate exactly one element in it as ‘large’. Corollary 12 restates this formally.

In this subsection, we say that an entry  $a_{ij}$  is *large* if  $|a_{ij}|$  is  $50\epsilon_1^{1/7}$  close to  $1$ . Otherwise, we say it is *small*. Note that, contrary to the usage in Lemma 11, small elements need not be close to  $0$ . While Lemma 11 allows us to deduce that most of the non-large elements in  $(a_{ij})$  are actually close to  $0$ , for what follows, it will be enough for us to just maintain a distinction between large elements and non-large elements.

Let  $(X, Y)$  be a restriction. Denote by  $A[X, Y]$  the submatrix  $(a_{ij})_{i \in X, j \in Y}$ . We say that a generalized diagonal in  $A[X, Y]$  is *good* if it contains exactly one large entry. We say that  $(X, Y)$  is *q-good* if with probability at least  $1 - q$ , a random generalized diagonal in  $A[X, Y]$  is good.

**Corollary 12.** *The restriction  $([n], [n])$  is  $50\epsilon_1^{1/7}$ -good.*

*Proof.* Immediate from Lemma 10 and Lemma 11.  $\square$

Our goal is to deduce that  $([n], [n])$  has a row or column which contains  $(1 - O(\epsilon_1^{1/7}))n$  large entries. The general plan of attack is to prove this by induction on  $n$ . We will have a separate argument for small values of  $n$ , and an inductive argument for large  $n$ . The latter will use the following lemma, which we will apply with  $|X'| = \lfloor |X|/2 \rfloor$ .

**Lemma 13.** *Suppose that  $(X, Y)$  is  $q$ -good for  $q < 1/2$ . For every  $X' \subset X$  there exists  $Y' \subset Y$  with  $|Y'| = |X'|$ , such that either  $(X', Y')$  or  $(X \setminus X', Y \setminus Y')$  is  $q$ -good. Similarly, for every  $Y' \subset Y$  there exists  $X' \subset X$  with  $|X'| = |Y'|$ , such that either  $(X', Y')$  or  $(X \setminus X', Y \setminus Y')$  is  $q$ -good.*

*Proof.* By symmetry, we need only prove the first statement. Fix  $X' \subset X$ . Since the sets  $(T_{X', Y'} : Y' \subset Y, |Y'| = |X'|)$  partition  $T_{X, Y}$ , the fact that  $(X, Y)$  is  $q$ -good implies that for some choice of  $Y'$ , the probability that a random generalized diagonal corresponding to a permutation in  $T_{X', Y'}$  is good is at least  $1 - q$ . Choose such a  $Y'$ .

Let  $p_1$  be the probability that a random generalized diagonal in  $A[X', Y']$  is good, and let  $p_2$  be the probability that a random generalized diagonal in  $A[X \setminus X', Y \setminus Y']$  is good. Then  $p_1(1 - p_2) + (1 - p_1)p_2 \geq 1 - q$ . We claim that this forces  $\max(p_1, p_2) \geq 1 - q$ .

Indeed, let  $p_1 = (1 + \delta_1)/2$  and  $p_2 = (1 + \delta_2)/2$ , where  $|\delta_1|, |\delta_2| \leq 1$ . We have

$$p_1(1 - p_2) + (1 - p_1)p_2 = \frac{1 - \delta_1\delta_2}{2}.$$

Since  $1 - q > 1/2$ , we must have  $\delta_1\delta_2 < 0$ . Without loss of generality, we may assume that  $\delta_1 > 0$ . Since  $1 \geq -\delta_2$ , we have

$$p_1 = \frac{1 + \delta_1}{2} \geq \frac{1 - \delta_1\delta_2}{2} \geq 1 - q,$$

proving the lemma.  $\square$

Let  $(X, Y)$  be a restriction. If  $i \in X$ , we say that row  $i$  is *p-strong* for  $(X, Y)$  if at least  $(1 - p)|Y|$  of the entries  $\{a_{ij} : j \in Y\}$  are large. If  $j \in Y$ , we say that column  $j$  is *p-strong* for  $(X, Y)$  if at least  $(1 - p)|X|$  of the entries  $\{a_{ij} : i \in X\}$  are large.

We say that  $(X, Y)$  has a  $p$ -strong row (resp. column) if some row (resp. column) is  $p$ -strong for  $(X, Y)$ . We say that  $(X, Y)$  has a  $p$ -strong line if it has either a  $p$ -strong row or a  $p$ -strong column.

Our goal is to show that  $([n], [n])$  has a strong line. We start by showing that two strong lines must coincide.

**Lemma 14.** *Suppose that  $(X, Y)$  is  $q$ -good. Let  $X_1, X_2 \subset X$ , and let  $Y_1, Y_2 \subset Y$  with  $|Y_1| = |X_1|$  and  $|Y_2| = |X_2|$ . Suppose that  $(X_1, Y_1)$  has a  $p_1$ -strong line, and that  $(X_2, Y_2)$  has a  $p_2$ -strong line. If  $(1 - p_1)|X_1| > 1$ ,  $(1 - p_2)|X_2| > 1$  and*

$$(1 - p_1)(1 - p_2) \frac{|X_1||X_2|}{|X|^2} \geq 4q$$

*then the strong lines must be the same (defined by the same row or by the same column).*

*Proof.* Suppose, for the sake of contradiction, that the two restrictions have different strong lines. Let  $L_1 \subset X_1 \times Y_1$  consist of the first  $t_1 = \lceil (1 - p_1)|X_1| \rceil$  indices of large elements in the strong line of  $(X_1, Y_1)$ , and let  $L_2$  consist of the first  $t_2 = \lceil (1 - p_2)|X_2| \rceil$  indices of large elements in the strong line of  $(X_2, Y_2)$ .

Say that  $(i_1, j_1) \in L_1$  and  $(i_2, j_2) \in L_2$  conflict if either  $i_1 = i_2$  or  $j_1 = j_2$  (or both). If  $L_1$  is row  $i$  and  $L_2$  is column  $j$ , then an entry on  $L_1$  not on column  $j$  never conflicts with an entry on  $L_2$  not on row  $i$ . Therefore, there are at least  $(t_1 - 1)(t_2 - 1)$  non-conflicting pairs. If both  $L_1$  and  $L_2$  are rows (resp. columns), then two entries conflict only if they are on the same column (resp. row). Therefore, the number of non-conflicting pairs is at least  $t_1 t_2 - \min(t_1, t_2) \geq (t_1 - 1)(t_2 - 1)$ .

For each non-conflicting pair, the probability that a random generalized diagonal in  $A[X, Y]$  goes through both entries of the pair is  $1/|X|(|X| - 1)$ . Since these events are all disjoint, it follows that

$$\frac{(t_1 - 1)(t_2 - 1)}{|X|^2} < \frac{(t_1 - 1)(t_2 - 1)}{|X|(|X| - 1)} \leq q.$$

Since  $t_1, t_2 \geq 2$ ,  $t_1 - 1 \geq t_1/2$  and  $t_2 - 1 \geq t_2/2$ . Using  $t_1/|X_1| \geq 1 - p_1$  and  $t_2/|X_2| \geq 1 - p_2$ , we deduce that

$$(1 - p_1)(1 - p_2) \frac{|X_1||X_2|}{|X|^2} < 4q,$$

contradicting our assumption.  $\square$

Note that the conditions  $(1 - p_1)|X_1| > 1$  and  $(1 - p_2)|X_2| > 1$  simply guarantee that each strong line has at least two large elements. (If one of the strong lines had only one large element, then it could be contained in the other strong line, and so there would be no contradiction.)

Our next result says that if there is one strong line, then there cannot be many large entries outside the line. For the proof, we need the simplest case of Bonferroni's inequality.

**Theorem (Bonferroni).** *Let  $A_1, \dots, A_h$  be events. Then*

$$\Pr[A_1 \vee \dots \vee A_h] \geq \sum_i \Pr[A_i] - \sum_{i < j} \Pr[A_i \wedge A_j].$$

**Lemma 15.** *Suppose that  $(X, Y)$  is  $q$ -good and has a  $p$ -strong line. Let  $m = |X|$ , and let  $\varrho = 2q/(1 - p)$ . If  $m \geq 6$ ,  $(1 - p)m > 1$ ,  $2\varrho m > 1$  and  $\varrho \leq 1/2$ , then that line is actually  $(q + 3\varrho)$ -strong.*

*Proof.* Without loss of generality, we may assume that the  $p$ -strong line is row  $i$ . Since  $(X, Y)$  is  $q$ -good, a random element in  $A[X, Y]$  is large with probability at least  $(1 - q)/m$ . Therefore,  $A[X, Y]$  contains at least  $(1 - q)m$  large entries.

Suppose that row  $i$  is not  $(q + 3\varrho)$ -strong. Then  $A[X, Y]$  contains at least  $3\varrho m$  large entries outside row  $i$ . Lemma 14 implies that no other line can be  $(1 - 2\varrho)$ -strong. Therefore, no column can contain more than  $2\varrho m$  large entries. So for any column  $j$ ,  $A[X, Y]$  contains at least  $\varrho m$  large entries outside row  $i$  and column  $j$ . The probability that a random generalized diagonal in  $A[X, Y]$  hits any single one of these, given that it hits a specific large entry in row  $i$ , is  $1/(m - 1)$ , and the probability that it hits any two of them is at most  $1/(m - 1)(m - 2)$ . Therefore, Bonferroni's inequality implies that the probability that a generalized diagonal in  $A[X, Y]$  contains at least two large elements is at least

$$(1 - p) \left( \frac{\varrho m}{m - 1} - \frac{(\varrho m + 1)(\varrho m)}{2(m - 1)(m - 2)} \right) = \frac{(1 - p)\varrho m}{m - 1} \left( 1 - \frac{\varrho m + 1}{2(m - 2)} \right) > q,$$

since  $\varrho m + 1 \leq m/2 + 1 \leq m - 2$ . But this probability must be at most  $q$ , a contradiction.  $\square$

The following sequence of lemmas shows the existence of a strong line in  $(X, Y)$ , given that  $(X, Y)$  is  $q$ -good for  $q$  sufficiently small depending on  $|X|$ .

**Lemma 16.** *Suppose that  $(X, Y)$  is  $q$ -good for some  $q < \frac{1}{m(m-1)}$ , where  $m = |X|$ . Then  $(X, Y)$  has a 0-strong line.*

*Proof.* Without loss of generality, we may assume that  $X = Y = [m]$ . Since  $q < 1/m$ , there must exist a permutation  $\pi \in S_m$  such that all generalized diagonals of the form  $\{(i, \pi(i) + j) : i \in [m]\}$  are good. Without loss of generality, we may assume that  $\pi$  is the identity, and that  $a_{1,1}$  is large. For  $j \in [m]$ , let  $a_{i,i+j}$  be the large entry on the corresponding diagonal. If  $i \neq 1$  and  $i+j \neq 1$ , then a random generalized diagonal in  $A[[m], [m]]$  passes through both  $a_{1,1}$  and  $a_{i,i+j}$  with probability  $1/(m(m-1))$ , contrary to our assumption. Therefore, all large entries are either on row 1 or on column 1. If there is an entry  $a_{i,1}$  not on row 1 and an entry  $a_{1,j}$  not on column 1, then we again reach a contradiction. It follows that either row 1 or column 1 consists of large elements only.  $\square$

We now improve this result using induction.

**Lemma 17.** *Suppose that  $(X, Y)$  is  $q$ -good for some  $q < \frac{1}{4m}$ , where  $m = |X|$ . Then  $(X, Y)$  has a 0-strong line.*

*Proof.* The proof is by induction on  $m$ . When  $m \leq 5$ , the claim follows from Lemma 16, so suppose that  $m \geq 6$ .

Let  $X' \subset X$  be an arbitrary subset of size  $s = \lfloor m/2 \rfloor$ . Lemma 13 shows that there exists  $Y' \subset Y$  with  $|Y'| = |X'|$ , such that either  $(X', Y')$  is  $q$ -good or  $(X \setminus X', Y \setminus Y')$  is  $q$ -good. The induction hypothesis implies that either  $(X', Y')$  or  $(X \setminus X', Y \setminus Y')$  has a 0-strong line. Similarly, if  $Y' \subset Y$  is an arbitrary subset of size  $s = \lfloor m/2 \rfloor$ , then there exists  $X' \subset X$  with  $|X'| = |Y'|$  such that either  $(X', Y')$  or  $(X \setminus X', Y \setminus Y')$  has a 0-strong line.

Since  $s \geq 3$ , we have

$$\left(\frac{s}{m}\right)^2 \geq \frac{9}{49} > \frac{1}{6} \geq \frac{1}{m} > 4q,$$

so Lemma 14 implies that all 0-strong lines arising from different choices of  $X'$  or  $Y'$  must be defined by the same row or column — say row  $i$ .

We claim that row  $i$  can have at most one small entry. If row  $i$  has at least two small entries  $a_{ij}, a_{ik}$ , then there exists  $Y' \subset Y$  with  $|Y'| = s$ , such that  $j \in Y'$  and  $k \in Y \setminus Y'$ . For any  $X' \subset X$  with  $|X'| = s$ , row  $i$  can be 0-strong in neither  $(X', Y')$  nor  $(X \setminus X', Y \setminus Y')$ , a contradiction. Thus, row  $i$  has at most one small entry.

Suppose that row  $i$  has exactly one small entry. Since  $q < 1/4m < 1/m$ , a random entry in  $A[X, Y]$  is large with probability at least  $(1-q)/m > 1/m - 1/m^2$ , and so there must be at least  $m$  large entries. Exactly  $m-1$  of these are on row  $i$ . Let  $a_{kl}$  be another large entry. The probability that a random generalized diagonal passes through both  $a_{kl}$  and one of the large entries on row  $i$  is at least

$$\frac{1}{m} \left(1 - \frac{1}{m-1}\right) \geq \frac{4/5}{m} > \frac{1}{4m},$$

a contradiction. Hence, row  $i$  must be 0-strong for  $A[X, Y]$ , completing the proof.  $\square$

We now use induction to tackle the case of large  $|X|$ .

**Lemma 18.** *Suppose that  $(X, Y)$  is  $q$ -good for  $q < 1/50$ . Then  $(X, Y)$  has a  $13q$ -strong line.*

*Proof.* The proof is by induction on  $m := |X|$ . When  $m < 1/(4q)$ , the statement of the lemma follows from Lemma 17, so suppose that  $m \geq 1/(4q) \geq 12$ .

Let  $X' \subset X$  be an arbitrary subset of size  $s = \lfloor m/2 \rfloor \geq 6$ . Lemma 13 shows that there exists  $Y' \subset Y$  with  $|Y'| = |X'|$ , such that either  $(X', Y')$  is  $q$ -good or  $(\overline{X'}, \overline{Y'})$  is  $q$ -good. The induction hypothesis implies that either  $(X', Y')$  or  $(X \setminus X', Y \setminus Y')$  has a  $13q$ -strong line. Similarly, if  $Y' \subset Y$  is an arbitrary subset of size  $s = \lfloor m/2 \rfloor$ , then there exists  $X' \subset X$  with  $|X'| = |Y'|$  such that either  $(X', Y')$  or  $(X \setminus X', Y \setminus Y')$  has a  $13q$ -strong line.

Since  $(1-13q)s > 1$  and  $(1-13q)^2(s/m)^2 \geq 4q$ , Lemma 14 implies that all the  $13q$ -strong lines arising from different choices of  $X'$  or  $Y'$  must be defined by the same row or column — say row  $i$ .

We claim that row  $i$  has at most  $\lfloor 13qm + 1 \rfloor$  small entries. Indeed, suppose it has at least  $\lfloor 13qm + 2 \rfloor$  small entries. Then there exists a subset  $Y' \subset Y$  with  $|Y'| = s$ , such that the  $Y'$ -part of row  $i$  contains at least  $\lfloor 13qs + 1 \rfloor$  small entries and the  $Y \setminus Y'$ -part of it contains at least  $\lfloor 13q(m-s) + 1 \rfloor$  small entries. For any  $X' \subset X$ , row  $i$  is not  $13q$ -strong in either  $(X', Y')$  nor  $(X \setminus X', Y \setminus Y')$ , a contradiction. Therefore, row  $i$  is a  $(13q + 1/m)$ -strong line for  $(X, Y)$ .

Let  $p = 13q + 1/m$ . We are going to apply Lemma 15. Clearly  $m \geq 12$ , and it is easy to check that  $(1-p)m > 1$  and  $\varrho \leq 1/2$ . Slightly more delicately, we have

$$2\varrho m = \frac{4qm}{1 - 13q - 1/m} > 4qm \geq 1.$$

Hence, by Lemma 15, row  $i$  is  $(q+3\varrho)$ -strong. Since  $p < 1/2$ , we have  $q+3\varrho < 13q$ , so row  $i$  is  $13q$ -strong, completing the proof.  $\square$

It might seem that the condition  $2\varrho m > 1$  is very tight. This will not matter for us, but in fact, one can prove a version of Lemma 15 with a weaker condition, at the cost of obtaining a worse guarantee on the strength of the line.

**Corollary 19.** *There exists an  $O(\epsilon_1^{1/7})$ -strong line for  $([n], [n])$ .*

*Proof.* Follows immediately from Corollary 12 and Lemma 18.  $\square$

## 2.6 Culmination of the proof

In this section, we will see what Corollary 19 implies in terms of the original family  $\mathcal{F}$ . Without loss of generality, we may assume for the rest of this section that the  $O(\epsilon_1^{1/7})$ -strong line whose existence is guaranteed by Corollary 19 is row 1.

What the corollary implicitly says is that the matrix  $(a_{ij})$  looks very like the canonical example shown in the introduction:

$$\begin{pmatrix} \overbrace{1 - \frac{1}{n} \quad \cdots \quad 1 - \frac{1}{n}}^{\frac{n}{2}} & \frac{1}{n} - 1 & \cdots & \frac{1}{n} - 1 \\ -\frac{1}{n} & \cdots & -\frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ -\frac{1}{n} & \cdots & -\frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & & \vdots & \vdots & & \vdots \\ -\frac{1}{n} & \cdots & -\frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}$$

Indeed, the corollary shows that, without loss of generality, the first row consists mainly of elements which are very close to  $\pm 1$ . Since the line must sum to zero, we know that roughly half of these are close to 1, and roughly half to  $-1$ . This information will enable us to deduce that  $\mathcal{F}$  is close to a disjoint union of roughly  $n/2$  cosets.

For  $i, j \in [n]$ , we define

$$\tau_{ij} := \frac{|\mathcal{F} \cap T_{ij}|}{(n-1)!}.$$

By (3), we have

$$\tau_{ij} = \frac{1}{2} + \frac{n}{2(n-1)} a_{ij}. \tag{8}$$

So we have  $\tau_{ij} \approx (a_{ij} + 1)/2$ . More precisely, we have the following.

**Lemma 20.** *Each  $\tau_{ij}$  is  $1/n$ -close to  $(a_{ij} + 1)/2$ . If  $a_{ij}$  is large, then  $\tau_{ij}$  is  $26\epsilon_1^{1/7}$ -close to  $\{0, 1\}$ .*

*Proof.* The formula (3) for  $a_{ij}$  implies that  $|a_{ij}| \leq 1$ . We have

$$\tau_{ij} = \frac{a_{ij} + 1}{2} + \frac{a_{ij}}{2(n-1)}.$$

The second term has absolute value at most  $\frac{1}{2(n-1)} \leq \frac{1}{n}$ , since  $n \geq 2$ .

A large entry is  $50\epsilon_1^{1/7}$  close in magnitude to  $\pm 1$ , and so  $(a_{ij} + 1)/2$  is  $25\epsilon_1^{1/7}$ -close to  $\{0, 1\}$ . Finally, assumption (2) implies that  $1/n \leq \epsilon_1^{1/7}$ .  $\square$

We are now almost ready to prove our main result.

**Lemma 21.** *The number of  $\tau_{1i}$  which are  $26\epsilon_1^{1/7}$ -close to 1 is  $O(\epsilon_1^{1/7})n$ -close to  $n/2$ .*

*Proof.* Let  $N_0$  be the number of  $\tau_{1i}$  which are  $26\epsilon_1^{1/7}$ -close to 0, and let  $N_1$  be the number of  $\tau_{1i}$  which are  $26\epsilon_1^{1/7}$ -close to 1. Lemma 3 shows that

$$T := \sum_{i=1}^n \tau_{1i} = \frac{n}{2}.$$

On the other hand,

$$(1 - 26\epsilon_1^{1/7})N_1 \leq T \leq 26\epsilon_1^{1/7}N_0 + (n - N_0) = n - (1 - 26\epsilon_1^{1/7})N_0.$$

Substituting  $T = n/2$ , we obtain

$$N_0, N_1 \leq (1 + O(\epsilon_1^{1/7}))n/2.$$

The fact that row 1 is an  $O(\epsilon_1^{1/7})$ -strong line implies that  $N_0 + N_1 = (1 - O(\epsilon_1^{1/7}))n$ , and so

$$N_1 \geq (1 - O(\epsilon_1^{1/7}))n - N_0 \geq (1 - O(\epsilon_1^{1/7}))n/2. \quad \square$$

The main result easily follows.

**Corollary 22.** *Suppose that  $n \geq 4$ , and*

$$\frac{1}{n^{7/3}} \leq \epsilon_1 < \epsilon_0,$$

where  $\epsilon_0 > 0$  is an absolute constant. Let  $\mathcal{F} \subset S_n$  be a family of permutations with size  $|\mathcal{F}| = n!/2$ , satisfying

$$\mathbb{E}[(f - f_1)^2] = \epsilon_1,$$

where  $f = 2\chi_{\mathcal{F}} - 1$ , and  $f_1$  is the orthogonal projection of  $f$  onto  $U_1$ . Then there exists a family  $\mathcal{G} \subset S_n$  which is a union of  $\lfloor n/2 \rfloor$  disjoint 1-cosets, satisfying

$$|\mathcal{G} \Delta \mathcal{F}| \leq O(\epsilon_1^{1/7})n!.$$

*Proof.* Lemma 21 implies that the set  $S = \{i : \tau_{1i} \geq 1 - 26\epsilon_1^{1/7}\}$  has cardinality which is  $O(\epsilon_1^{1/7})n$ -close to  $n/2$ . By assumption,  $\epsilon_1 \geq n^{-7/3}$ , and therefore

$$\| |S| - \lfloor n/2 \rfloor \| \leq \| |S| - n/2 \| + 1/2 \leq O(\epsilon_1^{1/7})n + 1/2 \leq O(\epsilon_1^{1/7})n + n^{2/3} \leq O(\epsilon_1^{1/7})n.$$

Define

$$\mathcal{G}' = \bigcup_{i \in S} T_{1i}.$$

Note that the  $T_{1i}$  are pairwise disjoint, and so  $|\mathcal{G}'| = (n-1)!|S|$ . By the definition of  $S$ , we have

$$|\mathcal{F} \cap \mathcal{G}'| \geq (1 - 26\epsilon_1^{1/7})(n-1)!|S| = (\frac{1}{2} - O(\epsilon_1^{1/7}))n!.$$

It follows that

$$\begin{aligned} |\mathcal{F} \Delta \mathcal{G}'| &= |\mathcal{F}| + |\mathcal{G}'| - 2|\mathcal{F} \cap \mathcal{G}'| \\ &\leq n!/2 + (1 + O(\epsilon_1^{1/7}))n!/2 - 2(\frac{1}{2} - O(\epsilon_1^{1/7}))n! \\ &= O(\epsilon_1^{1/7})n!. \end{aligned}$$

By adding or deleting

$$\| |S| - \lfloor n/2 \rfloor \| = O(\epsilon_1^{1/7})n$$

$T_{1i}$ 's from  $\mathcal{G}'$ , we may produce a family  $\mathcal{G} \subset S_n$  which is a union of  $\lfloor n/2 \rfloor$  disjoint 1-cosets, and satisfies  $|\mathcal{F} \Delta \mathcal{G}| = O(\epsilon_1^{1/7})n!$ , completing the proof.  $\square$

### 3 Proof in the general case

Up until now, we have only discussed the case  $|\mathcal{F}| = n!/2$ . Much of the argument remains intact for general values of  $c = |\mathcal{F}|/n!$ , although an additional argument is required in the culmination of the proof. Also,  $\epsilon_0$  will now depend upon  $c$ . More concretely, let

$$\eta = \min\{c, 1 - c\}. \quad (9)$$

As we shall see below, for the proof to go through, we will need  $\epsilon_0 = O(\eta^7)$ . Indeed, we shall make the following assumption.

$$\frac{1}{n^{7/3}} \leq \epsilon_1 \leq c_0 \eta^7, \quad (2)'$$

To explore all of these issues, let us follow the existing proof and see how it adapts for arbitrary  $c$ .

**Matrix representation (§ 2.1)** The coefficients  $a_{ij}$  for  $c = 1/2$  were defined so that the following holds.

$$f_1 = \sum_{i,j} a_{ij} T_{ij}. \quad (10)$$

As we remarked in the proof of Lemma 3, our definition of  $a_{ij}$  can be derived from this formula via Fourier inversion. For arbitrary  $c$ , the corresponding definition is

$$a_{ij} = (n-1)\langle f, T_{ij} \rangle - \frac{n-2}{n}(2c-1). \quad (3)'$$

Under this definition, (10) holds. Straightforward calculations yield the following updated versions of Lemma 3 and Lemma 4.

**Lemma 3'.** *We have*

$$f_1 = \sum_{i,j} a_{ij} T_{ij}.$$

Furthermore,

$$\sum_j a_{ij} = 2c - 1 \quad \forall i \in [n], \quad \sum_i a_{ij} = 2c - 1 \quad \forall j \in [n].$$

For each permutation  $\pi$ , we have:

$$f_1(\pi) = \sum_i a_{i\pi(i)}.$$

**Lemma 4'.** *We have*

$$\sum_{i,j} a_{ij}^2 = (n-1)(1-\epsilon_1) - (n-2)(2c-1)^2.$$

**Random restrictions (§ 2.2)** The proof of Lemma 5 becomes more cumbersome. Curiously enough, the variance of the random variable in question is actually maximized when  $c = 1/2$ , and so the bound on the variance holds true for arbitrary  $c$ . Here is the updated version.

**Lemma 5'.** *If  $(X, Y)$  is a restriction, define  $m(X, Y) = \mathbb{E}_{T_{X,Y}}[g_1(X, Y)]$ . Let  $(X, Y) \sim \mathcal{R}$ . Then  $\mathbb{E}_{\mathcal{R}}[m] = c - 1/2$  and  $\mathbb{V}_{\mathcal{R}}[m] \leq 1/2n$ .*

*Proof.* We only give the exact formula for  $\mathbb{V}_{\mathcal{R}}[m]$ :

$$\mathbb{V}_{\mathcal{R}}[m] = \frac{n+1}{4n(n-1)} - \frac{(2c-1)^2}{2n(n-1)} \leq \frac{1}{2n}. \quad \square$$

The proof of Lemma 6 remains the same, adjusting for the general value of  $\mathbb{E}_{\mathcal{R}}[m]$ .

**Lemma 6'.** *Let  $(X, Y) \sim \mathcal{R}$ . With probability at least  $1 - 3\epsilon_1^{1/7}$ ,  $(X, Y)$  satisfies the following properties:*

- (a)  $g(X, Y)$  is  $(\epsilon_1^{4/7}, \epsilon_1^{1/7})$ -almost Boolean.
- (b)  $\mathbb{E}[g_1(X, Y)]$  and  $\mathbb{E}[g_2(X, Y)]$  are  $\epsilon_1^{1/7}$ -close to  $c - 1/2$ .
- (c)  $\mathbb{E}[ (|g(X, Y)| - 1)^2 ] \leq \epsilon_1^{6/7}$ .

We redefine a *typical restriction* as one satisfying these updated properties.

**Decomposition under random restriction (§ 2.3)** Lemma 7 and Lemma 8 are not affected by the value of  $c$ . The value of  $c$  comes into play in the main result of this section, Lemma 9, in two ways. Firstly, the result is affected by the change in Lemma 5. Secondly, there is a hidden dependence of  $\epsilon_0$  on  $c$ , namely

$$\eta > 15\epsilon_1^{1/7}. \quad (11)$$

**Lemma 9'.** *Suppose that  $(X, Y)$  is a typical restriction. Then either  $g_1$  is  $(3\epsilon_1^{1/7}, 19\epsilon_1^{1/7})$ -almost close to  $c - 1/2$  and  $g_2 + c - 1/2$  is  $(4\epsilon_1^{1/7}, 24\epsilon_1^{1/7})$ -almost Boolean, or the same is true with the roles of  $g_1$  and  $g_2$  reversed.*

*Proof.* At the very end of the proof, we need to rule out possibility that for all permutations  $\pi \in S_n$  satisfying

$$g(\pi) \in \{C_4 \pm 1\} \pm 21\epsilon_1^{1/7} \text{ and } g(\pi) \in \{\pm 1\} \pm \epsilon_1^{1/7}, \quad (6)$$

the sign of  $g(\pi)$  is the same. This would imply that  $g$  is  $(8\epsilon_1^{1/7}, \epsilon_1^{1/7})$ -almost close to  $L \in \{\pm 1\}$ . Applying Lemma 7, this in turn would imply that  $\mathbb{E}[g]$  is  $28\epsilon_1^{1/7}$ -close to  $L$ . On the other hand, typicality implies that  $\mathbb{E}[g]$  is  $2\epsilon_1^{1/7}$ -close to  $2c - 1$ . In order to obtain a contradiction, we need to assume that  $2c - 1$  is not  $30\epsilon_1^{1/7}$ -close to  $\pm 1$ . This is equivalent to (11).  $\square$

**Random partitions (§ 2.4)** At the beginning of this section, we defined the concept of a *good partition*, which we now need to update. For a permutation  $\pi \in S_n$  and a partition  $X \subset [n]$ , define

$$P_1 := \sum_{i \in X} a_{i\pi(i)}, \quad P_2 := \sum_{i \in \bar{X}} a_{i\pi(i)}.$$

We say that the partition  $X$  is *good for  $\pi$*  if either  $P_1$  is  $25\epsilon_1^{1/7}$ -close to  $c - 1/2$  and  $P_2$  is  $25\epsilon_1^{1/7}$ -close to  $\{-c - 1/2, 3/2 - c\}$ , or the same is true with the roles of  $P_1$  and  $P_2$  reversed. We say that the permutation  $\pi \in S_n$  is *good* if with probability at least  $4/5$ , a random partition  $X \sim U(2^{[n]})$  is good for  $\pi$  (this is the same definition as before). With the updated definition, Lemma 10 remains the same.

As for Lemma 11, apart from slightly updating the statement, there is also a hidden dependence of  $\epsilon_0$  upon  $c$ , namely

$$\eta = \Omega(\epsilon_1^{1/7}). \quad (12)$$

**Lemma 11'.** *Suppose that  $\pi \in S_n$  is a good permutation. Then for some  $m \in [n]$ ,  $|a_{m\pi(m)}|$  is  $50\epsilon_1^{1/7}$ -close to  $\{2c, 2(1 - c)\}$ , and for  $i \neq m$ ,  $|a_{i\pi(i)}| \leq 50\epsilon_1^{1/7}$ .*

*Proof.* We redefine ‘large’ elements as those which are  $50\epsilon_1^{1/7}$ -close in magnitude to  $\{2c, 2(1 - c)\}$ . Analyzing what happens when a single  $s_i$  switches over, we deduce as in the original proof that each  $s_i$  is either small or large.

For the Berry-Esseen argument, we need a lower bound on  $\sigma^2$ . In the original proof, we deduced such a bound from the fact that with probability at least  $4/5$ , it holds that  $|T - S_0/2| \geq 1/2 - 50\epsilon_1^{1/7}$ . The same argument shows that with probability at least  $4/5$ , it holds that  $|T - S_0/2| \geq \eta - 50\epsilon_1^{1/7}$ , and therefore

$$\sigma^2 \geq \frac{4}{5}(\eta - 50\epsilon_1^{1/7})^2.$$

This implies that

$$\psi = \frac{O(\epsilon_1^{1/7})}{\eta - 50\epsilon_1^{1/7}},$$

where the implied constant does not depend upon  $c$ . Condition (12) guarantees that (say)  $\psi \leq 1/10$ .

The intervals  $I_1, I_2$  retain their length, while for  $I_3$  we get the guarantee

$$|I_3| \geq 2\eta - 50\epsilon_1^{1/7}.$$

Recall inequality (7), from which we would like to derive a contradiction:

$$\frac{\Pr[N \in I_3]}{|I_3|} \geq \min \left( \frac{\Pr[N \in I_1]}{|I_1|}, \frac{\Pr[N \in I_2]}{|I_2|} \right). \quad (7)$$

The left-hand side is at most (roughly)  $1/(5\eta)$ , while the right-hand side is  $\Omega(\epsilon_1^{-1/7})$ . We get a contradiction if  $1/\eta = O(\epsilon_1^{-1/7})$ , which is the same condition as (12).

The rest of the proof goes through without change.  $\square$

**Strong lines (§ 2.5)** As we mentioned in the introduction, this part is almost completely independent of the rest of the proof. All we have to do is redefine a *large entry* so that it conforms to the specification of Lemma 11', that is,  $|a_{ij}|$  is  $50\epsilon_1^{1/7}$ -close to  $\{2c, 2(1-c)\}$ . With this small change, all the results in this section carry through.

**Culmination of the proof (§ 2.6)** This section requires a small overhaul. Whereas for  $c = 1/2$ , a large element was always close to  $\pm 1$ , now all we know is that it is close in magnitude to  $\{2c, 2(1-c)\}$ . Its actual value is therefore close to one of the values  $\{2c, 2(1-c), -2c, -2(1-c)\}$ . Defining  $\tau_{ij}$  as before, this means that  $\tau_{ij}$  is close to one of the values  $\{0, 1, 2c, 2c-1\}$ . Of these, one is always outside  $[0, 1]$  and so cannot occur, and one is a 'medium' value,

$$\gamma := \begin{cases} 2c & \text{if } c < 1/2, \\ 2c-1 & \text{if } c > 1/2, \end{cases}$$

lying inside the interval  $(0, 1)$ . An additional argument is needed to show that such medium values do not actually occur in large quantities on the strong line.

We first rearrange the formula (3)':

$$\tau_{ij} := \frac{|\mathcal{F} \cap T_{ij}|}{(n-1)!} = \frac{2(n-2)c+1}{2(n-1)} + \frac{n}{2(n-1)}a_{ij}. \quad (8)$$

Roughly, we have  $\tau_{ij} \approx a_{ij}/2 + c$ . More precisely, we have the following analogue of Lemma 20.

**Lemma 20'.** *Each  $\tau_{ij}$  is  $2/n$ -close to  $a_{ij}/2 + c$ .*

*If  $a_{ij}$  is large, then  $\tau_{ij}$  is  $26\epsilon_1^{1/7}$ -close to  $\{0, 1, \gamma\}$ .*

As before, without loss of generality, we may assume that row 1 is the strong line. Before proving the analogue of Lemma 21, we need to show that for any two 'reasonable' large entries on row 1, either both are close to  $\{0, 1\}$ , or both are close to  $\gamma$ . Since most entries turn out to be 'reasonable', this implies a dichotomy: either most entries are close to  $\{0, 1\}$ , or most are close to  $\gamma$ . Since the row sums to roughly  $cn$ , the second case cannot occur.

For  $j \in [n]$ , let  $r(j)$  be the probability that a random generalized diagonal passing through  $a_{1j}$  is good. We say that an entry  $a_{1j}$  is *reasonable* if  $a_{1j}$  is large,  $r(j) \geq 4/5$  and  $g(\{1\}, \{j\})$  is  $(1/5, \epsilon_1^{1/7})$ -almost Boolean. (Recall that  $g(\{1\}, \{j\})$  is the function  $f_1$  restricted to permutations in  $T_{1j}$ .)

**Lemma 23.** *Assume that  $\gamma$  is  $156\epsilon_1^{1/7}$ -far from  $\{0, 1\}$ . Let  $j, k \in [n]$  be such that  $a_{1j}$  and  $a_{1k}$  are reasonable. Either both  $\tau_{1j}$  and  $\tau_{1k}$  are  $26\epsilon_1^{1/7}$ -close to  $\gamma$ , or neither of them are.*

*Proof.* If  $\pi \in T_{1j}$  then  $(jk)\pi \in T_{1k}$ . Since  $4 \cdot 1/5 < 1$ , there exists a permutation  $\pi \in T_{1j}$  such that both  $f_1(\pi)$  and  $f_1((jk)\pi)$  are  $\epsilon_1^{1/7}$ -close to  $\pm 1$ , and both  $\pi$  and  $(jk)\pi$  are good. Let  $i = \pi^{-1}(k)$ . Note that

$$(f_1(\pi) - a_{ik}) - (f_1((jk)\pi) - a_{ij}) = a_{1j} - a_{1k} = \frac{2(n-1)}{n}(\tau_{1j} - \tau_{1k}).$$

Since both  $a_{ik}$  and  $a_{ij}$  are small, the left-hand side is  $102\epsilon_1^{1/7}$ -close to  $\{0, \pm 2\}$ , and therefore  $\tau_{1j} - \tau_{1k}$  is  $102\epsilon_1^{1/7}$ -close to  $\{0, \pm n/(n-1)\}$ . Assumption (2)' implies that  $2\epsilon_1^{1/7} > 2/n > 1/(n-1)$ , and so  $\tau_{1j} - \tau_{1k}$  is  $104\epsilon_1^{1/7}$ -close to  $\{0, \pm 1\}$ .

Suppose for a contradiction that  $\tau_{1j}$  is  $26\epsilon_1^{1/7}$ -close to  $\gamma$ , and that  $\tau_{1k}$  is  $26\epsilon_1^{1/7}$ -close to  $b \in \{0, 1\}$ . Then  $\gamma$  is  $156\epsilon_1^{1/7}$ -close to  $b + \{0, \pm 1\} \in \{-1, 0, 1, 2\}$ . Since  $\gamma \in (0, 1)$ ,  $\gamma$  must be  $156\epsilon_1^{1/7}$ -close to  $\{0, 1\}$ , contradicting the assumption of the lemma.  $\square$

Then following lemma says that most of the entries  $a_{1j}$  are reasonable.

**Lemma 24.** *The probability that  $a_{1j}$  is not reasonable for a uniform random  $j \in [n]$  is  $O(\epsilon_1^{1/7})$ .*

*Proof.* Corollary 19 shows that the probability that  $a_{1j}$  is not large is  $O(\epsilon_1^{1/7})$ .

Corollary 12 shows that  $\mathbb{E}[1 - r(j)] \leq 50\epsilon_1^{1/7}$ , and so, by Markov's inequality,

$$\Pr_{j \in [n]} [1 - r(j) > 1/5] < \frac{50\epsilon_1^{1/7}}{1/5} = 250\epsilon_1^{1/7}.$$

Notice that  $f_1$  is  $(\epsilon_1^{5/7}, \epsilon_1^{1/7})$ -almost Boolean, since otherwise  $\mathbb{E}[(f_1 - f)^2] \geq \mathbb{E}[(|f_1| - 1)^2] > \epsilon_1^{5/7} \epsilon_1^{2/7} = \epsilon_1$ . Therefore, by Markov's inequality,

$$\Pr_{j \in [n]} [g(\{1\}, \{j\}) \text{ is not } (1/5, \epsilon_1^{1/7})\text{-almost Boolean}] \leq \frac{\epsilon_1^{5/7}}{1/5} = 5\epsilon_1^{5/7}.$$

The lemma follows from a union bound.  $\square$

We can now prove the analogue of Lemma 21.

**Lemma 21'.** *The number of  $\tau_{1i}$  which are  $51\epsilon_1^{1/7}$ -close to 1 is  $O(\epsilon_1^{1/7})n$ -close to  $cn$ .*

*Proof.* There are two cases, depending on whether  $\gamma$  is  $156\epsilon_1^{1/7}$ -close to  $\{0, 1\}$  or not.

Suppose first that  $\gamma$  is  $156\epsilon_1^{1/7}$ -close to  $\{0, 1\}$ . Let  $N_0$  be the number of  $\tau_{1i}$  which are  $182\epsilon_1^{1/7}$ -close to 0, and let  $N_1$  be the number of  $\tau_{1i}$  which are  $182\epsilon_1^{1/7}$ -close to 1. Lemma 3' shows that

$$T := \sum_{i=1}^n \tau_{1i} = cn.$$

On the other hand,

$$(1 - 182\epsilon_1^{1/7})N_1 \leq T \leq 182\epsilon_1^{1/7}N_0 + (n - N_0) = n - (1 - 182\epsilon_1^{1/7})N_0.$$

Substituting in  $T = cn$  gives:

$$\begin{aligned} N_0 &\leq (1 + O(\epsilon_1^{1/7}))(1 - c)n, \\ N_1 &\leq (1 + O(\epsilon_1^{1/7}))cn. \end{aligned}$$

Corollary 19 and Lemma 20' together show that  $N_0 + N_1 = (1 - O(\epsilon_1^{1/7}))n$ , and so

$$N_1 \geq (1 - O(\epsilon_1^{1/7}))n - N_0 \geq (c - O(\epsilon_1^{1/7}))n.$$

This completes the proof when  $\gamma$  is  $156\epsilon_1^{1/7}$ -close to  $\{0, 1\}$ .

Suppose next that  $\gamma$  is  $156\epsilon_1^{1/7}$ -far from  $\{0, 1\}$ , so that Lemma 23 applies. By Lemma 24, all but  $O(\epsilon_1^{1/7})n$  of the entries in the first row are reasonable. Lemma 23 implies that either all of the corresponding  $\tau_{1i}$  are  $26\epsilon_1^{1/7}$ -close to  $\gamma$ , or none are. In the latter case, they must be  $26\epsilon_1^{1/7}$ -close to  $\{0, 1\}$ , and so an argument similar to the preceding case proves the lemma. It remains to rule out the case that all reasonable  $\tau_{1i}$  are  $26\epsilon_1^{1/7}$ -close to  $\gamma$ .

Suppose for a contradiction that all the reasonable  $\tau_{1i}$  are  $26\epsilon_1^{1/7}$ -close to  $\gamma$ . Assume first that  $c < 1/2$ , so that  $\gamma = 2c$ . Since all but  $O(\epsilon_1^{1/7})n$  of the  $\tau_{1i}$  are reasonable, we have

$$T = \sum_i \tau_{1i} \geq (1 - O(\epsilon_1^{1/7}))2cn.$$

This contradicts the equation  $T = cn$  since  $\epsilon_1$  is small enough (not, here, depending on  $c$ ).

Assume now that  $c > 1/2$ , so that  $\gamma = 2c - 1$ . Then we have

$$(1 - c)n = n - T = \sum_i (1 - \tau_{1i}) \geq (1 - O(\epsilon_1^{1/7}))2(1 - c)n,$$

a contradiction.  $\square$

Note that the assumption  $\epsilon_1^{1/7} = O(\eta)$  does not imply that  $\gamma$  is  $156\epsilon_1^{1/7}$ -far from  $\{0, 1\}$ . Indeed, if  $c$  is close to  $1/2$  then  $\gamma$  is close to  $\{0, 1\}$ . Hence, it is necessary for us to split the proof of Lemma 21' into the two cases above.

The analogue of Corollary 22 now follows, just as before. We state it without any prior assumptions.

**Corollary 22'.** *Suppose that  $n \geq 4$  and*

$$\frac{1}{n^{7/3}} \leq \epsilon_1 \leq c_0 \eta^7, \quad (13)$$

where  $c_0 > 0$  is an absolute constant. Let  $\mathcal{F} \subset S_n$  be a family of permutations with size  $|\mathcal{F}| = c \cdot n!$ , satisfying

$$\mathbb{E}[(f - f_1)^2] = \epsilon_1,$$

where  $f = 2\chi_{\mathcal{F}} - 1$ , and  $f_1$  is the orthogonal projection of  $f$  onto  $U_1$ . Then there exists a family  $\mathcal{G} \subset S_n$  which is a union of  $\lfloor cn \rfloor$  disjoint 1-cosets, satisfying

$$|\mathcal{G} \Delta \mathcal{F}| \leq O(\epsilon_1^{1/7})n!.$$

**Getting rid of the assumptions on  $\epsilon_1$**  Corollary 22' has a drawback: it needs to assume that  $\epsilon_1$  is not too small and not too large. When  $\epsilon_1$  is large enough (depending on  $\eta$ ), the statement holds trivially, so we may focus our attention on the case where  $\epsilon_1$  is small. Intuitively, having  $\epsilon_1$  small should work in our favor. We shall introduce a few artificial errors to increase  $\epsilon_1$ , and then later on take account of them, by introducing an extra error term into our conclusion statement.

We start by showing how to artificially increase  $\epsilon_1$ .

**Lemma 25.** *Let  $\mathcal{F} \subset S_n$ , and let  $v \leq 1/16$ . Then there exists a family  $\mathcal{H} \subset S_n$  such that*

$$|\mathcal{F} \Delta \mathcal{H}| \leq \sqrt{vn}! \quad \text{and} \quad v \leq \mathbb{E}[(h - h_1)^2] \leq (\sqrt{\mathbb{E}[(f - f_1)^2]} + 2\sqrt{v})^2,$$

where  $f = 2\chi_{\mathcal{F}} - 1$ ,  $h = 2\chi_{\mathcal{H}} - 1$  and  $f_1, h_1$  are the projections of  $f, h$  into  $U_1$ .

Moreover, if  $|\mathcal{F}| \geq n!/2$  then  $\mathcal{H} \subset \mathcal{F}$ , and otherwise  $\mathcal{H} \supset \mathcal{F}$ .

*Proof.* By taking complements if necessary, we may assume that  $|\mathcal{F}| \geq n!/2$ . Let  $\text{sgn}(\pi)$  denote the sign of a permutation  $\pi$ . Since  $n \geq 3$ , the sign function is orthogonal to  $U_1$  (this is because the sign representation is not a constituent of the permutation representation), and so

$$\mathbb{E}[(f - f_1)^2] \geq \langle f, \text{sgn} \rangle^2 = \left( \mathbb{E}_{\pi} \text{sgn}(\pi) f(\pi) \right)^2. \quad (14)$$

First, assume that at least half of the permutations in  $\mathcal{F}$  are even. Then the number of these is at least  $n!/4 \geq \sqrt{vn}!$ . Define  $\mathcal{G}$  (and so  $g$ ) by removing  $\sqrt{vn}!$  of them. We have

$$\mathbb{E}_{\pi} \text{sgn}(\pi) g(\pi) = \mathbb{E}_{\pi} \text{sgn}(\pi) f(\pi) - 2\sqrt{v}.$$

Therefore, either  $\mathbb{E}_{\pi} \text{sgn}(\pi) f(\pi) \geq \sqrt{v}$ , or  $\mathbb{E}_{\pi} \text{sgn}(\pi) g(\pi) \leq -\sqrt{v}$ . In the former case, we take  $\mathcal{H} = \mathcal{F}$ , and we are done by the inequality (14). In the latter case, we take  $\mathcal{H} = \mathcal{G}$ . The inequality (14) shows that  $\mathbb{E}[(h - h_1)^2] \geq v$ . Moreover, since projections are contracting, we have

$$\|h - h_1\|_2 \leq \|f - f_1\|_2 + \|(h - h_1) + (f - f_1)\|_2 \leq \|f - f_1\|_2 + \|h - f\|_2 \leq \|f - f_1\|_2 + 2\sqrt{v}.$$

Similarly, if at least half of the permutations in  $\mathcal{F}$  are odd, then the number of these is at least  $n!/4 \geq \sqrt{vn}!$ . Define  $\mathcal{G}$  (and so  $g$ ) by removing  $\sqrt{vn}!$  of them. We have

$$\mathbb{E}_{\pi} \text{sgn}(\pi) g(\pi) = \mathbb{E}_{\pi} \text{sgn}(\pi) f(\pi) + 2\sqrt{v}.$$

Therefore, either  $\mathbb{E}_{\pi} \text{sgn}(\pi) f(\pi) \leq -\sqrt{v}$ , or  $\mathbb{E}_{\pi} \text{sgn}(\pi) g(\pi) \geq \sqrt{v}$ , so we may continue as before.  $\square$

Using this trick and Corollary 22', we get our main theorem in full generality.

**Theorem 2.** Let  $\mathcal{F} \subset S_n$  be a family of permutations with size  $|\mathcal{F}| = c \cdot n!$ , satisfying

$$\mathbb{E}[(f - f_1)^2] = \epsilon_1,$$

where  $f = 2\chi_{\mathcal{F}} - 1$ , and  $f_1$  is the orthogonal projection of  $f$  onto  $U_1$ . Then there exists a family  $\mathcal{G} \subset S_n$  which is a union of  $dn$  disjoint 1-cosets, such that

$$|d - c| = O\left(\epsilon_1^{1/7} + \frac{1}{n^{1/3}}\right) \quad \text{and} \quad \frac{|\mathcal{G} \Delta \mathcal{F}|}{n!} = O\left(\frac{1}{\eta} \left(\epsilon_1^{1/7} + \frac{1}{n^{1/3}}\right)\right),$$

where  $\eta = \min\{c, 1 - c\}$ .

*Proof.* If  $n < 4$  then the theorem is trivial (by taking the absolute constants implied by the  $O$ -terms to be sufficiently large), so we may assume that  $n \geq 4$ . If  $\epsilon_1$  satisfies (2)', then the theorem follows directly from Corollary 22'. Otherwise, there are two cases:  $\epsilon_1$  is too large, and  $\epsilon_1$  is too small. If  $\epsilon_1 > c_0\eta^7$  then the theorem holds, since  $\epsilon_1^{1/7}/\eta > c_0^{1/7}$ , so suppose  $\epsilon_1 < n^{-7/3}$ .

Apply Lemma 25 with  $v = n^{-7/3}$  to obtain a family  $\mathcal{H}$ . The value  $\epsilon_2 = \mathbb{E}[(h - h_1)]^2$  satisfies

$$\frac{1}{n^{7/3}} \leq \epsilon_2 \leq \left(\frac{1}{n^{7/6}} + \frac{2}{n^{7/6}}\right)^2 = \frac{9}{n^{7/3}}.$$

Moreover,  $|\mathcal{F} \Delta \mathcal{H}|/n! \leq n^{-7/6}$  and so  $c_2 := \mathbb{E}[\mathcal{H}]$  satisfies  $|c - c_2| \leq n^{-7/6}$ . Also,  $\eta_2 := \min(c_2, 1 - c_2)$  satisfies  $|\eta - \eta_2| \leq n^{-7/6}$  as well.

There are two cases: either  $\epsilon_2 > c_0\eta_2^7$ , or not. In the first case,  $9/n^{1/3} > c_0^{1/7}\eta_2$ , and so  $\eta_2 = O(n^{-1/3})$ . Hence  $\eta = O(n^{-1/3})$  and so the theorem holds, since  $n^{-1/3}/\eta = \Omega(1)$ .

The more interesting case is when  $\epsilon_2 < c_0\eta_2^7$ . Corollary 22' applies, and we get a family  $\mathcal{G} \subset S_n$  which is the union of  $\lfloor c_2n \rfloor = \lfloor cn \rfloor$  disjoint 1-cosets, where

$$|\mathcal{G} \Delta \mathcal{H}| \leq O(\epsilon_2^{1/7})n!.$$

Since  $\epsilon_2^{1/7} = O(n^{-1/3})$  and  $|\mathcal{F} \Delta \mathcal{H}| \leq n^{-7/6} \cdot n!$ , we have

$$|\mathcal{G} \Delta \mathcal{F}| \leq |\mathcal{G} \Delta \mathcal{H}| + |\mathcal{H} \Delta \mathcal{F}| \leq O(\epsilon_2^{1/7})n! + n^{-7/6} \cdot n! \leq O(n^{-1/3})n!,$$

completing the proof of the theorem. □

**Remark.** When  $\epsilon_1 > c_0\eta^7$ , the error terms  $\epsilon_1^{1/7}/\eta$  ensures that the theorem holds. Other error terms have the same effect, and so other versions of the theorem are possible. For example, instead of  $\epsilon_1^{1/7}/\eta$ , we could have  $\epsilon_1^{1/7} + (\epsilon_1^{1/7}/\eta)^2$ .

## 4 Almost extremal isoperimetric sets in the transposition graph

As explained in the introduction, the main reason for developing Fourier-theoretic stability results, such as the main theorem of this paper, is for applications in extremal combinatorics. Oftentimes, one must struggle to translate the combinatorial information in an extremal problem to the Fourier language, but there is one setting in which the translation is almost immediate (yet may demand certain non-trivial calculations.) That is the setting of normal Cayley graphs on groups, and characterization of the maximum-sized independent sets, or the sets of minimum edge-expansion, in those graphs. See [7] for a more complete description of this. In a nutshell, there are good characterizations relating edge-expansion in graphs to the eigenvalues and eigenvectors of the graph, namely, the theorems of Alon-Milman [1] and Dodziuk [5]. A Cayley graph whose generating set is closed under conjugation is known as a *normal* Cayley graph. For any normal Cayley graph on a group  $\Gamma$ , its eigenspaces are precisely the *isotypical subspaces* of  $\mathbb{C}^\Gamma$  (the subspaces consisting of functions whose Fourier transform is concentrated on a fixed irreducible representation of  $\Gamma$ ). Furthermore, the eigenvalues are given by a formula involving the average of the character of the corresponding representation on the generating set of the graph.

The example of the above phenomenon which we have in mind is the application of the Alon-Milman/Dodziuk theorems to the Cayley graph on  $S_n$  generated by the transpositions. In other words, the graph  $G$  with  $V(G) = S_n$ , and

$$E(G) = \{\{\sigma, \tau\} : \sigma\tau^{-1} \text{ is a transposition}\}$$

— two permutations are joined if they differ by a transposition. For any set  $\mathcal{A} \subset V(G)$ , we let  $\partial\mathcal{A}$  denote the edge-boundary of  $\mathcal{A}$ , i.e. the set of edges between  $\mathcal{A}$  and its complement. As explained in [7], by using Dodziuk/Alon-Milman, the work of Diaconis and Shashahani [4] yields the following theorem:

**Theorem 4** (Diaconis and Shashahani). *Let  $\mathcal{A} \subset S_n$  with  $|\mathcal{A}| = cn!$ . Then*

$$|\partial\mathcal{A}| \geq (1 - c)n|\mathcal{A}|, \tag{15}$$

*with equality if and only if the characteristic vector of  $\mathcal{A}$  belongs to  $U_1$ .*

The characterization of Boolean functions in  $U_1$  given in [9] immediately yields the following characterization of the extremal isoperimetric sets.

**Corollary 26.** *Let  $\mathcal{A} \subset S_n$ , with  $|\mathcal{A}| = cn!$ , and  $|\partial\mathcal{A}| = (1 - c)n|\mathcal{A}|$ . Then  $\mathcal{A}$  is a dictatorship.*

We now want a stability version of this. In [7], Lemma 13, we prove a stability version of Dodziuk/Alon-Milman, which, when combined with the eigenvalue estimates in [4], shows that any set which has edge-boundary close to the minimum, must have its characteristic vector very close (in  $L^2$  norm) to  $U_1$ :

**Theorem 5** (Lemma 13 in [7]). *Let  $\mathcal{A} \subset S_n$  with  $|\mathcal{A}| = cn!$ . If*

$$|\partial\mathcal{A}| \leq (1 - c + \delta_0)n|\mathcal{A}|,$$

*then*

$$E[(f - f_1)^2] \leq \frac{n}{n-2}c\delta_0,$$

*where  $f$  is the characteristic vector of  $\mathcal{A}$ , and  $f_1$  is its projection on  $U_1$ .*

Combining this with Theorem 2 immediately yields the following.

**Theorem 6.** *Let  $\mathcal{A} \subset S_n$  with  $|\mathcal{A}| = cn!$ . If*

$$|\partial\mathcal{A}| \leq (1 - c + \delta_0)n|\mathcal{A}|$$

*then there exists a dictatorship  $\mathcal{B} \subset S_n$  with*

$$\frac{|\mathcal{A} \Delta \mathcal{B}|}{n!} = O\left(\frac{1}{c(1-c)}\left((c\delta_0)^{1/7} + \frac{1}{n^{1/3}}\right)\right).$$

We may apply a perturbation argument similar to the one in [7] to prove the following strengthening of Theorem 6:

**Theorem 7.** *There exists  $\eta_0 > 0$  such that the following holds. Let  $\mathcal{A} \subset S_n$  with  $|\mathcal{A}| = cn!$ , where  $\min(c, 1 - c) \geq \eta_0$ . If*

$$|\partial\mathcal{A}| \leq (1 - c + \delta_0)n|\mathcal{A}|$$

*then there exists a dictatorship  $\mathcal{B} \subset S_n$  with*

$$\frac{|\mathcal{A} \Delta \mathcal{B}|}{n!} = O(\delta_0).$$

This is best possible up to an absolute constant factor, as can be seen by taking

$$\mathcal{A} = T_{11} \cup T_{12} \cup \dots \cup T_{1a} \cup T_{2,a+1} \cup T_{2,a+2} \cup \dots \cup T_{2,a+b},$$

where  $\min\{a/n, 1 - a/n\} = \Omega(1)$  and  $b/a = \Theta(\delta_0)$ .

Therefore, a subset of  $S_n$  with measure bounded away from 0 and 1, which has edge-boundary close to the lower bound (15), must be close in structure to a dictatorship. This is a ‘genuine’ stability result. One may contrast it with the ‘quasi-stability’ result in [7], where we prove that a subset of  $S_n$  with size  $\Theta((n-1)!)$  has edge-boundary close to the minimum iff it is close in structure to a *union* of dictatorships, as opposed to a single dictatorship.

## 5 Conclusion

### Remarks and open questions

The most obvious open question in the context of this trilogy is whether it is possible to prove the common generalization of the main theorems in all three papers. Is it true that no matter what the expectation of  $f$  is, if  $f$  is Boolean and close to  $U_t$ , then it is close to a union of  $t$ -cosets? This surely must be true, but our techniques fall short of proving it. We also believe the correct dependence between the two distances to be linear. We make the following conjecture.

**Conjecture 1.** *Let  $\mathcal{A} \subset S_n$ , and let  $t \in \mathbb{N}$ . Let  $f$  denote the characteristic function of  $\mathcal{A}$ , 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  $\mathcal{B} \subset S_n$  which is a union of  $t$ -cosets, such that*

$$|\mathcal{A} \Delta \mathcal{B}| \leq C_0(\epsilon + 1/n)|\mathcal{A}|,$$

*where  $C_0$  is an absolute constant.*

Another related question involves understanding the precise extremal isoperimetric sets in the transposition graph on  $S_n$ , for all set-sizes. Limor Ben Efraim conjectures that the minimum edge-boundary is always achieved by an initial segment of the lexicographical order on  $S_n$ . (If  $\sigma, \pi \in S_n$ , we say that  $\sigma < \pi$  in the lexicographic order if  $\sigma(j) < \pi(j)$ , where  $j = \min\{i \in [n] : \sigma(i) \neq \pi(i)\}$ .)

It would also be interesting to discover other groups where there is an elegant characterization of Boolean functions whose Fourier support is concentrated on certain irreducible representations.

### Acknowledgment

The authors wish to thank Gil Kalai for many useful conversations, and Café Karkur for its patient support of our research.

## References

- [1] N. Alon, V. D. Milman,  $\lambda_1$ , isoperimetric inequalities for graphs, and superconcentrators, *Journal of Combinatorial Theory, Series B*, 38 (1985), 73–88.
- [2] A. C. Berry, The accuracy of the Gaussian approximations to the sum of independent variates, *Transactions of the American Mathematical Society* 49(1) (1941), 122–136.
- [3] J. Bourgain, On the distribution of the Fourier spectrum of Boolean functions, *Israel Journal of Mathematics* 131 (2002), 269–276.
- [4] P. Diaconis, M. Shahshahani, Generating a random permutation with random transpositions, *Z. Wahrsch. Verw. Gebiete*, Volume 57, Issue 2 (1981), 159–179.
- [5] J. Dodziuk, Difference equations, isoperimetric inequality and transience of certain random walks, *Transactions of the American Mathematical Society* 284 (1984), 787–794.
- [6] D. Ellis, Stability for  $t$ -intersecting families of permutations, *Journal of Combinatorial Theory, Series A* 118 (2011), 208–227.
- [7] D. Ellis, Y. Filmus, E. Friedgut, A quasi-stability result for dictatorships in  $S_n$ , *Combinatorica* 35 (2015), 573–618.
- [8] D. Ellis, Y. Filmus, E. Friedgut, A quasi-stability result for low-degree Boolean functions on  $S_n$ , submitted. arXiv:1511.08694.
- [9] D. Ellis, E. Friedgut, H. Pilpel, Intersecting families of permutations, *Journal of the American Mathematical Society* 24 (2011), 649–682.

- [10] C. G. Esseen, On the Liapunoff limit of error in the theory of probability, *Arkiv för matematik, astronomi och fysik* A28 (1942), 1–19.
- [11] Y. Filmus, A comment on ‘Intersecting Families of Permutations’, manuscript.  
Available at <http://www.cs.toronto.edu/~yuvalf/EFP-comment.pdf>.
- [12] E. Friedgut, Boolean functions with low average sensitivity depend on few coordinates, *Combinatorica* 18(1) (1998), 27–36.
- [13] E. Friedgut, G. Kalai, A. Naor, Boolean functions whose Fourier transform is concentrated on the first two levels, *Advances in Applied Mathematics* 29 (2002), 427–437.
- [14] G. Kindler, R. O’Donnell, Gaussian noise sensitivity and Fourier tails, *IEEE Conference on Computational Complexity* (2012), 137–147.
- [15] G. Kindler, S. Safra, Noise resistant Boolean functions are juntas, manuscript.  
Available at <http://www.cs.huji.ac.il/~gkindler/papers/noise-stable-r-juntas.ps>.
- [16] Z. Lin, Z. Bai, *Probability inequalities*, Springer-Verlag, Berlin, 2010.
- [17] N. Nisan, M. Szegedy, On the degree of Boolean functions as real polynomials, *Computational Complexity* 4 (1994), 301–313.
- [18] B. E. Sagan, *The Symmetric Group: Representations, Combinatorial Algorithms and Symmetric Functions*, Springer-Verlag, New York, 1991. [2nd revised printing, 2001.]
- [19] A. Terras. *Fourier Analysis on Finite Groups and Applications*. London Mathematical Society Student Texts, 43, Cambridge University Press, 1999.