Spectral Methods in Extremal Combinatorics

by

Yuval Filmus

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Computer Science
University of Toronto

Abstract

Spectral Methods in Extremal Combinatorics

Yuval Filmus

Doctor of Philosophy

Graduate Department of Computer Science

University of Toronto

2013

Extremal combinatorics studies how large a collection of objects can be if it satisfies a given set of restrictions. Inspired by a classical theorem due to Erdős, Ko and Rado, Simonovits and Sós posed the following problem: determine how large a collection of graphs on the vertex set $\{1,\ldots,n\}$ can be, if the intersection of any two of them contains a triangle. They conjectured that the largest possible collection, containing $1/8$ of all graphs, consists of all graphs containing a fixed triangle (a *triangle-star*). The first major contribution of this thesis is a confirmation of this conjecture.

We prove the Simonovits–Sós conjecture in the following strong form: the only triangle-intersecting families of measure at least $1/8$ are triangle-stars (*uniqueness*), and every triangle-intersecting family of measure $1/8 - \epsilon$ is $O(\epsilon)$-close to a triangle-star (*stability*).

In order to prove the stability part of our theorem, we utilize a structure theorem for Boolean functions on $\{0,1\}^m$ whose Fourier expansion is concentrated on the first $t+1$ levels, due to Kindler and Safra. The second major contribution of this thesis consists of two analogs of this theorem for Boolean functions on S_m whose Fourier expansion is concentrated on the first two levels.

In the same way that the Kindler–Safra theorem is useful for studying triangle-intersecting families, our structure theorems are useful for studying intersecting families of permutations, which are families in which any two permutations agree on the image of at least one point. Using one of our theorems, we give a simple proof of the following result of Ellis, Friedgut and Pilpel: an intersecting family of permutations on S_m of size $(1-\epsilon)(m-1)!$ is $O(\epsilon)$-close to a double coset, a family which consists of all permutations sending some point i to some point j.

Acknowledgements

הֵן מִדְּרֹשׁ חָכְמָה בְּשָׂרִי נֶאֱכַל —
וּבְשַׂר אֲחֵרִים אַהֲבָה אוֹכֶלֶת.
(אבן גבירול)

I thank my supervisor, Toni Pitassi, for her support and encouragement, for collaborating with me on other research, for introducing me to the right people, and for her careful reading of the manuscript; Ehud Friedgut, for kindling the flame and for suggesting me to come to Toronto; Irit Dinur, for convincing me to stay in the game; Ehud Friedgut and David Ellis, for our joint collaboration, which forms the bulk of this thesis; Steve Cook, for the opportunity to work with him; Faith Ellen, for hosting me in the holidays; Jan Krajíček, for inviting me to Prague; Jakob Nordström, for inviting me to Stockholm; all my friends in Toronto, for making my time here so enjoyable; my American relatives, for providing me with a second home; my overseas family, for not complaining too much; and my uncle Dani, for giving me his copy of Herstein's Topics in Algebra.

Among my local friends, I would like to thank Joel Oren, Natan Dubitski, Justin Ward, Dai Tri Man Lê and Robert Robere for being simultaneously my coauthors and my best friends; my suitemate Jackie Cheung for introducing me to the lore of board gaming, and for linguistic discussions; Kaveh Ghasemloo for religious discussions; Siavosh Benabbas and Per Austrin for film suggestions; Sergei Gorbunov for concert suggestions; and Michalis Famelis for inviting me to Rhythms of Resistance.

Contents

Chapter 1

Introduction

How many edges can a graph on n vertices contain, if it has no triangles? How many sets can a family of subsets of $\{1,\dots,n\}$ consist of, if no set in the family is a subset of another set in the family? How many sets can a family of subsets of $\{1,\dots,n\}$ contain, if any two sets intersect? These are the sort of questions considered in the area of extremal combinatorics.

In 1938[1], Erdős, Ko and Rado [30] proved the following seminal theorem.

Theorem (Erdős–Ko–Rado)**.** *Suppose $\mathcal{F}$ is a family of subsets of $\{1,\dots,n\}$ consisting of sets of size k, such that the intersection of any two sets in $\mathcal{F}$ is non-empty. If $k \le n/2$ then $|\mathcal{F}| \le \binom{n-1}{k-1}$. If furthermore $k < n/2$, then $|\mathcal{F}| = \binom{n-1}{k-1}$ if and only if $\mathcal{F}$ consists of all subsets containing some element $x \in \{1,\dots,n\}$ (such a family is variously known as a star, sunflower, dictatorship, centered family, principal family, kernel system).*

The original theorem was the starting point of an entire research program in extremal combinatorics, which proceeded in various directions.

How big can a family of permutations of S_n be, if any two permutations agree on at least one point? How big can a family of graphs on n vertices be, if the intersection of any two graphs contains a triangle? Suppose $\mathcal{I}$ is a non-empty family of sets which is closed under taking subsets. Is it always the case that no intersecting subfamily of $\mathcal{I}$ is larger than the maximal star contained in $\mathcal{I}$? The last question, known as Chvátal's conjecture [11, Problem

[1] Although the paper [30] dates from 1961, Erdős [29] mentions that the result itself was proved in 1938.

25], remains open.

Focusing on families of subsets of $\{1, \ldots, n\}$, we could require the family to satisfy stronger properties than just being intersecting. What can we say about two families if any set in one family intersects every set in the other family? How large can the family be, if any three sets intersect? How large can an intersecting family be, if it is not a star? How large can the family be, if every two sets intersect in at least two points?

Another direction focuses on *stability*. Suppose that $\mathcal{F}$ is an intersecting family consisting of subsets of $\{1, \ldots, n\}$ of size k. If $k < n/2$ then we know that $|\mathcal{F}| \leq \binom{n-1}{k-1}$, and furthermore the extremal families are stars. What can we say about intersecting families of size $(1-\epsilon)\binom{n-1}{k-1}$? What does an arbitrary intersecting family approximately look like?

The Erdős–Ko–Rado theorem can be proved in various ways. The original proof uses the technique of *shifting*, which modifies the family into a form which is easier to analyze while maintaining its intersecting property. Various other proofs are known: Frankl and Graham [37] describe three additional proofs, and two other proofs appear in [44] and [36].

Friedgut [40] came up with a way of proving a variant of the Erdős–Ko–Rado theorem and some of its generalizations using Fourier analysis. The advantage of his approach is that it automatically yields stability: it shows that families whose size is almost maximal are themselves close to extremal families (families of maximal size). The idea is to derive properties of the Fourier expansion of the characteristic function of an intersecting family. These properties allow us to characterize the Fourier spectra of families whose size is maximal or almost maximal, and through the Fourier spectra, the structure of the families.

The Fourier coefficients of a function on n coordinates are divided into $n+1$ levels. The Fourier expansion of a star is particularly simple: only coefficients on the first two levels are non-zero. Friedgut's method starts by deriving an inequality on the Fourier coefficients of the characteristic function, which immediately implies a tight upper bound on the size of intersecting families. Furthermore, the bound can only be tight if the Fourier expansion of the characteristic function is supported on the first two levels. A simple argument shows that for Boolean functions, that can only be the case if the family is a star.

When the intersecting family has almost maximal size, the method implies that the Fourier expansion of the characteristic function is *concentrated* on the first two levels, that is, all the other coefficients are small in magnitude. Friedgut then applies a classical structure theorem for Boolean functions, the Friedgut–Kalai–Naor theorem, to conclude that the intersecting family is close to a family depending on one element only, which must be a star.

Deza and Frankl [13] proved a theorem which is the analog of Erdős–Ko–Rado for permutations. Two permutations in S_n are *intersecting* if they agree on at least one point. Their theorem shows that an intersecting family of permutations in S_n contains at most $(n-1)!$ permutations. Cameron and Ku [7] showed that this upper bound is achieved only by families of the form $\{\pi \in S_n : \pi(i) = j\}$, known as *double cosets*. Ellis [22] showed that an intersecting family of size $(1-\epsilon)(n-1)!$ must be $O(\epsilon)$-close to a double coset.

Ellis's method is very similar to the method used by Friedgut to prove the Erdős–Ko–Rado theorem. The main difference is that instead of using Fourier analysis on $\{0,1\}^n$, Ellis needs to use Fourier analysis on S_n, which is rather more complicated. At a very coarse level, the Fourier coefficients with respect to S_n can be divided into n levels. The Fourier expansion of double cosets is supported on the first two levels.

Ellis's proof follows the same steps as the previously described one. He starts with an inequality satisfied by the Fourier coefficients, which directly implies the upper bound $(n-1)!$. The Fourier expansion of the characteristic function of a family of size $(n-1)!$ must be supported on the first two levels, and a relatively simple argument shows that this can only happen if the family is a double coset. If the family has size $(1-\epsilon)(n-1)!$, then most of the Fourier expansion is concentrated on the first two levels. At this point Ellis invokes a bootstrapping argument which relies on the fact that the family is intersecting. One of our major results replaces this ad hoc argument with an analog of the Friedgut–Kalai–Naor theorem that works for arbitrary Boolean functions.

1.1 Main results of the thesis

1.1.1 Intersection theorems

A family of graphs on n vertices is called *triangle-intersecting* (respectively, *odd-cycle-intersecting*) if the intersection of any two graphs in the family contains a triangle (respectively, an odd cycle). Chung, Frankl, Graham and Shearer [10] showed that a triangle-intersecting family can contain at most 1/4 of the graphs, and mentioned a conjecture due to Sós and Simonovits that the correct upper bound is 1/8. We strengthen and confirm this conjecture by proving the following (Theorem 4.1):

Upper bound: Every odd-cycle-intersecting family of graphs on n vertices contains at most 1/8 of the graphs.

Uniqueness: The unique families achieving the bound 1/8 are *triangle-juntas*, families formed by taking all supergraphs of a fixed triangle.

Stability: If an odd-cycle-intersecting family contains $1/8-\epsilon$ of the graphs, then it is $O(\epsilon)$-close to a triangle-junta.

We also generalize these results to other settings (Theorems 4.2–4.5).

Sós and Simonovits's conjecture is very natural, but surprisingly its analog fails if we replace the triangle by a path of length 3 (see Section 10.2). Accordingly, our proof makes essential use of the graphical nature of the families.

1.1.2 Structure theorems

The second main part of the thesis concerns structure theorems for Boolean functions on S_n (the set of all permutations on n points), analogous to the theorem of Friedgut, Kalai and Naor.

Friedgut–Kalai–Naor. The *Fourier expansion* of a function $f\colon\{0,1\}^n \to \mathbb{R}$ is

$$f = \sum_{S\subseteq\{1,\dots,n\}} \hat{f}(S)\chi_S,$$

where the functions χ_S (which do not depend on f), known as the *Fourier characters*, form an orthonormal basis for the vector space of all functions on $\{0,1\}^n$ (for a suitable inner product).

The coefficients $\hat{f}(S)$ are known as the *Fourier coefficients*. The *level* of the Fourier coefficient $\hat{f}(S)$ is $|S|$.

If f depends only on the inputs $T \subseteq \{1, \ldots, n\}$, then the Fourier expansion of f is supported on coefficients $\hat{f}(S)$ such that $S \subseteq T$ (that is, the other Fourier coefficients all vanish), and so its Fourier expansion is supported on the first $|T|+1$ levels (including level zero). In particular, if f depends on only one input (we call f a *dictatorship*), then its Fourier expansion is supported on the first two levels. Conversely, if f is a Boolean function whose Fourier expansion is supported on the first two levels, then it is not hard to show that f must be a dictatorship.

Friedgut, Kalai and Naor proved that if f is a *balanced* Boolean function whose Fourier expansion is *concentrated* on the first two levels, then f is *close* to a dictatorship. More formally, suppose that for some $\delta > 0$,

$$\delta < \Pr_{x \in \{0,1\}^n} [f(x) = 1] < 1 - \delta$$

and

$$\sum_{|S|>1} \hat{f}(S)^2 < \delta.$$

Then there is a dictatorship g which differs from f on an $O(\delta)$ fraction of the inputs.

Our results. The Fourier expansion has an analog for functions defined on S_n. That is, there is an orthonormal basis B for the vector space of all functions on S_n, and the Fourier expansion of a function $f\colon S_n \to \mathbb{R}$ is its expansion in terms of the basis B. In contrast to the usual Fourier expansion, in this case there is no natural indexing of the basis functions in B. However, there is a way of partitioning them into n levels.

If the function f depends only on the value of the input permutation π on the indices $I \subseteq \{1, \ldots, n\}$, then its Fourier expansion is concentrated on the first $|I| + 1$ levels. The same is true if f depends only on the value of π^{-1} on the indices I. In particular, if f depends only on $\pi(i)$ or only on $\pi^{-1}(i)$ (we call f a dictator), then its Fourier expansion is supported on the first two levels. Ellis, Friedgut and Pilpel [28] proved the converse: if the Fourier expansion of a Boolean function is supported on the first two levels, then it is a dictatorship. (They also proved a more general result for functions supported on the first $k + 1$ levels.)

We prove two analogs of Friedgut–Kalai–Naor in this setting. Both results concern Boolean functions f whose Fourier expansion is concentrated on the first two levels. The first result (Theorem 7.1) tackles the case where the support of f (the number of permutations π such that $f(\pi) = 1$) is small, and the second result (Theorem 8.1) tackles the case where the support of f is large.

Our first result applies when the support of f has size $c(n-1)!$ for $c = o(n)$. We show that c must be close to an integer, and that f must be close to the characteristic function of a union of double cosets. We cannot conclude that f is close to a dictatorship since the double cosets need not be disjoint (in this context, a dictatorship is a function which, given a permutation π, depends only on $\pi(i)$ for some $i \in [n]$ or on $\pi^{-1}(j)$ for some $j \in [n]$). We also present an application to intersecting families of permutations.

Our second result, applies when the support of f has size $c(n-1)!$ where $\min(c, n-c) = \omega(n^{5/6})$. In contrast to our first result, in this case we are able to show that f must be close to a dictatorship.

1.2 Organization of the thesis

Following the introduction, we present some necessary background material in Chapter 2. The bulk of the thesis is composed of two main parts, *intersection theorems* and *structure theorems*.

The first main part of the thesis contains our result on triangle-intersecting families of graphs. This part begins in Chapter 3, which describes a method devised by Friedgut to prove intersection theorems via Fourier analysis. We describe Friedgut's method using two simple applications: the traffic light puzzle and Friedgut's Fourier-theoretic proof of a generalized version of the Erdős–Ko–Rado theorem. This chapter contains some original material, but mostly follows papers by Friedgut and his coauthors [4, 40]. Our result on triangle-intersecting families appears in Chapter 4. This chapter follows work by the author together with David Ellis and Ehud Friedgut [27]. The first part ends in Chapter 5, in which we prove a version of the Ahlswede–Khachatrian theorem, a generalization of the Erdős–Ko–Rado theorem. This chapter is expository in nature.

The second main part of the thesis contains our stability theorems for Boolean functions on S_n This part begins with Chapter 6, which introduces Fourier analysis on S_n from an unorthodox perspective friendly to theoretical computer science. The subsequent two chapters generalize the theorem of Friedgut, Kalai and Naor to Boolean functions on S_n. Chapter 7 contains our stability result for Boolean functions on S_n of small support. This chapter follows work by the author together with David Ellis and Ehud Friedgut [24]. Chapter 8 contains our stability result for balanced Boolean functions on S_n. This chapter follows work by the author together with David Ellis and Ehud Friedgut [26].

Following the two main parts, we describe in Chapter 9 several applications of extremal combinatorics to theoretical computer science. While the applications we present do not use any of the novel results proven in this thesis, they use results of similar types. This expository chapter serves to relate our results to theoretical computer science.

We conclude the thesis in Chapter 10, reporting on some subsequent research and describing some open problems.

Chapter 2

Preliminaries

2.1 Notation

We will use $\mathbb{N}$ to denote the set of natural numbers (including zero), $\mathbb{Z}$ to denote the set of integers, $\mathbb{R}$ to denote the set of real numbers, and $\mathbb{R}_{\geq 0}$ to denote the set of non-negative real numbers. The ring of integers modulo k is denoted $\mathbb{Z}_k$.

For n a natural number, we will use $[n]$ to denote the set $\{1, \ldots, n\}$. The power set of a set S will be denoted 2^S. We will use $[a, b], [a, b), (a, b], (a, b)$ to denote various intervals of real numbers. A square bracket indicates that the endpoint is contained in the interval. The notation $\lceil x \rfloor$ means the rounding of x, which is an integer m satisfying $|m - x| \leq 1/2$. When $x \in \mathbb{Z} + 1/2$, round arbitrarily down.

A *family of sets on n points* is a subset of $2^{[n]}$. If all sets in the family have cardinality k then the family is *k-uniform*.

A *star* is a family of sets on n points consisting of all sets containing some fixed $i \in [n]$. For a subset $S \subseteq [n]$, an *S-star* is the family of sets on n points consisting of all supersets of S. For an integer $t \geq 1$, a t-star is any S-star for $|S| = t$.

For a proposition P, we define $[\![P]\!]$ by

$$[\![P]\!] = \begin{cases} 1, & \text{if } P \text{ is true,} \\ 0, & \text{if } P \text{ is false.} \end{cases}$$

If $x, y \in \{0, 1\}$, then $x \oplus y$ is their sum modulo 2, also known as exclusive or (XOR). The corresponding operation on sets, symmetric difference, is denoted by Δ.

We denote the transpose of a matrix A by A', and the Hermitian (conjugate transpose) by A^*.

The group of all permutations on $[n]$ is denoted S_n.

Unless otherwise mentioned, all logarithms are natural (to the base e).

We say that a function f is *supported* on a set S if $f(x) = 0$ for all $x \notin S$. A related but distinct informal notion is of a function *concentrated* on a set, whose exact meaning depends on the context. For example, it could mean that $\sum_{x \notin S} f(x)^2$ is small. We will not use this notion when stating theorems, but only when discussing results.

We will use the standard asymptotic notations $O(\cdot)$, $\Omega(\cdot)$, $o(\cdot)$, $\omega(\cdot)$. Unless explicitly mentioned, all the quantities in question are non-negative. We use $\pm O(\cdot)$ to express a quantity whose *absolute* value is $O(\cdot)$. For example, $f = g \pm O(1)$ is the same as $|f - g| = O(1)$.

Unless stated otherwise, the underlying constant in asymptotic notation is universal. To express the fact that the constant depends on some other quantities, we will add them as subscripts. For example, if $f \le C_t$ where C_t is a constant depending only on t, then we can also write $f = O_t(1)$.

2.2 Probability theory

The probability of an event E will be denoted $\Pr[E]$. The expectation (or mean) of a random variable X will be denoted $\mathbb{E}\, X$ or $\mathbb{E}[X]$. If the expectation is taken over a variable A, we will write $\mathbb{E}_A X$. The variance of a random variable X will be denoted $\mathbb{V}\, X$ or $\mathbb{V}[X]$. The binomial distribution on n points with probability p will be denoted $\mathrm{Bin}(n, p)$.

We will use the following classical results.

Theorem (Bonferroni inequalities)**.** *For events $A_1, \ldots, A_n$ on the same probability space,*

$$\sum_i \Pr[A_i] - \sum_{i<j} \Pr[A_i \textit{ and } A_j] \le \Pr[A_1 \textit{ or } \cdots \textit{ or } A_n] \le \sum_i \Pr[A_i].$$

The second inequality is usually known as the *union bound*.

Theorem (Markov's inequality)**.** *If X is a non-negative random variable then*

$$\Pr[X \geq c] \leq \frac{\mathbb{E}[X]}{c}.$$

Theorem (Chebyshev's inequality)**.** *For every random variable X,*

$$\Pr[|X - \mathbb{E}\,X| \geq c] \leq \frac{\mathbb{V}[X]}{c^2}.$$

The following result is usually known in theoretical computer science as Chernoff's inequality or Hoeffding's inequality, and is also a special case of Azuma's inequality. For definiteness, we choose the name Chernoff's inequality.

Theorem (Chernoff's inequality)**.** *Let $X_1, \dots, X_n$ be independent random variables such that X_i lies in some interval of length d_i with probability 1, and let $X = \sum_{i=1}^n X_i$. Then*

$$\Pr[X - \mathbb{E}\,X \geq c] \leq \exp\left(-\frac{2c^2}{\sum_{i=1}^n d_i^2}\right)$$

and

$$\Pr[\mathbb{E}\,X - X \geq c] \leq \exp\left(-\frac{2c^2}{\sum_{i=1}^n d_i^2}\right).$$

All the results listed so far belong to the genre of concentration of measure. The next result belongs to the complementary genre of anti-concentration. Although named after both Berry and Esseen, the version stated below is due to Esseen (a slightly weaker version had been proved a year earlier by Berry, whose work wasn't known to Esseen).

Theorem (Berry–Esseen)**.** *There exists some constant $C > 0$ such that the following is true. Let $X_1, \dots, X_n$ be independent non-constant random variables, and let $X = \sum_{i=1}^n X_i$. Let Y be an independent normally distributed random variable with the same mean and variance as X. For every interval I,*

$$|\Pr[X \in I] - \Pr[Y \in I]| \leq C \frac{\sum_{i=1}^n \mathbb{E}\,|X_i - \mathbb{E}\,X_i|^3}{(\sum_{i=1}^n \mathbb{V}\,X_i)^{3/2}}.$$

2.3 Convex functions

We list two important inequalities on convex functions.

Definition 2.1. A function φ defined on an interval I is *convex* if for every $x, y \in I$ and $t \in [0,1]$, $\varphi(tx + (1-t)y) \le t\varphi(x) + (1-t)\varphi(y)$. ○

Lemma 2.1 (Jensen's inequality)**.** *Suppose φ is a convex function defined on an interval I, and X is a random variable supported on I. Then*

$$\varphi(\mathbb{E}[X]) \le \mathbb{E}[\varphi(X)].$$

Lemma 2.2. *Suppose φ is a convex function defined on an interval $I = [0, M]$. For each n and $x_1, \ldots, x_n \in I$ whose sum is also in I,*

$$\sum_{i=1}^{n} \varphi(x_i) \le \varphi\left(\sum_{i=1}^{n} x_i\right) + (n-1)\varphi(0).$$

Proof. It is enough to prove the case $n = 2$, which we can rewrite as

$$\varphi(x) - \varphi(0) \le \varphi(x+y) - \varphi(y).$$

Let $\psi_x(t) = \varphi(x+t) - \varphi(t)$. Since φ is convex, we have $\varphi'' \ge 0$, and so $\psi_x'(t) = \varphi'(x+t) - \varphi'(x) \ge 0$. Therefore $\psi_x(y) \ge \psi_x(0)$. □

2.4 Measures on families of sets

The Erdős–Ko–Rado theorem (stated in the introduction) concerns k-uniform families of sets on n points, for $k \le n/2$. From our point of view, it is more natural to consider unconstrained families of sets. To this end, we will consider various measures over $2^{[n]}$ which highlight sets of a certain size.

Let $n \ge 1$ be a natural number, and $p \in [0,1]$. The measure μ_p on $2^{[n]}$ (in the sense of measure theory) is defined by its value on singletons, $\mu_p(\{S\}) = p^{|S|}(1-p)^{n-|S|}$. This is a probability measure, that is $\mu_p(2^{[n]}) = 1$, and it has the following probabilistic interpretation. Let S be a random set chosen by putting each $x \in [n]$ in S with probability p independently. For every $\mathcal{F} \subseteq 2^{[n]}$, $\mu_p(\mathcal{F})$ is the probability that S is in $\mathcal{F}$. When $p = 1/2$, $\mu_{1/2}(\mathcal{F}) = 2^{-n}|\mathcal{F}|$. We define $\mu = \mu_{1/2}$ for short, and call this measure the *uniform measure*.

Intuitively, the measure $\mu_{k/n}$ highlights sets of size k: in the probabilistic interpretation, the size of the set S is distributed $\mathrm{Bin}(n, k/n)$, which is concentrated around k. We develop some formal connections between the two settings in the context of intersecting families in Section 3.5. Here we illustrate this connection by giving Katona's beautiful proof of the Erdős–Ko–Rado theorem [58] in both settings.

Classical setting. Let $\mathcal{F}$ be a k-uniform intersecting family of sets on n points, where $k \leq n/2$. Put the numbers 1 to n in a circle in random order, and let S be the set appearing in a random interval of length k on the circle. On the one hand, $\mu(\mathcal{F}) = \Pr[S \in \mathcal{F}]$. On the other hand, for each of the $n!$ possible orders, any two intervals of length k which outline a set in $\mathcal{F}$ must intersect (on the circle). This implies that $\Pr[S \in \mathcal{F}] \leq k/n$.

Probabilistic setting. Let $\mathcal{F}$ be an intersecting family of sets on n points, and let $p \leq 1/2$. Put each number from 1 to n at a random point on the unit-circumference circle, and let S be the set of points appearing in a random interval of length p. On the one hand, $\mu_p(\mathcal{F}) = \Pr[S \in \mathcal{F}]$. On the other hand, for each way of choosing the n points on the circle, any two intervals of length p which outline a set in $\mathcal{F}$ must intersect (on the circle). This implies that $\Pr[S \in \mathcal{F}] \leq p$.

An extension of Katona's argument to cross-intersecting families (pairs of families in which each set of the first family intersects each set of the second family) appears in Section 3.4. An extension to r-wise intersecting families (families in which every r sets have a common element) appears in [34]. Katona's argument does not seem to extend to 2-intersecting families (families in which any two sets intersect in at least two elements); see [53] for relevant work in that direction.

2.5 Fourier analysis

A function $f\colon \{0,1\}^n \to \mathbb{R}$ is called a *function on n bits*. If the range of f is $\{0,1\}$, then f is a *Boolean* function. For a subset $S \subseteq [n]$, its characteristic vector $\mathbf{1}_S$ is defined by $(\mathbf{1}_S)_i = [\![i \in S]\!]$. Similarly, for a family $\mathcal{F}$ of sets on n points, its characteristic function $\mathbf{1}_{\mathcal{F}}\colon \{0,1\}^n \to \{0,1\}$ is defined by $\mathbf{1}_{\mathcal{F}}(\mathbf{1}_S) = [\![S \in \mathcal{F}]\!]$. In the sequel, we identify a subset with its characteristic vector. This association between families of sets and Boolean functions allows us to analyze the former using the latter.

Let $n \geq 1$ be a natural number. The inner product between two functions on n bits is defined by

$$\langle f, g \rangle = \mathop{\mathbb{E}}_{x} f(x)g(x),$$

where x is a uniformly random point in $\{0,1\}^n$. The norm of a function f is $\|f\| = \sqrt{\langle f, f \rangle}$. The Fourier basis functions $\chi_S^{[n]}$ for $S \subseteq [n]$, also known as *Fourier characters*, are defined by

$$\chi_S^{[n]}(T) = (-1)^{|S \cap T|}.$$

In terms of bit vectors,

$$\chi_x^{[n]}(y) = (-1)^{\langle\langle x, y \rangle\rangle},$$

where the inner product between two bit vectors x, y is simply $\langle\langle x, y \rangle\rangle = \sum_i x_i y_i \pmod 2$. When n is understood from the context, as is the case for the rest of this section, we omit the superscript.

The Fourier characters enjoy three basic properties: they are multiplicative, they form a group, and they form a basis.

Lemma 2.3. *Let $x, y, z \in \{0,1\}^n$. We have $\chi_x(y \oplus z) = \chi_x(y)\chi_x(z)$.*

Proof. Easy calculation. □

The analog version for sets states that for all $X, Y, Z \subseteq [n]$, $\chi_X(Y \Delta Z) = \chi_X(Y)\chi_X(Z)$.

Lemma 2.4. *Let $x, y \in \{0,1\}^n$. We have $\chi_x \chi_y = \chi_{x \oplus y}$.*

Proof. Easy calculation. □

The analog version for sets states that for all $X, Y \subseteq [n]$, $\chi_X \chi_Y = \chi_{X \Delta Y}$.

Lemma 2.5. *Let $n \geq 1$ be a natural number. The functions χ_x for $x \in \{0,1\}^n$ form an orthonormal basis to the vector space of all real-valued functions on n bits.*

Proof. Let $x, y \in \{0,1\}^n$. We have

$$\langle \chi_x, \chi_y \rangle = 2^{-n} \sum_z (-1)^{\langle\langle x, z \rangle\rangle} (-1)^{\langle\langle y, z \rangle\rangle} = 2^{-n} \sum_z (-1)^{\langle\langle x \oplus y, z \rangle\rangle}.$$

If $x = y$ then $x \oplus y = 0$, and so $\langle \chi_x, \chi_x \rangle = 1$. If $x \neq y$ then $x \oplus y \neq 0$, say $(x \oplus y)_i = 1$. Therefore

$$\langle \chi_x, \chi_y \rangle = 2^{-n} \sum_{z_1,\ldots,z_{i-1},z_{i+1},\ldots,z_n} (-1)^{\sum_{j \neq i}(x_j \oplus y_j) z_j} \sum_{z_i} (-1)^{z_i} = 0.$$

Since the functions χ_x are orthogonal, they are linearly independent. Since there are 2^n of them, they form a basis. $\square$

Every function f on n bits can be expanded in terms of the Fourier basis (the *Fourier expansion*):

$$f = \sum_{S \subseteq [n]} \hat{f}(S) \chi_S.$$

The numbers $\hat{f}(S)$ are known as *Fourier coefficients*, and the function $\hat{f}$ is called the *Fourier transform*. The Fourier coefficient $\hat{f}(S)$ is said to belong to *level* $|S|$. The first level is level 0, and so on.

The Fourier transform, mapping f to $\hat{f}$, is a linear operator, and so for scalars α, β and functions f, g, the Fourier transform of $h = \alpha f + \beta g$ is $\hat{h} = \alpha \hat{f} + \beta \hat{g}$.

The following lemmas contain some basic properties of the Fourier transform.

Lemma 2.6 (Parseval's identity)**.** *Let f, g be functions on n bits. We have*

$$\langle f, g \rangle = \sum_S \hat{f}(S) \hat{g}(S).$$

In particular,

$$\|f\|^2 = \sum_S \hat{f}(S)^2.$$

Moreover, $\hat{f}(S) = \langle f, \chi_S \rangle$.

Proof. This follows directly from the fact that the Fourier characters form an orthonormal basis. $\square$

Lemma 2.7. *Let f be a Boolean function on n bits. Then*

$$\hat{f}(\varnothing) = \sum_S \hat{f}(S)^2 = \mu(f).$$

Proof. Since $\chi_\varnothing$ is the constant 1 vector, $\hat{f}(\varnothing) = \mathbb{E}_x f(x) = \mu(f)$. Since f is Boolean, $\|f\|^2 = \mathbb{E}_x f(x)^2 = \mathbb{E}_x f(x) = \mu(f)$. The proof is complete using Parseval's identity. $\square$

Lemma 2.8. *Let f be a function on n bits which depends only on a subset S of the coordinates. For any $T \nsubseteq S$, $\hat{f}(T) = 0$.*

Proof. Suppose that f does not depend on the ith coordinate, and $i \in T$. Then

$$\begin{aligned}\hat{f}(T) &= 2^{-n} \sum_x f(x)(-1)^{\sum_{j\in T} x_j} \\ &= 2^{-n} \sum_{x_1,\dots,x_{i-1},x_{i+1},\dots,x_n} (-1)^{\sum_{j\in T\setminus\{i\}} x_j} \\ &\quad (f(x_1,\dots,x_{i-1},0,x_{i+1},\dots,x_n) - f(x_1,\dots,x_{i-1},1,x_{i+1},\dots,x_n)) = 0,\end{aligned}$$

since f does not depend on x_i. □

2.5.1 Generalization to $\mathbb{Z}_k$

Up to now, we have considered functions whose domain was $\{0,1\}^n$. We can think of this domain also as $\mathbb{Z}_2^n$. Replacing 2 with an arbitrary $k \geq 2$, there is an analogous theory of the Fourier transform for functions on $\mathbb{Z}_k^n$. For any two vectors $x, y \in \mathbb{Z}_k^n$, define $\langle\!\langle x, y\rangle\!\rangle = \sum_i x_i y_i$ (all in $\mathbb{Z}_k$), and for any two functions $f, g\colon \mathbb{Z}_k^n \to \mathbb{C}$, define $\langle f, g\rangle = \mathbb{E}_x f(x)\overline{g(x)}$, where x is a random element of $\mathbb{Z}_k^n$, and the bar denotes complex conjugation (we need to consider complex functions since the Fourier characters are complex). The Fourier characters χ_x are indexed by elements $x \in \mathbb{Z}_k^n$, and defined by

$$\chi_x(y) = \omega^{\langle\!\langle x,y\rangle\!\rangle}, \quad \omega = e^{2\pi i/k}.$$

Here we could pick ω to be any primitive kth root of unity. When $k > 2$, the Fourier characters are complex functions. We have the following analogs of Lemma 2.3, Lemma 2.4 and Lemma 2.5, with very similar proofs.

Lemma 2.9. *Let $x, y, z \in \mathbb{Z}_k^n$. We have $\chi_x(y+z) = \chi_x(y)\chi_x(z)$.*

Lemma 2.10. *Let $x, y \in \mathbb{Z}_k^n$. We have $\chi_x\chi_y = \chi_{x+y}$.*

Lemma 2.11. *Let $n \geq 1$ be a natural number. The functions χ_x for $x \in \mathbb{Z}_k^n$ form an orthonormal basis to the vector space of all complex-valued functions on $\mathbb{Z}_k^n$.*

Since the Fourier characters form a basis, every function f can be expanded as

$$f = \sum_{x \in \mathbb{Z}_k^n} \hat{f}(x)\chi_x.$$

We have the following analogues of Lemma 2.6, Lemma 2.7 and Lemma 2.8.

Lemma 2.12 (Parseval's identity)**.** *Let f, g be functions on $\mathbb{Z}_k^n$. We have*

$$\langle f, g\rangle = \sum_S \hat{f}(S)\overline{\hat{g}(S)}.$$

In particular,

$$\|f\|^2 = \sum_S |\hat{f}(S)|^2.$$

Moreover, $\hat{f}(S) = \langle f, \chi_S\rangle$.

Lemma 2.13. *Let f be a Boolean function on $\mathbb{Z}_k^n$. Then*

$$\hat{f}(\varnothing) = \sum_S |\hat{f}(S)|^2 = \mu(f),$$

where $\varnothing$ is the zero vector of length n.

Lemma 2.14. *Let f be a function on $\mathbb{Z}_k^n$ which depends only on a subset S of the coordinates. For any $T \nsubseteq S$, $\hat{f}(T) = 0$.*

2.5.2 Generalization to μ_p

In order to get a bound on the size of intersecting families via Fourier analysis, we will use Lemma 2.7 to relate the Fourier expansion of a family of sets and its size. However, the relation given by the lemma only holds for the measure μ. In order to get an analogous result for μ_p, we will need to develop Fourier analysis with respect to a skewed inner product. While this material is somewhat non-standard, the reader should rest assured that it is only marginally more complicated than standard Fourier analysis.

Let $p \in (0,1)$, and let f, g be two real-valued functions on n bits. The p-skewed inner product of f and g is given by

$$\langle f, g\rangle_p = \mathop{\mathbb{E}}_{x \sim \mu_p} f(x)g(x).$$

The corresponding norm is $\|f\|_p = \sqrt{\langle f, f\rangle}$. Let $q = 1-p$. The Fourier characters $\chi_{S,p}$ for $S \subseteq [n]$ are given by

$$\chi_{S,p}(T) = \left(-\frac{q}{p}\right)^{|S\cap T|}\sqrt{\frac{p}{q}}^{|S|}.$$

This mysterious definition generalizes the usual definition while satisfying the following analog of Lemma 2.5.

Lemma 2.15. *Let $n \geq 1$ be a natural number. The functions $\chi_{S,p}$ for $S \subseteq [n]$ form an orthonormal basis to the vector space of all real-valued functions on n bits, with respect to the p-skewed inner product.*

Proof. Let $S, T \subseteq [n]$. We have

$$\langle \chi_{S,p}, \chi_{T,p}\rangle_p = \sum_U p^{|U|}q^{n-|U|}\left(-\frac{q}{p}\right)^{|S\cap U|+|T\cap U|}\sqrt{\frac{p}{q}}^{|S|+|T|}.$$

If $S = T$ then this simplifies to

$$\begin{aligned}
\|\chi_{S,p}\|_p^2 &= \sum_U p^{|U|}q^{n-|U|}\left(\frac{q}{p}\right)^{2|S\cap U|-|S|} \\
&= \sum_U p^{|U|+|S|-2|S\cap U|}q^{n-|U|-|S|+2|S\cap U|} \\
&= \sum_U p^{|U\Delta S|}q^{n-|U\Delta S|} = \sum_U p^{|U|}q^{n-|U|} = 1.
\end{aligned}$$

If $S \neq T$ then, without loss of generality, there is an element $i \in S \setminus T$. In this case we have

$$\begin{aligned}
\langle \chi_{S,p}, \chi_{T,p}\rangle_p = \sum_{U\subseteq[n]\setminus\{i\}} & p^{|U|}q^{n-|U|}\left(-\frac{q}{p}\right)^{|S\cap U|+|T\cap U|}\sqrt{\frac{p}{q}}^{|S|+|T|} \\
&+ p^{|U|+1}q^{n-|U|-1}\left(-\frac{q}{p}\right)^{|S\cap U|+1+|T\cap U|}\sqrt{\frac{p}{q}}^{|S|+|T|} \\
= \sum_{U\subseteq[n]\setminus\{i\}} & p^{|U|}q^{n-|U|}\left(-\frac{q}{p}\right)^{|S\cap U|+|T\cap U|}\sqrt{\frac{p}{q}}^{|S|+|T|}\left(1+\frac{p}{q}\cdot\left(-\frac{q}{p}\right)\right) = 0.
\end{aligned}$$

□

In Section 2.6.1 we present a construction of the p-skewed Fourier characters using tensor products, which will serve to demystify the explicit formula.

We have the following weakened form of Lemma 2.4.

Lemma 2.16. *Let $S, T \subseteq [n]$. If $S \cap T = \varnothing$ then $\chi_{S,p}\chi_{T,p} = \chi_{S\cup T,p}$.*

Proof. Easy calculation. □

As before, we can expand every function in the new Fourier basis:

$$f = \sum_{S \subseteq [n]} \hat{f}_p(S) \chi_{S,p}.$$

This operation is known as the *p-skewed Fourier transform*, and the functions χ_S are known as the p-skewed Fourier characters of basis vectors. When p is clear from context, we will just say skewed Fourier transform, and drop the subscript p from $\hat{f}_p$ and $\chi_{S,p}$.

Here are the analogues of Lemma 2.6, Lemma 2.7 and Lemma 2.8.

Lemma 2.17 (Parseval's identity). *Let f, g be functions on n bits. We have*

$$\langle f, g \rangle_p = \sum_S \hat{f}_p(S) \hat{g}_p(S).$$

In particular,

$$\|f\|_p^2 = \sum_S \hat{f}_p(S)^2.$$

Moreover, $\hat{f}(S) = \langle f, \chi_{S,p} \rangle_p$.

Proof. This follows directly from orthonormality. □

Lemma 2.18. *Let f be a Boolean function on n bits. Then*

$$\hat{f}(\varnothing) = \sum_S \hat{f}_p(S)^2 = \mu_p(f).$$

Proof. Since $\chi_\varnothing$ is the constant 1 vector, $\hat{f}_p(\varnothing) = \mathbb{E}_{x \sim \mu_p} f(x) = \mu_p(f)$. Since f is Boolean, $\|f\|_p^2 = \mathbb{E}_{x \sim \mu_p} f(x)^2 = \mathbb{E}_{x \sim \mu_p} f(x) = \mu_p(f)$. The proof is complete using Parseval's identity. □

Lemma 2.19. *Let f be a function on n bits which depends only on a subset S of the coordinates. For any $T \nsubseteq S$, $\hat{f}_p(T) = 0$.*

Proof. Suppose that f does not depend on the ith coordinate, and $i \in T$. Then

$$\begin{aligned}
\hat{f}_p(T) &= \sum_{S \subseteq [n]} \mu_p(S) f(S) \left(-\frac{q}{p}\right)^{|S \cap T|} \sqrt{\frac{p}{q}}^{|T|} \\
&= \sum_{S \subseteq [n] \setminus \{i\}} \mu_p(S) \left(-\frac{q}{p}\right)^{|S \cap T|} \sqrt{\frac{p}{q}}^{|T|} \\
&\quad \left[q f(S) + p \cdot \left(-\frac{q}{p}\right) f(S \cup \{i\}) \right] = 0,
\end{aligned}$$

since f does not depend on x_i. □

2.6 Tensor products

The mysterious construction of the Fourier characters with respect to the μ_p measure becomes clearer when its tensorial structure is revealed. In this section, we provide a short introduction to tensor products of vectors and matrices, culminating in the construction of the μ_p measure. Tensor products of matrices will be crucial in Chapters 3 and Chapter 4.

Suppose f is a function on one bit, which we identify with the column vector

$$f = \begin{pmatrix} f(0) \\ f(1) \end{pmatrix}.$$

Consider the linear operator of replacing f with a constant function whose value is always $(f(0) + f(1))/2$. This operator, which we call *averaging*, can be described by a matrix

$$A = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

Now consider a function g on two bits, which we identify with the column vector

$$g = \begin{pmatrix} g(0,0) \\ g(0,1) \\ g(1,0) \\ g(1,1) \end{pmatrix}.$$

The two operators of averaging over the first coordinate and over the second coordinate are described by matrices

$$A_1 = \begin{pmatrix} 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$

The operator A_1 operates on g to produce the function $g_1(x,y) = (g(0,y) + g(1,y))/2$, and A_2 operates on g to produce $g_2(x,y) = (g(x,0) + g(x,1))/2$. If we apply both A_1 and A_2 on g, in any order, we get a constant function whose value is always $(g(0,0)+g(0,1)+g(1,0)+g(1,1))/4$; we say that A_1 and A_2 *commute*, since $A_1A_2 = A_2A_1$.

We can think of the operator A_1 as splitting g into two functions $g(x,0)$ and $g(x,1)$, and applying A on each of the slices. Similarly, A_2 splits g into two functions $g(0,y)$ and $g(1,y)$, and applies A on both slices.

More generally, consider the space of functions on $n+m$ bits. Suppose we have a linear operator B which acts on functions on n bits, and a linear operator C which acts on functions on m bits. We can lift B to functions on $n+m$ bits by applying B separately to each of the 2^m slices formed by fixing the last m coordinates. This function is denoted $B \otimes I_{2^m}$ (here I_{2^m} represents the identity function on 2^m values). Similarly, we can lift C to functions on $n+m$ bits by applying it separately to each of the 2^n slices formed by fixing the first n coordinates. This function is denoted by $I_{2^n} \otimes C$. The two linear operators commute, and we define

$$B \otimes C = (B \otimes I_{2^m})(I_{2^n} \otimes C) = (I_{2^n} \otimes C)(B \otimes I_{2^m}).$$

The reader can check that if we take $C = I_{2^m}$ then we indeed get $B \otimes I_{2^m}$, per its previous definition. The linear operator $B \otimes C$ is called the *tensor product* of B and C.

Coming back to our previous example, $A_1 = A \otimes I_2$, $A_2 = I_2 \otimes A$, and $A_1 A_2 = A_2 A_1 = A \otimes A$. We also write $A \otimes A$ as $A^{\otimes 2}$, the second *tensor power* of A.

Suppose the function g of our running example depended only on the first coordinate, that is, $g(x,y) = f(x)$ for some function f on one bit. In that case, $A_1 g$ depends only on Af, and $(A_1 g)(x,y) = (Af)(x)$. This example can be generalized. Suppose f_1, f_2 are two functions on one bit, and let $g(x,y) = f_1(x) f_2(y)$. In this case we have $(A_1 g)(x,y) = (Af_1)(x) f_2(y)$. More generally, for any two linear operators B, C acting on one-bit functions,

$$((B \otimes C)g)(x,y) = (Bf_1)(x)(Cf_2)(y).$$

We say that g is the tensor product of f_1 and f_2, in symbols $g = f_1 \otimes f_2$.

The definition generalized easily into the case where f_1 is a function on n bits, f_2 is a function on m bits, B acts on n-bit functions, and C acts on m-bit functions. The function $f_1 \otimes f_2$ on $n+m$ bits is defined by

$$(f_1 \otimes f_2)(x,y) = f_1(x) f_2(y), \quad x \in \{0,1\}^n,\, y \in \{0,1\}^m,$$

and we have the identity

$$(B \otimes C)(f_1 \otimes f_2) = (Bf_1) \otimes (Cf_2).$$

A similar property holds for the (usual) inner product of two functions. If f_1, g_1 are functions on n bits and f_2, g_2 are functions on m bits then

$$\begin{aligned}\sum_{x,y}(f_1 \otimes f_2)(x,y)(g_1 \otimes g_2)(x,y) &= \sum_{x,y} f_1(x) f_2(y) g_1(x) g_2(y) \\ &= \left(\sum_x f_1(x) g_1(x)\right)\left(\sum_y f_2(y) g_2(y)\right).\end{aligned}$$

We are interested in the tensor product due to the following elementary result, which allows us to determine the eigenvalues and eigenvectors of a tensor product in terms of the eigenvalues and eigenvectors of its constituents.

Lemma 2.20. *If $Av = \lambda v$ and $Bu = \mu u$ then $(A \otimes B)(u \otimes v) = \lambda\mu(u \otimes v)$.*

Another property which will be useful is the following result concerning linear independence.

Lemma 2.21. *Suppose $v_1, \ldots, v_n$ are linearly independent vectors of length N, and $u_1, \ldots, u_m$ are linearly independent vectors of length M. The nm vectors $v_i \otimes u_j$ of length NM are also linearly independent.*

Proof. Let V be the $n \times N$ matrix whose rows are the vectors $v_1, \ldots, v_n$, and define U similarly. Since $v_1, \ldots, v_n$ are linearly independent, V has a right inverse R_V, which is an $N \times n$ matrix satisfying $VR_V = I_n$. Similarly, U has a right inverse R_U. The matrix $V \otimes U$ contains the nm vectors $v_i \otimes u_j$ as rows, and has a right inverse $R_V \otimes R_U$, showing that the vectors $v_i \otimes u_j$ are linearly independent. □

2.6.1 Tensorial construction of the Fourier transform

The Fourier basis vectors are examples of vectors constructed using tensor products. Consider the ordinary Fourier transform, described in the beginning of Section 2.5. The two building blocks are

$$\psi_0(x) = 1, \quad \psi_1(x) = (-1)^x.$$

For functions on n bits, the Fourier basis vectors $\chi_x^{[n]}$, where $x \in \{0,1\}^n$, are defined by

$$\chi_x^{[n]} = \bigotimes_{i=1}^{n} \psi_{x_i}.$$

Indeed, the resulting function is

$$\chi_x^{[n]}(y) = \prod_{i=1}^{n} \psi_{x_i}(y_i) = (-1)^{\sum_i x_i y_i}.$$

The fact that the functions $\chi_x^{[n]}$ are orthonormal follows from the fact that the functions ψ_0, ψ_1 are. The derivation uses the fact that the underlying measure μ itself is a tensor power (as a function from sets to $\mathbb{R}$): $\mu^{[n]} = (\mu^{[1]})^{\otimes n}$, where $\mu^{[n]}$ is the uniform measure for functions on n bits. Lemma 2.8, which describes the Fourier transform of functions that do not depend on all their arguments, also follows essentially from the tensorial construction of the Fourier basis.

This point of view allows us to explain where the skewed Fourier basis vectors, described in Section 2.5.2, come from. We want to find two one-bit functions $\psi_{0,p}, \psi_{1,}$ which form an orthonormal basis with respect to $\mu_p^{[1]}$. In order for the analog of Lemma 2.8 to hold, we need $\psi_{0,p}$ to be constant (that is clear from considering the case $n = 1$). Since $\mu_p^{[1]}$ is a probability measure and $\psi_{0,p}$ needs to have unit norm, we can conclude that $\psi_{0,p} = 1$. As for $\psi_{1,p}$, orthonormality imposes the two equations (recall $q = 1 - p$)

$$q\psi_{1,p}(0)^2 + p\psi_{1,p}(1)^2 = 1,$$
$$q\psi_{1,p}(0) + p\psi_{1,p}(1) = 0.$$

Substituting the second equation into the first, we deduce that $\psi_{1,p}(0) = \pm\sqrt{p/q}$. In order to be compatible with the case $p = 1/2$, we choose $\psi_{1,p}(0) = \sqrt{p/q}$, and then $\psi_{1,p}(1) = -\sqrt{q/p}$. The diligent reader can check that if we define

$$\chi_{x,p} = \bigotimes_{i} \psi_{x_i,p}$$

then we get the same skewed Fourier basis as in Section 2.5.2.

A generalization of this construction in terms of reversible Markov chains appears in [16].

2.7 Structure theorems for Boolean functions on the Boolean cube

Friedgut's method for intersection problems, described in Chapter 3, relies on structure theorems for Boolean functions on $\{0,1\}^n$ (which is also known as the Boolean cube) in order to prove stability. The simplest theorem, due to Friedgut, Kalai and Naor [42], states that a Boolean function whose Fourier spectrum is concentrated on the first two levels is close to a dictatorship. The original theorem concerned the μ measure. Kindler and Safra [64, 63] generalized the theorem to the μ_p measure for arbitrary p.

Theorem 2.22 (Friedgut–Kalai–Naor)**.** *For every $p \in (0,1)$ there is a constant C_p such that the following is true. If f is a Boolean function satisfying*

$$\sum_{|S|>1} \hat{f}_p(S)^2 < \epsilon,$$

then there is some Boolean function g which depends on at most one coordinate such that $\|f-g\|_p^2 < C_p\epsilon$.

This theorem has many proofs: the original paper [42] already contained two different proofs, and Kindler and Safra [64, 63] contributed another one.

Kindler and Safra [64, 63] extended Friedgut–Kalai–Naor to functions whose Fourier spectrum is concentrated on the first $k+1$ levels. In contrast to the case $k=1$, for $k \geq 2$ it is no longer the case that we can approximate such a function with a function depending on k coordinates. For example, the not-all-equal function on three inputs has the Fourier expansion

$$\mathrm{nae}_3(x,y,z) = \frac{3}{4}\chi_\varnothing - \frac{1}{4}\chi_{\{x,y\}} - \frac{1}{4}\chi_{\{x,z\}} - \frac{1}{4}\chi_{\{y,z\}}.$$

While the Fourier expansion is supported on the first $2+1$ levels, nae_3 depends strongly on all its coordinates. This phenomenon is not without bounds: Nisan and Szegedy [68] proved that a function whose Fourier spectrum is supported on the first $k+1$ levels depends on at most $k2^k$ coordinates. Moreover, it turns out that if the Fourier spectrum is concentrated on the first $k+1$ levels, we can always find a good approximation which depends on $O_k(1)$ coordinates.

Theorem 2.23 (Kindler–Safra [63]). *For every $p \in (0,1)$ and every $k \geq 1$ there are constants $C_{p,k}, M_{p,k}$ such that the following is true. If f is a Boolean function satisfying*

$$\sum_{|S|>k} \hat{f}_p(S)^2 < \epsilon,$$

then there is some Boolean function g which only depends on $M_{p,k}$ coordinates such that $\|f - g\|_p^2 < C_{p,k}\epsilon$. Furthermore, the constants $C_{p,k}, M_{p,k}$ are continuous functions of p.

Kindler and Safra do not make explicit the fact that the constants $C_{p,k}, M_{p,k}$ are continuous in p. However, both are continuous functions of the parameter $\delta_p^{-2} = 1 + (p(1-p))^{-1/2}$, which is the Bonami–Beckner constant for the μ_p measure [69].

Chapter 3

Friedgut's method

Ehud Friedgut has developed a principled method for proving Erdős–Ko–Rado-like results using Fourier analysis. The method first appears in this context in [40], in which Friedgut proves a generalization of Erdős–Ko–Rado: if $\mathcal{F}$ is a t-intersecting family of sets and $p \le 1/(t+1)$ then $\mu_p(\mathcal{F}) \le p^t$ (a t-intersecting family is one in which any two sets intersect in at least t points; an intersecting family is a 1-intersecting family). The second application is to intersecting families of permutations, together with Ellis and Pilpel [28]. They show that a t-intersecting family of permutations on n points contains at most $(n-t)!$ permutations, when n is large enough (two permutations are t-intersecting if they agree on at least t points). We discuss the case $t = 1$ of this result briefly in Chapter 7. The third application, to triangle-intersecting families of graphs, is the subject of the following chapter.

Friedgut's research program was initiated in his earlier paper with Alon, Dinur and Sudakov on graph products [4]. Since this is a particularly simple example, we start our exposition of Friedgut's method by solving the *traffic light puzzle* described in that paper. Following that, we reproduce his proof of the Erdős–Ko–Rado theorem. The chapter continues with a discussion on how results utilizing the μ_p measure are related to results on uniform families of sets (families in which each set contains the same prescribed number of elements). In particular, we relate Friedgut's method to a method due to Lovász [66, 37] and Wilson [79]. The chapter concludes with a short comparison of Friedgut's method and the Lovász theta function.

3.1 Traffic light puzzle

A (red-yellow-green) traffic light is controlled by n tristate switches, in such a way that if you change the state of all the switches, then the color of the traffic light always changes. Show that the light is controlled by a single switch. This puzzle appears in a paper by Greenwell and Lovász [46], where the solution uses induction on n. In this section, we solve the puzzle using Fourier analysis, following [4].

We can identify the set of states of the switches with $\mathbb{Z}_3^n$. Let f, g, h be the characteristic functions corresponding to the states in which the traffic light is red, yellow, green (respectively). Let A be a $3^n \times 3^n$ matrix indexed by $\mathbb{Z}_3^n$ in which $A_{xy} = [\![x_i \neq y_i \text{ for all } i]\!]$. If the current light is red and we change the state of all the switches, then the new light is not red. Therefore, if we regard f as a column vector,

$$f' A f = 0. \tag{3.1}$$

(Here f' is the transpose of f.) Another way of stating this is that any two switch positions in which the traffic light is red must agree on the position of at least one switch. This condition is similar to the condition satisfied by intersecting families of sets.

We are now going to show that the Fourier characters are the eigenvectors of A. This will enable us to deduce from (3.1) information on the Fourier expansion of f. Let $x \in \mathbb{Z}_3^n$, and recall from Section 2.5.1 that $\chi_x(y) = \omega^{\sum_i x_i y_i}$, where $\omega = e^{2\pi i/3}$ is a primitive third root of unity. We have

$$\begin{aligned}(A\chi_x)_y &= \sum_{z \in \mathbb{Z}_3^n} A_{yz} \chi_x(z) \\ &= \sum_{z_1 \neq y_1} \cdots \sum_{z_n \neq y_n} \chi_x(z) \\ &= \sum_{z_1 \neq y_1} \omega^{x_1 z_1} \cdots \sum_{z_n \neq y_n} \omega^{x_n z_n}.\end{aligned}$$

Each of the factors in the sum above is equal to

$$\sum_{z_i \neq y_i} \omega^{x_i z_i} = \left(\sum_{z_i \in \mathbb{Z}_3} \omega^{x_i z_i} \right) - \omega^{x_i y_i}.$$

Now there are two possibilities. If $x_i = 0$, then the sum is equal to $2 = 2\omega^{x_i y_i}$. Otherwise, it is equal to $-\omega^{x_i y_i}$. Let $|x|$ be the number of coordinates such that $x_i \neq 0$. In total, we conclude that

$$(A\chi_x)_y = (-1)^{|x|} 2^{n-|x|} \prod_{i=1}^{n} \omega^{x_i y_i} = 2^n \left(-\frac{1}{2}\right)^{|x|} \chi_x(y).$$

In other words, χ_x is an eigenvector of A with eigenvalue depending on $|x|$:

$$A\chi_x = \lambda_{|x|}\chi_x, \quad \lambda_k = 2^n \left(-\frac{1}{2}\right)^k.$$

Consider now what happens to the Fourier expansion of f after applying A:

$$\begin{aligned} Af &= A \sum_{x \in \mathbb{Z}_3^n} \hat{f}(x)\chi_x \\ &= \sum_{x \in \mathbb{Z}_3^n} \hat{f}(x) A\chi_x \\ &= 2^n \sum_{x \in \mathbb{Z}_3^n} \left(-\frac{1}{2}\right)^{|x|} \hat{f}(x)\chi_x. \end{aligned}$$

In other words, the Fourier expansion of Af is given by

$$\widehat{Af}(x) = 2^n \left(-\frac{1}{2}\right)^{|x|} \hat{f}(x). \tag{3.2}$$

Parseval's identity (Lemma 2.12), together with the fact that f is real, implies that

$$f' A f = \langle Af, f \rangle = \sum_{x \in \mathbb{Z}_3^n} 2^n \left(-\frac{1}{2}\right)^{|x|} |\hat{f}(x)|^2.$$

Combining this with equation (3.1), we get

$$\sum_{x \in \mathbb{Z}_3^n} \left(-\frac{1}{2}\right)^{|x|} |\hat{f}(x)|^2 = 0. \tag{3.3}$$

At this point, we appeal to the properties in Lemma 2.13:

$$\mu(f) = \hat{f}(\varnothing) = \sum_{x \in \mathbb{Z}_3^n} |\hat{f}(x)|^2.$$

(Recall $\varnothing$ is identified with the zero vector of length n.) The idea is to consider (3.3) stated in the following form:

$$|\hat{f}(\varnothing)|^2 = \sum_{x \neq \varnothing} -\left(-\frac{1}{2}\right)^{|x|} |\hat{f}(x)|^2. \tag{3.4}$$

The left-hand side is equal to $\mu(f)^2$. On the right hand side, we have the squared norms of the non-empty Fourier coefficients, scaled by the negated eigenvalues $-2^{-n}\lambda_{|x|}$. The right-hand side has to balance exactly the left-hand side. The squared norms in the right-hand side sum to $\mu(f) - \mu(f)^2$, and the negated eigenvalues are at most $1/2$. Therefore the maximum value that can be achieved by the right-hand side is $(\mu(f) - \mu(f)^2)/2$, and we get the inequality

$$\mu(f)^2 \leq \frac{\mu(f) - \mu(f)^2}{2}. \tag{3.5}$$

Algebra now yields $\mu(f) \leq 1/3$. The general idea of getting an upper bound on $\mu(f)$ via spectral methods, crucial to Friedgut's method, is due to Hoffman [50].

The same argument which gives $\mu(f) \leq 1/3$ also gives $\mu(g), \mu(h) \leq 1/3$. On the other hand, clearly $\mu(f) + \mu(g) + \mu(h) = 1$, and we conclude that $\mu(f) = \mu(g) = \mu(h) = 1/3$, that is, the function mapping switch states to traffic light color is balanced.

If $\mu(f) = 1/3$ then (3.5) is tight. This means that the right-hand side of (3.4) reaches its maximum. Following our reasoning, this can happen only if $\hat{f}(x) = 0$ for $|x| \geq 2$, since $-2^{-n}\lambda_k = -(-1/2)^k < 1/2$ for $k \geq 2$. In other words, the Fourier expansion of f is supported on the first two levels, and f has the general form

$$f = \frac{1}{3}\chi_\varnothing + \sum_{|x|=1} \hat{f}(x)\chi_x.$$

Since f is Boolean, $f^2 = f$. We can calculate the Fourier expansion of f^2 using Lemma 2.10. Let $e_i \in \mathbb{Z}_3^n$ denote the vector whose only non-zero coordinate is $(e_i)_i = 1$. For $i \neq j$ and $c, d \in \{1, 2\}$, the lemma shows that $\chi_{ce_i}\chi_{de_j} = \chi_{de_j}\chi_{ce_i} = \chi_{ce_i+de_j}$, and furthermore these are the only ways in which $\chi_{ce_i+de_j}$ arises in f^2. Therefore

$$\widehat{f^2}(ce_i + de_j) = 2\hat{f}(ce_i)\hat{f}(de_j).$$

Since $f = f^2$, this must be equal to zero. Taking $c = d = 1$, we get that for any two $i \neq j$, either $\hat{f}(e_i) = 0$ or $\hat{f}(e_j) = 0$. That can only be the case if $\hat{f}(e_i) \neq 0$ for at most one value of i. Considering other values of c, d, we get that f has the general form

$$f = \frac{1}{3}\chi_\varnothing + \hat{f}(e_i)\chi_{e_i} + \hat{f}(2e_i)\chi_{2e_i}.$$

Therefore f depends only on one coordinate, namely i. The functions g, h also depend only on one coordinate, which must be the same, since otherwise the corresponding sets are not disjoint. We conclude that the traffic light depends only on the switch i.

3.2 Erdős–Ko–Rado using Fourier analysis

The solution of the traffic light puzzle given in the preceding section is a particularly simple example of Friedgut's method. In this section, we present Friedgut's proof of the Erdős–Ko–Rado theorem and one of its generalizations. Ingredients of this proof will be used in Chapter 4.

Friedgut's method in a nutshell. Let us present the solution to the traffic light puzzle in more abstract terms. We started with a family $\mathcal{F}$ we wanted to analyze (the set of switch configurations resulting in a red light) and its characteristic function f. We found a matrix A such that if $\mathcal{F}$ is intersecting (doesn't contain two configurations that differ on all switches) then

$$f' A f = 0. \tag{3.6}$$

The reason that this equation holds is that for any two intersecting elements (switch configurations) x, y, we have

$$A_{xy} = \mathbf{1}'_{\{x\}} A \mathbf{1}_{\{y\}} = 0.$$

Equation (3.6), in turn, is equivalent to an equation

$$\sum_x \lambda_x |\hat{f}(x)|^2 = 0. \tag{3.7}$$

This is the case since the Fourier basis vectors are the eigenvectors of A. In fact, the symmetry of the construction of A implies that λ_x depends only on $|x|$. In order to derive an upper bound on $|\mathcal{F}|$, we apply the following elementary lemma (due to Hoffman [50]).

Lemma 3.1 (Hoffman's bound)**.** *Let $\lambda_S, x_S \in \mathbb{R}$ for S ranging over some arbitrary index set containing $\varnothing$. Suppose that the following two equations hold, for some $m \in \mathbb{R}$:*

$$m = x_\varnothing = \sum_S x_S^2,$$

$$\sum_S \lambda_S x_S^2 = 0.$$

Let $\lambda_{\min} = \min_S \lambda_S$ be the minimal eigenvalue and let $\lambda_2 = \min_{S:\lambda_S > \lambda_{\min}}$ be the second smallest eigenvalue.

Upper bound: *We have $m \le m_{\max}$, where $m_{\max} = \dfrac{-\lambda_{\min}}{\lambda_\varnothing - \lambda_{\min}}$.*

Uniqueness: *If $m = m_{\max}$ then $x_S = 0$ unless $S = \varnothing$ or $\lambda_S = \lambda_{\min}$.*

Stability: *If $m \neq 0$ then*

$$\sum_S{}' x_S^2 \leq \frac{-\lambda_{\min}}{\lambda_2 - \lambda_{\min}}(m_{\max} - m),$$

where the primed sum ranges over all $S \neq \varnothing$ such that $\lambda_S \neq \lambda_{\min}$.

Proof. The first equation implies that $\sum_{S \neq \varnothing} x_S^2 = m - m^2$. The second equation gives

$$0 = \sum_S \lambda_S x_S^2 \geq \lambda_\varnothing m^2 + \lambda_{\min}(m - m^2).$$

We can assume $m \neq 0$, and so we can divide by m to obtain

$$0 \geq \lambda_\varnothing m + \lambda_{\min}(1 - m) = \lambda_{\min} + (\lambda_\varnothing - \lambda_{\min})m.$$

The upper bound immediately follows, as does uniqueness. (Note that if $\lambda_\varnothing = \lambda_{\min}$ then the second equation cannot possibly hold.)

In order to derive stability, we split the sum in the second equation into three parts:

$$0 = \sum_S \lambda_S x_S^2 \geq \lambda_\varnothing m^2 + \lambda_{\min} \sum_{S : \lambda_S = \lambda_{\min}} x_S^2 + \lambda_2 \sum_S{}' x_S^2.$$

Let $\sum'_S x_S^2 = rm$. Dividing by m,

$$\begin{aligned} 0 &\geq \lambda_\varnothing m + \lambda_{\min}(1 - m - r) + \lambda_2 r \\ &= \lambda_{\min} + (\lambda_\varnothing - \lambda_{\min})m + (\lambda_2 - \lambda_{\min})r \\ &= (\lambda_\varnothing - \lambda_{\min})(m - m_{\max}) + (\lambda_2 - \lambda_{\min})r. \end{aligned}$$

Therefore

$$r \leq \frac{\lambda_\varnothing - \lambda_{\min}}{\lambda_2 - \lambda_{\min}}(m_{\max} - m).$$

The stability bound, which is a bound on rm, follows from

$$(\lambda_\varnothing - \lambda_{\min})m \leq (\lambda_\varnothing - \lambda_{\min})m_{\max} = -\lambda_{\min}. \qquad \square$$

In our case, $m = \mu(\mathcal{F})$ and $x_S = |\hat{f}(S)|$, and the first equation holds due to Lemma 2.6 or one of its analogues. For the traffic light puzzle, $\lambda_\varnothing = 1$ and $\lambda_{\min} = -1/2$, and so we get the bound

$m \le (1/2)/(1+1/2) = 1/3$. Also, $\lambda_S = \lambda_{\min}$ for $|S| = 1$, from which we deduce the structure of tight solutions.

Hoffman's bound also contains a third part, stability. While this part isn't needed for solving the traffic light puzzle, we will use it to derive the stability version of Erdős–Ko–Rado.

We proceed to give Friedgut's proof of the Erdős–Ko–Rado theorem, following his paper [40]. We will prove the Erdős–Ko–Rado theorem in the following form: if $\mathcal{F}$ is an intersecting family of sets (of arbitrary size) and $p \le 1/2$ then $\mu_p(\mathcal{F}) \le p$. (Recall that a family $\mathcal{F}$ is intersecting if any two sets in $\mathcal{F}$ intersect.) Section 3.5 explains the connection to the usual formulation of the Erdős–Ko–Rado theorem, which was given in Chapter 1.

Since we are interested in the μ_p-measure of $\mathcal{F}$, the eigenvectors of the operator A which we propose to construct should be the p-skewed Fourier basis vectors, described in Section 2.5.2. This will allow us to apply Lemma 2.17 and obtain a bound on $\mu_p(\mathcal{F})$. Following Friedgut's method, we need $A_{xy} = 0$ whenever x, y intersect.

In order to construct the operator A, we focus on the case $n = 1$. The operator $A^{[1]}$ needs to have the form

$$A^{[1]} = \begin{pmatrix} \alpha & \beta \\ \gamma & 0 \end{pmatrix}.$$

The two eigenvectors are

$$\chi^{[1]}_{\varnothing,p} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \chi^{[1]}_{\{1\},p} = \begin{pmatrix} \sqrt{p/q} \\ -\sqrt{q/p} \end{pmatrix}, \quad q = 1 - p.$$

Hoffman's bound is oblivious to multiplication by any positive constant, and intuitively we should get a better bound if $\lambda_\varnothing$ is positive. We arbitrarily set $\lambda_\varnothing = 1$, which implies that $\gamma = 1$ and $\beta = 1 - \alpha$. Therefore

$$A^{[1]}\chi^{[1]}_{\{1\},p} = \begin{pmatrix} 1-\beta & \beta \\ 1 & 0 \end{pmatrix}\begin{pmatrix} \sqrt{p/q} \\ -\sqrt{q/p} \end{pmatrix} = \begin{pmatrix} -(\sqrt{p/q}+\sqrt{q/p})\beta + \sqrt{p/q} \\ \sqrt{p/q} \end{pmatrix}.$$

For this to be a multiple of $\chi^{[1]}_{\{1\},p}$, we must have

$$-(\sqrt{p/q}+\sqrt{q/p})\beta + \sqrt{p/q} = \frac{\sqrt{p/q}}{-\sqrt{q/p}} \cdot \sqrt{p/q} = -(p/q)\sqrt{p/q}.$$

Carrying the calculation out,

$$\beta = \frac{\sqrt{p/q} + (p/q)\sqrt{p/q}}{\sqrt{p/q} + \sqrt{q/p}} = \frac{1 + p/q}{1 + q/p} = \frac{p}{q}.$$

The corresponding eigenvalue can be gleaned from the second entry, namely

$$\lambda_{\{1\}} = \frac{\sqrt{p/q}}{-\sqrt{q/p}} = -\frac{p}{q}.$$

Concluding, we have proved the following lemma.

Lemma 3.2. *Let $A^{[1]}$ be the matrix defined by*

$$A^{[1]} = \begin{pmatrix} 1 - p/q & p/q \\ 1 & 0 \end{pmatrix}.$$

The matrix has two eigenvectors: $\chi_{\varnothing}^{[1]}$ with eigenvalue $\lambda_{\varnothing} = 1$ and $\chi_{\{1\}}^{[1]}$ with eigenvalue $\lambda_{\{1\}} = -(p/q)$.

In order to construct the matrix A for general n, we simply take the nth tensor power of A, and use Lemma 2.20 to determine all the eigenvectors and eigenvalues of A.

Lemma 3.3. *Let $n \geq 1$ be an integer, and $A^{[n]} = (A^{[1]})^{\otimes n}$. The eigenvectors of $A^{[n]}$ are $\chi_{S,p}^{[n]}$ for $S \subseteq [n]$, with corresponding eigenvalues $\lambda_S = (-(p/q))^{|S|}$. Furthermore, if $S, T \subseteq [n]$ intersect then $\mathbf{1}_S' A^{[n]} \mathbf{1}_T = 0$.*

Proof. The claim about the eigenvectors and eigenvalues follows directly from Lemma 3.2 by way of Lemma 2.20 (see Section 2.6). For the other claim, notice that

$$\mathbf{1}_S^{[n]} = \bigotimes_{i=1}^{n} \mathbf{1}_{S \cap \{i\}}^{[1]}, \quad \mathbf{1}_T^{[n]} = \bigotimes_{i=1}^{n} \mathbf{1}_{T \cap \{i\}}^{[1]},$$

where we identify $\{i\}$ with $\{1\}$ on the right-hand side. The formulas for tensor products described in Section 2.6 imply that

$$(\mathbf{1}_S^{[n]})' A^{[n]} \mathbf{1}_T^{[n]} = \prod_{i=1}^{n} (\mathbf{1}_{S \cap \{i\}})' A^{[1]} \mathbf{1}_{T \cap \{i\}}.$$

The factor corresponding to any $i \in S \cap T$ in this product is zero, and the second claim follows. □

The fact that the eigenvectors of $A^{[n]}$ are the p-skewed Fourier basis vectors allows us to obtain an analogue of (3.7).

Lemma 3.4. *Let $\mathcal{F}$ be an intersecting family on n points, and f its characteristic function. For every $p \in [0,1)$ and $q = 1-p$,*

$$\sum_{S \subseteq [n]} \lambda_S \hat{f}_p^2(S) = 0, \text{ where } \lambda_S = \left(-\frac{p}{q}\right)^{|S|}. \tag{3.8}$$

Proof. First, we claim the analogue of (3.6), namely

$$\langle A^{[n]} f, f \rangle_p = 0. \tag{3.9}$$

Indeed,

$$\langle A^{[n]} f, f \rangle_p = \sum_{S,T \in \mathcal{F}} \langle A^{[n]} \mathbf{1}_S, \mathbf{1}_T \rangle_p = \sum_{S,T \in \mathcal{F}} \mu_p(T) \mathbf{1}_T' A^{[n]} \mathbf{1}_S = 0,$$

using Lemma 3.3 and the fact that $\mathcal{F}$ is intersecting.

Consider the Fourier expansion of f,

$$f = \sum_{S \subseteq [n]} \hat{f}_p(S) \chi_{S,p}^{[n]}.$$

Applying $A^{[n]}$, we get (using Lemma 3.3)

$$A^{[n]} f = \sum_{S \subseteq [n]} \lambda_S \hat{f}_p(S) \chi_{S,p}^{[n]}.$$

The orthonormality of the Fourier characters now immediately implies (3.8). □

At this stage, we can already derive the upper bound in Erdős–Ko–Rado.

Lemma 3.5. *Let $\mathcal{F}$ be an intersecting family on n points. For every $p \le 1/2$, $\mu_p(\mathcal{F}) \le p$.*

Proof. When $p \le 1/2$, $p \le q$, and so $\min_S \lambda_S = -p/q$. Applying Hoffman's bound using $x_S = \hat{f}_p(S)$ and $m = \mu_p(\mathcal{F})$, we deduce that

$$\mu_p(\mathcal{F}) \le \frac{p/q}{1 + p/q} = p.$$ □

When $p < 1/2$, the eigenvalue λ_S decreases in magnitude with $|S|$, and so we can apply Hoffman's bound to deduce the structure of the Fourier expansion of $\mathbf{1}_{\mathcal{F}}$ when $\mu_p(\mathcal{F}) = p$. In order to conclude that $\mathcal{F}$ is a star, we need the following result.

Lemma 3.6. *Suppose f is a Boolean function on n bits whose p-skewed Fourier expansion (for some $p \in (0,1)$) is supported on the first two levels, that is $\hat{f}_p(S) = 0$ for $|S| > 1$. Then either f is constant or it depends on one coordinate.*

Proof. The Fourier expansion of f is

$$f = \hat{f}(\varnothing) + \sum_{i=1}^{n} \hat{f}(\{i\})\chi_{\{i\}}.$$

Since f is Boolean, $f^2 = f$, where

$$f^2 = \hat{f}(\varnothing) + \sum_{i=1}^{n}[\hat{f}(\{i\})^2\chi_{\{i\}}^2 + 2\hat{f}(\varnothing)\hat{f}(\{i\})\chi_{\{i\}}] + 2\sum_{i<j}\hat{f}(\{i\})\hat{f}(\{j\})\chi_{\{i,j\}},$$

using Lemma 2.16. Since $\chi_{\{i\}}^2$ depends only on coordinate i, its Fourier expansion is supported by the coefficients at $\varnothing$ and $\{i\}$. Therefore the only contribution to $\widehat{f^2}(\{i,j\})$ is $\hat{f}(\{i\})\hat{f}(\{j\})$, and we conclude that $\hat{f}(\{i\})\hat{f}(\{j\}) = 0$. This implies that $\hat{f}(\{i\}) \neq 0$ for at most one index i, and so either f is constant or it depends only on i. □

Uniqueness now follows from the uniqueness part of Hoffman's bound.

Lemma 3.7. *Let $\mathcal{F}$ be an intersecting family on n points, and $p \in (0,1/2)$. Then $\mu_p(\mathcal{F}) = p$ if and only if $\mathcal{F}$ is a star.*

Proof. If $\mathcal{F}$ is a star then it is easy to check that $\mu_p(\mathcal{F}) = p$ for all p. Now suppose $p < 1/2$ and $\mu_p(\mathcal{F}) = p$. Since $p < 1/2$, $p/q < 1$, and so $\lambda_S = \lambda_{\min}$ only for singletons. The uniqueness part of Hoffman's bound implies that the p-skewed Fourier expansion of $\mathbf{1}_{\mathcal{F}}$ is supported on the first two levels. Lemma 3.6 implies that $\mathcal{F}$ depends on at most one point. Since $\mathcal{F}$ is a non-empty intersecting family, it must be a star. □

In order to derive stability, we simply replace Lemma 3.6 with the much stronger Friedgut–Kalai–Naor theorem (Theorem 2.22 on page 23).

Lemma 3.8. *Let $\mathcal{F}$ be an intersecting family on n points, and $p \in (0,1/2)$. If $\mu_p(\mathcal{F}) \geq p - \epsilon$ then $\mu_p(\mathcal{F}\Delta\mathcal{G}) = O_p(\epsilon)$ for some star $\mathcal{G}$.*

Proof. We can assume that $n \geq 3$ and that ϵ is "small enough". For large ϵ, we can choose the implied constant in $O_p(\epsilon)$ so that the result holds trivially for any choice of $\mathcal{G}$.

When $p < 1/2$, $p/q < 1$ and so $\lambda_2 = -(p/q)^3$. Since λ_2 and $\lambda_{\min}$ don't depend on n, the stability part of Hoffman's bound implies that

$$\sum_{|S|>1} \hat{f}_p(S)^2 = O_p(\epsilon),$$

where $f = \mathbf{1}_{\mathcal{F}}$. Theorem 2.22 implies that $\|f - g\|_p^2 = O_p(\epsilon)$ for some Boolean function g depending on at most one coordinate. Since $\|f - 0\|_p^2 = \mu_p(\mathcal{F}) \geq p - \epsilon$ and $\|f - 1\|_p^2 \geq 1 - p$, if ϵ is small enough then g cannot be constant. Thus either g or $1 - g$ is the characteristic function of a star. If $1 - g$ is the characteristic function of a star then $\|g\|_p^2 = 1 - p$ and so $\|f - g\|_p^2 \geq (\|f\|_p - \|g\|_p)^2 \geq (\sqrt{1-p} - \sqrt{p})^2$, which again cannot happen if ϵ is small enough. We conclude that g is the characteristic function of a star, and so we can take $\mathcal{G}$ to be the family satisfying $g = \mathbf{1}_{\mathcal{G}}$. □

Putting all three parts together, we obtain a stability version of Erdős–Ko–Rado.

Theorem 3.9. *Let $\mathcal{F}$ be an intersecting family on n points, and $p \leq 1/2$.*

Upper bound: *$\mu_p(\mathcal{F}) \leq p$.*

Uniqueness: *If $p < 1/2$, then $\mu_p(\mathcal{F}) = p$ if and only if $\mathcal{F}$ is a star.*

Stability: *If $p < 1/2$ and $\mu_p(\mathcal{F}) \geq p - \epsilon$ then $\mu_p(\mathcal{F}\Delta\mathcal{G}) = O_p(\epsilon)$ for some star $\mathcal{G}$.*

We remark that other stability versions exist, see for example Keevash and Mubayi [60] (who use a result of Frankl [35]) and Keevash [59].

3.3 t-intersecting families

Friedgut's paper [40] goes on to prove a generalization of Erdős–Ko–Rado to t-intersecting families, which are families in which any two sets contain at least t points in common. The generalization is as follows. For every $p \leq 1/(t+1)$, if $\mathcal{F}$ is t-intersecting then $\mu_p(\mathcal{F}) \leq p^t$; when $p < 1/(t+1)$, this is tight only for t-stars, and we have stability. For a discussion of what happens when $p > 1/(t+1)$, see Section 10.1.

Friedgut originally phrased his proof in terms of rings satisfying $X^t = 0$. Later on, he found a more elementary presentation [41], which we follow here.

The proof of Lemma 3.5 uses the following properties of A to deduce the upper bound part of Erdős–Ko–Rado:

- For every intersecting family $\mathcal{F}$, $f'Af = 0$, where $f = \mathbf{1}_{\mathcal{F}}$.

- The eigenvectors of $\mathcal{F}$ are the p-skewed Fourier characters $\chi_{S,p}$.
- The eigenvalue corresponding to $\chi_{\varnothing,p}$ is 1.
- When $p \le 1/2$, all other eigenvalues are at least $-p/q$.

In our case, we want to get a bound of p^t instead of p, and so we need to construct a matrix A_t with the following properties:

- For every t-intersecting family $\mathcal{F}$, $f' A_t f = 0$, where $f = \mathbf{1}_{\mathcal{F}}$.
- The eigenvectors of $\mathcal{F}$ are the p-skewed Fourier characters $\chi_{S,p}$.
- The eigenvalue corresponding to $\chi_{\varnothing,p}$ is 1.
- When $p \le 1/(t+1)$, all other eigenvalues are at least $-p^t/(1-p^t)$.

We construct A_t in two stages. First, we find a large linear space of matrices satisfying the first two properties. Then, we identify a matrix inside this linear space which satisfies the other two properties. To this end, we make the following definition.

Definition 3.1. Let $n \ge 1$ be an integer and $p \in (0,1)$. We say that a matrix B is *admissible* for n,p if it satisfies the following two properties:

Intersection property If $|S \cap T| \ge t$ for some $S,T \subseteq [n]$ then $\mathbf{1}_S' B \mathbf{1}_T = 0$.

Eigenvector property The eigenvectors of $\mathcal{F}$ are the p-skewed Fourier characters $\chi_{S,p}$. ○

Fix n and p, and let $q = 1-p$. We already know one matrix which is admissible, namely the matrix $A^{[n]} = (A^{[1]})^{\otimes n}$ constructed in the preceding section. The eigenvector property is satisfied because the eigenvectors of each of the factors $A^{[1]}$ are the one-dimensional p-skewed Fourier characters. Furthermore, the zero in the bottom-right corner of the ith factor $A^{[1]}$ guarantees that $\mathbf{1}_S' A^{[n]} \mathbf{1}_T = 0$ whenever $i \in S \cap T$.

If S,T are t-intersecting, then we know that for all sets $J \subseteq [n]$ of size at most $t-1$, the sets $S \setminus J, T \setminus J$ intersect. Therefore if we replace up to $t-1$ factors $A^{[1]}$ with arbitrary other factors, then the intersection property still holds. If the eigenvectors of each of these other factors are the one-dimensional p-skewed Fourier characters, as is the case for the identity matrix, for example, then the eigenvector property holds. This gives us a large collection of admissible matrices.

Lemma 3.10. *Let $J \subseteq [n]$ be a set of size $|J| < t$. Define*

$$B_{J,i} = \begin{cases} A^{[1]}, & \text{if } i \notin J, \\ I_2, & \text{if } i \in J, \end{cases}$$

$$B_J = \bigotimes_{i=1}^{n} B_{J,i}.$$

Here I_2 is the 2×2 identity matrix. The matrix B_J is admissible, and the eigenvalue corresponding to $\chi_{S,p}$ is

$$\lambda_{J,S} = \left(-\frac{p}{q}\right)^{|S \setminus J|}.$$

Proof. The eigenvector property and the formula for the eigenvalues follow directly from Lemma 3.2 by way of Lemma 2.20. For the intersection property, recall that

$$\mathbf{1}_S^{[n]} = \bigotimes_{i=1}^{n} \mathbf{1}_{S \cap \{i\}}^{[1]}, \quad \mathbf{1}_T^{[n]} = \bigotimes_{i=1}^{n} \mathbf{1}_{T \cap \{i\}}^{[1]}.$$

The formulas for tensor products described in Section 2.6 imply that

$$(\mathbf{1}_S^{[n]})' B_J \mathbf{1}_T^{[n]} = \prod_{i \in J} (\mathbf{1}_{S \cap \{i\}})' B_{J,i} \mathbf{1}_{T \cap \{i\}}.$$

If $|S \cap T| \geq t$ then there must be some $i \in (S \cap T) \setminus J$. Then $B_{J,i} = A^{[1]}$, and the factor corresponding to i in the product is zero, proving the intersection property. □

Given an admissible matrix B with given $\lambda_\varnothing$ and $\lambda_{\min}$, we can create another admissible matrix with the same (or better) properties using the process of *symmetrization*. For every permutation $\pi \in S_n$, define $\pi(B)$ in the natural way: $\pi(B)_{S,T} = B_{\pi(S),\pi(T)}$, and let $B' = \mathbb{E}_\pi \pi(B)$. It is easy to check that $\lambda_\varnothing(B') = \lambda_\varnothing(B)$ and $\lambda_{\min}(B') \geq \lambda_{\min}(B)$. We conclude that it is enough to focus on symmetric admissible matrices. Applying the same process to the B_J, we get a dramatically smaller collection of symmetric admissible matrices.

Lemma 3.11. *Let $k < t$, and define*

$$B_k = \binom{n}{k}^{-1} \sum_{J \subset [n] : |J| = k} B_J.$$

The matrix B_k is admissible, and the eigenvalue corresponding to $\chi_{S,p}$ is

$$\lambda_{k,S} = \left(-\frac{p}{q}\right)^{|S|} P_k(|S|),$$

where P_k is some polynomial of degree k (which depends on p).

Proof. Given S, enumerating over all possible sizes of $|S \cap J|$, we get

$$\lambda_{k,S} = \binom{n}{k}^{-1} \left(-\frac{p}{q}\right)^{|S|} \sum_{j=0}^{k} \binom{|S|}{j}\binom{n-|S|}{k-j}\left(-\frac{q}{p}\right)^{j},$$

which leads to

$$P_k(s) = \binom{n}{k}^{-1} \sum_{j=0}^{k} \binom{s}{j}\binom{n-s}{k-j}\left(-\frac{q}{p}\right)^{j}.$$

If we open up the binomial coefficients, we see that P_k is a polynomial of degree at most k. Moreover, the coefficient of s^k is

$$\binom{n}{k}^{-1} \sum_{j=0}^{k} \frac{1}{j!}\frac{(-1)^{k-j}}{(k-j)!}\left(-\frac{q}{p}\right)^{j} = \binom{n}{k}^{-1} (-1)^k \sum_{j=0}^{k} \frac{1}{j!(k-j)!}\left(\frac{q}{p}\right)^{j} \neq 0,$$

and so $\deg P_k = k$. □

This prompts the following definition.

Definition 3.2. Let B be an admissible matrix having the property that the eigenvalue $\lambda_S(B)$ corresponding to $\chi_{S,p}$ depends only on $|S|$. The function $\Lambda\colon \{0,\dots,n\} \to \mathbb{R}$ defined by $\Lambda(s) = \lambda_{\{1,\dots,s\}}(B)$ is called an *admissible spectrum.* ○

Using this terminology, we can rephrase our goal: find an admissible spectrum satisfying $\Lambda(0) = 1$ and $\Lambda(s) \geq -p^t/(1-p^t)$ for $s \in \{1,\dots,n\}$. Lemma 3.11 gives us a large collection of admissible spectra.

Lemma 3.12. *Let P be an arbitrary polynomial of degree smaller than t. Then*

$$\Lambda_P(s) = \left(-\frac{p}{q}\right)^{s} P(s)$$

is an admissible spectrum.

Proof. Any linear combination of $B_0,\dots,B_{t-1}$ is admissible, and so the lemma follows immediately from Lemma 3.11. □

We are now left with the daunting task of engineering a polynomial P for which Λ_P satisfies the two properties listed above. The task gets much simpler if we consider what happens when the family $\mathcal{F}$ is a t-star, say $\mathcal{F} = \{S \subseteq [n] : S \supseteq [t]\}$. In this case $\mu_p(\mathcal{F}) = p^t$, and so the uniqueness part in Hoffman's bound implies that the Fourier transform of $\mathbf{1}_{\mathcal{F}}$ is supported on $\chi_{\varnothing,p}$ and $\chi_{S,p}$ for those S such that $\lambda_{S,p}(B) = \lambda_{\min}(B) = -p^t/(1-p^t)$, where B is the matrix we are trying to construct. Let $\delta(x) = [\![x = 1]\!]$ be a one-bit delta function, and $c(x) = 1$ be the constant one-bit function. For $s \le t$,

$$\begin{aligned}\widehat{\mathbf{1}_{\mathcal{F}}}([s]) &= \langle \delta^{\otimes t} \otimes c^{\otimes n-t}, (\chi_{\{1\}}^{[1]})^s \otimes (\chi_{\varnothing}^{[1]})^{n-s} \rangle_p \\ &= \langle \delta, \chi_{\{1\}}^{[1]} \rangle_p^s \cdot \langle \delta, \chi_{\varnothing}^{[1]} \rangle_p^{t-s} \cdot \langle c, \chi_{\varnothing}^{[1]} \rangle_p^{n-t} = (-\sqrt{pq})^s p^{t-s}.\end{aligned}$$

In particular, $\widehat{\mathbf{1}_{\mathcal{F}}}([s]) \neq 0$, and so $\lambda_{[s],p}(B) = -p^t/(1-p^t)$. In terms of the spectrum Λ, it must satisfy

$$\Lambda(0) = 1, \quad \Lambda(1) = \cdots = \Lambda(t) = -\frac{p^t}{1-p^t}.$$

If Λ is of the form Λ_P, then this gives us $t+1$ points of the polynomial P, which has t degrees of freedom. So there is at most one polynomial P which satisfies all the equations, and our hope is that for this P (if it exists), $\Lambda_P(s) \ge -p^t/(1-p^t)$ for all s.

Lemma 3.13. *Let P be the unique polynomial of degree smaller than t satisfying $P(s) = (-q/p)^s(-p^t/(1-p^t))$ for $s = 1, \ldots, t$. Then $P(0) = 1$ and if $p \le 1/(t+1)$, $(-p/q)^s P(s) \ge -p^t/(1-p^t)$ for all integers $s \ge 0$.*

If furthermore $p < 1/(t+1)$, then $(-p/q)^s P(s) > -p^t/(1-p^t)$ for $s > t$. Moreover, there is a rational function R_t satisfying $R_t(p) > -p^t/(1-p^t)$ for all $p < 1/(t+1)$ such that for $s > t$, $(-p/q)^s P(s) \ge R_t(p)$.

Since the proof of the lemma is technical and unenlightening, we relegate it to Section 3.3.2.

Applying Hoffman's bound, we get a wealth of information on the Fourier coefficients of t-intersecting families.

Lemma 3.14. *Let $\mathcal{F}$ be a t-intersecting family with characteristic function $f = \mathbf{1}_{\mathcal{F}}$, and let $p \le 1/(t+1)$.*

Upper bound: $\mu_p(\mathcal{F}) \le p^t$.

Uniqueness: *If $p < 1/(t+1)$ and $\mu_p(\mathcal{F}) = p^t$ then the p-skewed Fourier expansion of f is supported on the first $t+1$ levels, that is $\hat{f}_p(S) = 0$ for $|S| > t$.*

Stability: *If $p < 1/(t+1)$ and $\mu_p(\mathcal{F}) \geq p^t - \epsilon$ then*

$$\sum_{|S|>t} \hat{f}_p^2(S) = O_{p,t}(\epsilon).$$

Moreover, for each t, the hidden constant depends continuously on p.

Proof. Let B be an admissible matrix whose admissible spectrum is Λ_P, where P is the polynomial given by Lemma 3.13. Since B is admissible and $\mathcal{F}$ is t-intersecting,

$$\langle Bf, f\rangle_p = \sum_{S,T\in\mathcal{F}} \langle B\mathbf{1}_S, \mathbf{1}_T\rangle_p = \sum_{S,T\in\mathcal{F}} \mu_p(T)\mathbf{1}_T' B\mathbf{1}_S = 0.$$

Writing f and Bf in the Fourier basis, we conclude

$$\sum_{S\subseteq[n]} \Lambda_P(|S|)\hat{f}_p^2(S) = 0.$$

Let $\lambda_S = \Lambda_P(|S|)$. Lemma 3.13 shows that $\lambda_\varnothing = 1$ and $\lambda_{\min} = \min_S \lambda_S = -p^t/(1-p^t)$, hence the upper bound in Hoffman's bound shows that $\mu_p(\mathcal{F}) \leq p^t$.

When $p < 1/(t+1)$, Lemma 3.13 shows that $\lambda_S = \lambda_{\min}$ only when $|S| \leq t$, and so the uniqueness part of Hoffman's bound implies uniqueness. Furthermore, the lemma shows that the second smallest eigenvalue is $\lambda_2 = R_t(p)$ for some rational function R_t satisfying $R_t(p) > -p^t/(1-p^t)$ for all $p < 1/(t+1)$. Hoffman's bound shows that if $\mu_p(\mathcal{F}) \geq p^t - \epsilon$ then

$$\sum_{|S|>t} \hat{f}_p^2(S) \leq \frac{-\lambda_{\min}}{\lambda_2 - \lambda_{\min}}\epsilon = \frac{p^t/(1-p^t)}{R_t(p) + p^t/(1-p^t)}\epsilon.$$

Since $R_t(p) > -p^t/(1-p^t)$ for all $p < 1/(t+1)$, the expression in front of ϵ is continuous.

□

In the preceding section, when $\mu_p(\mathcal{F}) = p$ we were able to conclude that $\mathcal{F}$ is a star by considering the Fourier expansion of $f = \mathbf{1}_\mathcal{F}$ and using $f^2 = f$. This time the argument is more complicated, since it is not the case that if $\hat{f}_p$ is supported on the first $t+1$ levels then f depends on at most t coordinates; we gave a counterexample in Section 2.7 for $t = 2$. However, it is the case if $\mathcal{F}$ is monotone and $\mu_p(\mathcal{F}) \leq p^t$.

Lemma 3.15. *Let $\mathcal{F}$ be a monotone family of sets on n points (if $S \in \mathcal{F}$ and $T \supseteq S$ then $T \in \mathcal{F}$), and define $f = \mathbf{1}_{\mathcal{F}}$. Suppose that for some $p \in (0,1)$ and integer t, the p-skewed Fourier expansion of f is supported on the first $t+1$ levels and $\mu_p(\mathcal{F}) \leq p^t$. If $\mu_p(\mathcal{F}) = p^t$ then $\mathcal{F}$ is a t-star, and otherwise $\mathcal{F} = \varnothing$.*

Proof. The proof is by induction on $n+t$. The claim trivially holds when $t = 0$, so suppose $t \geq 1$. Define $f_0(x) = f(x,0)$ and $f_1(x) = f(x,1)$, and let the corresponding monotone families be $\mathcal{F}_0, \mathcal{F}_1$. If $\mathcal{F}_0 = \mathcal{F}_1$ then f doesn't depend on the last coordinate, and the result follows by applying the inductive hypothesis to $\mathcal{F}_0$.

Otherwise, since $\mu_p(\mathcal{F}_0) \leq \mu_p(\mathcal{F}_1)$ by monotonicity, we deduce that $\mu_p(\mathcal{F}_0) < \mu_p(\mathcal{F}) \leq p^t$. Hence the induction hypothesis applied to $\mathcal{F}_0$ shows that $\mathcal{F}_0 = \varnothing$, and so $\mathcal{F} = \{S \cup \{n\} : S \in \mathcal{F}_1\}$. Therefore

$$\begin{aligned}
\hat{f}_p(S) &= \sum_{T \in \mathcal{F}} \mu_p(T) \chi^{[n]}_{S,p}(T) \\
&= \sum_{T \in \mathcal{F}_1} \mu_p(T \cup \{n\}) \chi^{[n]}_{S,p}(T \cup \{n\}) \\
&= \sum_{T \in \mathcal{F}_1} p\mu_p(T) \chi^{[n-1]}_{S \cap [n-1],p}(T) \chi^{[1]}_{[\![n \in S]\!]}(\{1\}) = p\chi^{[1]}_{[\![n \in S]\!]}(\{1\}) \widehat{f_1}(S \cap [n-1]).
\end{aligned}$$

In particular, if $\widehat{f_1}(T) \neq 0$ then $\hat{f}_p(T \cup \{n\}) \neq 0$. This implies that the p-skewed Fourier expansion of f_1 is supported on the first t levels. Also, $\mu_p(\mathcal{F}) = p\mu_p(\mathcal{F}_1)$ implies that $\mu_p(\mathcal{F}_1) \leq p^{t-1}$, with equality only if $\mu_p(\mathcal{F}) = p^t$. The result now follows by applying the inductive hypothesis to $\mathcal{F}_1$ and $t-1$. $\square$

This lemma enables us to deduce the structure of t-intersecting families $\mathcal{F}$ satisfying $\mu_p(\mathcal{F}) = p^t$, as well as stability. For stability, Kindler and Safra's theorem (Theorem 2.23 on page 24) replaces Friedgut–Kalai–Naor, and the argument also gets more complicated.

Theorem 3.16. *Let $\mathcal{F}$ be a t-intersecting family on n points, and $p \leq 1/(t+1)$.*

Upper bound: $\mu_p(\mathcal{F}) \leq p^t$.

Uniqueness: *If $p < 1/(t+1)$, then $\mu_p(\mathcal{F}) = p^t$ if and only if $\mathcal{F}$ is a t-star.*

Stability: *If $p < 1/(t+1)$ and $\mu_p(\mathcal{F}) \geq p^t - \epsilon$ then $\mu_p(\mathcal{F} \Delta \mathcal{G}) = O_{p,t}(\epsilon)$ for some t-star $\mathcal{G}$. Furthermore, for each t, the hidden constant is a continuous function of p.*

Proof. Our starting point is Lemma 3.14. The upper bound is already stated in the lemma. For uniqueness, suppose $\mathcal{F}$ is a t-intersecting family satisfying $\mu_p(\mathcal{F}) = p^t$. The monotone closure of $\mathcal{F}$ (defined as $\{S \subseteq [n] : S \supseteq T \text{ for some } T \in \mathcal{F}\}$) is also t-intersecting and its μ_p-measure is at least $\mu_p(\mathcal{F})$. In view of the upper bound, we can conclude that $\mathcal{F}$ is monotone. Hence Lemma 3.15 applies and shows that $\mathcal{F}$ is a t-star.

For stability, suppose first that $\mathcal{F}$ is monotone. Theorem 2.23 implies that $\mathcal{F}$ is $D_{p,t}\epsilon$-close to some family $\mathcal{G}$ on $M_{p,t}$ coordinates, where the constant $D_{p,t}$ is the product of $C_{p,t}$ and the hidden constant in Lemma 3.14. Moreover, for each t, $D_{p,t}$ depends continuously on p, since both its factors do. We show that if ϵ is small enough, then $\mathcal{G}$ must be a t-star. We assume without loss of generality that $\mathcal{G}$ depends only on the coordinates $[M_{p,t}]$.

Suppose first that $\mathcal{G}$ is not supported on the first $t+1$ levels. Let $g = \mathbf{1}_{\mathcal{G}}$, define

$$f^{>t} = \sum_{|S|>t} \hat{f}_p(S)\chi_{S,p},$$

and define $g^{>t}$ similarly. Orthonormality of the Fourier characters implies that $\|f - g\|_p^2 \geq \|f^{>t} - g^{>t}\|_p^2 \geq (\|f^{>t}\|_p - \|g^{>t}\|_p)^2 \geq (\sqrt{\epsilon} - \|g^{>t}\|_p|)^2$, and so $\|g^{>t}\|_p \leq (\sqrt{D_{p,t}}+1)\sqrt{\epsilon}$. Now, there are only finitely many possible families $\mathcal{G}$ (up to choice of coordinates), and so if ϵ is small enough, $\|g^{>t}\|_p > (\sqrt{D_{p,t}}+1)\sqrt{\epsilon}$ for all of them, showing that $\mathcal{G}$ has to be supported on the first $t+1$ levels.

Suppose next that $\mathcal{G}$ is not monotone, say for some $S \subset T \subseteq [M_{p,t}]$, $S \in \mathcal{G}$ and $T \notin \mathcal{G}$. For each $X \subseteq [n] \setminus [M_{p,t}]$, either $S \cup X \notin \mathcal{F}$, in which case $S \cup X \in \mathcal{F}\Delta\mathcal{G}$, or $S \cup X \in \mathcal{F}$, in which case $T \cup X \in \mathcal{F}\Delta\mathcal{G}$. Since $\mu_p^{[M_{p,t}]}(T) \geq \mu_p^{[M_{p,t}]} \geq p^{M_{p,t}}$, we deduce that $\mu_p(\mathcal{F}\Delta\mathcal{G}) \geq p^{M_{p,t}}$. If ϵ is small enough, this cannot be possible.

Consider now a monotone family $\mathcal{G}$ which is supported on the first $t+1$ levels. If $\mu_p(\mathcal{G}) \neq p^t$ then

$$|\mu_p(\mathcal{G}) - p^t| \leq |\mu_p(\mathcal{G}) - \mu_p(\mathcal{F})| + \epsilon \leq \|f - g\|_p + \epsilon \leq (D_{p,t}+1)\epsilon.$$

As before, if ϵ is small enough then this cannot happen. We conclude that $\mathcal{G}$ must be supported on the first $t+1$ levels and $\mu_p(\mathcal{G}) = p^t$. Lemma 3.15 implies that $\mathcal{G}$ must be a t-star.

Summarizing, there is a function $f_{p,t}$ such that if $\epsilon < f_{p,t}$, $\mathcal{G}$ must be a t-star. Moreover, since for each t, $f_{p,t}$ is obtained by considering finitely many families, $f_{p,t}$ is continuous in p.

Let $K_{p,t} = D_{p,t} + f_{p,t}^{-1}$, which for every t is a continuous function in p. We claim that there always exists a t-star $\mathcal{G}'$ such that $\mu_p(\mathcal{F}\Delta\mathcal{G}') \le K_{p,t}\epsilon$. If $\epsilon \le f_{p,t}$, then we can take $\mathcal{G}' = \mathcal{G}$. Otherwise, take $\mathcal{G}'$ to be any t-star. Then $\mu_p(\mathcal{F}\Delta\mathcal{G}') \le 1 \le f_{p,t}^{-1}\epsilon \le K_{p,t}\epsilon$.

If $\mathcal{F}$ is not monotone, then consider the monotone closure of $\mathcal{F}$, which we denote $\mathcal{F}^{\uparrow}$. Since $\mathcal{F}^{\uparrow}$ is t-intersecting, $\mu_p(\mathcal{F}^{\uparrow}) \le p$, which implies that $\mu_p(\mathcal{F}\Delta\mathcal{F}^{\uparrow}) \le \epsilon$. Applying the reasoning above to $\mathcal{F}^{\uparrow}$, we obtain a t-star $\mathcal{G}$ satisfying $\mu_p(\mathcal{F}^{\uparrow}\Delta\mathcal{G}) \le K_{p,t}\epsilon$. Therefore $\mu_p(\mathcal{F}\Delta\mathcal{G}) \le (K_{p,t} + 1)\epsilon$. □

3.3.1 More on admissible spectra

Lemma 3.12 gives a wide choice of admissible spectra. In this section we show that this lemma states all the admissible spectra. For the rest of this section, fix n, t and p, and define $q = 1 - p$.

We start by determining all admissible matrices. The first step is to find a nice basis for the space of functions satisfying the eigenvector property, and to this end the following lemma will be useful.

Definition 3.3. Let $\mathcal{B}$ be the vector space of all matrices whose eigenvectors are the p-skewed Fourier basis vectors. For $B \in \mathcal{B}$, let $\lambda_S(B)$ be the eigenvalue corresponding to $\chi_{S,p}$. ○

Lemma 3.17. *Let $J \subseteq [n]$. For every function $f\colon J \to \mathbb{R}$, the linear span $\mathcal{B}_J$ of $\{B_K : K \subseteq J\}$ (where B_K is given by Lemma 3.10) contains a matrix B whose eigenvalues satisfy $\lambda_S(B) = (-(p/q))^{|S|} f(S \cap J)$.*

Proof. We use a dimension argument. The vector space $\mathcal{B}'_J$ of all matrices $B \in \mathcal{B}$ such that $\lambda_S(B) = (-(p/q))^{|S|} f(S \cap J)$ for some f has dimension $2^{|J|}$. Lemma 3.10 shows that

$$\lambda_S(B_K) = \left(-\frac{p}{q}\right)^{|S|} \left(-\frac{q}{p}\right)^{|S\cap K|},$$

showing that $B_K \in \mathcal{B}'_J$ whenever $K \subseteq J$. To complete the proof, we show that the matrices $\{B_K : K \subseteq J\}$ are linearly independent. Suppose that $\sum_{K\subseteq J} c_K B_K = 0$. Then for all $S \subseteq [n]$,

$$0 = \lambda_S\left(\sum_{K\subseteq J} c_K B_K\right) = \left(-\frac{p}{q}\right)^{|S|} \sum_{K\subseteq J} c_K \left(-\frac{q}{p}\right)^{|S\cap K|}.$$

In particular, there is a linear dependency among the vectors x_K of length $2^{|J|}$ defined by

$$x_{K,S} = \left(-\frac{q}{p}\right)^{|S\cap K|}, \quad \text{where } S \subseteq J.$$

Let $y_0 = \begin{pmatrix} 1 & 1 \end{pmatrix}'$ and $y_1 = \begin{pmatrix} 1 & -q/p \end{pmatrix}'$. Then

$$x_K = \bigotimes_{j \in J} y_{[\![j \in K]\!]}.$$

Since the vectors y_0, y_1 are linearly independent, so are the vectors x_K by Lemma 2.21, and we conclude that $\mathcal{B}_J = \mathcal{B}'_J$. □

This lemma immediately implies that the matrices B_J form a basis for all of $\mathcal{B}$.

Corollary 3.18. *The vector space $\mathcal{B}$ is spanned by the matrices $\mathcal{B}_J$ for $J \subseteq [n]$.*

We proceed to determine the subspace of admissible matrices within $\mathcal{B}$.

Lemma 3.19. *The vector space of all admissible matrices is spanned by B_J for $|J| < t$.*

Proof. Lemma 3.10 states that all B_J are admissible. Now suppose that $B \in \mathcal{B}$ is admissible. We know by Corollary 3.18 that B must have the general form

$$B = \sum_{J \subseteq [n]} c_J B_J.$$

We show by reverse induction on $|J|$ that $c_J = 0$ for $|J| \geq t$. Given J such that $|J| \geq t$, B must satisfy $\mathbf{1}'_J B \mathbf{1}_J = 0$. Now

$$\mathbf{1}'^{[n]}_J B_K \mathbf{1}^{[n]}_J = \prod_{i \in K} \mathbf{1}'^{[1]}_{J \cap \{i\}} B_{K,i} \mathbf{1}^{[1]}_{J \cap \{i\}},$$

where $B_{J,i} = A^{[1]}$ if $i \notin J$, and $B_{J,i} = I_2$ otherwise. Looking at the matrix $A^{[1]}$ (given in Lemma 3.2), we deduce that

$$(\mathbf{1}^{[n]}_J)' B_K \mathbf{1}^{[n]}_J = \left(1 - \frac{p}{q}\right)^{([n] \setminus J) \cap ([n] \setminus K)} 0^{J \setminus K} = [\![J \subseteq K]\!] \left(1 - \frac{p}{q}\right)^{([n] \setminus J) \cap ([n] \setminus K)}.$$

Therefore

$$0 = \mathbf{1}'_J B \mathbf{1}_J = \sum_{J \subseteq K} c_K \left(1 - \frac{p}{q}\right)^{([n] \setminus J) \cap ([n] \setminus K)}.$$

By the induction hypothesis, $c_K = 0$ for all $K \supsetneq J$, and so we can conclude that $c_J = 0$ as well. □

Having determined the vector space of all admissible matrices, we can conclude that Lemma 3.12 is optimal.

Lemma 3.20. *Suppose B is an admissible matrix and $\lambda_S(B) = (-(p/q))^{|S|} f(|S|)$ for some function $f\colon \mathbb{N} \to \mathbb{R}$. Then f is a polynomial of degree less than t.*

Proof. Recall that for a permutation $\pi \in S_n$, $\pi(B)$ is defined by $\pi(B)_{S,T} = B_{\pi(S),\pi(T)}$. It is not difficult to see that $\pi(B)$ is also admissible and $\lambda_{\pi(S)}(\pi(B)) = \lambda_S(B)$. This implies that $B = (1/n!) \sum_\pi \pi(B)$, since both sides have the same eigenvalues. In view of Lemma 3.19, we see that B must be in the span of the matrices B_k defined in Lemma 3.11, and the lemma follows. ☐

3.3.2 Proof of Lemma 3.13

In order to complete the proof of Theorem 3.16, it remains to prove the technical Lemma 3.13. The general plan is to explicitly construct the polynomial P using Lagrange interpolation.

Lemma 3.21. *Let $(x_1, y_1), \dots, (x_m, y_m)$ be m pairs of real points such that $x_i \neq x_j$ for $i \neq j$. The unique polynomial Q of degree smaller than m satisfying $Q(x_i) = y_i$ for all $i \in [m]$ is*

$$Q(s) = \sum_{i=1}^m y_i \prod_{j \neq i} \frac{s - x_j}{x_i - x_j}.$$

Proof. To see that $Q(x_k) = y_k$,

$$Q(x_k) = \sum_{i=1}^m y_i \prod_{j \neq i} \frac{x_k - x_j}{x_i - x_j} = y_k \prod_{j \neq k} \frac{x_k - x_j}{x_k - x_j} = y_k,$$

since the terms $i \neq k$ all vanish. If there were another polynomial Q' of degree smaller than m satisfying $Q'(x_i) = y_i$ for all $i \in [m]$ then $Q - Q'$ would have been a polynomial of degree smaller than m having m roots, namely $x_1, \dots, x_m$, which is impossible. ☐

In our case, the polynomial P has the following form.

Lemma 3.22. *There is a unique polynomial P of degree smaller than t satisfying $(-p/q)^s P(s) = -p^t/(1-p^t)$ for $s \in [t]$, which is given by the following formula for $s \notin [t]$:*

$$P(s) = (-1)^{t+1} \frac{p^t}{1-p^t} \sum_{r=1}^t \left(\frac{q}{p}\right)^r \frac{(s-1)\cdots(s-t)}{(s-r)(r-1)!(t-r)!}. \tag{3.10}$$

Proof. According to Lemma 3.21, the unique polynomial P is given by

$$\begin{aligned} P(s) &= \frac{-p^t}{1-p^t}\sum_{r=1}^{t}\left(-\frac{q}{p}\right)^r \prod_{\substack{j=1\\ j\neq r}}^{t}\frac{s-j}{r-j} \\ &= \frac{-p^t}{1-p^t}\sum_{r=1}^{t}\left(-\frac{q}{p}\right)^r \frac{(s-1)\cdots(s-t)/(s-r)}{(r-1)\cdots 1\cdot(-1)\cdots(-(t-r))} \\ &= \frac{-p^t}{1-p^t}\sum_{r=1}^{t}\left(-\frac{q}{p}\right)^r \frac{(s-1)\cdots(s-t)/(s-r)}{(r-1)!(-1)^{t-r}(t-r)!} \\ &= (-1)^{t+1}\frac{p^t}{1-p^t}\sum_{r=1}^{t}\left(\frac{q}{p}\right)^r \frac{(s-1)\cdots(s-t)}{(s-r)(r-1)!(t-r)!}. \end{aligned}$$

□

Using the explicit form of P and some calculations, we can derive all the properties of P that are needed for Lemma 3.13. Part (d) of the ensuing lemma simplifies the original argument in [40].

Lemma 3.23. *Let P be the polynomial given by* (3.10)*, and suppose $p \in (0,1)$ and $q = 1-p$.*

(a) $(-1)^{t+1}P(s) > 0$ *for* $s > t$.

(b) $P(0) = 1$.

(c) $(-p/q)^{t+2}P(t+2) > -p^t/(1-p^t)$ *if and only if* $p < 1/(t+1)$.

(d) When $p < 1/(t+1)$, the sequence $y_s = |(-p/q)^s P(s)|$ is decreasing for $s \geq t+1$, that is, $y_{s+1} < y_s$ for $s \geq t+1$.

Proof. Item (a) is immediate from formula (3.10).

For item (b), calculation gives

$$\begin{aligned} P(0) &= (-1)^{t+1}\frac{p^t}{1-p^t}\sum_{r=1}^{t}\left(\frac{q}{p}\right)^r \frac{(-1)\cdots(-t)}{(-r)(r-1)!(t-r)!} \\ &= \frac{p^t}{1-p^t}\sum_{r=1}^{t}\left(\frac{q}{p}\right)^r \frac{t!}{r!(t-r)!} \\ &= \frac{1}{1-p^t}\sum_{r=1}^{t}\binom{t}{r}p^{t-r}q^r \\ &= \frac{1}{1-p^t}\left((p+q)^t - \binom{t}{0}p^{t-0}q^0\right) = 1, \end{aligned}$$

using the binomial theorem in the penultimate step.

For item (c), the first inequality is equivalent to $D < 1$, where

$$\begin{aligned} D &\triangleq \left(\frac{-p^t}{1-p^t}\right)^{-1} \left(-\frac{p}{q}\right)^{t+2} P(t+2) \\ &= \sum_{r=1}^{t} \left(\frac{p}{q}\right)^{t+2-r} \frac{(t+1)\cdots 2}{(t+2-r)(r-1)!(t-r)!} \\ &= \sum_{r=1}^{t} \left(\frac{p}{q}\right)^{t+2-r} \frac{(t+1)!}{(t+2-r)(r-1)!(t-r)!}. \end{aligned}$$

To continue the calculation, let $x = p/q$:

$$\begin{aligned} D &= \sum_{r=1}^{t} x^{t+2-r} \frac{(t+1)!}{(t+2-r)(r-1)!(t-r)!} \\ &= \sum_{r=1}^{t} x^{t+2-r} \frac{(t+1)!}{(r-1)!(t-r)!} \left(\frac{1}{t+1-r} - \frac{1}{(t+2-r)(t+1-r)}\right) \\ &= (t+1) \sum_{r=1}^{t} x^{t+2-r} \binom{t}{r-1} - \sum_{r=1}^{t} x^{t+2-r} \binom{t+1}{r-1} \\ &= (t+1)x \sum_{r=0}^{t-1} x^{t-r} \binom{t}{r} - \sum_{r=0}^{t-1} x^{t+1-r} \binom{t+1}{r} \\ &= (t+1)x((1+x)^t - 1) - ((1+x)^{t+1} - 1 - (t+1)x) \\ &= (t+1)x(1+x)^t - (1+x)^{t+1} + 1 = 1 + (1+x)^t(tx-1). \end{aligned}$$

Therefore $D < 1$ if and only if $p/q < 1/t$, which is equivalent to $p < 1/(t+1)$.

For item (d), write

$$y_s = \frac{p^t}{1-p^t} \sum_{r=1}^{t} \left(\frac{p}{q}\right)^{s-r} \frac{(s-1)\cdots(s-t)}{(s-r)(r-1)!(t-r)!}.$$

Let $Y_{s,r}$ be the rth term in the sum. When $p < 1/(t+1)$ and $s \geq t+1$, we have

$$\begin{aligned} \frac{Y_{s+1,r}}{Y_{s,r}} &= \frac{p}{q} \cdot \frac{s}{s-t} \cdot \frac{s-r}{s-r+1} < \frac{1}{t} \cdot \frac{s}{s-t} \cdot \frac{s-1}{s} \\ &= \frac{s-1}{t(s-t)} = 1 + \frac{(t-1)(t+1-s)}{t(s-t)} \leq 1. \end{aligned}$$

Hence $Y_{s+1,r} < Y_{s,r}$, and we conclude that $y_{s+1} < y_s$. □

Lemma 3.13 easily follows.

Lemma 3.13. *Let P be the unique polynomial of degree smaller than t satisfying $P(s) = (-q/p)^s(-p^t/(1-p^t))$ for $s = 1, \ldots, t$. Then $P(0) = 1$ and if $p \leq 1/(t+1)$, $(-p/q)^s P(s) \geq -p^t/(1-p^t)$ for all integers $s \geq 0$.*

If furthermore $p < 1/(t+1)$, then $(-p/q)^s P(s) > -p^t/(1-p^t)$ for $s > t$. Moreover, there is a rational function R_t satisfying $R_t(p) > -p^t/(1-p^t)$ for all $p < 1/(t+1)$ such that for $s > t$, $(-p/q)^s P(s) \geq R_t(p)$.

Proof. Let $\Lambda(s) = (-p/q)^s P(s)$. Lemma 3.23 already shows that $P(0) = 1$. If $p < 1/(t+1)$ then the lemma shows that $\Lambda(t+1) > 0$ and $|\Lambda(s)| < p^t/(1-p^t)$ for integers $s \geq t+2$. Since $|\Lambda(s)|$ decreases for $s \geq t+1$, the minimum of $\Lambda(s)$ for $s > t$ is attained at $s = t+2$. Lemma 3.22 shows that $\Lambda(t+2)$ is a rational function of p.

When $p = 1/(t+1)$, we get the required results by continuity. □

3.4 Cross-intersecting families

Recall that a family is t-intersecting if any two sets intersect in at least t points. If instead we take each of the two sets from a different family, we arrive at the concept of cross-t-intersecting families, which we explore in this section. Apart from the generalization of Hoffman's bound, all the material in this section is new. Similar results have been obtained by Tokushige [78] in the classical setting (see Section 3.5), who generalized a result of Wilson [79] to the cross-intersecting setting. His proof also uses the generalized Hoffman's bound, and we expand more on the connection in Section 3.5.3.

Definition 3.4. Let $\mathcal{F}, \mathcal{G}$ be two families of sets on n points. We say that $\mathcal{F}, \mathcal{G}$ are *cross-intersecting* if every $S \in \mathcal{F}$ and $T \in \mathcal{G}$ intersect. We say that $\mathcal{F}, \mathcal{G}$ are *cross-t-intersecting* if every $S \in \mathcal{F}$ and $T \in \mathcal{G}$ intersect in at least t points. ○

What can we say about the μ_p-measure of two families if they are cross-intersecting? The following extension of Katona's circle argument gives one possible answer.

Lemma 3.24. *Suppose $\mathcal{F}, \mathcal{G}$ are two cross-intersecting families of sets on n points. Then for all $p \leq 1/2$, $\mu_p(\mathcal{F})\mu_p(\mathcal{G}) \leq p^2$.*

Proof. Let $x_1, \ldots, x_n$ be n random points on the unit-circumference circle, and let F, G be two independent random intervals of length p on the circle. Let $S(F), S(G)$ denote the set of points falling in F, G, respectively. We have

$$\mu_p(\mathcal{F})\mu_p(\mathcal{G}) = \Pr[S(F) \in \mathcal{F} \text{ and } S(G) \in \mathcal{G}].$$

Fix some position of $x_1, \ldots, x_n$. Let $I_F = \{F : S(F) \in \mathcal{F}\}$ and $I_G = \{G : S(G) \in \mathcal{G}\}$. We show that $\mu(I_F \times I_G) \le p^2$, where μ is the Haar measure on the circle (we represent each interval by its starting point). If I_F or I_G is empty then $\mu(I_F \times I_G) = 0$, so assume both are non-empty. Take an arbitrary $F \in I_F$. Every interval in I_G must intersect F, and so $\mu(I_G) \le p$. Similarly, $\mu(I_F) \le p$. Therefore $\mu(I_F \times I_G) \le p^2$. □

We comment that in the literature other bounds are considered, though only in the classical regime of uniform families.

It turns out that we can extend the entire edifice considered in this chapter to cross-intersecting families. We start by generalizing Hoffman's bound, following Ellis, Friedgut and Pilpel [28] (a similar bound appears in Alon et al. [5]). For simplicity, we only consider the upper bound and uniqueness.

Lemma 3.25 (Hoffman's bound, cross-intersecting version). *Let $\lambda_S, x_S, y_S \in \mathbb{R}$ for S ranging over some arbitrary index set containing $\varnothing$, with $\lambda_\varnothing > 0$. Suppose that the following two equations hold, for some $m_x, m_y \in \mathbb{R}$:*

$$m_x = x_\varnothing = \sum_S x_S^2,$$

$$m_y = y_\varnothing = \sum_S y_S^2,$$

$$\sum_S \lambda_S x_S y_S = 0.$$

Let $\lambda_{\max} = \max_{S \neq \varnothing} |\lambda_S|$.

Upper bound: $\sqrt{m_x m_y} \le m_{\max}$*, where* $m_{\max} = \dfrac{\lambda_{\max}}{\lambda_\varnothing + \lambda_{\max}}$.

Uniqueness: *If* $\sqrt{m_x m_y} = m_{\max}$ *then* $m_x = m_y = m_{\max}$*, and moreover, for* $S \neq \varnothing$*, if* $\lambda_S = -\lambda_{\max}$ *then* $x_S = y_S$*, if* $\lambda_S = \lambda_{\max}$ *then* $x_S = -y_S$*, and* $x_S = y_S = 0$ *otherwise.*

Proof. The arithmetic-geometric mean inequality shows that

$$(1 - m_x)(1 - m_y) = 1 - m_x - m_y + m_x m_y \le 1 - 2\sqrt{m_x m_y} + m_x m_y = (1 - \sqrt{m_x m_y})^2.$$

Using that (in the last step) and the Cauchy–Schwartz inequality, we have

$$\begin{aligned}
\lambda_\varnothing m_x m_y &= \sum_{S\neq\varnothing} (-\lambda_S) x_S y_S \\
&\leq \sum_{S\neq\varnothing} \lambda_{\max} |x_S y_S| \\
&\leq \sqrt{\sum_{S\neq\varnothing} \lambda_{\max} x_S^2}\sqrt{\sum_{S\neq\varnothing} \lambda_{\max} y_S^2} \\
&\leq \sqrt{m_x(1-m_x)}\sqrt{m_y(1-m_y)}\lambda_{\max} \\
&\leq \sqrt{m_x m_y}(1-\sqrt{m_x m_y})\lambda_{\max}.
\end{aligned}$$

Rearrangement yields the desired bound.

If the inequality is tight then the arithmetic-geometric mean inequality is tight, showing that $m_x = m_y = m_{\max}$. Since the Cauchy–Schwartz inequality is tight, for some α, $|x_S| = \alpha|y_S|$ for all $S \neq \varnothing$. Since $\sum_{S\neq\varnothing} x_S^2 = \sum_{S\neq\varnothing} y_S^2$, we conclude that $\alpha = 1$. Furthermore, we must have $-\lambda_S x_S y_S = \lambda_{\max}|x_S y_S|$, implying the conditions on x_S, y_S. □

Using Hoffman's bound, we can provide another proof of Lemma 3.24.

Lemma 3.26. *Let $\mathcal{F},\mathcal{G}$ be cross-intersecting families and $p \leq 1/2$. Then $\mu_p(\mathcal{F})\mu_p(\mathcal{G}) \leq p^2$, with equality only if $\mu_p(\mathcal{F}) = \mu_p(\mathcal{G}) = p$. If furthermore $p < 1/2$, then equality is only possible if $\mathcal{F} = \mathcal{G}$ and $\mathcal{F}$ is a star.*

Proof. Let $f = \mathbf{1}_\mathcal{F}$ and $g = \mathbf{1}_\mathcal{G}$, and define $q = 1-p$. It is easy to generalize the proof of Lemma 3.4 to show that $\langle A^{[n]} f, g\rangle_p = 0$ and so

$$\sum_{S\subseteq[n]} \lambda_S \hat{f}_p(S)\hat{g}_p(S) = 0, \text{ where } \lambda_S = \left(-\frac{p}{q}\right)^{|S|}.$$

Since $\lambda_\varnothing = 1$ and $\lambda_{\max} = p/q$, Hoffman's bound shows that $\mu_p(\mathcal{F})\mu_p(\mathcal{G}) \leq p^2$. Furthermore, equality is only possible if $\mu_p(\mathcal{F}) = \mu_p(\mathcal{G}) = p$.

Now suppose that $p < 1/2$ and $\mu_p(\mathcal{F})\mu_p(\mathcal{G}) = p^2$. Since $|\lambda_S| = \lambda_{\max}$ only when $|S| = 1$, in which case $\lambda_S = -\lambda_{\max}$, Hoffman's bound shows that $\mathcal{F} = \mathcal{G}$. We conclude that $\mathcal{F}$ is an intersecting family of μ_p-measure p, and so a star. □

We proceed to generalize this lemma to cross-t-intersecting families. Let $q = 1 - p$. Recall that the proof of Theorem 3.16 goes by constructing an admissible matrix B which satisfies

$$\lambda_S(B) = \left(-\frac{p}{q}\right)^{|S|} P(|S|),$$

where P is given by Lemma 3.13. We already know that $\lambda_S(B) \geq -p^t/(1-p^t)$ for all S. In fact, our work in Section 3.3.2 implies that $|\lambda_S(B)| \leq p^t/(1-p^t)$ except, perhaps, for sets of size $t+1$. We proceed to determine what happens when $|S| = t+1$.

Lemma 3.27. *Let P be the polynomial given by Lemma 3.13. If $p \leq 1-2^{-1/t}$ then $(p/q)^s|P(s)| \leq p^t/(1-p^t)$ for all $s > 0$. If furthermore $p < 1-2^{-1/t}$ then there is equality only for $s \in \{1,\ldots,t\}$.*

Proof. We show below that for $p < 1 - 2^{-1/t}$, $(p/q)^{t+1}|P(t+1)| < p^t/(1-p^t)$. Lemma 3.23(d,e) shows that when $p < 1/(t+1)$, $(p/q)^s|P(s)| < p^t/(1-p^t)$ for all $s \geq t+2$. When $t = 1$, $1 - 2^{-1/t} = 1/(t+1)$. When $t \geq 2$,

$$\left(\frac{t}{t+1}\right)^{-t} = \left(1+\frac{1}{t}\right)^t \geq \left(1+\frac{1}{2}\right)^2 > 2,$$

implying that $1 - 2^{-1/t} < 1/(t+1)$. This implies the lemma for $p < 1 - 2^{-1/t}$. The result for $p = 1 - 2^{-1/t}$ follows by continuity.

It remains to determine when $(p/q)^{t+1}|P(t+1)| < p^t/(1-p^t)$. Lemma 3.22 shows that

$$\begin{aligned}|P(t+1)| &= \frac{p^t}{1-p^t}\sum_{r=1}^{t}\left(\frac{q}{p}\right)^r \frac{t!}{(r-1)!(t+1-r)!}\\ &= \frac{p^t}{1-p^t}\sum_{r=0}^{t-1}\left(\frac{q}{p}\right)^{r+1}\binom{t}{r}\\ &= \frac{p^t}{1-p^t}\frac{q}{p}\left(\left(1+\frac{q}{p}\right)^t - \left(\frac{q}{p}\right)^t\right)\\ &= \frac{p^t}{1-p^t}\frac{q-q^{t+1}}{p^{t+1}}.\end{aligned}$$

Therefore

$$\left(\frac{p}{q}\right)^{t+1}|P(t+1)| = \frac{p^t}{1-p^t}\left(\frac{1}{q^t}-1\right).$$

The value on the right is smaller than $p^t/(1-p^t)$ when $q^{-t} < 2$, which is the same as $p < 1 - 2^{-1/t}$. ∎

This implies the following theorem.

Theorem 3.28. *Let $\mathcal{F}, \mathcal{G}$ be cross-t-intersecting families and $p \le 1 - 2^{-1/t}$. Then $\mu_p(\mathcal{F})\mu_p(\mathcal{G}) \le p^{2t}$, with equality only if $\mu_p(\mathcal{F}) = \mu_p(\mathcal{G}) = p^t$. If furthermore $p < 1 - 2^{-1/t}$, then equality is only possible if $\mathcal{F} = \mathcal{G}$ and $\mathcal{F}$ is a t-star.*

Proof. Let $f = \mathbf{1}_{\mathcal{F}}$ and $g = \mathbf{1}_{\mathcal{G}}$, and define $q = 1 - p$. Let B be the admissible matrix satisfying $\lambda_S(B) = (-p/q)^{|S|}P(|S|)$, where P is given by Lemma 3.13. Such a matrix exists due to Lemma 3.12. Then

$$\sum_{S \subseteq [n]} \lambda_S(B) \hat{f}_p(S) \hat{g}_p(S) = 0.$$

Note that $\lambda_{\varnothing}(B) = 1$. When $p \le 1 - 2^{-1/t}$, Lemma 3.27 shows that $\lambda_{\max} = p^t/(1 - p^t)$, and Hoffman's bound shows that $\mu_p(\mathcal{F})\mu_p(\mathcal{G}) \le p^{2t}$. Furthermore, equality is only possible if $\mu_p(\mathcal{F}) = \mu_p(\mathcal{G}) = p^{2t}$.

Now suppose that $p < 1 - 2^{-1/t}$ and $\mu_p(\mathcal{F})\mu_p(\mathcal{G}) = p^t$. Lemma 3.27 shows that for $S \ne \varnothing$, $|\lambda_S| = \lambda_{\max}$ only when $|S| \le t$, in which case $\lambda_S = -\lambda_{\max}$. Hoffman's bound then shows that $\mathcal{F} = \mathcal{G}$. We conclude that $\mathcal{F}$ is a t-intersecting family of μ_p-measure p^t, and so a t-star. □

Tokushige [78] obtains a very similar result for the classical setting: if $\mathcal{F}, \mathcal{G}$ are k-uniform cross-t-intersecting families on n points and $k/n < 1 - 2^{-1/t}$ then $|\mathcal{F}||\mathcal{G}| \le \binom{n-t}{k-t}^2$, with equality only if $\mathcal{F} = \mathcal{G}$ is a t-star.

3.5 Classical versus probabilistic setting

The classical Erdős–Ko–Rado theorem is stated in terms of k-uniform intersecting families (families in which each set has size k): for $k \le n/2$, a k-uniform intersecting family on n points contains at most $\binom{n-1}{k-1}$ sets. In this section we explore the connection between Theorem 3.9 and the classical Erdős–Ko–Rado theorem. Section 3.5.1 shows how to convert results in the classical (k-uniform) setting to results in the probabilistic (μ_p) setting, and Section 3.5.2 discusses the other direction. Finally, Section 3.5.3 sketches how to translate Friedgut's method into the classical setting.

3.5.1 Classical to probabilistic

We already hinted at a connection between the classical setting and the probabilistic setting in Section 2.4, where we used Katona's circle argument to prove the Erdős–Ko–Rado theorem both in the k-uniform setting and in the μ_p-setting. For that proof, the connection between the parameters k, n, p is $p = k/n$. Indeed, following Frankl and Tokushige [38], Tokushige [77] and Dinur and Safra [19], it is easy to derive the μ_p versions of intersection theorems from k-uniform versions.

We start by defining two useful operations, *extension* and *slicing*, and the concept of *equivalence*.

Definition 3.5. Let $\mathcal{F}$ be a family of sets on n points. For $m > n$, its *extension* to m points is

$$\operatorname{Ex}(\mathcal{F}, [m]) = \{S \subseteq [m] : S \cap [n] \in \mathcal{F}\}.$$

Its kth slice consists of all sets of size k:

$$\operatorname{Sl}(\mathcal{F}, k) = \{S \in \mathcal{F} : |S| = k\}.$$

Two families on n points are *equivalent* if they are the same up to a permutation on the points. ○

It is easy to see that the extension operation preserves the μ_p measure for all p.

Definition 3.6. Let $\mathfrak{P} = (\mathfrak{P}_n)_{n=1}^{\infty}$ be a sequence of collections $\mathfrak{P}_n$, where $\mathfrak{P}_n$ is a collection of families of sets on n points. We say that $\mathfrak{P}$ has the *extension property* if whenever $\mathcal{F} \in \mathfrak{P}_n$ and $m \geq n$, $\operatorname{Ex}(\mathcal{F}, [m]) \in \mathfrak{P}_m$. ○

Examples of objects with the extension property are the objects $\mathfrak{P}^t$ corresponding to t-intersecting families: each $\mathfrak{P}^t_n$ is the collection of t-intersecting families on n points.

Definition 3.7. An object $\mathfrak{P}$ is *dominated* in an interval I by a family $\mathcal{G}$ on m points if there exists an integer $N \geq m$ such that for all $n \geq N$, $\mathcal{F} \in \mathfrak{P}_n$ and $k/n \in I$,

$$|\operatorname{Sl}(\mathcal{F}, k)| \leq |\operatorname{Sl}(\operatorname{Ex}(\mathcal{G}, [n]), k)|.$$

The domination is *unique* if equality is possible only if $\mathcal{F}$ and $\operatorname{Ex}(\mathcal{G}, [n])$ are equivalent.

The object $\mathfrak{P}$ is (uniquely) *weakly dominated* in an interval $I = (p_0, p_1)$ by a family $\mathcal{G}$ if it is (uniquely) dominated by $\mathcal{G}$ in the intervals $J = (p_0 + \epsilon, p_1 - \epsilon)$ for all $\epsilon > 0$. ○

As an example, the Erdős–Ko–Rado theorem states that $\mathfrak{P}^1$ is uniquely $[0,1/2)$-dominated. The relevant family $\mathcal{G}$ is the star on one point, $\{\{1\}\}$. The extension $\mathrm{Ex}(\mathcal{G},[n])$ is then simply a star.

The Ahlswede–Khachatrian theorem, a far-reaching generalization of Erdős–Ko–Rado described in Chapter 5, partitions $[0,1/2)$ into intervals I_r such that $\mathfrak{P}^r$ is weakly dominated in each of them. In the Ahlswede–Khachatrian theorem, the connection between n,k,r is $f_1(r) \le (k-t+1)/n \le f_2(r)$. The asymptotically negligible term $(t+1)/n$ forces us to use the more robust notion of weak domination.

The main result of this section shows that the fact that an object $\mathfrak{P}$ is dominated implies upper bounds on the μ_p measure of families in $\mathfrak{P}$. As an example, the classical version of the Erdős–Ko–Rado theorem directly implies Friedgut's formulation $\mu_p(\mathcal{F}) \le p$. The proof of the upper bound follows Dinur and Safra [19]; a similar proof appears in Tokushige [77]. The result on uniqueness is novel.

Theorem 3.29. *Let $\mathfrak{P}$ have the extension property, and suppose that $\mathfrak{P}$ is weakly dominated in an open interval I by some family $\mathcal{G}$. For every $p \in I$ and $\mathcal{F} \in \mathfrak{P}_n$, $\mu_p(\mathcal{F}) \le \mu_p(\mathcal{G})$.*

If furthermore the domination is unique, then equality is possible only if $\mathcal{F}$ is equivalent to $\mathrm{Ex}(\mathcal{G},[n])$.

Proof. Every $p \in I$ belongs to some open interval strictly contained in I, so we can assume that $\mathfrak{P}$ is dominated in I. Let $I = (p_0, p_1)$, let N be the integer promised by the definition of domination, and let $M \ge N$. Put $\mathcal{F}_M = \mathrm{Ex}(\mathcal{F},[M])$ and $\mathcal{G}_M = \mathrm{Ex}(\mathcal{G},[M])$. The extension property shows that $\mathcal{F}_M \in \mathfrak{P}_M$. Domination implies that

$$\begin{aligned}\mu_p(\mathcal{F}) = \mu_p(\mathcal{F}_M) &= \sum_{k=0}^{M} p^k (1-p)^{M-k} |\mathrm{Sl}(\mathcal{F}_M,k)| \\ &\le \sum_{k\in(p_0M,p_1M)} p^k(1-p)^{M-k}|\mathrm{Sl}(\mathcal{G}_M,k)| + \sum_{k\notin(p_0M,p_1M)} p^k(1-p)^{M-k}\binom{M}{k}. \qquad (3.11)\end{aligned}$$

Chernoff's inequality shows that if $X \sim \mathrm{Bin}(M,p)$ is a binomially-distributed random variable then

$$\Pr[X \notin (p_0 M, p_1 M)] < e^{-2(p_0-p)^2 M} + e^{-2(p_1-p)^2 M} = O(c^M),$$

for some $c < 1$. Hence as $M \longrightarrow \infty$, the second term in (3.11) tends to zero. The first term is clearly bounded by $\mu_p(\mathcal{G}_M) = \mu_p(\mathcal{G})$. By letting $M \longrightarrow \infty$, we deduce that $\mu_p(\mathcal{F}) \le \mu_p(\mathcal{G})$.

Suppose now that the domination is unique and $\mu_p(\mathcal{F}) = \mu_p(\mathcal{G})$. By extending $\mathcal{F}$ or $\mathcal{G}$, we can assume that both $\mathcal{F}$ and $\mathcal{G}$ are families on n points, and so for M and $k \in (p_0 M, p_1 M)$,

$$|\operatorname{Sl}(\mathcal{G}_M, k)| - |\operatorname{Sl}(\mathcal{F}_M, k)| = \sum_{i=0}^{n} (|\operatorname{Sl}(\mathcal{G}, i)| - |\operatorname{Sl}(\mathcal{F}, i)|) \binom{M-n}{k-i}.$$

The binomial coefficients are increasing: $\binom{M-n}{k-i+1} / \binom{M-n}{k-i} = (M-n-k+i)/(k-i+1) = \Theta(M)$. We want to show that $|\operatorname{Sl}(\mathcal{G}, i)| = |\operatorname{Sl}(\mathcal{F}, i)|$ for all i. If not, consider the smallest index such that $|\operatorname{Sl}(\mathcal{G}, i)| - |\operatorname{Sl}(\mathcal{F}, i)| = \alpha \neq 0$. Then

$$|\operatorname{Sl}(\mathcal{G}_M, k)| - |\operatorname{Sl}(\mathcal{F}_M, k)| = \alpha\Theta(M^{n-i}) \binom{M-n}{k-n} = \alpha\Theta(M^{-i}) \binom{M}{k}.$$

Therefore

$$\begin{aligned}
\mu_p(\mathcal{F}) = \mu_p(\mathcal{F}_M) &= \sum_{k=0}^{M} p^k (1-p)^{M-k} |\operatorname{Sl}(\mathcal{F}_M, k)| \\
&\le \sum_{k \in (p_0 M, p_1 M)} p^k (1-p)^{M-k} |\operatorname{Sl}(\mathcal{G}_M, k)| \\
&\quad - \alpha\Theta(M^{-i}) \sum_{k \in (p_0 M, p_1 M)} p^k (1-p)^{M-k} \binom{M}{k} + \sum_{k \notin (p_0 M, p_1 M)} p^k (1-p)^{M-k} \binom{M}{k} \\
&= \mu_p(\mathcal{G}) - O(c^{-M}) - \alpha\Theta(M^{-i})(1 - O(c^{-M})) + O(c^{-M}).
\end{aligned}$$

Here all the asymptotic terms are positive, but the constant α need not be. Since by assumption $\mu_p(\mathcal{F}) = \mu_p(\mathcal{G})$, we conclude that $\alpha\Theta(M^{-i})(1 - O(c^{-M})) = O(c^{-M}) - O(c^{-M})$. This is only possible when $\alpha = 0$, and so $|\operatorname{Sl}(\mathcal{F}, i)| = |\operatorname{Sl}(\mathcal{G}, i)|$ for all $i \in \{0, \ldots, n\}$.

Repeating the same argument for $\mathcal{F}_M, \mathcal{G}_M$, we deduce that $|\operatorname{Sl}(\mathcal{F}_M, i)| = |\operatorname{Sl}(\mathcal{G}_M, i)|$ for all $i \in \{0, \ldots, M\}$. Let $M \ge \max(2n+1, n/p_0)$, and let $k = \lceil pM \rceil \ge n$. For large enough M, $k \in (p_0 M, p_1 M)$ and so unique domination implies that $\operatorname{Sl}(\mathcal{F}_M, k)$ is equivalent to $\operatorname{Sl}(\mathcal{G}_M, k)$, say $\operatorname{Sl}(\mathcal{F}_M, k) = \pi(\operatorname{Sl}(\mathcal{G}_M, k))$ for some $\pi \in S_M$.

For a slice X, call two coordinates i, j *congruent* if X is invariant under permuting i and j. For example, the last $M - n$ coordinates in both $\operatorname{Sl}(\mathcal{F}_M, k)$ and $\operatorname{Sl}(\mathcal{G}_M, k)$ are congruent. Let

F be the congruence class of coordinates congruent to the last $M - n$ coordinates with respect to $\mathrm{Sl}(\mathcal{F}_M, k)$, and define G similarly with respect to $\mathrm{Sl}(\mathcal{G}_M, k)$. Since F and G are the unique congruence classes with at least $n + 1$ coordinates, we must have $F = \pi(G)$. Without loss of generality, we can assume that π fixes the last $M - n$ coordinates, and so $\pi([n]) = [n]$.

Let $l \in \{0, \ldots, n\}$, and let $S_l \subseteq [M] \setminus [n]$ be a set of size $k - l$. Notice that $\{S \subseteq [n] : S \cup S_l \in \mathcal{F}_M\}$ is equal to $\mathrm{Sl}(\mathcal{F}, l)$, and so $\mathrm{Sl}(\mathcal{F}, l) = \pi(\mathrm{Sl}(\mathcal{G}, l))$. We conclude that $\mathcal{F} = \pi(\mathcal{G})$, and so $\mathcal{F}$ is equivalent to $\mathcal{G}$. □

An application of this theorem appears in Section 10.1.

3.5.2 Probabilistic to classical

We now turn to the other direction. Following Friedgut [40], we will show how to deduce a stability version of the Erdős–Ko–Rado theorem from Theorem 3.9. Our techniques apply to *monotone* objects.

Definition 3.8. Let $\mathfrak{P} = (\mathfrak{P}_n)_{n=1}^{\infty}$ be a sequence of collections $\mathfrak{P}_n$, where $\mathfrak{P}_n$ is a collection of families of sets on n points. We say that $\mathfrak{P}$ has the *monotone property* if for every family $\mathcal{F} \in \mathfrak{P}_n$, every $S \in \mathcal{F}$ and every $T \supset S$, $\mathcal{F} \cup \{T\} \in \mathfrak{P}_n$. ○

The objects $\mathfrak{P}^t$ of t-intersecting families are monotone since if $\mathcal{F}$ is t-intersecting and $S \in \mathcal{F}$, then $\mathcal{F} \cup \{T\}$ is t-intersecting for all $T \supset S$. Next, we adapt Definition 3.7 to the μ_p setting.

Definition 3.9. An object $\mathfrak{P}$ is *μ-dominated* in an interval $I = (p_0, p_1)$ by a family $\mathcal{G}$ if for all $p \in I$ and $\mathcal{F} \in \mathfrak{P}_n$, $\mu_p(\mathcal{F}) \leq \mu_p(\mathcal{G})$.

The domination is *unique* if equality is possible only if $\mathcal{F}$ and $\mathrm{Ex}(\mathcal{G}, [n])$ are equivalent. It is *stable* if there is a constant C such that for every $p \in I$, whenever $\mu_p(\mathcal{F}) > (1 - \epsilon)\mu_p(\mathcal{G})$, there exists a family $\mathcal{H}$ equivalent to $\mathrm{Ex}(\mathcal{F}, [n])$ such that $\mu_p(\mathcal{G} \Delta \mathcal{H}) < C\epsilon$.

The domination is *weakly stable* if the domination is stable in the intervals $J = (p_0, p_1 - \epsilon)$ for all $\epsilon > 0$. ○

The notion of weak stability arises for technical reasons: in our applications, the constant C depends continuously on p, and as $p \longrightarrow p_1$, $C \longrightarrow \infty$. Therefore the domination is not stable on the entire interval (p_0, p_1), but it is stable on the smaller intervals $(p_0, p_1 - \epsilon)$ due to continuity.

Our argument makes use of the operation of taking an *upset*.

Definition 3.10. Let $\mathcal{F}$ be a family of sets on n points. Its *upset* is given by

$$\mathcal{F}^{\uparrow} = \{S \subseteq [n] : S \supseteq T \text{ for some } T \in \mathcal{F}\}.$$

A family $\mathcal{F}$ is *monotone* if whenever $T \in \mathcal{F}$ and $S \supseteq T$, also $S \in \mathcal{F}$. ○

As an example, the family $\mathcal{F}^{\uparrow}$ is always monotone. Monotone families are useful because of the following result and its corollary.

Lemma 3.30. *Let $\mathcal{F}$ be a monotone family on n points. For every $k \in \{0, \ldots, n-1\}$,*

$$\frac{|\operatorname{Sl}(\mathcal{F}, k+1)|}{\binom{n}{k+1}} \geq \frac{|\operatorname{Sl}(\mathcal{F}, k)|}{\binom{n}{k}}.$$

Proof. Consider the bipartite graph in which the left side is $\operatorname{Sl}(\mathcal{F}, k)$, the right side is $\operatorname{Sl}(\mathcal{F}, k+1)$, and two sets are connected if the left set is a subset of the right set. The degree of every set on the left is $n - k$, and the degree of every set on the right is at most $k + 1$. Considering the number of edges incident to both sides,

$$(k+1)|\operatorname{Sl}(\mathcal{F}, k+1)| \geq (n-k)|\operatorname{Sl}(\mathcal{F}, k)|,$$

which directly implies the statement of the theorem. □

Corollary 3.31. *Let $\mathcal{F}$ be a monotone family on n points, let $k < n$ and let $p = k/n + \eta < 1$. We have*

$$\mu_p(\mathcal{F}) \geq (1 - e^{-2\eta^2 n}) \frac{|\operatorname{Sl}(\mathcal{F}, k)|}{\binom{n}{k}}.$$

Proof. The lemma implies that

$$\begin{aligned}
\mu_p(\mathcal{F}) &\geq \sum_{l=k}^{n} p^l (1-p)^{n-l} |\operatorname{Sl}(\mathcal{F}, l)| \\
&\geq \sum_{l=k}^{n} p^l (1-p)^{n-l} \binom{n}{l} \frac{|\operatorname{Sl}(\mathcal{F}, k)|}{\binom{n}{k}} \\
&= \frac{|\operatorname{Sl}(\mathcal{F}, k)|}{\binom{n}{k}} \Pr[\operatorname{Bin}(n, p) \geq k].
\end{aligned}$$

Chernoff's inequality shows that

$$\Pr[\mathrm{Bin}(n,p) < k] \le e^{-2n(p-k/n)^2} = e^{-2\eta^2 n},$$

completing the proof. □

We will need the following estimate on binomial coefficients.

Lemma 3.32. *Let m,l be integers satisfying $0 \le l \le m$. For any $n \ge m$ and k in the range $l \le k \le n-(m-l)$ we have*

$$\frac{\binom{n-m}{k-l}}{\binom{n}{k}} = p^l(1-p)^{m-l} \pm \Theta_m\!\left(\frac{1}{n}\right), \quad p = \frac{k}{n}.$$

Proof. We have

$$\begin{aligned}\frac{\binom{n-m}{k-l}}{\binom{n}{k}} &= \frac{k\cdots(k-l+1)\cdot(n-k)\cdots(n-k-(m-l)+1)}{n\cdots(n-m+1)} \\ &= \frac{k}{n}\cdots\frac{k-l+1}{n-l+1}\cdot\frac{n-k}{n-l}\cdots\frac{n-k-(m-l)+1}{n-m+1}.\end{aligned}$$

We want to estimate these fractions using p and $1-p$. For $a,b \le m$ and $n \ge 2m$,

$$\left|\frac{k-a}{n-b} - \frac{k}{n}\right| = \left|\frac{bk-an}{n(n-b)}\right| \le \frac{mn}{n(n-m)} \le \frac{2m}{n}.$$

This shows that for some $\epsilon_1,\ldots,\epsilon_m$ satisfying $|\epsilon_i| \le 2m/n$,

$$\frac{\binom{n-m}{k-l}}{\binom{n}{k}} = \prod_{i=1}^{l}(p+\epsilon_i)\prod_{i=l+1}^{m}(1-p+\epsilon_i).$$

Expanding the right-hand side, each term which includes any ϵ_i is bounded in absolute value by $2m/n$, and there are 2^m-1 of these, and so

$$\left|\frac{\binom{n-m}{k-l}}{\binom{n}{k}} - p^l(1-p)^{m-l}\right| < \frac{2^{m+1}m}{n}.$$ □

The main result of this section shows that if an object $\mathfrak{P}$ is μ-dominated then it is almost (but not quite) dominated, and furthermore stability carries over. This extends Tokushige's result from [77] and Friedgut's result from [40]. Tokushige only proved the upper bound part. Friedgut only proved the stability part, and in a weaker form (bounding $|\mathrm{Sl}(\mathcal{F},k) \setminus \mathrm{Sl}(\mathcal{H},k)|$ instead of $|\mathrm{Sl}(\mathcal{F},k) \Delta \mathrm{Sl}(\mathcal{H},k)|$).

We will employ the shorthand notation $C(n,k) = \binom{n}{k}$ for the duration of the proof.

Theorem 3.33. *Let $\mathfrak{P}$ have the monotone property, and suppose that $\mathfrak{P}$ is weakly μ-dominated in an open interval $I = (p_0, p_1)$ by some family $\mathcal{G}$.*

Upper bound: *For every $\delta, \epsilon > 0$ there is a constant $N_{\delta,\epsilon}$ such that whenever $n \geq N_{\delta,\epsilon}$ and $k/n \in (p_0, p_1 - \delta)$, every family $\mathcal{F} \in \mathfrak{P}_n$ satisfies*

$$|\operatorname{Sl}(\mathcal{F}, k)| < (\mu_{k/n}(\mathcal{G}) + \epsilon)\binom{n}{k}.$$

Stability: *For every $\delta > 0$ there are constants C_δ, N_δ such that for $n \geq N_\delta$, $k/n \in (p_0, p_1 - \delta)$ and every $\epsilon > 0$ and family $\mathcal{F} \in \mathfrak{P}_n$ satisfying*

$$|\operatorname{Sl}(\mathcal{F}, k)| > (\mu_{k/n}(\mathcal{G}) - \epsilon)\binom{n}{k},$$

there exists a family $\mathcal{H}$ which is equivalent to $\operatorname{Ex}(\mathcal{G}, [n])$ and satisfies

$$|\operatorname{Sl}(\mathcal{F}, k) \Delta \operatorname{Sl}(\mathcal{H}, k)| < C_\delta\left(\epsilon + \sqrt{\frac{\log n}{n}}\right)\binom{n}{k}.$$

Proof. We will need the following easy fact: since $p \mapsto \mu_p(\mathcal{G})$ has a continuous derivative, there is some constant B such that $|\mu_{k/n+\eta}(\mathcal{G}) - \mu_{k/n}(\mathcal{G})| \leq B|\eta|$.

Upper bound. Suppose $|\operatorname{Sl}(\mathcal{F}, k)|/C(n,k) \geq \mu_{k/n}(\mathcal{G}) + \epsilon$. Since $\mu_{k/n+\eta}(\mathcal{G}) \leq \mu_{k/n}(\mathcal{G}) + B\eta$, for some $\eta < \delta$ we have $\mu_{k/n+\eta}(\mathcal{G}) \leq \mu_{k/n}(\mathcal{G}) + \epsilon/2$. Let $p = k/n + \eta$, and apply Corollary 3.31 to deduce

$$\mu_p(\mathcal{F}^\uparrow) \geq (1 - e^{-2\eta^2 n})\frac{|\operatorname{Sl}(\mathcal{F}, k)|}{\binom{n}{k}} \geq (1 - e^{-2\eta^2 n})(\mu_{k/n}(\mathcal{G}) + \epsilon).$$

When n is large enough,

$$\mu_p(\mathcal{F}^\uparrow) > \mu_{k/n}(\mathcal{G}) + \epsilon/2 \geq \mu_p(\mathcal{G}),$$

contradicting the fact that $\mathfrak{P}$ is μ-dominated by $\mathcal{G}$.

Stability. Suppose that the domination is stable on $(p_0, p_1 - \delta/2)$ with constant C, and for some $\epsilon > 0$, $|\operatorname{Sl}(\mathcal{F}, k)|/C(n,k) \geq \mu_{k/n}(\mathcal{G}) - \epsilon$. We let m denote the size of the support of $\mathcal{G}$.

Let $\eta = \sqrt{\log n/n}$, and $p = k/n + \eta$. When n is large enough, $p < p_1 - \delta/2$. Corollary 3.31 shows that

$$\mu_p(\mathcal{F}^\uparrow) \geq (1 - e^{-2\eta^2 n})\frac{|\mathrm{Sl}(\mathcal{F},k)|}{\binom{n}{k}} \geq (1 - e^{-2\eta^2 n})(\mu_{k/n}(\mathcal{G}) - \epsilon) \geq$$
$$(1 - e^{-2\eta^2 n})(\mu_p(\mathcal{G}) - B\eta - \epsilon) \geq \mu_p(\mathcal{G}) - B\eta - \epsilon - e^{-2\eta^2 n} \geq \mu_p(\mathcal{G}) - \epsilon - 2B\eta,$$

for large enough n, since $e^{-2\eta^2 n} = 1/n^2 = o(\sqrt{\log n/n})$. Let $\epsilon' = \epsilon + 2B\eta$. Stability on $(p_0, p_1 - \delta/2)$ implies that there is a family $\mathcal{H}$ which is equivalent to $\mathrm{Ex}(\mathcal{G},[n])$ and satisfies $\mu_p(\mathcal{H}\Delta\mathcal{F}^\uparrow) \leq C\epsilon'$.

We proceed to show that $|\mathrm{Sl}(\mathcal{F} \setminus \mathcal{H}, k)|$ is small. Without loss of generality, we can assume that the original coordinates of $\mathcal{G}$ form the first m coordinates of $\mathcal{H}$. For each $A \subseteq [m]$, let $\mathcal{H}_A = \{S \subseteq [n] \setminus [m] : S \cup A \in \mathcal{H}\}$, and define $\mathcal{F}_A$ similarly. We have

$$\mu_p(\mathcal{F} \setminus \mathcal{H}) = \sum_{A \subseteq [m]} p^{|A|}(1-p)^{m-|A|}\mu_p(\mathcal{F}_A \setminus \mathcal{H}_A).$$

Corollary 3.31, together with $p - (k - |A|)/n \geq \eta$, shows that

$$\mu_p(\mathcal{F}_A) \geq (1 - e^{-2\eta^2 n})\frac{|\mathrm{Sl}(\mathcal{F}_A, k - |A|)|}{\binom{n-m}{k-|A|}} = (1 - 1/n^2)\frac{|\mathrm{Sl}(\mathcal{F}_A, k - |A|)|}{\binom{n-m}{k-|A|}}.$$

When n is large enough, Lemma 3.32 shows that

$$(k/n)^{|A|}(1 - k/n)^{m-|A|}\mu_p(\mathcal{F}_A) \geq (1 - 2/n^2)\frac{|\mathrm{Sl}(\mathcal{F}_A, k - |A|)|}{\binom{n}{k}}.$$

As $n \longrightarrow \infty$, $p \longrightarrow k/n$, and so for large enough n,

$$p^{|A|}(1-p)^{m-|A|}\mu_p(\mathcal{F}_A) \geq (1 - 3/n^2)\frac{|\mathrm{Sl}(\mathcal{F}_A, k - |A|)|}{\binom{n}{k}}.$$

For each $A \subseteq [m]$, either $\mathcal{H}_A = \varnothing$ or $\mathcal{H}_A = 2^{[n]\setminus[m]}$. Therefore

$$\begin{aligned}
\frac{|\mathrm{Sl}(\mathcal{F} \setminus \mathcal{H}, k)|}{\binom{n}{k}} &= \sum_{\substack{A \subseteq [m]:\\ \mathcal{H}_A = \varnothing}} \frac{|\mathrm{Sl}(\mathcal{F}_A, k - |A|)|}{\binom{n}{k}} \\
&\leq (1 - 3/n^2)^{-1} \sum_{\substack{A \subseteq [m]:\\ \mathcal{H}_A = \varnothing}} p^{|A|}(1-p)^{m-|A|}\mu_p(\mathcal{F}_A) \\
&= (1 - 3/n^2)^{-1}\mu_p(\mathcal{F} \setminus \mathcal{H}) \leq 2C\epsilon',
\end{aligned}$$

for large enough n.

We would like to deduce that $|\mathrm{Sl}(\mathcal{H}\setminus\mathcal{F},k)|$ is also small. To that end, we show that $\mathrm{Sl}(\mathcal{F},k)$ and $\mathrm{Sl}(\mathcal{H},k)$ have roughly the same size. Corollary 3.31 shows that

$$\mu_{k/n}(\mathcal{G}) + B\eta \geq \mu_p(\mathcal{G}) \geq \mu_p(\mathcal{F}^{\uparrow}) \geq (1-e^{-2\eta^2 n})\frac{|\mathrm{Sl}(\mathcal{F},k)|}{\binom{n}{k}}.$$

Therefore for large enough n,

$$\epsilon < \mu_{k/n}(\mathcal{G}) - \frac{|\mathrm{Sl}(\mathcal{F},k)|}{\binom{n}{k}} < 2B\eta.$$

For the family $\mathcal{H}$, using Lemma 3.32 we get

$$\frac{|\mathrm{Sl}(\mathcal{H},k)|}{\binom{n}{k}} = \sum_{l=0}^{m} \frac{|\mathrm{Sl}(\mathcal{G},l)|\binom{n-m}{k-l}}{\binom{n}{k}} = \sum_{l=0}^{m}\left(\left(\frac{k}{n}\right)^{l}\left(1-\frac{k}{n}\right)^{m-l} \pm \Theta\left(\frac{1}{n}\right)\right)|\mathrm{Sl}(\mathcal{G},l)| = \mu_{k/n}(\mathcal{G}) \pm \Theta\left(\frac{1}{n}\right).$$

Therefore for large enough n,

$$\left|\frac{|\mathrm{Sl}(\mathcal{F},k)| - |\mathrm{Sl}(\mathcal{H},k)|}{\binom{n}{k}}\right| < \epsilon'.$$

Therefore

$$\frac{|\mathrm{Sl}(\mathcal{F}\Delta\mathcal{H},k)|}{\binom{n}{k}} \leq 2\frac{|\mathrm{Sl}(\mathcal{F}\setminus\mathcal{H},k)|}{\binom{n}{k}} + \epsilon' \leq (4C+1)\epsilon'.$$ □

As a consequence, Theorem 3.16 has the following implication in the classical setting. Recall that in a k-uniform family, every set has size exactly k.

Theorem 3.34. *For every $t \geq 1$ and $\delta > 0$ there are constants $C_{t,\delta}, N_{t,\delta}$ such that for any $k \in (0, n/(t+1) - \delta n)$ and any k-uniform t-intersecting family $\mathcal{F}$ on $n \geq N_{t,\delta}$ points satisfying*

$$|\mathcal{F}| > \binom{n-t}{k-t} - \epsilon\binom{n}{k}$$

there exists a t-star $\mathcal{H}$ such that

$$|\mathcal{F}\Delta\,\mathrm{Sl}(\mathcal{H},k)| < C_{t,\delta}\left(\epsilon + \sqrt{\frac{\log n}{n}}\right)\binom{n}{k}.$$

Proof. Theorem 3.16 shows that for every t-intersecting family $\mathcal{F}$ and $p < 1/(t+1)$, if $\mu_p(\mathcal{F}) \geq p^t - \epsilon$ then there exists a t-star $\mathcal{G}$ such that $\mu_p(\mathcal{F}\Delta\mathcal{G}) \leq K_{p,t}\epsilon$, where $K_{p,t}$ is continuous as a function of p. In every interval $(0, 1/(t+1) - \delta)$, $K_{p,t}$ is bounded, showing that the monotone object $\mathfrak{P}^t$ of t-intersecting families is weakly μ-dominated in $(0, 1/(t+1) - \delta/2)$ by $\mathcal{G} = \{[t]\}$.

Suppose $\mathcal{F}$ is a k-uniform t-intersecting family satisfying

$$|\mathcal{F}| > \binom{n-t}{k-t} - \epsilon\binom{n}{k}.$$

Lemma 3.32 shows that for some constant K_t depending only on t,

$$|\mathcal{F}| > \binom{n}{k}\left(\left(\frac{k}{n}\right)^3 - \frac{K_t}{n} - \epsilon\right).$$

Let $\epsilon' = \epsilon + K_t/n$. Theorem 3.33 shows that if n is large enough, there exists a t-star $\mathcal{H}$ such that

$$|\mathcal{F} \Delta \operatorname{Sl}(\mathcal{H}, k)| < C'_{t,\delta}\left(\epsilon' + \sqrt{\frac{\log n}{n}}\right)\binom{n}{k} \leq 2C'_{t,\delta}\left(\epsilon + \sqrt{\frac{\log n}{n}}\right)\binom{n}{k},$$

when n is large enough. Here $C'_{t,\delta}$ is the constant given by the theorem. The proof is complete by taking $C_{t,\delta} = 2C'_{t,\delta}$. □

3.5.3 Lovász's method

In his groundbreaking paper on the Shannon capacity of graphs [66], Lovász presents, as an application of his methods, a spectral proof of the classical Erdős–Ko–Rado theorem. He starts by noticing that a k-uniform intersecting family on n points is the same as an independent set in the Kneser graph $\operatorname{Kn}(n;k)$.

Definition 3.11. Let $n, k \geq 1$ be integers satisfying $n \geq 2k$. The *Kneser graph* $\operatorname{Kn}(n;k)$ has the vertex set $\binom{[n]}{k}$ consisting of all subsets of $[n]$ of size k, and two vertices A, B are connected if the sets A, B are disjoint. (When $2k > n$, the Kneser graph has no edges.) ○

Fact 3.35 ([37]). *The eigenvalues of the Kneser graph* $\operatorname{Kn}(n;k)$ *(that is, of its adjacency matrix) are* $\lambda_i = (-1)^i\binom{n-k-i}{k-i} = (-1)^i\binom{n-k-i}{n-2k}$ *for* $i \in \{0, \dots, k\}$. *The eigenvalue* λ_i *has multiplicity* $\binom{n}{i} - \binom{n}{i-1}$, *where* $\binom{n}{-1} = 0$. *Let the corresponding eigenspace be* V_i. *The subspace* $U_i = V_0 \oplus \cdots \oplus V_i$ *is spanned by the (linearly independent) characteristic vectors of all* i*-stars.*

We can define a Fourier transform with respect to the Kneser graph. In contrast to the usual Fourier transform, here each Fourier coefficient is a vector in one of the sets V_i. For a similar situation arising when considering functions on S_n, consult Chapter 6.

Definition 3.12. Let $f\colon\binom{[n]}{k} \to \mathbb{R}$ be a Boolean function. For $i \in \{0,\dots,k\}$, $\hat{f}(i)$ is the projection of f to V_i. ○

Lemma 3.36. *Let $f\colon\binom{[n]}{k} \to \mathbb{R}$ be a Boolean function. We have $\|\hat{f}(0)\|^2 = \mu(f)^2$ and*

$$\sum_{i=0}^{k}\|\hat{f}(i)\|^2 = \mu(f),$$

where $\mu(f) = \mathbb{E}_X f(X)$ and $\|f\|^2 = \mathbb{E}_X f(X)^2$.

More generally, for any two arbitrary functions $f,g\colon\binom{[n]}{k} \to \mathbb{R}$,

$$\langle f,g\rangle = \sum_{i=0}^{k}\langle \hat{f}(i),\hat{g}(i)\rangle,$$

where $\langle f,g\rangle = \mathbb{E}_X f(X)g(X)$.

Proof. Let $\mathbf{1}$ be the constant 1 vector. Since the V_i are eigenspaces corresponding to different eigenvalues and the adjacency matrix is symmetric, the V_i are orthogonal to each other. Since $\|\mathbf{1}\| = 1$, it is not hard to check that $\hat{f}(0) = \mu(f)\mathbf{1}$, and so $\|\hat{f}(0)\|^2 = \mu(f)^2$. The other identities follow directly from the orthogonality of the V_i together with $\|f\|^2 = \mu(f)$ for Boolean f. □

As a consequence, we can apply Hoffman's bound.

Theorem 3.37. *Let $n,k \geq 1$ be integers satisfying $n \geq 2k$. Every k-uniform intersecting family on n points contains at most $\binom{n-1}{k-1}$ sets. If furthermore $n > 2k$, then this bound is achieved only by stars.*

Furthermore, if $\mathcal{F},\mathcal{G}$ are k-uniform cross-intersecting families on n points then $|\mathcal{F}||\mathcal{G}| \leq \binom{n-1}{k-1}^2$. When $n > 2k$, this bound is achieved only if $\mathcal{F} = \mathcal{G}$ is a star.

Proof. Let f be the characteristic vector of a k-uniform intersecting family $\mathcal{F}$. Since f is intersecting, $f' \operatorname{Kn}(n;k) f = 0$, and Lemma 3.36 implies that

$$\sum_{i=0}^{k}\lambda_i\|\hat{f}(i)\|^2 = 0.$$

We identify $\lambda_\varnothing$ with $\lambda_0 = \binom{n-k}{k}$ in Hoffman's bound. Fact 3.35 implies that $\lambda_{\min} = \lambda_1 = -\binom{n-k-1}{k-1}$. Hoffman's bound implies that the measure of any intersecting family is at most

$$\frac{-\lambda_{\min}}{\lambda_\varnothing - \lambda_{\min}} = \frac{\binom{n-k-1}{k-1}}{\binom{n-k}{k} + \binom{n-k-1}{k-1}} = \frac{1}{\frac{n-k}{k} + 1} = \frac{k}{n}.$$

Hence such a family contains at most this many sets:

$$\frac{k}{n}\binom{n}{k} = \binom{n-1}{k-1}.$$

Since $\lambda_i = (-1)^i\binom{n-k-i}{n-2k}$, when $n > 2k$ the eigenvalues λ_i decrease in magnitude, and so $\mu(\mathcal{F}) = k/n$ is only possible if $f \in V_0 \oplus V_1 = U_1$. In view of Fact 3.35, f is of the form

$$f(A) = \sum_{i=1}^{n} c_i [\![i \in A]\!] = \sum_{i \in A} c_i.$$

Let A be a set such that $i \in A$ and $j \notin A$. Since $f(A\Delta\{i,j\}) = f(A) + c_j - c_i$ and f is Boolean, we conclude that $c_j - c_i \in \{0, \pm 1\}$. Let $c_{\min}, c_{\max}$ be the smallest and largest values among the c_i. Since f is not constant, $c_{\min} < c_{\max}$, and so $c_{\max} = c_{\min} + 1$, and every c_i is equal to either of these values.

If there are at least k of the c_i which are equal to $c_{\min}$ then we can form a set containing all of them, on which f attains its minimum value 0. This implies that $c_{\min} = 0$ and $c_{\max} = 1$. Since f is Boolean, we conclude that there is exactly one i for which $c_i = c_{\max} = 1$, and so $\mathcal{F}$ is an $\{i\}$-star.

Otherwise, since $2k < n$, there are at least k of the c_i which are equal to $c_{\max}$. Running the same argument as before, we deduce that $c_{\max} = 1/k$, and that there is exactly one i for which $c_i = c_{\min} = 1/k - 1$. The corresponding function is $f(A) = [\![i \notin A]\!]$, which corresponds to a family $\mathcal{F}$ with the wrong measure $\mu(\mathcal{F}) = 1 - k/n \neq k/n$.

The cross-intersecting version of the result follows using similar arguments by applying the cross-intersecting version of Hoffman's bound. □

The eigenvalues λ_i are related to λ_0 by the formula

$$\lambda_i = (-1)^i \frac{k}{n-k} \cdots \frac{k-i+1}{n-(k-i+1)} \lambda_0 \approx \left(-\frac{k}{n-k}\right)^i.$$

The eigenvalues we obtained in Lemma 3.3 were $(-p/(1-p))^i$. If we put $p = k/n$ then we obtain $\lambda_i/\lambda_0 \approx (-p/(1-p))^i$, showing that the eigenvalues in both cases behave in similar ways.

How do we generalize this proof method to t-intersecting families? It is natural to consider the extended Kneser graph whose edges connect vertices A, B such that $|A \cap B| < t$. However, the eigenvalues of this graph are somewhat awkward to compute. Instead, we will follow Wilson [79] and consider different weighted graphs.

Fact 3.38 ([79]). *Let $n, k \geq 1$ be integers satisfying $n \geq 2k$. For $s \in \{0, \ldots, k\}$, let $C^{(s)}$ be the matrix indexed by $\binom{[n]}{k}$ given by $C^{(s)}_{A,B} = \binom{s}{|A\cap B|}$ (which equals zero if $|A \cap B| > s$). The eigenspaces of $C^{(s)}$ are the subspaces V_i given in Fact 3.35, and the corresponding eigenvalues are*

$$\lambda_i^{(s)} = (-1)^i \binom{k-i}{s}\binom{n-k-i+s}{n-2k+s}.$$

(Wilson uses the notation $B^{(k-s)}$ for $C^{(s)}$.)

Note that $C^{(0)} = \mathrm{Kn}(n;k)$. The surprising property that all the matrices $C^{(s)}$ have the same eigenspaces is explained by the fact that these matrices commute. A deeper explanation lies in their forming a basis for the Bose–Mesner algebra of the Johnson association scheme (see Wilson [79] for references): Bose–Mesner algebras are always simultaneously diagonalizable.

The definition of $C^{(s)}$ makes it clear that $f'C^{(s)}f = 0$ whenever f is the characteristic function of a t-intersecting family and $s < t$. Therefore the admissible spectra (so to speak) are given by functions of the form

$$\lambda_i^* = \sum_{s=0}^{t-1} c_s (-1)^i \binom{k-i}{s}\binom{n-k-i+s}{n-2k+s}.$$

Since $\binom{n-k-i+s}{n-2k+s} = \binom{n-k-i+s}{k-i}$, if n is large then $\binom{n-k-i+s}{n-2k+s} \approx \binom{n-k-i}{n-2k}$ (recall that $i \leq k$ is small). Consequently,

$$\lambda_i^* \approx \sum_{s=0}^{t-1} c_s (-1)^i \binom{k-i}{s}\binom{n-k-i}{n-2k} = (-1)^i P(i)\binom{n-k-i}{n-2k},$$

where P is some polynomial of degree $\deg P < t$ depending on the coefficients c_s. Conversely, given such a polynomial P, we can calculate corresponding coefficients c_s. Putting $p = k/n$, we can further approximate

$$\lambda_i^* \approx P(i)\left(-\frac{p}{1-p}\right)^i.$$

This suggests that if we mimic the proof of Lemma 3.14, then we obtain an upper bound of roughly $p^t\binom{n}{k} = (k/n)^t\binom{n}{k}$ on the size of k-uniform t-intersecting families for large enough n. Wilson [79] laboriously carries out this idea, and shows that it works as long as $n \geq (t+1)(k-t+1)$, obtaining the correct upper bound $\binom{n-t}{k-t} \approx (k/n)^t\binom{n}{k}$. When $n > (t+1)(k-t+1)$, he is also able to prove uniqueness, that is the bound is achieved only for t-stars. Tokushige [78], whose work has already been mentioned in Section 3.4, extends this result to cross-t-intersecting families.

When $n < (t+1)(k-t+1)$, t-stars are no longer the optimal families. The optimal families are given by the Ahlswede–Khachatrian theorem, described in Chapter 10.

3.6 Friedgut's method and the Lovász theta function

Our solution of the traffic light puzzle in Section 3.1 involved proving the following theorem: if $\mathcal{F} \subseteq \{0,1,2\}^n$ is a family of vectors such that any two $x, y \in \mathcal{F}$ agree on some index, then $|\mathcal{F}| \leq 3^{n-1}$. Our proof went along the following lines:

1. Let $G = (V, E)$ be the graph whose vertex set is $V = \{0,1,2\}^n$, and two vectors $x, y \in V$ are connected if $x_i \neq y_i$ for all $i \in [n]$.
2. Let f be the characteristic vector of a family $\mathcal{F}$ as above, and let A be the adjacency matrix of G. Since $\mathcal{F}$ is an independent set of G, $f' A f = 0$.
3. The graph G is 2^n-regular and A has a minimal eigenvalue of -2^{n-1}, and so Hoffman's bound implies that $|\mathcal{F}|/|V| \leq 2^{n-1}/(2^n + 2^{n-1}) = 1/3$.

A similar argument is used to prove Theorem 3.16 on page 41, which states that a t-intersecting family has μ_p-measure at most p^t whenever $p \leq 1/(t+1)$:

1. Given n, we construct an edge-weighted directed graph $G = (V, E, w)$ on the vertex set $V = 2^{[n]}$ satisfying two properties:
   (a) For every edge $\{S, T\} \in E$, $|S \cap T| < t$.
   (b) The adjacency matrix A of G has a μ_p-orthonormal set of eigenvectors, and $A\mathbf{1} = \mathbf{1}$.
2. If f is the characteristic function of a t-intersecting family $\mathcal{F}$ then $f' A f = 0$, since $\mathcal{F}$ is an independent set in G.
3. Hoffman's bound implies that $\mu_p(\mathcal{F}) \leq -\lambda_{\min}/(1-\lambda_{\min})$, where $\lambda_{\min}$ is the minimal eigenvalue of A. Since $\lambda_{\min} = -p^t/(1-p^t)$, we conclude that $\mu_p(\mathcal{F}) \leq p^t$.

In other words, in both cases we reduce the problem of bounding the size of some constrained family to the problem of bounding the size of an independent set in an appropriate graph, and then use Hoffman's bound to obtain such a bound.

In the course of his exploration of the Shannon capacity of graphs, Lovász [66] came up with a different method of bounding the size of independent sets in graphs. As we show below, his method is stronger than Hoffman's bound. Chung and Richardson [9], adapting a method of Delsarte [12], were able to strengthen Lovász's method even further. We describe these various bounds and the relations among them in Section 3.6.1. We then explain in Section 3.6.2 how symmetry considerations lead to the precise way in which Friedgut's method employs Hoffman's bound.

3.6.1 Bounds on independent sets in graphs

In this section we explore the following problem: Given a graph G and a positive measure w on its vertices, what is the largest w-measure of an independent set in G? We will explore two different methods, one due to Hoffman and the other due to Lovász, leading to four different bounds. We will use the following two pieces of notation: $\mathbf{1}$ is the constant 1 vector (its length would be clear from context), and for a vector $x \in \mathbb{R}^n$, $\operatorname{diag}(x)$ is the $n \times n$ diagonal matrix satisfying $\operatorname{diag}(x)_{ii} = x_i$.

We start by defining weighted graphs and their independence number.

Definition 3.13. A *weighted graph* $G = (V, E, w)$ consists of a non-empty graph (V, E) along with a positive measure w on V. The *independence number* $\alpha(G)$ is the largest w-measure of an independent set in G. ○

Hoffman's method relies on the following version of Hoffman's bound [50, 51].

Lemma 3.39. *Let $G = (V, E, w)$ be a weighted graph. Suppose that A is a $V \times V$ matrix satisfying the following properties:*

1. *$A_{ij} \leq 0$ whenever $(i, j) \notin E$.*
2. *$A\mathbf{1} = \lambda\mathbf{1}$.*
3. *$A_{ij}w(i) = A_{ji}w(j)$ for all $i, j \in V$.*

If $\lambda_{\min}$ is the minimal eigenvalue of A then

$$\alpha(G) \leq \frac{-\lambda_{\min}}{\lambda - \lambda_{\min}} w(V).$$

Proof. Let $n = |V|$. Without loss of generality, we can assume that $w(V) = 1$. Let B be a $V \times V$ matrix given by $B_{ij} = A_{ij}\sqrt{w(i)/w(j)}$. Since

$$B_{ji} = A_{ji}\sqrt{\frac{w(j)}{w(i)}} = A_{ij}\frac{w(i)}{w(j)}\sqrt{\frac{w(j)}{w(i)}} = A_{ij}\sqrt{\frac{w(i)}{w(j)}} = B_{ij},$$

the matrix B is symmetric, and so has an orthonormal set of eigenvectors $v_1, \dots, v_n$ with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$. Writing $B = \sqrt{D}A\sqrt{D}^{-1}$, where $D = \mathrm{diag}(w)$, we see that v_i corresponds to an eigenvector u_i of A given by $(u_i)_k = (v_i)_k/\sqrt{w(k)}$. In particular, we can assume that v_1 corresponds to $\mathbf{1}$, and so $(v_1)_k = \sqrt{w(k)}$ and $\lambda_1 = \lambda$. Note that v_1 has unit norm since $w(V) = 1$.

Let $\mathcal{I}$ be any independent set in G, and let f be its characteristic function. Define a vector g by $g_k = \sqrt{w(k)}f_k$. Since $\mathcal{I}$ is an independent set,

$$0 \geq g'Bg = \sum_{i=1}^{n} \lambda_i \langle g, v_i\rangle^2 \geq \lambda\langle g, v_1\rangle^2 + \lambda_{\min}\sum_{i=2}^{n}\langle g, v_i\rangle^2 = \lambda\langle g, v_1\rangle^2 + \lambda_{\min}(\|g\|^2 - \langle g, v_1\rangle^2).$$

Easy calculation shows that $\langle g, v_1\rangle = \|g\|^2 = w(\mathcal{I})$. Therefore

$$0 \geq \lambda w(\mathcal{I})^2 + (\lambda - \lambda_{\min})(w(\mathcal{I}) - w(\mathcal{I})^2).$$

Dividing by $w(\mathcal{I})$ and rearranging, we get the stated bound. $\square$

The third property of A in Lemma 3.39 guarantees that A has a w-orthonormal set of eigenvectors.

As an example, in Section 3.2 we dealt with the following situation. The graph G has as vertex set all subsets of $[n]$. The weight of a subset $i \subseteq [n]$ is $w(i) = \mu_p(i)$, for some $p \in (0,1)$. Two subsets $i, j \subseteq [n]$ are connected if $i \cap j = \emptyset$, and so we have a self-loop on $\emptyset$. We considered the matrix

$$A = \begin{pmatrix} 1 - \frac{p}{q} & \frac{p}{q} \\ 1 & 0 \end{pmatrix}^{\otimes n}$$

with eigenvalues $\lambda = 1$ and $\lambda_{\min} = -p/q$ (where $q = 1 - p$), and concluded $\alpha(G) \leq p$. The corresponding matrix B appearing in the proof is

$$B = \begin{pmatrix} 1 - \frac{p}{q} & \sqrt{\frac{p}{q}} \\ \sqrt{\frac{p}{q}} & 0 \end{pmatrix}^{\otimes n},$$

showing that the third property of A is satisfied.

In all applications described in this chapter, we always had equality in the first property of A listed in Lemma 3.39. Indeed, if the bound is tight, then we must have $A_{ij} = 0$ whenever i, j both belong to some maximum independent set.

Given a graph G, the best possible bound obtainable using Hoffman's method can be determined by solving a semidefinite program. This was first noticed by Grötschel, Lovász and Schrijver [47] in the context of Lovász's bound.

Lemma 3.40. *Let $G = (V, E, w)$ be a weighted graph, and consider the following two semidefinite programs:*

$$\theta_H(G) = \min_{\substack{\text{symmetric } B\in\mathbb{R}^{V\times V}\\ \lambda_{\min}\in\mathbb{R}}} \frac{-\lambda_{\min}}{1-\lambda_{\min}} w(V) \qquad \theta'_H(G) = \min_{\substack{\text{symmetric } B\in\mathbb{R}^{V\times V}\\ \lambda_{\min}\in\mathbb{R}}} \frac{-\lambda_{\min}}{1-\lambda_{\min}} w(V)$$

$$\text{s.t. } B_{ij} = 0 \text{ whenever } (i,j) \notin E \qquad \text{s.t. } B_{ij} \le 0 \text{ whenever } (i,j) \notin E$$

$$B\sqrt{w} = \sqrt{w} \qquad B\sqrt{w} = \sqrt{w}$$

$$B \geq \lambda_{\min} I \qquad B \geq \lambda_{\min} I$$

Here $\sqrt{w}$ is the vector given by $\sqrt{w}_i = \sqrt{w(i)}$, and I is the $V \times V$ identity matrix.

Then $\alpha(G) \le \theta'_H(G) \le \theta_H(G)$.

Proof. Let A be a matrix satisfying the conditions of Lemma 3.39, and define a $V \times V$ matrix B by $B_{ij} = A_{ij}\sqrt{w(i)/w(j)}$. It is not hard to check that B satisfies the conditions of the program for θ'_H, and vice versa.

Since G is non-empty and $\theta_H(G) \ge \theta'_H(G) \ge \alpha(G)$, we get that in both cases $\lambda_{\min} \ge -\alpha(G)/(w(V) - \alpha(G))$, and so the minimum is actually obtained. □

The parameter $\theta'_H(G)$ equals the best possible bound which can be obtained using Lemma 3.39, while $\theta_H(G)$ corresponds to the version of Hoffman's bound used in the rest of this chapter.

An even better bound on the measure of independent sets is due to Lovász [66].

Lemma 3.41. *Let $G = (V, E, w)$ be a weighted graph. Suppose that A is a symmetric $V \times V$ matrix satisfying the following property: $A_{ij} \ge 0$ whenever $(i,j) \notin E$. Let W be the $V \times V$ matrix* *given by* $W_{ij} = \sqrt{w(i)w(j)}$*, and suppose that the maximal eigenvalue of* $A+W$ *is* λ*. Then*

$$\alpha(G) \le \lambda.$$

Proof. Since A and W are symmetric, $v'(A+W)v \le \lambda\|v\|^2$ for any vector v. Let $\mathcal{I}$ be an independent set, let f be its characteristic function, and define a vector g by $g_k = \sqrt{w(k)}f_k$. Then

$$\lambda w(\mathcal{I}) = \lambda\|g\|^2 \ge g'(A+W)g \ge g'Wg = w(\mathcal{I})^2.$$

Here we used $\|g\|^2 = w(\mathcal{I})$ and $g'Wg = w(\mathcal{I})^2$. For the latter, note that $W = \sqrt{w}\sqrt{w}'$ and so $g'Wg = g'\sqrt{w}\sqrt{w}'g = \langle\sqrt{w}, g\rangle^2 = w(\mathcal{I})^2$. We conclude that $w(\mathcal{I}) \le \lambda$. □

In fact, Lovász only considered the case in which the condition $A_{ij} \ge 0$ is always tight. The more general condition was used by Delsarte [12] in the context of error-correcting codes and ported to graphs by Chung and Richardson [9]; see also Schrijver [73]. As before, if Lovász's bound is tight then $A_{ij} = 0$ whenever i, j both belong to some maximum independent set.

Continuing our preceding example, the following matrix satisfies the conditions of Lemma 3.41:

$$A' = \begin{pmatrix} p-q & -\sqrt{pq} \\ -\sqrt{pq} & 0 \end{pmatrix}^{\otimes n}, \quad A' + W = \begin{pmatrix} p-q & -\sqrt{pq} \\ -\sqrt{pq} & 0 \end{pmatrix}^{\otimes n} + \begin{pmatrix} q & \sqrt{pq} \\ \sqrt{pq} & p \end{pmatrix}^{\otimes n}.$$

When $n = 1$, $A' + W = \operatorname{diag}(p, p)$, and in general, the maximal eigenvalue of $A' + W$ is p.

As in the case of Hoffman's method, the best bound obtainable using Lovász's method is the solution of a semidefinite program.

Lemma 3.42. *Let* $G = (V, E, w)$ *be a weighted graph, and consider the following two semidefinite programs:*

$$\begin{aligned} \theta_L(G) = &\min_{\substack{\text{symmetric } A\in\mathbb{R}^{V\times V} \\ \lambda\in\mathbb{R}}} \lambda \\ &\text{s.t. } A_{ij} = 0 \text{ whenever } (i,j) \notin E \\ &A + W \le \lambda I \end{aligned} \qquad \begin{aligned} \theta'_L(G) = &\min_{\substack{\text{symmetric } A\in\mathbb{R}^{V\times V} \\ \lambda\in\mathbb{R}}} \lambda \\ &\text{s.t. } A_{ij} \ge 0 \text{ whenever } (i,j) \notin E \\ &A + W \le \lambda I \end{aligned}$$

Here W *is the* $V \times V$ *matrix given by* $W_{ij} = \sqrt{w(i)w(j)}$*, and* I *is the* $V \times V$ *identity matrix.*

Then $\alpha(G) \le \theta'_L(G) \le \theta_L(G)$*.*

The usual Lovász theta function is $\theta_L(G)$. Sometimes the Lovász theta function is defined so that it coincides with $\theta_L(\overline{G})$, where $\overline{G}$ is the complemented graph.

It turns out that Lovász's method is stronger than Hoffman's method.

Lemma 3.43. *Let $G = (V, E, w)$ be a weighted graph. We have $\theta_L(G) \le \theta_H(G)$ and $\theta'_L(G) \le \theta'_H(G)$.*

Let $\sqrt{w}$ be the vector defined by $\sqrt{w}_i = \sqrt{w(i)}$. If the optimum of $\theta_L(G)$ is achieved for some matrix A having $\sqrt{w}$ as an eigenvector then $\theta_L(G) = \theta_H(G)$, and similarly for $\theta'_L(G)$ and $\theta'_H(G)$.

Proof. We start by proving that $\theta'_L(G) \le \theta'_H(G)$. Let B be any solution to the program for $\theta'_H(G)$, with minimal eigenvalue $\lambda_{\min}$. We will construct a matrix A which is a solution to the program for $\theta'_L(G)$ and whose maximal eigenvalue is at most $-\lambda_{\min}/(1-\lambda_{\min}) \cdot w(V)$. It follows that $\theta'_L(G) \le \theta'_H(G)$.

Let W be the matrix given by $W_{ij} = \sqrt{w(i)w(j)}$. It is easy to check that W is a rank 1 matrix satisfying $W\sqrt{w} = w(V)\sqrt{w}$. The matrix A will have the form $A = -xB$, for an $x > 0$ to be determined. Clearly $A_{ij} \ge 0$ whenever $(i,j) \notin E$. Since $B\sqrt{w} = \sqrt{w}$, the maximal eigenvector of $B + W$ is $\max(w(V) - x, -x\lambda_{\min})$. Choosing $x = w(V)/(1-\lambda_{\min})$, both expressions are equal to $-\lambda_{\min}/(1-\lambda_{\min}) \cdot w(V)$.

Next, suppose that the optimum of $\theta'_L(G)$ is achieved for some matrix A having $\sqrt{w}$ as an eigenvector with eigenvalue λ. Since G is non-empty, $\theta'_L(G) \le \theta_L(G) < w(V)$. Indeed, if M is the adjacency matrix then the maximal eigenvalue of $(-\epsilon M) + W$ is smaller than $w(V)$ for small enough $\epsilon > 0$.

Let λ' be the maximal eigenvalue of A among all eigenvectors different from $\sqrt{w}$. We claim that $\lambda' = \theta'_L(G)$. Indeed, since $W\sqrt{w} = v(W)\sqrt{w}$, the maximal eigenvalue of $xA + W$ for $x > 0$ is $\lambda(x) = \max(x\lambda', x\lambda + w(V))$, which implies (putting $x = 1$) that $\lambda \le \theta'_L(G) - w(V) < 0$. If $\lambda' < \lambda + w(V)$ then for small enough $\epsilon > 0$, $\lambda(1+\epsilon) < \lambda(1)$, contradicting the definition of A. Similarly, $\lambda + w(V) = \theta'_L(G)$, since otherwise $\lambda(1-\epsilon) < \lambda(1)$ for small enough $\epsilon > 0$.

Let $B = A/\lambda$. Clearly $B\sqrt{w} = \sqrt{w}$ and so B is feasible for the semidefinite program for

$\theta'_H(G)$. Moreover, the minimal eigenvalue of B is $\lambda_{\min} = \min(1, \lambda'/\lambda) = \theta'_L(G)/\lambda$, and so

$$\frac{-\lambda_{\min}}{1-\lambda_{\min}} w(V) = \frac{\theta'_L(G)}{-\lambda + \theta'_L(G)} w(V) = \theta'_L(G).$$

We conclude that $\theta'_H(G) \leq \theta'_L(G)$ and so $\theta'_H(G) = \theta'_L(G)$.

The argument for $\theta_L(G), \theta_H(G)$ is completely analogous. □

3.6.2 Symmetry considerations

When constructing the matrix used to prove Theorem 3.16 on the maximum μ_p-measure of t-intersecting families, we were looking for a matrix whose eigenvectors are the p-skewed Fourier characters. Why is this condition meaningful? While we are not able to provide a satisfactory answer for the p-skewed case, when $p = 1/2$ we can show that the bounds $\theta_H(G), \theta'_H(G), \theta_L(G), \theta'_L(G)$ are attained for matrices whose eigenvectors are the Fourier basis vectors. As we show below, this implies that $\theta_H(G) = \theta_L(G)$ and $\theta'_H(G) = \theta'_L(G)$, and so in this case Lovász's bound is as strong as Hoffman's bound.

Our argument will apply to *agreement graphs*.

Definition 3.14. An *agreement graph* is a non-empty graph $G = (V, E)$ where $V = \mathbb{Z}_m^n$ for some n, m and $E = \{\{S, T\} : S - T \in \mathcal{G}\}$ for some $\mathcal{G} \subseteq V$.

We view an agreement graph G as a weighted graph by giving all the vertices weight 1. ○

Here are two examples:

1. The graph considered in Section 3.1 on the traffic light puzzle has $V = \mathbb{Z}_3^n$ and $E = \{\{S, T\} : S - T \in \{1, 2\}^n\}$.

2. Chapter 4 considers *triangle-agreeing* families of graphs. These are families of graphs on the vertex set $[n]$ such that the agreement $G_1 \nabla G_2 = \overline{G_1 \Delta G_2}$ of any two graphs G_1, G_2 in the family contains a triangle. In this case $V = \mathbb{Z}_2^{\binom{n}{2}}$ and $\mathcal{G}$ consists of all triangle-free graphs.

   Since every triangle-intersecting family (in which the intersection of any two graphs contains a triangle) is a fortiori triangle-agreeing, a bound on the size of triangle-agreeing families is stronger than a bound on the size of triangle-intersecting families. In fact, Theorem 4.27 on page 100 shows that the maximal size of a triangle-intersecting family is the same as the maximal size of a triangle-agreeing family (the theorem works for any choice of $\mathcal{G}$).

An agreement graph has many automorphisms: for every $x \in V$, $S \mapsto S + x$ is an automorphism. This allows us to symmetrize the matrices appearing in the various semidefinite programs considered in the preceding section.

Lemma 3.44. *Let $G = (V, E)$ be an agreement graph, and let $\theta \in \{\theta_H, \theta'_H, \theta_L, \theta'_L\}$. There is an optimal matrix for the semidefinite program for $\theta(G)$ whose eigenvectors are the Fourier basis vectors (defined in Section 2.5.1).*

Proof. We only prove the case $\theta = \theta_H$, the other cases being similar. Let B be any matrix which is feasible for the semidefinite program for $\theta_H(G)$. For every $x \in V$, define a matrix B^x by $B^x_{S,T} = B_{S+x,T+x}$, and note that B^x is similar to B. It is not hard to check that B^x is also feasible with the same minimal eigenvalue $\lambda_{\min}$. Therefore $C = \mathbb{E}_{x\in V} B^x$ is also feasible. Furthermore, for every vector f we have $f'Cf = \mathbb{E}_{x\in V} f'B^x f \geq \lambda_{\min}\|f\|^2$, showing that the minimal eigenvalue of C is at least $\lambda_{\min}$. Clearly $C_{S,T} = C_{S+x,T+x}$, and we conclude that the optimum in the semidefinite program for $\theta_H(G)$ is obtained for some matrix C satisfying $C_{S,T} = C_{S+x,T+x}$ for all $x \in V$, since the objective value $-\lambda_{\min}/(1 - \lambda_{\min}) = 1 - 1/(1 - \lambda_{\min})$ is monotone decreasing with $\lambda_{\min}$.

In order to complete the proof, we show that if C satisfies $C_{S,T} = C_{S+x,T+x}$ for all $x \in V$ then its eigenvectors are the Fourier basis vectors. Let $V = \mathbb{Z}_m^n$, and let $\omega = e^{2\pi i/m}$ be a primitive mth root of unity. Recall that the Fourier basis vectors are given by $\chi_x(y) = \omega^{\langle\!\langle x,y\rangle\!\rangle}$, where $\langle\!\langle x, y\rangle\!\rangle = \sum_{i=1}^n x_i y_i$. We have

$$(C\chi_x)(y) = \sum_{z\in V} C_{y,z}\chi_x(z) = \sum_{z\in V} C_{\mathbf{0},z-y}\chi_x(z) = \sum_{z\in V} C_{\mathbf{0},z}\chi_x(z + y)$$
$$= \chi_x(y) \sum_{z\in V} C_{\mathbf{0},z}\chi_x(z) = \chi_x(y)(C\chi_x)(\mathbf{0}),$$

since $\chi_x(z + y) = \chi_x(z)\chi_x(y)$; here $\mathbf{0}$ is the zero vector. □

Corollary 3.45. *Let $G = (V, E)$ be an agreement graph. Then $\theta_L(G) = \theta_H(G)$ and $\theta'_L(G) = \theta'_H(G)$.*

Proof. The lemma shows that each of the semidefinite programs $\theta_L(G), \theta'_L(G)$ has an optimal solution for which $\mathbf{1}$ is an eigenvector, and so the claim follows from Lemma 3.43. $\square$

If $V = \mathbb{Z}_2^n$ and $\mathcal{G}$ is downwards-closed (using the identification $\mathbb{Z}_2^n = 2^n$), then the development in Section 3.3.1 restricts the structure of an optimal matrix for $\theta_H(G)$ even further.

Lemma 3.46. *Let $G = (V, E)$ be an agreement graph with $V = 2^n$, and suppose that $\mathcal{G} = \{S \in V : \{V, \varnothing\} \in E\}$ is downwards-closed. Then there is a matrix B achieving the optimum in the semidefinite program for $\theta_H(G)$ which is in the span of the matrices $\{B_J : J \in \mathcal{G}\}$, where B_J is the matrix defined in Lemma 3.10 on page 37 for $p = 1/2$.*

Proof. The proof is the same as the proof of Lemma 3.19 on page 44. $\square$

Corollary 3.47. *Let $G = (V, E)$ be an agreement graph with $V = 2^n$, and suppose that $\mathcal{G} = \{S \in V : \{V, \varnothing\} \in E\}$ is downwards-closed. Then $\theta_H(G)$ is given by the following linear program:*

$$\min_{c_J \text{ for } J \in \mathcal{G}} \frac{-\lambda_{\min}}{1 - \lambda_{\min}} |V|$$

$$\text{s.t. } c_\varnothing = 1$$

$$(-1)^{|S|} \sum_{J \in \mathcal{G}} c_J [\![J \subseteq S]\!] \geq \lambda_{\min} \text{ for all } S \subseteq V.$$

Equivalently, $\theta_H(G) = -\lambda_{\min}/(1 - \lambda_{\min}) \cdot |V|$, where

$$\lambda_{\min} = \max_{\substack{c_J \text{ for } J \in \mathcal{G} \\ c_\varnothing = 1}} \min_{S \subseteq V} (-1)^{|S|} \sum_{J \in \mathcal{G}} c_J [\![J \subseteq S]\!].$$

Proof. For a matrix B whose eigenvectors are the Fourier basis vectors, let $\lambda_S(B)$ be the eigenvalue corresponding to χ_S. Lemma 3.17 on page 43 shows that for each $J \in \mathcal{G}$ and for each function f on J, there is a matrix B in the span of $\{B_K : K \in \mathcal{G}\}$ such that $\lambda_S(B) = (-1)^{|S|} f(S \cap J)$. Lemma 3.10 on page 37 shows that $\lambda_S(B_J) = (-1)^{|S|}(-1)^{|S \cap J|}$. Together, these lemmas imply (using inclusion-exclusion) that for B in the span of $\{B_K : K \in \mathcal{G}\}$, $\lambda_S(B)$ is always of the form

$$\lambda_S(B) = (-1)^{|S|} \sum_{J \in \mathcal{G}} c_J [\![J \subseteq S]\!],$$

and furthermore all such functions are achievable. Since $\sqrt{w} = \mathbf{1} = \chi_\varnothing$, the condition $B\sqrt{w} = \sqrt{w}$ corresponds to $\lambda_\varnothing(B) = 1$, which is the same as $c_\varnothing = 1$. $\square$

A similar situation arises in the context of Section 3.5.3 on page 62, in which a similar argument shows that the optimal matrix has the subspaces V_i given by Lemma 3.35 as eigenspaces, and we can use a similar argument to reduce the computation of θ_H, θ_L to a linear program.

Chapter 4

Triangle-intersecting families of graphs

How big can a family of graphs on n vertices be, if the intersection of any two of them contains a triangle? Simonovits and Sós were the first to raise this question, around 1976, in the context of their studies of graphical intersection problems. They conjectured that such a family can contain at most $2^{\binom{n}{2}-3}$ graphs, the optimal families being $\triangle$-*stars*, supersets of a fixed triangle.

A decade later, Chung, Frankl, Graham and Shearer [10] were the first to prove a non-trivial upper bound on the size of triangle-intersecting families of graphs. Using Shearer's lemma, they gave an upper bound of $2^{\binom{n}{2}-2}$. Finally, 25 years later, together with David Ellis and Ehud Friedgut, we were able to settle the conjecture in the affirmative [27].

Surprisingly, all known upper bounds on triangle-intersecting families apply to the wider class of odd-cycle-intersecting families, in which we only require the intersection of any two graphs in the family to contain an odd cycle. In other words, the intersection of any two graphs must be non-bipartite. (A similar phenomenon happens with respect to triangle-free graphs: a maximum triangle-free graph is also bipartite.)

Moreover, all upper bounds apply to odd-cycle-*agreeing* families, in which instead of looking at the intersection $G_1 \cap G_2$ of any two graphs, we look at their *agreement* $G_1 \nabla G_2 = \overline{G_1 \Delta G_2} = (G_1 \cap G_2) \cup (\overline{G_1} \cap \overline{G_2})$. Chung, Frankl, Graham and Shearer showed that this phenomenon is general, as we discuss in Section 4.4.

Let us summarize all the definitions we have made so far.

Definition 4.1. The *agreement* of two sets $A, B \subseteq U$ with respect to U is

$$A \nabla B = \overline{A \Delta B} = (A \cap B) \cup (\overline{A} \cap \overline{B}),$$

where all complements are with respect to U. ○

Definition 4.2. A *family of graphs on n vertices* is a collection of graphs on the vertex set $[n]$, considered as sets of edges.

A family of graphs is *triangle-intersecting* if the intersection of any two graphs in the family contains a triangle. It is *odd-cycle-intersecting* if the intersection of any two graphs in the family contains an odd cycle. It is *triangle-agreeing* if the agreement of any two graphs in the family (with respect to the complete graph on $[n]$) contains a triangle. It is *odd-cycle-agreeing* if the agreement of any two graphs in the family contains an odd-cycle.

A $\triangle$-*star* is a family of graphs of the form $\{G \subseteq K_n : G \supseteq T\}$, where T is a triangle. A $\triangle$-*semistar* is a family of graphs of the form $\{G \subseteq K_n : G \cap T = S\}$, where T is a triangle and $S \subseteq T$. ○

We can now state the main theorem of this chapter, settling the Simonovits–Sós conjecture.

Theorem 4.1. *Let $\mathcal{F}$ be an odd-cycle-agreeing family of graphs on n vertices.*

Upper bound: $\mu(\mathcal{F}) \leq 1/8$.

Uniqueness: $\mu(\mathcal{F}) = 1/8$ *if and only if $\mathcal{F}$ is a $\triangle$-semistar.*

Stability: *If $\mu(\mathcal{F}) \geq 1/8 - \epsilon$ then there is a $\triangle$-semistar $\mathcal{G}$ such that $\mu(\mathcal{F} \Delta \mathcal{G}) = O(\epsilon)$.*

Our proof generalizes to families of hypergraphs. A hypergraph is a collection of non-empty subsets of $[n]$, called hyperedges. The counterpart of a triangle is a Schur triple of hyperedges, $\{A, B, A \Delta B\}$. Equivalently, it is a triple of hyperedges A, B, C satisfying $A \Delta B \Delta C = \varnothing$. The counterpart of an odd cycle is an odd number of hyperedges $A_1, \ldots, A_{2k+1}$ whose symmetric difference vanishes.

Definition 4.3. A *hypergraph on n points* is a collection of non-empty subsets of $[n]$, called *hyperedges*. A *family of hypergraphs on n points* is a family consisting of hypergraphs on n points.

A *Schur triple* consists of three sets $A, B, A\Delta B$. An *odd circuit* consists of $2k+1$ sets $A_1, \ldots, A_{2k+1}$ satisfying $A_1 \Delta \cdots \Delta A_{2k+1} = \varnothing$. (A Schur triple is an odd circuit of length 3.)

A family of hypergraphs is *Schur-triple-intersecting* if the intersection of any two hypergraphs contains a Schur triple. The concepts of *odd-circuit-intersecting*, *Schur-triple-agreeing* and *odd-circuit-agreeing* are defined similarly.

A *Schur-star* is a family of hypergraphs of the form $\{H \subseteq 2^{[n]} \setminus \{\varnothing\} : H \supseteq T\}$, where T is a Schur triple. A *Schur-semistar* is a family of hypergraphs of the form $\{H \subseteq 2^{[n]} \setminus \{\varnothing\} : H \cap T = S\}$, where T is a Schur triple and $S \subseteq T$. ○

We can now state the analog of Theorem 4.1 for hypergraphs.

Theorem 4.2. *Let $\mathcal{F}$ be an odd-circuit-agreeing family of hypergraphs on n points.*

Upper bound: $\mu(\mathcal{F}) \leq 1/8$.

Uniqueness: $\mu(\mathcal{F}) = 1/8$ *if and only if $\mathcal{F}$ is a Schur-semistar.*

Stability: *If $\mu(\mathcal{F}) \geq 1/8 - \epsilon$ then there is a Schur-semistar $\mathcal{G}$ such that $\mu(\mathcal{F}\Delta\mathcal{G}) = O(\epsilon)$.*

This theorem in fact generalizes Theorem 4.1: given a family of graphs, we can extend it to a family of hypergraphs with the same measure by replacing each graph with all possible hypergraphs containing it. If the original family is odd-cycle-agreeing then the new family will be odd-circuit-agreeing.

Both of our main theorems generalize to the μ_p measure for $p < 1/2$ with p^3 replacing $1/8$, at the cost of applying to intersecting families rather than agreeing families (when $p \neq 1/2$, the symmetry between edges and non-edges is lost); the μ_p measure of a graph G with m edges is $\mu_p(G) = p^m(1-p)^{\binom{n}{2}-m}$.

Theorem 4.3. *Let $\mathcal{F}$ be an odd-circuit-intersecting family of hypergraphs on n points, and suppose* $0 < p < 1/2$.

Upper bound: $\mu_p(\mathcal{F}) \leq p^3$.

Uniqueness: $\mu_p(\mathcal{F}) = p^3$ *if and only if* $\mathcal{F}$ *is a Schur-star.*

Stability: *If* $\mu_p(\mathcal{F}) \geq p^3 - \epsilon$ *then* $\mu_p(\mathcal{F} \Delta \mathcal{G}) \leq K_p \epsilon$ *for some Schur-star* $\mathcal{G}$*, where* K_p *is a constant depending continuously on* p *in the interval* $(0, 1/2)$.

This theorem enables us to use Theorem 3.33 to obtain results concerning odd-circuit-intersecting families of hypergraphs with prescribed number of edges.

Theorem 4.4. *For every* $\delta > 0$ *there are constants* C_δ, N_δ *such that for any* $k \in (0, n/2 - \delta n)$ *and any odd-circuit-intersecting* k*-uniform family* $\mathcal{F}$ *of hypergraphs on* $n \geq N_{t,\delta}$ *points,*

$$|\mathcal{F}| < \binom{n-3}{k-3} + \epsilon \binom{n}{k}.$$

If furthermore $\mathcal{F}$ *satisfies*

$$|\mathcal{F}| > \binom{n-3}{k-3} - \epsilon \binom{n}{k},$$

then there exists a Schur-star $\mathcal{H}$ *such that*

$$|\mathcal{F} \Delta \operatorname{Sl}(\mathcal{H}, k)| < C_\delta \left(\epsilon + \sqrt{\frac{\log n}{n}} \right) \binom{n}{k}.$$

A similar theorem is true for odd-cycle-intersecting families of graphs.

Theorem 4.5. *For every* $\delta > 0$ *there are constants* C_δ, N_δ *such that for any* $k \in (0, n/2 - \delta n)$ *and any odd-cycle-intersecting* k*-uniform family* $\mathcal{F}$ *of graphs on* $n \geq N_{t,\delta}$ *points,*

$$|\mathcal{F}| < \binom{n-3}{k-3} + \epsilon \binom{n}{k}.$$

If furthermore $\mathcal{F}$ *satisfies*

$$|\mathcal{F}| > \binom{n-3}{k-3} - \epsilon \binom{n}{k},$$

then there exists a $\triangle$*-star* $\mathcal{H}$ *such that*

$$|\mathcal{F} \Delta \operatorname{Sl}(\mathcal{H}, k)| < C_\delta \left(\epsilon + \sqrt{\frac{\log n}{n}} \right) \binom{n}{k}.$$

Roadmap. We start our exposition with the bound from [10], proven in Section 4.1. We then present in Section 4.2 an unpublished proof of the Simonovits–Sós conjecture for families of graphs on 8 vertices. Section 4.3 is devoted to our proof of the Simonovits–Sós conjecture for odd-cycle-intersecting families of graphs, and is the main section of this chapter. While the proof in Section 4.3 already gives the correct upper bound for odd-cycle-agreeing families of graphs, uniqueness and stability need a further argument which appears in Section 4.4. The extension to hypergraphs appears in Section 4.5. The chapter closes with the extension to μ_p measures for $p < 1/2$. Relevant open problems are discussed in Chapter 10.

Unless otherwise specified, all the material is taken from our joint paper with David Ellis and Ehud Friedgut [27].

We will use K_A to denote the complete graph on the vertex set A and $K_{A,B}$ to denote the complete bipartite graph with bipartitions A and B. Also, $K_n = K_{[n]}$.

4.1 Bound using Shearer's lemma

Prior to our work, the best known upper bound on the size of triangle-intersecting families of graphs was $2^{\binom{n}{2}-2}$. The proof, which consists of a simple application of Shearer's lemma, brings forth some ideas which will be useful later. The contents of this section are taken mainly from Chung, Frankl, Graham and Shearer [10].

We begin by stating a generalization of Shearer's lemma due to Friedgut [72, 39].

Lemma 4.6 (Shearer's lemma)**.** *Let S be a finite set, and let $A_1, \ldots, A_m$ be subsets of S which cover every element of S exactly k times. For a family $\mathcal{F}$ of subsets of S and $A \subseteq S$, let the* projection of $\mathcal{F}$ to A *be the family*

$$\operatorname{proj}(\mathcal{F}, A) = \{B \subseteq A : X \cap A = B \textit{ for some } X \in \mathcal{F}\}.$$

For every $p \in (0,1)$,

$$\mu_p(\mathcal{F})^k \leq \prod_{i=1}^{m} \mu_p(\operatorname{proj}(\mathcal{F}, A_i)).$$

The idea of the upper bound is that if $\mathcal{F}$ is odd-cycle-intersecting and G is a bipartite graph, then if we remove the edges in G, $\mathcal{F}$ must still be intersecting, as a family of unstructured sets.

Since G can be chosen to contain roughly half the edges, Shearer's lemma results in a good upper bound.

Theorem 4.7. *Let $\mathcal{F}$ be an odd-cycle-intersecting family of graphs on n points. Then for all $p \le 1/2$, $\mu_p(\mathcal{F}) \le p^2$. If $p = 1/2$, then it is enough to assume that $\mathcal{F}$ is odd-cycle-agreeing.*

Proof. Let $\mathcal{F}$ be an odd-cycle-intersecting family of graphs on n points. Let $\mathcal{A}$ be the collection of all *complements* of $K_{L,R}$, where L, R form an unordered partition of $[n]$. For every $A \in \mathcal{A}$, the projection $\text{proj}(\mathcal{F}, A)$ must be intersecting, and so the Erdős–Ko–Rado theorem implies that $\mu_p(\text{proj}(\mathcal{F}, A)) \le p$ for all $p \le 1/2$. Each edge appears in exactly half of the graphs in $\mathcal{A}$ (those for which one endpoint is in L and the other is in R). Therefore Shearer's lemma immediately gives $\mu_p(\mathcal{F}) \le p^2$.

If $\mathcal{F}$ is odd-cycle-agreeing then the projection $\text{proj}(\mathcal{F}, A)$ is an agreeing family: for every two sets $S, T \in \text{proj}(\mathcal{F}, A)$, $S \nabla T \neq \varnothing$. Since $S \nabla \overline{S} = \varnothing$, $\text{proj}(\mathcal{F}, A)$ contains at most half of the sets, that is $\mu(\text{proj}(\mathcal{F}, A)) \le 1/2$. Applying Shearer's lemma again, we conclude that $\mu(\mathcal{F}) \le 1/4$. □

4.2 Families of graphs on eight vertices

In this section we prove that an odd-cycle-agreeing family of graphs on 8 vertices contains at most 1/8 of the graphs, following our unpublished manuscript [33]. Unfortunately, we have not been able to extend the proof beyond 8 vertices.

The basic idea is to divide all graphs on 8 vertices into sets of size 8. Each set S satisfies the following property: the agreement of any two distinct graphs in S is bipartite. Therefore an odd-cycle-agreeing family can contain at most one graph from S.

If we identify a graph on 8 vertices with a vector in $\mathbb{Z}_2^{28}$, the set of all graphs on 8 vertices becomes a vector space of dimension 28. Our sets will be cosets of a single vector space V of dimension 3. The construction of V utilizes a permutation with a special property.

Definition 4.4. A permutation $\pi \in S_{\{0,\ldots,7\}}$ is *antilinear* if $\pi(0) = 0$ and $\pi(x) \oplus \pi(y) \neq \pi(x \oplus y)$. Here $\oplus$ is the familiar XOR operation, which corresponds to addition in $\mathbb{Z}_2^3$; we identify $\mathbb{Z}_2^3$ with $\{0, \ldots, 7\}$ via the mapping $(a, b, c) \mapsto 4a + 2b + c$. ○

Lemma 4.8. *The permutation* $\pi = (1234)$ *is antilinear.*

Proof. In order to verify the second constraint, we go over all possible unordered triplets $x, y, x \oplus y$, which are given by the seven lines of the Fano plane:

$$\pi(2) \oplus \pi(4) \oplus \pi(6) = 3 \oplus 1 \oplus 6 = 4,$$

$$\pi(1) \oplus \pi(4) \oplus \pi(5) = 2 \oplus 1 \oplus 5 = 6,$$

$$\pi(3) \oplus \pi(4) \oplus \pi(7) = 4 \oplus 1 \oplus 7 = 2,$$

$$\pi(1) \oplus \pi(2) \oplus \pi(3) = 2 \oplus 3 \oplus 4 = 5,$$

$$\pi(2) \oplus \pi(5) \oplus \pi(7) = 3 \oplus 5 \oplus 7 = 1,$$

$$\pi(1) \oplus \pi(6) \oplus \pi(7) = 2 \oplus 6 \oplus 7 = 3,$$

$$\pi(3) \oplus \pi(5) \oplus \pi(6) = 4 \oplus 5 \oplus 6 = 7.$$

□

Given an antilinear permutation, the construction of V is very simple.

Lemma 4.9. *Let* $\pi \in S_{\{0,\ldots,7\}}$ *be an antilinear permutation. Define a set* $V \subset \mathbb{Z}_2^{28}$ *by*

$$V = \{v_k : k \in \{0, \ldots, 7\}\}, \quad v_k(i,j) = \langle\!\langle \pi(i \oplus j), k \rangle\!\rangle.$$

(Here $\langle\!\langle 4a_1 + 2b_1 + c_1, 4a_2 + 2b_2 + c_2 \rangle\!\rangle = a_1 a_2 + b_1 b_2 + c_1 c_2 \in \mathbb{Z}_2$ *is an inner product on* $\mathbb{Z}_2^3$*.) The set* V *is a vector space of dimension* 3*, and the agreement between any two distinct graphs in* V *is a cube.*

Proof. Since $v_i + v_j = v_{i \oplus j}$, V is a vector space of dimension 3. Hence it is enough to show that for $k \neq 0$, $v_k \nabla v_0 = \overline{v_k}$ is a cube. The graph $\overline{v_k}$ contains the edge (i,j) whenever $\langle\!\langle \pi(i \oplus j), k \rangle\!\rangle = 0$, or equivalently $i \oplus j \in \pi^{-1}(k^\perp)$. Here $k^\perp$ is the orthogonal complement of k. Therefore i has three neighbors $i \oplus \pi^{-1}(k^\perp \setminus \{0\})$. We show below that the three points $\pi^{-1}(k^\perp \setminus \{0\})$ are linearly independent, and so $\overline{v_k}$ is a cube. Indeed, suppose $\pi^{-1}(k^\perp \setminus \{0\}) = \{x, y, z\}$. If the vectors are not linearly independent then $z = x \oplus y$, but then $\pi(x) \oplus \pi(y) \oplus \pi(z) \neq 0$ by antilinearity, whereas the points in $k^\perp$ sum to zero. □

The upper bound easily follows.

Theorem 4.10. *Suppose $\mathcal{F}$ is an odd-cycle-agreeing family of graphs on $n \le 8$ vertices. Then $\mu(\mathcal{F}) \le 1/8$.*

Proof. By extending $\mathcal{F}$ if necessary, we can assume that $n = 8$. Let V be the vector space constructed in Lemma 4.9 using $\pi = (1234)$. For every graph G (regarded as a vector in $\mathbb{Z}_2^{28}$), it is easy to see that the coset $V + G$ also satisfies the property that the agreement between any two distinct graphs in the coset is a cube. Hence $\mathcal{F}$ contains at most one graph from each coset, implying $\mu(\mathcal{F}) \le 1/8$. □

This proof idea cannot be extended beyond 8 vertices. Indeed, suppose V is a subspace of dimension 3 of the vector space of all graphs on $n \ge 9$ vertices, with the property that the agreement between any two distinct graphs is triangle-free. Pick a basis for V. Assign to each edge a color in $\mathbb{Z}_2^3$ according to its 0/1 status in each of the basis vectors.

Let a, b, c be the colors assigned to some triangle. For each k, there is a graph in V in which the status of the edges in the triangle is $\langle\!\langle a, k\rangle\!\rangle, \langle\!\langle b, k\rangle\!\rangle, \langle\!\langle c, k\rangle\!\rangle$. These can never be all zero, since otherwise the complement of the respective graph in V contains a triangle. Hence a, b, c must be linearly independent, and in particular different and non-zero.

Since there are only 7 non-zero colors but at least 8 edges incident to any vertex, the pigeonhole principle shows that there must be some triangle whose colors are not linearly independent (due to either a zero color or a repeated color). This shows that V cannot exist.

There are other ways to extend the argument. For example, we can drop the assumption that V is a subspace, or we could demand that V have size $8k$ and that any set of $k+1$ distinct graphs in it contains two whose agreement is triangle-free. Using similar but more elaborate arguments, one can show that even with these extensions, the object V can only exist for bounded n. More details can be found in [33].

4.3 Proof of the Simonovits–Sós conjecture

In this section, we prove Theorem 4.1 for odd-cycle-intersecting families. In fact, our proof will already yield the upper bound for odd-cycle-agreeing families, but for uniqueness and stability more work is necessary (this work is taken up in the next section). The proof, which employs

Friedgut's method, follows closely the steps outlined in Section 3.3, and the reader is advised to read that section prior to the present one. For the rest of this section, fix the number of vertices n.

Following our previous footsteps, the idea of the proof is to find a matrix A of dimension $\binom{n}{2}$ satisfying the following properties:

- For every odd-cycle-intersecting family $\mathcal{F}$, $f'Af = 0$, where $f = \mathbf{1}_{\mathcal{F}}$.
- The eigenvectors of $\mathcal{F}$ are the Fourier characters χ_S.
- The eigenvalue corresponding to $\chi_\varnothing$ is 1.
- All other eigenvalues are at least $-1/7$.

Since our goal is to get a bound of 1/8, Hoffman's bound tells us that the eigenvalues need to be at least $-(1/8)/(1-1/8) = -1/7$, which explains the last item. As before, our first step is identifying the admissible matrices.

Definition 4.5. Let $n \geq 1$ be an integer. We say that a matrix A is *odd-cycle-admissible* (*admissible* for short) if it satisfies the following two properties:

Intersection property If G, H are graphs whose intersection contains an odd cycle then $\mathbf{1}_G' A \mathbf{1}_H = 0$.

Eigenvector property The eigenvectors of $\mathcal{F}$ are the Fourier characters χ_S.

If A is admissible then we use $\lambda_G(A)$ to denote the eigenvalue corresponding to χ_G. ○

Lemma 4.11. *Let J be a set of edges. Define*

$$B_{J,i} = \begin{cases} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, & \text{if } i \notin J, \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, & \text{if } i \in J, \end{cases}$$

$$B_J = \bigotimes_{i=1}^{n} B_{J,i}.$$

If J is bipartite then the matrix B_J is admissible, and the eigenvalue corresponding to χ_G is

$$\lambda_G(B_J) = (-1)^{|S \setminus J|}.$$

Furthermore, the vector space of all admissible matrices is spanned by B_J for all bipartite J.

Proof. Note that when $p = 1/2$, the matrix $A^{[1]}$ of Lemma 3.2 becomes

$$A^{[1]} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The proof that B_J is admissible for bipartite J is a straightforward modification of the proof of Lemma 3.10. The crucial point is that if the intersection of two graphs G, H contains an odd cycle, then $(G \cap H) \setminus J$ cannot be empty since J contains no odd cycles.

The other statement follows from a straightforward modification of the proof of Lemma 3.19. Here the crucial point is that a supergraph of a non-bipartite graph is non-bipartite. This allows the reverse induction argument to go through. □

We are looking for an admissible matrix A such that $\lambda_\varnothing(A) = 1$ and $\lambda_G(A) \geq -1/7$ for all graphs G. Following our reasoning in Section 3.3, the existence of odd-cycle-intersecting families of measure $1/8$ (namely, $\triangle$-stars) implies that $\lambda_G(A) = -1/7$ for non-empty subgraphs G of any triangle. In Section 3.3, this was enough data to determine the matrix A. What makes the present problem much more difficult is that these constraints are not enough to determine A. Instead, we will restrict ourselves to a smaller supply of building blocks.

The proof of the upper bound $1/4$ in Section 4.1 relied on projecting the family to the complement of a random bipartite graph. The same construction will serve us here as well.

Definition 4.6. For a graph H, define a function q_H on graphs by

$$q_H(G) = \Pr_{L,R}[G \cap K_{L,R} \text{ is isomorphic to } H],$$

where L, R is a random bipartition of $[n]$ chosen by putting each $i \in [n]$ independently in L or in R with equal probability $1/2$. Similarly, for an integer k, define

$$q_k(G) = \Pr_{L,R}[|G \cap K_{L,R}| = k].$$

○

Lemma 4.12. *For every graph H there is an admissible matrix E_H such that* $\lambda_G(E_H) = (-1)^{|G|} q_H(G)$.

For every integer k there is an admissible matrix E_k such that $\lambda_G(E_k) = (-1)^{|G|} q_k(G)$.

Proof. We show that for every graph H there is an admissible matrix E'_H such that

$$\lambda_G(E'_H) = (-1)^{|G|} \Pr_{L,R}[G \cap K_{L,R} = H].$$

Given the matrices E'_H, it is easy to construct the matrices E_H and E_k.

We construct E'_H by taking an average over admissible matrices $E'_{H,L,R}$ satisfying

$$\lambda_G(E'_{H,L,R}) = (-1)^{|G|} [\![G \cap K_{L,R} = H]\!].$$

The existence of admissible matrices $E'_{H,L,R}$ satisfying this formula follows directly from Lemma 3.17.

□

It turns out that in order to get an upper bound of 1/8, it is enough to take a linear combination of the matrices E_k. In order to get uniqueness, we will have to throw in some of the matrices E_H.

In order to construct the matrix A, consider the following table:

G	$q_0(G)$	$q_1(G)$	$q_2(G)$	$q_3(G)$	$q_4(G)$
$\varnothing$	1	0	0	0	0
$-$	1/2	1/2	0	0	0
$\wedge$	1/4	1/2	1/4	0	0
$\triangle$	1/4	0	3/4	0	0
F_4	1/16	4/16	6/16	4/16	1/16
K_4^-	1/8	0	1/4	1/2	1/8

In this table, $-$ is a single edge, $\wedge$ is a path of length 2, $\triangle$ is a triangle, F_4 is any forest having 4 edges, and K_4^- is the diamond graph, obtained from K_4 by removing one edge. Suppose we are looking for a matrix A of the particularly simple form

$$A = c_0 E_0 + c_1 E_1 + c_2 E_2 + c_3 E_3 + c_4 E_4,$$

$$\lambda_G(A) = (-1)^{|G|}(c_0 q_0(G) + c_1 q_1(G) + c_2 q_2(G) + c_3 q_3(G) + c_4 q_4(G)).$$

The matrix A should satisfy $\lambda_{\varnothing}(A) = 1$, $\lambda_{-}(A) = \lambda_{\wedge}(A) = \lambda_{\triangle}(A) = -1/7$, and $\lambda_G(A) \geq -1/7$ for all graphs G. The first constraint already gives us $c_0 = 1$. The second constraint gives $c_1 = -5/7$ and $c_2 = -1/7$. Substituting c_0, c_1, c_2 into the inequalities corresponding to F_4 and K_4^- gives us a lower bound and an upper bound (respectively) on $4c_3 + c_4$. Both bounds coincide, implying that $4c_3 + c_4 = 3/7$. This prompts the following choice for A:

$$A = E_0 - \frac{5}{7}E_1 - \frac{1}{7}E_2 + \frac{3}{28}E_3. \tag{4.1}$$

Amazingly, this matrix A fits the bill.

Lemma 4.13. *The matrix A given by* (4.1) *is admissible, and satisfies the following properties:*

(a) $\lambda_{\varnothing}(A) = 1$.

(b) $\lambda_G(A) \geq -1/7$ *for all graphs G, with equality only for the following graphs: forests of one, two or four edges; triangles; diamonds.*

(c) If $\lambda_G(A) > -1/7$ *then in fact* $\lambda_G(A) \geq -1/8$.

The proof of this rather technical lemma appears in Section 4.3.2. Using this lemma, Hoffman's bound immediately implies that the measure of any odd-cycle-intersecting family is at most $1/8$ (the argument is sketched below). However, the matrix A is not enough to prove uniqueness, that is, that the unique maximal families are $\triangle$-stars. The problem is that $\lambda_G(A)$ is tight for graphs other than subgraphs of triangles. However, this is easy to fix by perturbing A by another matrix.

Lemma 4.14. *Let B be the admissible matrix*

$$B = \sum_F E_F - E_{\square}, \tag{4.2}$$

where the sum goes over all forests containing 4 *edges, and* $\square$ *is a cycle of length* 4.

(a) $\lambda_G(B) = 0$ *whenever G contains less than* 4 *edges.*

(b) $\lambda_F(B) = 1/16$ *whenever F is a forest containing* 4 *edges.*

(c) $\lambda_{K_4^-}(B) = 1/8$.

(d) $|\lambda_G(B)| \leq 1$ *for all graphs G.*

Proof. Item (a) is immediate since $G \cap K_{L,R}$ contains less than 4 edges.

For item (b), note first that $\lambda_F(E_\square) = 0$ since F contains no $\square$. Similarly $\lambda_F(E_{F'})$ for any forest $F' \neq F$. If L, R is a random partition of $[n]$, then each edge in F belongs to $F \cap K_{L,R}$ with probability 1/2, independently of the other edges (we show this formally in Section 4.3.2). Therefore $\lambda_F(E_F) = 1/16$.

For item (c), note first that $\lambda_{K_4^-}(E_F) = 0$. In order to calculate $\lambda_{K_4^-}(E_\square) = -q_\square(K_4^-)$, label K_4^- with $\{a, b, c, d\}$ so that the missing edge is (a, c). We have $K_4^- \cap K_{L,R} = \square$ exactly when a, c belong to the same side of the partition L, R and b, d to the other, which happens with probability 1/8. Hence $\lambda_{K_4^-}(E_\square) = -1/8$.

Item (d) follows from $0 \leq q_\square(G) \leq 1$ and $0 \leq \sum_F q_F(G) \leq 1$, which in turn follow from $q_\square, q_F$ being probabilities, and the different events considered in $\sum_F q_F$ being disjoint. $\square$

Now all we have to do is perturb A by an appropriate multiple of B.

Lemma 4.15. *Let $C = A + (2/119)B$, where A is given by* (4.1) *and B is given by* (4.2)*. The matrix C is admissible and satisfies the following properties:*

(a) $\lambda_\varnothing(C) = 1$.

(b) $\lambda_G(C) \geq -1/7$ *for all graphs G, with equality only for the following graphs: forests of one or two edges; triangles.*

(c) If $\lambda_G(C) > -1/7$ *then in fact* $\lambda_G(C) \geq -135/952$.

Proof. The first item is immediate. We now consider several cases. If G is a graph for which $\lambda_G(A) > -1/7$ then in fact $\lambda_G(A) \geq -1/8$, and so $|\lambda_G(B)| \leq 1$ implies that

$$\lambda_G(C) \geq -1/8 - 2/119 = -135/952.$$

If G is a forest on one or two edges or a triangle then $\lambda_G(B) = 0$ and so $\lambda_G(C) = -1/7$. If G is a forest on four edges then

$$\lambda_G(C) = -1/7 + (2/119)(1/16) = -135/952.$$

Finally, if G is a diamond then

$$\lambda_G(C) = -1/7 + (2/119)(1/8) > -135/952.$$

$\square$

Applying Hoffman's bound, we get an analog of Lemma 3.14 on page 39, and from that an analog of Theorem 3.16 on page 41, which is the main result of this section.

Lemma 4.16. *Let $\mathcal{F}$ be an odd-cycle-intersecting family of graphs with characteristic function $f = \mathbf{1}_{\mathcal{F}}$.*

Upper bound: $\mu(\mathcal{F}) \leq 1/8$.

Uniqueness: *If $\mu(\mathcal{F}) = 1/8$ then the Fourier expansion of f is supported on the first 4 levels, that is $\hat{f}(S) = 0$ for $|S| > 3$.*

Stability: *If $\mu(\mathcal{F}) \geq 1/8 - \epsilon$ then*

$$\sum_{|S|>3} \hat{f}^2(S) = O(\epsilon).$$

Proof. The proof is very similar to the proof of Lemma 3.14, using the matrix C given by Lemma 4.15. Since C is admissible, $\langle Cf, f\rangle = 0$, and so

$$\sum_G \lambda_G(C)\hat{f}^2(G) = 0.$$

Lemma 4.15 shows that $\lambda_{\min} = \min_G \lambda_G(C) = -1/7$ and $\lambda_2 = \min_{G:\lambda_G(C)>-1/7} \lambda_G(C) \geq -135/192$. Hoffman's bound implies that $\mu(\mathcal{F}) \leq -\lambda_{\min}/(\lambda_\varnothing(C) - \lambda_{\min}) = 1/8$, hence the upper bound. When $\mu(\mathcal{F}) = -1/7$, Hoffman's bound implies that $\hat{f}(G) \neq 0$ only for $G = \varnothing$ or whenever $\lambda_G(C) = -1/7$. Uniqueness follows from the fact that $\lambda_G(C) = -1/7$ only for graphs containing at most three edges. Finally, stability follows from Hoffman's bound since $\lambda_2 - \lambda_{\min}$ is bounded from below by a positive constant. □

Theorem 4.17. *Let $\mathcal{F}$ be an odd-cycle-intersecting family of graphs.*

Upper bound: $\mu(\mathcal{F}) \leq 1/8$.

Uniqueness: $\mu(\mathcal{F}) = 1/8$ *if and only if $\mathcal{F}$ is a $\triangle$-star.*

Stability: *If $\mu(\mathcal{F}) \geq 1/8 - \epsilon$ then $\mu(\mathcal{F}\Delta\mathcal{G}) = O(\epsilon)$ for some $\triangle$-star $\mathcal{G}$.*

Proof. The upper bound is already given by Lemma 4.16. For uniqueness, the argument in the proof of Theorem 3.16 shows that if $\mu(\mathcal{F}) = 1/8$ then $\mathcal{F}$ is a 3-star. Since $\mathcal{F}$ is odd-cycle-intersecting, it must be a $\triangle$-star.

For stability, suppose $\mu(\mathcal{F}) \geq 1/8 - \epsilon$. The stability part of Lemma 4.16 combined with Theorem 2.23 on page 24 shows that $\mathcal{F}$ is $D\epsilon$-close to a family $\mathcal{G}$ depending on M coordinates, where D depends on the hidden constant in Lemma 4.16 as well as the constant $C_{1/2,3}$ given by Theorem 2.23.

We start by showing that if ϵ is small enough, then $\mathcal{G}$ must be odd-cycle-intersecting. Suppose $\mathcal{G}$ is not odd-cycle-intersecting. Then there are two graphs $G_1, G_2 \in \mathcal{G}$ whose intersection contains no odd cycles. We can assume furthermore that $|G_1|, |G_2| \leq M$, say both are supported on the edge set X of size M. Let $\mathcal{F}_1 = \{G \subseteq \overline{X} : G \cup G_1 \in \mathcal{F}\}$, and define $\mathcal{F}_2$ similarly. If $G \in \mathcal{F}_1$ then $\overline{X} \setminus G \notin \mathcal{F}_2$, since $\mathcal{F}$ is odd-cycle-intersecting. Hence $\mu(\mathcal{F}_1) + \mu(\mathcal{F}_2) \leq 1$. Since all the corresponding graphs belong to $\mathcal{G}$, this shows that $\mu(\mathcal{F}\Delta\mathcal{G}) \geq 2^{-M}$, which is impossible if ϵ is small enough.

Next, suppose $\mathcal{G}$ is odd-cycle-intersecting. Then $\mu(\mathcal{G}) \leq 1/8$. Among all odd-cycle-intersecting families on M coordinates with $\mu(\mathcal{G}) < 1/8$, let the one with largest measure have measure $1/8 - \alpha$. Then $\mu(\mathcal{F}\Delta\mathcal{G}) \geq |\mu(\mathcal{F}) - \mu(\mathcal{G})| \geq \alpha - \epsilon$, and so $\alpha \leq (D+1)\epsilon$. If ϵ is small enough, this is impossible, and we conclude that $\mu(\mathcal{G}) = 1/8$. By uniqueness, $\mathcal{G}$ must be a $\triangle$-star.

Concluding, we have shown that for some ϵ_0, if $\epsilon \leq \epsilon_0$ then the family $\mathcal{G}$ given by Theorem 2.23 is a $\triangle$-star and $\mu(\mathcal{F}\Delta\mathcal{G}) \leq D\epsilon$. Otherwise, for any $\triangle$-star $\mathcal{G}$, $\mu(\mathcal{F}\Delta\mathcal{G}) \leq \epsilon_0^{-1}\epsilon$. In both cases $\mu(\mathcal{F}\Delta\mathcal{G}) \leq \max(D, \epsilon_0^{-1})\epsilon$. □

How did we know to choose A of the form $c_0E_0 + c_1E_1 + c_2E_2 + c_3E_3 + c_4E_4$? We were looking for a matrix A whose eigenvalues are easy to analyze. Analyzing the eigenvalues of a matrix of this form reduces to understanding the functions q_k for small k, which are amenable to analysis. Another possible choice with the same properties is

$$c_0E_0 + c_1E_1 + c_2E_2 + c_3E_3 + c_4'(B_\varnothing - E_0 - E_1 - E_2 - E_3).$$

It turns out that there is a matrix A' of this form which satisfies a version of Lemma 4.13. However, A' has more tight graphs (graphs for which the eigenvalue is $-1/7$) than A, and so we preferred to work with A.

4.3.1 Cut statistics

The present and the following section are devoted to the proof of Lemma 4.13. To this end, we develop a theory of *cut statistics* of graphs. Recall that the function $q_k(G)$ is the probability that $|G \cap K_{L,R}| = k$, where L, R is a random partition of the vertices of G. We think of L, R as a random cut in the graph, and $G \cap K_{L,R}$ is the set of edges crossing the cut. The *cut distribution* of a graph G is the distribution of $|G \cap K_{L,R}|$. It will be useful to represent the cut distribution by a generating function.

Definition 4.7. For a graph G, its *cut function* is

$$Q_G(x) = \sum_{k=0}^{\infty} q_k(G) x^k. \qquad \bigcirc$$

Since G is finite, the cut function is a polynomial. As an example,

$$Q_{-}(x) = \frac{1}{2} + \frac{1}{2}x,$$

since the probability that a random cut separates an edge is exactly $1/2$.

Generating functions are useful because if cuts in G_1 and G_2 are independent (say G_1 and G_2 are disjoint) then $Q_{G_1+G_2} = Q_{G_1} Q_{G_2}$.

Definition 4.8. Let G be a graph. The vertex set of G is denoted $V(G)$, and the number of vertices is denoted $v(G) = |V(G)|$. The edge set of G is denoted $E(G)$, and the number of edges is denoted $e(G) = |E(G)|$. For $U \subseteq V(G)$, $G[U]$ is the graph induced by U. $\bigcirc$

Lemma 4.18. *Suppose a graph G is composed of two subgraphs G_1, G_2 having disjoint vertex sets. Then $Q_G = Q_{G_1} Q_{G_2}$.*

More generally, if the connected component of G are $G_1, \ldots, G_r$, then $Q_G = Q_{G_1} \cdots Q_{G_r}$.

Proof. A random partition of $V[G]$ can be generated by joining together two independent random partitions of $V[G_1]$ and $V[G_2]$. Therefore

$$q_k(G) = \sum_{l=0}^{k} q_l(G_1) q_{k-l}(G_2).$$

This directly implies the formula $Q_G = Q_{G_1} Q_{G_2}$. $\square$

This lemma reduces understanding cut distributions to the case of connected graphs. The following lemmas reduce it further to the case of biconnected graphs (graphs in which removing any vertex does not disconnect the graph).

Lemma 4.19. *Let G be a connected graph and v a vertex of G. Suppose that removing v disconnects the graph into components $G_1, \ldots, G_r$. Let $H_i = G[V(G_i) \cup \{v\}]$ be the graph resulting from reintroducing v into G_i, preserving the original edges. Then $Q_G = Q_{H_1} \cdots Q_{H_r}$.*

Proof. Let $q_k^v(G)$ be $\Pr[|G \cap K_{L\cup\{v\},R}| = k]$, where L, R is a random partition of $V(G) \setminus \{v\}$. It is not hard to see that $q_k^v(G) = q_k(G)$. The same argument employed in the proof of Lemma 4.18 now shows that

$$q_k(G) = q_k^v(G) = q_k^v(H_1) \cdots q_k^v(H_r) = q_k(H_1) \cdots q_k(H_r). \qquad \square$$

Definition 4.9. Let G be a connected graph. A *bridge* is an edge of G whose removal disconnects the graph. A *biconnected component* of G is a maximal bridge-less biconnected subgraph of G. A *block* of G is either a bridge or a biconnected component.

The *split* of a graph G is obtained by replacing each block of each connected component of G with a disjoint copy. So each block in G becomes a connected component in the split of G. ○

Lemma 4.20. *Let G_s be the split of G. Then $Q_{G_s} = Q_G$.*

Proof. Using Lemma 4.18, we can assume that G is connected. If G is biconnected, we are done. Otherwise, choose a cut vertex v of G (a vertex whose removal disconnects the graph). Applying Lemma 4.19, G has the same cut function as the graph resulting from taking each connected component of G and attaching to it its own copy of v. Keep applying this process until all connected components are single edges or biconnected. The result is the split of G. $\square$

We need one more simple observation, and then we can state a formula for the first few coefficients of the cut function of a graph (recall we are only interested in $q_0, \ldots, q_3$).

Lemma 4.21. *Let G be a bridge-less graph. Then $q_1(G) = 0$.*

Proof. Suppose there is a partition L, R such that $G \cap K_{L,R} = \{e\}$. Then e is a bridge. $\square$

Corollary 4.22. *Let G be a graph with m bridges, and let H be the union of disjoint copies of its biconnected components. Suppose*

$$Q_H(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots.$$

Then $a_1 = 0$ and

$$Q_G(x) = \frac{1}{2^m}\left(a_0 + m a_0 x + \left(\binom{m}{2} a_0 + a_2\right) x^2 + \left(\binom{m}{3} a_0 + m a_2 + a_3\right) x^3\right) + \cdots. \tag{4.3}$$

Proof. The lemma shows that $a_1 = 0$. The formula follows from Lemma 4.20 using the formula $Q_-(x) = 1/2 + (1/2)x$ by expanding the product:

$$\begin{aligned} Q_G(x) &= \left(\frac{1}{2} + \frac{1}{2}x\right)^m \left(a_0 + a_2 x^2 + a_3 x^3 + \cdots\right) \\ &= \frac{1}{2^m}\left(1 + mx + \binom{m}{2}x^2 + \binom{m}{3}x^3 + \cdots\right)\left(a_0 + a_2 x + a_3 x^3 + \cdots\right) \\ &= \frac{1}{2^m}\left(a_0 + m a_0 x + \left(\binom{m}{2} a_0 + a_2\right) x^2 + \left(\binom{m}{3} a_0 + m a_2 + a_3\right) x^3\right) + \cdots. \end{aligned}$$ □

4.3.2 Proof of Lemma 4.13

Armed with the theory of cut statistics developed in the preceding section, we are ready to prove Lemma 4.13. Before starting the proof proper, we need two auxiliary results.

Lemma 4.23. *Let G be a graph.*

(a) If G has N connected components then $q_0(G) = 2^{N - v(G)}$. (Recall $v(G)$ is the number of vertices in G.)

(b) If G has exactly m bridges then $q_1(G) = m q_0(G)$.

(c) If G has a vertex of odd degree then $q_k(G) \le 1/2$ for all k.

(d) For any odd k, $q_k(G) \le 1/2$.

(e) Always $q_2(G) \le 3/4$.

Proof. For item (a), let the connected components of G be $G_1, \ldots, G_N$. If $G \cap K_{L,R} = \varnothing$ then each of G_i lies entirely in L or in R. The probability of this is

$$q_0(G) = \prod_{i=1}^{N} \frac{2}{2^{v(G_i)}} = \frac{2^N}{2^{v(G)}}.$$

Item (b) follows directly from (4.3).

For item (c), suppose G is a graph with a vertex v of odd degree. Let L', R' be a partition of $V(G) \setminus \{v\}$. The set $(G \cap K_{L' \cup \{v\}, R'}) \setminus (G \cap K_{L',R'})$ contains all neighbors of v belonging to R'. Similarly, the set $(G \cap K_{L', R' \cup \{v\}}) \setminus (G \cap K_{L',R'})$ contains all neighbors of v belonging to L'. Therefore

$$(|G \cap K_{L' \cup \{v\}, R'}| - |G \cap K_{L',R'}|) + (|G \cap K_{L', R' \cup \{v\}}| - |G \cap K_{L',R'}|) = \deg(v).$$

Since $\deg(v)$ is odd, we conclude that $|G \cap K_{L' \cup \{v\}, R'}| \neq |G \cap K_{L', R' \cup \{v\}}|$. Hence the probability that $|G \cap K_{L,R}| = k$, conditioned on $L' \subseteq L$ and $R' \subseteq R$, is at most $1/2$. Averaging over all L', R', the item follows.

For item (d), in view of item (c), we can assume that all vertices have even degree. This implies that G can be partitioned into cycles (since a connected graph with even degrees has an Eulerian tour). Each cut of G cuts either 0 or 2 edges of each cycle, and therefore an even number of edges overall, showing that $q_k(G) = 0$ for odd k.

For item (e), note that the average number of edges cut in a random cut is $e(G)/2$ (recall that $e(G)$ is the number of edges in G), and so

$$\frac{e(G)}{2} = \sum_{k=0}^{e(G)} k q_k(G) < 2q_2(G) + e(G)(1 - q_2(G)) = e(G) + (2 - e(G)) q_2(G).$$

The inequality is strict since $q_0(G) > 0$. If $e(G) = 2$ then

$$Q_G(x) = \left(\frac{1}{2} + \frac{1}{2}x\right)^2 = \frac{1}{4} + \frac{1}{2}x + \frac{1}{4}x^2,$$

so we can assume that $e(G) > 2$. This implies that

$$q_2(G) < \frac{e(G)/2}{e(G) - 2} = \frac{(e(G) - 2)/2 + 1}{e(G) - 2} = \frac{1}{2} + \frac{1}{e(G) - 2}.$$

Therefore $q_2(G) < 3/4$ whenever $e(G) \geq 6$. So we can assume that $e(G) \leq 5$.

Let G_s be the split of G, which has the same cut function as G by Lemma 4.20. If G has any bridges then G_s has vertices of degree 1, and so $q_2(G) \leq 1/2$ by part (c). Otherwise, since each block of G contains at least 3 edges, G must be biconnected. So G is one of C_3, C_4, C_5, K_4^- (here C_l is the cycle of length l). One can check that $q_2(C_3) = q_2(C_4) = 3/4$, $q_2(C_5) = 5/8$ and $q_2(K_4^-) = 1/4$. □

The following lemma focuses exclusively on q_0.

Lemma 4.24. *Let G be a graph with m bridges, and let H be the union of its biconnected components.*

(a) We have $q_0(\varnothing) = 1$, $q_0(-) = 1/2$, and $q_0(G) \le 1/4$ for all other graphs. (Here $-$ is a single edge.)

(b) If $m = 0$ and $e(G)$ is odd then either $q_0(G) \le 1/16$ or G is a triangle or a diamond.

(c) If H is non-empty then $q_0(H) \le 1/4$.

Proof. For item (a), if G is connected and $|G| \ge 2$ then $v(G) \ge 3$ and so Lemma 4.23(a) shows that $q_0(G) = 2^{1-v(G)} \le 1/4$. If G has $N \ge 2$ connected components, then since every connected component contains at least two vertices, the same item shows that $q_0(G) = 2^{N-v(G)} \le 2^{-N} \le 1/4$.

For item (b), notice that since $m = 0$, every connected component of G contains at least three vertices. Lemma 4.23(a) implies that $q_0(G) \le (1/4)^N$, where N is the number of connected components. If $N \ge 2$ then $q_0(G) \le 1/16$, so we can assume G is connected. Lemma 4.23(a) again implies that $q_0(G) \le 2^{1-v(G)}$, hence $q_0(G) \le 1/16$ if $v(G) \ge 5$. The remaining case is that G is a connected bridge-less graph on at most 4 vertices. Since $e(G)$ is odd, G is either a triangle or a diamond.

Item (c) follows directly from item (a). □

We are finally ready to prove Lemma 4.13.

Lemma 4.13. *The matrix A given by (4.1) is admissible, and satisfies the following properties:*

(a) $\lambda_\varnothing(A) = 1$.

(b) $\lambda_G(A) \ge -1/7$ for all graphs G, with equality only for the following graphs: forests of one, two or four edges; triangles; diamonds.

(c) If $\lambda_G(A) > -1/7$ then in fact $\lambda_G(A) \ge -1/8$.

The idea of the proof is to consider the matrix

$$A = E_0 - \frac{5}{7}E_1 - \frac{1}{7}E_2 + \frac{3}{28}E_3 \tag{4.1}$$

which has been engineered to satisfy the lemma for small graphs. For large graphs, all q_k are small, and so the eigenvalues tend to zero. It remains to consider what happens for graphs in the medium range. Because of the $(-1)^{|G|}$ factor (we remind the reader that $|G| = e(G)$), we consider graphs with an even and an odd number of edges separately. Formula (4.3) shows that the q_k decay fast if G contains many bridges, so the proof will consist of a case analysis, where the cases correspond to different numbers of bridges.

It is possible to considerably reduce the number of cases by computing the eigenvalues for all graphs with a small number of edges. For graphs with many edges, it is relatively easy to show that the eigenvalues are close enough to zero. However, we opted to present a completely human-verifiable proof. This will also come in handy later on, when we generalize the framework to $p < 1/2$.

Proof. In view of Lemma 4.12, the eigenvalues of A satisfy the formula

$$\lambda_G(A) = (-1)^{|G|}\left(q_0(G) - \frac{5}{7}q_1(G) - \frac{1}{7}q_2(G) + \frac{3}{28}q_3(G)\right).$$

This formula already shows that $\lambda_\varnothing(A) = 1$. It will be less confusing to consider instead of $\lambda_G(A)$ the function

$$f(G) = q_0(G) - \frac{5}{7}q_1(G) - \frac{1}{7}q_2(G) + \frac{3}{28}q_3(G).$$

We now split the proof into two cases: $|G|$ is odd and $|G|$ is even.

Graphs with an odd number of edges. Suppose $|G|$ is odd. We show that $f(G) = 1/7$ if G is a single edge, a triangle or a diamond, and $f(G) \le 1/8$ otherwise.

Let m be the number of bridges in G. Lemma 4.23(b) shows that $q_1(G) = mq_0(G)$, and so

$$f(G) = (1 - \frac{5}{7}m)q_0(G) - \frac{1}{7}q_2(G) + \frac{3}{28}q_3(G).$$

We will use the bound $q_3(G) \le 1/2$ given by Lemma 4.23(d).

When $m = 0$,

$$f(G) = q_0(G) - \frac{1}{7}q_2(G) + \frac{3}{28}q_3(G).$$

If $q_0(G) \le 1/16$ then using $q_3(G) \le 1/2$,

$$f(G) \le \frac{1}{16} + \frac{3}{28}\cdot\frac{1}{2} = \frac{13}{112} < \frac{1}{8}.$$

If $q_0(G) > 1/16$ then Lemma 4.24(b) shows that G is either a triangle or a diamond. In both cases, we can explicitly compute $f(\triangle) = f(K_4^-) = 1/7$.

When $m = 1$,

$$f(G) = \frac{2}{7}q_0(G) - \frac{1}{7}q_2(G) + \frac{3}{28}q_3(G).$$

If $G = -$ then $q_0(G) = 1/2$ and so $f(G) = 1/7$. Otherwise, Lemma 4.24(a) shows that $q_0(G) \le 1/4$. Therefore using $q_3(G) \le 1/2$,

$$f(G) \le \frac{2}{7} \cdot \frac{1}{4} + \frac{3}{28} \cdot 12 = 1/8.$$

When $m \ge 2$,

$$f(G) \le -\frac{3}{7}q_0(G) - \frac{1}{7}q_2(G) + \frac{3}{28}q_3(G) \le \frac{3}{56} < \frac{1}{8},$$

using $q_3(G) \le 1/2$.

Graphs with an even number of edges. Suppose $|G|$ is even. We show that $f(G) = -1/7$ if G is a forest on two or four edges, and $f(G) \ge -3/28 > -1/8$ otherwise.

Let m be the number of bridges, let H be the union of all biconnected components of G, and let $a_k = q_k(H)$. Corollary 4.22 shows that

$$\begin{aligned} f(G) &= \frac{1}{2^m}\left(a_0 - \frac{5}{7}ma_0 - \frac{1}{7}\left(\binom{m}{2}a_0 + a_2\right) + \frac{3}{28}\left(\binom{m}{3}a_0 + ma_2 + a_3\right)\right) \\ &= \frac{1}{2^m}\left(\left(1 - \frac{5}{7}m - \frac{1}{7}\binom{m}{2} + \frac{3}{28}\binom{m}{3}\right)a_0 + \left(-\frac{1}{7} + \frac{3}{28}m\right)a_2 + \frac{3}{28}a_3\right). \end{aligned}$$

When $m = 0$,

$$f(G) = a_0 - \frac{1}{7}a_2 + \frac{3}{28}a_3.$$

Lemma 4.23(e) shows that $a_2 \le 3/4$, and so $f(G) \ge -(1/7)(3/4) = -3/28$.

When $m = 1$,

$$f(G) = \frac{1}{7}a_0 - \frac{1}{56}a_2 + \frac{3}{56}a_3.$$

Again, $a_2 \le 3/4$ implies $f(G) \ge -(1/56)(3/4) = -3/223 > -3/28$.

When $m \ge 2$, the coefficients in front of a_2 and a_3 are positive, and so

$$f(G) \ge \frac{1}{2^m}\left(1 - \frac{5}{7}m - \frac{1}{7}\binom{m}{2} + \frac{3}{28}\binom{m}{3}\right)a_0.$$

Denote the coefficient in front of $f(G)$ by $r(m)$. Lemma 4.24(c) shows that either $H = \varnothing$ or $a_0 \leq 1/4$; the former case can happen only when m is even, since $|G|$ is even. We list some values of $r(m)$:

m	2	3	4	5	6	7
$-r(m)$	$\frac{1}{7}$	$\frac{41}{224}$	$\frac{1}{7}$	$\frac{41}{448}$	$\frac{23}{448}$	$\frac{13}{512}$

Recall $2^m r(m)$ is a third degree polynomial. It is not hard to check that the polynomial is increasing in the range $m \geq (7 + \sqrt{151})/3 \approx 6.4$. Therefore for $m \geq 7$, $r(m) \geq 2^{7-m} r(7)$. If $r(m)$ is negative then this implies that $r(m) \geq r(7)$, and otherwise $r(m) \geq r(7)$ trivially.

The table shows that for $m \geq 5$, $f(G) \geq -41/448 > -3/28$. If $H \neq \varnothing$, then $a_0 \leq 1/4$ and so the table shows that for $m \geq 2$, $f(G) \geq r(m)/4 \geq -41/896 > -3/28$. It remains to consider the case that $m \in \{2,3,4\}$ and $H = \varnothing$. Since $|G|$ is even, m must be even, and so G is a forest with two or four edges. Direct calculation shows that in both cases, $f(G) = 1/7$. $\square$

4.4 Agreeing families

The matrices constructed in Section 4.3 are admissible not only for odd-cycle-intersecting families but also for odd-cycle-agreeing families. The reason is that

$$\chi_S^{[1]\prime} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \chi_T^{[1]} = 0$$

not only when $S = T = \{1\}$, but also when $S = T = \varnothing$.

Definition 4.10. Let $n \geq 1$ be an integer. We say that a matrix A is *odd-cycle-agreeing-admissible* (*agreeing-admissible* for short) if it satisfies the following two properties:

Intersection property If G, H are graphs whose agreement $G \nabla H$ contains an odd cycle then $\mathbf{1}_G' B \mathbf{1}_H = 0$.

Eigenvector property The eigenvectors of $\mathcal{F}$ are the Fourier characters χ_S.

If A is agreeing-admissible then we use $\lambda_G(A)$ to denote the eigenvalue corresponding to χ_G. $\bigcirc$

Lemma 4.25. *Let J be a set of edges. Define*

$$B_{J,i} = \begin{cases} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, & \text{if } i \notin J, \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, & \text{if } i \in J, \end{cases}$$

$$B_J = \bigotimes_{i=1}^{n} B_{J,i}.$$

If J is bipartite then the matrix B_J is agreeing-admissible, and the eigenvalue corresponding to χ_G is

$$\lambda_G(B_J) = (-1)^{|S \setminus J|}.$$

Furthermore, the vector space of all agreeing-admissible matrices is spanned by B_J for all bipartite J.

Proof. The proof uses the observation that for $S, T \subseteq [1]$,

$$\chi_S^{[1]\prime} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \chi_T^{[1]} = 0$$

whenever $S \nabla T \neq \emptyset$. In all other respects, the proof is identical to that of Lemma 4.11. □

We conclude that Lemma 4.16 is in fact true for odd-cycle-agreeing families as well.

Lemma 4.26. *Let $\mathcal{F}$ be an odd-cycle-agreeing family of graphs with characteristic function $f = \mathbf{1}_{\mathcal{F}}$.*

Upper bound: $\mu(\mathcal{F}) \leq 1/8$.

Uniqueness: *If $\mu(\mathcal{F}) = 1/8$ then the Fourier expansion of f is supported on the first 4 levels, that is $\hat{f}(S) = 0$ for $|S| > 3$.*

Stability: *If $\mu(\mathcal{F}) \geq 1/8 - \epsilon$ then*

$$\sum_{|S|>3} \hat{f}^2(S) = O(\epsilon).$$

Proof. Follow the proof of Lemma 4.16, replacing Lemma 4.11 with Lemma 4.25. □

This already gives us the correct upper bound. However, uniqueness doesn't follow since our argument for uniqueness relies heavily on the fact that maximal families are monotone. Our proof of stability relies on uniqueness, so this also does not follow.

In order to amend this situation, we take inspiration from the following result of Chung, Frankl, Graham and Shearer [10], which shows that upper bounds for intersection problems *always* imply upper bounds for agreement problems. The proof uses the classical technique of shifting.

Definition 4.11. Let X be a finite set, and let $\mathcal{H}$ be a family of subsets of X. A family of subsets of X is *$\mathcal{H}$-intersecting* if the intersection of any two members of the family contains some $H \in \mathcal{H}$. A family of subsets of X is *$\mathcal{H}$-agreeing* if the agreement of any two members of the family contains some $H \in \mathcal{H}$. ○

Theorem 4.27. *Let X be a finite set, and let $\mathcal{H}$ be a family of subsets of X. Then the maximal size of an $\mathcal{H}$-intersecting family is equal to the maximal size of an $\mathcal{H}$-agreeing family.*

Proof. Every $\mathcal{H}$-intersecting family is also $\mathcal{H}$-agreeing, and so it is enough to show that if $\mathcal{F}$ is an $\mathcal{H}$-agreeing family then there is an $\mathcal{H}$-intersecting family of size $|\mathcal{F}|$. We do that by applying several cardinality-preserving operations on $\mathcal{F}$ which will make it $\mathcal{H}$-intersecting.

For $i \in X$ and a family $\mathcal{G}$, the monotonization operator $C_i(\mathcal{G})$ is defined as follows. Partition 2^X into pairs $A, A \cup \{i\}$. Whenever $\mathcal{G} \cap \{A, A \cup \{i\}\} = \{A\}$, replace A with $A \cup \{i\}$ in $C_i(\mathcal{G})$.

Clearly $|C_i(\mathcal{G})| = |\mathcal{G}|$, and $C_i(\mathcal{G})$ is *i-monotone*: if $A \in C_i(\mathcal{G})$ then $A \cup \{i\} \in C_i(\mathcal{G})$. Moreover, if $\mathcal{G}$ is j-monotone then so is $C_i(\mathcal{G})$. Indeed, consider any $A \in C_i(\mathcal{G})$. If $i \notin A$ then $A, A \cup \{i\} \in \mathcal{G}$ and so $A \cup \{j\}, A \cup \{i,j\} \in \mathcal{G}$. This shows that $A \cup \{j\} \in C_i(\mathcal{G})$. If $i \in A$ and $A \in \mathcal{G}$ then $A \cup \{j\} \in \mathcal{G}$ and so $A \cup \{j\} \in C_i(\mathcal{G})$. Finally, if $i \in A$ and $A \notin \mathcal{G}$ then $A \setminus \{i\}, A \setminus \{i\} \cup \{j\} \in \mathcal{G}$. Therefore $A \cup \{j\} \in C_i(\mathcal{G})$.

The crucial property is that if $\mathcal{G}$ is $\mathcal{H}$-agreeing then so is $C_i(\mathcal{G})$. Indeed, let $A, B \in C_i(\mathcal{G})$. If $A, B \in \mathcal{G}$ then A, B are certainly $\mathcal{H}$-agreeing. If $A \setminus \{i\}, B \setminus \{i\} \in \mathcal{G}$ then $A \nabla B = (A \setminus \{i\}) \nabla (B \setminus \{i\})$, and again A, B are $\mathcal{H}$-agreeing. The remaining case is when $i \in A$ and $A \setminus \{i\}, B \in \mathcal{G}$. If $i \in B$ then $A \nabla B \supset (A \setminus \{i\}) \nabla B$, and we're again done. If $i \notin B$ then necessarily $B \cup \{i\} \in \mathcal{G}$, since otherwise we would have replaced B with $B \cup \{i\}$, and so $A \nabla B = (A \setminus \{i\}) \nabla (B \setminus \{i\})$

again shows that A, B are $\mathcal{H}$-agreeing.

Let $\mathcal{F}'$ result from applying the operators C_i on $\mathcal{F}$ in sequence for all $i \in X$. The family $\mathcal{F}'$ has the same size as $\mathcal{F}$, it is $\mathcal{H}$-agreeing, and it is monotone. These properties imply that for any $A, B \in \mathcal{F}'$, $(A \cup \overline{B}) \nabla B = A \cap B$ contains some $H \in \mathcal{H}$, and so $\mathcal{F}'$ is $\mathcal{H}$-intersecting. $\square$

Using the same monotonization operations C_i, we are able to prove a uniqueness counterpart of Theorem 4.27.

Definition 4.12. Let X be a finite set, and let $\mathcal{H}$ be a family of subsets of X. An $\mathcal{H}$*-star* is an H-star for some $H \in \mathcal{H}$. An $\mathcal{H}$*-semistar* is an H-semistar for some $H \in \mathcal{H}$, which is a family of the form $\{A \subseteq X : A \cap H = J\}$ for some $J \subseteq H$. $\bigcirc$

Theorem 4.28. *Let X be a finite set, and let $\mathcal{H}$ be a family of subsets of X, and $\mathcal{I} \subseteq \mathcal{H}$ be a subfamily of $\mathcal{H}$. Suppose that all maximal $\mathcal{H}$-intersecting families are $\mathcal{I}$-stars, and that for all $I \in \mathcal{I}$, $x \in I$ and $y \notin I$, neither $I \setminus \{x\}$ nor $I \setminus \{x\} \cup \{y\}$ contain any $H \in \mathcal{H}$. Then all maximal $\mathcal{H}$-agreeing families are $\mathcal{I}$-semistars.*

Proof. Let C_i be the monotonization operators defined in the proof of Theorem 4.27. Suppose $\mathcal{F}$ is a maximal $\mathcal{H}$-agreeing family. The proof of that theorem shows that if we apply the operators C_i to $\mathcal{F}$ in sequence for all $i \in X$, then we get a maximal $\mathcal{H}$-intersecting family, which is an $\mathcal{I}$-star by assumption. Therefore the proof will be complete if we show that whenever $\mathcal{G}$ is $\mathcal{H}$-agreeing and $C_i(\mathcal{G})$ is an $\mathcal{I}$-semistar, then $\mathcal{G}$ is an $\mathcal{I}$-semistar as well.

By possibly complementing some of the coordinates, we can further assume that $C_i(\mathcal{G})$ is an $\mathcal{I}$-star, say it is an I-star, where $I \in \mathcal{I}$. If $i \notin I$ then $\mathcal{G} = C_i(\mathcal{G})$, and so $\mathcal{G}$ is also an I-star. Otherwise, we know that for each $A \in C_i(\mathcal{G})$, either $A \in \mathcal{G}$ or $A \setminus \{i\} \in \mathcal{G}$, but not both. Define

$$\mathcal{G}_+ = \{A \subseteq X \setminus I : A \cup I \in \mathcal{G}\},$$

$$\mathcal{G}_- = \{A \subseteq X \setminus I : A \cup (I \setminus \{i\}) \in \mathcal{G}\}.$$

We will show that whenever $|A \nabla B| \leq 1$, either $A, B \in \mathcal{G}_+$ or $A, B \in \mathcal{G}_-$. If $|A \Delta B| = 1$ then $|A \nabla \overline{A}| = 0$ while $|\overline{A} \nabla B| = 1$, showing that either $A, B \in \mathcal{G}_+$ or $A, B \in \mathcal{G}_-$. It follows that either $\mathcal{G}_+ = \varnothing$ or $\mathcal{G}_- = \varnothing$, and in both cases $\mathcal{G}$ is an H-semistar.

Suppose to the contrary that $|A \nabla B| \leq 1$ and $A \in \mathcal{G}_+$, $B \in \mathcal{G}_-$. Since $A \cup I, B \cup (I \setminus \{i\}) \in \mathcal{G}$, their agreement must contain some $H \in \mathcal{H}$. However,

$$D \triangleq (A \cup I) \nabla (B \cup I \setminus \{i\}) = (A \nabla B) \cup (I \setminus \{i\}).$$

By assumption, D does not contain any $H \in \mathcal{H}$. This contradiction shows that either $A, B \in \mathcal{G}_+$ or $A, B \in \mathcal{G}_-$. □

In our case, $\mathcal{H}$ is the family of all odd cycles and $\mathcal{I}$ is the family of all triangles. The condition in the theorem then states that the only way to introduce a cycle into a path of length 2 with a single edge is to complete the triangle. This observation allows us to prove Theorem 4.1.

Lemma 4.29. *Let T be a triangle, $x \in T$ and $y \notin T$. Then neither $T \setminus \{x\}$ nor $T \setminus \{x\} \cup \{y\}$ contains any cycle.*

Proof. Obvious. □

Theorem 4.1. *Let $\mathcal{F}$ be an odd-cycle-agreeing family of graphs on n vertices.*

Upper bound: $\mu(\mathcal{F}) \leq 1/8$.

Uniqueness: $\mu(\mathcal{F}) = 1/8$ *if and only if $\mathcal{F}$ is a $\triangle$-semistar.*

Stability: *If $\mu(\mathcal{F}) \geq 1/8 - \epsilon$ then there is a $\triangle$-semistar $\mathcal{G}$ such that $\mu(\mathcal{F} \Delta \mathcal{G}) = O(\epsilon)$.*

Proof. Lemma 4.26 already shows the upper bound. Alternatively, the upper bound follows from Theorem 4.17 via Theorem 4.27. Uniqueness follows from Theorem 4.17 via Theorem 4.28.

For stability, we use an argument very similar to the proof of Theorem 4.17. The stability part of Lemma 4.26 together with Theorem 2.23 shows that there is a family $\mathcal{G}$ depending on $M_{1/2,3}$ coordinates which is $O(\epsilon)$-close to $\mathcal{F}$. As in the proof of Theorem 4.17 (replacing intersection with agreement), we argue that if ϵ is small enough, then $\mathcal{G}$ must be odd-cycle-agreeing. Since $\mu(\mathcal{F} \Delta \mathcal{G}) \geq |\mu(\mathcal{F}) - \mu(\mathcal{G})|$ and there are only finitely many possible families $\mathcal{G}$, if ϵ is small enough then $\mu(\mathcal{G}) = 1/8$. Then $\mathcal{G}$ is a $\triangle$-semistar by uniqueness. □

4.5 Extension to hypergraphs

In this section we generalize the work done in the previous sections to the hypergraph setting, proving Theorem 4.2. For the rest of this section, fix the size n of the ground set. It will be easier to consider the hypergraphs as collections of non-zero vectors in $\mathbb{Z}_2^n$ rather than non-empty subsets of $[n]$. Under this view, a hypergraph is a subset of $\{x \in \mathbb{Z}_2^n : x \neq \mathbf{0}\}$ (where $\mathbf{0}$ is the zero vector), and an odd circuit consists of an odd number of vectors $x_1, \ldots, x_{2k+1}$ summing to $\mathbf{0}$.

Our goal is to show that every odd-circuit-agreeing family contains at most 1/8 of the hypergraphs. In view of Section 4.4, we can focus on odd-circuit-intersecting families. If hypergraphs were allowed to contain the zero vector, then the $\mathbf{0}$-star is odd-circuit-intersecting and contains at most 1/2 of the graphs, which is one reason to outlaw the zero vector. Even if we insist that the odd circuit have size at least 3, the zero vector causes problems. First, it is now important that the vectors in a circuit are all different: otherwise the $\{\mathbf{0}, x\}$-star, which contains 1/4 of the graphs, is odd-circuit-intersecting for all $x \neq \mathbf{0}$, since $\mathbf{0} + x + x = \mathbf{0}$. Even if we disallow that, the zero vector can turn an odd circuit into an even circuit. While we believe that the theorem should remain true even for circuit-agreeing families, this seems much harder to prove. For all these reasons, we do not allow our hypergraphs to contain the zero vector.

We will mostly retrace our steps in Section 4.3. We start with some basic definitions.

Definition 4.13. We say that a matrix A is *odd-circuit-admissible* (*admissible* for short) if it satisfies the following two properties:

Intersection property If G, H are hypergraphs whose intersection contains an odd circuit then $\mathbf{1}_G' A \mathbf{1}_H = 0$.

Eigenvector property The eigenvectors of $\mathcal{F}$ are the Fourier characters χ_S.

If A is admissible then we use $\lambda_G(A)$ to denote the eigenvalue corresponding to χ_G. ○

Lemma 4.30. *Let J be a set of edges. Define*

$$B_{J,i} = \begin{cases} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, & \text{if } i \notin J, \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, & \text{if } i \in J, \end{cases}$$

$$B_J = \bigotimes_{i=1}^{n} B_{J,i}.$$

If J contains no odd circuits then the matrix B_J is admissible, and the eigenvalue corresponding to χ_G is

$$\lambda_G(B_J) = (-1)^{|S \setminus J|}.$$

Furthermore, the vector space of all admissible matrices is spanned by B_J for all bipartite J.

Proof. The proof is the same as the proof of Lemma 4.11. □

In the case of graphs, we made an essential use of randomly generated bipartite graphs, which are graphs that we know do not contain odd cycles. The counterpart in the hypergraph setting is hyperplanes.

Definition 4.14. Let $y \in \mathbb{Z}_2^n$ be an arbitrary vector. The *hyperplane* defined by y is

$$P_y = \{x \in \mathbb{Z}_2^n : \langle\langle x, y \rangle\rangle = 1\}.$$

Here we understand $\langle\langle x, y \rangle\rangle$ as a number in $\mathbb{Z}_2$. ○

Lemma 4.31. *Hyperplanes contain no odd circuits.*

Proof. Let P_y be a hyper plane and $x_1, \ldots, x_{2k+1} \in P_y$. Then

$$\langle\langle y, x_1 + \cdots + x_{2k+1} \rangle\rangle = \langle\langle y, x_1 \rangle\rangle + \cdots + \langle\langle y, x_{2k+1} \rangle\rangle = 1,$$

and so $x_1 + \cdots + x_{2k+1} \neq \mathbf{0}$. □

We can think of every graph G as a hypergraph in which each vector has Hamming weight 2. The complete bipartite graph $K_{L,R}$ corresponds naturally to a vector $y \in \mathbb{Z}_2^n$ in which $y_i = 0$ whenever $i \in L$, $y_i = 1$ whenever $i \in R$. Under this correspondence, $K_{L,R}$ is the subset of P_y consisting of vectors of Hamming weight 2. This prompts the following generalization of Lemma 4.12.

Definition 4.15. Two hypergraphs H_1, H_2 are *isomorphic* if there is an invertible linear operator L on $\mathbb{Z}_2^n$ that maps H_1 to H_2.

For a hypergraph H, define a function q_H on hypergraphs by

$$q_H(G) = \Pr_{y \in \mathbb{Z}_2^n}[G \cap P_y \text{ is isomorphic to } H].$$

Similarly, for an integer k, define

$$q_k(G) = \Pr_{y \in \mathbb{Z}_2^n}[|G \cap P_y| = k],$$

The functions q_H, q_k are invariant under isomorphism. ○

We reuse the same notation q_k, q_H since the two functions have the same values in both contexts when the input G is a graph.

Lemma 4.32. *For every hypergraph H there is an admissible matrix E_H such that $\lambda_G(E_H) = (-1)^{|G|} q_H(G)$.*

For every integer k there is an admissible matrix E_k such that $\lambda_G(E_k) = (-1)^{|G|} q_k(G)$.

Proof. The proof is the same as the proof of Lemma 4.12. □

From this point on, the proof is substantially the same as in the graphical case. Before embarking on the proof, we give suggestive names to some hypergraphs.

Definition 4.16. An *edge* is a non-zero vector.

A *forest* is a hypergraph in which all vectors are linearly independent. The (unique up to isomorphism) forest containing n edges is denoted F_n. We sometimes call it an n-forest.

A *cycle* is a hypergraph of the form $\{x_1, \ldots, x_n, x_1 + \cdots + x_n\}$ such that $x_1, \ldots, x_n$ are linearly independent. A cycle of size n is denoted C_n (the above cycle has size $n+1$). The cycle $\triangle = C_3$ is also known as a *triangle*, and $\square = C_4$ is also known as a *square*.

A *diamond* is a hypergraph of the form $\{x, y, z, x + y, x + z\}$, where x, y, z are linearly independent. We denote this hypergraph K_4^-. ○

All these definitions agree with their graphical counterparts when the hypergraph in question is indeed a graph. Also, note that all forests of a fixed size are now isomorphic to each other.

We construct the matrix A exactly as before (physically, though, the matrices are different):

$$A = E_0 - \frac{5}{7}E_1 - \frac{1}{7}E_2 + \frac{3}{28}E_3 \tag{4.1}$$

The proof of Lemma 4.13 can be modified to yield the following.

Lemma 4.33. *The matrix A given by* (4.1) *is admissible, and satisfies the following properties:*

(a) $\lambda_\varnothing(A) = 1$.

(b) $\lambda_H(A) \geq -1/7$ *for all hypergraphs H, with equality only for the following hypergraphs: forests of size one, two, or four; triangles; diamonds.*

(c) If $\lambda_H(A) > -1/7$ *then in fact* $\lambda_H(A) \geq -1/8$.

The proof appears in Section 4.5.1 below. The next step is to take care of the tight hypergraphs.

Lemma 4.34. *Let B be the admissible matrix*

$$B = E_{F_n} - E_\square. \tag{4.4}$$

(a) $\lambda_G(B) = 0$ *whenever G contains less than 4 edges.*

(b) $\lambda_{F_4}(B) = 1/16$.

(c) $\lambda_{K_4^-}(B) = 1/8$.

(d) $|\lambda_G(B)| \leq 1$ *for all hypergraphs G.*

Proof. Items (a) is immediate.

For item (b), $\lambda_{F_4}(B)$ is the probability that $F_4 \subseteq P_y$ when y is chosen randomly. Since all vectors in F_4 are linearly independent, it is easy to check that the probability is $1/16$.

For item (c), we show that the probability that $K_4^- \cap P_w$ is a square is $1/8$ when w is chosen randomly. Let $K_4^- = \{x, y, z, x + y, x + z\}$, where x, y, z are linearly independent. The only square contained in K_4^- is $\{y, z, x + y, x + z\}$ (this is easy to check: these are the only four vectors summing to zero). The vectors $y, z, x+y$ are linearly independent, and so the probability that all of them belong to $K_4^- \cap P_w$ is 1/8. Since $x + z = y + z + (x + y)$, in this case $x + z$ belongs to the intersection as well.

Item (d) is clear since $0 \le \lambda_G(E_X) \le 1$ for all hypergraphs X. □

This shows that the counterpart of Lemma 4.15 is true.

Lemma 4.35. *Let $C = A + (2/119)B$, where A is given by (4.1) and B is given by (4.4). The matrix C is admissible and satisfies the following properties:*

(a) $\lambda_\varnothing(C) = 1$.

(b) $\lambda_H(C) \ge -1/7$ *for all hypergraphs H, for equality only for the following graphs: forests of size one or two; triangles.*

(c) If $\lambda_H(C) > -1/7$ *then in fact* $\lambda_H(C) \ge -135/952$.

Proof. The proof is the same as the proof of Lemma 4.15. □

At this point, we can prove Theorem 4.2.

Lemma 4.36. *Let $\mathcal{F}$ be an odd-circuit-intersecting family of hypergraphs with characteristic function $f = \mathbf{1}_\mathcal{F}$.*

Upper bound: $\mu(\mathcal{F}) \le 1/8$.

Uniqueness: *If $\mu(\mathcal{F}) = 1/8$ then the Fourier expansion of f is supported on the first 4 levels, that is $\hat{f}(S) = 0$ for $|S| > 3$.*

Stability: *If $\mu(\mathcal{F}) \ge 1/8 - \epsilon$ then*

$$\sum_{|S|>3} \hat{f}^2(S) = O(\epsilon).$$

Proof. Proved like Lemma 4.16. □

Theorem 4.37. *Let $\mathcal{F}$ be an odd-circuit-intersecting family of hypergraphs.*

Upper bound: $\mu(\mathcal{F}) \le 1/8$.

Uniqueness: *$\mu(\mathcal{F}) = 1/8$ if and only if $\mathcal{F}$ is a Schur-star.*

Stability: *If $\mu(\mathcal{F}) \geq 1/8 - \epsilon$ then $\mu(\mathcal{F}\Delta\mathcal{G}) = O(\epsilon)$ for some Schur-star $\mathcal{G}$.*

Proof. Proved like Theorem 4.17. Note that a Schur-star is the same as a $\triangle$-star. □

Theorem 4.2. *Let $\mathcal{F}$ be an odd-circuit-agreeing family of hypergraphs on n points.*

Upper bound: *$\mu(\mathcal{F}) \leq 1/8$.*

Uniqueness: *$\mu(\mathcal{F}) = 1/8$ if and only if $\mathcal{F}$ is a Schur-semistar.*

Stability: *If $\mu(\mathcal{F}) \geq 1/8 - \epsilon$ then there is a Schur-semistar $\mathcal{G}$ such that $\mu(\mathcal{F}\Delta\mathcal{G}) = O(\epsilon)$.*

Proof. Proved like Theorem 4.1, since the condition in Theorem 4.28 is satisfied: if $H = \{x, y, x+y\}$ is a triangle and $z \notin H$ then neither $\{x, y\}$ nor $\{x, y, z\}$ contain any cycles. □

4.5.1 Proof of Lemma 4.33

We start by generalizing our development of cut statistics. The cut function is defined as before.

Definition 4.17. For a hypergraph H, its *cut function* is

$$Q_H(X) = \sum_{k=0}^{\infty} q_k(G) X^k.$$

○

In the case of graphs, we decomposed a given graph into its connected components and then into its blocks. Here, it will be enough to consider a much coarser decomposition, in which only the bridges are distinguished.

Definition 4.18. Let H be a hypergraph. A hyperedge $x \in H$ is a *bridge* if no cycle in H contains x. Alternatively, $x \notin \operatorname{Span}(H \setminus \{x\})$. Denote the set of bridges in H by $B(H)$, and let $\bar{B}(H) = H \setminus B(H)$. ○

Lemma 4.38. *Let H be a hypergraph, and $m = |B(H)|$. We have*

$$Q_H(X) = \left(\frac{1}{2} + \frac{1}{2}X\right)^m Q_{\bar{B}(H)}.$$

Proof. It is easy to check that $Q_-(X) = 1/2 + (1/2)X$, where $-$ is an edge. The proof will be complete if we show that whenever H is a hypergraph and $v \notin \mathrm{Span}(H)$ then $Q_{H\cup\{v\}} = Q_H Q_{\{v\}}$.

Let $u \in \mathrm{Span}(H)^\perp$ be some vector satisfying $\langle\langle u, v\rangle\rangle = 1$. Let $y = y_1 + y_2$, where y_1 is chosen randomly from $\{v\}^\perp$ and y_2 is chosen randomly from $\{0, u\}$. The vector y is distributed uniformly on $\mathbb{Z}_2^n$, $H \cap Q_y$ depends only on y_1, and $\{v\} \cap Q_y$ depends only on y_2. Hence the formula $Q_{H\cup\{v\}} = Q_H Q_{\{v\}}$ follows from basic properties of generating functions (like in the proof of Lemma 4.18). □

Corollary 4.39. *Let H be a hypergraph with $m = |B(H)|$, and let $a_k = q_k(\bar{B}(H))$. Then $a_1 = 0$ and*

$$Q_H(X) = \frac{1}{2^m}\left(a_0 + ma_0X + \left(\binom{m}{2}a_0 + a_2\right)X^2 + \left(\binom{m}{3}a_0 + ma_2 + a_3\right)X^3\right) + \cdots. \tag{4.3'}$$

Proof. We start by proving that $a_1 = 0$. Suppose $\bar{B}(H) \cap P_y = \{z\}$. If there were a cycle $C \subseteq \bar{B}(H)$ containing z then

$$0 = \left\langle\!\!\left\langle y, \sum_{c\in C} c \right\rangle\!\!\right\rangle = \sum_{c\in C} \langle\langle y, c\rangle\rangle = 1.$$

We conclude that z is a bridge, yet by construction $\bar{B}(H)$ has no bridges. Therefore $a_1 = 0$.

The rest of the proof follows from Lemma 4.38 like in the proof of Corollary 4.22. □

We now generalize the two auxiliary lemmas Lemma 4.23 and Lemma 4.24. The only major difference is that a new tight case appears in the counterpart of Lemma 4.24(b).

Definition 4.19. Let H be a hypergraph. A *vertex* is any element of $[n]$. The *neighborhood* of a vertex i in H is $N(i) = \{x \in H : x_i = 1\}$. The *degree* of a vertex is the size of its neighborhood. ○

Lemma 4.40. *Let H be a hypergraph.*

(a) $q_0(H) = 2^{-\mathrm{rank}(H)}$.

(b) $q_1(H) = |B(H)|q_0(H)$.

(c) If H has a vertex of odd degree then $q_k(H) \le 1/2$ for all k.

(d) For any odd k, $q_k(H) \le 1/2$.

(e) Always $q_2(H) \le 3/4$.

Proof. For item (a), let R be a basis of H. Clearly $H \cap Q_y = \emptyset$ if and only if $R \cap Q_y = \emptyset$. The formula immediately follows.

Item (b) follows directly from (4.3').

For item (c), let $i \in [n]$ have odd degree, and notice that

$$(H \cap Q_y)\Delta(H \cap Q_{y+e_i}) = N(i),$$

where e_i is the vector whose only non-zero coordinate is i. Since $|N(i)|$ is odd, $|H \cap Q_y| \neq |H \cap Q_{y+e_i}|$, and so at most one of them can be equal to k.

For item (d), we can assume that all vertices have even degree. In that case, for each $y \in \mathbb{Z}_2^n$ we have

$$\sum_{x \in H} \langle\!\langle x, y \rangle\!\rangle = \sum_{i \in [n]: y_i = 1} \sum_{x \in H: x_i = 1} 1 = \sum_{i \in [n]: y_i = 1} \deg(i) = 0,$$

(the calculation taking place in $\mathbb{Z}_2$), and so $|H \cap P_y|$ is always even.

For item (e), follow the proof of Lemma 4.23(e) to show that $q_2(H) < 3/4$ whenever $|H| \ge 6$. We can further restrict ourselves to the case in which all vertices have even degree, and so the vectors in H sum to zero. In particular, H is linearly dependent. Let H' be the smallest subset of H which is linearly dependent. Thus $|H'| \ge 3$ and $H \setminus H'$ also sums to zero. Since $|H \setminus H'| \le 2$, we conclude that H is minimally linearly dependent, and so a cycle. One checks that $q_2(C_3) = q_2(C_4) = 3/4$ and $q_2(C_5) = 5/8$. □

Definition 4.20. A hypergraph is a *k-hyperclique* if it consists of a linear subspace of dimension k, minus $\mathbf{0}$. We denote a k-hyperclique by $\mathbf{K}_k$. ○

Note that a clique on k vertices corresponds to a $(k-1)$-hyperclique.

Lemma 4.41. *Let H be a hypergraph.*

(a) We have $q_0(\emptyset) = 1$, $q_0(-) = 1/2$, *and* $q_0(H) \le 1/4$ *for all other hypergraphs.*

(b) If $|B(H)| = 0$ *and* $|H|$ *is odd then either* $q_0(H) \le 1/16$ *or* H *is one of the following: a triangle, a diamond or a 3-hyperclique.*

(c) If $\bar{B}(H)$ *is non-empty then* $q_0(\bar{B}(H)) \le 1/4$.

Note that a 3-hyperclique is not isomorphic to any graph.

Proof. Item (a) follows directly from Lemma 4.23(a).

For item (b), note that if $\operatorname{rank} H \geq 4$ then $q_0(H) \leq 1/16$ by Lemma 4.23(a). Otherwise, H is a bridge-less hypergraph of rank at most 3 with an odd number of vectors. A hypergraph of rank 1 is a bridge. The only bridge-less hypergraph of rank 2 is a triangle. A hypergraph of rank 3 with an odd number of vectors contains either 3, 5 or 7 vectors. In the first case, the hypergraph consists of bridges. In the second case, it consists either of a 3-cycle and two additional vectors, which must be bridges, or of a 4-cycle and an additional vertex contained in their span, a hypergraph which is isomorphic to a diamond. In the third case, it is a 3-hyperclique.

Item (c) follows directly from item (a). □

We are now ready to prove Lemma 4.33.

Lemma 4.33. *The matrix A given by* (4.1) *is admissible, and satisfies the following properties:*

(a) $\lambda_\varnothing(A) = 1$.

(b) $\lambda_H(A) \geq -1/7$ *for all hypergraphs H, with equality only for the following hypergraphs: forests of size one, two, or four; triangles; diamonds.*

(c) If $\lambda_H(A) > -1/7$ *then in fact* $\lambda_H(A) \geq -1/8$.

Proof. The proof of Lemma 4.13 relies only on formula (4.3), Lemma 4.23 and Lemma 4.24. Formula (4.3) has its exact analog in formula (4.3'), and the lemmas have their analogs in Lemma 4.40 and Lemma 4.41. The only difference is the additional tight case in Lemma 4.41(b), a 3-hyperclique. In addition, some translation is needed: m should be replaced by $|B(H)|$, and H by $\bar{B}(H)$.

Going through the proof of Lemma 4.13, the only point in which Lemma 4.24(b) is used is the case where $|H|$ is odd, $|B(H)| = 0$ and $q_0(H) = 1/16$. The proof is complete if we verify that $f(\mathbf{K}_3) \leq 1/8$. Recall that

$$f(\mathbf{K}_3) = q_0(\mathbf{K}_3) - \frac{1}{7}q_2(\mathbf{K}_3) + \frac{3}{28}q_3(\mathbf{K}_3).$$

It is easy to check that either $\mathbf{K}_3 \cap P_y = \varnothing$ (if y is in the orthogonal complement) or $|\mathbf{K}_3 \cap P_y| = 4$. The first event happens with probability 1/8 by Lemma 4.40, and so $f(\mathbf{K}_3) = 1/8$. □

4.6 Extension to skewed measures

We now generalize all the work done so far to the μ_p-setting for $p < 1/2$. Our aim is to prove an upper bound of p^3 on odd-circuit-intersecting families, along with the related uniqueness and stability results. We already have the upper bound for $p \le 1/4$ by Theorem 3.16 on page 41, and uniqueness and stability follow for $p < 1/4$ as in the proof of Theorem 4.17 on page 89. In this section we bridge the gap between $p = 1/4$ and $p = 1/2$. This will enable us to use Theorem 3.33 to get results on uniform odd-cycle-intersecting families of graphs and uniform odd-circuit-intersecting families of hypergraphs.

For the rest of this section, fix some $p < 1/2$, and let $q = 1 - p$. We start by generalizing the construction of the matrix A.

Definition 4.21. We say that a matrix A is *odd-circuit-admissible* (*admissible* for short) if it satisfies the following two properties:

Intersection property If G, H are hypergraphs whose intersection contains an odd circuit then $\mathbf{1}_G' A \mathbf{1}_H = 0$.

Eigenvector property The eigenvectors of $\mathcal{F}$ are the Fourier characters $\chi_{S,p}$.

If A is admissible then we use $\lambda_H(A)$ to denote the eigenvalue corresponding to χ_H. ○

Lemma 4.42. *Let J be a set of vectors. Define*

$$B_{J,i} = \begin{cases} \begin{pmatrix} 1 - \frac{p}{q} & \frac{p}{q} \\ 1 & 0 \end{pmatrix}, & \text{if } i \notin J, \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, & \text{if } i \in J, \end{cases}$$

$$B_J = \bigotimes_{i=1}^{n} B_{J,i}.$$

If J contains no odd circuits then the matrix B_J is admissible, and the eigenvalue corresponding to χ_H is

$$\lambda_H(B_J) = \left(-\frac{p}{q}\right)^{|S \setminus J|}.$$

Furthermore, the vector space of all admissible matrices is spanned by B_J for all bipartite J.

Proof. The proof is the same as the proof of Lemma 4.11, generalizing to arbitrary p. □

Lemma 4.43. *For every hypergraph H there is an admissible matrix E_H such that $\lambda_G(E_H) = (-p/q)^{|G|} q_H(G)$.*

For every integer k there is an admissible matrix E_k such that $\lambda_G(E_k) = (-p/q)^{|G|} q_k(G)$.

Proof. The proof is the same as the proof of Lemma 4.12. □

The definition of A has to be adapted. This time we are opting for a minimal eigenvalue of $-p^3/(1-p^3)$. We are again looking for a matrix of the form

$$\begin{aligned} A_p &= c_0 E_0 + c_1 E_1 + c_2 E_2 + c_3 E_3 + c_4 E_4, \\ \lambda_G(A_p) &= \left(-\frac{p}{q}\right)^{|G|} (c_0 q_0(G) + c_1 q_1(G) + c_2 q_2(G) + c_3 q_3(G) + c_4 q_4(G)). \end{aligned}$$

As in Section 4.3, by considering small graphs we get the following constraints on the coefficients:

$$\begin{aligned} c_0 &= 1, \\ c_1 &= \frac{p^2 - p - 1}{p^2 + p + 1}, \\ c_2 &= \frac{p^2 - 3p + 1}{p^2 + p + 1}, \\ \frac{5p^2 - 27p + 45 - 16/p}{p^2 + p + 1} &\le 4c_3 + c_4 \le \frac{5p^2 - 27p + 45 - 32/p + 8/p^2}{p^2 + p + 1}. \end{aligned}$$

When $p = 1/2$, the two bounds on $4c_3 + c_4$ coincide. When $p > 1/2$, they contradict one another, so the method fails. When $p < 1/2$, there is a gap, and choosing any value inside the gap, the corresponding eigenvalues are tight on neither 4-forests nor K_4^-. As before, we choose $c_4 = 0$. A judicious choice of c_3 is:

$$c_3 = \frac{5p^2 - 27p + 45 - 28/p + 6/p^2}{4(p^2 + p + 1)}.$$

This choice guarantees that $c_3 > 0$ for all $p \in (0, 1/2]$. Summarizing, we define A_p as follows:

$$A_p = E_0 + \frac{p^2 - p - 1}{p^2 + p + 1} E_1 + \frac{p^2 - 3p + 1}{p^2 + p + 1} E_2 + \frac{5p^2 - 27p + 45 - 28/p + 6/p^2}{4(p^2 + p + 1)} E_3. \tag{4.5}$$

We have the following generalization of Lemma 4.33.

Lemma 4.44. *Let $\tau = 0.248$, and suppose $\tau \le p < 1/2$. The matrix A_p given by (4.5) is admissible, and satisfies the following properties:*

(a) $\lambda_\varnothing(A) = 1$.

(b) $\lambda_H(A_p) \ge -p^3/(1-p^3)$ *for all hypergraphs H, with equality only for the following hypergraphs: forests of size one or two; triangles.*

The reason we only prove the lemma for $p \ge \tau$ is that it is false for p below some critical point smaller than $\tau = 0.248$. We prove the lemma in Section 4.6.1. The proof is very similar to the earlier proofs, but is complicated by the fact that we have to care about a range of values of p.

Using Lemma 4.44 we can generalize Theorem 4.17 to all $p < 1/2$. First we prove a generalization of Lemma 4.36.

Lemma 4.45. *Let $\mathcal{F}$ be an odd-circuit-intersecting family of hypergraphs with characteristic function $f = \mathbf{1}_{\mathcal{F}}$, and suppose $\tau \le p < 1/2$.*

Upper bound: $\mu_p(\mathcal{F}) \le p^3$.

Uniqueness: *If $\mu_p(\mathcal{F}) = p^3$ then the Fourier expansion of f is supported on the first 4 levels, that is $\hat{f}_p(S) = 0$ for $|S| > 3$.*

Stability: *If $\mu_p(\mathcal{F}) \ge p^3 - \epsilon$ then*

$$\sum_{|S|>3} \hat{f}_p^2(S) \le T_p \epsilon,$$

where T_p is a continuous function of p which doesn't depend on n.

Proof. The lemma is proved much like Lemma 4.16. In order to show that C_p is a continuous function of p, recall that C_p derives from Hoffman's bound:

$$T_p = \frac{-\lambda_{\min}}{\lambda_2 - \lambda_{\min}}.$$

In our case $\lambda_{\min} = -p^3/(1-p^3)$, and λ_2 is the second minimal eigenvalue of A_p.

First we show that for each p there is some constant T_p that doesn't depend on n. Lemma 4.44 shows that $\lambda_H(A_p) > \lambda_{\min}$ for all graphs other than the graphs listed there. As $|H| \longrightarrow \infty$,

$|\lambda_H(A_p)| \longrightarrow 0$, and so λ_2 is bounded away from $\lambda_{\min}$, showing that we can choose T_p that doesn't depend on n.

Second, we show that T_p is continuous on each interval $[\tau, 1/2-\delta]$. For each δ we can find N_δ such that if $|H| > N_\delta$ then $|\lambda_H(A_p)| \le |\lambda_{\min}|/2$ for all p in the interval. Hence λ_2 depends only on finitely many hypergraphs, and because $\lambda_H(A_p)$ is continuous for every H, we deduce that T_p is continuous. □

As $p \longrightarrow 1/2$, the constant T_p tends to ∞ due to the hypergraphs $\triangle, K_4^-$ which are tight for $p = 1/2$. We can perturb A_p as in Section 4.3, using the same matrix B and constructing C_p in a similar way: $C_p = A_p + (16/17)(\lambda_2(A_p) - \lambda_{\min}(A_p))B$. As a result, we can obtain a version of Lemma 4.45 in which T_p is bounded as $p \longrightarrow 1/2$. The interested reader can consult the details in [27]. We omit this step since it is not needed for any of our results.

Now we are ready to prove the generalization of Theorem 4.17.

Theorem 4.3. *Let $\mathcal{F}$ be an odd-circuit-intersecting family of hypergraphs on n points, and suppose* $0 < p < 1/2$.

Upper bound: $\mu_p(\mathcal{F}) \le p^3$.

Uniqueness: $\mu_p(\mathcal{F}) = p^3$ *if and only if $\mathcal{F}$ is a Schur-star.*

Stability: *If $\mu_p(\mathcal{F}) \ge p^3 - \epsilon$ then $\mu_p(\mathcal{F}\Delta\mathcal{G}) \le K_p\epsilon$ for some Schur-star $\mathcal{G}$, where K_p is a constant depending continuously on p in the interval* $(0, 1/2)$.

Proof. The upper bound is given by Lemma 4.45 for $p \ge \tau$ and by Theorem 3.16 for $p \le \tau$. For uniqueness, the argument in the proof of Theorem 3.16 shows that if $\mu(\mathcal{F}) = p^3$ then $\mathcal{F}$ is a 3-star. Since $\mathcal{F}$ is odd-circuit-intersecting, it must be a Schur-star.

For stability, suppose $\mu(\mathcal{F}) \ge p^3 - \epsilon$. Suppose first that $p \ge \tau$. The stability part of Lemma 4.45 combined with Theorem 2.23 shows that $\mathcal{F}$ is $D_p\epsilon$-close to a family $\mathcal{G}$ depending on M_p coordinates, where $D_p = T_pC_{p,3}$ (T_p coming from the lemma and $C_{p,3}$ from the theorem) and M_p are continuous.

We claim that if ϵ is small enough then $\mathcal{G}$ is odd-circuit-intersecting. Suppose $\mathcal{G}$ is not odd-circuit-intersecting. Thus there are two hypergraphs $H_1, H_2 \in \mathcal{G}$ whose intersection contains no odd cycles. We can assume furthermore that $|H_1|, |H_2| \le M_p$, say both are supported on the edge set X of size M_p. Let $\mathcal{F}_1 = \{H \subseteq \overline{X} : H \cup H_1 \in \mathcal{F}\}$, and define $\mathcal{F}_2$ similarly. If $H \in \mathcal{F}_1$ then $\overline{X} \setminus H \notin \mathcal{F}_2$, since $\mathcal{F}$ is odd-cycle-intersecting. Therefore $\mathcal{F}_1, \mathcal{F}_2$ are cross-intersecting, and Lemma 3.24 shows that $\mu_p(\mathcal{F}_1)\mu_p(\mathcal{F}_2) \le p^2$. Thus either $\mu_p(\mathcal{F}_1) \le p$ or $\mu_p(\mathcal{F}_2) \le p$. Without loss of generality, assume $\mu_p(\mathcal{F}_1) \le p$. Since all the corresponding graphs belong to $\mathcal{G}$, this shows that $\mu_p(\mathcal{F}\Delta\mathcal{G}) \ge (1-p)\mu_p^{[X]}(H_1) \ge (1-p)p^{M_p}$, which is impossible if $\epsilon < (1-p)p^{M_p}$.

There are only finitely many odd-circuit-intersecting families $\mathcal{G}$ depending on M_p coordinates such that $\mu_p(\mathcal{G}) < p^3$. Let the maximal μ_p-measure among all of them be $p^3 - \alpha$. Note that α is continuous in p. If $\mu_p(\mathcal{G}) \neq p^3$ then $\mu_p(\mathcal{F}\Delta\mathcal{G}) \ge |\mu_p(\mathcal{F}) - \mu_p(\mathcal{G})| \ge \alpha - \epsilon$ and so $\alpha \le (D_p+1)\epsilon$, which is impossible if $\epsilon < \alpha/(D_p+1)$. By uniqueness, the only odd-circuit-intersecting family with measure p^3 is a Schur-star.

Taking $K'_p = \max(D_p, ((1-p)p^{M_p})^{-1}, (\alpha/(D_p+1))^{-1})$, we deduce that for some Schur-star $\mathcal{G}$, $\mu_p(\mathcal{F}\Delta\mathcal{G}) \le K'_p\epsilon$, for any value of ϵ. Notice that K'_p is continuous in p.

Suppose next that $p \le \tau$. Replacing Lemma 4.45 with Lemma 3.14, the same argument shows that for some Schur-star $\mathcal{G}$, $\mu_p(\mathcal{F}\Delta\mathcal{G}) \le K''_p\epsilon$, where K''_p is continuous in p.

We can find a continuous function K_p defined for all $p \in (0, 1/2)$ larger than both K'_p and K''_p in their respective intervals, completing the proof of stability. □

Applying Theorem 3.33, we deduce the following version of our main theorem for uniform families of odd-circuit-intersecting hypergraphs.

Theorem 4.4. *For every $\delta > 0$ there are constants C_δ, N_δ such that for any $k \in (0, n/2 - \delta n)$ and any odd-circuit-intersecting k-uniform family $\mathcal{F}$ of hypergraphs on $n \ge N_{t,\delta}$ points,*

$$|\mathcal{F}| < \binom{n-3}{k-3} + \epsilon\binom{n}{k}.$$

If furthermore $\mathcal{F}$ satisfies

$$|\mathcal{F}| > \binom{n-3}{k-3} - \epsilon\binom{n}{k},$$

then there exists a Schur-star $\mathcal{H}$ such that

$$|\mathcal{F}\Delta\,\mathrm{Sl}(\mathcal{H}, k)| < C_\delta\left(\epsilon + \sqrt{\frac{\log n}{n}}\right)\binom{n}{k}.$$

Proof. Let $\mathfrak{P}^{\triangle}$ be the monotone object of odd-circuit-intersecting families of hypergraphs. Theorem 4.3 shows that $\mathfrak{P}^{\triangle}$ is weakly μ-dominated in $(0, 1/2)$. The result now follows as in the proof of Theorem 3.34 on page 61. □

A similar result can be proved for uniform families of odd-cycle-intersecting graphs.

Theorem 4.5. *For every $\delta > 0$ there are constants C_δ, N_δ such that for any $k \in (0, n/2 - \delta n)$ and any odd-cycle-intersecting k-uniform family $\mathcal{F}$ of graphs on $n \geq N_{t,\delta}$ points,*

$$|\mathcal{F}| < \binom{n-3}{k-3} + \epsilon\binom{n}{k}.$$

If furthermore $\mathcal{F}$ satisfies

$$|\mathcal{F}| > \binom{n-3}{k-3} - \epsilon\binom{n}{k},$$

then there exists a $\triangle$-star $\mathcal{H}$ such that

$$|\mathcal{F} \Delta \operatorname{Sl}(\mathcal{H}, k)| < C_\delta\left(\epsilon + \sqrt{\frac{\log n}{n}}\right)\binom{n}{k}.$$

Proof. The proof is the same as Theorem 4.4. □

4.6.1 Proof of Lemma 4.44

The proof of Lemma 4.44 is complicated by the fact that instead of the arithmetic inequalities appearing in the proof of Lemma 4.13, this time we get polynomial inequalities. Furthermore, the sign of some of the coefficients will depend on p. To handle the latter problem, when we want to lower bound an expression $c_k\alpha$ given bounds $0 \leq \alpha \leq B$, we will replace it by $\min(c_k B, 0)$. If $c_k \geq 0$ then $c_k\alpha \geq 0$, and otherwise $c_k\alpha \geq c_k B$.

As a result, we will get expressions involving (at times) multiple invocations of min. Each such inequality is equivalent to an inequality of the form $\min S \geq 0$, where S is a finite set of polynomials. In order to verify these inequalities, we check that $P \geq 0$ for each $P \in S$. In order to check that $P \geq 0$, we check that $P(3/8) > 0$ and that P has no zeroes in $[\tau, 1/2)$; the latter can be done formally using Sturm chains [76]. This tedious verification has been done for all the inequalities appearing in this section, and so the reader can rest assured that the proof is correct.

The proof requires one additional auxiliary lemma.

Lemma 4.46. *Let H be a hypergraph.*

(a) *If $|B(H)| = 1$ and $|H| > 1$ then $|H| \leq 4$.*

(b) *If $|B(H)| = 0$ and $|H| \leq 5$ then H is one of the following hypergraphs: the empty hypergraph, C_3, C_4, C_5, K_4^-.*

Proof. For item (a), if $|H| > 1$ then $\bar{B}(H) \neq \emptyset$ and so $|\bar{B}(H)| \geq 3$, since $\bar{B}(H)$ must contain a cycle.

For item (b), notice that if H is non-empty then H must contain a cycle. If $|H| = 3$ then $H = C_3$. If $|H| = 4$ then either $H = C_4$, or H contains a triangle, and so it is of the form $H = \{x, y, x+y, z\}$; clearly z must be a bridge. If $|H| = 5$ then either $H = C_5$, H contains a triangle, or H contains a square. If H contains a triangle then since H is bridge-less, it must be of the form $H = \{x, y, x+y, z, x+z\}$ which is a diamond. If H contains a square then it must be of the form $H = \{x, y, z, x+y+z, x+y\}$, again a diamond. □

Lemma 4.44. *Let $\tau = 0.248$, and suppose $\tau \leq p < 1/2$. The matrix A_p given by (4.5) is admissible, and satisfies the following properties:*

(a) *$\lambda_\emptyset(A) = 1$.*

(b) *$\lambda_H(A_p) \geq -p^3/(1-p^3)$ for all hypergraphs H, with equality only for the following hypergraphs: forests of size one or two; triangles.*

Proof. Let $c_0 = 1, c_1, c_2, c_3$ be the coefficients of E_0, E_1, E_2, E_3 in (4.5). One checks that c_1 is always negative on $[\tau, 1/2)$ and c_3 is always positive. The coefficient c_2 is more troublesome: it changes sign from positive to negative at $(3-\sqrt{5})/2 \approx 0.382$.

In view of Lemma 4.43, the eigenvalues of A satisfy the formula

$$\lambda_H(A_p) = \left(-\frac{p}{1-p}\right)^{|H|}(q_0(H) + c_1 q_1(H) + c_2 q_2(H) + c_3 q_3(H)).$$

This formula already shows that $\lambda_\emptyset(A_p) = 1$. Let $m = |B(H)|$. We split the proof into two cases: $|H|$ is odd and $|H|$ is even.

Hypergraphs with an odd number of vectors. Suppose $|H|$ is odd. We show that $\lambda_H(A_p) \geq -p^3/(1-p^3)$, with equality only if H is a single vector or a triangle. Lemma 4.40(d,e) implies the general bound

$$\lambda_H(A_p) \geq -\left(\frac{p}{1-p}\right)^{|H|}\left[q_0(H)(1+mc_1) + \max\left(\frac{3}{4}c_2, 0\right) + \frac{1}{2}c_3\right].$$

It can be checked that $1 + c_1 > 0$, whereas $1 + mc_1 < 0$ for $m \geq 2$.

When $m = 0$, Lemma 4.46(b) shows that either H is a triangle, C_5 or K_4^-, or $|H| \geq 7$. If H is a triangle then $\lambda_H(A_p) = -p^3/(1-p^3)$. If H is C_5 or K_4^-, we can verify that $\lambda_H(A_p) > -p^3/(1-p^3)$ by direct calculation, except that for K_4^-, we get equality when $p = 1/2$. If $|H| \geq 7$ then Lemma 4.41(b) shows that either $H = \mathbf{K}_3$ or $q_0(H) \leq 1/16$. In the former case, one can verify that $\lambda_H(A_p) > -p^3/(1-p^3)$ directly. In the latter case,

$$\lambda_H(A_p) \geq -\left(\frac{p}{1-p}\right)^{7}\left[\frac{1}{16} + \max\left(\frac{3}{4}c_2, 0\right) + \frac{1}{2}c_3\right].$$

One can check that the right-hand side is always larger than $-p^3/(1-p^3)$.

When $m = 1$, Lemma 4.46(a) implies that either $|H| = 1$ or $|H| \geq 5$. In the former case, $\lambda_H(A_p) = -p^3/(1-p^3)$. In the latter case, Lemma 4.41(a) implies that $q_0(H) \leq 1/4$, and therefore

$$\lambda_H(A_p) \geq -\left(\frac{p}{1-p}\right)^{5}\left[\frac{1}{4}(1+c_1) + \max\left(\frac{3}{4}c_2, 0\right) + \frac{1}{2}c_3\right].$$

It can be checked that the right-hand side is always larger than $-p^3/(1-p^3)$.

When $m \geq 2$, since $1 + mc_1 < 0$, we have the sharper estimate

$$\lambda_H(A_p) \geq -\left(\frac{p}{1-p}\right)^{|H|}\left[\max\left(\frac{3}{4}c_2, 0\right) + \frac{1}{2}c_3\right].$$

If $|H| = 3$ then $H = F_3$, and we can verify that $\lambda_H(A_p) > -p^3/(1-p^3)$ directly. Otherwise, $-(p/(1-p))^{|H|} \geq -(p/(1-p))^5$, and so

$$\lambda_H(A_p) \geq -\left(\frac{p}{1-p}\right)^{5}\left[\max\left(\frac{3}{4}c_2, 0\right) + \frac{1}{2}c_3\right].$$

It can be checked that the right-hand side is always larger than $-p^3/(1-p^3)$.

Hypergraphs with an even number of vectors. Suppose $|H|$ is even. We show that $\lambda_H(A_p) \geq -p^3/(1-p^3)$, with equality only if H consists of two vectors. Formula (4.3') implies that

$$\lambda_H(A_p) = \left(\frac{p}{1-p}\right)^{|H|}(d_0(m)a_0 + d_2(m)a_2 + d_3(m)a_3),$$

where d_0, d_2, d_3 are defined by

$$\begin{aligned} d_0(m) &= 2^{-m}\left[1 + mc_1 + \binom{m}{2}c_2 + \binom{m}{3}c_3\right], \\ d_2(m) &= 2^{-m}(c_2 + mc_3), \\ d_3(m) &= 2^{-m}c_3. \end{aligned}$$

Since $c_3 > 0$, we know that $d_3(m) > 0$. We can further check that $d_2(m) > 0$ when $m \geq 2$; this just involves checking that $c_2 + 2c_3 > 0$.

We claim that $d_0(m) > 0$ for $m \geq 10$. To see this, check first that $c_1 + 7c_3 > 0$ and $c_2 + 2c_3 > 0$. Note that

$$\begin{aligned} 2^{m+1}d_0(m+1) - 2^m d_0(m) &= c_1 + mc_2 + \binom{m}{2}c_3 \\ &\geq (c_1 + 7c_3) + m(c_2 + 2c_3) > 0, \end{aligned}$$

using $\binom{m}{2} \geq 2m + 7$, which is true for $m \geq 7$. It remains to check by direct calculation that $d_1(10) > 0$.

We have shown that when $m \geq 10$, $\lambda_H(A_p) > 0$. If $m < 10$ and H is a forest, then H is either a 2-forest, a 4-forest, a 6-forest or an 8-forest. If H is a 2-forest, then $\lambda_H(A_p) = -p^3/(1-p^3)$. For the other forests listed, direct calculation shows that $\lambda_H(A_p) > -p^3/(1-p)^3$, except that for 4-forests, we get equality when $p = 1/2$.

The remaining case is when $m < 10$ and H is not a forest. Lemmas 4.40(e) and 4.41(c) give the following bound:

$$\lambda_H(A_p) \geq \left(\frac{p}{1-p}\right)^2\left[\min\left(\frac{1}{4}d_0(m), 0\right) + \min\left(\frac{3}{4}d_2(m), 0\right)\right].$$

It can be checked that for all $m < 10$, the right-hand side is larger than $-p^3/(1-p^3)$. $\square$

Chapter 5

The Ahlswede–Khachatrian theorem

The Erdős–Ko–Rado theorem determines the largest μ_p-measure of an intersecting family of sets. In this chapter, we consider the analogue of this theorem to t-intersecting families (families in which any two sets have at least t elements in common), following Ahlswede and Khachatrian [2, 3]. We present a proof of the μ_p version of their theorem, which is adapted from the earlier proofs. Due to the simpler nature of the μ_p setting, our proof is simpler and cleaner.

We have already considered t-intersecting families in Section 3.3, in which we proved a theorem of Friedgut showing that if $\mathcal{F}$ is a t-intersecting family of sets and $p \le 1/(t+1)$ then $\mu_p(\mathcal{F}) \le p^t$. The upper bound on p came naturally from the proof. This limitation is not arbitrary. Indeed, when $p > 1/(t+1)$, the bound p^t is incorrect. The correct bound was found by Ahlswede and Khachatrian [2, 3] in the k-uniform setting. We state it in the language of *slices*, defined in Section 3.5.1 on page 53: for a family of sets $\mathcal{F}$, $\mathrm{Sl}(\mathcal{F}, k) = \{A \in \mathcal{F} : |A| = k\}$.

Definition 5.1. The (t,r) Frankl family $\mathcal{F}_{t,r}$ is the t-intersecting family defined by

$$\mathcal{F}_{t,r} = \{S \subseteq [t+2r] : |S| \ge t+r\}. \qquad \bigcirc$$

Theorem 5.1 (Ahlswede–Khachatrian)**.** *Let $1 \le t \le k \le n$ and $r \ge 0$, and let $\mathcal{F}$ be a t-intersecting family. When*

$$(k-t+1)\left(2+\frac{t-1}{r+1}\right) < n < (k-t+1)\left(2+\frac{t-1}{r}\right),$$

we have $|\mathrm{Sl}(\mathcal{F}, k)| \le |\mathrm{Sl}(\mathcal{F}_{t,r}, k)|$, with equality only if the slices are equivalent.

When

$$n = (k - t + 1)\left(2 + \frac{t-1}{r+1}\right),$$

we have $|\mathrm{Sl}(\mathcal{F}, k)| \le |\mathrm{Sl}(\mathcal{F}_{t,r}, k)| = |\mathrm{Sl}(\mathcal{F}_{t,r+1}, k)|$*, with equality only if* $\mathrm{Sl}(\mathcal{F}, k)$ *is equivalent to either* $\mathrm{Sl}(\mathcal{F}_{t,r}, k)$ *or* $\mathrm{Sl}(\mathcal{F}_{t,r+1}, k)$*.*

Theorem 3.29 on page 54 implies the following counterpart in the μ_p setting.

Corollary 5.2. *If* $\mathcal{F}$ *is* t*-intersecting then for* $r \ge 0$*, when*

$$\frac{r}{t + 2r - 1} < p < \frac{r+1}{t + 2r + 1},$$

we have $\mu_p(\mathcal{F}) \le \mu_p(\mathcal{F}_{t,r})$ *with equality only if* $\mathcal{F}$ *is equivalent to* $\mathcal{F}_{t,r}$*.*

If $p = (r+1)/(t+2r+1)$ *then* $\mu_p(\mathcal{F}) \le \mu_p(\mathcal{F}_{t,r}) = \mu_p(\mathcal{F}_{t,r+1})$*.*

Corollary 5.2 covers all $p < 1/2$ (and for $t = 1$, all $p \le 1/2$). For $p > 1/2$, there is no meaningful bound in sight: the μ_p-measure of the t-intersecting family consisting of all sets of size at least $(n+t)/2$ approaches 1. For $p = 1/2$, the measure of this family approaches 1/2.

Theorem 3.29 isn't strong enough to handle equality when there are two different optimal families. In the rest of this chapter, we adapt the proof of the Ahlswede–Khachtrian theorem to the μ_p setting, thereby settling the cases $p = (r+1)/(t+2r+1)$. We will prove the following version of the Ahlswede–Khachatrian theorem, which uses the notion of *extension*, also defined in Section 3.5.1: for a family of sets $\mathcal{F}$ on m points, $U^n(\mathcal{F}) = \{A \subseteq [n] : A \cap [m] \in \mathcal{F}\}$.

Theorem 5.3. *Let* $\mathcal{F}$ *be a* t*-intersecting family on* n *points for* $t \ge 2$*. If* $r/(t+2r-1) < p < (r+1)/(t+2r+1)$ *for some* $r \ge 0$ *then* $\mu_p(\mathcal{F}) \le \mu_p(\mathcal{F}_{t,r})$*, with equality if and only if* $\mathcal{F}$ *is equivalent to* $U^n(\mathcal{F}_{t,r})$*.*

If $p = (r+1)/(t+2r+1)$ *for some* $r \ge 0$ *then* $\mu_p(\mathcal{F}) \le \mu_p(\mathcal{F}_{t,r}) = \mu_p(\mathcal{F}_{t,r+1})$*, with equality if and only if* $\mathcal{F}$ *is equivalent to either* $U^n(\mathcal{F}_{t,r})$ *or* $U^n(\mathcal{F}_{t,r+1})$*.*

5.1 Proof overview

Our proof of the Ahlswede–Khachatrian theorem in the μ_p setting combines the approaches in the two papers [2, 3] in which Ahlswede and Khachatrian proved their theorem in the classical setting (the two papers present two different proofs). Theorem 3.9 on page 35 covers the case of intersecting families (the classical Erdős–Ko–Rado theorem), and therefore we will concentrate on t-intersecting families for $t \geq 2$.

Given $t \geq 2$ and $p \in (0, 1/2)$, our goal is to determine the t-intersecting families of maximum μ_p-measure. In general, the maximum μ_p-measure of a t-intersecting family depends on the size of its support: for example, the maximum μ_p-measure of a 2-intersecting family on 2 points is p^2 for all $p < 1/2$, but for any $p > 1/3$ there is a 2-intersecting family of larger measure $4p^3 - 3p^4$ on 4 points, namely the Frankl family $\mathcal{F}_{2,1}$. We will not be interested in the maximum μ_p-measure of a t-intersecting family on n points. Rather, we will be interested in the supremum of the μ_p-measures of t-intersecting families on any number of points; we will show that for all $p < 1/2$, the supremum is attained at one of the Frankl families.

The proof uses the technique of *shifting*. A t-intersecting family $\mathcal{F}$ on n points is *left-compressed* if for all $A \in \mathcal{F}$, $j \in A$ and $i \in [n] \setminus A$ satisfying $i < j$, we have $A \setminus \{j\} \cup \{i\} \in \mathcal{F}$. Using shifting, we can show that given any t-intersecting family, there is a left-compressed t-intersecting family with the same μ_p-measure for all p. Therefore as far as upper bounds are concerned, it is enough to consider left-compressed families.

Let $\mathcal{F}$ be a left-compressed t-intersecting family, let $r \geq 0$ be an integer, and suppose that $r/(t + 2r - 1) < p < (r + 1)/(t + 2r + 1)$. We can also assume that $\mathcal{F}$ is monotone (if $A \in \mathcal{F}$ and $B \supseteq A$ then $B \in \mathcal{F}$). The proof consists of two steps. In the first step, we show that if $\mathcal{F}$ depends (as a Boolean function) on some $i > t + 2r$ then we can construct from $\mathcal{F}$ a t-intersecting family of larger μ_p-measure. This implies that the maximum μ_p-measure of a t-intersecting family is attained at some family on $t+2r$ points. In the second step, we show that if $\mathcal{F}$ is not symmetric with respect to its first $t + 2r$ coordinates then we can construct from $\mathcal{F}$ a t-intersecting family of larger μ_p-measure. This implies that the maximum μ_p-measure of a t-intersecting family is attained (uniquely) at a family of the form $\{A \subseteq [t + 2r] : |A| \geq k\}$, and so at the Frankl family $\mathcal{F}_{t,r}$.

A similar but more delicate argument handles the case $p = (r + 1)/(t + 2r + 1)$, and this completes the proof for left-compressed t-intersecting families. The upper bound on the μ_p-measure holds for arbitrary t-intersecting families. An argument similar in spirit to the one in Section 4.4 shows that t-intersecting families of maximum μ_p-measure are equivalent to the corresponding Frankl family or families.

For the duration of the proof, we will use $\mu_p^X(\mathcal{F})$ to denote the μ_p-measure of a family $\mathcal{F}$ as a subset of 2^X.

5.2 Shifting

In this section we develop formally the classical technique of shifting. We start by defining the shifting operator.

Definition 5.2. Let $\mathcal{F}$ be a family of sets on n points, and let $i, j \in [n]$, $i \neq j$. For $A \in \mathcal{F}$, let $\mathbb{S}_{i \leftarrow j}(A) = A \setminus \{j\} \cup \{i\}$ if $j \in A$, $i \notin A$ and $A \setminus \{j\} \cup \{i\} \notin \mathcal{F}$, and let $\mathbb{S}_{i \leftarrow j}(A) = A$ otherwise. The *shifted family* $\mathbb{S}_{i \leftarrow j}(\mathcal{F})$ consists of the sets $\mathbb{S}_{i \leftarrow j}(A)$ for all $A \in \mathcal{F}$. ○

As an example, let $\mathcal{F} = \{\{2\}, \{13\}, \{23\}\}$. Then $\mathbb{S}_{1 \leftarrow 2}(\mathcal{F}) = \{\{1\}, \{13\}, \{23\}\}$. Since $|\mathbb{S}_{i \leftarrow j}(A)| = |A|$, shifting doesn't change the μ_p-measure of a family. Shifting also maintains the property of being t-intersecting.

Lemma 5.4. *Let $\mathcal{F}$ be a family of sets on n points, and let $i, j \in [n]$, $i \neq j$. If $\mathcal{F}$ is t-intersecting for some $t \geq 1$ then $\mathbb{S}_{i \leftarrow j}(\mathcal{F})$ is also t-intersecting.*

Proof. Let $A' = \mathbb{S}_{i \leftarrow j}(A), B' = \mathbb{S}_{i \leftarrow j}(B) \in \mathbb{S}_{i \leftarrow j}(\mathcal{F})$, where $A, B \in \mathcal{F}$. We consider several cases. If $A' = A$ and $B' = B$ then $|A' \cap B'| = |A \cap B| \geq t$ since $\mathcal{F}$ is t-intersecting. If $A' \neq A$ and $B' \neq B$ then $i \in A', B'$ and $j \in A, B$, and so $|A' \cap B'| = |(A \cap B) \setminus \{j\} \cup \{i\}| = |A \cap B| \geq t$. The remaining case is when $A' \neq A$ and $B' = B$. If $j \notin B$ then $|A' \cap B'| \geq |(A \setminus \{j\}) \cap B| = |A \cap B| \geq t$. If $j \in B$ and $i \in B$ then $|A' \cap B'| = |(A \setminus \{j\} \cup \{i\}) \cap B| = |A \cap B| \geq t$. If $j \in B$ and $i \notin B$ then by the definition of $\mathbb{S}_{i \leftarrow j}(B)$, we must have $B'' = B \setminus \{j\} \cup \{i\} \in \mathcal{F}$. Hence $|A' \cap B| = |(A' \setminus \{i\} \cup \{j\}) \cap (B \setminus \{j\} \cup \{i\})| = |A \cap B''| \geq t$. Therefore $\mathbb{S}_{i \leftarrow j}(\mathcal{F})$ is t-intersecting. □

By shifting a given family toward smaller elements, we can obtain a left-compressed family.

Definition 5.3. A family $\mathcal{F}$ on n points is *left-compressed* if $\mathbb{S}_{i \leftarrow j}(\mathcal{F}) = \mathcal{F}$ for all $i, j \in [n]$ such that $i < j$. ○

Lemma 5.5. *Let $\mathcal{F}$ be a t-intersecting family on n points. There is a left-compressed t-intersecting family $\mathcal{G}$ on n points such that $\mu_p(\mathcal{G}) = \mu_p(\mathcal{F})$ for all $p \in [0,1]$. Furthermore, $\mathcal{G}$ can be obtained from $\mathcal{F}$ by a sequence of applications of the operators $\mathbb{S}_{i\leftarrow j}$ for various i,j.*

Proof. Let $\Phi(\mathcal{F})$ be the sum of all elements in all sets in $\mathcal{F}$. It is easy to see that $\Phi(\mathbb{S}_{i\leftarrow j}(\mathcal{F})) \leq \Phi(\mathcal{F})$ whenever $i < j$, with equality only if $\mathbb{S}_{i\leftarrow j}(\mathcal{F}) = \mathcal{F}$. Let $\mathbb{S}(\mathcal{F})$ result from applying in sequence the operators $\mathbb{S}_{i\leftarrow j}$ for all $i,j \in [n]$ such that $i < j$, and define a sequence $\mathcal{F}_0 = \mathcal{F}$, $\mathcal{F}_{s+1} = \mathbb{S}(\mathcal{F}_s)$. Since $\Phi(\mathcal{F}_{s+1}) \leq \Phi(\mathcal{F}_s)$ and $\Phi(\mathcal{F}_s)$ is a non-negative integer, $\Phi(\mathcal{F}_s)$ reaches its minimum at some $s = T$. Since $\Phi(\mathcal{F}_{T+1}) = \Phi(\mathcal{F}_T)$ and so $\mathcal{F}_{T+1} = \mathcal{F}_T$, we conclude that $\mathbb{S}_{i\leftarrow j}(\mathcal{F}_T) = \mathcal{F}_T$ for all $i,j \in [n]$ such that $i < j$, and so $\mathcal{F}_T$ is left-compressed. Lemma 5.4 shows that $\mathcal{F}_T$ is t-intersecting. Finally, it is easy to check that shifting preserves the μ_p-measure for all $p \in [0,1]$. □

From now on until Section 5.5 we will only be interested in left-compressed families.

5.3 Generating sets

In this section we implement the first step of the proof, following [2]. In this step, we show that if $\mathcal{F}$ is a monotone left-compressed t-intersecting family and $p < (r+1)/(t+2r+1)$, then either $\mathcal{F}$ depends only on the first $t+2r$ points, or we can modify $\mathcal{F}$ to obtain a t-intersecting family of larger measure. The tool we will use is *generating sets*.

Definition 5.4. Let $\mathcal{F}$ be a family of sets on n points. Its *generating set* $G(\mathcal{F})$ is the family of inclusion-minimal sets in $\mathcal{F}$. Its *extent* $m(\mathcal{F})$ is the largest integer appearing in any set in $G(\mathcal{F})$.

Let G be a family of sets on n points. Its *upset* $U^n(G)$ is the family $\mathcal{F} = \{A \subseteq [n] : A \supseteq B \text{ for some } B \in G\}$.

A family of sets $\mathcal{F}$ on n points is *monotone* if for all $B \in \mathcal{F}$, we have $A \in \mathcal{F}$ whenever $B \subseteq A \subseteq [n]$. An upset is always monotone. If $\mathcal{F}$ is monotone then $\mathcal{F} = U^n(G(\mathcal{F}))$. ○

For example, $G(\mathcal{F}_{t,r}) = \{A \subseteq [t+2r] : |A| = t+r\}$ and $m(\mathcal{F}_{t,r}) = t+2r$. In the language of monotone Boolean functions, if $\mathcal{F}$ is monotone then $G(\mathcal{F})$ is its set of minterms.

Our goal in this section is to show that if $\mathcal{F}$ is a monotone t-intersecting family, $p < (r+1)/(t+2r+1)$ and $m(\mathcal{F}) > t+2r$ then there is another t-intersecting family $\mathcal{G}$ with $\mu_p(\mathcal{G}) > \mu_p(\mathcal{F})$. We will construct $\mathcal{G}$ by modifying the generating set of $\mathcal{F}$, guided by the following easy lemma.

Lemma 5.6. *Let $\mathcal{F}$ be a left-compressed t-intersecting family with $m = m(\mathcal{F})$, and suppose that $A, B \in \mathcal{F}$ both contain m. If $|A \cap B| = t$ then $A \cup B = [m]$ and so $|A| + |B| = m + t$.*

Proof. Let $A, B \in \mathcal{F}$ be as indicated. Clearly $A \cup B \subseteq [m]$. Suppose that for some $i \in [m]$, $i \notin A \cup B$. By assumption, $i < m$. Since $\mathcal{F}$ is left-compressed, $A' = A \setminus \{m\} \cup \{i\} \in \mathcal{F}$. However, $|A' \cap B| = |A \cap B| - 1 = t - 1$, contradicting the assumption that $\mathcal{F}$ is t-intersecting. We conclude that $A \cup B = [m]$ and so $|A| + |B| = |A \cup B| + |A \cap B| = m + t$. □

This lemma suggests separating the sets in $G(\mathcal{F})$ containing m according to their size.

Definition 5.5. Let $\mathcal{F}$ be a family of sets with $m = m(\mathcal{F})$. We define $G^*(\mathcal{F}) = \{A \in G(\mathcal{F}) : m \in A\}$ and $G^*_a(\mathcal{F}) = \{A \in G^*(\mathcal{F}) : |A| = a\}$. In words, $G^*(\mathcal{F})$ consists of those sets in $G(\mathcal{F})$ containing m, and $G^*_a(\mathcal{F})$ consists of those sets in $G(\mathcal{F})$ containing m and of size a.

For a family G on n points and $m \in [n]$, we define $G \setminus m = \{A \setminus \{m\} : A \in G\}$. ○

Suppose $a + b = m(\mathcal{F}) + t$ and $a \neq b$. Lemma 5.6 implies that $U^n\big(G(\mathcal{F}) \setminus (G^*_a(\mathcal{F}) \cup G^*_b(\mathcal{F})) \cup (G^*_a(\mathcal{F}) \setminus m(\mathcal{F}))\big)$ is t-intersecting. Moreover, it turns out that this transformation can be used to increase the μ_p-measure.

We start by proving two easy auxiliary results.

Lemma 5.7. *Let $\mathcal{F}$ be a monotone left-compressed family on n points with $m = m(\mathcal{F})$ and let $A \in G^*(\mathcal{F})$. Then*

$$\mathcal{F} \setminus U^n(G(\mathcal{F}) \setminus \{A\}) = \{A\} \times 2^{[n]\setminus[m]}.$$

In words, if $A \in G^*(\mathcal{F})$ then the sets generated by A are exactly $\{A\} \times 2^{[n]\setminus[m]}$.

Proof. Suppose $B \in \mathcal{F} \setminus U^n(G(\mathcal{F}) \setminus \{A\})$. Clearly $B \supseteq A$. We would like to show that $B \cap [m] = A$. If not, then let $x \in (B \cap [m]) \setminus A$. Since $\mathcal{F}$ is left-compressed, $C = \mathbb{S}_{x \leftarrow m}(A) \in \mathcal{F}$. Clearly $C \in U^n(G(\mathcal{F}) \setminus \{A\})$, and since $B \supseteq C$, also $B \in U^n(G(\mathcal{F}) \setminus \{A\})$, contrary to the assumption. Hence $B \cap [m] = A$.

For the other direction, let $B = A \cup C$, where $C \subseteq [n] \setminus [m]$. If $B \in U^n(G(\mathcal{F}) \setminus \{A\})$ then $B \supseteq D$ for some $D \in G(\mathcal{F}) \setminus \{A\}$. Since $\max D \le m$, necessarily $D \subseteq B \cap [m] = A$, contradicting the fact that A is inclusion-minimal. This completes the proof of the lemma. □

Lemma 5.8. *Let $\mathcal{F}$ be a family of sets on n points with $m = m(\mathcal{F})$ and let $A \in G^*(\mathcal{F})$. If $B \in \mathcal{F}$ and $B \cap [m-1] = A \setminus \{m\}$ then $m \in B$.*

Proof. Suppose that $m \notin B$. Since $B \in \mathcal{F}$, $B \supseteq C$ for some $C \in G(\mathcal{F})$. Since $\max C \le m$ and $m \notin B$, $C \subseteq B \cap [m] = A \setminus \{m\}$, contradicting the fact that A is inclusion-minimal. □

Next, we describe the transformation itself.

Lemma 5.9. *Let $\mathcal{F}$ be a monotone left-compressed t-intersecting family on n points with $m = m(\mathcal{F})$, and let $a + b = m + t$ for some non-negative integers $a \ne b$. Define*

$$H_a = G(\mathcal{F}) \setminus (G_a^*(\mathcal{F}) \cup G_b^*(\mathcal{F})) \cup (G_a^*(\mathcal{F}) \setminus m), \qquad \mathcal{G}_a = U^n(H_a),$$

$$H_b = G(\mathcal{F}) \setminus (G_a^*(\mathcal{F}) \cup G_b^*(\mathcal{F})) \cup (G_b^*(\mathcal{F}) \setminus m), \qquad \mathcal{G}_b = U^n(H_b).$$

The families $\mathcal{G}_a, \mathcal{G}_b$ are t-intersecting. Furthermore, if $G_a^*(\mathcal{F}) \ne \varnothing$ or $G_b^*(\mathcal{F}) \ne \varnothing$ then for all $p < 1/2$, $\max(\mu_p(\mathcal{G}_a), \mu_p(\mathcal{G}_b)) > \mu_p(\mathcal{F})$.

Proof. In order to show that $\mathcal{G}_a$ is t-intersecting, it is enough to show that H_a is t-intersecting. Let $A, B \in H_a$. If $A, B \notin G_a^*(\mathcal{F}) \setminus m$ then $A, B \in G(\mathcal{F})$ and so $|A \cap B| \ge t$, so suppose that $A \in G_a^*(\mathcal{F}) \setminus m$. Notice that $A \cup \{m\} \in G_a^*(\mathcal{F})$. If $B \notin G^*(\mathcal{F})$ then $m \notin B$ and $B \in G(\mathcal{F})$, and so $|A \cap B| = |(A \cup \{m\}) \cap B| \ge t$. If $B \in G_c^*(\mathcal{F})$ then $c \ne b$ and so $|A \cup \{m\}| + |B| = a + c \ne a + b = m + t$. Therefore Lemma 5.6 implies that $|(A \cup \{m\}) \cap B| \ge t + 1$ and so $|A \cap B| \ge t$. A similar argument applies if $B \in G_a^*(\mathcal{F}) \setminus \{m\}$ (with a in place of c), and we conclude that $\mathcal{G}_A$ is t-intersecting. The proofs for $\mathcal{G}_b$ are analogous.

Let $p < 1/2$. We proceed to calculate $\mu_p(\mathcal{G}_a)$ and $\mu_p(\mathcal{G}_b)$. Lemma 5.7 shows that

$$\mathcal{F} \setminus \mathcal{G}_a = G_b^*(\mathcal{F}) \times 2^{[n] \setminus [m]},$$

and Lemma 5.8 shows that

$$\mathcal{G}_a \setminus \mathcal{F} = (G_a^*(\mathcal{F}) \setminus m) \times 2^{[n] \setminus [m]}.$$

Therefore

$$\begin{aligned}\mu_p(\mathcal{G}_a) &= \mu_p(\mathcal{F}) - \mu_p^{[m]}(G_b^*(\mathcal{F})) + \mu_p^{[m]}(G_a^*(\mathcal{F}) \setminus m) \\ &= \mu_p(\mathcal{F}) - \mu_p^{[m]}(G_b^*(\mathcal{F})) + \frac{1-p}{p}\mu_p^{[m]}(G_a^*(\mathcal{F})).\end{aligned}$$

Without loss of generality, suppose that $\mu_p^{[m]}(G_a^*(\mathcal{F})) \geq \mu_p^{[m]}(G_b^*(\mathcal{F}))$, which implies that $\mu_p^{[m]}(G_a^*(\mathcal{F})) > 0$ by assumption. Then

$$\mu_p(\mathcal{G}_a) \geq \mu_p(\mathcal{F}) + \left(\frac{1-p}{p} - 1\right)\mu_p^{[m]}(G_a^*(\mathcal{F})) > 0,$$

since $p < 1/2$ implies $(1-p)/p > 1$. □

This lemma allows us to achieve our goal whenever $G_a^*(\mathcal{F}) \neq \varnothing$ for some $a \neq (m(\mathcal{F}) + t)/2$. When $a = (m(\mathcal{F}) + t)/2$, the construction in the lemma doesn't result in a t-intersecting family. In order to fix the construction, we will focus on a subset of $G_a^*(\mathcal{F})$ *not* containing some common element. This property will guarantee that the result is t-intersecting. If p is small enough (depending on $m(\mathcal{F})$), then the construction still increases the μ_p-measure.

Lemma 5.10. *Let $\mathcal{F}$ be a monotone left-compressed t-intersecting family on n points with $m = m(\mathcal{F}) > t + 2r$ for some $r \geq 0$, and let $a = (m+t)/2$ be integral. For $i \in [m-1]$, define*

$$H_i = G(\mathcal{F}) \setminus G_a^*(\mathcal{F}) \cup \{A \in G_a^*(\mathcal{F}) \setminus m : i \notin A\}, \qquad \mathcal{G}_i = U^n(H_i).$$

The families $\mathcal{G}_i$ are t-intersecting. Furthermore, if $p < (r+1)/(t+2r+1)$ and $G_a^*(\mathcal{F}) \neq \varnothing$ then $\max_{i \in [m-1]} \mu_p(\mathcal{G}_i) > \mu_p(\mathcal{F})$.

Proof. Let $i \in [m-1]$. We proceed to show that $\mathcal{G}_i$ is t-intersecting. As in the proof of the corresponding part of Lemma 5.9, it is enough to show that H_i is t-intersecting. If $A, B \in H_i$ and not both $A, B \in G_a^*(\mathcal{F}) \setminus m$ then the argument in Lemma 5.9 shows that $|A \cap B| \geq t$, so suppose that $A, B \in G_a^*(\mathcal{F}) \setminus m$. Note that $i \notin A, B$. Lemma 5.6 shows that $|(A \cup \{m\}) \cap (B \cup \{m\})| > t$, and so $|A \cap B| \geq t$, unless $(A \cup \{m\}) \cup (B \cup \{m\}) = [m]$. However, the latter is impossible since $i \notin A \cup B$. This shows that $\mathcal{G}_i$ is t-intersecting.

Let $K_i = \{A \in G_a^*(\mathcal{F}) : i \notin A\}$. We proceed to calculate $\mu_p(\mathcal{G}_i)$. Lemma 5.7 shows that

$$\mathcal{F} \setminus \mathcal{G}_i = (G_a^*(\mathcal{F}) \setminus K_i) \times 2^{[n] \setminus [m]},$$

and Lemma 5.8 shows that

$$\mathcal{G}_i \setminus \mathcal{F} = (K_i \setminus m) \times 2^{[n]\setminus[m]}.$$

Therefore

$$\begin{aligned}\mu_p(\mathcal{G}_i) &= \mu_p(\mathcal{F}) - \mu_p^{[m]}(G_a^*(\mathcal{F}) \setminus K_i) + \frac{1-p}{p}\mu_p^{[m]}(K_i) \\ &= \mu_p(\mathcal{F}) - \mu_p^{[m]}(G_a^*(\mathcal{F})) + \frac{1}{p}\mu_p^{[m]}(K_i).\end{aligned} \tag{5.1}$$

In view of this, we would like to maximize $\mu_p^{[m]}(K_i)$. Since all sets in K_i have Hamming weight a, $\mu_p^{[m]}(K_i) = |K_i|p^a(1-p)^{m-a}$, and similarly $\mu_p^{[m]}(G_a^*(\mathcal{F})) = |G_a^*(\mathcal{F})|p^a(1-p)^{m-a}$. We therefore want to maximize $|K_i|$. Since each $A \in G_a^*(\mathcal{F})$ satisfies $|A \cap [m-1]| = a-1$, it is easy to see that

$$\mathop{\mathbb{E}}_{i\in[m-1]}|K_i| = \frac{m-a}{m-1}|G_a^*(\mathcal{F})|.$$

There must be some $i \in [m-1]$ which satisfies $|K_i| \geq (m-a)/(m-1)\cdot|G_a^*(\mathcal{F})|$, and so $\mu_p^{[m]}(K_i) \geq (m-a)/(m-1)\cdot\mu_p^{[m]}(G_a^*(\mathcal{F}))$. Substituting this in (5.1), we obtain

$$\begin{aligned}\mu_p(\mathcal{G}_i) - \mu_p(\mathcal{F}) &\geq \left(\frac{1}{p}\cdot\frac{m-a}{m-1} - 1\right)\mu_p^{[m]}(G_a^*(\mathcal{F})) \\ &= \frac{m-a-p(m-1)}{p(m-1)}\mu_p^{[m]}(G_a^*(\mathcal{F})).\end{aligned}$$

The proof will be complete if we show that $m-a > p(m-1)$. Since $m > t+2r$ and $m+t$ is even, $m \geq t+2r+2$, and so

$$\begin{aligned}2[m-a-p(m-1)] &= m-t-2p(m-1) \\ &= (1-2p)m - t + 2p \\ &\geq (1-2p)(t+2r+2) - t + 2p \\ &= 2r+2-2p(t+2r+1) \\ &= 2[r+1-p(t+2r+1)] > 0.\end{aligned}$$

□

Combining Lemma 5.9 and Lemma 5.10, we obtain the following result.

Lemma 5.11. *Let $\mathcal{F}$ be a monotone left-compressed t-intersecting family on n points with $m = m(\mathcal{F}) > t+2r$ for some $r \geq 0$. If $p < (r+1)/(t+2r+1)$ then there exists a t-intersecting family $\mathcal{G}$ on n points satisfying $\mu_p(\mathcal{G}) > \mu_p(\mathcal{F})$.*

Proof. By definition, $G^*(\mathcal{F}) \neq \varnothing$, and so $G_a^*(\mathcal{F}) \neq \varnothing$ for some a. If $a \neq (m+t)/2$ then the result follows from Lemma 5.9, otherwise it follows from Lemma 5.10. $\square$

We can conclude an important corollary.

Corollary 5.12. *Let $t \geq 1$, $r \geq 0$ and $p < (r+1)/(t+2r+1)$. There exists a monotone left-compressed t-intersecting family $\mathcal{F}$ on $t+2r$ points such that for every t-intersecting family $\mathcal{G}$, $\mu_p(\mathcal{G}) \leq \mu_p(\mathcal{F})$. Furthermore, equality is only possible if $m(\mathcal{G}) \leq t+2r$.*

Proof. Lemma 5.5 implies that it is enough to construct a (not necessarily left-compressed) t-intersecting family $\mathcal{F}$ on $t+2r$ points. We let $\mathcal{F}$ be a t-intersecting family of maximal μ_p-measure among those on $t+2r$ points.

Now let $\mathcal{G}$ be a t-intersecting family on n points. In order to show that $\mu_p(\mathcal{G}) \leq \mu_p(\mathcal{F})$, we can assume that $\mathcal{G}$ has maximal μ_p-measure among t-intersecting families on n points. Lemma 5.11 implies that $m(\mathcal{G}) \leq t+2r$, and so $\mu_p(\mathcal{G}) \leq \mu_p(\mathcal{F})$ by definition. The lemma also implies that equality is only possible if $m(\mathcal{G}) \leq t+2r$. $\square$

At this point, [2] considers the complemented family $\bar{\mathcal{F}} = \{[n] \setminus A : A \in \mathcal{F}\}$. When $\mathcal{F}$ is a k-uniform t-intersecting family, $\bar{\mathcal{F}}$ is an $(n-k)$-uniform $(n-2k+t)$-intersecting family, and we can apply Corollary 5.12 to $\bar{\mathcal{F}}$. However, in our setting $\bar{\mathcal{F}}$ need not even be intersecting. Instead, we turn to the argument in [3].

5.4 Pushing-pulling

In this section we implement the second step of the proof, following [3]. We will show that if $\mathcal{F}$ is a left-compressed t-intersecting family of maximal μ_p-measure, where $p > r/(t+2r-1)$, then the first $t+2r$ coordinates of $\mathcal{F}$ are symmetric. We start by formalizing the notion of symmetry.

Definition 5.6. A family of sets $\mathcal{F}$ is *ℓ-invariant* if for all $i \neq j$ in the range $1 \leq i,j \leq \ell$, $\mathbb{S}_{i\leftarrow j}(\mathcal{F}) = \mathcal{F}$.

The *symmetric extent* $\ell(\mathcal{F})$ of a family of sets $\mathcal{F}$ on n points is the maximal $\ell \leq n$ such that $\mathcal{F}$ is ℓ-invariant. $\bigcirc$

Our goal in this section is to show that if $p > r/(t+2r-1)$ and $\ell(\mathcal{F}) < t+2r$ for some t-intersecting family $\mathcal{F}$ then we can come up with a t-intersecting family of larger μ_p-measure.

Since we are focusing on left-compressed families, the only way in which ℓ-invariance can fail is if $\mathbb{S}_{\ell\leftarrow i}(\mathcal{F}) \neq \mathcal{F}$ for some $i < \ell$. The following definition singles out the sets which determine the symmetric extent of a family.

Definition 5.7. Let $\mathcal{F}$ be a family of sets on n points with $\ell = \ell(\mathcal{F})$. If $n > \ell$ then its *boundary sets* are given by

$$X(\mathcal{F}) = \{A \in \mathcal{F} : \mathbb{S}_{\ell+1\leftarrow i}(A) \notin \mathcal{F} \text{ for some } i \leq \ell\}.$$

if $n = \ell$ then we define $X(\mathcal{F}) = \varnothing$. ○

Our starting point is the following analog of Lemma 5.6.

Lemma 5.13. *Let $\mathcal{F}$ be a left-compressed t-intersecting family with $\ell = \ell(\mathcal{F})$, and let $A, B \in X(\mathcal{F})$. If $|A \cap B| = t$ then $A \cap B \subseteq [\ell]$ and $A \cup B \supseteq [\ell]$, and so $|A \cap [\ell]| + |B \cap [\ell]| = \ell + t$.*

Proof. Let $A, B \in X(\mathcal{F})$ be as given, and note that $\ell + 1 \notin A, B$. We start by showing that $A \cap B \subseteq [\ell]$. Suppose that $x \in A \cap B$ satisfies $x > \ell$. Since $\ell + 1 \notin A, B$, in fact $x > \ell + 1$. Since $\mathcal{F}$ is left-compressed, $\mathbb{S}_{\ell+1\leftarrow x}(A) \in \mathcal{F}$. However, $|\mathbb{S}_{\ell+1\leftarrow x}(A) \cap B| = |A \cap B| - 1 = t - 1$, contrary to assumption. We conclude that $A \cap B \subseteq [\ell]$.

Next, we show that $A \cup B \supseteq [\ell]$. Suppose that $x \notin A \cup B$ for some $x \in [\ell]$. Since $t \geq 1$ and $A \cap B \subseteq [\ell]$, there is some $y \in A \cap B \cap [\ell]$. Since $\mathcal{F}$ is ℓ-invariant, $\mathbb{S}_{x\leftarrow y}(A) \in \mathcal{F}$. However, $|\mathbb{S}_{x\leftarrow y}(A) \cap B| = |A \cap B| - 1 = t - 1$, contrary to assumption. We conclude that $A \cup B \supseteq [\ell]$.

Finally, let $A' = A \cap [\ell]$ and $B' = B \cap [\ell]$. We have $A' \cup B' = [\ell]$ and $|A' \cap B'| = |A \cap B| = t$, and so $|A'| + |B'| = |A' \cup B'| + |A' \cap B'| = \ell + t$. □

This suggests breaking down $X(\mathcal{F})$ according to the size of the intersection with $[\ell]$.

Definition 5.8. Let $\mathcal{F}$ be a family of sets on n points with $\ell = \ell(\mathcal{F})$. Its ith *boundary marginal* is given by

$$X_i(\mathcal{F}) = \{B \subseteq [n] \setminus [\ell+1] : [i] \cup B \in X(\mathcal{F})\}. \quad ○$$

The part played by the sets $[i]$ is arbitrary. Indeed, we have the following easy lemma.

Definition 5.9. For a set X and an integer i, we define

$$\binom{X}{i} = \{A \subseteq X : |A| = i\}. \qquad \bigcirc$$

Lemma 5.14. *Let $\mathcal{F}$ be a family of sets on n points with $\ell = \ell(\mathcal{F})$. Then*

$$X(\mathcal{F}) = \bigcup_{i=1}^{\ell} \binom{[\ell]}{i} \times X_i(\mathcal{F}).$$

Proof. If $A \in X(\mathcal{F})$ then $\mathbb{S}_{\ell+1\leftarrow i}(A) \neq A$ for some $i \leq \ell$, and in particular $i \in A$. This shows that $X_0(\mathcal{F}) = \varnothing$. Also, clearly $\ell + 1 \notin A$ for all $A \in X(\mathcal{F})$. The lemma now follows directly from the ℓ-equivalence of $\mathcal{F}$. □

We now present two different constructions that attempt to increase the μ_p-measure of a t-intersecting family. The first construction is the counterpart of Lemma 5.9.

Lemma 5.15. *Let $\mathcal{F}$ be a t-intersecting left-compressed family on n points with $\ell = \ell(\mathcal{F})$, and let $a + b = \ell + t$ for some non-negative integers $a \neq b$. Define*

$$\mathcal{G}_a = \mathcal{F} \setminus \binom{[\ell]}{b} \times X_b(\mathcal{F}) \cup \binom{[\ell]}{a-1} \times \{\ell+1\} \times X_a(\mathcal{F}),$$
$$\mathcal{G}_b = \mathcal{F} \setminus \binom{[\ell]}{a} \times X_a(\mathcal{F}) \cup \binom{[\ell]}{b-1} \times \{\ell+1\} \times X_b(\mathcal{F}).$$

The families $\mathcal{G}_a, \mathcal{G}_b$ are t-intersecting. Furthermore, if $G_a^*(\mathcal{F}) \neq \varnothing$ or $G_b^*(\mathcal{F}) \neq \varnothing$ and $t \geq 2$ then for all $p \in (0,1)$, $\max(\mu_p(\mathcal{G}_a), \mu_p(\mathcal{G}_b)) > \mu_p(\mathcal{F})$.

Proof. We start by showing that $\mathcal{G}_a$ is t-intersecting. Let $A, B \in \mathcal{G}_a$. If $A, B \notin \binom{[\ell]}{a-1} \times \{\ell+1\} \times X_a(\mathcal{F})$ then $A, B \in \mathcal{F}$ and so $|A \cap B| \geq t$, so assume that $A \in \binom{[\ell]}{a-1} \times \{\ell+1\} \times X_a(\mathcal{F})$. Pick some $x \in [\ell]$ such that $x \notin A$, and notice that $A' = A \setminus \{\ell+1\} \cup \{x\} \in \mathcal{F}$.

Suppose first that $B \in \mathcal{F}$. If $\ell + 1 \in B$ or $x \notin B$ then $|A \cap B| \geq |A' \cap B| \geq t$, so suppose that $\ell + 1 \notin B$ and $x \in B$. If $B' = \mathbb{S}_{\ell+1\leftarrow x}(B) \in \mathcal{F}$ then $|A \cap B| = |A' \cap B'| \geq t$. Otherwise, $B \in X(\mathcal{F})$ and since $\ell + 1 \notin B$, $|B \cap [\ell]| \neq b$. Since $|A' \cap [\ell]| = a$, $|A' \cap [\ell]| + |B \cap [\ell]| \neq a + b = \ell + t$, and so Lemma 5.13 shows that $|A' \cap B| \geq t + 1$, which implies $|A \cap B| \geq |A' \cap B| - 1 \geq t$.

Finally, suppose that $A, B \notin \mathcal{F}$. Pick some $y \in [\ell]$ such that $y \notin B$, and notice that $B' = B \setminus \{\ell+1\} \cup \{y\} \in \mathcal{F}$. Since $|A' \cap [\ell]| + |B' \cap [\ell]| = 2a \neq a + b = \ell + t$, Lemma 5.13 shows that $|A' \cap B'| \geq t + 1$. Therefore $|A \cap B| = |[(A' \setminus \{x\}) \cap (B' \setminus \{y\})] \cup \{\ell+1\}| \geq t$. We conclude that $\mathcal{G}_a$ is t-intersecting.

It is straightforward to compute the μ_p-measures of $\mathcal{G}_a$ and $\mathcal{G}_b$:

$$\mu_p(\mathcal{G}_a) = \mu_p(\mathcal{F}) - \binom{\ell}{b} p^b (1-p)^{\ell+1-b} \mu_p^{[n]\setminus[\ell+1]}(X_b(\mathcal{F})) + \binom{\ell}{a-1} p^a (1-p)^{\ell+1-a} \mu_p^{[n]\setminus[\ell+1]}(X_a(\mathcal{F})),$$
$$\mu_p(\mathcal{G}_b) = \mu_p(\mathcal{F}) - \binom{\ell}{a} p^a (1-p)^{\ell+1-a} \mu_p^{[n]\setminus[\ell+1]}(X_a(\mathcal{F})) + \binom{\ell}{b-1} p^b (1-p)^{\ell+1-b} \mu_p^{[n]\setminus[\ell+1]}(X_b(\mathcal{F})).$$

These formulas become simpler if we put

$$\gamma_a = \binom{\ell}{a} p^a (1-p)^{\ell+1-a} \mu_p^{[n]\setminus[\ell+1]}(X_a(\mathcal{F})), \quad \gamma_b = \binom{\ell}{b} p^b (1-p)^{\ell+1-b} \mu_p^{[n]\setminus[\ell+1]}(X_b(\mathcal{F})).$$

By assumption, either $\gamma_a > 0$ or $\gamma_b > 0$. Substituting these variables, we get

$$\mu_p(\mathcal{G}_a) = \mu_p(\mathcal{F}) - \gamma_a + \frac{a}{\ell - a + 1}\gamma_b, \quad \mu_p(\mathcal{G}_b) = \mu_p(\mathcal{F}) - \gamma_b + \frac{b}{\ell - b + 1}\gamma_a.$$

Multiply the first equation by $\ell - a + 1$, the second equation by $\ell - b + 1$, and sum to get

$$(\ell-a+1)(\mu_p(\mathcal{G}_a)-\mu_p(\mathcal{F}))+(\ell-b+1)(\mu_p(\mathcal{G}_b)-\mu_p(\mathcal{F})) = (a+b-\ell-1)(\gamma_a+\gamma_b) = (t-1)(\gamma_a+\gamma_b) > 0.$$

We conclude that either $\mu_p(\mathcal{G}_a) > \mu_p(\mathcal{F})$ or $\mu_p(\mathcal{G}_b) > \mu_p(\mathcal{F})$. □

The second construction, which is the counterpart of Lemma 5.10, concerns $a = (\ell + t)/2$, and works by adjoining a new element, which ensures that the resulting family is t-intersecting.

Lemma 5.16. *Let $\mathcal{F}$ be a t-intersecting left-compressed family on n points with $\ell = \ell(\mathcal{F})$, and let $a = (\ell + t)/2$ be integral. Define*

$$\mathcal{G} = \mathcal{F} \setminus \binom{[\ell]}{a} \times X_a(\mathcal{F}) \times 2^{\{n+1\}} \cup \binom{[\ell+1]}{a} \times X_a(\mathcal{F}) \times \{n+1\}.$$

Note that $\mathcal{G}$ is a family on $n+1$ points. The family $\mathcal{G}$ is t-intersecting. Moreover, if $X_a(\mathcal{F}) \neq \varnothing$, $t \geq 2$ and $r/(t+2r-1) < p < 1/2$, $\ell < t + 2r$ for some $r \geq 0$, then $\mu_p(\mathcal{G}) > \mu_p(\mathcal{F})$.

Proof. Put $\mathcal{F}' = \mathcal{G} \times 2^{\{n+1\}}$, and note that $\mathcal{F}'$ is t-intersecting and $\mu_p(\mathcal{F}') = \mu_p(\mathcal{F})$. We start by showing that $\mathcal{G}$ is t-intersecting. Let $A, B \in \mathcal{G}$. If $A, B \in \mathcal{F}'$ then clearly $|A \cap B| \geq t$, so suppose that $A \in \binom{[\ell+1]}{a} \times X_a(\mathcal{F}) \times \{n+1\}$ and $\ell + 1 \in A$. Pick some $x \in [\ell]$ such that $x \notin A$, and notice that $A' = A \setminus \{\ell+1, n+1\} \cup \{x\} \in \mathcal{F}'$.

Suppose first that $B \in \mathcal{F}'$. If $\ell + 1 \in B$ or $x \notin B$ then $|A \cap B| \geq |A' \cap B| \geq t$, so suppose that $\ell + 1 \notin B$ and $x \in B$. If $B' = \mathbb{S}_{\ell+1\leftarrow x}(B) \in \mathcal{F}'$ then $|A \cap B| = |A' \cap B'| \geq t$. Otherwise, $B \in X(\mathcal{F}')$.

We distinguish between two cases. If $|B \cap [\ell]| \neq a$ then $|A' \cap [\ell]| + |B \cap [\ell]| \neq 2a = \ell + t$, and so Lemma 5.13 shows that $|A' \cap B| \geq t+1$, which implies $|A \cap B| \geq |A' \cap B| - 1 \geq t$. If $|B \cap [\ell]| = a$ then necessarily $n+1 \in B$, and so $B' = B \setminus \{n+1\} \in \mathcal{F}'$. Therefore $|A \cap B| \geq |[(A' \setminus \{x\}) \cap B'] \cup \{n+1\}| \geq |A' \cap B'| \geq t$.

Finally, suppose that $A, B \notin \mathcal{F}'$. Pick some $y \in [\ell]$ such that $y \notin B$, and notice that $B' = B \setminus \{\ell+1, n+1\} \cup \{y\} \in \mathcal{F}'$. We have $|A \cap B| = |[(A' \setminus \{x\}) \cap (B' \setminus \{y\})] \cup \{\ell+1, n+1\}| \geq |A' \cap B'| \geq t$. We conclude that $\mathcal{G}$ is t-intersecting.

It is straightforward to compute the μ_p-measure of $\mathcal{G}$:

$$\begin{aligned}\mu_p(\mathcal{G}) &= \mu_p(\mathcal{F}) - \binom{\ell}{a} p^a (1-p)^{\ell-a+1} \mu_p^{[n]\setminus[\ell+1]}(X_a(\mathcal{F})) + \binom{\ell+1}{a} p^{a+1} (1-p)^{\ell-a+1} \mu_p^{[n]\setminus[\ell+1]}(X_a(\mathcal{F})) \\ &= \mu_p(\mathcal{F}) + \left(-1 + \frac{\ell+1}{\ell-a+1} p\right) \binom{\ell}{a} p^a (1-p)^{\ell-a+1} \mu_p^{[n]\setminus[\ell+1]}(X_a(\mathcal{F})).\end{aligned}$$

Since $X_a(\mathcal{F}) \neq \varnothing$, in order to complete the proof we need to show that the expression inside the parentheses is positive. Since $\ell < t + 2r$ and $\ell + t$ is even, $\ell \leq t + 2r - 2$. Clearly $a \leq \ell$ and so $\ell - a + 1 > 0$, hence the parenthesized expression is positive if the following expression is:

$$\begin{aligned}2[(\ell+1)p - (\ell - a + 1)] &= 2a - 2(1-p)(\ell+1) \\ &= t - 1 - (1-2p)(\ell+1) \\ &\geq t - 1 - (1-2p)(t + 2r - 1) \\ &= 2p(t + 2r - 1) - 2r > 0,\end{aligned}$$

using in the third line the assumption $p < 1/2$. □

Combining Lemma 5.15 and Lemma 5.16, we obtain the following result.

Lemma 5.17. *Let $\mathcal{F}$ be a left-compressed t-intersecting family on n points with $\ell = \ell(\mathcal{F}) < t+2r$ for some $r \geq 0$. If $t \geq 2$ and $r/(t+2r-1) < p < 1/2$ then there exists a t-intersecting family $\mathcal{G}$ on $n+1$ points satisfying $\mu_p(\mathcal{G}) > \mu_p(\mathcal{F})$.*

Proof. By definition, $X(\mathcal{F}) \neq \varnothing$, and so $X_a(\mathcal{F}) \neq \varnothing$ for some a. If $a \neq (\ell + t)/2$ then the result follows from Lemma 5.15, otherwise it follows from Lemma 5.16. □

Combining this result with Corollary 5.12, we can prove the Ahlswede–Khachatrian theorem for left-compressed families.

Theorem 5.18. *Let $\mathcal{F}$ be a left-compressed t-intersecting family on n points for $t \geq 2$. If $r/(t+2r-1) < p < (r+1)/(t+2r+1)$ for some $r \geq 0$ then $\mu_p(\mathcal{F}) \leq \mu_p(\mathcal{F}_{t,r})$, with equality if and only if $\mathcal{F} = U^n(\mathcal{F}_{t,r})$.*

If $p = (r+1)/(t+2r+1)$ for some $r \geq 0$ then $\mu_p(\mathcal{F}) \leq \mu_p(\mathcal{F}_{t,r}) = \mu_p(\mathcal{F}_{t,r+1})$, with equality if and only if either $\mathcal{F} = U^n(\mathcal{F}_{t,r})$ or $\mathcal{F} = U^n(\mathcal{F}_{t,r+1})$.

Proof. Suppose first that $r/(t+2r-1) < p < (r+1)/(t+2r+1)$ for some $r \geq 0$. Corollary 5.12 gives a monotone left-compressed t-intersecting family $\mathcal{F}^*$ on $t+2r$ points such that $\mu_p(\mathcal{F}) \leq \mu_p(\mathcal{F}^*)$, with equality only if $m(\mathcal{F}) \leq t+2r$. Lemma 5.17 shows that $\ell(\mathcal{F}^*) = t+2r$, and so $\mathcal{F}^*$ must be of the form

$$\mathcal{F}^*_s = \{A \subseteq [t+2r] : |A| \geq s\}$$

for some s. This family is t-intersecting for $s \geq t+r$, and the optimal choice $s = t+r$ shows that $\mathcal{F}^* = \mathcal{F}^*_{t+r} = \mathcal{F}_{t,r}$. The corollary and the lemma together show that $\mu_p(\mathcal{F}) = \mu_p(\mathcal{F}^*)$ is only possible if $m(\mathcal{F}) = \ell(\mathcal{F}) = t+2r$, and so $\mathcal{F} = U^n(\mathcal{F}^*_s)$ for some s. This readily implies that $\mathcal{F} = U^n(\mathcal{F}^*)$.

Suppose next that $p = (r+1)/(t+2r+1)$ for some $r \geq 0$. Corollary 5.12 gives a monotone left-compressed t-intersecting family $\mathcal{F}^*$ on $t+2r+2$ points such that $\mu_p(\mathcal{F}) \leq \mu_p(\mathcal{F}^*)$, with equality only if $m(\mathcal{F}) \leq t+2r+2$. Since μ_p is continuous and there are finitely many families on $t+2r+2$ points, we see that $\mu_p(\mathcal{F}^*) = \mu_p(\mathcal{F}_{t,r}) = \mu_p(\mathcal{F}_{t,r+1})$. Corollary 5.12 and Lemma 5.17 show that $\mu_p(\mathcal{F}) = \mu_p(\mathcal{F}^*)$ is only possible if $m(\mathcal{F}) \leq t+2r+2$ and $\ell(\mathcal{F}) \geq t+2r$. Assume for simplicity that $n = t+2r+2$. The family $\mathcal{F}$ has the following general form:

$$\mathcal{F} = \mathcal{F}^*_a \cup \mathcal{F}^*_b \times \{t+2r+1\} \cup \mathcal{F}^*_c \times \{t+2r+2\} \cup \mathcal{F}^*_d \times \{t+2r+1, t+2r+2\}.$$

Some of these parts may be missing, in which case we use $\mathcal{F}^*_\infty$. Since $\mathcal{F}$ is t-intersecting, $d \geq t+r-1$. If $d = t+r-1$ then since $\mathcal{F}$ is t-intersecting, $a \geq t+r+1$ and $b, c \geq t+r$. Therefore $\mathcal{F} \subseteq \mathcal{F}_{t,r+1}$ and so $\mathcal{F} = \mathcal{F}_{t,r+1}$. Otherwise, $d \geq t+r$, and so monotonicity shows that $a, b, c \geq t+r$. Therefore $\mathcal{F} \subseteq U^n(\mathcal{F}_{t,r})$ and so $\mathcal{F} = U^n(\mathcal{F}_{t,r})$. □

5.5 Culmination of the proof

Combined with Lemma 5.5, Theorem 5.18 already provides a tight upper bound on the μ_p-measure of arbitrary t-intersecting families. In order to complete the proof of the Ahlswede–Khachatrian theorem, it remains to prove uniqueness.

Recall that two families $\mathcal{F}, \mathcal{G}$ on n points are *equivalent* if they differ by a permutation of the coordinates. We start by showing that the families $\mathcal{F}_{t,r}$ are resilient to shifting in the case of t-intersecting families, using an argument from [2]. We need a preparatory lemma.

Lemma 5.19. *Let $t, r \geq 0$, and consider the following graph. The vertices are subsets of $[t+2r]$ of size $[t+r]$. Two subsets A, B are connected if $|A \cap B| = t$ (note that $|A \cap B| \geq t$). Then the graph is connected.*

Proof. If $r = 0$ then the graph contains a single vertex and there is nothing to prove, so suppose $r \geq 1$. We start by showing that $A = [t+r]$ and $B = [t+r]\Delta\{1, t+r+1\} = \{2, \ldots, t+r+1\}$ are connected. Let $C = [t] \cup \{t+r+1, \ldots, t+2r\}$. Then

$$|A \cap C| = |[t]| = t,$$
$$|B \cap C| = |\{2, \ldots, t\} \cup \{t+r+1\}| = t.$$

Hence A and B are connected via C. This shows that any two sets A, B with $|A\Delta B| = 2$ are connected, and so the graph is connected. □

Now we can prove the desired result on shifting.

Lemma 5.20. *Let $\mathcal{F}$ be a t-intersecting family on n points, and suppose that for some $i, j \in [n]$, $\mathbb{S}_{i\leftarrow j}(\mathcal{F})$ is equivalent to $\mathcal{F}_{t,r}$. Then $\mathcal{F}$ is equivalent to $\mathcal{F}_{t,r}$.*

Proof. We can assume that $\mathbb{S}_{i\leftarrow j}(\mathcal{F}) = U^n(\mathcal{F}_{t,r})$. If $j \in [t+2r]$ then since $\mathbb{S}_{i\leftarrow j}(\mathcal{F})$ depends only on the first $t+2r$ coordinates, necessarily $i \in [t+2r]$ and so $\mathbb{S}_{i\leftarrow j}(\mathcal{F}) = \mathcal{F}$. Similarly, if $i \notin [t+2r]$ then necessarily $j \notin [t+2r]$ and again $\mathbb{S}_{i\leftarrow j}(\mathcal{F}) = \mathcal{F}$. In both cases the lemma trivially holds. So without loss of generality, suppose that $n = t+2r+1$, $i = t+2r$ and $j = t+2r+1$. The following two subfamilies are involved in the shift:

$$\mathcal{F}_1 = \{A \in \mathcal{F} : j \in A, i \notin A, A\Delta\{i,j\} \notin \mathcal{F}\},$$

$$\mathcal{F}_2 = \{A \in \mathcal{F} : i \in A, j \notin A, A\Delta\{i,j\} \notin \mathcal{F}\}.$$

We have

$$\mathbb{S}_{i\leftarrow j}(\mathcal{F}) = \mathcal{F} \setminus \mathcal{F}_1 \cup \{A\Delta\{i,j\} : A \in \mathcal{F}_1\}.$$

If $\mathcal{F}_1 = \varnothing$ then $\mathbb{S}_{i\leftarrow j}(\mathcal{F}) = \mathcal{F}$, and the lemma clearly holds. If $\mathcal{F}_2 = \varnothing$ then $\mathbb{S}_{i\leftarrow j}(\mathcal{F})$ results from $\mathcal{F}$ by switching the coordinates i and j, and again the lemma holds. It remains to consider the case $\mathcal{F}_1, \mathcal{F}_2 \neq \varnothing$. Consider the family

$$\mathcal{G} = \{A \subseteq [t+2r-1] : |A| = t+r-1\}.$$

For every $A \in \mathcal{G}$, $A \cup \{i\} = A \cup \{t+2r\} \in \mathcal{F}_{t,r}$, and so either $A \cup \{i\} \in \mathcal{F}_2$ or $A \cup \{j\} \in \mathcal{F}_1$ (but not both). Form a graph whose vertices are the sets in $\mathcal{G}$, and two sets A, B are connected if $|A \cap B| = t-1$. Color a vertex A with 1 if $A \cup \{j\} \in \mathcal{F}_1$, and with 2 if $A \cup \{i\} \in \mathcal{F}_2$. Since $\mathcal{F}_1, \mathcal{F}_2 \neq \varnothing$, the coloring is not monochromatic. Lemma 5.19 shows that the graph is connected, and so there is some bichromatic edge (A, B), say $A' = A \cup \{j\} \in \mathcal{F}_1$ and $B' = B \cup \{i\} \in \mathcal{F}_2$. However, $|A' \cap B'| = |A \cap B| = t-1$, contradicting the fact that $\mathcal{F}$ is t-intersecting. We conclude that either $\mathcal{F}_1 = \varnothing$ or $\mathcal{F}_2 = \varnothing$. □

The Ahlswede–Khachatrian theorem is an easy corollary.

Theorem 5.3. *Let $\mathcal{F}$ be a t-intersecting family on n points for $t \geq 2$. If $r/(t+2r-1) < p < (r+1)/(t+2r+1)$ for some $r \geq 0$ then $\mu_p(\mathcal{F}) \leq \mu_p(\mathcal{F}_{t,r})$, with equality if and only if $\mathcal{F}$ is equivalent to $U^n(\mathcal{F}_{t,r})$.*

If $p = (r+1)/(t+2r+1)$ for some $r \geq 0$ then $\mu_p(\mathcal{F}) \leq \mu_p(\mathcal{F}_{t,r}) = \mu_p(\mathcal{F}_{t,r+1})$, with equality if and only if $\mathcal{F}$ is equivalent to either $U^n(\mathcal{F}_{t,r})$ or $U^n(\mathcal{F}_{t,r+1})$.

Proof. Let $\mathcal{G}$ be the left-compressed family satisfying $\mu_p(\mathcal{G}) = \mu_p(\mathcal{F})$ given by Lemma 5.5. Theorem 5.18 implies the upper bounds. Together with Lemma 5.20, the theorem implies the cases of equality. □

Chapter 6

Fourier analysis on the symmetric group

In this chapter we introduce Fourier analysis on the symmetric group, which is the group S_n of all permutations on $[n]$. Our goal in this chapter is to sketch the Fourier-theoretic proof of the Deza–Frankl conjecture [13] on intersecting families of permutations, due to Ellis, Friedgut and Pilpel [28]. These are families in which any two permutations agree on at least one point. Deza and Frankl proved that such families contain at most $(n-1)!$ permutations, and conjectured that the optimal families are of the form $\{\pi \in S_n : \pi(i) = j\}$ (we call these families *cosets*).

While the material in this chapter is not strictly used in the rest of the thesis, the chapter serves as a bridge between the first part of the thesis, *intersection theorems*, and the second part, *stability theorems*. The investigations in Chapter 7 and Chapter 8 were motivated by the work of Ellis, Friedgut and Pilpel, and the main theorem in Chapter 7 can be used to prove the stability part of their main result.

We start the chapter by highlighting in Section 6.1 some properties of the Fourier transform on $\{0,1\}^n$ that are exploited by Friedgut's method. The Fourier transform on S_n satisfies similar properties, and this explains why it is useful for extremal combinatorics. Section 6.2 describes the Fourier transform on S_n from a slightly unorthodox perspective. The more orthodox perspective is sketched in Section 6.3, in which we also prove that our description in Section 6.2 matches the classical description; to this effect, we use a method of Ellis [23]. The

Deza–Frankl conjecture and its Fourier-analytic proof are sketched in Section 6.4.

The only possibly original contribution of this chapter is the unorthodox presentation of the Fourier transform on S_n. The material on the Deza–Frankl conjecture is taken from Ellis, Friedgut and Pilpel [28], and the rest is classical representation theory.

For permutations $\alpha, \beta \in S_n$, we follow the convention that $\alpha\beta$ is the permutation resulting from applying β then α. For example, $(13)(12) = (123)$.

6.1 Friedgut's method and the symmetric group

Our proof of the Simonovits–Sós conjecture in Chapter 4 (for the usual μ measure) uses the following general plan. Start with a triangle-intersecting family of graphs $\mathcal{F}$ and its characteristic function $f = \mathbf{1}_{\mathcal{F}}$. Construct a matrix A such that $f'Af = 0$, whose eigenvectors are the Fourier characters. Conclude an equation of the form

$$\sum_G \lambda_G \hat{f}(G)^2 = 0,$$

and use Hoffman's bound to deduce a bound on $\mu(\mathcal{F})$. The last step uses the identities

$$\mu(\mathcal{F}) = \hat{f}(\varnothing) = \sum_G \hat{f}(G)^2,$$

which follow from two properties of the Fourier characters: $\chi_\varnothing$ is the constant 1 vector, and the Fourier characters are orthonormal. Hoffman's bound also implies that when $\mu(\mathcal{F}) = 1/8$, the Fourier expansion of $\mathcal{F}$ must be supported on the first four levels, and this implies that $\mathcal{F}$ is a $\triangle$-star.

The matrix A, in turn, is a linear combination of the matrices B_J described in Lemma 4.11 for various bipartite graphs J. Looking closely at the matrices B_J, we find that their effect on a vector v is shifting by $\overline{J}$:

$$(B_J v)_G = v_{G \oplus \overline{J}},$$

where $\oplus$ is the exclusive-or operation. Therefore

$$f' B_j f = |\mathcal{F} \cap (\mathcal{F} \oplus \overline{J})|, \quad \text{where } \mathcal{F} \oplus \overline{J} = \{G \oplus \overline{J} : G \in \mathcal{F}\}.$$

Since $\mathcal{F}$ is triangle-intersecting, whenever J is bipartite, the two families $\mathcal{F}$ and $\mathcal{F} \oplus \overline{J}$ are disjoint, since

$$G \cap (G \oplus \overline{J}) \subseteq G \nabla (G \oplus \overline{J}) = G \nabla (G \nabla J) = J.$$

(Recall that ∇ is the agreement operator, $G \nabla H = (G \cap H) \cup (\overline{G} \cap \overline{H})$.)

Summarizing, let us take stock of the features of the Fourier transform which are used in our proof:

- If $\mathcal{F}$ is triangle-intersecting then $\mathcal{F} \cap (\mathcal{F} \oplus \overline{J}) = \emptyset$ for every bipartite J, and the operator taking $\mathbf{1}_{\mathcal{F}}$ to $\mathbf{1}_{\mathcal{F} \oplus \overline{J}}$ has the Fourier basis vectors as eigenvectors (this is used to get the fundamental equation $f' A f = 0$).
- The Fourier basis is orthonormal, and the constant 1 vector is one of the basis vectors (this is used for the upper bound).
- If a Boolean function is supported on the first $k+1$ levels and has measure 2^{-k} then it is a k-star (this is used for uniqueness).
- If a Boolean function is concentrated on the first $k+1$ levels then it is close to a Boolean function depending on $O(1)$ coordinates (Kindler–Safra; this is used for stability).

The Fourier basis which we will construct for S_n will satisfy similar properties:

- If $\mathcal{F} \subseteq S_n$ is intersecting then $\pi\mathcal{F} = \{\pi\alpha : \alpha \in \mathcal{F}\}$ is disjoint from $\mathcal{F}$ whenever π is fixed-point free, and the operator taking $\mathbf{1}_{\mathcal{F}}$ to $\sum_{\pi \in \mathrm{FPF}_n} \mathbf{1}_{\pi\mathcal{F}}$ has the Fourier basis vectors as eigenvectors (here $\mathrm{FPF}_n \subseteq S_n$ is the set of all fixed-point free permutations).
- The Fourier basis for S_n is orthonormal, and the constant 1 vector is a basis vector.
- If a Boolean function on S_n is supported on the first two levels then it is the disjoint union of cosets.
- If a Boolean function on S_n is concentrated on the first two levels and has measure roughly $1/n$ then it is close to a coset.

The last property is the subject of Chapter 7.

6.2 Fourier analysis on the symmetric group

Our goal in this section is to construct a Fourier basis for S_n, which is an orthonormal basis for the space $\mathbb{R}[S_n]$ of all real functions on S_n. This basis will enjoy the properties listed in the previous section. We endow $\mathbb{R}[S_n]$ with the inner product

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

Our basis will be orthonormal with respect to this inner product.

Recall that for a function f on $\{0,1\}^n$, the kth level of the Fourier transform consists of those Fourier coefficients $\hat{f}(S)$ of size $|S| = k$.

Definition 6.1. A function f on $\{0,1\}^n$ is *k-simple* if its Fourier expansion is supported on the first $k+1$ levels.

A *k-semistar* is a family of sets on n points of the form $\tau_{A,B} = \{X \subseteq [n] : X \cap A = B\}$ where $|A| = k$ and $B \subseteq A$. ○

Lemma 6.1. *The space $L_k^{\{0,1\}^n}$ of k-simple functions on $\{0,1\}^n$ is spanned by the characteristic functions of all k-semistars.*

Proof. If f is an A-semistar for $|A| = k$ then Lemma 2.8 shows that $\hat{f}(T) \neq 0$ only for $B \subseteq A$, hence f is k-simple. For the other direction, we show that χ_A can be written as a linear combination of k-semistars whenever $|A| \leq k$. First, note that every l-semistar for $l \leq k$ can be written as a combination of k-semistars. Second,

$$\chi_A = \sum_{B\subseteq A}(-1)^{|B|}\mathbf{1}_{\tau_{A,B}}.$$ □

We will define the levels of the Fourier transform of S_n by analogy, and through this the Fourier basis. We start with the analogs of k-stars.

Definition 6.2. A *k-coset* is a family of permutations of the form $T_{(a_1,b_1),\ldots,(a_k,b_k)} = \{\pi \in S_n : \pi(a_1) = b_1, \ldots, \pi(a_k) = b_k\}$, where $a_i \neq a_j$ and $b_i \neq b_j$ whenever $i \neq j$.

The space $L_k^{S_n} \subseteq \mathbb{R}[S_n]$ is the linear span of the characteristic functions of all k-cosets. We say that a function on S_n is *k-simple* if it belongs to $L_k^{S_n}$. ○

Like their analogs in the Boolean cube, the spaces L_k (we omit the superscript S_n for clarity) get increasingly bigger: $L_0 \subset L_1 \subset \cdots \subset L_n = \mathbb{R}[S_n]$. It is not hard to verify directly that $L_k \subset L_{k+1}$ by writing a k-coset as the disjoint union of $(k+1)$-cosets. We can write L_k as an orthogonal sum of two subspaces $L_k = L_{k-1} \oplus R'_k$ (in other words, R'_k is the orthogonal complement of L_{k-1} with respect to L_k). The kth level of the Fourier transform corresponds to the subspace R'_k.

We can now describe a coarse version of the Fourier transform: given a function $f \in \mathbb{R}[S_n]$, define $\hat{f}'(k)$ to be the projection of f to R'_k. The coarse Fourier expansion of f is

$$f = \sum_{k=0}^{n} \hat{f}'(k).$$

This looks quite different from the Fourier expansion of functions on $\{0,1\}^n$. There, the Fourier coefficients were scalars, while here, each Fourier coefficient is a function belonging to some subspace. The two views can be unified if we replace the scalar $\hat{f}(A)$ with the function $\hat{f}(A)\chi_A$ which belongs to the one-dimensional subspace $\{\alpha\chi_A : \alpha \in \mathbb{R}\}$.

6.2.1 A finer decomposition

One crucial property of the Fourier transform on $\{0,1\}^n$ is the convolution property: the operator taking $\mathbf{1}_{\mathcal{F}}$ to $\mathbf{1}_{\mathcal{F}\oplus A}$ has the Fourier basis as eigenvectors. In order to obtain a similar property for S_n, we need to refine the Fourier transform constructed heretofore.

Definition 6.3. A *partition* of n is a non-increasing sequence of positive integers summing to n, for example $\lambda = (2,1,1)$ is a partition of 4. We also write $\lambda = (2,1^2)$ in this case. The *parts* of λ are $\lambda_1 = 2$ and $\lambda_2 = \lambda_3 = 1$. If λ has k parts, we use the convention that $\lambda_\ell = 0$ whenever $\ell > k$. The expression $\lambda \vdash n$ means that λ is a partition of n.

For a partition λ of n, a *(Young) tabloid of* $[n]$ *of shape* $\lambda = \lambda_1, \ldots, \lambda_k$ is an ordered partition $A_1, \ldots, A_k$ of $[n]$ (collectively known as $\mathbf{A}$) with $|A_i| = \lambda_i$. We also say that $\mathbf{A}$ is a λ-tabloid.

For a partition $\lambda = \lambda_1, \ldots, \lambda_k$ of n, a λ-*coset* is a family of permutations of the form $T_{\mathbf{A},\mathbf{B}} = \{\pi \in S_n : \pi(A_1) = B_1, \ldots, \pi(A_k) = B_k\}$, where $\mathbf{A}, \mathbf{B}$ have shape λ.

For a partition λ of n, the space $L_\lambda^{S_n} \subseteq \mathbb{R}[S_n]$ is the linear span of the characteristic functions of all λ-cosets. We say that a function on S_n is λ-*simple* if it belongs to $L_\lambda^{S_n}$. ○

Here are some examples. A k-coset is the same as an $(n-k,1^k)$-coset. An $(n-k,k)$-coset is a family of the form $\{\pi \in S_n : \pi(A) = B\}$, where $|A| = |B| = k$. It is easy to see that every $(n-k,k)$-coset is a disjoint union of k-cosets, and so $L_{(n-k,k)} \subseteq L_{(n-k,1^k)}$. For two partitions λ, μ, there is a general condition for the containment $L_\lambda \subseteq L_\mu$.

Definition 6.4. Let λ, μ be two partitions of n. We say that λ *dominates* μ, in symbols $\lambda \trianglerighteq \mu$, if $\sum_{i=1}^{j} \lambda_i \geq \sum_{i=1}^{j} \mu_i$ for all j (recall $\lambda_i = 0$ if λ has fewer than i parts). If $\lambda \trianglerighteq \mu$ and $\lambda \neq \mu$, then we say that λ *strictly dominates* μ, in symbols $\lambda \triangleright \mu$. ◯

Claim 6.2. *Let λ, μ be two partitions of n. Then $L_\lambda \subseteq L_\mu$ if and only if $\lambda \trianglerighteq \mu$, and $L_\lambda \subsetneq L_\mu$ if and only if $\lambda \triangleright \mu$.*

For example, $(n-k,1^k) \triangleright (n-(k+1),1^{k+1})$, and so $L_{(n-k,1^k)} \subset L_{(n-(k+1),1^{k+1})}$. Similarly, $(n-k,k) \triangleright (n-k,1^k)$ implies $L_{(n-k,k)} \subset L_{(n-k,1^k)}$. The domination order on partitions is only a partial order: for example the two partitions (3^2) and $(4,1^2)$ are incomparable. Claim 6.2, like Claim 6.3 below, is proven in Section 6.3 by appealing to the representation theory of S_n.

Mirroring our definition of the kth level R'_k, we define the subspaces R_λ and note their crucial property.

Definition 6.5. For a partition λ of n, let R_λ be the subspace of L_λ consisting of all vectors orthogonal to L_μ for all $\mu \triangleright \lambda$. The *Fourier expansion* of a function $f \in \mathbb{R}[S_n]$ is

$$f = \sum_{\lambda \vdash n} \hat{f}(\lambda),$$

where $\hat{f}(\lambda)$ is the projection of f to R_λ. ◯

Definition 6.6. For a function $f \in \mathbb{R}[S_n]$ and a permutation $\pi \in S_n$, the function $\pi f \in \mathbb{R}[S_n]$ is defined by $(\pi f)(\pi\alpha) = f(\alpha)$ for all $\alpha \in S_n$. ◯

Claim 6.3. *Let $P \subseteq S_n$ be a set of permutations which is closed under conjugation (we say that P is* conjugation-invariant*). For each partition λ of n there is a constant $C_{P,\lambda}$ such that for all $f \in \mathbb{R}[S_n]$,*

$$\hat{g}(\lambda) = C_{P,\lambda} \hat{f}(\lambda), \text{ where } g = \sum_{\pi \in P} \pi f.$$

As an example, we could take P to be all transpositions or all fixed-point free permutations. When P is not conjugation-invariant, the relation between the Fourier transforms of f and g is more complicated, and outside the scope of this thesis.

The subspaces R_λ are orthogonal by construction, and so we get Parseval's identity.

Lemma 6.4. *Let $f, g \in \mathbb{R}[S_n]$. We have*

$$\|f\|^2 = \sum_{\lambda \vdash n} \|\hat{f}(\lambda)\|^2$$

and

$$\langle f, g \rangle = \sum_{\lambda \vdash n} \langle \hat{f}(\lambda), \hat{g}(\lambda) \rangle.$$

Corollary 6.5. *Let $P \subseteq S_n$ be a conjugation-invariant set of permutations. Let B_P be the operator taking f to $\sum_{\pi \in P} \pi f$. For every function $f \in \mathbb{R}[S_n]$,*

$$f' B_P f = \sum_{\lambda \vdash n} C_{P,\lambda} \|\hat{f}(\lambda)\|^2,$$

where $C_{P,\lambda}$ are the constants given by Claim 6.3.

To complete the picture, we show how to extract $\mu(\mathcal{F})$ from the Fourier expansion of its characteristic function.

Lemma 6.6. *Let $\mathcal{F} \subseteq S_n$, and let $f = \mathbf{1}_{\mathcal{F}}$ be its characteristic function. Then*

$$\mu(\mathcal{F}) = \|\hat{f}((n))\| = \sum_{\lambda \vdash n} \|\hat{f}(\lambda)\|^2.$$

Proof. The subspace $R_{(n)} = L_{(n)}$ consists of all constant functions, and so $\hat{f}((n)) = \mu(\mathcal{F}) \mathbf{1}_{S_n}$. Therefore $\mu(\mathcal{F}) = \|\hat{f}((n))\|$. The second equality follows from Parseval's identity since $\|f\|^2 = \langle f, f \rangle = \mu(f)$. $\square$

Finally, in order to obtain an orthonormal Fourier basis, it suffices to choose an orthonormal basis for each subspace R_λ. This choice is arbitrary to some extent, and the literature contains several different canonical choices. For our purposes, the decomposition of $\mathbb{R}[S_n]$ into orthogonal subspaces R_λ suffices.

6.3 Representation theory of the symmetric group

We proceed to broadly outline classical representation theory, and some elements of the representation theory of S_n. Our account follows standard textbook treatments. For general representation theory, the standard text is Serre [74], and an introductory textbook is James and Liebeck [56]. For the representation theory of S_n, the standard text is James and Kerber [55]. For a quick overview of both, see Diaconis [14].

The purpose of this section is to connect our somewhat unorthodox description in Section 6.2 to the classical point of view, thereby proving the two unproved claims in that section. The reader who is willing to believe these claims can safely skip this section. Conversely, our account assumes familiarity with the rudiments of representation theory.

6.3.1 General representation theory

Representations The traditional way of developing the Fourier transform over an arbitrary group is through representation theory. A *representation* ρ of a group G is a homomorphism from G to the group of linear transformations over some finite vector space V. In other words, ρ assigns to each $\alpha \in G$ a linear operator $\rho(\alpha)$ acting on V in such a way that $\rho(\alpha\beta) = \rho(\alpha)\rho(\beta)$. The dimension of ρ is $\dim \rho = \dim V$. We sometimes write the representation as (ρ, V), to emphasize the role of V.

For example, let $G = S_n$ and let V be an abstract vector space of dimension n with basis $e_1, \ldots, e_n$. For each $\pi \in S_n$, we define $\rho(\pi)$ by sending e_i to $e_{\pi(i)}$. This is known as the *permutation representation*, since in matrix form, $\rho(\pi)$ is the permutation matrix corresponding to π.

Reducibility A representation ρ acting on V is *reducible* if there is a non-trivial subspace $0 \subsetneq W \subsetneq V$ such that $\rho(\alpha)W = W$ for each $\alpha \in G$ (this means that for each $w \in W$, $\rho(\alpha)w \in W$). Such a subspace is called an *invariant subspace*. If no such subspace exists, the representation is called *irreducible*.

For example, the permutation representation of S_n is reducible since the subspace W_1 spanned by $e_1 + \cdots + e_n$ is invariant. Another invariant subspace W_2 is spanned by the vectors $e_i - e_n$ for $i < n$, and we have $V = W_1 \oplus W_2$. The representation ρ thus decomposes to two *subrepresentations* ρ_1, ρ_2 acting on W_1, W_2, respectively. The representation ρ_1, known as the *trivial representation*, is trivially irreducible (any representation of dimension 1 is irreducible), and it turns out that ρ_2 is also irreducible.

Every representation can be decomposed into irreducible subrepresentations. More explicitly, given any representation ρ on V, there is a way of decomposing V into orthogonal invariant subspaces $V_1, \ldots, V_k$, such that the restriction of ρ to each V_i is irreducible. Here orthogonality is with respect to the inner product

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

Equivalence We say that two representations (ρ_1, V_1) and (ρ_2, V_2) are *equivalent* if there is a bijection T between V_1 and V_2 such that for all $\alpha \in G$ and $v \in V_1$, $T(\rho_1(\alpha)v) = \rho_2(\alpha)(Tv)$.

For example, take $n = 2$ in the previous example. The representation ρ_1 is equivalent to the representation $\rho_1'(e) = \rho_1'((12)) = 1$ acting on $\mathbb{C}$ by multiplication. The representation ρ_2 is equivalent to the representation $\rho_2'(e) = 1$, $\rho_2'((12)) = -1$ acting on $\mathbb{C}$. The two representations are inequivalent.

In general, the decomposition of a representation into irreducible subrepresentations need not be unique (we give an example below). However, if we lump all equivalent subrepresentations together, we get a coarser decomposition which is unique.

Group algebra We can turn $\mathbb{C}[G]$ into an algebra by defining a multiplication operation between any two $f, g \in \mathbb{C}[G]$:

$$(fg)(\alpha) = \sum_{\beta \in G} f(\alpha\beta^{-1})g(\beta).$$

In Section 6.2, we defined a similar operation αf, which is equal to $\mathbf{1}_\alpha f$.

The regular representation The *regular representation* ρ_{reg} on $\mathbb{C}[G]$ is defined by $\rho_{\text{reg}}(\alpha)f = \alpha f$. A subrepresentation of $\mathbb{C}[G]$ is determined by a subspace $W \subseteq \mathbb{C}[G]$ which is invariant, that is, satisfies $\alpha w \in W$ for all $\alpha \in G$, $w \in W$. From now on, we identify the subrepresentation with W.

For example, consider the group $G = S_3$. Define

$$W_1 = \{f \in \mathbb{C}[S_3] : f(\pi) = f(\pi(23)), \sum_\pi f(\pi) = 0\}.$$

The first condition states that $f(\pi)$ depends only on $\pi(1)$. It is not hard to check that $\dim W_1 = 2$ and that W_1 is invariant. Similarly we can define subspaces W_2 and W_3. It is not hard to check that $W_1(132) = W_2$ and $W_1(123) = W_3$, and so W_1, W_2, W_3 are all equivalent.

More generally, if W is an invariant subspace of $\mathbb{C}[G]$, then for each $f \in \mathbb{C}[G]$, $Wf = \{wf : w \in W\}$ is another invariant subspace. It turns out that if W, W' are equivalent subrepresentations of $\mathbb{C}[G]$ then $W' = Wf$ for some $f \in \mathbb{C}[G]$.

An important property of the regular representation ρ_{reg} is that each irreducible representation is equivalent to some subrepresentation of ρ_{reg}. Moreover, in any decomposition of $\mathbb{C}[S_n]$ into irreducible invariant subrepresentations, there are $\dim \rho$ factors equivalent to each irreducible representation ρ.

Continuing our example, for $G = S_3$, every decomposition of ρ_{reg} into irreducible subrepresentations contains two subrepresentations equivalent to W_1. In any such decomposition which contains W_1, the other subrepresentation equivalent to W_1 is $W_1(\mathbf{1}_{(132)} - \mathbf{1}_{(123)})$. Different decompositions are available, for example there is one containing W_2 as a factor and another containing W_3 as a factor. In all such decompositions, the sum of the two subrepresentations equivalent to W_1 is the invariant subspace

$$W = \{f \in \mathbb{C}[S_3] : \sum_\pi f(\pi) = \sum_\pi (-1)^\pi f(\pi) = 0\},$$

where $(-1)^\pi$ is the sign of π.

Fourier transform The *Fourier transform* of a function $f \in \mathbb{C}[G]$ at a representation ρ on V is

$$\hat{f}(\rho) = \sum_{\alpha \in G} f(\alpha)\rho(\alpha),$$

which is a linear transformation on V. By choosing a basis for V, we can think of $\hat{f}(\rho)$ as a matrix of dimension $\dim V$. This definition is different than the one we gave in Section 6.2; we explain the connection below.

For example, if ρ is the one-dimensional trivial representation (available at any group), then $\hat{f}(\rho) = \sum_\alpha f(\alpha)$.

The Fourier transform of f consists of $\hat{f}(\rho)$ for all irreducible representations ρ. The Fourier transform of f determines f, that is, given matrices A_ρ of dimensions $\dim \rho$, there is a unique function f such that $\hat{f}(\rho) = A_\rho$ for all ρ. This is proved by recourse to the regular representation.

The homomorphism property of ρ shows that

$$\widehat{fg}(\rho) = \sum_{\alpha \in G} (fg)(\alpha)\rho(\alpha) = \sum_{\alpha,\beta \in G} f(\alpha\beta^{-1})g(\beta)\rho(\alpha\beta^{-1})\rho(\beta) = \hat{f}(\rho)\hat{g}(\rho).$$

For each irreducible representation ρ, $\hat{f}(\rho)$ is a matrix with $(\dim \rho)^2$ entries. In each decomposition of ρ_{reg} into irreducible subrepresentations, there are $\dim \rho$ subrepresentations which are equivalent to ρ, each of dimension ρ. The sum of all the corresponding subspaces is a subspace U_ρ of $\mathbb{C}[S_n]$ of dimension $(\dim \rho)^2$. One basis of U_ρ is given by the functions $f_{ij}(\alpha) = \rho(\alpha)_{ij}$. In this way, we can think of $\hat{f}(\rho)$ as an element of U_ρ. This connects the description of the Fourier transform in this section to its definition in Section 6.2. The advantage of the latter definition is that it is basis-free. Moreover, the formula $\widehat{fg}(\rho) = \hat{f}(\rho)\hat{g}(\rho)$ remains true under the latter definition.

Class functions A *class function* $f \in \mathbb{C}[G]$ is a function which is constant on conjugacy classes, that is $f(\alpha) = f(\beta)$ if $\alpha, \beta \in G$ are conjugate. If ρ is an irreducible representation then $\hat{f}(\rho)$ is a scalar (multiple of the identity). This is the property behind Claim 6.3.

Characters Each representation ρ has an associated character χ_ρ defined by $\chi_\rho(\alpha) = \operatorname{Tr} \rho(\alpha)$. It is easy to see that χ_ρ is a class function:

$$\chi_\rho(\beta\alpha\beta^{-1}) = \operatorname{Tr} \rho(\beta)\rho(\alpha)\rho(\beta)^{-1} = \operatorname{Tr} \rho(\alpha) = \chi_\rho(\alpha),$$

using the homomorphism property of ρ. It turns out that the vector space of all class functions is spanned by the characters of all irreducible representations. Therefore the number of inequivalent irreducible representations is equal to the number of conjugacy classes of G.

It is easy to see that $\dim \rho = \chi_\rho(e)$, where $e \in G$ is the identity element. Also, calculation shows that for every $\alpha \in G$,

$$\hat{\mathbf{1}}_{\{\beta\alpha\beta^{-1}:\beta\in G\}}(\rho) = \frac{\chi_\rho(\alpha)}{\dim\rho}|\{\beta\alpha\beta^{-1} : \beta \in G\}|.$$

Therefore, in order to calculate the constants $C_{P,\lambda}$ in Claim 6.3, it is enough to know all the characters of irreducible representations of G.

If a representation ρ decomposes into subrepresentations ρ_i, then it is easy to see that $\chi_\rho = \sum_i \chi_{\rho_i}$. We will use this fact to calculate the characters of S_n.

6.3.2 The symmetric group

In the previous section, we considered the space of all complex-valued functions over a group G. Such generality is needed, for example, in the case of the groups $\mathbb{Z}_k^n$ for $k > 2$. This is manifested by the fact that the Fourier characters in this case are complex. However, for $G = S_n$, there is a real-valued Fourier basis, and so we consider for simplicity $\mathbb{R}[S_n]$ instead of $\mathbb{C}[S_n]$.

The irreducible representations of the symmetric group are indexed by partitions of n. This is not surprising, since the conjugacy classes of S_n (known in this context as *cycle types*) are also indexed by partitions of n, and we know that the number of inequivalent irreducible representations is the same as the number of conjugacy classes.

Before constructing the actual irreducible representations, we construct the *permutation modules* M^λ. (A module here is the same as a vector space.) For each partition λ of n, consider the set $P(\lambda)$ of all λ-tabloids. A permutation $\pi \in S_n$ acts on a λ-tabloid $\mathbf{A} = (A_1, \dots, A_k)$ by permuting the numbers: $\mathbf{A}^\pi = (\pi(A_1), \dots, \pi(A_k))$. In this way, we get a representation ρ on a vector space of dimension $|P(\lambda)|$: given a basis $\{e_{\mathbf{A}} : \mathbf{A} \in P(\lambda)\}$, $\rho(\pi)(e_{\mathbf{A}}) = e_{\mathbf{A}^\pi}$.

The representation corresponding to M^λ is reducible, but it contains an irreducible subrepresentation which corresponds to a submodule S^λ known as a *Specht module*. We do not construct S^λ explicitly here. The decomposition of M^λ into irreducible representations contains representations isomorphic to S^μ if and only if $\mu \trianglerighteq \lambda$, and furthermore S^λ appears only once, and this fact can be used to construct the modules S^λ.

The regular representation decomposes into irreducible representations, and there is a corresponding decomposition of $\mathbb{R}[S_n]$ into subspaces which are equivalent to irreducible representations. For a partition λ of n, let $U_\lambda \subseteq \mathbb{R}[S_n]$ be the sum of all such subspaces equivalent to S^λ, and let V_λ be the sum of U_μ for all $\mu \trianglerighteq \lambda$. While the decomposition of $\mathbb{R}[S_n]$ into irreducible subspaces is not unique, the subspaces U_λ and V_λ are uniquely defined. Following Ellis [23], the following lemma shows that $V_\lambda = L_\lambda$ for all λ, where L_λ is the space defined in Section 6.2.

Lemma 6.7. *For every partition λ of n, $V_\lambda = L_\lambda$.*

Proof. For a tabloid $\mathbf{A} \in P(\lambda)$, let $V(\mathbf{A})$ be the subspace of $\mathbb{R}[S_n]$ spanned by $\{T_{\mathbf{A},\mathbf{B}} : \mathbf{B} \in P(\lambda)\}$. It is not hard to check that for $\pi \in S_n$, $\pi T_{\mathbf{A},\mathbf{B}} = T_{\mathbf{A},\mathbf{B}^\pi}$. From the definition it is obvious that the restriction of ρ_{reg} to $V(\mathbf{A})$ is equivalent to M^λ. If we decompose $V(\mathbf{A})$ into irreducible subrepresentations, all the irreducible representations we obtain in this way are equivalent to S^μ for some $\mu \trianglerighteq \lambda$ (this is a property of M^λ). Hence $V(\mathbf{A}) \subseteq V_\lambda$. Since L_λ is the sum of $V(\mathbf{A})$ over all $\mathbf{A} \in P(\lambda)$, we conclude that $L_\lambda \subseteq V_\lambda$.

For the other direction, take any $\mu \trianglerighteq \lambda$, and pick an arbitrary $\mathbf{A} \in P(\lambda)$. Since $V(\mathbf{A})$ is equivalent to M^λ, it contains some subrepresentation W equivalent to S^μ. Every subrepresentation of $\mathbb{R}[S_n]$ equivalent to S^μ is of the form Wf for some $f \in \mathbb{R}[S_n]$. Since $V(\mathbf{A})\mathbf{1}_{\{\pi^{-1}\}} = V(\mathbf{A}^\pi)$, we see that $V(\mathbf{A})f \subseteq L_\lambda$, and so $Wf \subseteq L_\lambda$. Therefore $U_\mu \subseteq L_\lambda$. Considering all $\mu \trianglerighteq \lambda$, we conclude that $V_\lambda \subseteq L_\lambda$. □

Corollary 6.8. *For every partition λ of n, $U_\lambda = R_\lambda$.*

Another immediate corollary is Claim 6.2. To prove Claim 6.3, recall that for every class function g and irreducible representation λ, $\hat{g}(\lambda)$ is equal to some scalar multiple of the identity $C_{P,\lambda}$. Therefore $\widehat{gf}(\lambda) = \hat{g}(\lambda)\hat{f}(\lambda) = C_{P,\lambda}\hat{f}(\lambda)$. In other words, the effect of multiplying by g from the left on the projection to $U_\lambda = R_\lambda$ is to multiply the projection by $C_{P,\lambda}$. Taking $g = \mathbf{1}_P$ for some conjugation-invariant family of permutations, we get Claim 6.3, since

$$\mathbf{1}_P f = \sum_{\pi \in P} \pi f.$$

6.3.3 Some formulas

In order to satisfy the curiosity of the reader, we now give some formulas for the characters of S_n. For a partition λ, let ξ_λ be the character of M^λ, and let χ_λ be the character of S^λ.

Lemma 6.9 (Young's rule)**.** *Let μ be a partition of n. The module M^μ decomposes as*

$$M^\mu = \bigoplus_{\lambda \trianglerighteq \mu} K_{\lambda,\mu} S^\lambda,$$

that is, there are $K_{\lambda,\mu}$ submodules equivalent to S^λ, where the Kostka number *$K_{\lambda,\mu}$ is the number of solutions of the following problem. For each part λ_i, choose λ_i integers $N_{i,1} < \ldots < N_{i,\lambda_i}$ under the following constraints: $N_{i+1,j} \geq N_{i,j}$ for all i, j such that $N_{i,j}, N_{i+1,j}$ exist; and each integer k is chosen exactly μ_k times.*

Corollary 6.10. *Let μ be a partition of n. We have*

$$\xi_\mu = \sum_{\lambda \trianglerighteq \mu} K_{\lambda,\mu} \chi_\lambda.$$

Young's rule gives us a way of calculating the characters χ_λ from the characters ξ_μ. The latter have an explicit formula.

Lemma 6.11. *Let μ be a partition of n, and $\pi \in S_n$. The character $\xi_\mu(\pi)$ is equal to the number of μ-tabloids fixed by π.*

There are many other formulas which help calculating the characters S_n. For a list, consult for example [28].

6.4 Intersecting families of permutations

Having developed Fourier analysis over S_n, we now put it to use to study intersecting families of permutations.

Definition 6.7. A family of permutations $\mathcal{F} \subseteq S_n$ is *intersecting* if every two permutations $\alpha, \beta \in S_n$ agree on at least one point. ○

Lemma 6.12. *An intersecting family of permutations in S_n contains at most $(n-1)!$ permutations.*

Proof. Let $\sigma = (12\ldots n)$. For $i \neq j \in \{0,\ldots,n-1\}$ and $\pi \in S_n$, the permutations $\pi\sigma^i, \pi\sigma^j$ do not intersect. Indeed, $\pi\sigma^i(t) = \pi(i+t)$ whereas $\pi\sigma^j(t) = \pi(j+t)$, where addition is modulo n. For each permutation π, define $S(\pi) = \{\pi\sigma^i : i \in \{0, \dots, n-1\}\}$. It is easy to check that the sets $\{S(\pi) : \pi(1) = 1\}$ partition S_n. Every intersecting family contains at most one permutation from each of these $(n-1)!$ sets. □

This proof comes from Deza and Frankl's paper [13]. Deza and Frankl conjectured that the only intersecting families of permutations of size $(n-1)!$ are cosets. This turned out to be much harder to prove, and the first proofs, due to Cameron and Ku [7] and to Larose and Malvenuto [65] appeared only 25 years later.

In this section we present an alternative proof of this uniqueness result due to Ellis, Friedgut and Pilpel [28]. Their proof follows Friedgut's method, and has the advantage of providing a stability result: if an intersecting family of permutations has size $(1-\epsilon)(n-1)!$, then it is close to a coset.

Definition 6.8. For an integer n, let $\mathrm{FPF}_n \subseteq S_n$ consist of all permutations having no fixed points. ○

Lemma 6.13. *A family $\mathcal{F} \subseteq S_n$ is intersecting if and only if $\mathcal{F}$ is disjoint from $\pi\mathcal{F}$ for every $\pi \in \mathrm{FPF}_n$.*

Proof. If $\mathcal{F}$ intersects $\pi\mathcal{F}$ for some $\pi \in \mathrm{FPF}_n$, say $\alpha = \pi\beta$ for some $\alpha, \beta \in \mathcal{F}$, then for all $i \in [n]$, $\alpha(i) = \pi\beta(i) \neq \beta(i)$, and so $\mathcal{F}$ is not intersecting. Conversely, if $\mathcal{F}$ is not intersecting then there are some $\alpha, \beta \in \mathcal{F}$ which agree on no point. Hence $\alpha\beta^{-1} \in \mathrm{FPF}_n$, and $\mathcal{F}$ intersects $\alpha\beta^{-1}\mathcal{F}$. □

The family FPF_n is conjugation-invariant, and so we can apply Claim 6.3 to obtain a condition on the Fourier expansion of the characteristic function of every intersecting family.

Lemma 6.14. *Let $\mathcal{F}$ be an intersecting family of permutations, and let $f = \mathbf{1}_{\mathcal{F}}$. There are coefficients C_λ such that*

$$\sum_{\lambda \vdash n} C_\lambda \|\hat{f}(\lambda)\|^2 = 0.$$

The coefficients C_λ satisfy the following properties:

(a) $C_{(n)} = |\mathrm{FPF}_n|$.

(b) When $n \geq 4$, $\max_{\lambda \neq (n)} |C_\lambda| = |\mathrm{FPF}_n|/(n-1)$.

(c) *When* $n \geq 5$, $|C_\lambda| = |\mathrm{FPF}_n|/(n-1)$ *is attained only at* $C_{(n-1,1)} = -|\mathrm{FPF}_n|/(n-1)$.

(d) *We have* $\max_{\lambda \neq (n),(n-1,1)} |C_\lambda| = O((n-2)!)$. *(Note that* $|\mathrm{FPF}_n| = \Theta(n!)$.*)*

Proof. Let B_π denote the operator mapping f to πf, and define

$$A = \sum_{\pi \in \mathrm{FPF}_n} B_\pi.$$

Lemma 6.13 shows that

$$f' A f = 0.$$

Claim 6.3 implies that for some constants C_λ,

$$\widehat{Af}(\lambda) = C_\lambda \hat{f}(\lambda).$$

Hence the claimed formula follows from Parseval's identity.

Clearly $C_{(n)} = |\mathrm{FPF}_n|$. All other properties of C_λ are proved in Ellis [22]. See also Renteln [71], who proves the first three properties. □

Applying Hoffman's bound, we get an alternative proof of the upper bound $(n-1)!$ on intersecting families.

Theorem 6.15. *Let* $\mathcal{F}$ *be an intersecting family of permutations.*

Upper bound: *If* $n \geq 4$ *then* $\mu(\mathcal{F}) \leq 1/n$.

Uniqueness: *If* $n \geq 5$ *and* $\mu(\mathcal{F}) = 1/n$ *then* $\mathbf{1}_\mathcal{F} \in L_{(n-1,1)}$.

Stability: *If* $\mu(\mathcal{F}) \geq 1/n - \epsilon$ *then* $\|f - f_1\|^2 = O(\epsilon)$, *where* $f = \mathbf{1}_\mathcal{F}$ *and* f_1 *is the projection of* f *to* $L_{(n-1,1)}$.

Proof. All items follows from Lemma 6.14 via Hoffman's bound, with ∅ replaced by the partition (n). To avoid confusion, we replace λ with C in Hoffman's bound. We have $C_{(n)} = |\mathrm{FPF}_n|$. When $n \geq 4$, $C_{\min} = -|\mathrm{FPF}_n|/(n-1)$. Hence

$$\mu(\mathcal{F}) \leq m_{\max} = \frac{|\mathrm{FPF}_n|/(n-1)}{|\mathrm{FPF}_n| + |\mathrm{FPF}_n|/(n-1)} = \frac{1}{n}.$$

When $n \geq 5$, $\mu(\mathcal{F}) = m_{\max}$ only if $\|\hat{f}(\lambda)\| = 0$ unless $\lambda \in \{(n),(n-1,1)\}$, proving uniqueness. For stability, note that $C_2 = O((n-2)!) = O(|\mathrm{FPF}_n|/n^2)$. Hence $\mu(\mathcal{F}) \geq (1-\epsilon)/n$ implies that

$$\sum_{\lambda \neq (n),(n-1,1)} \|\hat{f}(\lambda)\|^2 \leq \frac{|\mathrm{FPF}_n|/(n-1)}{|\mathrm{FPF}_n|/(n-1) - O(|\mathrm{FPF}_n|/n^2)}\epsilon = O(\epsilon). \qquad \square$$

Using the cross-intersecting version of Hoffman's bound, we get in the same way that for $n \geq 4$, if $\mathcal{F}$ and $\mathcal{G}$ are cross-intersecting then $|\mathcal{F}||\mathcal{G}| \leq (n-1)!^2$. This is false for $n = 3$, as the example $\mathcal{F} = \{e,(123),(132)\}$, $\mathcal{G} = \{(12),(13),(23)\}$ given by Ellis [22] demonstrates.

Ellis, Friedgut and Pilpel [28] proved an analog of Theorem 6.15 for t-intersecting families of permutations (families in which every two permutations agree on t points), and Ellis [23] proved an analog of the theorem for t-set-intersecting families of permutations (families in which every two permutations agree on the image of some set of size t).

We use an argument due to Ellis, Friedgut and Pilpel to show that cosets are the unique maximal intersecting families.

Lemma 6.16. *Suppose $f \in L_{(n-1,1)}$ satisfies $f \geq 0$. Then f can be expressed as a non-negative linear combination of the cosets $\mathbf{1}_{T_{i,j}}$.*

Proof. Since $f \in L_{(n-1,1)}$, for some coefficients $\alpha_{i,j}$,

$$f = \sum_{i,j=1}^{n} \alpha_{i,j} \mathbf{1}_{T_{i,j}}. \tag{6.1}$$

While the functions $\mathbf{1}_{\alpha_{i,j}}$ span $L_{(n-1,1)}$, they don't form a basis since they are not linearly independent. In other words, there are many different ways of expressing f as a linear combination of $\mathbf{1}_{\alpha_{i,j}}$. Indeed, for any β_i, γ_j satisfying $\sum_i \beta_i + \sum_j \gamma_j = 0$, another such linear combination is

$$f = \sum_{i,j=1}^{n} (\alpha_{i,j} + \beta_i + \gamma_j) \mathbf{1}_{T_{i,j}}. \tag{6.2}$$

We claim that there is such a linear combination in which all coefficients are non-negative.

Suppose to the contrary that (6.1) is a linear combination maximizing the value $m = \min_{i,j} \alpha_{i,j}$, and $m < 0$ (the maximum is achieved since m is the solution to a bounded linear program). Let G be a bipartite graph with n vertices on both sides, in which i and j are connected if $\alpha_{i,j} = m$. We claim that G has no perfect matching. Indeed, if $\{(i,\pi(i)) : i \in [n]\}$ is a perfect matching, then

$$f(\pi) = \sum_{i=1}^{n} \alpha_{i,j} = nm < 0,$$

contradicting the fact that $f \geq 0$. Here we use the fact that $\pi \in T_{i,j}$ if and only if $j = \pi(i)$.

Choose an arbitrary entry $\alpha_{k,l} = m$, and consider the graph H obtaining by deleting the two corresponding vertices from G. The graph H has no perfect matching, and so there is some set A of rows whose neighbor set B has cardinality $|B| < |A|$. For some small $\epsilon > 0$, let $\beta_i = -[\![i \in A]\!]\epsilon$ and let $\gamma_j = [\![j \in B]\!]\epsilon + [\![j = l]\!](|A| - |B|)\epsilon$. Note that $\sum_i \beta_i + \sum_j \gamma_j = 0$. If ϵ is small enough, the linear combination (6.2) has $\min_{i,j}(\alpha_{i,j} + \beta_i + \gamma_j) \geq m$, and m occurs less times among the coefficients as before (since $\alpha_{k,l} + \beta_k + \gamma_l > m$). Repeating this operation enough times, eventually we get a linear combination whose minimal coefficient is strictly larger than m, contradicting the choice of $\alpha_{i,j}$. □

Lemma 6.17. *Suppose $f \in L_{(n-1,1)}$ is a function with non-negative integer values. Then f can be expressed as a non-negative integral combination of the functions $\mathbf{1}_{T_{i,j}}$ for $i, j \in [n]$.*

In particular, if f is Boolean, then f is the characteristic function of a disjoint union of cosets.

Proof. The proof is by induction on $\sum_\pi f(\pi)$. If $f = 0$ then the result is trivial. Otherwise, let $f = \sum_{i,j} \alpha_{i,j} \mathbf{1}_{T_{i,j}}$ be the non-negative linear combination given by Lemma 6.16. Since $f \neq 0$, there must be some positive coefficient $\alpha_{k,l} > 0$. Hence $f(\pi) > 0$ for every $\pi \in T_{k,l}$. Since all values of f are integers, this shows that $f(\pi) \geq 1$ for all $\pi \in T_{k,l}$. Hence $f - \mathbf{1}_{T_{k,l}}$ is again a function with non-negative integer values. Applying the induction hypothesis, we deduce that f is a non-negative integral combination of the functions $\mathbf{1}_{T_{i,j}}$ for $i, j \in [n]$.

If f is Boolean then f must be a sum of distinct $\mathbf{1}_{T_{i,j}}$. Furthermore, the cosets $T_{i,j}$ appearing in this sum must be disjoint, for otherwise the sum is not Boolean. □

Ellis, Friedgut and Pilpel [28] prove a similar result for $L_{(n-k,1^k)}$. Ellis [21] shows that a similar result is not true for A_n, the alternating group (for the analog of $L_{(n-1,1)}$).

As a simple corollary, we obtain a proof of the Deza–Frankl conjecture.

Theorem 6.18. *For $n \geq 5$, if $\mathcal{F} \subseteq S_n$ is an intersecting family of permutations of size $(n-1)!$ then $\mathcal{F}$ is a coset.*

Proof. Theorem 6.15 shows that in this case, $\mathbf{1}_{\mathcal{F}} \in L_{(n-1,1)}$. Lemma 6.17 shows that $\mathcal{F}$ is a disjoint union of cosets. Since $|\mathcal{F}| = (n-1)!$, this disjoint union must consist of a single coset. □

The generalization of Lemma 6.17 to $L_{(n-t,1^t)}$ implies a similar generalization of Theorem 6.18 to t-intersecting families of permutations. Ellis [23] proved a similar result for t-set-intersecting families of permutations, which are families in which any two permutations agree on the unordered image of a set of size t. His proof does not involve the analog of Lemma 6.17 for $L_{(n-t,t)}$, though it does use the analog for $L_{(n-t,1^t)}$.

In the next chapter, we prove a stability result which will allow us to conclude that an intersecting family of size $(1-\epsilon)(n-1)!$ has to be $O(\epsilon/n)$-close to a coset. For the proof as well as other results in this vein, consult Section 7.6.

Chapter 7

A structure theorem for small dictatorships on S_n

The classical theorem by Friedgut, Kalai and Naor [42], Theorem 2.22, shows that a Boolean function on $\{0,1\}^n$ whose Fourier expansion is concentrated on the first two levels must be close to a Boolean function whose Fourier expansion is supported on the first two levels (which must be a function depending on at most one coordinate). In this chapter and the next, we prove similar results for Boolean functions on S_n.

In Chapter 3 we showed that an intersecting family of sets of maximal μ_p-measure (for $p < 1/2$) has the property that the Fourier expansion of its characteristic function is supported on the first two levels. Similarly, if the family has almost maximal measure, then the Fourier expansion must be concentrated on the first two levels. The Friedgut–Kalai–Naor theorem then implies stability: such a family must be close to a star. Similarly, in Chapter 6 we showed that an intersecting family of permutations of maximal size has the property that its characteristic function belongs to the space $L_{(n-1,1)}$ spanned by $\{\mathbf{1}_{T_{i,j}} : i, j \in [n]\}$. Moreover, if the family has almost maximal size, then the characteristic function is close to its projection to $L_{(n-1,1)}$. Using the main theorem proved in this chapter, we will prove stability: such a family must be close to a coset. In other words, we will show that if a Boolean family of permutations of size roughly $(n-1)!$ is close to a function in $L_{(n-1,1)}$ then it must be close to a Boolean function in $L_{(n-1,1)}$.

Our goal in this chapter is to prove the following theorem.

Theorem 7.1. *There is an $\epsilon_0 > 0$ such that the following holds.*

Let $\mathcal{F} \subseteq S_n$ be a family of permutations of size $c(n-1)!$, where $c \leq n/2$. Let $f = \mathbf{1}_{\mathcal{F}}$ (so $\mathbb{E}[f] = c/n$) and let $f_1 = \hat{f}((n)) + \hat{f}((n-1,1))$ be the projection of f to $L_{(n-1,1)}$.

If $\mathbb{E}[(f - f_1)^2] = \epsilon c/n$, where $\epsilon \leq \epsilon_0$, then there exists a family $\mathcal{G} \subseteq S_n$ which is the union of $\lceil c \rfloor$ cosets satisfying

$$|\mathcal{F} \Delta \mathcal{G}| = O\left(\sqrt{\epsilon} + \frac{1}{n}\right) c^2 (n-1)!.$$

Moreover,

$$|c - \lceil c \rfloor| = O\left(\sqrt{\epsilon} + \frac{1}{n}\right) c^2.$$

Here and in the rest of the chapter, $\lceil c \rfloor$ is the integer closest to c, breaking ties arbitrarily, and $x = O(y)$ means that for some constant $C > 0$, $x \leq Cy$ for large enough n.

The cosets comprising the family $\mathcal{G}$ need not be disjoint, and so it need not be the case that $\mathbf{1}_{\mathcal{G}} \in L_{(n-1,1)}$. This is unavoidable, for consider the family $\mathcal{F} = T_{1,1} \cup T_{2,2}$. We have

$$\|f - f_1\| \leq \|f - \mathbf{1}_{T_{1,1}} - \mathbf{1}_{T_{2,2}}\| = \|\mathbf{1}_{T_{(1,1),(2,2)}}\| = \frac{1}{n(n-1)},$$

where $T_{(1,1),(2,2)} = T_{(1,1)} \cap T_{(2,2)}$. Therefore we can take $\epsilon \approx 1/(2n)$. On the other hand, it is not hard to check that $|\mathcal{F} \Delta \mathcal{G}| = \Omega((n-1)!)$ for every family $\mathcal{G}$ which is the disjoint union of cosets. So on the one hand, the Fourier expansion of $f = \mathbf{1}_{\mathcal{F}}$ is concentrated on the first two levels, and on the other hand, it cannot be approximated with an error which is $o(|\mathcal{F}|)$ by a family which is the disjoint union of cosets.

Because the cosets comprising the family $\mathcal{G}$ are not necessarily disjoint, Theorem 7.1 is not a classical stability result. A classical stability result has the following form: if a Boolean family is close to a collection L, then it is close to a Boolean member of L. As shown in the preceding paragraph, such a result isn't true in our case (unless $c = 1$). In the following chapter, we prove a similar result which corresponds to the case $c = \Theta(n)$. In that regime, we are able to approximate $\mathcal{F}$ with a family which is the disjoint union of cosets, and so the theorem we obtain there is a classical stability result.

In subsequent work [25], we generalize Theorem 7.1 to Boolean functions of size $c(n-t)!$ supported on the first $t+1$ levels (in other words, close to $L_{n-t,1^t}$). This enables us to deduce stability for t-intersecting families of permutations.

This chapter follows our joint work with David Ellis and Ehud Friedgut [24].

7.1 Overview of the proof

The rest of this chapter, up to Section 7.6, contains the proof of Theorem 7.1. We will assume throughout that the family $\mathcal{F}$ and the related functions and quantities n, f, f_1, c, ϵ are given.

The idea of the proof is to analyze the $n \times n$ matrix B whose entries are given by

$$b_{ij} = \frac{\|\mathcal{F} \cap T_{i,j}\|}{(n-1)!} - \frac{\|\mathcal{F}\|}{n!} = n\langle f, \mathbf{1}_{T_{i,j}} \rangle - \frac{c}{n}.$$

When $\mathcal{F}$ is a disjoint union of c cosets, the matrix B takes one of the following forms:

$$\begin{array}{l} \quad\overbrace{\qquad\qquad\qquad\qquad\qquad}^{c} \\ \begin{pmatrix} 1-\frac{c}{n} & \cdots & 1-\frac{c}{n} & -\frac{c}{n} & \cdots & -\frac{c}{n} \\ -\frac{n-c}{n(n-1)} & \cdots & -\frac{n-c}{n(n-1)} & \frac{c}{n(n-1)} & \cdots & \frac{c}{n(n-1)} \\ \vdots & & \vdots & \vdots & & \vdots \\ -\frac{n-c}{n(n-1)} & \cdots & -\frac{n-c}{n(n-1)} & \frac{c}{n(n-1)} & \cdots & \frac{c}{n(n-1)} \end{pmatrix} \end{array} \quad \text{or} \quad \begin{array}{c} c\left\{\begin{matrix} \\ \\ \\ \\ \end{matrix}\right. \\ \\ \\ \\ \end{array} \begin{pmatrix} 1-\frac{c}{n} & -\frac{n-c}{n(n-1)} & \cdots & -\frac{n-c}{n(n-1)} \\ \vdots & \vdots & & \vdots \\ 1-\frac{c}{n} & -\frac{n-c}{n(n-1)} & \cdots & -\frac{n-c}{n(n-1)} \\ -\frac{c}{n} & \frac{c}{n(n-1)} & \cdots & \frac{c}{n(n-1)} \\ \vdots & & & \\ -\frac{c}{n} & \frac{c}{n(n-1)} & \cdots & \frac{c}{n(n-1)} \end{pmatrix}. \tag{7.1}$$

As we can see, in these cases B contains c "strong" entries (close to 1), and all other entries are of order $O(c/n)$. While we cannot expect B to be of similar form in general (since $\mathcal{F}$ might not be close to a family which is a disjoint union of cosets), we will show that B has roughly c strong entries, and all other entries are small.

In order to analyze the entries of B, we will estimate their sum of squares and their sum of cubes. It will turn out that both quantities are roughly equal to c, which implies (together with easy bounds on the individual entries) that B has roughly c strong entries.

It remains to analyze the moments of the entries of B. To this end, we define an auxiliary function

$$h = \sum_{i,j} b_{ij} \mathbf{1}_{T_{i,j}} = \frac{n}{n-1} f_1 - \frac{c}{n-1}.$$

The moments of this function are directly related to the moments of the entries of B:

$$\mathbb{E}[h^2] = \frac{1}{n-1}\sum_{i,j} b_{ij}^2,$$
$$\mathbb{E}[h^3] = \frac{n}{(n-1)(n-2)}\sum_{i,j} b_{ij}^3.$$

We will be able to estimate $\mathbb{E}[h^2]$ using $\mathbb{E}[f^2] = \mathbb{E}[f] = c/n$ and $\mathbb{E}[(f-f_1)^2] \le \epsilon c/n$. Estimating $\mathbb{E}[h^3]$ is the most technically difficult part of the proof.

7.2 Basic definitions

In this section we start the proof proper. We make the following crucial definitions:

$$a_{ij} = \frac{|\mathcal{F}\cap T_{i,j}|}{(n-1)!}, \tag{7.2}$$
$$b_{ij} = a_{ij} - \frac{c}{n}, \tag{7.3}$$
$$g = \sum_{i,j=1}^{n} a_{ij}\mathbf{1}_{T_{i,j}}, \tag{7.4}$$
$$h = \sum_{i,j=1}^{n} b_{ij}\mathbf{1}_{T_{i,j}}. \tag{7.5}$$

The entries a_{ij} and b_{ij} form two $n\times n$ matrices A and B, respectively. We start by calculating the row and column sums of these matrices.

Lemma 7.2. *Each row and each column in A sums to c. Each row and each column in B sums to zero.*

Proof. We have
$$\sum_{i=1}^{n} a_{1i} = \sum_{i=1}^{n}\frac{|\mathcal{F}\cap T_{1,i}|}{(n-1)!} = \frac{|\mathcal{F}|}{(n-1)!} = c.$$
A similar calculation shows that each other row and each column in A sums to c. The other claim follows directly from $b_{ij} = a_{ij} - c/n$. □

Corollary 7.3. *For each $i,j \in [n]$,*
$$\sum_{k\neq i}\sum_{l\neq j} b_{kl} = b_{ij}.$$
(The first sum is over $k\in[n]\setminus\{i\}$, and the second is over $l\in[n]\setminus\{j\}$.)

Proof. Using the fact that rows and columns sum to zero,

$$\sum_{k \neq i} \sum_{l \neq j} b_{kl} = \sum_{k \neq i} (-b_{kj}) = b_{ij}. \qquad \square$$

Next, we relate g and h to f_1 (the projection of f to $L_{(n-1,1)}$). We need first an easy result on the size of the intersection of two cosets.

Lemma 7.4. *Let T_{ij}, T_{kl} be two cosets and $S = |T_{ij} \cap T_{kl}|$. If $i = k$ and $j = l$ then $S = (n-1)!$. If $i \neq k$ and $j \neq l$ then $S = (n-2)!$. Otherwise, $S = 0$.*

Lemma 7.5. *We have*

$$g = \frac{n}{n-1} f_1 + \frac{n-2}{n-1} c,$$
$$h = \frac{n}{n-1} f_1 - \frac{c}{n-1}.$$

Proof. It is easy to see that $h = g - c$, and so it is enough to prove the second formula. Since both sides of the formula are in $L_{(n-1,1)}$, it is enough to show that both sides have the same inner product with each $\mathbf{1}_{T_{i,j}}$ (since these functions span $L_{(n-1,1)}$). We have

$$\begin{aligned}
\langle h, \mathbf{1}_{T_{i,j}} \rangle &= \sum_{k,l=1}^{n} b_{kl} \langle \mathbf{1}_{T_{k,l}}, \mathbf{1}_{T_{i,j}} \rangle \\
&= \frac{1}{n!} \sum_{k,l=1}^{n} b_{kl} |T_{k,l} \cap T_{i,j}| \\
&= \frac{b_{ij}}{n} + \frac{1}{n(n-1)} \sum_{k \neq i} \sum_{l \neq j} b_{kl} \\
&= \frac{b_{ij}}{n} + \frac{b_{ij}}{n(n-1)} = \frac{b_{ij}}{n-1},
\end{aligned}$$

using Lemma 7.4. On the other hand,

$$b_{ij} = \frac{|\mathcal{F} \cap T_{i,j}|}{(n-1)!} - \frac{c}{n} = n \langle f, \mathbf{1}_{T_{i,j}} \rangle - \frac{c}{n} = n \langle f_1, \mathbf{1}_{T_{i,j}} \rangle - \frac{c}{n}.$$

Therefore

$$\langle h, \mathbf{1}_{T_{i,j}} \rangle = \frac{b_{ij}}{n-1} = \frac{n}{n-1} \langle f_1, \mathbf{1}_{T_{i,j}} \rangle - \frac{c}{n(n-1)} = \frac{n}{n-1} \langle f_1, \mathbf{1}_{T_{i,j}} \rangle - \frac{c}{n-1} \langle \mathbf{1}_{S_n}, \mathbf{1}_{T_{i,j}} \rangle. \qquad \square$$

7.3 Moment formulas

We proceed to relate the moments of h and the moments of B, making repeated use of the fact that the rows and columns of B sum to zero.

Lemma 7.6. *We have*

$$\mathbb{E}[h^2] = \frac{1}{n-1}\sum_{i,j=1}^{n} b_{ij}^2.$$

Proof. Squaring the defining expression (7.5) of h, we get

$$h^2 = \sum_{i,j,k,l=1}^{n} b_{ij}b_{kl}\mathbf{1}_{T_{i,j}}\mathbf{1}_{T_{k,l}} = \sum_{i,j=1}^{n} b_{ij}^2\mathbf{1}_{T_{i,j}} + \sum_{\substack{i,j,k,l:\\ i\neq k, j\neq l}} b_{ij}b_{kl}\mathbf{1}_{T_{(i,j),(k,l)}}.$$

Taking expectations,

$$\begin{aligned}\mathbb{E}[h^2] &= \frac{1}{n}\sum_{i,j=1}^{n} b_{ij}^2 + \frac{1}{n(n-1)}\sum_{i,j=1}^{n} b_{ij}\sum_{k\neq i}\sum_{l\neq j} b_{kl}\\ &= \frac{1}{n}\sum_{i,j=1}^{n} b_{ij}^2 + \frac{1}{n(n-1)}b_{ij}^2 = \frac{1}{n-1}\sum_{i,j=1}^{n} b_{ij}^2,\end{aligned}$$

using Corollary 7.3. □

The calculation for the third power is similar but longer.

Lemma 7.7. *We have*

$$\mathbb{E}[h^3] = \frac{n}{(n-1)(n-2)}\sum_{i,j=1}^{n} b_{ij}^3.$$

Proof. Cubing the defining expression (7.5) of h, we get

$$h^3 = \sum_{i,j=1}^{n} b_{ij}^3\mathbf{1}_{T_{i,j}} + 3\sum_{i,j=1}^{n}\sum_{\substack{k\neq i\\ l\neq j}} b_{ij}^2 b_{kl}\mathbf{1}_{T_{(i,j),(k,l)}} + \sum_{i,j=1}^{n}\sum_{\substack{k\neq i\\ l\neq j}}\sum_{\substack{p\neq i,k\\ q\neq j,l}} b_{ij}b_{kl}b_{pq}\mathbf{1}_{T_{(i,j),(k,l),(p,q)}}.$$

Taking expectations,

$$\mathbb{E}[h^3] = \frac{1}{n}\sum_{i,j=1}^{n} b_{ij}^3 + \frac{3}{n(n-1)}\sum_{i,j=1}^{n}\sum_{\substack{k\neq i\\ l\neq j}} b_{ij}^2 b_{kl} + \frac{1}{n(n-1)(n-2)}\sum_{i,j=1}^{n}\sum_{\substack{k\neq i\\ l\neq j}}\sum_{\substack{p\neq i,k\\ q\neq j,l}} b_{ij}b_{kl}b_{pq}. \tag{7.6}$$

The sum in the second term simplifies to

$$\sum_{i,j=1}^{n}\sum_{\substack{k\neq i\\ l\neq j}} b_{ij}^2 b_{kl} = \sum_{i,j=1}^{n} b_{ij}^3,$$

using Corollary 7.3. Simplifying the sum in the third term requires more work. First,

$$\sum_{p\neq i,k}\sum_{q\neq j,l} b_{pq} = -\sum_{p\neq i,k}(b_{pj}+b_{pl}) = b_{ij}+b_{il}+b_{kj}+b_{kl}.$$

Therefore for each i,j,

$$\sum_{\substack{k\neq i\\ l\neq j}}\sum_{\substack{p\neq i,k\\ q\neq j,l}} b_{kl}b_{pq} = \sum_{\substack{k\neq i\\ l\neq j}} b_{kl}(b_{ij}+b_{il}+b_{kj}+b_{kl}).$$

Summing over all i,j,

$$\begin{aligned}\sum_{i,j=1}^{n}\sum_{\substack{k\neq i\\ l\neq j}}\sum_{\substack{p\neq i,k\\ q\neq j,l}} b_{ij}b_{kl}b_{pq} &= \sum_{i,j=1}^{n} b_{ij}\sum_{\substack{k\neq i\\ l\neq j}} b_{kl}(b_{ij}+b_{il}+b_{kj}+b_{kl})\\ &= \sum_{i,j,k,l=1}^{n} b_{ij}b_{kl}(b_{ij}+b_{il}+b_{kj}+b_{kl})\\ &- \sum_{i,j,k=1}^{n} b_{ij}b_{kj}(2b_{ij}+2b_{kj})\\ &- \sum_{i,j,l=1}^{n} b_{ij}b_{il}(2b_{ij}+2b_{il})\\ &+ \sum_{i,j=1}^{n} b_{ij}^2(4b_{ij}) = 4\sum_{i,j=1}^{n} b_{ij}^3.\end{aligned}$$

All other terms other than the last vanish due to Corollary 7.3, since in each case we can find at least one index which appears only once. For example, in the first term $b_{ij}^2 b_{kl}$, both k,l appear only once, and in the second term $b_{ij}b_{kl}b_{il}$, both j,k appear only once. Substituting everything in (7.6), we get

$$\mathbb{E}[h^3] = \left(\frac{1}{n}+\frac{3}{n(n-1)}+\frac{4}{n(n-1)(n-2)}\right)\sum_{i,j=1}^{n} b_{ij}^3 = \frac{n}{(n-1)(n-2)}\sum_{i,j=1}^{n} b_{ij}^3. \qquad \square$$

7.4 Bounding the moments

Having related the moments of B to the moments of h, we estimate the moments of h.

Lemma 7.8. *We have* $\mathbb{E}[f_1] = c/n$, $\mathbb{E}[g] = c$, $\mathbb{E}[h] = 0$.

Proof. The first formula follows from the fact that the constant vector is in $L_{(n-1,1)}$. For the second formula, Lemma 7.5 implies that

$$\mathbb{E}[g] = \frac{n}{n-1}\mathbb{E}[f_1] + \frac{n-2}{n-1}c = \frac{1}{n-1}c + \frac{n-2}{n-1}c = c.$$

The third formula follows from $h = g - c$. $\square$

Lemma 7.9. *We have*

$$\mathbb{E}[g^2] = (1-\epsilon)\frac{n^2}{(n-1)^2}\frac{c}{n} + \frac{n(n-2)}{(n-1)^2}c^2,$$
$$\mathbb{E}[h^2] = (1-\epsilon)\frac{n^2}{(n-1)^2}\frac{c}{n} - \frac{c^2}{(n-1)^2}.$$

Proof. Since f_1 is a projection,

$$\mathbb{E}[f_1^2] = \mathbb{E}[f^2] - \mathbb{E}[(f-f_1)^2] = (1-\epsilon)\frac{c}{n}.$$

Using Lemma 7.5,

$$\begin{aligned}
\mathbb{E}[g^2] &= \mathbb{E}\left[\left(\frac{n}{n-1}f_1 + \frac{n-2}{n-1}c\right)^2\right] \\
&= \frac{n^2}{(n-1)^2}\mathbb{E}[f_1^2] + 2\frac{n(n-2)}{(n-1)^2}c\,\mathbb{E}[f_1] + \frac{(n-2)^2}{(n-1)^2}c^2 \\
&= (1-\epsilon)\frac{n^2}{(n-1)^2}\frac{c}{n} + 2\frac{n-2}{(n-1)^2}c^2 + \frac{(n-2)^2}{(n-1)^2}c^2 \\
&= (1-\epsilon)\frac{n^2}{(n-1)^2}\frac{c}{n} + \frac{n(n-2)}{(n-1)^2}c^2.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbb{E}[h^2] &= \mathbb{E}\left[\left(\frac{n}{n-1}f_1 - \frac{c}{n-1}\right)^2\right] \\
&= \frac{n^2}{(n-1)^2}\mathbb{E}[f_1^2] - 2\frac{n}{(n-1)^2}c\,\mathbb{E}[f_1] + \frac{c^2}{(n-1)^2} \\
&= (1-\epsilon)\frac{n^2}{(n-1)^2}\frac{c}{n} - 2\frac{c^2}{(n-1)^2} + \frac{c^2}{(n-1)^2} \\
&= (1-\epsilon)\frac{n^2}{(n-1)^2}\frac{c}{n} - \frac{c^2}{(n-1)^2}.
\end{aligned}$$

□

Estimating $\mathbb{E}[h^3]$ is more difficult. Instead of estimating $\mathbb{E}[h^3]$, we estimate the related quantity $\mathbb{E}[g^3]$, using the fact that $g \geq 0$ and that $g = n/(n-1)f_1 + (n-2)/(n-1)c$ is close to the function $n/(n-1)f + (n-2)/(n-1)c$, which takes only two different values. We will use the following inequality.

Lemma 7.10. *Let $\theta \in (0, 1/2]$ and let $H, L, \eta \geq 0$, where $H > L$. Suppose F is a measurable function defined on $[0,1]$ such that the measure of $F^{-1}(H)$ is θ and the measure of $F^{-1}(L)$ is $1-\theta$. If $G \geq 0$ is a measurable function defined on $[0,1]$ satisfying $\mathbb{E}[G] = \mathbb{E}[F]$ and*

$\mathbb{E}[(G-F)^2] \le \eta$, where $\eta \le \theta(1-\theta)(H-L)^2$, then

$$\mathbb{E}[G^3] \ge \mathbb{E}[F^3] - 3(H^2-L^2)\sqrt{\theta\eta} - \frac{1}{\sqrt{\theta/2}}\eta^{3/2}.$$

Here the expectations are over a random $x \in [0,1]$.

Proof. Without loss of generality, we can assume that F is given by

$$F(x) = \begin{cases} H & \text{if } 0 \le x < \theta, \\ L & \text{if } \theta \le x \le 1. \end{cases}$$

Let $\tilde{G}$ be another function on $[0,1]$ given by

$$\tilde{G}(x) = \begin{cases} \frac{1}{\theta}\int_0^\theta G(x)\,\mathrm{d}x & \text{if } 0 \le x < \theta, \\ \frac{1}{1-\theta}\int_\theta^1 G(x)\,\mathrm{d}x & \text{if } \theta \le x \le 1. \end{cases}$$

Clearly $\mathbb{E}[\tilde{G}(x)] = \mathbb{E}[G(x)]$. The Cauchy–Schwarz inequality (or the convexity of x^2 via Jensen's inequality) shows that

$$\begin{aligned}
\mathbb{E}[(G-F)^2] &= \int_0^\theta (G(x)-H)^2\,\mathrm{d}x + \int_\theta^1 (G(x)-L)^2\,\mathrm{d}x \\
&\ge \frac{1}{\theta}\left(\int_0^\theta (G(x)-H)\,\mathrm{d}x\right)^2 + \frac{1}{1-\theta}\left(\int_\theta^1 (G(x)-L)\,\mathrm{d}x\right)^2 \\
&= \theta\left(\frac{1}{\theta}\int_0^\theta G(x)\,\mathrm{d}x - H\right)^2 + (1-\theta)\left(\frac{1}{1-\theta}\int_\theta^1 G(x)\,\mathrm{d}x - L\right)^2 = \mathbb{E}[(\tilde{G}-F)^2].
\end{aligned}$$

Moreover, since x^3 is convex, Jensen's inequality shows that

$$\begin{aligned}
\mathbb{E}[G^3] &= \int_0^\theta G(x)^3\,\mathrm{d}x + \int_\theta^1 G(x)^3\,\mathrm{d}x \\
&= \theta\cdot\frac{1}{\theta}\int_0^\theta G(x)^3\,\mathrm{d}x + (1-\theta)\cdot\frac{1}{1-\theta}\int_\theta^1 G(x)^3\,\mathrm{d}x \\
&\ge \theta\left(\frac{1}{\theta}\int_0^\theta G(x)\,\mathrm{d}x\right)^3 + (1-\theta)\left(\frac{1}{1-\theta}\int_\theta^1 G(x)\,\mathrm{d}x\right)^3 = \mathbb{E}[\tilde{G}^3].
\end{aligned}$$

Since $\tilde{G} \ge 0$, we see that the minimum of $\mathbb{E}[G^3]$ is attained on functions of the form $\tilde{G}$.

The function $\tilde{G}$ attains only two values: $\tilde{G}(0)$ in the range $[0,\theta)$, and $\tilde{G}(1)$ in the range $[\theta,1]$. Since $\mathbb{E}[\tilde{G}] = \mathbb{E}[F] = \theta H + (1-\theta)L$, if we put $\delta = (H-\tilde{G}(0))/(1-\theta)$ then

$$\begin{aligned}
\tilde{G}(0) &= H - (1-\theta)\delta, \\
\tilde{G}(1) &= L + \theta\delta.
\end{aligned}$$

The first formula is by definition, and to verify the second, it is enough to check that these formulas agree with $\theta\tilde{G}(0) + (1-\theta)\tilde{G}(1) = \theta H + (1-\theta)L$. We get a bound on δ from the second moment:

$$\eta \geq \mathbb{E}[(\tilde{G} - F)^2] = \theta((1-\theta)\delta)^2 + (1-\theta)(\theta\delta)^2 = \theta(1-\theta)\delta^2.$$

Moreover,

$$\mathbb{E}[G^3] \geq \mathbb{E}[\tilde{G}^3] = \theta(H - (1-\theta)\delta)^3 + (1-\theta)(L + \theta\delta)^3.$$

The derivative of the right-hand side with respect to δ is

$$\begin{aligned}
&3\theta(1-\theta)[-(H-(1-\theta)\delta)^2 + (L+\theta\delta)^2] \\
=&3\theta(1-\theta)[(\theta^2 - (1-\theta)^2)\delta^2 + 2(\theta L + (1-\theta)H)\delta + L^2 - H^2] \\
=&3\theta(1-\theta)(\delta - (H-L))(H + L - (1-2\theta)\delta).
\end{aligned}$$

In particular, the right-hand side is decreasing in the range

$$-\frac{H+L}{1-2\theta} \leq \delta \leq H - L.$$

Hence, provided $\sqrt{\eta/(\theta(1-\theta))} \leq H - L$, we can conclude that

$$\mathbb{E}[G^3] \geq \theta(H - (1-\theta)\delta_0)^3 + (1-\theta)(L + \theta\delta_0)^3, \quad \text{where } \delta_0 = \sqrt{\frac{\eta}{\theta(1-\theta)}}.$$

(Note that $-(H+L)/(1-2\theta) \leq -(H-L)$ since $L \geq 0$ and $\theta \leq 1/2$.) Calculating,

$$\begin{aligned}
\mathbb{E}[G^3] \geq\ & \theta H^3 + (1-\theta)L^3 - 3\theta(1-\theta)(H^2 - L^2)\delta_0 \\
& + 3\theta(1-\theta)((1-\theta)H + \theta L)\delta_0^2 + \theta(1-\theta)(\theta^2 - (1-\theta)^2)\delta_0^3 \\
& = \mathbb{E}[F^3] - 3\sqrt{\theta(1-\theta)}(H^2 - L^2)\sqrt{\eta} + 3((1-\theta)H + \theta L)\eta - \frac{1-2\theta}{\sqrt{\theta(1-\theta)}}\eta^{3/2}.
\end{aligned}$$

The third term is positive, and the lemma follows using $0 \leq \theta \leq 1/2$. □

We can now get a lower bound on $\mathbb{E}[g^3]$.

Lemma 7.11. *Assuming $\epsilon_0 \leq 1/2$,*

$$\mathbb{E}[g^3] \geq \frac{n^2}{(n-1)^3}c + 3\frac{n(n-2)}{(n-1)^3}c^2 + \frac{(n-2)^2(n+1)}{(n-1)^3}c^3 - O\left(\sqrt{\epsilon}(1+c)\frac{c}{n}\right).$$

Proof. Define a function F by

$$F = \frac{n}{n-1} f + \frac{n-2}{n-1} c.$$

Although F and g are functions on S_n rather than on $[0,1]$, we can think of them as functions on $[0,1]$ by assigning each permutation $\pi \in S_n$ an interval of length $1/n!$; a random $x \in [0,1]$ corresponds to a random $\pi \in S_n$. Clearly $\mathbb{E}[F] = \mathbb{E}[g]$ and $g \geq 0$, and so we can apply Lemma 7.10. The relevant parameters are

$$\begin{aligned}
\theta &= \frac{c}{n}, \\
H &= \frac{n}{n-1} + \frac{n-2}{n-1} c, \\
L &= \frac{n-2}{n-1} c, \\
\eta &= \mathbb{E}[(F-g)^2] = \frac{n^2}{(n-1)^2} \mathbb{E}[(f-f_1)^2] = \epsilon \frac{n^2}{(n-1)^2} \frac{c}{n}.
\end{aligned}$$

Note that

$$\theta(1-\theta)(H-L)^2 \geq \frac{c}{2n} \frac{n^2}{(n-1)^2},$$

and so $\eta \leq \theta(1-\theta)(H-L)^2$ whenever $\epsilon \leq 1/2$. Therefore the lemma applies, and we get

$$\begin{aligned}
\mathbb{E}[g^3] &\geq \mathbb{E}[F^3] - 3\left(\frac{n^2}{(n-1)^2} + 2\frac{n(n-2)}{(n-1)^2} c \right) \frac{n}{n-1} \frac{c}{n} \sqrt{\epsilon} - \sqrt{\frac{2n}{c}} \frac{n^3}{(n-1)^3} \frac{c^{3/2}}{n^{3/2}} \epsilon^{3/2} \\
&= \mathbb{E}[F^3] - O\left(\sqrt{\epsilon}(1+c) \frac{c}{n} \right).
\end{aligned}$$

To complete the proof of the lemma, we compute $\mathbb{E}[F^3]$:

$$\begin{aligned}
\mathbb{E}[F^3] &= \theta H^3 + (1-\theta) L^3 \\
&= \frac{c}{n} \left(\frac{n^3}{(n-1)^3} + 3\frac{n^2(n-2)}{(n-1)^3} c + 3\frac{n(n-2)^2}{(n-1)^3} c^2 + \frac{(n-2)^3}{(n-1)^3} c^3 \right) + \left(1 - \frac{c}{n}\right) \frac{(n-2)^3}{(n-1)^3} c^3 \\
&= \frac{n^2}{(n-1)^3} c + 3\frac{n(n-2)}{(n-1)^3} c^2 + 3\frac{(n-2)^2}{(n-1)^3} c^3 + \frac{(n-2)^3}{(n-1)^3} c^3 \\
&= \frac{n^2}{(n-1)^3} c + 3\frac{n(n-2)}{(n-1)^3} c^2 + \frac{(n-2)^2(n+1)}{(n-1)^3} c^3.
\end{aligned}$$

□

Using our calculation of $\mathbb{E}[h^2]$, we can conclude a bound on $\mathbb{E}[h^3]$.

Lemma 7.12. *Assuming $\epsilon_0 \leq 1/2$,*

$$\mathbb{E}[h^3] \geq \frac{n^2}{(n-1)^3} c - \frac{3n}{(n-1)^3} c^2 + \frac{2}{(n-1)^3} c^3 - O\left(\sqrt{\epsilon}(1+c) \frac{c}{n} \right).$$

Proof. The starting point is

$$\mathbb{E}[g^3] = \mathbb{E}[(h+c)^3] = \mathbb{E}[h^3] + 3c\,\mathbb{E}[h^2] + 3c^2\,\mathbb{E}[h] + c^3.$$

Lemma 7.8 shows that $\mathbb{E}[h] = 0$, and Lemma 7.9 gives the value of $\mathbb{E}[h^2]$. Using the lower bound on $\mathbb{E}[g^3]$ given by Lemma 7.11, we get a lower bound on $\mathbb{E}[h^3]$:

$$\begin{aligned}
\mathbb{E}[h^3] &= \mathbb{E}[g^3] - 3c\,\mathbb{E}[h^2] - c^3 \\
&\geq \frac{n^2}{(n-1)^3}c + 3\frac{n(n-2)}{(n-1)^3}c^2 + \frac{(n-2)^2(n+1)}{(n-1)^3}c^3 - O\Big(\sqrt{\epsilon}(1+c)\frac{c}{n}\Big) \\
&- 3(1-\epsilon)\frac{n}{(n-1)^2}c^2 + 3\frac{c^3}{(n-1)^2} - c^3 \\
&= \frac{n^2}{(n-1)^3}c - \frac{3n}{(n-1)^3}c^2 + \frac{2}{(n-1)^3}c^3 - O\Big(\sqrt{\epsilon}(1+c)\frac{c}{n}\Big).
\end{aligned}$$

□

Applying Lemma 7.6 and Lemma 7.7, we get good estimates on the moments of the matrix B.

Lemma 7.13. *Assuming $\epsilon_0 \leq 1/2$,*

$$\begin{aligned}
\sum_{i,j=1}^n b_{ij}^2 &= (1-\epsilon)\frac{n}{n-1}c - \frac{1}{n-1}c^2, \\
\sum_{i,j=1}^n b_{ij}^3 &\geq \frac{n(n-2)}{(n-1)^2}c - \frac{3(n-2)}{(n-1)^2}c^2 + \frac{2(n-2)}{n(n-1)^2}c^3 - O\big(\sqrt{\epsilon}(1+c)c\big).
\end{aligned}$$

Proof. The first formula follows from Lemma 7.6 and Lemma 7.9. The second formula follows from Lemma 7.7 and Lemma 7.12. □

If $\mathcal{F}$ is a sum of c disjoint cosets then $\epsilon = 0$, and in this case

$$\begin{aligned}
\sum_{i,j=1}^n b_{ij}^2 &= \frac{n}{n-1}c - \frac{1}{n-1}c^2 = \frac{c(n-c)}{n-1}, \\
\sum_{i,j=1}^n b_{ij}^3 &= \frac{n(n-2)}{(n-1)^2}c - \frac{3(n-2)}{(n-1)^2}c^2 + \frac{2(n-2)}{n(n-1)^2}c^3 = \frac{(n-2)(n-c)(n-2c)c}{n(n-1)^2}.
\end{aligned}$$

This can be checked either directly or by going through the proof: we have been careful not to drop any term which isn't multiplied by a power of ϵ. For general $\mathcal{F}$, the values of these moments are somewhat smaller. When c is small, the difference is not significant, and it is this fact that drives our proof.

7.5 Culmination of the proof

We start by giving a lower bound on c. This will allow us to replace $1 + c$ with $O(c)$ in the statement of Lemma 7.13.

Lemma 7.14. *Assuming ϵ_0 is small enough and n is large enough, $c \geq 1/2$.*

Proof. Suppose that $c \leq 1$. Lemma 7.13 shows that the second moment of B is at most $(1 + O(1/n))c$. The convexity of $x^{3/2}$ implies (via Lemma 2.2) that

$$\sum_{i,j} b_{ij}^3 = \sum_{i,j} (b_{ij}^2)^{3/2} \leq \left(\sum_{i,j} b_{ij}^2\right)^{3/2} = (1 + O(1/n))^{3/2} c^{3/2} = (1 + O(1/n)) c^{3/2}.$$

On the other hand, Lemma 7.13 shows that

$$\sum_{i,j} b_{ij}^3 \geq (1 - O(1/n) - O(\sqrt{\epsilon}))c.$$

Together, we get $\sqrt{c} \geq 1 - O(1/n) - O(\sqrt{\epsilon})$, and so for ϵ_0 small enough and n large enough, $c \geq 1/2$. (Here $1/2$ is an arbitrary constant smaller than 1.) □

This allows us to rewrite Lemma 7.13 more succinctly.

Lemma 7.15. *Assuming ϵ_0 is small enough and n is large enough,*

$$c - O\left(\epsilon c + \frac{c^2}{n}\right) \leq \sum_{i,j=1}^{n} b_{ij}^2 \leq c + O\left(\frac{c}{n}\right),$$
$$\sum_{i,j=1}^{n} b_{ij}^3 \geq c - O\left(\sqrt{\epsilon} c^2 + \frac{c^2}{n}\right).$$

The idea now is to subtract the two moments, and conclude that there must be roughly c strong values.

Lemma 7.16. *Assuming ϵ_0 is small enough and n is large enough, there exists a set $M \subseteq [n]^2$ of size m such that*

$$\sum_{(i,j) \in M} b_{ij} \geq m - O\left(\sqrt{\epsilon} c^2 + \frac{c^2}{n}\right).$$

Furthermore,

$$|m - c| = O\left(\sqrt{\epsilon} c^2 + \frac{c^2}{n}\right).$$

Proof. Let $E = \sqrt{\epsilon}c^2 + c^2/n$. By Lemma 7.14, we can assume that $c \geq 1/2$. Lemma 7.15 implies that

$$\sum_{i,j=1}^{n} b_{ij}^2(1 - b_{ij}) = O(E).$$

Let $M = \{(i,j) \in [n]^2 : b_{ij} \geq 1/2\}$ be the set of strong entries, and let $m = |M|$. We have

$$\sum_{(i,j)\notin M} b_{ij}^2 \leq 2 \sum_{(i,j)\notin M} b_{ij}^2(1 - b_{ij}) = O(E).$$

Therefore

$$m \geq \sum_{(i,j)\in M} b_{ij}^2 \geq c - O(E), \tag{7.7}$$

since $b_{ij} \leq a_{ij} \leq 1$. On the other hand,

$$\sum_{(i,j)\in M} (1 - b_{ij}) \leq 4 \sum_{(i,j)\in M} b_{ij}^2(1 - b_{ij}) = O(E).$$

Therefore

$$\sum_{(i,j)\in M} b_{ij} \geq m - O(E).$$

Using $x^2 \geq 2x - 1$, we also get

$$\sum_{(i,j)\in M} b_{ij}^2 \geq \sum_{(i,j)\in M} (2b_{ij} - 1) = m - O(E).$$

On the other hand,

$$\sum_{(i,j)\in M} b_{ij}^2 \leq \sum_{i,j=1}^{n} b_{ij}^2 = c + O\left(\frac{c}{n}\right).$$

The last two equations together show that $m - c = O(E)$. Equation (7.7) shows that $c - m = O(E)$, and we conclude that $|m - c| = O(E)$. □

By modifying M slightly, we deduce a version of Lemma 7.16 in which $|m - c| < 1$.

Lemma 7.17. *Assuming ϵ_0 is small enough and n is large enough, there exists a set $M \subseteq [n]^2$ of size m such that*

$$\sum_{(i,j)\in M} b_{ij} \geq m - O\left(\sqrt{\epsilon}c^2 + \frac{c^2}{n}\right).$$

Furthermore,

$$|m - c| < \min\left(1, O\left(\sqrt{\epsilon}c^2 + \frac{c^2}{n}\right)\right).$$

Proof. Let M' be the set of size m' given by Lemma 7.16. If $|m' - c| < 1$ then we take $M = M'$. Otherwise, it is enough to find a set M whose size m satisfies $|m - c| < 1$.

If $m' \le c$ then we let M consist of M' along with $d = \lfloor c \rfloor - m'$ additional arbitrary elements. Note that $d \le c - m' = O(\sqrt{\epsilon}c^2 + c^2/n)$. Since $b_{ij} \ge -c/n$, we have

$$\begin{aligned}\sum_{(i,j)\in M} b_{ij} &\ge m' - O\left(\sqrt{\epsilon}c^2 + \frac{c^2}{n}\right) \\ &\ge m - d - \frac{cd}{n} - O\left(\sqrt{\epsilon}c^2 + \frac{c^2}{n}\right) \\ &\ge m - O\left(\sqrt{\epsilon}c^2 + \frac{c^2}{n}\right).\end{aligned}$$

If $m' \ge c$ then let M contain the $m = \lceil c \rceil$ largest elements in M'. Since $m \le m'$, we have

$$\sum_{(i,j)\in M} b_{ij} \ge \frac{m}{m'}m' - \frac{m}{m'}O\left(\sqrt{\epsilon}c^2 + \frac{c^2}{n}\right) \ge m - O\left(\sqrt{\epsilon}c^2 + \frac{c^2}{n}\right). \qquad \square$$

Using this lemma, we can essentially prove our main theorem.

Lemma 7.18. *There is an $\epsilon_0 > 0$ such that the following holds.*

There exists a family $\mathcal{G}$ which is the union of m cosets satisfying

$$|\mathcal{F}\Delta\mathcal{G}| = O\left(\sqrt{\epsilon} + \frac{1}{n}\right)c^2(n-1)!.$$

Furthermore,

$$|m - c| < \min\left(1, O\left(\sqrt{\epsilon}c^2 + \frac{c^2}{n}\right)\right).$$

Proof. Pick ϵ_0 so that the conditions in Lemma 7.14 and Lemma 7.17 hold given $n \ge N$, for some N. For $n < N$, if $c = 0$ then the lemma trivially holds, and otherwise we can choose the constant in the first $O(\cdot)$ so that the theorem becomes trivial. From now on, assume $n \ge N$.

Let M be the set of m indices given by Lemma 7.17, and let the family $\mathcal{G}$ be given by

$$\mathcal{G} = \bigcup_{(i,j)\in M} T_{i,j}.$$

Lemma 7.16 shows that

$$\begin{aligned}\sum_{(i,j)\in M} |\mathcal{F} \cap T_{i,j}| \ge (n-1)! \sum_{(i,j)\in M} b_{ij} &= (n-1)!\left(m - O\left(\sqrt{\epsilon}c^2 + \frac{c^2}{n}\right)\right). \\ &= (n-1)!\left(c - O\left(\sqrt{\epsilon}c^2 + \frac{c^2}{n}\right)\right).\end{aligned}$$

Since $|T_{i,j} \cap T_{k,l}| \le (n-2)!$ whenever $(i,j) \neq (k,l)$,

$$\begin{aligned}|\mathcal{F} \cap \mathcal{G}| &\ge (n-1)!\left(c - O\left(\sqrt{\epsilon}c^2 + \frac{c^2}{n}\right)\right) - \binom{m}{2}(n-2)! \\ &= (n-1)!\left(c - O\left(\sqrt{\epsilon}c^2 + \frac{c^2}{n} + \frac{m^2}{n}\right)\right) \\ &= (n-1)!\left(c - O\left(\sqrt{\epsilon}c^2 + \frac{c^2}{n}\right)\right).\end{aligned}$$

We used $m < c + 1 = O(c)$, since $c \ge 1/2$ by Lemma 7.14. Therefore

$$\begin{aligned}|\mathcal{F} \Delta \mathcal{G}| &= |\mathcal{F}| + |\mathcal{G}| - 2(|\mathcal{F} \cap \mathcal{G}|) \\ &\le (m+c)(n-1)! - 2c(n-1)! + O\left(\sqrt{\epsilon}c^2 + \frac{c^2}{n}\right)(n-1)! \\ &\le O\left(\sqrt{\epsilon}c^2 + \frac{c^2}{n}\right)(n-1)!.\end{aligned}$$

□

The main theorem easily follows

Theorem 7.1. *There is an $\epsilon_0 > 0$ such that the following holds.*

Let $\mathcal{F} \subseteq S_n$ be a family of permutations of size $c(n-1)!$, where $c \le n/2$. Let $f = \mathbf{1}_{\mathcal{F}}$ (so $\mathbb{E}[f] = c/n$) and let $f_1 = \hat{f}((n)) + \hat{f}((n-1,1))$ be the projection of f to $L_{(n-1,1)}$.

If $\mathbb{E}[(f - f_1)^2] = \epsilon c/n$, where $\epsilon \le \epsilon_0$, then there exists a family $\mathcal{G} \subseteq S_n$ which is the union of $\lceil c \rfloor$ cosets satisfying

$$|\mathcal{F} \Delta \mathcal{G}| = O\left(\sqrt{\epsilon} + \frac{1}{n}\right)c^2(n-1)!.$$

Moreover,

$$|c - \lceil c \rfloor| = O\left(\sqrt{\epsilon} + \frac{1}{n}\right)c^2.$$

Proof. Let $\mathcal{G}'$ be the family given by Lemma 7.18. Since m is an integer, the inequality on $|m - c|$ implies that

$$|c - \lceil c \rfloor| = O\left(\sqrt{\epsilon}c^2 + \frac{c^2}{n}\right).$$

If $m = \lceil c \rfloor$ then we can take $\mathcal{G} = \mathcal{G}'$. Otherwise, form $\mathcal{G}$ by taking $\mathcal{G}'$ and either adding or removing $|m - \lceil c \rfloor|$ cosets (note that $|m - \lceil c \rfloor| = 1$). Since

$$|m - \lceil c \rfloor| \le |m - c| + |c - \lceil c \rfloor| = O\left(\sqrt{\epsilon}c^2 + \frac{c^2}{n}\right),$$

we deduce

$$|\mathcal{F} \Delta \mathcal{G}| \le |\mathcal{F} \Delta \mathcal{G}'| + |m - \lceil c \rfloor|(n-1)! = O\left(\sqrt{\epsilon} + \frac{1}{n}\right)c^2(n-1)!.$$

□

7.6 Intersecting families of permutations

As an application of Theorem 7.1, we prove two stability results for intersecting families of permutations. Our first stability result is a straightforward application of Theorem 7.1.

Theorem 7.19. *Let $\mathcal{F} \subseteq S_n$ be an intersecting family of permutations of measure $(1-\delta)/n$, where δ is small enough. Then there is a coset $\mathcal{G}$ such that*

$$|\mathcal{F} \triangle \mathcal{G}| = O\left(\sqrt{\delta} + \frac{1}{n}\right)(n-1)!.$$

Proof. Let $f = \mathbf{1}_{\mathcal{F}}$. Theorem 6.15 shows that $\|f - f_1\|^2 = O(\delta/n)$. We apply Theorem 7.1 with $c = 1-\delta$ and $\epsilon = O(\delta)$. If δ is small enough then $\epsilon \le \epsilon_0$. Theorem 7.1 then gives us a family $\mathcal{G}$ which is the union of $\lceil c \rfloor = 1$ cosets satisfying

$$|\mathcal{F}\Delta\mathcal{G}| = O\left(\sqrt{\delta} + \frac{1}{n}\right)(n-1)!. \qquad \square$$

Using a slightly more complicated argument, we get a stronger stability result.

Theorem 7.20. *There exists $\delta > 0$ such that for large enough n, every intersecting family of permutations of measure at least $(1-\delta)/n$ is contained in a coset.*

Proof. Let $\mathcal{F}$ be an intersecting family of permutations of size at least $(1-\delta)/n$. If δ is small enough, then Theorem 7.19 applies, and there is a coset $T_{i,j}$ such that

$$|\mathcal{F} \triangle T_{i,j}| = O\left(\sqrt{\delta} + \frac{1}{n}\right)(n-1)!.$$

Suppose $\mathcal{F}$ is not contained in $\mathcal{G}$, and pick some $\pi \in \mathcal{F} \setminus \mathcal{G}$. Without loss of generality, assume $i = j = 1$ and $\pi = (12)$. Every fixed-point free permutation in $S_{\{2,\ldots,n\}}$ lifts to a permutation in $T_{1,1}$ which does not intersect π. Since there are $(1 - 1/e \pm O(1/n))(n-1)!$ of these,

$$|\mathcal{F} \triangle T_{i,j}| \ge \left(\frac{1}{e} - O\left(\frac{1}{n}\right)\right)(n-1)!.$$

If δ is small enough and n is large enough, this leads to a contradiction. $\square$

Using more complicated arguments (and before Theorem 7.1 was proved), Ellis [22] was able to prove a much sharper result: the largest intersecting family of permutations not contained in a coset has size $(1 - 1/e + O(1/n))(n-1)!$. In fact, for $n \ge 6$, the maximal family is

$$\mathcal{F} = \{\pi \in S_n : \pi(1) = 1 \text{ and } \pi(i) = i \text{ for some } i > 2\} \cup \{(12)\}.$$

Moreover, all maximal families are of the form $\alpha\mathcal{F}\beta$. For $n \le 5$, another optimal family is the one in which every permutation has at least two fixed points.

Ellis [21] proved a similar result for t-intersecting families of permutations. Such a family must be contained in a t-coset (a family of the form $T_{(i_1,j_1),\ldots,(i_t,j_t)}$) unless it is smaller than $(1 - 1/e + O(1/n))(n-t)!$, and for large enough n, the maximal families are of the form $\alpha\mathcal{F}\beta$, where

$$\mathcal{F} = \{\pi \in S_n : \pi(1) = 1 \text{ and } \pi(i) = i \text{ for some } i > t+1\} \cup \{(i\,(t+1)) : 1 \le i \le t\}.$$

These results are analogous to a classical result by Hilton and Milner [49] in the context of k-uniform intersecting families of sets on n points. They proved that for $k < n/2$, the maximal families not contained in a star are of the form

$$\mathcal{F} = \{A \subseteq [n] : |A| = k, 1 \in A, A \cap \{2,\ldots,k+1\} \neq \varnothing\} \cup \{2,\ldots,k+1\}.$$

Ahlswede and Khachatrian [1] proved a similar theorem for t-intersecting uniform families of sets. In this case, the optimal family is always one of the families described in Section 10.1 (other than the restriction of a t-star).

Chapter 8

A structure theorem for balanced dictatorships on S_n

In this chapter we continue our study of Boolean functions on S_n whose Fourier expansion is concentrated on the first two levels. In the previous chapter, we considered functions which contained $c(n-1)!$ permutations, where c is small. In this chapter we consider the case where $c = \Theta(n)$. We will use the slightly different parametrization $\mathfrak{c} = c/n$. Our goal in this chapter is to prove the following theorem.

Theorem 8.1. *Let $\mathcal{F} \subseteq S_n$ be a family of permutations of size $\mathfrak{c}n!$, and let $\eta = \min(\mathfrak{c}, 1-\mathfrak{c})$. Let $f = 2\mathbf{1}_{\mathcal{F}} - 1$, and let $f_1 = \hat{f}((n)) + \hat{f}((n-1,1))$ be the projection of f to $L_{(n-1,1)}$.*

If $\mathbb{E}[(f-f_1)^2] = \epsilon$ then there exists a family $\mathcal{G} \subseteq S_n$ which is the union of $\lceil \mathfrak{c}n \rfloor$ disjoint cosets satisfying

$$|\mathcal{F} \Delta \mathcal{G}| = O\left(\frac{1}{\eta}\left(\epsilon^{1/7} + \frac{1}{n^{1/3}}\right)\right) n!.$$

Moreover,

$$|\mathfrak{c}n - \lceil \mathfrak{c}n \rfloor| = O\left(\left(\epsilon^{1/7} + \frac{1}{n^{1/3}}\right) n\right).$$

A minor difference from Theorem 7.1 is our definition of f: there, we defined $f = \mathbf{1}_{\mathcal{F}}$, whereas here, we define $f = 2\mathbf{1}_{\mathcal{F}} - 1$. Let $f^{(1)}, f_1^{(1)}$ correspond to the definition in Theorem 7.1, and let $f^{(2)}, f_1^{(2)}$ correspond to the definition in Theorem 8.1. We have $f^{(2)} = 2f^{(1)} - 1$ and

$f_1^{(2)} = 2f_1^{(1)} - 1$, and so

$$\mathbb{E}[(f^{(2)} - f_1^{(2)})^2] = 4\,\mathbb{E}[(f^{(1)} - f_1^{(1)})]^2.$$

Therefore if we replace the definitions of f and f_1 with their definitions in Theorem 8.1, we get essentially the same result. We use the present definition so that f becomes a ± 1 function, a convenient device often used in Fourier analysis.

In contrast to Theorem 7.1, in this theorem we are guaranteed that the family $\mathcal{G}$ is a union of *disjoint* cosets. Therefore $\mathcal{G}$ is either of the form

$$\mathcal{G} = \{\pi \in S_n : \pi(i) \in A\}$$

or of the form

$$\mathcal{G} = \{\pi \in S_n : \pi^{-1}(i) \in A\},$$

where $|A| = \lceil \mathfrak{c} n \rfloor$.

In the case of Theorem 7.1, we were not able to guarantee the disjointness of the cosets forming $\mathcal{G}$ since a family of the form $T_{1,1} \cup T_{2,2}$ satisfies the hypotheses of the theorem. In the regime considered in the present chapter, a union of $\mathfrak{c}n$ arbitrary cosets will be far from its projection to $L_{(n-1,1)}$: $\mathbb{E}[(f - f_1)^2] = \Theta(\mathfrak{c}^2)$. This is intuitively why we are able to guarantee the disjointness in the present case.

The proofs of Theorem 7.1 and Theorem 8.1 are rather different. While the proof of Theorem 7.1 consists of analyzing the first few moments of a matrix consisting, essentially, of entries $|\mathcal{F} \cap T_{i,j}|/(n-1)!$, in the present case these moments are not enough to distinguish a matrix corresponding to the union of roughly $\mathfrak{c}n$ cosets from an arbitrary matrix. The argument is therefore completely different.

The material in this chapter is taken from our joint paper with David Ellis and Ehud Friedgut [26].

8.1 Overview of the proof

For simplicity in this overview, we assume throughout that $\mathfrak{c} = n/2$. For technical reasons, for most of the proof we make the assumption $\epsilon \geq 1/n^{7/3}$.

Our starting point is the matrix A given by

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

This is different from the matrix A appearing in the proof of Theorem 8.1, though it is an affine shift of the matrices A and B appearing there. When $\mathcal{F} = T_{1,1} \cup \cdots \cup T_{1,n/2}$, the matrix A looks like this:

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

The matrix has a strong first row, consisting of entries which are close to ± 1, and all other entries are small, of order $1/n$.

The coefficients a_{ij} satisfy the equation

$$f_1 = \sum_{i,j=1}^{n} a_{ij} \mathbf{1}_{T_{i,j}}.$$

Indeed, this is the reason behind choosing this particular formula for a_{ij}. This equation implies the fundamental formula

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

which we will use over and over again.

The proof breaks down into three main parts. In the first part, we show that for almost all $\pi \in S_n$, the generalized diagonal defined by π in A, namely $\{a_{i\pi(i)} : i \in [n]\}$, has precisely one entry which is large (close to ± 1), and all the rest of its entries are small. In the second part, we deduce that A has either a row or a column in which almost all entries are large. In the third part, we extract a good approximation to $\mathcal{F}$ from the strong row or column highlighted by the second part.

The first part, comprising Sections 8.4 to 8.6, consists of two steps. In the first step, we consider a restriction of f_1 into permutations belonging to a set of the form $T_{X,Y} = \{\pi \in S_n :$

$\pi(X) = Y\}$. The value of f_1 on permutations from $T_{X,Y}$ depends only on entries a_{ij} for which $(i,j) \in X \times Y$ or $(i,j) \in \overline{X} \times \overline{Y}$ (here $\overline{X} = [n] \setminus X$). We separate the effects of these two submatrices by defining

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

Here $\pi_1 = \pi|_X$ and $\pi_2 = \pi|_{\overline{X}}$, and the domain of all functions is $T_{X,Y}$. Using straightforward estimates, we show in Section 8.4 that for most choices of X, Y, the functions g_1, g_2, g are well-behaved: the function g is close to ± 1, and $\mathbb{E}\, g_1 \approx \mathbb{E}\, g_2 \approx 0$. Using the decomposition $g = g_1 + g_2$, we conclude in Section 8.5 that one of g_1, g_2 must be essentially constant, and the other must be close to ± 1. This is intuitively obvious: there is no other way for g to be close to ± 1.

In the second step, we turn the table around: instead of first choosing the restriction (X, Y) and then looking at all permutations $\pi \in T_{X,Y}$, we first choose the permutation $\pi \in S_n$ and then consider all restrictions (X, Y) such that $\pi \in T_{X,Y}$. Essentially using the Friedgut–Kalai–Naor theorem, we conclude in Section 8.6 that for almost all $\pi \in S_n$, the generalized diagonal defined by π in A has precisely one entry which is close to ± 1, and $n - 1$ small entries.

The second part, consisting of Section 8.7, is largely independent of the first part. We consider the following abstract problem. We are given a matrix in which entries are either *large* or *small*. Furthermore, a $1 - \delta$ fraction of the generalized diagonals in this matrix contain exactly one large entry. We show that the matrix must contain a strong line: either a row or a column in which $1 - O(\delta)$ of the entries are large.

In the third part, consisting of Sections 8.8 and 8.9, we combine the results in both parts to prove the main theorem. Section 8.8 applies the result of the second part to the matrix A, and obtains the desired approximating family $\mathcal{G}$ in a straightforward manner. Finally, in Section 8.9 we discharge the assumption $\epsilon \geq 1/n^{7/3}$ using a perturbation argument, as well as other assumptions detailed in Section 8.3.

8.2 Nomenclature

In this section we collect some simple definitions which will be during the proof.

The notation $x \pm \gamma$ is shorthand for the closed interval $[x-\gamma, x+\gamma]$ (γ will usually be a small positive number). We say that x is *γ-close to y* if $|x-y| \le \gamma$, which is the same as $x \in y \pm \gamma$ or $y \in x \pm \gamma$. Otherwise, we say that x is *γ-far from y*. We say that x is *γ-close to y in magnitude* if $|x|$ is γ-close to y. We say that x is γ-close to $y \pm z$ if either x is γ-close to $y+z$ or x is γ-close to $y-z$. For a set S, we say that x is γ-close to S if x is γ-close to some $y \in S$.

Given a function ϕ and a probability distribution X over its domain (which will always be clear from the context), we say that ϕ is *(δ,γ)-almost close to C* if

$$\Pr_{x \sim X}[\phi(x) \in C \pm \gamma] \ge 1-\delta.$$

In words, with probability $1-\delta$, $\phi(x)$ is γ-close to C. Similarly, we say that ϕ is *(δ,γ)-almost Boolean* if $|\phi|$ is (δ,γ)-close to 1, that is

$$\Pr_{x \sim X}[|\phi(x)| \in 1 \pm \gamma] \ge 1-\delta.$$

We will apply this terminology to restrictions of f_1, which is why by almost Boolean we actually mean close to ± 1.

For a permutation $\pi \in S_n$ and $X \subseteq [n]$, $\pi|_X$ is the restriction of π to the set X, which is a bijection from X to $\pi(X)$.

Other notation will be defined along the way.

8.3 Basic definitions

We now begin the proof proper. For most of the proof (until Section 8.9), we make the following assumptions:

$$n \ge 4, \quad \frac{1}{n^{7/3}} \le \epsilon \le c_0 \eta^7. \tag{8.1}$$

Here c_0 is a constant which arises from the proof. In fact, during the proof, we will use the phrase *since ϵ is small enough compared to η^7*. By that we mean that the proposition qualified by the phrase holds if ϵ is small enough compared to η^7, and by choosing c_0 appropriately, we can assume that the required condition on ϵ is satisfied. Since $\eta \le 1/2$, these assumptions also imply that $\epsilon \le c_0/2^7$. When we use the phrase *since ϵ is small enough*, we mean that the qualified proposition holds if ϵ is small enough, which again can be guaranteed by choosing c_0 appropriately.

We also assume throughout the proof (until Section 8.9) that $\mathcal{F}, f, f_1, \mathfrak{c}, \eta, \epsilon$ are all given.

The rest of this section mirrors Section 7.2. However, since the basic definitions are slightly different, we repeat all of the proofs.

Since $f_1 \in L_{(n-1,1)}$, it can be written as a linear combination of $T_{i,j}$. We single out one such way. Let

$$a_{ij} = (n-1)\langle f, \mathbf{1}_{T_{i,j}} \rangle - \frac{n-2}{n}(2\mathfrak{c}-1). \tag{8.2}$$

The entries a_{ij} form an $n \times n$ matrix A.

Lemma 8.2. *Each row and each column of A sums to $2\mathfrak{c}-1$.*

Proof. Since the cosets $T_{1,j}$ partition S_n, we have

$$\sum_{j=1}^{n} a_{1j} = (n-1)\langle f, \mathbf{1}_{S_n} \rangle - (n-2)(2\mathfrak{c}-1) = (n-1)(2\mathfrak{c}-1) - (n-2)(2\mathfrak{c}-1) = 2\mathfrak{c}-1.$$

A similar argument works for the sum of any other row or of any column. □

This implies the following formula for double sums.

Lemma 8.3. *For each $i, j \in [n]$,*

$$\sum_{k \neq i} \sum_{l \neq j} a_{kl} = a_{ij} + (n-2)(2\mathfrak{c}-1).$$

Proof. Using Lemma 8.2,

$$\sum_{k \neq i} \sum_{l \neq j} a_{kl} = \sum_{k \neq i} (2\mathfrak{c}-1-a_{kj}) = (n-1)(2\mathfrak{c}-1) - (2\mathfrak{c}-1-a_{ij}) = a_{ij} + (n-2)(2\mathfrak{c}-1).$$ □

We can now prove the formula for f_1.

Lemma 8.4. *We have*

$$f_1 = \sum_{i,j=1}^{n} a_{ij} \mathbf{1}_{T_{i,j}},$$

and so for all $\pi \in S_n$,

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

Proof. Since both sides in the first equation belong to $L_{(n-1,1)}$, it is enough to show that both sides have the same inner product with each $\mathbf{1}_{T_{k,l}}$. Using Lemma 7.4 and Lemma 8.3, we have

$$\begin{aligned}
\left\langle \sum_{i,j=1}^{n} a_{ij}\mathbf{1}_{T_{i,j}}, \mathbf{1}_{T_{k,l}} \right\rangle &= \sum_{i,j=1}^{n} a_{ij}\langle \mathbf{1}_{T_{i,j}}, \mathbf{1}_{T_{k,l}} \rangle \\
&= \frac{a_{kl}}{n} + \frac{1}{n(n-1)} \sum_{i\neq k}\sum_{j\neq l} a_{ij} \\
&= \frac{a_{kl}}{n} + \frac{a_{kl}}{n(n-1)} + \frac{n-2}{n(n-1)}(2\mathfrak{c}-1) \\
&= \frac{a_{kl}}{n-1} + \frac{n-2}{n(n-1)}(2\mathfrak{c}-1) = \langle f, \mathbf{1}_{T_{k,l}} \rangle = \langle f_1, \mathbf{1}_{T_{k,l}} \rangle.
\end{aligned}$$

The second equation follows from the first since $\pi \in T_{i,j}$ if and only if $j = \pi(i)$. □

We proceed to calculate the sum of squares of the entries of A.

Lemma 8.5. *We have*

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

Proof. Since f_1 is an orthogonal projection of f and $f(\pi) \in \{\pm 1\}$,

$$\|f_1\|^2 = \|f\|^2 - \|f - f_1\|^2 = 1 - \epsilon.$$

Using Lemma 8.4, Lemma 7.4 and Lemma 8.2, we get

$$\begin{aligned}
\|f_1\|^2 &= \sum_{i,j,k,l=1}^{n} a_{ij}a_{kl}\langle \mathbf{1}_{T_{i,j}}, \mathbf{1}_{T_{k,l}} \rangle \\
&= \frac{1}{n}\sum_{i,j=1}^{n} a_{ij}^2 + \frac{1}{n(n-1)}\sum_{i,j=1}^{n} a_{ij}\sum_{k\neq i}\sum_{l\neq j} a_{kl} \\
&= \frac{1}{n}\sum_{i,j=1}^{n} a_{ij}^2 + \frac{1}{n(n-1)}\sum_{i,j=1}^{n} \left(a_{ij}^2 + (n-2)(2\mathfrak{c}-1)a_{ij}\right) \\
&= \frac{1}{n-1}\sum_{i,j=1}^{n} a_{ij}^2 + \frac{n(n-2)(2\mathfrak{c}-1)^2}{n(n-1)} = \frac{1}{n-1}\left(\sum_{i,j=1}^{n} a_{ij}^2 + (n-2)(2\mathfrak{c}-1)^2\right).
\end{aligned}$$ □

8.4 Random restrictions

In this section, we analyze the behavior of f_1 under random restrictions[1].

[1]Pun intended.

Definition 8.1. A *restriction* (X, Y) is composed of two subsets $X, Y \subseteq [n]$ of the same magnitude. We define a distribution $\mathcal{R}$ over restrictions as follows: choose X by including each $i \in [n]$ with probability $1/2$ independently, and choose Y at random among all sets of size $|X|$. If $(X, Y) \sim \mathcal{R}$ then we call (X, Y) a *random restriction.*

Given a restriction (X, Y), its *complement* $(\overline{X}, \overline{Y})$ is given by $\overline{X} = [n] \setminus X$, $\overline{Y} = [n] \setminus Y$. ○

The idea is to look at the behavior of f_1 for all permutations $\pi \in S_n$ satisfying $\pi(X) = Y$. The behavior of f_1 on permutations of that form depends only on the two submatrices of A supported by $X \times Y$ and by $\overline{X} \times \overline{Y}$. We decouple the effects of the two submatrices by decomposing f_1 on this set of permutations as a sum of two functions.

Definition 8.2. For a restriction (X, Y), the associated set of permutations is

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

We define the following functions, whose domain is $T_{X,Y}$:

$$g_1^{X,Y} = \sum_{\substack{i \in X \\ j \in Y}} a_{ij} \mathbf{1}_{T_{i,j}}, \quad g_2^{X,Y} = \sum_{\substack{i \in \overline{X} \\ j \in \overline{Y}}} a_{ij} \mathbf{1}_{T_{i,j}}, \quad g^{X,Y} = g_1^{X,Y} + g_2^{X,Y}.$$

Let $m_{X,Y} = \mathbb{E}[g_1^{X,Y}]$, where the expectation is taken over all $\pi \in T_{X,Y}$. ○

The function $g^{X,Y}$ is in fact identical to f_1 on $T_{X,Y}$.

Lemma 8.6. *For a restriction (X, Y) and $\pi \in T_{X,Y}$, $g^{X,Y}(\pi) = f_1(\pi)$.*

Proof. Follows immediately from Lemma 8.4. □

The goal of this section is to determine some typical properties of the functions $g_1^{X,Y}, g_2^{X,Y}, g^{X,Y}$ which hold for a random restriction (X, Y). We start by determining the expectation and variance of $m_{X,Y} = \mathbb{E}[g_1^{X,Y}]$. This will allow us to determine the value of $m_{X,Y}$ to a large accuracy.

Lemma 8.7. *Let $(X, Y) \sim \mathcal{R}$. Then $\mathbb{E}[m_{X,Y}] = \mathfrak{c} - 1/2$ and $\mathbb{V}[m_{X,Y}] \le 1/(2n)$.*

Proof. We start with an explicit formula for $m_{X,Y}$:

$$m_{X,Y} = \mathbb{E}[g_1^{X,Y}] = \sum_{\substack{i \in X \\ j \in Y}} a_{ij} \mathop{\mathbb{E}}_{T_{X,Y}}[\mathbf{1}_{T_{i,j}}] = \frac{1}{|X|} \sum_{\substack{i \in X \\ j \in Y}} a_{ij}.$$

Conditioned on $|X| = x$, the probability that $i \in X$ is x/n, as is the probability that $j \in Y$, and both events are independent. Therefore

$$\mathbb{E}[m_{X,Y} \| X| = x] = \frac{1}{x} \cdot \frac{x^2}{n^2} \sum_{i,j=1}^{n} a_{ij} = \frac{x}{n^2} n(2\mathfrak{c} - 1) = \frac{x}{n}(2\mathfrak{c} - 1),$$

using Lemma 8.2. Since $\mathbb{E}[|X|] = n/2$, we deduce that

$$\mathbb{E}[m_{X,Y}] = \sum_{x=0}^{n} \Pr[|X| = x] \frac{x}{n}(2\mathfrak{c} - 1) = \frac{n/2}{n}(2\mathfrak{c} - 1) = \mathfrak{c} - \frac{1}{2}.$$

The variance requires a similar but longer calculation. The formula for $m_{X,Y}$ implies the following formula for its square:

$$m_{X,Y}^2 = \frac{1}{|X|^2} \left(\sum_{\substack{i \in X \\ j \in Y}} a_{ij}^2 + \sum_{\substack{i \neq k \in X \\ j \in Y}} a_{ij} a_{kj} + \sum_{\substack{i \in X \\ j \neq l \in Y}} a_{ij} a_{il} + \sum_{\substack{i \neq k \in X \\ j \neq l \in Y}} a_{ij} a_{kl} \right).$$

Given $|X| = x$, the probability that $i, k \in X$ for $i \neq k$ is $x(x-1)/n(n-1)$. The same formula applies for the probability that $j, l \in Y$ for $j \neq l$, and the two events are independent. Therefore

$$\begin{aligned}
\mathbb{E}[m_{X,Y}^2 \| X| = x] &= \frac{1}{x^2} \left(\frac{x^2}{n^2} \sum_{i,j=1}^{n} a_{ij}^2 + \frac{x^2(x-1)}{n^2(n-1)} \sum_{i,j=1}^{n} a_{ij} \sum_{k \neq i} a_{kj} + \right. \\
&\quad \left. \frac{x^2(x-1)}{n^2(n-1)} \sum_{i,j=1}^{n} a_{ij} \sum_{l \neq j} a_{il} + \frac{x^2(x-1)^2}{n^2(n-1)^2} \sum_{i,j=1}^{n} a_{ij} \sum_{k \neq i} \sum_{l \neq j} a_{kl} \right) \\
&= \frac{1}{n^2} \sum_{i,j=1}^{n} a_{ij}^2 + 2 \frac{x-1}{n^2(n-1)} \sum_{i,j=1}^{n} a_{ij}(2\mathfrak{c} - 1 - a_{ij}) \\
&\quad + \frac{(x-1)^2}{n^2(n-1)^2} \sum_{i,j=1}^{n} a_{ij}(a_{ij} + (n-2)(2\mathfrak{c} - 1)) \\
&= \frac{(n-x)^2}{n^2(n-1)^2} \sum_{i,j=1}^{n} a_{ij}^2 + \frac{2(x-1)(n-1) + (n-2)(x-1)^2}{n^2(n-1)^2} n(2\mathfrak{c} - 1)^2,
\end{aligned}$$

using Lemma 8.2 and Lemma 8.3. Routine calculation gives $\mathbb{E}[(n - |X|)^2] = n(n+1)/4$ and $\mathbb{E}[(|X| - 1)^2] = n(n-3)/4 + 1$ (these can be verified by checking the cases $n = 0, 1, 2$), and so

$$\begin{aligned}
\mathbb{E}[m_{X,Y}^2] &= \frac{n+1}{4n(n-1)^2} \sum_{i,j=1}^{n} a_{ij}^2 + \frac{4(n-2)(n-1) + (n-2)(n(n-3)+4)}{4n(n-1)^2} (2\mathfrak{c} - 1)^2 \\
&= \frac{n+1}{4n(n-1)^2} \sum_{i,j=1}^{n} a_{ij}^2 + \frac{(n-2)(n+1)}{4(n-1)^2} (2\mathfrak{c} - 1)^2 \\
&\leq \frac{n+1}{4n(n-1)^2} ((n-1) - (n-2)(2\mathfrak{c} - 1)^2) + \frac{(n-2)(n+1)}{4(n-1)^2} (2\mathfrak{c} - 1)^2 \\
&= \frac{n+1}{4n(n-1)} + \frac{(n-2)(n+1)}{4n(n-1)} (2\mathfrak{c} - 1)^2,
\end{aligned}$$

using Lemma 8.5. Subtracting $\mathbb{E}[m_{X,Y}]^2 = (2\mathfrak{c}-1)^2/4$, we deduce

$$\begin{aligned}\mathbb{V}[m_{X,Y}] &\le \frac{n+1}{4n(n-1)} + \frac{(n-2)(n+1)-n(n-1)}{4n(n-1)}(2\mathfrak{c}-1)^2 \\ &= \frac{n+1}{4n(n-1)} - \frac{(2\mathfrak{c}-1)^2}{2n(n-1)} \le \frac{n+1}{4n(n-1)} \le \frac{1}{2n},\end{aligned}$$

using the assumption $n \ge 4$. □

The main result of this section is that for almost all restrictions (X,Y), the resulting functions $g_1^{X,Y}, g_2^{X,Y}, g^{X,Y}$ are nicely behaved, in a way which is summarized by the ensuing definition.

Definition 8.3. A restriction (X,Y) is *typical* if the functions $g_1^{X,Y}, g_2^{X,Y}, g^{X,Y}$ satisfy the following properties:

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

(b) $\mathbb{E}[g_1^{X,Y}]$ and $\mathbb{E}[g_2^{X,Y}]$ are $\epsilon^{1/7}$-close to $\mathfrak{c} - 1/2$.

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

For the definitions of almost Boolean and γ-close, see Section 8.2. The specific powers of ϵ were chosen by optimizing the parameters, taking into account the rest of the proof.

Lemma 8.8. *The probability that a random restriction is typical is at least* $1 - 3\epsilon^{1/7}$.

Proof. We will show that each of the three properties listed in Definition 8.3 fails with probability at most $\epsilon^{1/7}$, and the lemma then follows via a union bound. We will repeatedly use the fact that choosing $(X,Y) \sim \mathcal{R}$ and then choosing $\pi \in T_{X,Y}$ randomly is the same as choosing a random permutation $\pi \in S_n$. The reason is that given X, the sets $T_{X,Y}$ partition S_n and they all have the same size.

Since f is ± 1-valued,

$$\mathbb{E}[(|f_1|-1)^2] \le \mathbb{E}[(f_1 - f)^2] = \epsilon. \tag{8.3}$$

Property (a). For a restriction (X,Y), let

$$\alpha_{X,Y} = \Pr[|g^{X,Y}| \notin 1 \pm \epsilon^{1/7}].$$

Using the observation about choosing $\pi \in T_{X,Y}$ for a random restriction (X,Y) together with Lemma 8.6,

$$\mathop{\mathbb{E}}_{(X,Y)\sim\mathcal{R}}[\alpha_{X,Y}] = \Pr[|f_1| \notin 1 \pm \epsilon^{1/7}] = \Pr[(|f_1|-1)^2 > \epsilon^{2/7}] < \epsilon^{5/7},$$

using (8.3) and Markov's inequality. Another application of Markov's inequality shows that

$$\Pr[\alpha_{X,Y} > \epsilon^{4/7}] < \epsilon^{1/7}.$$

Property (b). This is a straightforward application of Lemma 8.7. Chebyshev's inequality implies that

$$\Pr[|\mathbb{E}[g_1^{X,Y}] - (\mathfrak{c} - 1/2)| > \epsilon^{1/7}] < \frac{\epsilon^{-2/7}}{2n} \leq \frac{\epsilon^{1/7}}{2},$$

since $1/n \leq \epsilon^{3/7}$ by assumption (8.1). Since $g_2^{X,Y} = g_1^{\overline{X},\overline{Y}}$ and $(\overline{X},\overline{Y})$ has the same distribution as (X,Y), we get a similar bound for $\mathbb{E}[g_2^{X,Y}]$.

Property (c). For a restriction (X,Y), let

$$\beta_{X,Y} = \mathbb{E}[(|g^{X,Y}|-1)^2].$$

Again using the observation on choosing $\pi \in T_{X,Y}$ for a random restriction (X,Y),

$$\mathop{\mathbb{E}}_{(X,Y)\sim\mathcal{R}}[\beta_{X,Y}] = \mathbb{E}[(|f_1|-1)^2] \leq \epsilon,$$

using (8.3). Markov's inequality now implies

$$\Pr[\beta_{X,Y} > \epsilon^{6/7}] < \epsilon^{1/7}. \qquad \square$$

8.5 Decomposition under a typical restriction

In this section we analyze the functions $g_1^{X,Y}, g_2^{X,Y}, g^{X,Y}$ for a typical restriction (X,Y) (see Definition 8.3). If (X,Y) is typical then $g^{X,Y}$ is almost Boolean. Now $g^{X,Y}$ can be written as $g^{X,Y} = g_1^{X,Y} + g_2^{X,Y}$, where $g_1^{X,Y}(\pi)$ only depends on $\pi|_X$, and $g_2^{X,Y}(\pi)$ only depends on $\pi|_{\overline{X}}$. How can an independent sum of two functions be almost Boolean? Intuitively, it is obvious that one of $g_1^{X,Y}, g_2^{X,Y}$ must be almost constant C, and the other must be almost close to two values $-C \pm 1$. Since $\mathbb{E}[g_1^{X,Y}] \approx \mathbb{E}[g_2^{X,Y}] \approx \mathfrak{c} - 1/2$, we can actually determine the constants involved, namely $C \approx \mathfrak{c} - 1/2$.

Turning this intuitive argument into a formal proof, we reach the following difficulty: given that (say) $g_1^{X,Y}$ is almost constant and $\mathbb{E}[g_1^{X,Y}] \approx \mathfrak{c} - 1/2$, can we conclude that $g_1^{X,Y}$ is almost close to $\mathfrak{c} - 1/2$? In general, the answer is false: it could be that most of the contribution to $\mathbb{E}[g_1^{X,Y}]$ arises from some rare extreme values. Property (c) in the definition of a typical restriction will be used to show that this cannot happen in our case. To that end, we need the following technical lemma.

Lemma 8.9. *Suppose that a function ϕ on a probability space satisfies the following two properties:*

(a) ϕ is (p,γ)-almost close to 0.

(b) For some $C \in \mathbb{R}$, $\mathbb{E}[(|\phi + C| - 1)^2] \leq \delta$, where $\delta \leq 1$.

Then

$$|E[\phi]| \leq 3\gamma + 3p + 6\sqrt{\frac{\delta}{1-p}}. \tag{8.4}$$

Moreover,

$$||C| - 1| \leq \gamma + \sqrt{\frac{\delta}{1-p}}. \tag{8.5}$$

Proof. Without loss of generality, we assume for the entire proof that $C \geq 0$. We start by proving (8.5). If $|C-1| \leq \gamma$ then (8.5) clearly holds. Otherwise, there are two possibilities: $C < 1-\gamma$ and $C > 1+\gamma$.

Suppose first that $C < 1-\gamma$. Whenever $|\phi| \leq \gamma$ we have $C + \phi \leq C + \gamma$ and $C + \phi \geq C - \gamma \geq -(C+\gamma)$, and so $|C+\phi| \leq C+\gamma$. Therefore $1 - |C+\phi| \geq 1 - C - \gamma > 0$. Since $|\phi| \leq \gamma$ happens with probability at least $1-p$, we conclude that

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

Therefore $|1 - C - \gamma| \leq \sqrt{\delta/(1-p)}$, implying (8.5).

The case $C > 1+\gamma$ is similar. Whenever $|\phi| \leq \gamma$ we have $C + \phi \geq C - \gamma$. Since $C - \gamma > 0$, we conclude that $|C+\phi| \geq C - \gamma$. Therefore $|C+\phi| - 1 \geq C - 1 - \gamma > 0$. Since $|\phi| \leq \gamma$ happens with probability at least $1-p$, we conclude that

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

Therefore $|C - 1 - \gamma| \le \sqrt{\delta/(1-p)}$, implying (8.5).

Concluding, in all cases (8.5) holds. We now turn to prove (8.4). The idea is to consider separately the points in which $|\phi + C| \ge 2$, the points in which $|\phi + C| < 2$ and $|\phi| > \gamma$, and the points in which $|\phi| < \gamma$ (these categories are exhaustive but need not be mutually exclusive). For points of the first type, we use the inequality $x \le x^2$ (valid for $x \ge 1$) to deduce that their contribution to $\mathbb{E}[|\phi + C - 1|]$ is small, using assumption (b). For points of the second type, we use assumption (a) to show that their contribution to $\mathbb{E}[|\phi + C - 1|]$ is small. Finally, the contribution of points of the third type to $\mathbb{E}[|\phi + C - 1|]$ is trivially small. Equation (8.5) then completes the proof.

We start with points of the first type. Assumption (b) implies that

$$\mathbb{E}[||\phi + C| - 1| \cdot [\![\phi + C \ge 2]\!]] \le \mathbb{E}[(|\phi + C| - 1)^2 \cdot [\![\phi + C \ge 2]\!]] \le \delta.$$

(Recall that $[\![\phi + C \ge 2]\!]$ is equal to 1 if $\phi + C \ge 2$ and 0 otherwise.) Therefore

$$\mathbb{E}[|\phi + C - 1| \cdot [\![\phi + C \ge 2]\!]] = \mathbb{E}[||\phi + C| - 1| \cdot [\![\phi + C \ge 2]\!]] \le \delta.$$

When $\phi + C < 0$, it is no longer the case that $|\phi + C - 1| = ||\phi + C| - 1|$, and the first term is actually larger than the second. However, if $\phi + C \le -2$ then

$$|\phi + C - 1| = 1 - (\phi + C) = -3 + (4 - (\phi + C)) \le -3 - 3(\phi + C) = 3(|\phi + C| - 1) = 3||\phi + C| - 1|.$$

Therefore as before,

$$\mathbb{E}[|\phi + C - 1| \cdot [\![\phi + C \ge 2]\!]] \le 3\,\mathbb{E}[||\phi + C| - 1| \cdot [\![\phi + C \ge 2]\!]] \le 3\delta.$$

Putting both cases together, we get

$$\mathbb{E}[|\phi + C - 1| \cdot [\![|\phi + C| \ge 2]\!]] \le 4\delta.$$

We continue with points of the second type. When $|\phi + C| \le 2$, the triangle inequality shows that $|\phi + C - 1| \le 3$. Assumption (a) implies that

$$\mathbb{E}[|\phi + C - 1| \cdot [\![|\phi + C| \le 2 \text{ and } |\phi| > \gamma]\!]] \le 3p.$$

Finally, we handle points of the third type. If $|\phi| \le \gamma$ then (8.5) shows that $|\phi + C - 1| \le 2\gamma + \sqrt{\delta/(1-p)}$. Therefore

$$\mathbb{E}[|\phi + C - 1| \cdot [\![|\phi| < \gamma]\!]] \le 2\gamma + \sqrt{\frac{\delta}{1-p}}.$$

Since all three types of points are exhaustive, we conclude that

$$\mathbb{E}[|\phi + C - 1|] \le 4\delta + 3p + 2\gamma + \sqrt{\frac{\delta}{1-p}}.$$

Using the triangle inequality (twice) and (8.5),

$$|\mathbb{E}[\phi]| \le \mathbb{E}[|\phi|] \le \mathbb{E}[|\phi + C - 1|] + |1 - C| \le 4\delta + 3p + 3\gamma + 2\sqrt{\frac{\delta}{1-p}}.$$

Since $\delta \le 1$, $\delta \le \sqrt{\delta} \le \sqrt{\delta/(1-p)}$, which completes the proof. □

During the rest of this section, we will always have some typical restriction (X, Y) in mind. Therefore we use the shorthand notations g_1, g_2, g for $g_1^{X,Y}, g_2^{X,Y}, g^{X,Y}$. Other useful pieces of notation are $\pi_1 = \pi|_X$ and $\pi_2 = \pi|_{\overline{X}}$. Since g_1 depends only on π_1, we can think of g_1 as a function whose domain is the set B_1 of bijection from X to Y. Similarly, g_2 depends only on π_2, and we can think of it as a function whose domain is the set B_2 of bijections from $\overline{X}$ to $\overline{Y}$. Under these conventions, we have the identity

$$g(\pi) = g_1(\pi_1) + g_2(\pi_2), \quad \text{where } \pi \in T_{X,Y}. \tag{8.6}$$

Given $\pi_1 \in B_1$ and $\pi_2 \in B_2$, we will denote by $\pi_1; \pi_2$ the permutation which restricts to π_1 on X and to π_2 on $\overline{X}$.

Our end goal in this section is showing that for a typical restriction (X, Y), one of g_1, g_2 is almost constant, and the other is (up to a shift) almost Boolean. The next lemma shows that this is the case locally: if we sample two permutations $\alpha, \beta \in T_{X,Y}$, then the values of g_1 and g_2 on these two permutations behave as if one of these functions were almost constant, and the other almost Boolean (up to a shift).

Lemma 8.10. *Suppose (X, Y) is a typical restriction. Choose α, β uniformly and independently from $T_{X,Y}$. Let*

$$\delta_1 = g_1(\alpha_1) - g_1(\beta_1), \quad \delta_2 = g_2(\alpha_2) - g_2(\beta_2).$$

With probability at least $1 - 8\epsilon^{2/7}$, one of the following three cases holds:

(a) $|\delta_1| \le 2\epsilon^{1/7}$ *and* $|\delta_2| \le 2\epsilon^{1/7}$.

(b) $|\delta_1| \le 2\epsilon^{1/7}$ *and* $|\delta_2| \in 2 \pm 2\epsilon^{1/7}$.

(c) $|\delta_1| \in 2 \pm 2\epsilon^{1/7}$ *and* $|\delta_2| \le 2\epsilon^{1/7}$.

(Since ϵ is small enough, these cases are mutually exclusive.)

Proof. Property (c) in Definition 8.3 states that

$$\mathop{\mathbb{E}}_{\pi \in T_{X,Y}}[(|g(\pi)| - 1)^2] \le \epsilon^{6/7}.$$

Markov's inequality implies that

$$\Pr_{\pi \in T_{X,Y}}[||g(\pi)| - 1| > \epsilon^{1/7}] < \epsilon^{4/7}.$$

We can choose the permutation π by choosing $\pi_1 \in B_1$ and $\pi_2 \in B_2$ independently. Markov's inequality shows that

$$\Pr_{\pi_1 \in B_1}\left[\Pr_{\pi_2 \in B_2}[||g(\pi_1;\pi_2)| - 1| > \epsilon^{1/7}] > \epsilon^{2/7}\right] < \epsilon^{2/7}.$$

We will choose α and β by choosing $\alpha_1, \beta_1 \in B_1$ and $\alpha_2, \beta_2 \in B_2$ independently. With probability at least $1 - 2\epsilon^{2/7}$ over the choice of α_1, β_1, we have

$$\Pr_{\pi_2 \in B_2}[||g(\alpha_1;\pi_2)| - 1| > \epsilon^{1/7}] \le \epsilon^{2/7},$$

$$\Pr_{\pi_2 \in B_2}[||g(\beta_1;\pi_2)| - 1| > \epsilon^{1/7}] \le \epsilon^{2/7}.$$

Hence with probability at least $1 - 4\epsilon^{2/7}$ over the choice of $\alpha_1, \beta_1, \alpha_2, \beta_2$, all the following quantities are $\epsilon^{1/7}$-close in magnitude to 1:

$$g(\alpha_1;\alpha_2), g(\alpha_1;\beta_2), g(\beta_1;\alpha_1), g(\beta_1;\beta_2).$$

We now consider several cases. Without loss of generality, suppose that $g(\alpha_1;\alpha_2)$ is $\epsilon^{1/7}$-close to 1. If $g(\alpha_1;\beta_2)$ and $g(\beta_1;\alpha_2)$ are also $\epsilon^{1/7}$-close to 1, then

$$|\delta_1| = |g_1(\alpha_1) - g_1(\beta_1)| = |g(\alpha_1;\alpha_2) - g(\beta_1;\alpha_2)| \le 2\epsilon^{1/7},$$

$$|\delta_2| = |g_2(\alpha_2) - g_2(\beta_2)| = |g(\alpha_1;\alpha_2) - g(\alpha_1;\beta_2)| \le 2\epsilon^{1/7}.$$

This is case (a). If $g(\alpha_1;\beta_2)$ is $\epsilon^{1/7}$-close to -1 and $g(\beta_1;\alpha_2)$ is $\epsilon^{1/7}$-close to 1 then

$$|\delta_1| = |g_1(\alpha_1) - g_1(\beta_1)| = |g(\alpha_1;\alpha_2) - g(\beta_1;\alpha_2)| \le 2\epsilon^{1/7},$$
$$|\delta_2| = |g_2(\alpha_2) - g_2(\beta_2)| = |g(\alpha_1;\alpha_2) - g(\alpha_1;\beta_2)| \in 2 \pm 2\epsilon^{1/7}.$$

This is case (b). Similarly, if $g(\alpha_1;\beta_2)$ is $\epsilon^{1/7}$-close to 1 and $g(\beta_1;\alpha_2)$ is $\epsilon^{1/7}$-close to -1 then we are in case (c). It remains to rule out the case that both $g(\alpha_1;\beta_2)$ and $g(\beta_1;\alpha_2)$ are $\epsilon^{1/7}$-close to -1. In this case

$$g(\beta_1;\beta_2) = g(\beta_1;\alpha_2) + g(\alpha_1;\beta_2) - g(\alpha_1;\alpha_2) \in -3 \pm 3\epsilon^{1/7},$$

which contradicts $|g(\beta_1;\beta_2)| \in 1 \pm \epsilon^{1/7}$ since $\epsilon^{1/7}$ is small enough. □

We are now ready to prove the main lemma of this section, which states that if (X,Y) is typical then either g_1 is almost close to $\mathfrak{c} - 1/2$ and $g_2 + (\mathfrak{c} - 1/2)$ is almost Boolean, or the same is true with the roles of g_1 and g_2 reversed. The appearance of the constant $\mathfrak{c} - 1/2$ stems from the fact that $\mathbb{E}[g_1] \approx \mathbb{E}[g_2] \approx \mathfrak{c} - 1/2$. As mentioned in the beginning of the section, Lemma 8.9 is crucial to translate this fact into a statement about values which g_1, g_2 are close to.

Lemma 8.11. *Suppose (X,Y) is a typical restriction. Either g_1 is $(3\epsilon^{1/7}, 19\epsilon^{1/7})$-close to $\mathfrak{c}-1/2$ and $g_2 + (\mathfrak{c} - 1/2)$ is $(4\epsilon^{1/7}, 24\epsilon^{1/7})$-almost Boolean, or the same is true with the roles of g_1 and g_2 reversed.*

Here and elsewhere in this chapter, the constants $3, 19, 4, 24$ are not optimal, and are stated explicitly only for later convenience.

Proof. The first step is to determine which of g_1, g_2 is almost constant and which is almost Boolean (up to a shift). To that end, define the following two probabilities:

$$p_1 = \Pr_{\alpha_1,\beta_1 \in B_1}[|g_1(\alpha_1) - g_1(\beta_1)| > 2\epsilon^{1/7}],$$
$$p_2 = \Pr_{\alpha_2,\beta_2 \in B_2}[|g_2(\alpha_2) - g_2(\beta_2)| > 2\epsilon^{1/7}].$$

Lemma 8.10 shows that $p_1 p_2 \le 8\epsilon^{2/7}$, since choosing $\alpha, \beta \in T_{X,Y}$ is equivalent to choosing $\alpha_1, \beta_1 \in B_1$ and $\alpha_2, \beta_2 \in B_2$ independently. Therefore either $p_1 \le 3\epsilon^{1/7}$ or $p_2 \le 3\epsilon^{1/7}$ (or both).

Without loss of generality, suppose for the rest of the proof that $p_1 \le 3\epsilon^{1/7}$. This will imply that g_1 is almost constant and that g_2 is almost Boolean (up to a shift).

Analyzing $\mathbf{g_1}$. Since $p_1 \le 3\epsilon^{1/7}$, a simple averaging argument shows that for some $\sigma_1 \in B_1$,

$$\Pr_{\alpha_1 \in B_1}[|g_1(\alpha_1) - g_1(\sigma_1)| > 2\epsilon^{1/7}] \le 3\epsilon^{1/7}.$$

We conclude that g_1 is $(3\epsilon^{1/7}, 2\epsilon^{1/7})$-close to $C_1 = g_1(\sigma_1)$. We would like to show that $C_1 \approx \mathfrak{c} - 1/2$, and to that effect we appeal to Lemma 8.9. We will apply this lemma to the function $\phi = g_1 - C_1$, which is $(3\epsilon^{1/7}, 2\epsilon^{1/7})$-close to 0. This satisfies the first condition of the lemma, with $p = 3\epsilon^{1/7}$ and $\gamma = 2\epsilon^{1/7}$. For the second condition, we rewrite property (c) in Definition 8.3 as

$$\mathop{\mathbb{E}}_{\pi_2 \in B_2} \mathop{\mathbb{E}}_{\pi_1 \in B_1}[(|g_1(\pi_1) + g_2(\pi_2)| - 1)^2] \le \epsilon^{6/7}.$$

Thus for some $\sigma_2 \in B_2$,

$$\mathop{\mathbb{E}}_{\pi_1 \in B_1}[(|g_1(\pi_1) + g_2(\sigma_2)| - 1)^2] \le \epsilon^{6/7}.$$

This shows that the second condition of Lemma 8.9 is satisfied, with $C = C_1 + g_2(\sigma_2)$ and $\delta = \epsilon^{6/7}$. Note that $\delta \le 1$ since ϵ is small enough. The lemma shows that

$$|\mathbb{E}[g_1] - C_1| = |\mathbb{E}[g_1 - C_1]| \le 3 \cdot 2\epsilon^{1/7} + 3 \cdot 3\epsilon^{1/7} + 6\sqrt{\frac{\epsilon^{6/7}}{1 - 3\epsilon^{1/7}}} \le 16\epsilon^{1/7},$$

since ϵ is small enough. On the other hand, property (b) of Definition 8.3 shows that $\mathbb{E}[g_1]$ is $\epsilon^{1/7}$-close to $\mathfrak{c} - 1/2$. We conclude that C_1 is $17\epsilon^{1/7}$-close to $\mathfrak{c} - 1/2$, and so g_1 is $(3\epsilon^{1/7}, 19\epsilon^{1/7})$-close to $\mathfrak{c} - 1/2$.

Analyzing $\mathbf{g_2}$. The argument for g_2 is similar, but the details are more complicated. Lemma 8.10 shows for random $\alpha_2, \beta_2 \in B_2$, with probability at least $1 - 8\epsilon^{2/7}$, either $|g_2(\alpha_2) - g_2(\beta_2)| \le 2\epsilon^{1/7}$ or $|g_2(\alpha_2) - g_2(\beta_2)| \in 2 \pm 2\epsilon^{1/7}$. A simple averaging argument shows that for some $\sigma_2 \in B_2$,

$$\Pr_{\alpha_2 \in B_2}[|g_2(\alpha_2) - g_2(\sigma_2)| \le 2\epsilon^{1/7} \text{ or } |g_2(\alpha_2) - g_2(\sigma_2)| \in 2 \pm 2\epsilon^{1/7}] \ge 1 - 8\epsilon^{2/7}.$$

In other words, g_2 is concentrated around the three values $C_2 - 2, C_2, C_2 + 2$, where $C_2 = g_2(\sigma_2)$. Since we want to show that (up to a shift) g_2 is almost Boolean, we need to eliminate one of these values, either $C_2 - 2$ or $C_2 + 2$. To that end, define

$$q_- = \Pr_{\alpha_2 \in B_2}[g_2(\alpha_2) \in C_2 - 2 \pm 2\epsilon^{1/7}],$$
$$q_+ = \Pr_{\beta_2 \in B_2}[g_2(\beta_2) \in C_2 + 2 \pm 2\epsilon^{1/7}].$$

If both events whose probabilities are measure by q_-, q_+ happen then $|g_2(\alpha_2) - g_2(\beta_2)| \in 4 \pm 4\epsilon^{1/7}$. Since ϵ is small enough, Lemma 8.10 implies that $q_- q_+ \le 8\epsilon^{2/7}$. Therefore either $q_- \le 3\epsilon^{1/7}$ or $q_+ \le 3\epsilon^{1/7}$ (or both). Without loss of generality, we assume that $q_- \le 3\epsilon^{1/7}$, and so g_2 is concentrated around the two values $C_2, C_2 + 2$. Putting $D_2 = C_2 + 1$, we conclude that with probability at least $1 - 8\epsilon^{2/7} - 3\epsilon^{1/7} \ge 1 - 4\epsilon^{1/7}$ (since ϵ is small enough), g_2 is $2\epsilon^{1/7}$-close to either $D_2 - 1$ or $D_2 + 1$. In other words, $g_2 - D_2$ is $(4\epsilon^{1/7}, 2\epsilon^{1/7})$-almost Boolean.

It remains to show that D_2 is close to $-(\mathfrak{c} - 1/2)$. Since g_1 is $(3\epsilon^{1/7}, 19\epsilon^{1/7})$-close to $\mathfrak{c} - 1/2$, we conclude that $g - (\mathfrak{c} - 1/2) - D_2 = (g_1 - (\mathfrak{c} - 1/2)) + (g_2 - D_2)$ is $(7\epsilon^{1/7}, 21\epsilon^{1/7})$-almost Boolean. On the other hand, property (a) in Definition 8.3 states that g is $(\epsilon^{4/7}, \epsilon^{1/7})$-almost Boolean. Therefore with probability at least $1 - 7\epsilon^{1/7} - \epsilon^{4/7} \ge 1 - 8\epsilon^{1/7}$ (since ϵ is small enough) over $\pi \in T_{X,Y}$, $g(\pi)$ is simultaneously $21\epsilon^{1/7}$ close to $(D_2 + (\mathfrak{c} - 1/2)) \pm 1$ and $\epsilon^{1/7}$-close to ± 1. Call a π satisfying this property *reasonable*. So $\pi \in T_{X,Y}$ is reasonable with probability at least $1 - 8\epsilon^{1/7}$.

We would like to conclude that $D_2 + (\mathfrak{c} - 1/2)$ is $22\epsilon^{1/7}$-close to 0. However, this need not be the case: it could (a priori) be that whenever π is reasonable, $g(\pi)$ is always positive or always negative. We rule out these cases using Lemma 8.9. Suppose that whenever π is reasonable, $g(\pi)$ is always $\epsilon^{1/7}$-close to 1. Therefore $g - 1$ is $(8\epsilon^{1/7}, \epsilon^{1/7})$-close to 0. Thus the first hypothesis of Lemma 8.9 is satisfied for $\phi = g - 1$ with $p = 8\epsilon^{1/7}$ and $\gamma = \epsilon^{1/7}$. Property (c) in Definition 8.3 shows that the second hypothesis of the lemma is satisfied with $C = 1$ and $\delta = \epsilon^{6/7}$ (note $\delta \le 1$ since ϵ is small enough). Therefore the lemma shows that

$$|\mathbb{E}[g] - 1] = |\mathbb{E}[g - 1]| \le 3\epsilon^{1/7} + 3 \cdot 8\epsilon^{1/7} + 6\sqrt{\frac{\epsilon^{6/7}}{1 - 8\epsilon^{1/7}}} \le 28\epsilon^{1/7},$$

since ϵ is small enough. On the other hand, Property (b) in the definition shows that $\mathbb{E}[g]$ is $2\epsilon^{1/7}$-close to $2\mathfrak{c} - 1$. We conclude that $2\mathfrak{c} - 1$ is $30\epsilon^{1/7}$-close to 1, or in other words, $\eta \le |\mathfrak{c} - 1| \le 15\epsilon^{1/7}$. This contradicts the fact that ϵ is small enough compared to η^7. Therefore there must exist some reasonable π_- such that $g(\pi)$ is $\epsilon^{1/7}$-close to -1.

Using very similar reasoning, we conclude that there must exist some reasonable π_+ such that $g(\pi)$ is $\epsilon^{1/7}$-close to 1 (this time the contradiction is $\eta \le |\mathfrak{c}| \le 15\epsilon^{1/7}$). From π_- we get that one of $(D_2 + (\mathfrak{c} - 1/2)) \pm 1$ is $22\epsilon^{1/7}$-close to -1, and from π_+ we get that one of $(D_2 + (\mathfrak{c} - 1/2)) \pm 1$ is $22\epsilon^{1/7}$-close to 1. Since ϵ is small enough, we conclude that $|D_2 + (\mathfrak{c} - 1/2)| \le 22\epsilon^{1/7}$. Since $g_2 - D_2$ is $(4\epsilon^{1/7}, 2\epsilon^{1/7})$-almost Boolean, we conclude that $g_2 + (\mathfrak{c} - 1/2)$ is $(4\epsilon^{1/7}, 24\epsilon^{1/7})$-almost Boolean. □

8.6 Random partitions

The main result of the preceding section states that for most random partitions (X, Y), either $g_1 - (\mathfrak{c} - 1/2)$ is almost zero and $g_2 + (\mathfrak{c} - 1/2)$ is almost Boolean, or the same is true with the roles of g_1, g_2 reversed. In other words, if we choose a random partition (X, Y) and a random $\pi \in T_{X,Y}$, then for P_1, P_2 given by

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

either $P_1 - (\mathfrak{c} - 1/2)$ is close to zero and $P_2 + (\mathfrak{c} - 1/2)$ is close to ± 1, or the reverse is true. In this section we reverse the order of the random choice: we first choose $\pi \in S_n$ and then (X, Y) compatible with π. For most permutations π, it will be the case the for most choice of (X, Y), P_1, P_2 satisfy the given property. For such a permutation π, we will be able to deduce strong information on the values $a_{i\pi(i)}$.

We start by formally defining the property of permutations outlined above, and proving that almost all permutations satisfy this property.

Definition 8.4. A *random partition*[2] X is a uniformly random subset of $[n]$. We denote the corresponding probability distribution by $\mathcal{P}$. (Note that if $(X, Y) \sim \mathcal{R}$ then $X \sim \mathcal{P}$).

For a permutation $\pi \in S_n$ and a partition X, define

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

[2]The reader might object that calling X a partition is a misnomer. However, every X defines a partition $X, \overline{X}$.

A partition X is *good for* π if either P_1 is $25\epsilon^{1/7}$-close to $\mathfrak{c}-1/2$ and P_2 is $25\epsilon^{1/7}$-close to $-(\mathfrak{c}-1/2)\pm 1$, or the same is true with the roles of P_1, P_2 reversed. Otherwise, X is *bad for* π.

A permutation $\pi \in S_n$ is *good* if a random partition X is bad for π with probability at most $1/5$. Otherwise, π is *bad*. ○

The next lemma shows that almost all permutations are good. This is an easy consequence of Lemma 8.11.

Lemma 8.12. *The probability that a random permutation $\pi \in S_n$ is good is at least $1-50\epsilon^{1/7}$.*

Proof. Let (X,Y) be a random restriction. Lemma 8.8 shows that (X,Y) is typical with probability at least $1-3\epsilon^{1/7}$. If (X,Y) is typical, then Lemma 8.11 shows that either g_1 is $(3\epsilon^{1/7}, 19\epsilon^{1/7})$-close to $\mathfrak{c}-1/2$ and $g_2+(\mathfrak{c}-1/2)$ is $(4\epsilon^{1/7}, 24\epsilon^{1/7})$-almost Boolean, or the same is true with the roles of g_1, g_2 reversed. In particular, with probability at least $1-7\epsilon^{1/7}$ over the choice of $\pi \in T_{X,Y}$, X is good for π.

Summarizing, if we choose a random restriction (X,Y) and a random permutation $\pi \in T_{X,Y}$, then with probability at least $1-10\epsilon^{1/7}$, X is good for π. Reversing the order of the random choices,

$$\mathop{\mathbb{E}}_{\pi\in S_n} \Pr_{X\sim\mathcal{P}}[X \text{ is bad for } \pi] \le 10\epsilon^{1/7}.$$

Markov's inequality shows that

$$\Pr_{\pi\in S_n}\left[\Pr_{X\sim\mathcal{P}}[X \text{ is bad for } \pi] > 1/5\right] < 50\epsilon^{1/7}. \qquad \square$$

If a permutation is good, then if we randomly partition the values $a_{i\pi(i)}$ into two parts, with probability at least $4/5$ one of them will sum roughly to $\mathfrak{c}-1/2$, and the other will sum roughly to $-(\mathfrak{c}-1/2)\pm 1$. Intuitively, that can only happen if most of the values $a_{i\pi(i)}$ are small, and one of them is large. The large value determines which part sums to $\mathfrak{c}-1/2$ and which to $-(\mathfrak{c}-1/2)\pm 1$. To make this intuition precise, we use the Berry–Esseen theorem.

Lemma 8.13. *Suppose $\pi \in S_n$ is a good permutation. For some $m \in [n]$, $|a_{m\pi(m)}|$ is $50\epsilon^{1/7}$-close to $2\mathfrak{c}$ or to $2(1-\mathfrak{c})$, and for $i \neq m$, $|a_{i\pi(i)}| \le 50\epsilon^{1/7}$.*

Proof. The proof is inspired by one of the proofs in [64]. Considering the effect of an element $a_{i\pi(i)}$ "switching sides", we can group the elements $a_{i\pi(i)}$ into two types: "small" elements (close to 0) and "large" elements (close to $2\mathfrak{c}$ or to $2(1-\mathfrak{c})$, depending on $\sum_i a_{i\pi(i)}$). For similar reasons, there can be at most one large element. The difficult part is showing that not all elements can be small. The intuitive reason is that if all elements are small then the distribution of P_1 is approximately normal, whereas P_1 should be bimodal.

For the rest of the proof, we use the shorthands

$$\zeta = \mathfrak{c} - \frac{1}{2}, \quad s_i = a_{i\pi(i)} \quad \text{and} \quad S = \sum_{i=1}^{n} s_i.$$

Since π is good, S must be $50\epsilon^{1/7}$-close to ± 1. Let $K \in \{1, -1\}$ be the value which S is $50\epsilon^{1/7}$-close to. We conclude that whenever X is good for π, the values P_1, P_2 are $25\epsilon^{1/7}$-close to $\zeta, K - \zeta$.

For a partition X, let

$$T(X) = \sum_{i \in X} s_i = \frac{S}{2} + \sum_{i=1}^{n} W_i(x), \text{ where } W_i = \begin{cases} s_i/2 & \text{if } i \in X, \\ -s_i/2 & \text{if } i \notin X. \end{cases}$$

We start by showing that each element s_i is either small or large (we define these properties formally below). Consider an arbitrary element s_i. For a random partition Y, the probability that both $Y \setminus \{i\}$ and $Y \cup \{i\}$ are good for π is at least $1 - 2/5 = 3/5$. Hence for some partition Y, both $Y \setminus \{i\}$ and $Y \cup \{i\}$ are good for π. Therefore $T(Y \setminus \{i\})$ and $T(Y \cup \{i\})$ are both $25\epsilon^{1/7}$-close to one of $\zeta, K - \zeta$. Since $|T(Y \setminus \{i\}) - T(Y \cup \{i\})| = |s_i|$, this shows that either $|s_i| \le 50\epsilon^{1/7}$ (we say that s_i is *small*) or $|s_i|$ is $50\epsilon^{1/7}$-close to $|\zeta - (K - \zeta)| = |2\zeta - K|$ (we say that s_i is *large*). When $K = 1$, a large element is $50\epsilon^{1/7}$-close to $2(1 - \mathfrak{c})$, and when $K = -1$, a large element is $50\epsilon^{1/7}$-close to $2\mathfrak{c}$. In both cases,

$$|2\zeta - K| \ge 2\eta. \tag{8.7}$$

Having grouped the elements s_i into small and large elements, we proceed to show that there is at most one large element. Intuitively, if there are two large elements, then by considering all possible ways of their switching sides, we eventually reach a value of T which is close to neither ζ nor $K - \zeta$.

Suppose s_i, s_j are both large. For a random partition Y, the probability that all of $Y \setminus \{i,j\}, Y \setminus \{i\} \cup \{j\}, Y \setminus \{j\} \cup \{i\}, Y \cup \{i,j\}$ are good for π is at least $1 - 4/5 = 1/5$. Hence for some partition Y, all these partitions are good for π. To simplify notation, suppose that $i, j \notin Y$. All of the following values must be $25\epsilon^{1/7}$-close to either ζ or $K - \zeta$:

$$T(Y),\ T(Y) + s_i,\ T(Y) + s_j,\ T(Y) + s_i + s_j.$$

Furthermore, $|s_i|$ and $|s_j|$ are $50\epsilon^{1/7}$-close to $|2\zeta - K|$. Intuitively, s_i and s_j must have the same sign, and therefore $T(Y), T(Y) + s_i, T(Y) + s_i + s_j$ cannot all be $25\epsilon^{1/7}$-close to the two values ζ and $K - \zeta$.

Without loss of generality, suppose that $T(Y)$ is $25\epsilon^{1/7}$-close to ζ (the argument for the other case is similar). If $T(Y) + s_i$ were $25\epsilon^{1/7}$-close to ζ then $|s_i| \le 50\epsilon^{1/7}$, which contradicts the assumption that $|s_i|$ is $50\epsilon^{1/7}$-close to $|2\zeta - k| \ge 2\eta$ (since ϵ is small enough compared to η^7). Therefore $T(Y) + s_i$ is $25\epsilon^{1/7}$-close to $K - \zeta$, showing that s_i is $50\epsilon^{1/7}$-close to $K - 2\zeta$. A similar arguments shows that s_j is $50\epsilon^{1/7}$-close to $K - 2\zeta$. Therefore $T(Y) + s_i + s_j$ is $125\epsilon^{1/7}$-close to $\zeta + 2(K - 2\zeta) = 2K - 3\zeta$. However,

$$|(2K - 3\zeta) - \zeta| = |2K - 4\zeta| \ge 4\eta,$$

$$|(2K - 3\zeta) - (K - \zeta)| = |K - 2\zeta| \ge 2\eta.$$

Since ϵ is small enough compared to η^7, we deduce that $T(Y) + s_i + s_j$ is $25\epsilon^{1/7}$-far from both ζ and $K - \zeta$, reaching a contradiction. We conclude that there is at most one large element.

It remains to rule out the case that all elements are small. For the rest of the proof, suppose that all elements s_i are small. Note that a partition X is good for π if and only if $\overline{X}$ is good for π, and both partitions $X, \overline{X}$ have the same probability under $\mathcal{P}$. Therefore for a random partition X, with probability at least $2/5$, $T(X)$ is $25\epsilon^{1/7}$-close to ζ, and with probability at least $2/5$, $T(X)$ is $25\epsilon^{1/7}$-close to $K - \zeta$.

Using the Berry–Esseen theorem, we will show that for a random partition X, $T(X)$ is close to a normal distribution. This contradicts the observation that $2/5$ of the time, $T(X)$ is close to ζ, and $2/5$ of the time, it is close to $K - \zeta$.

Let $X \sim \mathcal{P}$ be a random partition, and let $T = T(X)$, $W_i = W_i(X)$ be random variables. (The functions $W_i(X)$, attaining the values $\pm s_i/2$, were defined above.) Applying the Berry–Esseen theorem (see Section 2.2), X is $C\psi$*-close in distribution* (see below) to a normally distributed random variable N with mean $S/2$ and variance σ^2 matching those of T, where

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

Here $C\psi$-close in distribution means that for every interval I, $|\Pr[X \in I] - \Pr[N \in I]| \le \psi$, and C is the constant in the statement of Berry–Esseen. Since each s_i is small, we have

$$\psi \le \frac{\sum_{i=1}^n |s_i|^2 \cdot 50\epsilon^{1/7}}{\left(\sum_{i=1}^n s_i^2\right)^{3/2}} = \frac{50\epsilon^{1/7}}{\sqrt{\sum_{i=1}^n s_i^2}} = \frac{25\epsilon^{1/7}}{\sigma}.$$

In order to get a lower bound on $\sigma^2 = \mathbb{V}[T]$, we use the fact that with probability $2/5$, T is $25\epsilon^{1/7}$-close to ζ, and with probability $2/5$, it is $25\epsilon^{1/7}$-close to $K - \zeta$. Recall that $\mathbb{E}[T] = S/2$ is $25\epsilon^{1/7}$-close to $K/2$. Also,

$$|\zeta - K/2| = |(K - \zeta) - K/2| \ge \eta.$$

Since ϵ is small enough compared to η^7, this shows that

$$\sigma^2 = \mathbb{E}[(T - \mathbb{E}\, T)^2] \ge \frac{4}{5}(\eta - 50\epsilon^{1/7})^2.$$

Substituting this into the expression we got for ψ,

$$\psi \le \frac{25\epsilon^{1/7}}{\sigma} \le \frac{\sqrt{5}}{2}\frac{25\epsilon^{1/7}}{\eta - 50\epsilon^{1/7}}.$$

Since ϵ is small enough compared to η^7, we get that $C\psi \le 1/5$. In other words, the normal distribution approximates T reasonably well.

The distribution of T has two peaks, around ζ and around $K-\zeta$, while a normal distribution has a single peak. We want to use this fact to rule out the possibility that T is close to a normal distribution. To that end, we use the following property of the normal distribution: its density is bitonic (increasing and then decreasing).

Consider the following three intervals: I_1 is the closed interval of radius $25\epsilon^{1/7}$ around ζ, I_2 is the closed interval of radius $25\epsilon^{1/7}$ around $K - \zeta$, and I_3 is the open interval of radius $|2\zeta - K|/2 - 25\epsilon^{1/7}$ around $K/2$ (here we are using the fact that $|2\zeta - K| \ge 2\eta$ and ϵ is small enough compared to η^7). The interval I_3 lies just between I_1 and I_2. The properties of T imply that

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

Since $C\psi \le 1/5$ and N is $C\psi$-close to T in distribution,

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

Since the density of the normal distribution is bitonic and I_3 lies between I_1 and I_2,

$$\frac{\Pr[N \in I_3]}{|I_3|} \ge \min\left(\frac{\Pr[N \in I_1]}{|I_1|}, \frac{\Pr[N \in I_2]}{|I_2|}\right), \tag{8.8}$$

where $|I|$ is the length of an interval I. However, on the one hand

$$\frac{\Pr[N \in I_3]}{|I_3|} \le \frac{2/5}{|2\mathfrak{c} - 1 - K|/2 - 25\epsilon^{1/7}} \le \frac{2/5}{\eta - 25\epsilon^{1/7}},$$

and on the other hand

$$\frac{\Pr[N \in I_1]}{|I_1|}, \frac{\Pr[N \in I_2]}{|I_2|} \ge \frac{1/5}{25\epsilon^{1/7}}.$$

Plugging these estimates into (8.8), we get

$$\frac{2/5}{\eta - 25\epsilon^{1/7}} \ge \frac{1/5}{25\epsilon^{1/7}},$$

or equivalently

$$\eta - 25\epsilon^{1/7} \le 50\epsilon^{1/7},$$

which is impossible if ϵ is small enough compared to η^7.

We conclude that not all elements can be small. Therefore there is exactly one large element, and all other elements are small. □

Lemma 8.13 is similar in spirit to the Friedgut–Kalai–Naor theorem, but the parameters are somewhat different. To see the connection, compare the formula for T in the proof of the lemma

$$T = \frac{S}{2} + \sum_{i=1}^{n} (-1)^{[\![i \in X]\!]} \frac{s_i}{2}$$

to the Fourier expansion of a function supported on the first two levels: if we put $x_i = [\![i \in X]\!]$, then $(-1)^{[\![i \in X]\!]}$ becomes a Fourier character.

Combining Lemma 8.12 with Lemma 8.13, we get the following corollary, which summarizes all our work in the first part of the proof.

Corollary 8.14. *With probability at least $1 - 50\epsilon^{1/7}$, a random permutation $\pi \in S_n$ satisfies the following property. For some $m \in [n]$, $|a_{m\pi(m)}|$ is $50\epsilon^{1/7}$-close to $2\mathfrak{c}$ or to $2(1-\mathfrak{c})$, and for $i \ne m$, $|a_{i\pi(i)}| \le 50\epsilon^{1/7}$.*

8.7 Strong lines

Corollary 8.14 shows that almost all generalized diagonals in the matrix A (see definition below) contain exactly one large element and $n-1$ small elements. Intuitively, that can only happen if the matrix A contains a row or a column which consists almost entirely of large elements, and the rest of the matrix contains small elements. Our proof of this fact will be inductive, and to that end we define a generalized property of this form.

Definition 8.5. An element a_{ij} is *large* if it is $50\epsilon^{1/7}$-close in magnitude to $2\mathfrak{c}$ or to $2(1-\mathfrak{c})$. Otherwise, it is *small*.

For a restriction (X,Y) (that is, $X,Y \subseteq [n]$ and $|X| = |Y|$), we denote by $A[X,Y]$ the submatrix consisting of the entries a_{ij} for $i \in X$ and $j \in Y$. A *generalized diagonal* in $A[X,Y]$ is a set of the form $D_\pi = \{a_{i\pi(i)} : i \in X\}$, where π is a bijection from X to Y. A *random generalized diagonal* is obtained by choosing π randomly from the set of all bijections from X to Y.

A generalized diagonal is *good* if it contains exactly one large entry.

A restriction (X,Y) is *q-good* if with probability at least $1-q$, a random generalized diagonal in $A[X,Y]$ is good. ○

In this language, Corollary 8.14 states that $([n],[n])$ is $50\epsilon^{1/7}$-good. The corollary distinguishes two kinds of elements: those that are close to $2\mathfrak{c}$ or to $2(1-\mathfrak{c})$, and those that are close to 0. However, since not every permutation satisfies the conclusion of the corollary, the matrix A could also contain other elements. For our work in this section, it is enough to distinguish large elements from all other elements, which for convenience we call small.

Having stated our premises in abstract form, we proceed to state our goal in an abstract form.

Definition 8.6. For a restriction (X,Y), a row $i \in X$ is *p-strong for* (X,Y) if at least $(1-p)|Y|$ of the entries $\{a_{ij} : j \in Y\}$ are large. Similarly, a column $j \in Y$ 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 a restriction (X,Y) *has a p-strong row (column)* if some row $i \in X$ (some column $j \in Y$) 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. ○

Using these definition, we can state the main result of this section.

Lemma 8.15. *If (X, Y) is q-good, where $q < 1/50$, then (X, Y) has a $13q$-strong line.*

(We haven't tried to optimize the constants 1/50 and 13.) We will apply this lemma to $([n], [n])$ to deduce that A has a strong line.

Our work in this section is oblivious of the particular classification into large and small elements. All we need to know is that most generalized diagonals contain exactly one large element. Therefore, the main result of this section, Lemma 8.15, is of independent interest.

We start by estimating the number of large entries in a q-good restriction.

Lemma 8.16. *If (X, Y) is q-good then $A[X, Y]$ has at least $(1 - q)|X|$ large entries.*

Proof. We can pick a random entry from $A[X, Y]$ by picking a random generalized diagonal D_π and picking a random entry from D_π. Since $A[X, Y]$ is q-good, we deduce that a random entry in $A[X, Y]$ is large with probability at least $(1 - q)/|X|$. Therefore $A[X, Y]$ contains at least $(1 - q)|X|$ large entries. □

We will prove Lemma 8.15 by induction on $|X|$. The base case is the following lemma, in which the size of X is not fixed but depends on q.

Lemma 8.17. *If (X, Y) is q-good, where $|X| > 1$ and $q < 1/(|X|(|X| - 1))$, then (X, Y) has a 0-strong line.*

Proof. Let $|X| = m$. For simplicity of notation, assume $X = Y = [m]$.

For a permutation $\pi \in S_m$, let $\pi + i \in S_m$ be defined by $(\pi + i)(x) = \pi(x) + i$, where addition is modulo m. For a random permutation $\pi \in S_m$, the probability that $D_\pi, D_{\pi+1}, \ldots, D_{\pi+(m-1)}$ are all good is at least $1 - mq > 1 - 1/(m - 1) \geq 0$, and so there is some $\pi \in S_m$ for which all of $D_\pi, \ldots, D_{\pi+(m-1)}$ are good. Without loss of generality, suppose that π is the identity permutation.

The generalized diagonal D_π contains a unique large entry. Without loss of generality, suppose a_{11} is large. For each $j \in [m]$, the generalized diagonal $D_{\pi+j}$ contains some strong entry $a_{i(i+j)}$. If $i, i + j \neq 1$ then a random generalized diagonal in $A[X, Y]$ passes through both large entries $a_{11}, a_{i(i+j)}$ with probability $1/m(m - 1)$, contrary to the assumption that (X, Y)

is q-good. Therefore either $i = 1$ or $i + j = 1$. In other words, all large entries are either on the first row or on the first column.

We claim that either all large entries are on the first row, or all of them are on the first column. Otherwise, there are large entries a_{1j}, a_{k1}, where $j, k \neq 1$. As before, the probability that a random generalized diagonal passes through both of them is $1/m(m-1)$, contrary to the assumption that (X, Y) is q-good. We conclude that either the first row or the first column consists entirely of large entries. In other words, (X, Y) has a 0-strong line. $\square$

The inductive step uses the following simple lemma, which we will apply with $|X'| = \lfloor |X|/2 \rfloor$.

Lemma 8.18. *Suppose that (X, Y) is q-good, where $q < 1/2$. Every $X' \subset X$ can be completed to a restriction (X', Y') such that either (X', Y') or $(X \setminus X', Y \setminus Y')$ is q-good. A similar claim holds for every $Y' \subset Y$.*

Proof. Let $X' \subset X$ be an arbitrary subset of X. For $Y' \subset Y$ of size $|Y'| = |X'|$, let $P_{X',Y'}$ be the set of bijections from X to Y that send X' to Y'. The sets $P_{X',Y'}$ partition the set of all bijections from X to Y, and so a simple averaging argument shows that for some set Y',

$$p = \Pr_{\pi \in P_{X',Y'}} [D_\pi \text{ is good}] \geq 1 - q.$$

For $\pi \in P_{X',Y'}$, let $\pi_1 = \pi|_{X'}$ and $\pi_2 = \pi|_{Y'}$. If D_π is good then exactly one of D_{π_1}, D_{π_2} is good. We want to show that one of these events happens with probability $1 - q$ for random $\pi \in P_{X',Y'}$. Let

$$p_1 = \Pr_{\pi \in P_{X',Y'}} [D_{\pi_1} \text{ is good}], \quad p_2 = \Pr_{\pi \in P_{X',Y'}} [D_{\pi_2} \text{ is good}].$$

Notice that if $p_1 \geq 1 - q$ then (X', Y') is q-good, and if $p_2 \geq 1 - q$ then $(X \setminus X', Y \setminus Y')$ is q-good. Our goal is to show that one of these inequalities holds.

Clearly $p \leq p_1(1 - p_2) + (1 - p_1)p_2$. We want to show that $p \geq 1 - q$ forces $\max(p_1, p_2) \geq 1 - q$. To that end, let $p_1 = (1 + \delta_1)/2$ and $p_2 = (1 + \delta_2)/2$. Notice that $|\delta_1|, |\delta_2| \leq 1$. Calculation shows that

$$p \leq p_1(1 - p_2) + (1 - p_2)p_2 = \frac{(1 + \delta_1)(1 - \delta_2) + (1 - \delta_1)(1 + \delta_2)}{4} = \frac{1 - \delta_1\delta_2}{2}.$$

Since $p \geq 1 - q > 1/2$, we see that $\delta_1\delta_2 < 0$, and so one of δ_1, δ_2 is positive and the other is negative. If $\delta_1 > 0$ then using $\delta_2 \geq -1$,

$$p \leq \frac{1 - \delta_1\delta_2}{2} \leq \frac{1 + \delta_1}{2} = p_1,$$

and so $p_1 \geq p \geq 1 - q$. Similarly, if $\delta_2 > 0$ then $p_2 \geq p \geq 1 - q$, completing the proof. □

Using Lemma 8.18 we will come up with several candidates for a strong line in $A[X, Y]$. The next lemma will be used to show that all these candidates point at the same line.

Lemma 8.19. *Suppose that (X, Y) is q-good, and let $(X_1, Y_1), (X_2, Y_2)$ be subrestrictions of (X, Y), that is $X_1, X_2 \subset X$ and $Y_1, Y_2 \subset Y$. Suppose that (X_1, Y_1) has a p_1-strong line and that (X_2, Y_2) has a p_2-strong line. Let $t_1 = \lceil (1 - p_1)|X_1| \rceil$ and $t_2 = \lceil (1 - p_2)|X_2| \rceil$. If*

$$t_1, t_2 \geq 2 \quad \textit{and} \quad t_1 t_2 \geq 4q|X|^2$$

then the two strong lines are the same (that is, they are defined by the same row of X or by the same column of Y).

Note that t_1, t_2 are the number of large entries in the strong lines whose existence is assumed.

Proof. Suppose, for the sake of contradiction, that the two strong lines are not the same. Let $L_1 \subseteq X_1 \times Y_1$ consist of the first t_1 indices of large entries in the first strong line, and let $L_2 \subseteq X_2 \times Y_2$ consist of the first t_2 indices of large entries in the second strong line. 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 $j_1 = j_2$. A generalized diagonal in $A[X, Y]$ never contains two conflicting entries, but has a chance of $1/(|X|(|X| - 1))$ of containing two non-conflicting entries. We want to show that the number of conflicting pairs of entries, one from L_1 and one from L_2, is small, and so the probability that a generalized diagonal contains two large entries is large. That will imply the desired contradiction.

If L_1 corresponds to row i and L_2 corresponds to column j then an entry from L_1 not on column j doesn't conflict with an entry from L_2 not on row i. Therefore there are at least $(t_1 - 1)(t_2 - 1)$ non-conflicting pairs. If L_1 and L_2 both correspond to rows then an entry from L_1 conflicts with an entry from L_2 only if they are both on the same column, and so there are at least $t_1t_2 - \min(t_1,t_2)$ non-conflicting pairs. Without loss of generality, $t_1 \le t_2$, and so

$$t_1t_2 - \min(t_1,t_2) = t_1t_2 - t_1 \ge t_1t_2 - t_1 - (t_2-1) = (t_1-1)(t_2-1).$$

Summarizing, there must be at least $(t_1-1)(t_2-1)$ non-conflicting pairs. The probability that a generalized diagonal contains each non-conflicting pair is $1/(|X|(|X|-1))$. Since L_1, L_2 lie on lines, these events are disjoint, and so the probability that a generalized diagonal contains at least two large entries is at least

$$\frac{(t_1-1)(t_2-1)}{|X|(|X|-1)} > \frac{(t_1-1)(t_2-1)}{|X|^2} \ge \frac{t_1t_2}{4|X|^2} \ge q,$$

using $t_1, t_2 \ge 2$. This contradicts the assumption that (X,Y) is q-good, completing the proof.

□

As a warm-up illustrating how to use the preceding two lemmas, we prove another base case of our general inductive claim, which is itself proved by induction, with Lemma 8.17 as the base case.

Lemma 8.20. *If (X,Y) is q-good and $q < 1/(4|X|)$, then (X,Y) has a 0-strong line.*

Proof. The proof is by induction on $m = |X|$. When $m = 1$, the claim is trivial. When $2 \le m \le 5$, the claim follows from Lemma 8.17, so assume $m \ge 6$.

Let X' be an arbitrary subset of X of size $s = \lfloor m/2 \rfloor$. Use Lemma 8.18 to complete X' to a restriction (X',Y') such that either (X',Y') or $(X \setminus X', Y \setminus Y')$ is q-good. Since $q < 1/(4m) < 1/(4s)$, the induction hypothesis implies that either (X',Y') or $(X \setminus X', Y \setminus Y')$ has a 0-strong line. Similarly, each $Y' \subset Y$ of size s can be extended to a restriction (X',Y') such that either (X',Y') or $(X \setminus X', Y \setminus Y')$ has a 0-strong line.

Each of the 0-strong lines obtained in this way contains either $s \ge 3$ or $m - s \ge s \ge 3$ strong entries. Since $m \ge 6$ and $s/m \ge 3/7$,

$$s^2 \ge \frac{9}{49}m^2 > m > 4qm^2.$$

Therefore Lemma 8.19 shows that all the 0-strong lines are the same. Without loss of generality, we can assume that they are all defined by row i. We will show that row i is 0-strong for (X,Y).

We start by showing that row i can contain at most one small entry. If it contained two small entries a_{ij}, a_{ik}, then we could choose $Y' \subset Y$ of size s that contains j but not k. There is no way to complete Y' to a restriction (X', Y') such that row i is strong for either (X', Y') or $(X \setminus X', Y \setminus Y')$, since if $i \in X'$ then $A[X', Y']$ contains the small entry a_{ij}, and if $i \in X \setminus X'$ then $A[X \setminus X', Y \setminus Y]$ contains the small entry a_{ik}. This contradiction shows that row i can contain at most one small entry.

Suppose now that row i contains exactly one small entry. Lemma 8.16 shows that $A[X, Y]$ must contain at least $(1-q)m > m - 1/4$ large entries, and so at least m. By assumption, $m-1$ of them are on row i. Let a_{kl} be a large entry not on row i. The probability that a random generalized diagonal in $A[X, Y]$ contains both a_{kl} and one of the large entries on row i is at least

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

contradicting the fact that (X, Y) is q-good. Here $1 - 1/(m-1)$ accounts for the single small entry on row i. This contradiction shows that row i must consist entirely of large entries, completing the proof. □

The proof of the general inductive step is similar, with a crucial difference towards the end. Using Lemma 8.18 and Lemma 8.19, we can locate a strong line in $A[X, Y]$. However, the reasoning used in the preceding proof shows that the quality of this strong line potentially deteriorates by a single element. The argument used in the proof to overcome this difficulty only works when $q < 1/m$. For the general case, we will use the following lemma.

Lemma 8.21. *Suppose that (X, Y) is q-good and has a p-strong line. Let $m = |X|$ and $\varrho = 2q/(1-p)$. If $m \geq 6$, $(1-p)m > 1$, $2\varrho m > 1$ and $\varrho \leq 1/2$, then that line is in fact $(q + 3\varrho)$-strong.*

Proof. Without loss of generality, suppose that row i is p-strong for (X, Y). Lemma 8.16 shows that $A[X, Y]$ contains at least $(1-q)m$ large entries. Suppose for the sake of contradiction that row i is not $(q + 3\varrho)$-strong. Then $A[X, Y]$ must contain at least $3\varrho m$ large entries outside of row i.

We claim that no column j can contain $2\varrho m$ or more large entries. Indeed, otherwise that line would be $(1-2\varrho)$-strong. Since $(1-p)m > 1$, $2\varrho m > 1$ and

$$(1-p)m \cdot 2\varrho m = 4qm^2,$$

Lemma 8.19 rules out this case. As a consequence, for each column j, there must be at least ϱm large entries outside of row i and column j. Intuitively, this shows that a random generalized diagonal passing through any large a_{ij} has a reasonable chance of passing through another large entry.

Let P_j consist of the first $\lceil \varrho m \rceil$ large entries outside of row i and column j. For each large entry a_{ij} on row i, the probability that a random generalized diagonal in $A[X,Y]$ contains both a_{ij} and any specific entry in P_j is $1/(m(m-1))$, and the probability that it contains a_{ij} as well as two specific entries in P_j is at most $1/(m(m-1)(m-2))$. Using the Bonferroni inequality (see Section 2.2), we get that a random generalized diagonal contains both a_{ij} and an entry in P_j with probability at least

$$\frac{|P_j|}{m(m-1)} - \frac{|P_j|(|P_j|-1)}{2m(m-1)(m-2)} > \frac{\varrho m}{m(m-1)} - \frac{(\varrho m+1)(\varrho m)}{2m(m-1)(m-2)} = \frac{\varrho}{m-1}\left(1 - \frac{\varrho m+1}{2(m-2)}\right).$$

Since $\varrho \le 1/2$ and $m \ge 6$, $\varrho m + 1 \le m/2 + 1 \le m-2$, and so the probability that a random generalized diagonal contains both a_{ij} and an entry in P_j is at least $\varrho/(2(m-1))$. Since there are at least $(1-p)m$ large entries on row i, we conclude that a random generalized diagonal contains at least two large entries with probability at least

$$\frac{\varrho}{2(m-1)} \cdot (1-p)m > \frac{\varrho(1-p)}{2} = q.$$

This contradicts the assumption that (X,Y) is q-good. $\square$

Using the preceding lemma, we can prove the main result of this section. This time we use Lemma 8.20 as the base case.

Lemma 8.15. *If (X,Y) is q-good, where $q < 1/50$, then (X,Y) has a $13q$-strong line.*

Proof. The proof is by induction on $m = |X|$. If $m < 1/(4q)$ then Lemma 8.20 shows that (X,Y) has a 0-strong line, so we can assume that $m \ge 1/(4q) > 12$.

Let X' be an arbitrary subset of X of size $s = \lfloor m/2 \rfloor$. Use Lemma 8.18 to complete X' to a restriction (X',Y') such that either (X',Y') 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 $13q$-strong line. Similarly, each $Y' \subset Y$ of size s can be extended to a restriction (X', Y') such that either (X', Y') or $(X \setminus X', Y \setminus Y')$ has a $13q$-strong line.

We claim that all these strong lines are the same. Since $s \geq m - s$, each such strong line contains at least $(1 - 13q)s > (1 - 13/50)7 > 2$ large entries. Moreover, since $s/m \geq 7/13$,

$$((1 - 13q)s)^2 \geq \left(1 - \frac{13}{50}\right)^2 \left(\frac{7}{13}\right)^2 m^2 > \frac{4}{50} m^2 > 4qm^2.$$

Therefore Lemma 8.19 shows that all these strong lines are the same. Without loss of generality, we can assume that they are all defined by row i. We will show that row i is $13q$-strong for (X, Y).

We first show that row i can contain at most $\lfloor 13qm + 1 \rfloor$ small entries. Indeed, suppose that row i contained at least $\lfloor 13qm + 2 \rfloor$ small entries. Then there exists a subset $Y' \subset Y$ of size s such that row i contains at least $\lfloor 13qs + 1 \rfloor$ small entries in columns corresponding to Y', and at least $\lfloor 13q(m - s) + 1 \rfloor$ small entries in columns corresponding to $Y \setminus Y'$. It is not hard to check that for each restriction (X', Y'), row i is $13q$-strong for neither (X', Y') nor $(X \setminus X', Y \setminus Y')$. This contradiction shows that row i contains at most $\lfloor 13qm + 1 \rfloor$ small entries. Therefore row i is $(13q + 1/m)$-strong for (X, Y).

To complete the proof, we need to improve $(13q + 1/m)$-strong to $13q$-strong, and to that end we use Lemma 8.21. Let $p = 13q + 1/m$ and $\varrho = 2q/(1 - p)$. We need to check all the hypotheses of the lemma:

$$\begin{aligned}
&(1 - p)m = (1 - 13q)m - 1 > \frac{37}{50} \cdot 13 - 1 > 1, \\
&2\varrho m = \frac{4qm}{1 - 13q - 1/m} > 4qm \geq 1, \\
&\varrho = \frac{2q}{1 - p} < \frac{2/50}{1 - 13/50 - 1/13} < 1/2.
\end{aligned}$$

All the hypotheses are satisfied, and so the lemma shows that row i is in fact $(q + 3\varrho)$-strong. Since

$$q + 3\varrho = \left(1 + \frac{6}{1 - p}\right) q < \left(1 + \frac{6}{1 - 13/50 - 1/13}\right) q < 13q,$$

we are done. $\square$

Putting this together with Corollary 8.14, we can summarize all the work we have done so far.

Corollary 8.22. *The matrix A contains a line (either a row or a column) which contains $(1 - O(\epsilon^{1/7}))m$ large entries, where an entry is large if it is $50\epsilon^{1/7}$-close in magnitude to either $2\mathfrak{c}$ or $2(1 - \mathfrak{c})$.*

Proof. Corollary 8.14 shows that $([n], [n])$ is $50\epsilon^{1/7}$-good. Therefore Lemma 8.15 shows that $([n], [n])$ has a $650\epsilon^{1/7}$-strong line. □

8.8 Constructing the approximation

To see where we stand, let us focus for a moment on the case $\mathfrak{c} = 1/2$. Corollary 8.22 shows that A contains a line in which almost all entries are close to ± 1. In other words, assuming that the strong line is the first row, A looks quite similar to the canonical example shown in the introduction, which corresponds to a sum of $n/2$ disjoint cosets:

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

When $\mathfrak{c} = 1/2$, Lemma 8.2 shows that the first row sums to zero, and so roughly half of the entries are 1, and half are -1. This will be enough to construct a good approximation for $\mathcal{F}$.

For the rest of this section, we make the following simplifying assumption: *the strong line given by Corollary 8.22 is row* 1. In other words, $1 - O(\epsilon^{1/7})$ *of the entries in row* 1 *are large (in the sense of Corollary 8.22)*.

For general $\mathfrak{c}$, the situation becomes slightly more complicated, since now the corollary shows that each large entry in the strong line is close to one of the four values $\pm 2\mathfrak{c}, \pm 2(1 - \mathfrak{c})$. To understand what that means, we use the relation between a_{ij} and the more immediately relevant quantities

$$\tau_{ij} = \frac{|\mathcal{F} \cap T_{i,j}|}{(n-1)!}. \tag{8.9}$$

An easy calculation using the definition of a_{ij} (which we do below) will show that if a_{ij} is large then τ_{ij} is close to one of the four values $0, 1, 2\mathfrak{c}, 2\mathfrak{c}-1$. Among the two last values, one is outside of $[0,1]$, and we name the other γ:

$$\gamma_{ij} = \begin{cases} 2\mathfrak{c} & \text{if } \mathfrak{c} \le 1/2, \\ 2\mathfrak{c}-1 & \text{if } \mathfrak{c} \ge 1/2. \end{cases} \tag{8.10}$$

Lemma 8.23. *We have*

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

Moreover, each τ_{ij} is $2/n$-close to $a_{ij}/2+\mathfrak{c}$. Furthermore, if a_{ij} is large (in the sense of Corollary 8.22) then τ_{ij} is $26\epsilon^{1/7}$-close to $\{0,1,\gamma\}$.

Proof. Equation (8.10) at the beginning of Section 8.3 defines a_{ij} as

$$\begin{aligned} a_{ij} &= (n-1)\langle f, \mathbf{1}_{T_{ij}}\rangle - \frac{n-2}{n}(2\mathfrak{c}-1) \\ &= (n-1)\langle 2\mathbf{1}_{\mathcal{F}} - \mathbf{1}_{S_n}, \mathbf{1}_{T_{ij}}\rangle - \frac{n-2}{n}(2\mathfrak{c}-1) \\ &= (n-1)\left(2\frac{|\mathcal{F}\cap T_{ij}|}{n!} - \frac{1}{n}\right) - \frac{n-2}{n}(2\mathfrak{c}-1) \\ &= 2\frac{n-1}{n}\tau_{ij} - \frac{(n-2)(2\mathfrak{c})+1}{n}. \end{aligned}$$

Since $\tau_{ij} \le 1$, this shows that $a_{ij} < 2$, while $\mathfrak{c} \le 1$ shows that $a_{ij} \ge -(2n-3)/n > -2$. Therefore $|a_{ij}| < 2$.

Writing τ_{ij} as a function of a_{ij},

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

Since $|a_{ij}| < 2$ and $0 \le \mathfrak{c} \le 1$, $|a_{ij}+1-2\mathfrak{c}| \le 3$. Therefore the error term has magnitude at most $3/(2(n-1)) \le 2/n$, since by assumption (8.1), $n \ge 4$. Thus τ_{ij} is $2/n$-close to $a_{ij}/2+\mathfrak{c}$.

If a_{ij} is large then a_{ij} is $50\epsilon^{1/7}$-close to $\{\pm 2\mathfrak{c}, \pm 2(1-\mathfrak{c})\}$. Therefore τ_{ij} is $(25\epsilon^{1/7}+2/n)$-close to $\{0, 2\mathfrak{c}, 2\mathfrak{c}-1, 1\}$. Assumption (8.1) states that $1/n^{1/3} \le \epsilon^{1/7}$, and so $2/n \le 2\epsilon^{3/7} \le \epsilon^{1/7}$, since ϵ is small enough. Therefore τ_{ij} is $26\epsilon^{1/7}$-close to $\{0, 2\mathfrak{c}, 2\mathfrak{c}-1, 1\}$.

If $\mathfrak{c} \le 1/2$ then $2\mathfrak{c} - 1 \le 0$. Since $\tau_{ij} \ge 0$, if τ_{ij} is $26\epsilon^{1/7}$-close to $2\mathfrak{c} - 1$ then it is actually $26\epsilon^{1/7}$-close to 0. Similarly, if $\mathfrak{c} \ge 1/2$ then $2\mathfrak{c} \ge 1$, and since $\tau_{ij} \le 1$ we can omit $2\mathfrak{c}$ from the set. We conclude that τ_{ij} is $26\epsilon^{1/7}$-close to $\{0, \gamma, 1\}$. □

If $\mathcal{F}$ is indeed close to a family which is the union of cosets, then we expect the large entries in A to be close to $\{0, 1\}$ rather than to γ. Intuitively, this should hold for the following reason. The value of f at any permutation π depends strongly on the value of the large entry in the generalized diagonal D_π. Since for almost all π, $f(\pi) \in \{1, -1\}$, any two large entries in the strong line should either be roughly equal, or roughly at distance two. This rules out large entries which are close to γ.

To formalize this argument, we look at a large subset of the strong line given by Corollary 8.22.

Definition 8.7. Let a_{1j} be an entry on row 1. (Recall that we are assuming that row 1 is the strong line given by Corollary 8.22.) Let $r(j)$ be the probability that a random generalized diagonal passing through a_{1j} is *bad* (a generalized diagonal is *good* if contains exactly one large entry, and all other entries are *small*, of magnitude at most $50\epsilon^{1/7}$).

The entry a_{1j} is *reasonable* if all the following conditions hold:

(a) a_{1j} is large.

(b) $r(j) \le 1/5$.

(c) The function $g^{\{1\},\{j\}}$ is $(1/5, \epsilon^{1/7})$-almost Boolean. (Recall $g^{\{1\},\{j\}}$ is the restriction of f_1 to $T_{1,j}$.) ○

We first show that most entries on row 1 are reasonable.

Lemma 8.24. *Row 1 contains $(1 - O(\epsilon^{1/7}))m$ reasonable entries.*

Proof. We calculate the probability that a random entry chosen from row 1 is unreasonable. Corollary 8.22 shows that Property (a) of Definition 8.7 fails with probability $O(\epsilon^{1/7})$.

Corollary 8.14 shows that a random generalized diagonal is not good with probability at most $50\epsilon^{1/7}$. In other words,

$$\mathop{\mathbb{E}}_{j \in [n]}[r(j)] \le 50\epsilon^{1/7}.$$

Markov's inequality shows that

$$\Pr_{j \in [n]}[r(j) > 1/5] < 250\epsilon^{1/7}.$$

Therefore Property (b) in Definition 8.7 fails with probability at most $250\epsilon^{1/7}$.

Finally, recall that since f is ± 1-valued, $\mathbb{E}[(|f_1|-1)^2] \le \mathbb{E}[(|f_1|-f)^2] = \epsilon$. Markov's inequality shows that

$$\Pr_{\pi \in S_n}[(|f_1(\pi)|-1)^2 > \epsilon^{2/7}] < \epsilon^{5/7}.$$

Since the sets $T_{1,j}$ partition S_n,

$$\mathop{\mathbb{E}}_{j \in [n]} \Pr_{\pi \in T_{1,j}}[(|f_1(\pi)|-1)^2 > \epsilon^{2/7}] < \epsilon^{5/7},$$

Markov's inequality again shows that

$$\Pr_{j \in [n]}\left[\Pr_{\pi \in T_{1,j}}[(|f_1(\pi)|-1)^2 > \epsilon^{2/7}] > 1/5\right] < 5\epsilon^{5/7} = O(\epsilon^{1/7}),$$

since ϵ is small enough. This implies that Property (c) in Definition 8.7 fails with probability $O(\epsilon^{1/7})$. Using a union bound completes the proof. □

Second, we show that for any two reasonable entries, either both of the corresponding τ values are close to γ, or neither of them are.

Lemma 8.25. *Assume that γ is $156\epsilon^{1/7}$-far from $\{0,1\}$. Suppose a_{1j}, a_{1k} are two reasonable entries. Either both τ_{1j}, τ_{1k} are $26\epsilon^{1/7}$-close to γ, or neither of them are.*

Proof. Choose a permutation $\pi \in T_{1j}$ randomly. Property (b) in Definition 8.7 states that $D_\pi = \{a_{i\pi(i)} : i \in [n]\}$ is bad with probability at most $1/5$. Property (c) in the definition states that $f_1(\pi)$ is $\epsilon^{1/7}$-far from ± 1 with probability at most $1/5$. Since $(jk)\pi$ is a random permutation from T_{1k}, we get similar properties for $D_{(jk)\pi}$ and $f_1((jk)\pi)$. Since $4/5 < 1$, there is a choice of $\pi \in T_{1j}$ such that $D_\pi, D_{(jk)\pi}$ are both good, and $f_1(\pi), f_1((jk)\pi)$ are both $\epsilon^{1/7}$-close to ± 1.

Let $i = \pi^{-1}(k)$. Lemma 8.4 shows that

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

using Lemma 8.23. Since a_{ik}, a_{ij} are small, the left-hand side is $102\epsilon^{1/7}$-close to $\{0, \pm 2\}$, and therefore $\tau_{1j} - \tau_{1k}$ is $102\epsilon^{1/7}$-close to $\{0, \pm n/(n-1)\}$. Assumption (8.1) implies that $1/(n-1) < 2/n < 2/n^{1/3} \leq 2\epsilon^{1/7}$, and so $\tau_{1j} - \tau_{1k}$ is $104\epsilon^{1/7}$-close to $\{0, \pm 1\}$.

Property (a) in Definition 8.7 shows that τ_{1j}, τ_{1k} are large, and so Lemma 8.23 shows that for some $t_{1j}, t_{1k} \in \{0, \gamma, 1\}$, τ_{1j}, τ_{1k} are $26\epsilon^{1/7}$-close to t_{1j}, t_{1k} (respectively). Therefore $t_{1j} - t_{1k}$ is $156\epsilon^{1/7}$-close to $\{0, \pm 1\}$. We complete the proof by showing that this is impossible if $t_{1j} = \gamma$ and $t_{1k} \neq \gamma$ (the case $t_{1j} \neq \gamma$ and $t_{1k} = \gamma$ is symmetric).

If $t_{1j} = \gamma$ and $t_{1k} \in \{0, 1\}$, then we get that γ must be $156\epsilon^{1/7}$-close to $\{-1, 0, 1\} \cup \{0, 1, 2\} = \{-1, 0, 1, 2\}$. Since $\gamma \in [0, 1]$, it must be $156\epsilon^{1/7}$-close to $\{0, 1\}$, contrary to our assumption. □

If γ is $156\epsilon^{1/7}$-close to some $\{0, 1\}$, then we might as well treat it as that number (either 0 or 1), which is why the assumption made at the beginning of the lemma is reasonable.

At this point we are ready to state the main lemma of this section, which shows that row 1 has (roughly) the correct number of entries τ_{1j} which are close to 1, and the correct number of entries which are close to 0. The idea is to use the fact that the row sums (roughly) to $\mathfrak{c}$. Lemma 8.25 shows that the number of entries on row 1 which are close to γ is either very small or very large. We can rule out the latter case using the constraint on the sum, and then the same constraint allows us to estimate the number of entries which are close to 1 and to 0.

Lemma 8.26. *The number of entries τ_{1j} which are $51\epsilon^{1/7}$-close to 1 is $O(\epsilon^{1/7})n$-close to $\mathfrak{c}n$, and the number of entries τ_{1j} which are $51\epsilon^{1/7}$-close to 0 is $O(\epsilon^{1/7})n$-close to $(1-\mathfrak{c})n$.*

Proof. We distinguish between two cases, depending on whether γ is $156\epsilon^{1/7}$-close to $\{0, 1\}$ or not. We will use the simple formula

$$T = \sum_{j=1}^{n} \tau_{1j} = \sum_{j=1}^{n} \frac{|\mathcal{F} \cap T_{1j}|}{(n-1)!} = \frac{|\mathcal{F}|}{(n-1)!} = \mathfrak{c}n,$$

which follows from the fact that the sets T_{1j} partition S_n.

The easier case is when γ is $156\epsilon^{1/7}$-close to $\{0, 1\}$. In that case, Lemma 8.23 shows that each large entry a_{ij} corresponds to τ_{ij} which is $182\epsilon^{1/7}$-close to $\{0, 1\}$. Let N_0 be the number of large entries a_{1j} such that τ_{1j} is $182\epsilon^{1/7}$-close to 0, and let N_1 be the number of large entries a_{1j} such that τ_{1j} is $182\epsilon^{1/7}$-close to 1.

Since $0 \le \tau_{1j} \le 1$, we have

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

Since $T = \mathfrak{c}n$, this shows that

$$N_0 \le (1 + O(\epsilon^{1/7}))(1 - \mathfrak{c})n, \quad N_1 \le (1 + O(\epsilon^{1/7}))\mathfrak{c}n.$$

On the other hand, Corollary 8.22 shows that $(1 - O(\epsilon^{1/7}))n$ of the entries on row 1 are large, and so

$$N_0 + N_1 \ge (1 - O(\epsilon^{1/7}))n.$$

Therefore

$$N_1 \ge (1 - O(\epsilon^{1/7}))n - N_0 \ge (\mathfrak{c} - O(\epsilon^{1/7}))n,$$

showing that N_1 is $O(\epsilon^{1/7})n$-close to $\mathfrak{c}n$. Similarly we get that N_0 is $O(\epsilon^{1/7})n$-close to $(1 - \mathfrak{c})n$. This completes the proof when γ is $156\epsilon^{1/7}$-close to $\{0, 1\}$.

The more complicated case is when γ is $156\epsilon^{1/7}$-far from $\{0, 1\}$. Let

$$R = \{j : a_{1j} \text{ is reasonable}\}.$$

Lemma 8.24 shows that $|R| = (1 - O(\epsilon^{1/7}))n$. Lemma 8.25 shows that either all τ_{1j} for $j \in R$ are $26\epsilon^{1/7}$-close to γ, or none are. We want to rule out the first case.

Suppose that for all $j \in R$, τ_{1j} is $26\epsilon^{1/7}$-close to γ. Then

$$(\gamma - 26\epsilon^{1/7})n \le T \le (\gamma + 26\epsilon^{1/7})n + (n - |R|).$$

Since $n - |R| = O(\epsilon^{1/7})n$, we conclude that T is $O(\epsilon^{1/7})n$-close to γn. On the other hand, $T = \mathfrak{c}n$. Therefore γ is $O(\epsilon^{1/7})$-close to $\mathfrak{c}$. However, $|\gamma - \mathfrak{c}| \in \{\mathfrak{c}, 1 - \mathfrak{c}\}$, and so $|\gamma - \mathfrak{c}| \ge \eta$. Since ϵ is small enough compared to η^7, we reach a contradiction. We conclude that for all $j \in R$, τ_{1j} is $26\epsilon^{1/7}$-close to $\{0, 1\}$.

Since $|R| = (1 - O(\epsilon^{1/7}))n$, we conclude that all but $O(\epsilon^{1/7})n$ of the entries on row 1 are $26\epsilon^{1/7}$-close to $\{0, 1\}$. Therefore we can repeat the argument for the previous case (replacing large entries with reasonable entries), reaching similar conclusions. This completes the proof of the lemma. □

At this point we discharge our simplifying assumption that the strong line promised by Corollary 8.22 is row 1. Without this assumption, Lemma 8.26 reads as follows.

Lemma 8.26'. *Let τ be the $n \times n$ matrix whose entries are τ_{ij}. There is a line L of the matrix (either a row or a column) such that the following holds.*

The number of entries in L which are $51\epsilon^{1/7}$-close to 1 is $O(\epsilon^{1/7})n$-close to $\mathfrak{c}n$, and the number of entries in L which are $51\epsilon^{1/7}$-close to 0 is $O(\epsilon^{1/7})n$-close to $(1-\mathfrak{c})n$.

Using the lemma, we can state a qualified version of the main theorem, summarizing our work so far. In the statement of the following corollary, we restate the assumptions (8.1).

Corollary 8.27. *There exists $c_0 > 0$ such that the following holds.*

Let $\mathcal{F} \subseteq S_n$ be a family of permutations of size $\mathfrak{c}n!$, where $n \geq 4$, *and let $\eta = \min(\mathfrak{c}, 1-\mathfrak{c})$. Let $f = 2\mathbf{1}_{\mathcal{F}} - 1$, and let $f_1 = \hat{f}((n)) + \hat{f}((n-1,1))$ be the projection of f to $L_{(n-1,1)}$.*

If $\mathbb{E}[(f-f_1)^2] = \epsilon$, where ϵ satisfies

$$\frac{1}{n^{7/3}} \leq \epsilon \leq c_0\eta^7,$$

then there exists a family $\mathcal{G} \subseteq S_n$ which is the union of $\mathfrak{d}n$ disjoint cosets satisfying

$$|\mathcal{F}\Delta\mathcal{G}| = O\left(\epsilon^{1/7}\right)n!.$$

Moreover,

$$|\mathfrak{c} - \mathfrak{d}| = O\left(\epsilon^{1/7}\right).$$

Proof. We choose c_0 so that all statements in the proof so far stating that ϵ is small enough compared to η^7 hold. Under this choice of c_0, the assumptions (8.1) are satisfied, and so Lemma 8.26' applies. Let L be the line whose existence is given by the lemma, and define

$$\mathcal{G} = \bigcup\{T_{i,j} : \tau_{ij} \in L \text{ and } \tau_{ij} \text{ is } 51\epsilon^{1/7}\text{-close to } 1\}.$$

The lemma shows that $\mathcal{G}$ is a union of $\mathfrak{d}n$ disjoint cosets, where $|\mathfrak{c}n - \mathfrak{d}n| = O(\epsilon^{1/7})n$. Since $\tau_{ij} = |\mathcal{F} \cap T_{i,j}|/(n-1)!$,

$$|\mathcal{F} \cap \mathcal{G}| \geq \mathfrak{d}n \cdot (1 - 51\epsilon^{1/7})(n-1)! \geq (\mathfrak{c} - O(\epsilon^{1/7}))(1 - 51\epsilon^{1/7})n! = (\mathfrak{c} - O(\epsilon^{1/7}))n!.$$

Therefore

$$|\mathcal{F}\Delta\mathcal{G}| = |\mathcal{F}| + |\mathcal{G}| - 2|\mathcal{F}\cap\mathcal{G}| \le (\mathfrak{c}+\mathfrak{d})n! - 2(\mathfrak{c} - O(\epsilon^{1/7}))n! = O(\epsilon^{1/7})n!. \qquad \square$$

Note that apart from the assumptions on n and ϵ, the corollary differs from theorem also with respect to the size of $\mathcal{G}$. While this issue is not hard to fix, we defer it for later.

8.9 Culmination of the proof

In this section, we extend Corollary 8.27 to the full theorem. Apart from the issue regarding the size of $\mathcal{G}$, we have to handle the cases where $n < 4$, and where ϵ is too small or too large. If $n < 4$ or ϵ is too large then the theorem becomes trivial by choosing the constants in $O(\cdot)$ appropriately. Handling the issue that ϵ is too small is more delicate.

Looking at the proof, we used the assumption that $\epsilon \ge 1/n^{7/3}$ in two points: in the proof of Lemma 8.8, and twice in the preceding section: in Lemma 8.23 and in Lemma 8.25. In all these cases, we could have got rid of the assumption by replacing ϵ with $\max(\epsilon, 1/n^{7/3})$. However, that would have made the proof more cumbersome. Instead, we use a perturbation argument to ensure that ϵ is large enough.

Lemma 8.28. *Let $\mathcal{F} \subseteq S_n$ be a family of permutations, where $n \ge 3$, and let $\kappa \le 1/16$. If $|\mathcal{F}| \ge n!/2$, then there exists a family $\mathcal{H} \subseteq \mathcal{F}$ satisfying*

$$|\mathcal{F}\Delta\mathcal{H}| \le \sqrt{\kappa}n! \quad \text{and} \quad \kappa \le \mathbb{E}[(h-h_1)^2] \le \left(\sqrt{\mathbb{E}[(f-f_1)^2]} + 2\sqrt{\kappa}\right)^2,$$

where $f = 2\mathbf{1}_F - 1$, $h = 2\mathbf{1}_H - 1$, and f_1, h_1 are the projections of f, h into $L_{(n-1,1)}$.

If $|\mathcal{F}| \le n!/2$, then there exists a family $\mathcal{H} \supseteq \mathcal{F}$ satisfying the same properties.

Proof. Without loss of generality, we can assume that $|\mathcal{F}| \ge n!/2$ (otherwise apply the same argument to the complement of $\mathcal{F}$).

For $\pi \in S_n$, let $(-1)^\pi$ be the sign of π. Since $n \ge 3$, in each coset $T_{i,j}$ half of the permutations are even and half are odd. In other words, if we define a function $s \in \mathbb{R}[S_n]$ by $s(\pi) = (-1)^\pi$, then s is orthogonal to $L_{(n-1,1)}$. Since $\|s\| = 1$,

$$\mathbb{E}[(f-f_1)^2] \ge \langle f, s\rangle^2 = \left(\mathop{\mathbb{E}}_{\pi\in S_n} (-1)^\pi f(\pi)\right)^2. \tag{8.11}$$

We consider two cases: at least half of the permutations in $\mathcal{F}$ are even, and at least half of the permutations in $\mathcal{F}$ are odd. Suppose first that half of the permutations in $\mathcal{F}$ are even. There are at least $n!/4 \ge \sqrt{\kappa}n!$ of them. Define a family $\mathcal{G}$ by removing $\sqrt{\kappa}n!$ even permutations from $\mathcal{F}$, and let $g = 2\mathbf{1}_{\mathcal{G}} - 1$. We have

$$\langle g, s\rangle = \mathop{\mathbb{E}}_{\pi\in S_n}(-1)^\pi g(\pi) = \mathop{\mathbb{E}}_{\pi\in S_n}(-1)^\pi f(\pi) - 2\sqrt{\kappa} = \langle f, s\rangle - 2\sqrt{\kappa}.$$

Therefore, either $\langle f, s\rangle \ge \sqrt{\kappa}$ or $\langle g, s\rangle \le -\sqrt{\kappa}$. In the first case, (8.11) shows that $\mathbb{E}[(f-f_1)^2] \ge \kappa$, and so we can take $\mathcal{H} = \mathcal{F}$. In the second case, (8.11) shows that $\mathbb{E}[(g-g_1)^2] \ge \kappa$. Moreover, since $g - g_1$ and $f - f_1$ are projections to $L^\perp_{(n-1,1)}$,

$$\|g - g_1\| \le \|f - f_1\| + \|(g - g_1) - (f - f_1)\| \le \|f - f_1\| + \|g - f\| = \|f - f_1\| + 2\sqrt{\kappa}.$$

Therefore $\mathbb{E}[(g-g_1)]^2 = \|g - g_1\|^2 \le (\sqrt{\mathbb{E}[(f-f_1)^2]} + 2\sqrt{\kappa})^2$, showing that we can take $\mathcal{H} = \mathcal{G}$.

When at least half the permutations in $\mathcal{F}$ are odd, the reasoning is similar. There are at least $\sqrt{\kappa}n!$ of then, and we define a family $\mathcal{G}$ by removing $\sqrt{\kappa}n!$ odd permutations from $\mathcal{F}$. This time we have

$$\langle g, s\rangle = \mathop{\mathbb{E}}_{\pi\in S_n}(-1)^\pi g(\pi) = \mathop{\mathbb{E}}_{\pi\in S_n}(-1)^\pi f(\pi) + 2\sqrt{\kappa} = \langle f, s\rangle + 2\sqrt{\kappa},$$

and so either $\langle f, s\rangle \le -\sqrt{\kappa}$ or $\langle g, s\rangle \ge \sqrt{\kappa}$. In the first case we can take $\mathcal{H} = \mathcal{F}$ and in the second $\mathcal{H} = \mathcal{G}$, as before. □

Using this perturbation lemma, we can extend Corollary 8.27 to the general case.

Lemma 8.29. *Let $\mathcal{F} \subseteq S_n$ be a family of permutations of size $\mathfrak{c}n!$ and let $\eta = \min(\mathfrak{c}, 1-\mathfrak{c})$. Let $f = 2\mathbf{1}_{\mathcal{F}} - 1$, and let $f_1 = \hat{f}((n)) + \hat{f}((n-1,1))$ be the projection of f to $L_{(n-1,1)}$.*

If $\mathbb{E}[(f-f_1)^2] = \epsilon$ then there exists a family $\mathcal{G} \subseteq S_n$ which is the union of $\mathfrak{d}n$ disjoint cosets satisfying

$$|\mathcal{F}\Delta\mathcal{G}| = O\left(\frac{1}{\eta}\left(\epsilon^{1/7} + \frac{1}{n^{1/3}}\right)\right)n!.$$

Moreover,

$$|\mathfrak{c} - \mathfrak{d}| = O\left(\epsilon^{1/7} + \frac{1}{n^{1/3}}\right).$$

Proof. If $n \geq 4$ and $1/n^{7/3} \leq \epsilon \leq c_0\eta^7$, then the lemma follows directly from Corollary 8.27, since $\eta \leq 1/2$.

If $n < 4$ then there are two cases: either $\epsilon = 0$ or not. If $\epsilon = 0$ then the lemma trivially holds by taking $\mathcal{G} = \mathcal{F}$. Otherwise, since $n < 4$, there is a finite number of possible families $\mathcal{F}$, and so $\epsilon \geq \epsilon_0$ for some constant $\epsilon_0 > 0$. Therefore by choosing the constants in $O(\cdot)$ appropriately, the lemma trivially holds for any family $\mathcal{G}$.

If $\epsilon > c_0\eta^7$ then $\epsilon^{1/7}/\eta > c_0^{1/7}$, and so choosing the constants in $O(\cdot)$ appropriately, the lemma trivially holds.

Finally, suppose that $\epsilon < 1/n^{7/3}$. Using Lemma 8.28 with $\kappa = 1/n^{7/3}$, we get a family $\mathcal{H}$ such that $|\mathcal{F}\Delta\mathcal{H}|/n! \leq 1/n^{7/6}$ and the value $\epsilon' = \mathbb{E}[(h - h_1)^2]$ satisfies

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

Furthermore, if $\mathfrak{c}' = |\mathcal{H}|/n!$ and $\eta' = \min(\mathfrak{c}', 1-\mathfrak{c}')$ then $|\mathfrak{c}-\mathfrak{c}'| \leq 1/n^{7/6}$ implies that $|\eta-\eta'| \leq 1/n^{7/6}$.

We now consider two cases: either $\epsilon' > c_0\eta'^7$ or not. In the first case, $c_0\eta'^7 < \epsilon' \leq 9/n^{7/3}$ and so $\eta' = O(1/n^{1/3})$. We conclude that $\eta = O(1/n^{1/3})$ or $1/\eta \cdot 1/n^{1/3} = \Omega(1)$, and so by choosing the constants in $O(\cdot)$ appropriately, the theorem trivially holds.

The more interesting case is when $\epsilon' \leq c_0\eta'^7$. In this case we can apply Corollary 8.27 to the perturbed family $\mathcal{H}$ to get a family $\mathcal{G}$ which is the union of $\mathfrak{d}n$ disjoint cosets satisfying

$$|\mathcal{H}\Delta\mathcal{G}| = O(\epsilon'^{1/7})n! \quad \text{and} \quad |\mathfrak{c}' - \mathfrak{d}| = O(\epsilon'^{1/7}).$$

Since $\epsilon'^{1/7} = O(1/n^{1/3})$, the triangle inequality shows that

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

completing the proof. □

Finally, we deduce Theorem 8.1 by adding or removing an appropriate number of cosets to $\mathcal{G}$.

Theorem 8.1. *Let $\mathcal{F} \subseteq S_n$ be a family of permutations of size $\mathfrak{c}n!$, and let $\eta = \min(\mathfrak{c}, 1-\mathfrak{c})$. Let $f = 2\mathbf{1}_{\mathcal{F}} - 1$, and let $f_1 = \hat{f}((n)) + \hat{f}((n-1,1))$ be the projection of f to $L_{(n-1,1)}$.*

If $\mathbb{E}[(f - f_1)^2] = \epsilon$ *then there exists a family* $\mathcal{G} \subseteq S_n$ *which is the union of* $\lceil \mathfrak{c}n \rfloor$ *disjoint cosets satisfying*

$$|\mathcal{F} \Delta \mathcal{G}| = O\left(\frac{1}{\eta}\left(\epsilon^{1/7} + \frac{1}{n^{1/3}}\right)\right) n!.$$

Moreover,

$$|\mathfrak{c}n - \lceil \mathfrak{c}n \rfloor| = O\left(\left(\epsilon^{1/7} + \frac{1}{n^{1/3}}\right) n\right).$$

Proof. Let $\mathcal{G}'$ be the family given by Lemma 8.29. Let $e = |\mathfrak{d}n - \lceil \mathfrak{c}n \rfloor|$. Since $|\lceil \mathfrak{c}n \rfloor - \mathfrak{c}n| \leq 1/2$,

$$e \leq |\mathfrak{d}n - \mathfrak{c}n| + \frac{1}{2} = O\left(\epsilon^{1/7} + \frac{1}{n^{1/3}}\right) n.$$

We can form a family $\mathcal{G}$ consisting of $\lceil \mathfrak{c}n \rfloor$ disjoint cosets by adding or removing e appropriate cosets to $\mathcal{G}'$. The triangle inequality shows that

$$|\mathcal{F} \Delta \mathcal{G}| = O\left(\frac{1}{\eta}\left(\epsilon^{1/7} + \frac{1}{n^{1/3}}\right)\right) n!,$$

since $\eta \leq 1/2$.

Finally, since $\lceil \mathfrak{c}n \rfloor$ is the integer closest to $\mathfrak{c}n$,

$$|\mathfrak{c}n - \lceil \mathfrak{c}n \rfloor| \leq |\mathfrak{c}n - \mathfrak{d}n| = O\left(\epsilon^{1/7} + \frac{1}{n^{1/3}}\right) n. \qquad \square$$

Chapter 9

Extremal combinatorics in theoretical computer science

In this chapter we explain how our work is connected to theoretical computer science, by presenting several applications of extremal combinatorics to theoretical computer science. While we do not present any applications of the results proved in this thesis, we give applications of similar types of results: Erdős–Ko–Rado type theorems and stability results.

We start by presenting Kalai's proof of a quantitative version of Arrow's theorem [57], a result for which the Friedgut–Kalai–Naor theorem was conceived and proved. Arrow's theorem is a result in social choice theory, aspects of which nowadays form a respected sub-area of algorithmic game theory.

Next, we explain how extremal combinatorics is used to prove inapproximability results for vertex cover and its extension to k-uniform hypergraphs. We focus on a simple $(2-\epsilon)$-inapproximability result for 4-uniform hypergraph vertex cover which employs the Ahlswede–Khachatrian theorem, which we proved in Chapter 5.

Finally, we explain how the Friedgut–Kalai–Naor theorem is used when analyzing dictatorship tests, another device used to prove inapproximability results. While oftentimes dictatorship tests are analyzed using other means, the analysis of assignment testers in Dinur's proof of the PCP theorem [15] does use the Friedgut–Kalai–Naor theorem.

None of the results presented in this expository chapter are original. However, we have tried to highlight the way in which extremal combinatorics is used in these results.

9.1 Arrow's theorem

One of the earliest applications of Fourier analysis in computer science is Kalai's quantitative version of Arrow's theorem [57]. The seminal result of Friedgut, Kalai and Naor [42] appeared naturally in the context of the proof, and it was Kalai's work that motivated it. To this day, much of the terminology used in the analysis of Boolean functions comes from voting.

Arrow's theorem considers an election in which each voter has to rank *all* candidates (a similar result for the more usual situation in which each voter chooses only one candidate is the Gibbard–Satterthwaite theorem).

Definition 9.1. Consider an election with n voters and m candidates, in which each voter ranks all candidates. A *profile* is an element of S_m^n which represents all the votes. A *voting rule* is a function $\varphi\colon S_m^n \to S_m$ giving the results of the election, which are a ranking of the candidates. ○

A reasonable requirement for the voting rule is that if we restrict ourselves to two candidates i, j, then the results of the election only depend on the votes restricted to i, j. Another reasonable requirement is that if all voters prefer candidate i to j, then the voting rule produce an order in which i is preferred over j.

Definition 9.2. Let $\pi \in S_m$ be a permutation, and $i, j \in [m]$. We define

$$\pi|_{i,j} = [\![\pi^{-1}(i) < \pi^{-1}(j)]\!].$$

If π is a voter's ranking of the candidates, then $\pi|_{i,j} = 1$ if the voter prefers i over j.

A voting rule φ satisfies *independence of irrelevant alternatives* (IIA) if for any two candidates i, j, $\varphi(\alpha_1, \ldots, \alpha_n)|_{i,j}$ depends only on $\alpha_1|_{i,j}, \ldots, \alpha_n|_{i,j}$.

A voting rule φ satisfies *unanimity* if whenever all voters prefer i over j, the results produced by the rule also prefer i over j. In symbols, if $\alpha_k|_{i,j} = 1$ for all $k \in [n]$ then $\varphi(\alpha_1, \ldots, \alpha_n)|_{i,j} = 1$. ○

Arrow's theorem states that the only reasonable voting rule is a dictatorship.

Theorem 9.1 (Arrow [6])**.** *Let $m \geq 3$. Every voting rule φ which satisfies IIA and unanimity is of the form $\varphi(\alpha_1, \dots, \alpha_n) = \alpha_s$ for some $s \in [n]$.*

We will be especially interested in *symmetric* voting rules, in which all candidates are equal.

Definition 9.3. A voting rule φ is *symmetric* if for all $\pi \in S_m$, $\varphi(\pi\alpha_1, \dots, \pi\alpha_n) = \pi\varphi(\alpha_1, \dots, \alpha_n)$. (Multiplication is defined so that, for example, if α is the ranking $1 > 2 > \cdots > m$ then $\pi\alpha$ is the ranking $\pi(1) > \pi(2) > \cdots > \pi(m)$.) ○

If a symmetric rule satisfies IIA, then it is in effect given by a function $f\colon \{0,1\}^n \to \{0,1\}$, which satisfies the following property: for each $i, j \in [m]$,

$$\varphi(\alpha_1, \dots, \alpha_n)|_{i,j} = f(\alpha_1|_{i,j}, \dots, \alpha_n|_{i,j}).$$

By considering $i = 1, j = 2$ and then $i = 2, j = 1$, we deduce that f is anti-symmetric:

$$f(1 - t_1, \dots, 1 - t_n) = 1 - f(t_1, \dots, t_n).$$

This leads to an alternative formulation of Arrow's theorem for the symmetric case.

Definition 9.4. A *binary voting rule* is a function $f\colon \{0,1\}^n \to \{0,1\}$ satisfying

$$f(1 - t_1, \dots, 1 - t_n) = 1 - f(t_1, \dots, t_n).$$

We call the latter property *antisymmetry*, and it implies that f is *balanced*, that is $\Pr[f = 0] = \Pr[f = 1] = 1/2$, where the probability is taken with respect to the uniform distribution on $\{0,1\}^n$.

Given vectors $\alpha_1, \dots, \alpha_n \in S_m$ and a binary voting rule f, define a binary relation $R_f^{(m)}$ on $[m]$ which represents the outcome of the election by $i \; R_f^{(m)} \; j$ if $f(\alpha_1|_{i,j}, \dots, \alpha_n|_{i,j}) = 1$. The relation $R_f^{(m)}$ is *rational* if R_f is transitive, that is if $i \; R_f^{(m)} \; j$ and $j \; R_f^{(m)} \; k$ imply $i \; R_f^{(m)} \; k$. ○

A rational outcome is simply one that represents an actual ranking of the candidates. In this language, we can restate Arrow's theorem for the symmetric case as follows.

Theorem 9.2. *Let f be a binary voting rule. If $m \geq 3$ and $R_f^{(m)}$ is always rational then f depends on one coordinate only.*

Following Kalai, we prove Theorem 9.2 using Fourier analysis, for the case in which there are three candidates. Many other proofs are known, combinatorial, topological and geometric. The advantage of this method of proof is that it generalizes to give a stability version of Arrow's theorem: if $R_f^{(m)}$ is almost always rational, then f is close to a function which depends on one coordinate only.

Let $\alpha_1, \dots, \alpha_n \in S_3$ be the voting profile, and define

$$x_s = \alpha_s|_{12}, \quad y_s = \alpha_s|_{23}, \quad z_s = \alpha_s|_{31}.$$

Since each α_s is a permutation, $(x_s, y_s, z_s) \notin \{(0,0,0),(1,1,1)\}$. Similarly, the outcome is rational if

$$(f(x_1,\dots,x_n), f(y_1,\dots,y_n), f(z_1,\dots,z_n)) \notin \{(0,0,0),(1,1,1)\}.$$

The starting point of Kalai's proof is the following formula for the probability that the outcome is rational.

Lemma 9.3. *Let f be a binary voting rule. The probability that $R_f^{(3)}$ is irrational is*

$$\frac{1}{4} + 3\sum_{X \neq \varnothing} \left(-\frac{1}{3}\right)^{|X|} \hat{f}(X)^2.$$

Proof. Let p be the required probability. We start with the formula

$$p = \frac{1}{6^n} \sum_{x,y,z \in \{0,1\}^n} \Psi(x,y,z)(f(x)f(y)f(z) + (1-f(x))(1-f(y))(1-f(z))),$$

where Ψ is the characteristic function of all voting profiles:

$$\Psi(x,y,z) = \prod_{s=1}^{n} [\![(x_s,y_s,z_s) \notin \{(0,0,0),(1,1,1)\}]\!].$$

Applying Parseval's identity, we deduce

$$p = \frac{8^n}{6^n} \sum_{X,Y,Z \subseteq [n]} \hat{\Psi}(X,Y,Z)(\hat{f}(X)\hat{f}(Y)\hat{f}(Z) + \widehat{1-f}(X)\widehat{1-f}(Y)\widehat{1-f}(Z)).$$

We can compute the Fourier expansion of Ψ explicitly. The Fourier characters are $\chi_{X,Y,Z}$ for $X, Y, Z \subseteq [n]$. For every $s \in [n]$,

$$\begin{aligned}[\![(x_s,y_s,z_s) \notin \{(0,0,0),(1,1,1)\}]\!] &= \frac{3}{4} - \frac{1}{4}((-1)^{x_s+y_s} + (-1)^{x_s+z_s} + (-1)^{y_s+z_s}) \\ &= \frac{3}{4}\chi_{\varnothing,\varnothing,\varnothing} - \frac{1}{4}\chi_{\{s\},\{s\},\varnothing} - \frac{1}{4}\chi_{\{s\},\varnothing,\{s\}} - \frac{1}{4}\chi_{\varnothing,\{s\},\{s\}}.\end{aligned}$$

The Fourier expansion of Ψ is obtained by multiplying this expression for all $s \in [n]$. The Fourier coefficient $\hat{\Psi}(X,Y,Z)$ is non-zero if for each $s \in [n]$,

$$(X \cap \{s\}, Y \cap \{s\}, Z \cap \{s\}) \in \{(\varnothing,\varnothing,\varnothing), (\{s\},\{s\},\varnothing), (\{s\},\varnothing,\{s\}), (\varnothing,\{s\},\{s\})\}.$$

In other words, each element $s \in [n]$ belongs to either 0 or 2 of X,Y,Z. Since each element belongs to an even number of sets X,Y,Z, we deduce that $X \Delta Y \Delta Z = \varnothing$. Furthermore, for each $s \in [n]$ such that $s \notin X,Y,Z$, there is a factor of 3/4, and for each other s, there is a factor of $-1/4$. We have exactly $|X \cup Y \cup Z|$ factors of the second type, and so

$$\hat{\Psi}(X,Y,Z) = \left(\frac{3}{4}\right)^{n-|X\cup Y\cup Z|} \left(-\frac{1}{4}\right)^{|X\cup Y\cup Z|}.$$

Canceling a factor of $(3/4)^n$,

$$p = \sum_{\substack{X,Y,Z\subseteq[n] \\ X\Delta Y\Delta Z=\varnothing}} \left(-\frac{1}{3}\right)^{|X\cup Y\cup Z|} (\hat{f}(X)\hat{f}(Y)\hat{f}(Z) + \widehat{1-f}(X)\widehat{1-f}(Y)\widehat{1-f}(Z)).$$

Next, note that $\widehat{1-f}(X) = [\![X = \varnothing]\!] - \hat{f}(X)$ due to the linearity of the Fourier transform. Hence if none of X,Y,Z are empty, then

$$\hat{f}(X)\hat{f}(Y)\hat{f}(Z) + \widehat{1-f}(X)\widehat{1-f}(Y)\widehat{1-f}(Z) = \hat{f}(X)\hat{f}(Y)\hat{f}(Z) - \hat{f}(X)\hat{f}(Y)\hat{f}(Z) = 0.$$

The only non-zero summands correspond therefore to $X = Y$, $Y = Z$ and $X = Z$. When $X = Y \neq \varnothing$, we get

$$\hat{f}(X)\hat{f}(Y)\hat{f}(Z) + \widehat{1-f}(X)\widehat{1-f}(Y)\widehat{1-f}(Z) = \hat{f}(X)^2(\hat{f}(\varnothing) + 1 - \hat{f}(\varnothing)) = \hat{f}(X)^2.$$

Therefore the sum simplifies to

$$p = \hat{f}(\varnothing)^3 + (1-\hat{f}(\varnothing))^3 + 3\sum_{X\neq\varnothing} \left(-\frac{1}{3}\right)^{|X|} \hat{f}(X)^2.$$

Finally, the fact that f is antisymmetric implies that it is balanced and so $\hat{f}(\varnothing) = \mathbb{E} f = 1/2$. Substituting this into the formula, we deduce

$$p = \frac{1}{4} + 3\sum_{X\neq\varnothing} \left(-\frac{1}{3}\right)^{|X|} \hat{f}(X)^2. \qquad \square$$

Theorem 9.2 almost immediately follows, using essentially Hoffman's bound.

Proof of Theorem 9.2 when $m = 3$. If $R_f^{(3)}$ is always rational then according to Lemma 9.3,

$$3 \sum_{X \neq \varnothing} \left(-\frac{1}{3}\right)^{|X|} \hat{f}(X)^2 = -\frac{1}{4}.$$

On the other hand,

$$3 \sum_{X \neq \varnothing} \left(-\frac{1}{3}\right)^{|X|} \hat{f}(X)^2 \geq -\sum_{X \neq \varnothing} \hat{f}(X)^2, \tag{9.1}$$

with equality only if the Fourier expansion of f is supported on the first two levels. Since f is balanced, Parseval's identity shows that

$$\sum_{X \neq \varnothing} \hat{f}(X)^2 = \sum_{X \subseteq [n]} \hat{f}(X)^2 - \hat{f}(\varnothing)^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$$

Therefore we must have equality in (9.1), and we deduce that the Fourier expansion of f is supported on the first two levels. Lemma 3.6 implies that f depends on exactly one coordinate. □

By using the Friedgut–Kalai–Naor theorem instead of Lemma 3.6, we obtain a stability version of Theorem 9.2.

Theorem 9.4. *Let f be a binary voting rule. If $R_f^{(3)}$ is rational with probability $1-\epsilon$ then there is a function g depending on one coordinate only such that $\Pr[f = g] = 1 - O(\epsilon)$.*

Proof. According to Lemma 9.3,

$$3 \sum_{X \neq \varnothing} \left(-\frac{1}{3}\right)^{|X|} \hat{f}(X)^2 = -\frac{1}{4} + \epsilon.$$

Equation (9.1) generalizes to

$$3 \sum_{X \neq \varnothing} \left(-\frac{1}{3}\right)^{|X|} \hat{f}(X)^2 \geq -\sum_{|X|=1} \hat{f}(X)^2 - \frac{1}{9} \sum_{|X|>1} \hat{f}(X)^2. \tag{9.2}$$

Let $\delta = \sum_{|X|>1} \hat{f}(X)^2$. As in the proof of Theorem 9.2, equation (9.2) implies that

$$-\frac{1}{4} + \epsilon \geq -\left(\frac{1}{4} - \delta\right) - \frac{1}{9}\delta = -\frac{1}{4} + \frac{8}{9}\delta.$$

In other words, $\delta \leq (9/8)\epsilon$. The theorem now follows directly from Theorem 2.22. □

In order to extend Kalai's proof to the setting where there are more than three candidates, we need to handle higher powers of $\hat{f}(X)$, which seems hard. Falik and Friedgut [31] extended Theorem 9.4 to arbitrary $m \geq 3$ as well as to the setting of the Gibbard–Satterthwaite theorem (in which each voter selects one candidate) by analyzing the voting rule φ directly using the representation theory of S_m. Other results in the same vein (which use Fourier analysis on the Boolean cube) are [43, 54, 67, 61].

9.2 Inapproximability of k-uniform hypergraph vertex cover

Erdős–Ko–Rado type results have been used to prove inapproximability results for vertex cover and its extension to hypergraphs [17, 18, 19, 62, 48], as a means of analyzing the so-called *biased long code*. As an illustration of this method, we explain in full a $(2-\epsilon)$-inapproximability result for 4-uniform hypergraph vertex cover, following [17]. The key combinatorial fact used in this result is the following corollary of the Ahlswede–Khachtrian theorem.

Lemma 9.5. *For every $p < 1/2$ and $\epsilon > 0$ there exists t such that $\mu_p(\mathcal{F}) < \epsilon$ for every t-intersecting family.*

Proof. For $t > 1$, let $r(t)$ be the unique integer r such that

$$\frac{r}{t+2r-1} \leq p < \frac{r+1}{t+2r+1}.$$

The integer r exists since $(r+1)/(t+2r+1) \to 1/2$ as $r \to \infty$. Theorem 5.3 on page 122 shows that a t-intersecting family has μ_p-measure at most

$$\mu_p(\mathcal{F}_{t,r(t)}) = \Pr[\mathrm{Bin}(t+2r(t),p) \geq t+r(t)] \leq \Pr\Big[\mathrm{Bin}(t+2r(t),p) \geq \frac{t+2r(t)}{2}\Big].$$

Since $p < 1/2$, the latter probability tends to 0 as $t+2r(t) \to \infty$, which is certainly the case when $t \to \infty$. Hence for large enough t, $\mu_p(\mathcal{F}_{t,r(t)}) < \epsilon$. □

9.2.1 k-uniform hypergraph vertex cover

A *vertex cover* of a graph G is a set of vertices that touches every edge of G. The problem of *minimum vertex cover* is, given a graph G, determine the minimum size of a vertex cover of G.

The decision version of vertex cover is one of the classical NP-complete problems. There is a simple 2-approximation algorithm for vertex cover, which consists of greedily selecting a maximal matching in the graph, and taking all vertices appearing in the matching. More complicated algorithms achieve a better approximation ratio of $2 - o(1)$, but no known polynomial-time algorithm achieves a $(2 - \epsilon)$-approximation for any constant $\epsilon > 0$. It is therefore believed that the inapproximability threshold of vertex cover is 2, and this is indeed the case assuming the unique games conjecture.

Vertex cover is related to another classical NP-complete problem, *maximum independent set*: the complement of a vertex cover is an independent set (a set of vertices containing no edges), and vice versa. In terms of approximation, however, independent set is much harder, being $n^{1-\epsilon}$-hard to approximate on graphs having n vertices.

We also consider the generalizations of vertex cover and independent set to hypergraphs. A vertex cover of a hypergraph H is a set of vertices that touches every hyperedge of H, and an independent set is a set of vertices I such that no hyperedge of H contains only vertices of I. Under these definitions, it is still true that the complement of a vertex cover is an independent set and vice versa.

We will be interested in *k-uniform hypergraph vertex cover*, which is the problem of finding the minimum size of a vertex cover in a *k-uniform* hypergraph, a hypergraph in which each hyperedge contains exactly k vertices. Classical vertex cover is the case $k = 2$. The greedy algorithm for k-uniform hypergraph vertex cover (a generalization of the algorithm for vertex cover) achieves a k-approximation, which is conjectured to be optimal.

Our exposition will be clearer by considering the slightly more general problem in which vertices have non-negative weights, and the goal is to find a vertex cover of minimal total weight. Inapproximability results for the corresponding problem of weighted k-uniform hypergraph vertex cover easily translate to matching results for the unweighted version, by duplicating vertices according to their weight.

9.2.2 An inapproximability result

In this section we focus on the inapproximability of 4-uniform hypergraph vertex cover, proving the following theorem of Holmerin [52]. We will follow the proof technique of Dinur, Guruswami and Khot [17].

Theorem 9.6. *For every $\epsilon > 0$ it is NP-hard to approximate weighted 4-uniform hypergraph vertex cover to within a factor $2 - \epsilon$.*

The general plan is to reduce from *label cover* (defined below). Given an instance L of label cover, we create an instance H of weighted 4-uniform hypergraph vertex cover with total weight 1 having the following property: if L is a YES instance then H has an independent set of weight $1/2 - \epsilon/6$, while if L is a NO instance then H does not have any independent set of weight $\epsilon/6$ or more. Since the complement of an independent set is a vertex cover, if L is a YES instance then H has a vertex cover of weight $1/2 + \epsilon/6$, while if L is a NO instance then H has no vertex cover of weight $1 - \epsilon/6$ or less. Since $(1 - \epsilon/6)/(1/2 + \epsilon/6) \geq 2 - \epsilon$, an algorithm approximating weighted 4-uniform hypergraph vertex cover to within a factor $2 - \epsilon$ can tell YES instances of label cover from NO instances, which is an NP-hard task.

We will use the following version of label cover, which can be obtained from the PCP theorem through an application of Raz's parallel repetition theorem [70]. A much simpler proof has been obtained recently by Dinur and Steurer [20].

Definition 9.5. An instance of *s-label cover* is given by the following data:

- A finite set of labels Σ of size s.
- Two disjoint sets of variables X, Y.
- For some $x \in X, y \in Y$, constraints $\varphi_{x \to y} : \Sigma \to \Sigma$.

For an instance of label cover, an *assignment* is a function $\alpha : X \cup Y \to \Sigma$. The assignment α *satisfies* a constraint $\varphi_{x \to y}$ if $\varphi_{x \to y}(\alpha(x)) = \alpha(y)$. ○

Theorem 9.7 ([20, Theorem 8.2])**.** *There are some absolute constants $a > 1$ and $\beta \in (0, 1)$ such that for every $k \in \mathbb{N}$ there is a polytime reduction that takes an instance ψ of 3SAT and outputs an instance $L = (\Sigma, X, Y, \varphi)$ of a^k-label cover such that:*

YES: If ψ is satisfiable then there exists an assignment which satisfies all constraints in L.

NO: If ψ is not satisfiable then no assignment satisfies more than β^k of the constraints in L.

Furthermore, the constraint graph *whose vertices are $X \cup Y$ and whose edges are (x, y) whenever $\varphi_{x\to y}$ exists is* bi-regular*: all the X vertices have the same degree, and all the Y vertices have the same degree*[1]*.*

The idea of the reduction is to encode an assignment of X, using consistency hyperedges to enforce the existence of a complementing assignment of Y. The reduction will focus on *consistency triples*.

Definition 9.6. Let $L = (\Sigma, X, Y, \varphi)$ be an instance of label cover. A *consistency triple* (x, x', y) consists of two different $x, x' \in X$ and $y \in Y$ such that both constraints $\varphi_{x\to y}, \varphi_{x'\to y}$ exist. For a subset $X' \subseteq X$, an X'-*consistency triple* is a consistency triple (x, x', y) such that $x, x' \in X'$. ◯

Reduction Let $\epsilon > 0$ be given, and let $k \in \mathbb{N}$ and $p \in (0, 1/2)$ be parameters to be chosen later. Define $s = a^k$ and $\delta = \beta^k$, where a, β are the parameters given by Theorem 9.7. Given an instance $L = (\Sigma, X, Y, \varphi)$ of s-label cover, we construct a weighted 4-uniform hypergraph $H = (V, E, w)$ as follows:

- Vertices: $V = X \times 2^\Sigma$.
- Weights: For $x \in X$ and $F \subseteq \Sigma$, $w(x, F) = \mu_p(F)/|X|$.
- Edges: For every consistency triple (x, x', y) and $F_1, F_2, F_1', F_2' \subseteq \Sigma$, there is an edge $\{(x, F_1), (x, F_2), (x', F_1'), (x', F_2')\}$ whenever

$$\varphi_{x\to y}(F_1 \cap F_2) \cap \varphi_{x'\to y}(F_1' \cap F_2') = \varnothing.$$

The basic idea is that each independent set of H encodes (in some sense) an assignment to X. If L is a YES instance then we can use a satisfying assignment to construct an independent set of large weight. Conversely, we can decode any independent set of non-negligible weight

[1]This can be guaranteed by starting with the NP-complete problem 3SAT-5 in which each variable appears exactly 5 times; see Feige [32].

into an assignment for L satisfying a non-negligible fraction of the constraints. To that end, we will use Lemma 9.5.

We start by explaining the structure of independent sets in H.

Lemma 9.8. *Let $A \subseteq V$ be a subset of the vertices of H. For each $x \in X$, define*

$$A_x = \{F \subseteq \Sigma : (x, F) \in A\}$$

and

$$A_x^{(2)} = \{F_1 \cap F_2 : (x, F_1), (x, F_2) \in A\}.$$

Then $w(A) = \mathbb{E}_{x \in X} \mu_p(A_x)$, and A is independent if and only if for every consistency triple (x, x', y), the families $\varphi_{x \to y}(A_x^{(2)}), \varphi_{x' \to y}(A_{x'}^{(2)})$ are cross-intersecting (every set in the first family intersects every set in the second family).

When L is a YES instance, H has a large independent set in which each A_x is a star.

Lemma 9.9. *If L is a YES instance then H has an independent set of weight p.*

Proof. Let α be a satisfying assignment for L, and define $A = \{(x, F) : \alpha(x) \in F\}$. Using the terminology of Lemma 9.8, it is easy to see that for all $x \in X$, $A_x = A_x^{(2)} = \{F : \alpha(x) \in F\}$ is a star. For every consistency triple (x, x', y), the families $\varphi_{x \to y}(A_x^{(2)}), \varphi_{x' \to y}(A_{x'}^{(2)})$ are cross-intersecting since every set in both families contains $\alpha(y)$. Therefore A in an independent set according to Lemma 9.8. The lemma also implies that $w(A) = \mathbb{E}_{x \in X} \mu_p(A_x) = p$. □

The independent set given by Lemma 9.9 encodes an assignment to X in a very straightforward way. This kind of encoding, in which an element $\sigma \in \Sigma$ is encoded by the family of all sets $F \subseteq \Sigma$ containing it, is known as the *biased long code* (it is biased since a set F is given weight $\mu_p(F)$). In the following part of the proof, we list-decode the biased long code. Given an independent set of non-negligible weight, we identify a non-negligible subset of the vertices, and for each of them construct a short list of possible assignments. We then use these lists to construct an assignment for L satisfying a non-negligible fraction of the constraints.

The first step, list-decoding the biased long code, is accomplished using the following lemma, which makes essential use of Lemma 9.5.

Lemma 9.10. *For every $\tau > 0$ there is a constant $t = t(p,\tau)$ such that the following holds. Suppose that H has an independent set of weight 2τ. There is a subset $X' \subseteq X$ of size $|X'| \geq \tau|X|$, and for each $x \in X'$ a non-empty list $\boldsymbol{\alpha}(x)$ of size $|\boldsymbol{\alpha}(x)| < t$, such that whenever (x,x',y) is an X'-consistency triple, $\varphi_{x\to y}(\boldsymbol{\alpha}(x)) \cap \varphi_{x'\to y}(\boldsymbol{\alpha}(x')) \neq \emptyset$.*

Proof. We let t be the constant given by Lemma 9.5 (with τ replacing ϵ). Let A be an independent set of weight 2τ. Lemma 9.8 shows that $\mathbb{E}_{x\in X}\, \mu_p(A_x) = 2\tau$. Therefore

$$\Pr_{x\in X}[\mu_p(A_x) \geq \tau] = 1 - \Pr_{x\in X}[1-\mu_p(A_x) \geq 1-\tau] \geq 1 - \frac{1-2\tau}{1-\tau} = \frac{\tau}{1-\tau} \geq \tau,$$

where the first inequality is Markov's inequality. We let $X' = \{x \in X : \mu_p(A_x) \geq \tau\}$. Lemma 9.5 shows that for every $x \in X'$, A_x is *not* t-intersecting. Therefore we can choose for each $x \in X'$ some $\boldsymbol{\alpha}(x) \in A_x^{(2)}$ of size $|\boldsymbol{\alpha}(x)| < t$. Since A is an independent set, Lemma 9.8 shows that whenever (x,x',y) is an X'-consistency triple, $\varphi_{x\to y}(\boldsymbol{\alpha}(x)) \cap \varphi_{x'\to y}(\boldsymbol{\alpha}(x')) \neq \emptyset$. This shows that if some X'-consistency triple (x,x',y) exists, $\boldsymbol{\alpha}(x)$ is non-empty. If $\boldsymbol{\alpha}(x)$ is empty, let $\boldsymbol{\alpha}(x) = \{\sigma\}$ for some arbitrary $\sigma \in \Sigma$. □

Given the lists produced by Lemma 9.10, we construct an assignment for L by randomly choosing a value from each of the lists.

Lemma 9.11. *Suppose that for some τ, t there is a subset $X' \subseteq X$ of size $|X'| \geq \tau|X|$, and for each $x \in X'$ a non-empty list $\boldsymbol{\alpha}(x)$ of size $|\boldsymbol{\alpha}(x)| < t$, such that whenever (x,x',y) is an X'-consistency triple, $\varphi_{x\to y}(\boldsymbol{\alpha}(x)) \cap \varphi_{x'\to y}(\boldsymbol{\alpha}(x')) \neq \emptyset$.*

There is an assignment α for L which satisfies at least a fraction τ/t^2 of constraints.

Proof. Let $Y' \subseteq Y$ consist of those $y \in Y$ for which a constraint $\varphi_{x\to y}$ exists for some $x \in X'$. For every $y \in Y'$, we arbitrarily choose some canonical such $\chi_y \in X'$. We define a random assignment α for L as follows. For $x \in X'$, let $\alpha(x)$ be a random element of $\boldsymbol{\alpha}(x)$. For $y \in Y'$, let $\alpha(y)$ be a random element of $\varphi_{\chi_y\to y}(\boldsymbol{\alpha}(\chi_y))$. Define α arbitrarily on the rest of its domain.

We claim that the probability that α satisfies $\varphi_{x\to y}$ is at least $1/t^2$ whenever $x \in X'$. Indeed, if $x = \chi_y$ then given $\alpha(y)$, the probability that $\varphi_{x\to y}(\alpha(x)) = \alpha(y)$ is at least $1/t$. If $x \neq \chi_y$ then we are given that $S = \varphi_{x\to y}(\boldsymbol{\alpha}(x)) \cap \varphi_{\chi_y\to y}(\boldsymbol{\alpha}(\chi_y))$ is non-empty. The probability that $\alpha(y) \in S$

is at least $1/t$, and given that $\alpha(y) \in S$, the probability that $\varphi_{x\to y}(\alpha(x)) = \alpha(y)$ is at least $1/t$. Overall, α satisfies $\varphi_{x\to y}$ with probability at least $1/t^2$.

In expectation, α satisfies $1/t^2$ of the constraints $\varphi_{x\to y}$ in which $x \in X'$. Due to the bi-regularity of the constraint graph (see Definition 9.5), α satisfies τ/t^2 of all constraints in expectation. Therefore there must exist some assignment satisfying at least τ/t^2 of all constraints. □

We are now ready to prove the main result of this section, Theorem 9.6.

Theorem 9.6. *For every $\epsilon > 0$ it is NP-hard to approximate weighted 4-uniform hypergraph vertex cover to within a factor $2 - \epsilon$.*

Proof. Define $p = 1/2 - \epsilon/6$ and $\tau = \epsilon/12$, let $t = t(p,\tau)$ be the constant in Lemma 9.10, and choose k so that $\beta^k < \tau/t^2$. We will show that if there is a $(2 - \epsilon)$-approximation algorithm for weighted 4-uniform hypergraph vertex cover then we can use it to distinguish between YES instances and NO instances of Theorem 9.7, a task which the theorem states is NP-hard.

Given an instance $L = (\Sigma, X, Y, \varphi)$ of a^k-label cover, we can construct the hypergraph H described in this section in polynomial time (since k is constant, $|V| = |X| \cdot 2^{|\Sigma|} = 2^{a^k}|X|$ is linear in the size of L). When L is a YES instance, Lemma 9.9 shows that there is an independent set of weight $1/2 - \epsilon/6$, and so a vertex cover of weight $1/2 + \epsilon/6$ (recall that $w(V) = 1$). Conversely, we claim that if L is a NO instance then it has no vertex cover of weight at most $1 - \epsilon/6$. Indeed, otherwise there would be an independent set of weight $\epsilon/6 = 2\tau$. Lemma 9.10 combined with Lemma 9.11 then implies that there is an assignment for L satisfying a fraction $\tau/t^2 > \beta^k$ of the constraints, and so L cannot be a NO instance, contrary to assumption.

Finally, applying a $(2 - \epsilon)$-approximation algorithm to H would be able to tell the two cases apart since

$$\frac{1 - \epsilon/6}{1/2 + \epsilon/6} = 2\frac{1 - \epsilon/6}{1 + \epsilon/3} \geq 2(1 - \epsilon/6)(1 - \epsilon/3) \geq 2 - \epsilon.$$ □

The inapproximability threshold that we obtain is roughly $1/(1 - p)$. Therefore the best inapproximability threshold is obtained by choosing the largest possible value of p. Here we are limited by the fact that Lemma 9.5 only holds for $p < 1/2$. So in a sense, Lemma 9.5 determines the inapproximability threshold obtained by this method.

9.2.3 Other inapproximability results

The proof in the preceding section can be adapted to give a $(k/2-\epsilon)$-inapproximability result for weighted k-uniform hypergraph vertex cover for any even k. Lemma 9.5 has to be replaced by the following result, which is proved by Dinur, Guruswami, Khot and Regev [18], following Graham, Grötschel and Lovász [45].

Claim 9.12. *For every $p < 1-1/r$, $\epsilon > 0$ and $r \geq 2$ there exists t such that $\mu_p(\mathcal{F}) < \epsilon$ for every r-wise t-intersecting family (a family in which every r sets contain at least t elements in common).*

Dinur, Guruswami, Khot and Regev use this claim along with a different construction to prove a $(k-1-\epsilon)$-inapproximability result for weighted k-uniform hypergraph vertex cover for any $k > 2$. The case $k = 2$ is tackled in Dinur and Safra's classical paper [19], who prove the following result.

Theorem 9.13. *Suppose that $p < 1/2$ satisfies $(1-p)^2 \geq p$, and let $p^\bullet$ be the maximal μ_p-measure of a 2-intersecting family. It is NP-hard to approximate vertex cover to within a factor $(1-p^\bullet)/(1-p)$.*

The value of $p^\bullet$ can be deduced from the Ahlswede–Khachtrian theorem, see Corollary 5.2 and Theorem 5.3 in Chapter 10. The best choice of p is $p = (3-\sqrt{5})/2$, in which case $p^\bullet = 4p^3-3p^4$ and the resulting approximation ratio is $10\sqrt{5}-21 \approx 1.36$.

Optimal inapproximability results have been proven by Khot and Regev [62] assuming the unique games conjecture. Assuming the conjecture, they prove that weighted k-uniform hypergraph vertex cover is NP-hard to approximate to within a factor k. They essentially use the following fact, which can be proved using Katona's circle method: for $p \leq 1-1/k$, a k-wise intersecting family has μ_p-measure at most p.

9.3 Property testing

Suppose we are given a function $f\colon \{0,1\}^n \to \{0,1\}$. Is the function f a dictatorship[2] of the form $f(x) = x_i$? To be certain that f is a dictatorship, we would have to examine all values of f. However, in some circumstances, we are only prepared to examine a small number of values, and in return we are willing to accept some error. Such a situation arises in inapproximability, for example, as we describe below.

What kind of error are we willing to tolerate? One possibility is that our test is correct on most functions f. Another possibility is that our test is randomized, and is correct for *each* function f with high probability. We will be interested in the latter option.

Can we achieve this requirement? Suppose that our test examines only a constant number of values of f, say C of them. Starting with a dictatorship f, construct a new function g by changing m random coordinates. When running our test on g, the probability that it samples any of the changed coordinates is only roughly $Cm/2^n$, and so the test only has a chance to notice the difference between f and g if $m = \Theta(2^n)$. We therefore revise our requirements for the test:

YES: If f is a dictatorship, then the test should always accept.

NO: If f is ϵ-far from every dictatorship (that is, $\Pr[f \neq g] \geq \epsilon$ for every dictatorship g), then the test should reject with probability close to 1.

We won't be able to achieve quite these parameters, but we will come close.

When proving the correctness of the test, in the negative case, we are given that the test succeeds with moderate probability, and want to conclude that f is close to a dictatorship. To that end, we can use the Friedgut–Kalai–Naor theorem.

In the rest of this section, we start by explaining a simpler test, which tests for f being linear. Then we modify this test to test for dictatorships. Next we come up with an even simpler test, whose analysis requires the Friedgut–Kalai–Naor theorem. Finally, we explain how variants of this test are used in PCP theory.

[2] In this section, conforming with common usage in the field, a dictatorship is a function of the form $f(x) = x_i$, whereas in the rest of the thesis, a dictatorship is a function determined by a single coordinate.

9.3.1 Linearity testing

We start by showing how to test that a function $f\colon\{0,1\}^n \to \{0,1\}$ is linear, that is of the form $f(x) = \langle\!\langle x, w\rangle\!\rangle$ for some $w \in \{0,1\}^n$; here $\langle\!\langle x, w\rangle\!\rangle = \sum_{i=1}^n x_i w_i \pmod 2$. The analysis will become simpler if instead of f we consider the related function $F(x) = (-1)^{f(x)}$. If f is linear then F is a Fourier character. Abusing notation, we call F linear as well.

The basic idea is very simple: if f is linear then $f(x \oplus y) = f(x) \oplus f(y)$. In terms of the function F, $F(x \oplus y) = F(x)F(y)$ and so $F(x)F(y)F(x \oplus y) = 1$, and this is our test.

Test L: Given $F\colon\{0,1\}^n \to \{\pm 1\}$, choose random $x, y \in \{0,1\}^n$, and accept if $F(x)F(y)F(x \oplus y) = 1$.

The test always succeeds for linear F. We can express its success probability on general F using the Fourier coefficients of F.

Lemma 9.14. *The test L accepts with probability*

$$\frac{1}{2} + \frac{1}{2}\sum_{U \subseteq [n]} \hat{F}(U)^3.$$

Proof. We can express the acceptance probability as

$$\mathop{\mathbb{E}}_{x,y} \frac{F(x)F(y)F(x \oplus y) + 1}{2} = \frac{1}{2} + \frac{1}{2}\mathop{\mathbb{E}}_{x,y} F(x)F(y)F(x \oplus y).$$

Substituting the Fourier expansion of F, we get

$$\begin{aligned} F(x)F(y)F(x \oplus y) &= \sum_{S,T,U \subseteq [n]} \hat{F}(S)\chi_S(x)\hat{F}(T)\chi_T(y)\hat{F}(U)\chi_U(x \oplus y) \\ &= \sum_{S,T,U \subseteq [n]} \hat{F}(S)\hat{F}(T)\hat{F}(U)\chi_S(x)\chi_T(y)\chi_U(x)\chi_U(y) \\ &= \sum_{S,T,U \subseteq [n]} \hat{F}(S)\hat{F}(T)\hat{F}(U)\chi_{S\Delta U}(x)\chi_{T\Delta U}(y). \end{aligned}$$

Since $\mathbb{E}_x \chi_A(x) = [\![A = \varnothing]\!]$, if we take expectation with respect to x, y then all terms but those in which $S = T = U$ disappear, and so

$$\mathop{\mathbb{E}}_{x,y} F(x)F(y)F(x \oplus y) = \sum_{U \subseteq [n]} \hat{F}(U)^3.$$

The lemma immediately follows. $\square$

If f is linear then $F = \chi_A$ for some $A \subseteq [n]$, and so $\hat{F}(U) = [\![U = A]\!]$, and the test always succeeds. To understand what happens when f is far from linear, we need to calculate the distance between F and the set of linear functions. (Here distance is the fraction of different entries.)

Lemma 9.15. *The distance between F and the set of linear functions is*

$$\frac{1}{2} - \frac{1}{2} \max_{U \subseteq [n]} \hat{F}(U).$$

Therefore F is ϵ-far from being linear (that is, the distance between F and every linear function is at least ϵ) if and only if for all $U \subseteq [n]$,

$$\hat{F}(U) \leq 1 - 2\epsilon.$$

Proof. The distance between F and χ_A is

$$\begin{aligned} d(F, \chi_A) &= \mathop{\mathbb{E}}_x \left(\frac{F(x) - \chi_A(x)}{2} \right)^2 \\ &= \frac{1}{4} \sum_{S \subseteq [n]} (\hat{F}(S) - \widehat{\chi_A}(S))^2 \\ &= \frac{1}{4} \sum_{S \subseteq [n]} \hat{F}(S)^2 + \frac{1}{4} \left[(\hat{F}(A) - 1)^2 - \hat{F}(A)^2 \right] \\ &= \frac{1}{2} - \frac{1}{2} \hat{F}(A), \end{aligned}$$

using Parseval's identity twice: in the second equality, and in the last one to conclude that $\sum_{S \subseteq [n]} \hat{F}(S)^2 = \mathbb{E} F^2 = 1$. □

As a conclusion, we can analyze test L.

Theorem 9.16. *Let $F \colon \{0,1\}^n \to \{0,1\}$.*

1. *If F is linear then L always accepts.*
2. *If F is ϵ-far from being linear then L accepts with probability at most $1 - \epsilon$.*

Proof. The first part follows directly from Lemma 9.14. For the second part, Lemma 9.15 shows that $\hat{F}(U) \leq 1 - 2\epsilon$, and so according to Lemma 9.14, L succeeds with probability

$$\frac{1}{2} + \frac{1}{2} \sum_{U \subseteq [n]} \hat{F}(U)^3 \leq \frac{1}{2} + \frac{1 - 2\epsilon}{2} \sum_{U \subseteq [n]} \hat{F}(U)^2 = \frac{1}{2} + \frac{1 - 2\epsilon}{2} = 1 - \epsilon,$$

using Parseval's identity. □

We can amplify the success probability by repeating the test and taking a majority vote. In this way, for every ϵ, δ we can devise a test with a constant number of queries that always accepts linear F, and accepts functions ϵ-far from being linear with probability at most δ.

9.3.2 Dictatorship testing

The test L developed in the preceding section behaves in the same way for every linear function. In this section we modify it so that it highlights linear functions of the form $F = \chi_{\{i\}}$. The idea is that such functions are resilient to changing all other coordinates. Instead of testing F at points $x, y, x \oplus y$, we will test it at points x, y, z where z is obtained from $x \oplus y$ by randomly modifying some of its coordinates.

Test $D(p)$: Given $F: \{0,1\}^n \to \{\pm 1\}$, choose random $x, y \in \{0,1\}^n$. Let $z \in \{0,1\}^n$ be defined by $z_i = x_i \oplus y_i$ with probability $1 - p$ and $z_i = x_i \oplus y_i \oplus 1$ with probability p (independently for each i). Accept if $F(x)F(y)F(z) = 1$.

This time, the test succeeds for dictatorships only with probability $1 - p$. Again, we can express its success probability in terms of the Fourier coefficients of F.

Lemma 9.17. *The test $D(p)$ accepts with probability*

$$\frac{1}{2} + \frac{1}{2} \sum_{U \subseteq [n]} (1 - 2p)^{|U|} \hat{F}(U)^3.$$

Proof. We can express the acceptance probability as

$$\mathop{\mathbb{E}}_{x,y,z} \frac{F(x)F(y)F(z) + 1}{2} = \frac{1}{2} + \frac{1}{2} \mathop{\mathbb{E}}_{x,y,z} F(x)F(y)F(z).$$

Let $z = x \oplus y \oplus w$, where $w_i = 1$ with probability p. Substituting the Fourier expansion of F, we get

$$
\begin{aligned}
F(x)F(y)F(z) &= F(x)F(y)F(x \oplus y \oplus w) \\
&= \sum_{S,T,U \subseteq [n]} \hat{F}(S)\chi_S(x)\hat{F}(T)\chi_T(y)\hat{F}(U)\chi_U(x \oplus y \oplus w) \\
&= \sum_{S,T,U \subseteq [n]} \hat{F}(S)\hat{F}(T)\hat{F}(U)\chi_S(x)\chi_T(y)\chi_U(x)\chi_U(y)\chi_U(w) \\
&= \sum_{S,T,U \subseteq [n]} \hat{F}(S)\hat{F}(T)\hat{F}(U)\chi_{S\Delta U}(x)\chi_{T\Delta U}(y)\chi_U(w).
\end{aligned}
$$

Since $\mathbb{E}_x \chi_A(x) = [\![A = \varnothing]\!]$, we deduce that

$$\mathop{\mathbb{E}}_{x,y,z} F(x)F(y)F(z) = \sum_{U \subseteq [n]} \hat{F}(U)^3 \mathop{\mathbb{E}}_w \chi_U(w).$$

It remains to calculate $\mathbb{E}_w \chi_U(w)$:

$$\mathop{\mathbb{E}}_w \chi_U(w) = \prod_{i \notin U} 1 \prod_{i \in U} \mathop{\mathbb{E}}_{w_i} (-1)^{w_i} = \prod_{i \in U} [p(-1) + (1-p)(1)] = (1-2p)^{|U|}.$$

The lemma immediately follows. □

There is one problem with this test: it succeeds not only for dictatorships, but also for the function $F = 1$. To handle this situation, we assume that the function F is *odd*, an assumption which is justified in certain circumstances.

Definition 9.7. A function $F\colon \{0,1\}^n \to \{0,1\}$ is *odd* if

$$F(x) = -F(1-x),$$

where $1-x$ is the vector defined by $(1-x)_i = 1 - x_i$. ○

Dictatorships, for example, are odd. The Fourier expansion of odd functions is supported on odd-sized Fourier coefficients.

Lemma 9.18. *If $F\colon \{0,1\}^n \to \{0,1\}$ is odd then $\hat{F}(A) = 0$ whenever $|A|$ is even.*

Proof. We have

$$
\begin{aligned}
\hat{F}(A) = \langle F, \chi_A \rangle = \mathop{\mathbb{E}}_x F(x)\chi_A(x) &= \mathop{\mathbb{E}}_x \left(\frac{F(x) - F(1-x)}{2} \right) \chi_A(x) \\
&= \frac{1}{2} \mathop{\mathbb{E}}_x F(x)\chi_A(x) - \frac{1}{2} \mathop{\mathbb{E}}_x F(x)\chi_A(1-x) = 0,
\end{aligned}
$$

since $\chi_A(1-x) = \chi_A(x)$ due to $|A|$ being even. □

We can modify the proof of Lemma 9.15 to obtain a similar result for dictatorships.

Lemma 9.19. *The distance between F and the set of dictatorships is*

$$\frac{1}{2} - \frac{1}{2}\max_{i\in[n]} \hat{F}(\{i\}).$$

Therefore F is ϵ-far from being a dictatorship if and only if for all $i \in [n]$,

$$\hat{F}(\{i\}) \leq 1 - 2\epsilon.$$

Putting everything together, we can analyze the test $D(p)$ for odd functions.

Theorem 9.20. *Let $F\colon \{0,1\}^n \to \{0,1\}$ be an odd function.*

1. *If F is a dictatorship then $D(p)$ accepts with probability $1-p$.*
2. *If F is ϵ-far from being a dictatorship then $D(p)$ accepts with probability at most $1-p-(1-2p)\epsilon$, assuming $\epsilon \leq 2p(1-p)$.*

Proof. If F is a dictatorship, say $F = \chi_{\{i\}}$, then Lemma 9.17 shows that $D(p)$ succeeds with probability

$$\frac{1}{2} + \frac{1}{2}\sum_{U\subseteq[n]} (1-2p)^{|U|}\hat{F}(U)^3 = \frac{1}{2} + \frac{1}{2}(1-2p) = 1-p,$$

since the only non-zero Fourier coefficient of F is $\hat{F}(\{i\}) = 1$.

For the second part, Lemma 9.18 shows that $\hat{F}(U) \neq 0$ only when $|U|$ is odd. Lemma 9.19 shows that $\hat{F}(\{i\}) \leq 1-2\epsilon$ for all $i \in [n]$. The success probability of $D(p)$, given by Lemma 9.17, is

$$q = \frac{1}{2} + \frac{1}{2}\sum_{U\subseteq[n]} (1-2p)^{|U|}\hat{F}(U)^3.$$

If $|U| = 1$ then $(1-2p)^{|U|}\hat{F}(U) \leq (1-2p)(1-2\epsilon)$, and otherwise $|U| \geq 3$ and so $(1-2p)^{|U|}\hat{F}(U) \leq (1-2p)^3$. When $\epsilon \leq 2p(1-p)$, $(1-2p)^2 \leq 1-2\epsilon$, and so

$$q \leq \frac{1}{2} + \frac{1}{2}\sum_{U\subseteq[n]} (1-2p)(1-2\epsilon)\hat{F}(U)^2 = \frac{1}{2} + \frac{1}{2}(1-2p)(1-2\epsilon) = 1-p-(1-2p)\epsilon,$$

using Parseval's identity. □

As in the case of test L, for any ϵ, δ we can repeat the test enough times to boost the probabilities to $1-\delta$ in the positive case and δ in the negative case.

9.3.3 Dictatorship testing using two queries

Test $D(p)$ uses three queries. Can we get any meaningful test with only two? The idea is to get rid entirely of the linearity apparatus, and concentrate on the noise stability of F.

> Test $D_2(p)$: Given $F\colon\{0,1\}^n \to \{\pm 1\}$, choose random $x \in \{0,1\}^n$. Let $z \in \{0,1\}^n$ be defined by $z_i = x_i$ with probability $1-p$ and $z_i = x_i \oplus 1$ with probability p. Accept if $F(x)F(z) = 1$.

Once again, the test succeeds for dictatorships only with probability $1-p$. It also works for anti-dictatorships, which are functions of the form $F(x) = -(-1)^{x_i}$. We can express its success probability in terms of the Fourier coefficients of F.

Lemma 9.21. *The test $D_2(p)$ accepts with probability*

$$\frac{1}{2} + \frac{1}{2}\sum_{U\subseteq[n]} (1-2p)^{|U|}\hat{F}(U)^2.$$

Proof. We can express the acceptance probability as

$$\mathop{\mathbb{E}}_{x,z} \frac{F(x)F(z)+1}{2} = \frac{1}{2} + \frac{1}{2}\mathop{\mathbb{E}}_{x,z} F(x)F(z).$$

Let $z = x \oplus w$, where $w_i = 1$ with probability p. Substituting the Fourier expansion of F, we get

$$\begin{aligned} F(x)F(z) &= F(x)F(x\oplus w) \\ &= \sum_{S,U\subseteq[n]} \hat{F}(S)\chi_S(x)\hat{F}(U)\chi_U(x\oplus w) \\ &= \sum_{S,U\subseteq[n]} \hat{F}(S)\hat{F}(U)\chi_{S\Delta U}(x)\chi_U(w). \end{aligned}$$

Since $\mathbb{E}_x \chi_A(x) = [\![A = \varnothing]\!]$, we deduce that

$$\mathop{\mathbb{E}}_{x,z} F(x)F(z) = \sum_{U\subseteq[n]} \hat{F}(U)^2 \mathop{\mathbb{E}}_{w}\chi_U(w).$$

In the proof of Lemma 9.17, we calculated $\mathbb{E}_w \chi_U(w) = (1-2p)^{|U|}$, implying the statement of the lemma. □

When analyzing $D(p)$ in the case that F is ϵ-far from dictatorships, we used the inequality $\hat{F}(\{i\}) \le 1-2\epsilon$ to estimate $\hat{F}(\{i\})^3 \le (1-2\epsilon)\hat{F}(\{i\})^2$, and then applied Parseval's inequality. The matching expression for $D_2(p)$ is $\hat{F}(\{i\})^2$, and so this approach fails. Instead, we use the Friedgut–Kalai–Naor theorem.

Theorem 9.22. *Let $F\colon \{0,1\}^n \to \{0,1\}$ be an odd function.*

1. *If F is a dictatorship or an anti-dictatorship then $D_2(p)$ accepts with probability $1-p$.*
2. *If F is ϵ-far from being a dictatorship or an anti-dictatorship then $D_2(p)$ accepts with probability at most $1-p-\Omega(p(1-2p)\epsilon)$, assuming $p,\epsilon < 1/2$.*

Proof. If F is a dictatorship or an anti-dictatorship, then it is easy to check that $D_2(p)$ accepts with probability $1-p$. For the other direction, suppose that $D_2(p)$ accepts with probability q. According to Lemma 9.21,

$$q = \frac{1}{2} + \frac{1}{2}\sum_{U\subseteq[n]} (1-2p)^{|U|}\hat{F}(U)^2.$$

Let $\gamma = \sum_{|U|>1} \hat{F}(U)^2$. Lemma 9.18 shows that $\hat{F}(U)=0$ for even $|U|$, and so Parseval's identity implies

$$q \le \frac{1}{2} + \frac{1}{2}(1-2p)(1-\gamma) + \frac{1}{2}(1-2p)^3\gamma = 1-p-2p(1-p)(1-2p)\gamma.$$

We conclude that $\gamma = (1-p-q)/(2p(1-p)(1-2p))$. The Friedgut–Kalai–Naor theorem shows that F is $C\gamma$-close to some function G depending on at most one coordinate, for some universal constant C. If G is constant then since F is odd, $C\gamma \ge 1/2$. Otherwise, by assumption $C\gamma \ge \epsilon$. Assuming $\epsilon \le 1/2$, in both cases $C\gamma \ge \epsilon$, implying $1-p-q \ge 2p(1-p)(1-2p)\epsilon/C$. □

9.3.4 Applications to inapproximability

In what sense is a test using two queries better than a test using three? The difference can be essential in applications to inapproximability. In order to prove that a certain problem P is inapproximable within some ratio R, we reduce from some version of the PCP theorem, say Theorem 9.7. Given an instance I of label cover, we construct some instance $f(I)$ of P, with the following property: if I is a YES instance, then $f(I)$ has value V, while if I is a NO instance, then $f(I)$ has value VR. When P is a problem like MAX-3SAT, we can think of $f(I)$

as a test for I being a YES instance. Each clause of $f(I)$ is some predicate that depends on three variables, which could be (for example) three values of a Boolean function F related to the original instance I. Therefore a test querying three locations corresponds to MAX-3SAT or to the more general MAX-3CSP (where instead of clauses we can use arbitrary Boolean functions on three variables), and a test querying two locations corresponds to MAX-2SAT or MAX-2CSP.

In the usual context in which dictatorship testing is used, YES instances are dictatorships, but NO instances, rather than being ϵ-far from being a dictatorship, satisfy other quasirandomness properties (see for example Håstad [75]). However, in the context of assignment testers, used for example in Dinur's proof of the PCP theorem [15], NO instances are ϵ-far from a certain collection of functions, and a test similar to the three-query dictatorship test is used. Because of the specific nature of the test, the analysis requires the Friedgut–Kalai–Naor theorem.

Moreover, in the context of inapproximability, we can enforce the restriction that F is odd using an appropriate encoding: F is defined explicitly only for inputs x for which $x_1 = 0$, say, and $F(x) = -F(1-x)$ for inputs on which F is not explicitly defined. (A similar device is used when encoding a symmetric matrix using its upper triangular part only.)

Chapter 10

Open problems

10.1 The Ahlswede–Khachatrian theorem

In Chapter 5 we proved the Ahlswede–Khachatrian theorem, using the technique of shifting.

Theorem 5.3. *Let $\mathcal{F}$ be a t-intersecting family on n points for $t \geq 2$. If $r/(t+2r-1) < p < (r+1)/(t+2r+1)$ for some $r \geq 0$ then $\mu_p(\mathcal{F}) \leq \mu_p(\mathcal{F}_{t,r})$, with equality if and only if $\mathcal{F}$ is equivalent to $U^n(\mathcal{F}_{t,r})$.*

If $p = (r+1)/(t+2r+1)$ for some $r \geq 0$ then $\mu_p(\mathcal{F}) \leq \mu_p(\mathcal{F}_{t,r}) = \mu_p(\mathcal{F}_{t,r+1})$, with equality if and only if $\mathcal{F}$ is equivalent to either $U^n(\mathcal{F}_{t,r})$ or $U^n(\mathcal{F}_{t,r+1})$.

(Recall that for a family of sets $\mathcal{F}$ on m points, $U^n(\mathcal{F}) = \{A \subseteq [n] : A \cap [m] \in \mathcal{F}\}$.)

In Chapter 3 we proved the special case $r = 0$ of this theorem using Friedgut's method. The benefit of this method is that it automatically yields stability. However, the same method doesn't work for $r \geq 1$, since the analog of Lemma 3.13 on page 39 is false. That is, there is no polynomial of degree at most $t-1$ which has the correct eigenvalues.

For example, consider the simplest case $t = 2$, $r = 1$. The Ahlswede–Khachatrian theorem shows that for p in the range $1/3 \leq p \leq 2/5$, the maximal μ_p-measure of a 2-intersecting family is $m_{AK}(p) = 4p^3 - 3p^4$. On the other hand, for Friedgut's method to give an upper bound of m, we need to find a linear polynomial P such that $P(0) = 1$ and $(-p/(1-p))^s P(s) \geq -m/(1-m)$ for all $s \geq 1$ (cf. Lemma 3.13).

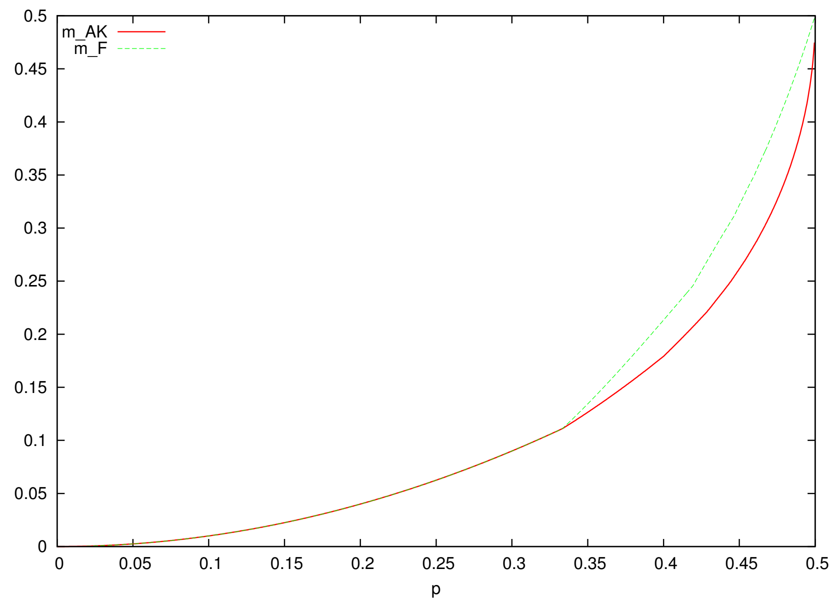


Figure 10.1: Optimal bound m_{AK} versus bound given by Friedgut's method m_F.

Letting $P(s) = as + 1$, finding the best polynomial P for given p reduces to the following (infinite) linear program:

$$\min \frac{m}{1-m}$$

s.t.

$$\left(-\frac{p}{1-p}\right)^s (as+1) \geq -\frac{m}{1-m} \qquad \text{for all } s \geq 1$$

The solution to this program (in the range $1/3 \leq p \leq 3/(5+\sqrt[3]{10}) \approx 0.42$) is $m = m_F(p) = 3p^4/(1-4p+6p^2)$ rather than $m_{AK}(p)$ (the tight constraints are for $s = 1$ and $s = 4$). Figure 10.1 plots the optimal bound m_{AK} against the bound m_F achieved by Friedgut's method.

The reason that Friedgut's method fails for $p > 1/(t+1)$ is that there exists some non-Boolean function f which satisfies all the admissible constraints $f'B_0f = f'B_1f = 0$ given by Lemma 3.11 but has measure $m_F(p) > m_{AK}(p)$. An important open question is modifying the method to prove the general case $r \geq 1$.

k-wise t-intersecting families. A related open question concerns generalizing the Ahlswede–Khachatrian theorem to k-wise t-intersecting families, which are families in which every k sets intersect in at least t elements (the usual case corresponds to $k = 2$). It has been conjectured that the Ahlswede–Khachatrian theorem generalizes as follows.

Definition 10.1. The (k,t,r) Frankl family $\mathcal{F}_{k,t,r}$ is the k-wise t-intersecting family defined by

$$\mathcal{F}_{k,t,r} = \{S \subseteq [t+kr] : |S| \geq t + (k-1)r\}.$$

Conjecture 10.1. *If $\mathcal{F}$ is k-wise t-intersecting then for $p < 1/2$,*

$$\mu_p(\mathcal{F}) \leq \sup_{r \geq 0} \mu_p(\mathcal{F}_{k,t,r}).$$

Furthermore, the supremum is attained for at most two values of r, and equality is possible only if $\mathcal{F}$ is equivalent to the corresponding families.

A proof of this conjecture would yield a more direct proof of Claim 9.12 on page 231, in the same way that the easier Lemma 9.5 follows from the Ahlswede–Khachatrian theorem (see page 224).

Cross-intersecting families. A related question regards cross-t-intersecting families in the regime $p \leq 1/(t+1)$. In Chapter 3, we were only able to prove that $\mu_p(\mathcal{F})\mu_p(\mathcal{G}) \leq p^{2t}$ for cross-t-intersecting families $\mathcal{F}, \mathcal{G}$ when $p \leq 1 - 2^{-1/t}$. For $t \geq 2$, $1 - 2^{-1/t}$ is smaller than $1/(t+1)$. Can we close this gap?

10.2 Graphical intersection problems

Generalization to $p > 1/2$. In Chapter 4 we proved that if $\mathcal{F}$ is an odd-cycle-intersecting family of graphs then $\mu_p(\mathcal{F}) \leq p^3$ for all $p \leq 1/2$. As we commented in Section 4.6, our proof breaks for $p > 1/2$. However, we expect the theorem to hold up to $p = 3/4$. When $p > 3/4$, the μ_p-measure of the family of all graphs containing more than $\frac{3}{4}\binom{n}{2} + 4n$ edges tends to 1, and the intersection of any two such graphs contains more than $n^2/4$ edges, and so some triangle (by Mantel's theorem). When $p = 3/4$, the μ_p-measure of the family tends to some minuscule constant which is much smaller than p^3.

Cycle-intersecting families. Another problem left open by our approach is that of cycle-intersecting families. We expect cycle-intersecting families to have measure at most p^3 for $p \le 1/2$. For $p > 1/2$, the μ_p-measure of all graphs containing at least $\frac{1}{2}\binom{n}{2} + n/2$ edges tends to 1, and the intersection of any two such graphs contains at least n edges, and so a cycle. This problem appears much harder than the one we solved in Chapter 4. The reason is that our proof relies heavily on random bipartite graphs, which are very dense. In contrast, cycle-free graphs (in other words, forests) are very sparse, and that makes controlling the eigenvalues of cycle-admissible matrices much harder, as we explain below.

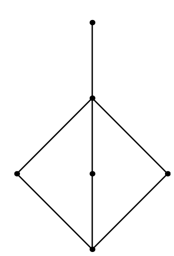
In Chapter 4 we construct an operator A whose eigenvalues $\lambda_G(A)$ are given by linear combinations of functions of the form $q_k(G) = \Pr_H[|G \cap H| = k]$, where H is a random bipartite graph. For concreteness, let us consider $k = 1$. On the one hand $q_1(-) = 1/2$, where $-$ is the graph with a single edge, and on the other hand $q_1(G) \le 1$ for all G. The reason that $q_1(-) = 1/2$ is that the expected density of H is $1/2$.

Consider what happens when instead H is a random forest. A random forest has density at most $2/n$, and so $q_1(-) \le 2/n$. On the other hand, $q_1(K_n) = 1$. This large dynamic range makes it difficult to ensure that $\lambda_-(A) = -p^3/(1-p^3)$ while at the same time $\lambda_G(A) \ge -p^3/(1-p^3)$ for all dense graphs G (consider what coefficient $q_1(G)$ should get in the expression for $\lambda_G(A)$).

H-intersecting families for general H. In general, for every given graph H, one can ask what is the maximal size of an H-intersecting family. It is tempting to conjecture that the maximal family always has μ-measure 2^{-H}, but that is false even for P_3, the path of length 3, as shown by Christofides [8].

Lemma 10.2. *There is a P_3-intersecting family of μ-measure* 17/128.

Proof. Consider the following graph G with 7 edges:



The graph can be decomposed into three disjoint copies of P_2 and an extra edge (the top edge in the diagram). Define a family $\mathcal{F}$ of subgraphs of G which consists of the following 17 graphs: G (1 graph), G minus an edge (7 graphs), G minus the extra edge and any other edge (6 graphs), G minus any of the three copies of P_2 (3 graphs). The intersection of any two graphs in $\mathcal{F}$ always contains at least one copy of P_2 and at least one additional edge, which together form a copy of P_3. Since G has 7 edges, the μ-measure of $\mathcal{F}$ is $17/2^7 = 17/128$. □

We do not know what is the largest μ-measure of a P_3-intersecting family, nor do we have any conjecture. We conjecture, however, that the maximal K_k-intersecting families are K_k-stars.

Cross-intersecting families. As in the case of cross-t-intersecting families, we were not able to prove the cross-intersecting version of our result on odd-circuit-intersecting families. The problem is that the inequality $|\lambda_G| \le p^3/(1-p^3)$ is violated for forests of three edges. We expect the cross-intersecting version of the theorem to hold for all $p \le 1/2$ and perhaps even for all $p \le 3/4$.

10.3 Stability theorems for Boolean functions on S_n

Sharper results. In Chapter 7 and Chapter 8, we proved two stability theorems of the following form: if $\mathcal{F}$ is a family of size $c(n-1)!$ (where $c \le n/2$) whose characteristic function f is close to its projection f_1 to $L_{(n-1,1)}$, then $\mathcal{F}$ is close to a family $\mathcal{G}$ which is the union of $\lceil c \rfloor$ cosets (the cosets can be assumed to be disjoint in the case of Chapter 8). Succinctly put, Theorem 7.1 states that

$$\mathbb{E}[(f - f_1)^2] = \epsilon\frac{c}{n} \Longrightarrow \frac{|\mathcal{F}\Delta\mathcal{G}|}{n!} = O\left(\left(\epsilon^{1/2} + \frac{1}{n}\right)c\right)\frac{c}{n},$$

and Theorem 8.1 states that

$$\mathbb{E}[(f - f_1)^2] = \epsilon \Longrightarrow \frac{|\mathcal{F}\Delta\mathcal{G}|}{n!} = O\left(\frac{n}{c}\left(\epsilon^{1/7} + \frac{1}{n^{1/3}}\right)\right).$$

In order to get an approximation error which is $o(c(n-1)!)$, we need $c = o(n)$ in the first theorem, and $c = \omega(n^{5/6})$ in the second theorem. The theorems thus complement each other.

Together with David Ellis and Ehud Friedgut, we conjecture a sharper result.

Conjecture 10.3. *Let $\mathcal{F} \subseteq S_n$ be a family of permutations of size $c(n-1)!$. Let $f = \mathbf{1}_{\mathcal{F}}$ (so $\mathbb{E}[f] = c/n$) and let $f_1 = \hat{f}((n)) + \hat{f}((n-1,1))$ be the projection of f to $L_{(n-1,1)}$.*

If $\mathbb{E}[(f-f_1)^2] = \epsilon\|f\|^2$, then there exists a family $\mathcal{G} \subseteq S_n$ which is the union of $\lceil c \rfloor$ cosets satisfying

$$|\mathcal{F}\Delta\mathcal{G}| = O(\epsilon|\mathcal{F}|).$$

Generalizations. Together with David Ellis and Ehud Friedgut, we proved a generalization of Theorem 7.1 to families close to $L_{(n-t,1^t)}$ [25]. As in the case of Theorem 7.1, this generalization is only meaningful when $c = o(n)$, where the family in question has size $c(n-t)!$. We conjecture that Conjecture 10.3 extends to this case.

Bibliography

[1] Rudolf Ahlswede and Levon H. Khachatrian. The complete nontrivial-intersection theorem for systems of finite sets. *J. Comb. Theory A*, 76(1):121–138, 1996.

[2] Rudolf Ahlswede and Levon H. Khachatrian. The complete intersection theorem for systems of finite sets. *Eur. J. Comb.*, 18(2):125–136, 1997.

[3] Rudolf Ahlswede and Levon H. Khachatrian. A pushing-pulling method: New proofs of intersection theorems. *Combinatorica*, 19(1):1–15, 1999.

[4] Noga Alon, Irit Dinur, Ehud Friedgut, and Benny Sudakov. Graph products, Fourier analysis and spectral techniques. *Geom. Func. Analysis*, 14(5):913–940, 2004.

[5] Noga Alon, Haim Kaplan, Michael Krivelevich, Dalia Malkhi, and Julien Stern. Scalable secure storage when half the system is faulty. In *ICALP*, pages 576–587, 2000.

[6] K. Arrow. A difficulty in the theory of social welfare. *J. of Political Economy*, 58:328–346, 1950.

[7] Peter J. Cameron and C.Y. Ku. Intersecting families of permutations. *Euro. J. Comb.*, 24:881–890, 2003.

[8] Demetres Christofides. A counterexample to a conjecture of Simonovits and Sós. *submitted*.

[9] Fan Chung and Ross M. Richardson. Weighted laplacians and the sigma functions of a graph. *Contemp. Math.*, 415:93–107, 2006.

[10] Fan R. K. Chung, Ronald L. Graham, Peter Frankl, and James B. Shearer. Some intersection theorems for ordered sets and graphs. *J. Comb. Theory, Ser. A*, 43(1):23–37, 1986.

[11] Vašek Chvátal, David A. Klarner, and Donald E. Knuth. Selected combinatorial research problems. Technical Report CS-72-292, Stanford University, June 1972.

[12] Philippe Delsarte. An algebraic approach to the association schemes of coding theory. *Philips Res. Reps. Suppl.*, 10, 1973.

[13] Michel Deza and Péter Frankl. On the maximum number of permutations with given maximal or minimal distance. *J. Comb. Theory A*, 22:352–360, 1977.

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

[15] Irit Dinur. The PCP theorem by gap amplification. *J. ACM*, 54(3):12, 2007.

[16] Irit Dinur, Ehud Friedgut, and Oded Regev. Independent sets in graph powers are almost contained in juntas. *Geom. and Func. Analysis*, 18(1):77–97, 2008.

[17] Irit Dinur, Venkatesan Guruswami, and Subhash Khot. Vertex cover on k-uniform hypergraphs is hard to approximate within factor $(k - 3 - \epsilon)$. *ECCC*, 9:27, 2002.

[18] Irit Dinur, Venkatesan Guruswami, Subhash Khot, and Oded Regev. A new multilayered PCP and the hardness of hypergraph vertex cover. *SIAM J. Comput.*, 34(5):1129–1146, 2005.

[19] Irit Dinur and Shmuel Safra. On the hardness of approximating minimum vertex cover. *Ann. Math.*, 162(1):439–485, 2005.

[20] Irit Dinur and David Steurer. Analytical approach to parallel repetition. *ArXiv e-prints*, 2013.

[21] David Ellis. Stability for t-intersecting families of permutations. *J. Comb. Theory A*, 118(1):208–227, 2011.

[22] David Ellis. A proof of the Cameron-Ku conjecture. *J. London Math. Soc.*, 85:165–190, 2012.

[23] David Ellis. Setwise intersecting families of permutations. *J. Comb. Theory A*, 119:825–849, 2012.

[24] David Ellis, Yuval Filmus, and Ehud Friedgut. A quasi-stability result for dictatorships in S_n. *Submitted*.

[25] David Ellis, Yuval Filmus, and Ehud Friedgut. A quasi-stability result for low-degree boolean functions on S_n. *Manuscript*.

[26] David Ellis, Yuval Filmus, and Ehud Friedgut. A stability result for balanced dictatorships in S_n. *Submitted*.

[27] David Ellis, Yuval Filmus, and Ehud Friedgut. Triangle-intersecting families of graphs. *J. Eur. Math. Soc.*, 14(3):841–885, 2012.

[28] David Ellis, Ehud Friedgut, and Haran Pilpel. Intersecting families of permutations. *J. Am. Math. Soc.*, 24(3):649–682, 2011.

[29] Paul Erdős. My joint work with Richard Rado. In *Proc. 11th British Combinat. Conf., Cambridge*, pages 53–80. University Press, 1987.

[30] Paul Erdős, Chao Ko, and Richard Rado. Intersection theorems for systems of finite sets. *Quart. J. Math. Oxford (2)*, 12:313–320, 1961.

[31] Dvir Falik and Ehud Friedgut. Between Arrow and Gibbard-Satterthwaite: a representation-theoretic approach. *Submitted*.

[32] Uriel Feige. A threshold of ln n for approximating set cover. *J. ACM*, 45:634–652, July 1998.

[33] Yuval Filmus. Triangle-intersecting families on eight vertices. *ArXiv e-prints*, February 2011.

[34] Péter Frankl. On Sperner families satisfying an additional condition. *J. Comb. Theory A*, 20:1–11, 1976.

[35] Péter Frankl. Erdős-Ko-Rado theorem with conditions on the maximal degree. *J. Comp. Theory A*, 46:252–263, 1987.

[36] Péter Frankl and Zoltán Füredi. A new short proof of the EKR theorem. *J. Comb. Theory A*, 119(6):1388–1390, 2012.

[37] Péter Frankl and Ron L. Graham. Old and new proofs of the Erdős-Ko-Rado theorem. *J. of Sichuan Univ. Natural Science Edition*, 26:112–122, 1989.

[38] Péter Frankl and Norihide Tokushige. Weighted multiply intersecting families. *Studia Sci. Math. Hungar.*, 40(3):287–291, 2003.

[39] Ehud Friedgut. Hypergraphs, entropy and inequalities. *Amer. Math. Month.*, 111(9):749–760, 2004.

[40] Ehud Friedgut. On the measure of intersecting families, uniqueness and stability. *Combinatorica*, 28(5):503–528, 2008.

[41] Ehud Friedgut. Personal communication, 2009.

[42] Ehud Friedgut, Gil Kalai, and Assaf Naor. Boolean functions whose Fourier transform is concentrated on the first two levels. *Adv. App. Math.*, 29(3):427–437, 2002.

[43] Ehud Friedgut, Gil Kalai, and Noam Nisan. Elections can be manipulated often. In *FOCS*, pages 243–249, 2008.

[44] Zoltán Füredi, Kyung-Won Hwang, and Paul M. Weichsel. A proof and generalizations of the Erdős-Ko-Rado theorem using the method of linearly independent polynomials. *J. Combin. Theory Ser. A*, 53(2):306–315, 1990.

[45] Ronald L. Graham, Martin Grötschel, and László Lovász, editors. *Handbook of combinatorics*. Elsevier Science B.V., Amsterdam, 1995.

[46] D. Greenwell and László Lovász. Applications of product colorings. *Acta Math. Acad. Sci. Hungar.*, 25(3–4):335–340, 1974.

[47] Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric Algorithms and Combinatorial Optimization*, volume 2 of *Algorithms and Combinatorics*. Springer, 1988.

[48] Venkatesan Guruswami, Sushant Sachdeva, and Rishi Saket. Inapproximability of minimum vertex cover on k-uniform k-partite hypergraphs. *ECCC*, 20:71, 2013.

[49] Anthony J. W. Hilton and Eric C. Milner. Some intersection theorems for systems of finite sets. *Quart. J. Math. Oxford (2)*, 18:369–384, 1967.

[50] Alan J. Hoffman. On eigenvalues and colorings of graphs, I. In *Graph Theory and its Applications*, Proc. Advanced Sem., Math. Research Center, Univ. of Wisconsin, Madison, Wis., pages 79–91. Academic Press, New York, 1969.

[51] Alan J. Hoffman and Leonard Howes. On eigenvalues and colorings of graphs, II. *Ann. N. Y. Acad. Sci.*, pages 238–242, 1970.

[52] Jonas Holmerin. Vertex cover on 4-regular hypergraphs is hard to approximate within $2-\epsilon$. *ECCC*, 8:94, 2001.

[53] Ralph Howard, Gyula Károlyi, and László A. Székely. Towards a Katona type proof for the 2-intersecting Erdős-Ko-Rado theorem. *Electr. J. Comb.*, 8(1), 2001.

[54] Marcus Isaksson, Guy Kindler, and Elchanan Mossel. The geometry of manipulation — A quantitative proof of the Gibbard-Satterthwaite theorem. *Combinatorica*, 32(2):221–250, 2012.

[55] Gordon James and Adalbert Kerber. *The Representation Theory of the Symmetric Group*. Encyclopedia of Mathematics and its Applications. Addison-Wesley, 1981.

[56] Gordon James and Martin Liebeck. *Representations and Characters of Groups.* Cambridge University Press, 1993.

[57] Gil Kalai. A fourier-theoretic perspective for the Condorcet paradox and Arrow's theorem. *Adv. in Appl. Math.*, 29:412–426, 2002.

[58] Gyula O. H. Katona. A simple proof of the Erdős-Chao Ko-Rado theorem. *J. Comb. Theory, Ser. B*, 13:183–184, 1972.

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

[60] Peter Keevash and Dhruv Mubayi. Set systems without a simplex or a cluster. *Combinatorica*, 30:175–200, 2010.

[61] Nathan Keller. A tight quantitative version of Arrow's impossibility theorem. *J. Eur. Math. Soc.*, 14:1331–1355, 2012.

[62] Subhash Khot and Oded Regev. Vertex cover might be hard to approximate to within $2-\epsilon$. *J. Comput. Syst. Sci.*, 74(3):335–349, 2008.

[63] Guy Kindler. *Property testing, PCP and Juntas.* PhD thesis, Tel-Aviv University, 2002.

[64] Guy Kindler and Shmuel Safra. Noise-resistant Boolean functions are juntas. *Manuscript.*

[65] Benoit Larose and Claudia Malvenuto. Stable sets of maximal size in Kneser-type graphs. *Euro. J. Comb.*, 25(5):657–673, 2004.

[66] László Lovász. On the Shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25:1–7, 1979.

[67] Elchanan Mossel. A quantitative Arrow theorem. *Prob. Th. Rel. Fields*, 154(1–2):49–88, 2012.

[68] Noam Nisan and Mario Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4(4):301–313, 1994.

[69] Krzysztof Oleszkiewicz. On a nonsymmetric version of the Khinchine-Kahane inequality. In Evariste Giné, Christian Houdré, and David Nualart, editors, *Stochastic Inequalities and Applications*, volume 56 of *Progress in Probability*, pages 157–168. Birkhäuser Basel, 2003.

[70] Ran Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27:763–803, 1998.

[71] Paul Renteln. On the spectrum of the derangement graph. *Electr. J. Comb.*, 14(1), 2007.

[72] Vojtech Rödl and Ehud Friedgut. Proof of a hypercontractive estimate via entropy. *Israel J. Math.*, 125:369–380, 2001.

[73] Alexander Schrijver. A comparison of the Delsarte and Lovász bounds. *IEEE Trans. Inform. Theory*, 25:425–429, 1979.

[74] Jean-Pierre Serre. *Linear Representations of Finite Groups.* Graduate Texts in Mathematics. Springer-Verlag, 1977.

[75] Johan Håstad. Some optimal inapproximability results. *J. ACM*, 48(4):798–859, 2001.

[76] Jacques Charles François Sturm. Mémoire sur la résolution des équations numériques. *Bulletin des sciences de Férussac*, 11:419–425, 1829.

[77] Norihide Tokushige. Intersecting families — uniform versus weighted. *Ryukyu Math. J.*, 18:89–103, 2005.

[78] Norihide Tokushige. The eigenvalue method for cross t-intersecting families. *J. Algebr. Comb.*, pages 1–10, 2013.

[79] Richard M. Wilson. The exact bound in the Erdős-Ko-Rado theorem. *Combinatorica*, 4:247–257, 1984.