INSTITUTO DE MATEMÁTICA PURA E APLICADA

Random Matrices, Additive Combinatorics, and Convex Geometry

by

Marcelo Campos

Advisor: Robert Morris

INSTITUTO DE MATEMÁTICA PURA E APLICADA

Abstract

Let A be drawn uniformly at random from the set of all $n \times n$ symmetric matrices with entries in $\{-1, 1\}$. In Chapter 2 we show that

$$\mathbb{P}(\det(A) = 0) \leqslant e^{-cn},$$

where $c > 0$ is an absolute constant, thereby resolving a long-standing conjecture.

In Chapter 3 we show that

$$\mathbb{P}(\sigma_{\min}(A) \leqslant \varepsilon/\sqrt{n}) \leqslant C\varepsilon + e^{-cn},$$

where $\sigma_{\min}(A)$ denotes the least singular value of A and $C, c > 0$ are absolute constants. This result confirms a folklore conjecture on the lower tail of the least singular value of such matrices and is best possible up to the value of $C, c > 0$. We also prove that the probability that A has a repeated eigenvalue is $e^{-\Omega(n)}$, confirming a conjecture of Nguyen, Tao and Vu.

In Chapter 4 we provide a new proof of the efficient container lemma. The method of hypergraph containers was introduced in 2015 by Balogh, Morris and Samotij, and independently by Saxton and Thomason, and since then it has been used to resolve a large number of open problems in extremal and probabilistic combinatorics. One weakness of the original container lemma is that it can only be applied to hypergraphs whose uniformity is at most logarithmic in the number of vertices. In order to remedy this shortcoming, Balogh and Samotij introduced an 'efficient' container lemma which can be applied to hypergraphs whose uniformity is polynomial in the number of vertices. We give a new, much simpler proof of the efficient container lemma of Balogh and Samotij, with improved bounds. The statement of our new container lemma is inspired by the recent proof of the Kahn–Kalai conjecture by Park and Pham.

In Chapter 5, we study the number of k-element subsets J of G, an abelian group, such that $|J + J| \leqslant \lambda|J|$. Proving a conjecture of Alon, Balogh, Morris and Samotij, and improving a result of Green and Morris, who proved the conjecture for λ fixed, we provide an upper bound on the number of such sets which is tight up to a factor of $2^{o(k)}$, when $G = \mathbb{Z}$ and $\lambda = o(k/(\log n)^3)$. We also provide a generalization of this result to arbitrary abelian groups which is tight up to a factor of $2^{o(k)}$ in many cases. The main tool used in the proof is the asymmetric container lemma, introduced by Morris, Samotij and Saxton.

In Chapter 6 we determine the number and typical structure of sets of integers with bounded doubling. In particular, improving results of Green and Morris, and of Mazur, we show that the following holds for every fixed $\lambda > 2$ and every $k \geqslant (\log n)^4$: if $\omega \to \infty$ as $n \to \infty$ (arbitrarily slowly), then almost all sets $J \subset [n]$ with $|J| = k$ and $|J+J| \leqslant \lambda k$ are contained in an arithmetic progression of length $\lambda k/2 + \omega$.

In Chapter 7 we prove a refinement of a result of Green and Ruzsa about the number of sumsets in $\mathbb{Z}_n$, for n prime. In particular, we determine up to a factor of $2^{o(m)}$ the number of sets of the form $J + J$, with $J \subset \mathbb{Z}_n$, $|J| = k$ and $|J + J| = m$.

In Chapter 8 we improve the best known bound for a famous problem about covering in convex geometry. In 1957, Hadwiger conjectured that every convex body in $\mathbb{R}^d$ can be covered by 2^d translates of its interior. For over 60 years, the best known bound was of the form $O(4^d\sqrt{d}\log d)$, but this was recently improved by a factor of $e^{\Omega(\sqrt{d})}$ by Huang, Slomka, Tkocz and Vritsiou. We take another step towards Hadwiger's conjecture by deducing an almost-exponential improvement from the recent breakthrough work of Chen, Klartag and Lehec on Bourgain's slicing problem. More precisely, we prove that, for any convex body $K \subset \mathbb{R}^d$,

$$\exp\left(-\Omega\left(\frac{d}{(\log d)^8}\right)\right) \cdot 4^d$$

translates of $\mathrm{int}(K)$ suffice to cover K. We also show that a positive answer to Bourgain's slicing problem would imply an exponential improvement for Hadwiger's conjecture.

The work in Chapters 2 and 3, as well as in Appendices A and B, is joint with Matthew Jenssen, Marcus Michelen and Julian Sahasrabudhe. The work in Chapter 6 is joint with Maurício Collares, Robert Morris, Natasha Morrison and Victor Souza. The work in Chapter 8 is joint with Peter van Hintum, Robert Morris and Marius Tiba.

Resumo

Tome A uniformemente ao acaso do conjunto de todas as matrizes simétricas $n \times n$ com entradas em $\{-1, 1\}$. No Capítulo 2 mostramos que

$$\mathbb{P}(\det(A) = 0) \leqslant e^{-cn},$$

onde $c > 0$ é uma constante absoluta, resolvendo assim uma conjectura conhecida na área.

No Capítulo 3 mostramos que

$$\mathbb{P}(\sigma_{\min}(A) \leqslant \varepsilon/\sqrt{n}) \leqslant C\varepsilon + e^{-cn},$$

onde $\sigma_{\min}(A)$ denota o menor valor singular de A e $C, c > 0$ são constantes absolutas. Este resultado confirma uma conjectura sobre a cauda inferior do menor valor singular de tais matrizes e é melhor possível a menos valor de $C, c > 0$. Também provamos que a probabilidade de A ter um autovalor repetido é $e^{-\Omega(n)}$, confirmando uma conjectura de Nguyen, Tao e Vu.

No Capítulo 4 fornecemos uma nova prova do lema dos containers eficiente. O método dos containers de hipergrafos foi introduzido em 2015 por Balogh, Morris e Samotij, e independentemente por Saxton e Thomason, e desde então tem sido usado para resolver vários problemas em aberto em combinatória extremal e probabilística. Uma fraqueza do lema dos containers original é que ele só pode ser aplicado a hipergrafos cuja uniformidade é no máximo logarítmica no número de vértices. Para resolver esse problema, Balogh e Samotij introduziram um lema dos containers 'eficiente' que pode ser aplicado a hipergrafos cuja uniformidade é polinomial no número de vértices. Damos uma prova nova e muito mais simples do lema do containers eficiente de Balogh e Samotij. O enunciado do nosso novo lema dos containers é inspirado na demonstração recente da conjectura de Kahn-Kalai por Park e Pham.

No Capítulo 5, estudamos o número de subconjuntos J, com k elementos, de G, um grupo abeliano, tais que $|J + J| \leqslant \lambda|J|$. Provando uma conjectura de Alon, Balogh, Morris e Samotij, e melhorando um resultado de Green e Morris, que provaram a conjectura para λ fixo, provamos uma cota superior para o número desses conjuntos que é ótima a menos de um fator de $2^{o(k)}$, quando $G = \mathbb{Z}$ e $\lambda = o(k/(\log n)^3)$. Nós também fornecemos uma generalização deste resultado para grupos abelianos, também ótimo a menos de um fator de $2^{o(k)}$ em muitos casos. A principal ferramenta utilizada na prova é o lema dos containers assimétrico, introduzido por Morris, Samotij e Saxton.

No Capítulo 6 determinamos o número e a estrutura típica de conjuntos de inteiros com 'constante de doubling' limitada. Em particular, melhorando resultados de Green e Morris e de Mazur, mostramos que o seguinte vale para todo $\lambda > 2$ fixo e todo $k \geqslant (\log n)^4$: se $\omega \to \infty$ como $n \to \infty$ (arbitrariamente devagar), então quase todos os conjuntos $J \subset [n]$ com $|J| = k$ e $|J + J| \leqslant \lambda k$ estão contidos em uma progressão aritmética de comprimento $\lambda k/2 + \omega$.

No Capítulo 7 provamos um refinamento de um resultado de Green e Ruzsa sobre o número de sumsets em $\mathbb{Z}_n$, para n linha. Em particular, determinamos a menos de um fator de $2^{o(m)}$ o número de conjuntos da forma $J + J$, com $J \subset \mathbb{Z}_n$, $|J| = k$ e $|J + J| = m$.

No Capítulo 8 melhoramos a melhor cota conhecida para um problema famoso em geometria convexa. Em 1957, Hadwiger conjecturou que todo corpo convexo em $\mathbb{R}^d$ pode ser coberto por translações 2^d de seu interior. Por mais de 60 anos, a melhor cota conhecida era da forma $O(4^d\sqrt{d}\log d)$, mas ela foi recentemente melhorada por um fator de $e^{\Omega(\sqrt{d})}$ por Huang, Slomka, Tkocz e Vritsiou. Damos mais um passo em direção à conjectura de Hadwiger ao deduzir uma melhoria quase exponencial usando trabalhos inovadores de Chen, Klartag e Lehec para o 'Bourgain Slicing Problem'. Mais precisamente, provamos que, para qualquer corpo convexo $K \subset \mathbb{R}^d$,

$$\exp\left(-\Omega\left(\frac{d}{(\log d)^8}\right)\right) \cdot 4^d$$

translados de $\mathrm{int}(K)$ são suficientes para cobrir K. Também mostramos que uma resposta positiva para o 'Bourgain Slicing Problem' implicaria uma melhoria exponencial para essas cotas.

Os trabalhos apresentados nos Capítulos 2 e 3, e nos Apêndices A e B, foram realizados em colaboração com Matthew Jenssen, Marcus Michelen e Julian Sahasrabudhe. O trabalho apresentado no Capítulo 6 foi feito em colaboração com Maurício Collares, Robert Morris, Natasha Morrison e Victor Souza. O trabalho apresentado no Capítulo 8 foi feito em colaboração com Peter van Hintum, Robert Morris e Marius Tiba.

Contents

Chapter 1

Introduction

1.1 Singularity Probability

Let B be a random $n \times n$ matrix whose entries are chosen independently and uniformly from $\{-1, 1\}$. It is an old problem, likely stemming from multiple origins, to determine the probability that B is singular. While a moment's thought reveals the lower bound of $(1 + o(1))2n^2 2^{-n}$, the probability that two rows or columns are equal up to sign, establishing the corresponding *upper bound* remains an extremely challenging open problem. Indeed, it is widely believed that

$$\mathbb{P}(\det(B) = 0) = (1 + o(1))2n^2 2^{-n}. \tag{1.1}$$

While this precise asymptotic has so far eluded researchers, a huge amount is now known about this fascinating problem. The first advances were made by Komlós [95] in the 1960s, who showed that the singularity probability is $O(n^{-1/2})$ (see also [96] and [22]). He used Erdős [55] solution to Littlewood-Offord problem thus connecting random matrices to anti-concentration problems in probability.

Nearly 30 years later Kahn, Komlós and Szemerédi [88], in a remarkable paper, showed that the singularity probability is exponentially small. At the heart of their paper is an ingenious argument with the Fourier transform that allows them to give vastly more efficient descriptions of "structured" subspaces of $\mathbb{R}^n$ that are spanned by $\{-1, 1\}$-vectors.

Their method was then developed by Tao and Vu [152, 153] who showed a bound of $(3/4+o(1))^n$, by proving an interesting link between the ideas of [88] and the structure of set addition and, in particular, Freiman's theorem. This trajectory was then developed further by Bourgain, Vu and Wood [30], who proved a bound of $(2^{-1/2} + o(1))^n$, and by Tao and Vu [155], who pioneered the development of "inverse Littlewood-Offord theory", now an integral aspect of random matrix theory (see Section 1.1.1).

In 2007, Rudelson and Vershynin, in an important and influential paper [129], gave a different proof of the exponential upper bound on the singularity probability of B. The key idea was to construct efficient ε-nets for points on the sphere that have special anti-concentration properties and are thus more likely to be in the kernel of B. This then led them to prove an elegant inverse Littlewood-Offord type result, inspired by [155], in a geometric setting.

This perspective was then developed further in the 2018 breakthrough work of Tikhomirov [165], who proved

$$\mathbb{P}(\det(B) = 0) = (1/2 + o(1))^n,$$

thereby essentially proving the conjectured upper bound. One of the key innovations in [165] was to adopt a probabilistic viewpoint of the (discretized) sphere: instead of directly proving that efficient nets exist by latching onto some sort of structure, he shows that the probability of randomly selecting a "structured" point on the discrete sphere is incredibly unlikely. While this change in perspective may not immediately sound useful, Tikhomirov's "inversion of randomness" gives him access to a whole host of probabilistic tools.

Another major advance on the problem was made recently by Jain, Sah and Sawhney [86], who (building on the recent work of Litvak and Tikhomirov [102]), proved the natural analogue of (1.1) for random matrices with independent entries chosen from a finite set S, for any *non-uniform* distribution on S. For the case of $\{-1, 1\}$-matrices, however, they do not improve on the bound of Tikhomirov.

While the problem for matrices B with all entries independent is now very well understood, the situation for *symmetric* random matrices remains somewhat more mysterious. Indeed all of the previously mentioned works on random matrices depend deeply on the fact that the entries of B are independent, and often treat B as n independent copies of a row, thus allowing for an essentially "one-dimensional" treatment of the problem. In the symmetric case, no such perspective is available.

Let A be a random $n \times n$ symmetric matrix, uniformly drawn from all symmetric matrices with entries in $\{-1, 1\}$. Again, it is generally believed that $\mathbb{P}(\det A = 0) = \Theta(n^2 2^{-n})$ (see, e.g. [41, 169, 170]) but progress has come more slowly. The problem of showing that A is almost surely non-singular goes back, at least, to Weiss in the early 1990s but was not resolved until 2005 by Costello, Tao and Vu [41], who obtained the bound

$$\mathbb{P}(\det(A) = 0) \leqslant n^{-1/8+o(1)}. \tag{1.2}$$

The first super-polynomial bounds were obtained by Nguyen [113] and, simultaneously, Vershynin [166], the latter obtaining a bound of the form $\exp(-n^c)$. Nguyen [113] developed the

quadratic Littlewood-Offord theory introduced in [41], while Vershynin [166] worked in the geometric framework pioneered in his work with Rudelson [129–131].

In 2019, a more combinatorial perspective on this problem was introduced by Ferber, Jain, Luh and Samotij [63] and applied by Ferber and Jain [62] to show

$$\mathbb{P}(\det A = 0) \leqslant \exp(-cn^{1/4}(\log n)^{1/2}).$$

In a similar spirit, the author in joint work Mattos, Morris and Morrison [37] then improved this bound to

$$\mathbb{P}(\det A = 0) \leqslant \exp(-cn^{1/2}), \tag{1.3}$$

by proving a"rough" inverse Littlewood-Offord theorem, inspired by the theory of hypergraph containers (see [15, 142]). This bound was then improved by Jain, Sah and Sawhney [87], who improved the exponent to $-cn^{1/2}\log^{1/4} n$, and, simultaneously, by the author in joint work with Jenssen, Michelen and Sahasrabudhe [35], who improved the exponent to $-c(n\log n)^{1/2}$.

The convergence of these results onto the exponent of $-c(n\log n)^{1/2}$ is no coincidence and in fact represents a natural barrier in the problem. Indeed, all of the results up to now have treated "structured" vectors by only using the top-half of the matrix (i.e. the half above the diagonal), which conveniently consists of independent entries. However, as pointed out in [37], if one is restricted to working in the top-half of A one cannot obtain an exponent better than $-c(n\log n)^{1/2}$. Thus to get beyond this obstruction, somehow the randomness of the matrix must "reused".

In Chapter 2 we prove an exponential upper-bound on the singularity probability of a symmetric random matrix, thereby breaking though this barrier and giving the optimal bound, up to the constant in the exponent.

Theorem 1.1.1 (C., Jenssen, Michelen, Sahasrabudhe)**.** *Let A be uniformly drawn from all $n \times n$ symmetric matrices with entries in $\{-1, 1\}$. Then*

$$\mathbb{P}(\det(A) = 0) \leqslant e^{-cn}, \tag{1.4}$$

where $c > 0$ is an absolute constant.

The main technical innovations in the proof are a new inverse Littlewood-Offord type theorem for "conditioned" random walks and a new "inversion of randomness" technique that allows us to "reuse" the randomness of our matrix by pushing some of the randomness onto the random selection of a vector from our discretized sphere. In fact, there is a delicate tradeoff between these two ingredients; a loss in the second ingredient allows for an improvement in the first, *unless* some specific "arithmetic" structure arises.

1.1.1 Inverse Littlewood-Offord theory

For $v \in \mathbb{R}^n$, we define the concentration function (one of several to come) as

$$\rho(v) := \max_{b \in \mathbb{R}} \mathbb{P}\left(\sum_{i=1}^{n} \varepsilon_i v_i = b\right),$$

where $\varepsilon_1, \dots, \varepsilon_n \in \{-1, 1\}$ are uniform and independent. The study of $\rho(v)$ goes back at least to the classical work of Littlewood and Offord [100, 101] on the zeros of random polynomials, but perhaps begins in earnest with the beautiful 1945 result of Erdős [55]: if $v \in \mathbb{R}^n$ has all non-zero coordinates then

$$\rho(v) \leqslant 2^{-n}\binom{n}{\lfloor n/2 \rfloor} = O(n^{-1/2}).$$

This was then developed by Szemerédi and Sárközy[141], who showed that if all of the v_i are *distinct* then one can obtain the much stronger bound of $O(n^{-3/2})$, and by Stanley [147] who determined the *exact* maximum, using algebraic tools. A higher-dimensional version of this problem also received considerable attention (going under the name of *the* Littlewood-Offord problem) and was studied by several authors [75, 90, 94, 138] before it was ultimately resolved in the work of Frankl and Füredi [66] (see also [161]).

Of these early results, the most important for us here is the work of Halász [79] who made an important connection with the Fourier transform to prove (among other things) the following beautiful result: if there are N_k solutions to $x_1 + \dots + x_k = x_{k+1} + \dots + x_{2k}$ among the entries of v, then $\rho(v) = O(n^{-2k-1/2}N_k)$.

More recently the question has been turned on its head by Tao and Vu [155], who pioneered the study of "inverse" Littlewood-Offord theory. They suggested that if $\rho(v)$ is "large" then v must exhibit some particular arithmetic structure. For example, Tao and Vu [155, 157], and Nguyen and Vu [117, 118] proved that if v is such that $\rho(v) > n^{-C}$ then $O(n^{1-\varepsilon})$ of the elements v_i of v can be efficiently covered with a generalized arithmetic progression of rank $r = O_{\varepsilon,C}(1)$.

While these results provide a very detailed picture in the range $\rho(v) > n^{-C}$, they begin to break down[1] if $\rho(v) = n^{-\omega(1)}$ and therefore are of limited direct use in showing that the singularity probability is exponentially small. Inverse results which work for smaller ρ bring us to the "counting" Littlewood-Offord theorem of Ferber, Jain, Luh and Samotij [63], and the "weak" inverse Littlewood-Offord theorems of Campos, Mattos, Morris and Morrison [37] and of the author with Jenssen, Michelen and Sahasrabudhe in [35], which are useful for $\rho(v)$ as small as $\exp(-c(n \log n)^{1/2})$, but afford less structure.

[1] Technically these results break down if $\rho(v) < n^{-\log\log n}$.

Our novel inverse Littlewood-Offord theorem in Chapter 2 is most similar to that of Rudelson and Vershynin [129], also developed in [166] and [87], who showed that if $\rho(v) \gg e^{-cn}$ then there exists $\phi > 0$ with[2] $\phi \approx 1/\rho(v)$ for which the dilated vector $\phi \cdot v$ is exceptionally close to the integer lattice $\mathbb{Z}^n$. These Littlewood-Offord theorems, styled after Rudelson-Vershynin, tend to be a little bit subtler; instead of determining the structure of the whole vector, we only show there is some "correlation" with the rigid object $\mathbb{Z}^n$.

To state our inverse Littlewood-Offord theorem, we formulate an important notion introduced by Rudelson and Vershynin. We switch to working in $\mathbb{R}^d$, for $d \approx cn$, hinting at the later context of these results. If $A \subseteq \mathbb{R}^d$ and $x \in \mathbb{R}^d$ then define $d(x, A) := \inf_{a \in A}\{\|x - a\|_2\}$. Now, for $\alpha \in (0, 1)$, define the *least common denominator* of a vector $v \in \mathbb{R}^d$ to be the smallest $\phi > 0$ for which $\phi \cdot v$ is within $\sqrt{\alpha d}$ of a non-zero integer point. That is,

$$D_\alpha(v) = \inf\left\{\phi > 0 : \ d(\phi \cdot v, \mathbb{Z}^d \setminus \{0\}) \leqslant \sqrt{\alpha d}\right\}.$$

Note here that we have excluded the origin from $\mathbb{Z}^d$ in the definition since $\phi \cdot v \approx 0$ does not tell us any interesting about v. Indeed, given *any* $v \in \mathbb{S}^{d-1}$, one could always set $\phi < \sqrt{\alpha d}$ and obtain $d(\phi \cdot v, \mathbb{Z}^d) \leqslant d(\phi \cdot v, 0) \leqslant \sqrt{\alpha d}$, and so this degenerate case needs to be excluded somehow. In fact, in Chapter 2, we will work with a slightly different non-degeneracy condition (see (2.2)).

Our Littlewood-Offord theorem shows that a similar conclusion to that of [129] can be obtained in the presence of a large number ($k \approx n$) of additional "soft" constraints on the random walk. In particular we prove the following result, which is in fact weaker than what we really need (see Lemma 2.3.1), but captures its essence. We say that a random vector with entries in $\{-1, 0, 1\}$ is μ-lazy if each entry is independent and is equal to 0 with probability $1 - \mu$ and is equal to each of $-1, 1$ with probability $\mu/2$.

Theorem 1.1.2 (C., Jenssen, Michelen, Sahasrabudhe)**.** *There exist $R, c_1, c_2 > 0$, for which the following holds for every $d \in \mathbb{N}$, $\alpha \in (0, 1)$, $0 \leqslant k \leqslant c_1 \alpha d$ and $t \geqslant \exp(-c_1 \alpha d)$. Let $v \in \mathbb{S}^{d-1}$, let $w_1, \dots, w_k \in \mathbb{S}^{d-1}$ be orthogonal and let W be the matrix with rows $w_1, \dots, w_k$.*

If $\tau \in \{-1, 0, 1\}^d$ is a $1/4$-lazy random vector and

$$\mathbb{P}\left(|\langle \tau, v\rangle| \leqslant t \ \text{ and } \ \|W\tau\|_2 \leqslant c_2\sqrt{k}\right) \geqslant Rte^{-c_1 k}, \tag{1.5}$$

then $D_\alpha(v) \leqslant 16/t$.

Here we interpret $\|W\tau\|_2 \leqslant c_2\sqrt{k}$ as encoding the "soft" constraints and $|\langle \tau, v\rangle| \leqslant t$ as the "hard" constraint. It is useful to think of $t \approx \rho(v)$, although we actually set t relative to a related notion tailored specifically to our application.

[2] In what follows, we will be somewhat vague with our use of $\approx$.

To understand the quantitative aspect of Theorem 1.1.2, it is best to consider the contrapositive of Theorem 1.1.2, which roughly says that if v is "unstructured at scale t" (that is, $D_\alpha(v) > 16/t$) then the soft and hard constraints are roughly negatively dependent[3]. Indeed, if v is sufficiently "unstructured at scale t" then we might expect $\langle \tau, v\rangle$ to approximate a Gaussian and, in particular, to have

$$\mathbb{P}(|\langle \tau, v\rangle| \leqslant t) \approx t.$$

On the other hand, since $w_1, \dots, w_k \in \mathbb{S}^{d-1}$ are orthogonal, it turns out that (see Lemma 2.5.7)

$$\mathbb{P}(\|W\tau\|_2 \leqslant c_2\sqrt{k}) \leqslant e^{-c_1 k},$$

where $c_1 > 0$ is a suitably small constant depending on $c_2 > 0$. If these two events were negatively dependent then we would expect a bound of

$$\mathbb{P}\left(|\langle \tau, v\rangle| \leqslant t \text{ and } \|W\tau\|_2 \leqslant c_2\sqrt{k}\right) \leqslant te^{-c_1 k}.$$

Theorem 1.1.2 says something *almost* as strong as this, giving the inequality up to a constant R and the value of c_1.

For us, the main difficulty lies in "decoupling" the soft and hard constraints, which is ultimately achieved by a somewhat complicated geometric argument on the Fourier side. However, we should point out that Theorem 1.1.2 is non-trivial even in the case of $k = 0$ and in fact reduces, in this case, to the inverse Littlewood-Offord result proved by Rudelson and Vershynin in [129].

In fact, the $k = 0$ case is useful for understanding the sort of structure that the conclusion $D_\alpha(v) < c/t$ provides. It is not hard to show that if one chooses $v \in \mathbb{S}^{n-1}$ very close to a point on the lattice $(Ct)\mathbb{Z}^n$, where $C \gg 1$, then v satisfies

$$\mathbb{P}(|\langle v, \tau\rangle| \leqslant t) \gg t. \tag{1.6}$$

Thus the inverse theorem of [129, 131] says, roughly speaking, that *all* vectors satisfying (1.6) must have this structure. Our Theorem 1.1.2 says the same is true even in the presence of a large number of additional constraints.

1.2 Least Singular Value

Let A be a $n \times n$ random symmetric matrix whose entries on and above the diagonal $(A_{i,j})_{i \leqslant j}$ are i.i.d. with mean 0 and variance 1. This matrix model, sometimes called the Wigner matrix

[3]Here, we say events S, T are negatively dependent if $\mathbb{P}(S \cap T) \leqslant \mathbb{P}(S)\mathbb{P}(T)$.

ensemble, was introduced in the 1950s in the seminal work of Wigner [173], who established the famous "semi-circular law" for the eigenvalues of such matrices.

In Chapter 3 we study the extreme behavior of the *least singular value* of A, which we denote by $\sigma_{\min}(A) := \inf_{v \in \mathbb{S}^{n-1}} \|Av\|_2$. Notice that since A is symmetric we also have $\sigma_{\min}(A) = \min_{1 \leqslant i \leqslant n} |\lambda_i(A)|$, where $\lambda_i(A)$ are the eigenvalues of A. Since by Wigner's "semi-circular law" almost all eigenvalues of A lie in the interval $[-2\sqrt{n}, +2\sqrt{n}]$ with high probability, if they were evenly distributed we would expect that $\sigma_{\min}(A) = \Theta(n^{-1/2})$. Thus it is natural to consider

$$\mathbb{P}(\sigma_{\min}(A) \leqslant \varepsilon n^{-1/2}), \tag{1.7}$$

for all $\varepsilon \geqslant 0$ (see Section 1.2.2). In Chapter 3 we prove a bound on this quantity which is optimal up to constants, for all random symmetric matrices with i.i.d. *subgaussian* entries. This confirms the folklore conjecture, explicitly stated by Vershynin in [166].

Theorem 1.2.1 (C., Jenssen, Michelen, Sahasrabudhe)**.** *Let ζ be a subgaussian random variable with mean 0 and variance 1 and let A be a $n \times n$ random symmetric matrix whose entries above the diagonal $(A_{i,j})_{i \leqslant j}$ are independent and distributed according to ζ. Then for every $\varepsilon \geqslant 0$,*

$$\mathbb{P}_A(\sigma_{\min}(A) \leqslant \varepsilon n^{-1/2}) \leqslant C\varepsilon + e^{-cn}, \tag{1.8}$$

where $C, c > 0$ depend only on ζ.

This conjecture is sharp up to the value of the constants $C, c > 0$ and resolves the "up-to-constants" analogue of the Spielman–Teng conjecture for random symmetric matrices (see Section 1.2.2). Also note that the special case $\varepsilon = 0$ tells us that the singularity probability of any random symmetric A with subgaussian entry distribution is exponentially small, generalizing our previous work presented in Chapter 2 on the $\{-1, 1\}$ case.

1.2.1 Repeated eigenvalues

Before we discuss the history of the least singular value problem, we highlight one further contribution presented in Chapter 3: a proof that a random symmetric matrix has no repeated eigenvalues with probability $1 - e^{-\Omega(n)}$.

In the 1980s Babai conjectured that the adjacency matrix of the binomial random graph $G(n, 1/2)$ has no repeated eigenvalues with probability $1 - o(1)$ (see [163]). Tao and Vu [163] proved this conjecture in 2014 and, in subsequent work on the topic with Nguyen [116], went on to conjecture the probability that a random symmetric matrix with i.i.d. subgaussian entries has no repeated eigenvalues is $1 - e^{-\Omega(n)}$. In Chapter 3 we prove this conjecture en route to proving Theorem 1.2.1, our main theorem.

Theorem 1.2.2 (C., Jenssen, Michelen, Sahasrabudhe)**.** *Let ζ be a subgaussian random variable with mean 0 and variance 1 and let A be a $n \times n$ random symmetric matrix where $(A_{i,j})_{i \leqslant j}$ are independent and distributed according to ζ. Then A has no repeated eigenvalues with probability $1 - e^{-cn}$, where $c > 0$ is a constant depending only on ζ.*

Theorem 1.2.2 is easily seen to be sharp whenever $A_{i,j}$ is discrete: consider the event that three rows of A are identical; this event has probability $e^{-\Theta(n)}$ and results in two 0 eigenvalues. Also note that the constant in Theorem 1.2.2 can be made arbitrary small; consider the entry distribution ζ which takes value 0 with probability $1 - p$ and each of $\{-p^{-1/2}, p^{-1/2}\}$ with probability $p/2$. Here the probability of 0 being a repeated root is $\geqslant e^{-(3+o(1))pn}$.

We in fact prove a more refined version Theorem 1.2.2 which gives an upper bound on the probability that two eigenvalues of A fall into an interval of length ε. This is the main result of Section 3.7 and an important step in the proof of Theorem 1.2.1. For this, we let $\lambda_1(A) \geqslant \ldots \geqslant \lambda_n(A)$ denote the eigenvalues of the $n \times n$ real symmetric matrix A.

Theorem 1.2.3. *[C., Jenssen, Michelen, Sahasrabudhe] Let ζ be a subgaussian random variable with mean 0 and variance 1 and let A be a $n \times n$ random symmetric matrix where $(A_{i,j})_{i \leqslant j}$ are independent and distributed according to ζ. Then for each $\ell < cn$ and all $\varepsilon \geqslant 0$ we have*

$$\max_{k \leqslant n-\ell} \mathbb{P}(|\lambda_{k+\ell}(A) - \lambda_k(A)| \leqslant \varepsilon n^{-1/2}) \leqslant (C\varepsilon)^{\ell} + 2e^{-cn},$$

where $C, c > 0$ are constants, depending only on ζ.

1.2.2 History of the least singular value problem

The behavior of the least singular value was first studied for random matrices B_n with *i.i.d.* coefficients, rather than for *symmetric* random matrices. For this model, the history goes back to von Neumann [168] who suggested that one typically has

$$\sigma_{\min}(B_n) \approx n^{-1/2},$$

while studying approximate solutions to linear systems. This was then more rigorously conjectured by Smale [144] and proved by Szarek [148] and Edelman [43] in the case that $B_n = G_n$ is a random matrix with i.i.d. *standard gaussian* entries. Edelman found an exact expression for the density of the least singular value in this case. By analyzing this expression, one can deduce that

$$\mathbb{P}(\sigma_{\min}(G_n) \leqslant \varepsilon n^{-1/2}) \leqslant \varepsilon, \tag{1.9}$$

for all $\varepsilon \geqslant 0$ (see e.g. [146]) as well as an asymptotic expansion for this probability when ε is fixed and $n \to \infty$. While this gives a very satisfying understanding of the gaussian case,

one encounters serious difficulties when trying to extend this result to other distributions, as Edelman's proof relies crucially on the special tools available only in the gaussian case. In the last 20 or so years, intense study of the least singular value of i.i.d. random matrices has been undertaken with the overall goal of proving an appropriate version of (1.9) for different entry distributions and models of random matrices.

An important and challenging feature of the more general problem arises in the case of *discrete* distributions, where the matrix B_n can become singular with non-zero probability. This singularity event will affect the quantity (1.7) for very small ε and thus estimating the probability that $\sigma_{\min}(B_n) = 0$ is a crucial aspect of generalizing (1.9). This is reflected in the famous and influential Spielman–Teng conjecture [145] which stipulates the bound

$$\mathbb{P}(\sigma_{\min}(B_n) \leqslant \varepsilon n^{-1/2}) \leqslant \varepsilon + 2e^{-cn}, \tag{1.10}$$

where B_n is a Bernoulli random matrix. Here this added exponential term "comes from" the singularity probability of B_n.

In this direction, a key breakthrough was made by Rudelson [128] who proved that if B_n has i.i.d. *subgaussian* entries then

$$\mathbb{P}(\sigma_{\min}(B_n) \leqslant \varepsilon n^{-1/2}) \leqslant C\varepsilon n + n^{-1/2}\,.$$

This result was extended in a series of works [130, 154, 155, 171] ultimately terminating in the influential work of Rudelson and Vershynin [129] who showed the "up-to-constants" version of Spielman-Teng:

$$\mathbb{P}(\sigma_{\min}(B_n) \leqslant \varepsilon n^{-1/2}) \leqslant C\varepsilon + e^{-cn}, \tag{1.11}$$

where B_n is a matrix with i.i.d. entries that follow any subgaussian distribution and $C, c > 0$ depend only on ζ. A key ingredient in the proof of (1.11) is a novel approach to the "inverse Littlewood-Offord problem," a perspective pioneered by Tao and Vu [155] (discussed in section 1.1.1).

Another very different approach was taken by Tao and Vu [156] who showed that the distribution of the least singular value of B_n is identical to the least singular value of the Gaussian matrix G_n, up to scales of size n^{-c}. In particular they prove that

$$\left|\mathbb{P}(\sigma_{\min}(B_n) \leqslant \varepsilon n^{-1/2}) - \mathbb{P}(\sigma_{\min}(G_n) \leqslant \varepsilon n^{-1/2})\right| = O(n^{-c_0}), \tag{1.12}$$

thus resolving the Speilman-Teng conjecture for $\varepsilon \geqslant n^{-c_0}$, in a rather strong form.

While falling just short of the Spielman-Teng conjecture, the work Tao and Vu [156], Rudelson and Vershynin [129] and subsequent refinements by Tikhomirov [104] and Livshyts, Tikhomirov

and Vershynin [104] (see also [103, 124]) leave us with a very strong understanding of the least singular value for *i.i.d.* matrix models. However, progress on the analogous problem for random symmetric matrices, or *Wigner random matrices*, has come somewhat more slowly and more recently: in the symmetric case, even proving that A_n is non-singular with probability $1 - o(1)$ was not resolved until the important 2006 paper of Costello, Tao and Vu [41].

Progress on the symmetric version of Spielman–Teng continued with Nguyen [113] and, independently, Vershynin [166]. Nguyen proved that for any $B > 0$ there exists an $A > 0$ for which[4]

$$\mathbb{P}(\sigma_{\min}(A_n) \leqslant n^{-A}) \leqslant n^{-B}.$$

Vershynin [166] proved that if A_n is a matrix with subgaussian entries then, for all $\varepsilon > 0$, we have

$$\mathbb{P}(\sigma_{\min}(A_n) \leqslant \varepsilon n^{-1/2}) \leqslant C_\eta \varepsilon^{1/8-\eta} + 2e^{-n^c}, \tag{1.13}$$

for all $\eta > 0$, where the constants $C_\eta, c > 0$ may depend on the underlying subgaussian random variable. He went on to conjecture that ε should replace $\varepsilon^{1/8}$ as the correct order of magnitude and that e^{-cn} should replace e^{-n^c}.

After Vershynin, a series of works [35, 37, 62, 63, 87] made progress on singularity probability (i.e. the $\varepsilon = 0$ case of Vershynin's conjecture), and we, in Chapter 2, ultimately showed that the singularity probability is exponentially small, when $A_{i,j}$ is uniform in $\{-1, 1\}$:

$$\mathbb{P}_{A_n}(\det(A_n) = 0) \leqslant e^{-cn},$$

which is sharp up to the value of $c > 0$.

However, for general ε the state of the art is due to Jain, Sah and Sawhney [87], who improved on Vershynin's bound (1.13) by showing

$$\mathbb{P}(\sigma_{\min}(A_n) \leqslant \varepsilon n^{-1/2}) \leqslant C\varepsilon^{1/8} + e^{-\Omega(n^{1/2})},$$

under the subgaussian hypothesis on A_n.

For large ε, for example $\varepsilon \geqslant n^{-c}$, another very different and powerful set of techniques have been developed, which in fact apply more generally to the distribution of other "bulk" eigenvalues. The works of Tao and Vu [154, 159], Erdős, Schlein and Yau [50, 51, 57], Erdős, Ramírez, Schlein, Tao, Vu, Yau [49], and specifically Bourgade, Erdős, Yau and Yin [24] tell us that

$$\mathbb{P}(\sigma_{\min}(A_n) \leqslant \varepsilon n^{-1/2}) \leqslant \varepsilon + o(1), \tag{1.14}$$

[4]We note that this result is actually proved for all matrices of the form $A_n + F$, where F is any fixed $n \times n$ matrix and the entries of A_n have mean 0, but need not be identically distributed and may have large variances.

thus obtaining the correct dependence on ε asymptotically[5]. These results are similar in flavor to (1.12) in that they show the distribution various eigenvalue statistics are closely approximated by the corresponding statistics in the Gaussian case. We note however that it appears these techniques are limited to these large ε and different ideas are required for $\varepsilon < n^{-C}$, and certainly for ε as small as $e^{-\Theta(n)}$.

Our main theorem, Theorem 1.2.1, proves Vershynin's conjecture and thus proves the optimal dependence on ε for all $\varepsilon > e^{-cn}$, up to constants.

1.2.2.1 Smoothed analysis

The least singular value of random matrices also has significant application to theoretical computer science due its central role in *smoothed analysis*, a notion introduced by Spielman and Teng [145]. Inspired by the observation that certain algorithms have theoretically slow worst-case performance but efficient performance in practice, Spielman and Teng proved that if one perturbs any linear program by Gaussian noise, then the simplex method typically runs quickly. Here, the Gaussian noise represents various errors in the data input and provides theoretical groundwork for the observation that the simplex algorithm typically runs quickly even on examples for which it theoretical exhibits exponential run-time. At the core of their proof is the study of the least singular value of random perturbations of arbitrary matrices. Together with Sankar, Spielman and Teng [139] later proved that Edelman's bound (1.9) remains essentially unchanged if the Gaussian random matrix is perturbed by an arbitrary matrix; in particular they showed that if G_n is an $n \times n$ matrix of i.i.d. standard Gaussians then for any (deterministic) $n \times n$ matrix F we have

$$\mathbb{P}(\sigma_n(F + G_n) \leqslant \varepsilon/\sqrt{n}) \leqslant 1.823\varepsilon\,. \tag{1.15}$$

This bound is then used to show that the condition number—i.e. the ratio of greatest and least singular values—of various perturbed matrices is well-behaved, thus allowing for efficient behavior of many algorithms. Since its introduction, smoothed analysis has been applied to understand the behavior of various algorithms (e.g. [9, 20] and the references therein).

In practice, many numerical errors are of a more discrete nature, and so one may ask if the behavior of (1.15) still holds if G_n is replaced by, say B_n, a matrix with i.i.d. entries from

[5] Tao and Vu in [159] treat the least singular value with their Corollary 24. There they prove up the distribution of $\sigma_{\min}(A_n)$ agrees with $\sigma_{\min}$ of a symmetic gaussian matrix up to a polynomial error assuming certain moment-matching assumptions on the distribution of the entries of A_n. A follow-up work [49] joint with Erdős, Ramírez, Schlein, and Yau describes an approach to combine ideas from the works [50, 51, 57] to remove the moment matching assumptions of [159], but does not explicitly address the problem of the least singular value. Finally, the work [24] proves the non-quantitative (1.14) (see the discussion below Theorem 2.2 of [24]).

$\{-1, 1\}$. Perhaps surprisingly, Tao and Vu [158] proved that this is not the case, and showed that a careful choice of F can yield an extremely small singular value of $F + B_n$ with probability $\Omega(n^{-1/2})$. These matrices, however, have increasing operator norm (see [85, 158] for results depending on F). It was in this context that in the case of $F = 0$, Spielman and Teng conjectured (1.10).

Work on smoothed analysis with non-gaussian noise continues with work by Tikhomirov [164] and Jain, Sah and Sahwney [85]. The smoothed analysis of random *symmetric* matrices has also recently received significant attention (see e.g.[29, 60, 114, 166]).

1.2.2.2 Semicircular Laws and Universality

Among the earliest mathematical treatments of random matrices are Wigner's groundbreaking works [172, 174] on random symmetric matrices—also called Wigner matrices—in which he proved the celebrated semicircular law for certain ensembles of random matrices. Over a decade later, Pastur [120] generalized Wigner's work to show that this behavior is *universal* meaning that the semicircular law holds for an arbitrary (sufficiently well-behaved) distribution of matrix entries. Even further generalizations and extensive connections to free probability have since been explored (see the books [6, 12, 149] for more context).

A seminal sequence of works [50, 51, 57] by Erdős, Schlein and Yau developed the theory of *local* semicircular laws, which show quantitative rates of convergence to the semicircular distribution on small windows for a general class of Wigner matrices (see [10, 11] for earlier works and more historical context). Work on the semicircular law continues still, for instance in the works [69, 70] and the recent survey [77].

The above results belong to the wider study of *universality*: the idea that certain statistics of the spectrum of a random matrix should not depend too heavily on the precise distribution of its entries. Of particular interest in this area is showing that the k-point correlation functions in the bulk converge and have the same limit as those of the Gaussian Orthogonal (or Unitary) ensemble. The problem of universality in this context is often referred to as the Wigner-Dyson-Mehta conjecture (see Conjectures 1.2.1 and 1.2.2 of Mehta's text [107] for precise statements). Many results in this context focus on relaxing the assumptions on the distribution of the matrix entries. Rather than delve into the literature here, we refer the reader to the works [1, 24, 47–49, 52–54, 84, 159, 160, 162] for results in this direction and more context.

1.2.2.3 Eigenvalue gaps

Another well-studied spectral statistic of Wigner matrices is the size of the *gaps* between consecutive eigenvalues. This study has been inspired in part by Wigner's bold postulate [172] that the spacing between the discrete energy levels of a heavy atomic nucleus should resemble the the spacing between the eigenvalues of a random Hermitian matrix.

Given a random Wigner matrix A_n, a natural problem is to determine the limiting distribution of a given gap $\delta_i(A_n) := \lambda_{i+1}(A_n) - \lambda_i(A_n)$. Even for the GUE, this was only recently computed (in the bulk) by Tao [150]. Subsequent work of Tao and Vu [159] and Erdős and Yau [58] shows that this gap distribution is in fact universal within a large class of Wigner matrices. Another statistic of interest is the *minimum gap* $\delta_{\min}(A_n) := \min_{1 \leqslant i \leqslant n-1} \lambda_{i+1}(A_n) - \lambda_i(A_n)$. For the GUE, the limiting distribution of $\delta_{\min}$ was determined by Bourgade and Ben-Arous [8]. As noted by Nguyen-Tao-Vu [116], the Wegner estimates of Erdős, Schlein, and Yau [57] show that under certain smoothness assumptions of the entries A_n one has the following bound:

$$\mathbb{P}(\delta_{\min}(A_n) \leqslant \varepsilon/\sqrt{n}) \lesssim n\varepsilon^3 + e^{-cn},$$

for $\varepsilon > 0$, matching the behavior of the GUE up to the implicit constant and exponential error term.

In the case where the entries of A_n are discrete, even showing that $\delta_{\min}(A_n) > 0$ (i.e. the spectrum of A_n is *simple*) with high probability is non-trivial. Recently Tao and Vu [163] showed that the spectrum of A_n is simple with high probability under very mild assumptions on the distribution of the entries of A_n. In particular, their result applies to the case where A_n is the adjacency matrix of the Erdős-Renyi random graph $G(n, 1/2)$, which resolved a conjecture of Babai (motivated by the graph isomorphism question).

In the case where the entries of A_n are subgaussian, Nguyen, Tao and Vu [116] showed that $\delta_{\min}(A_n) = 0$ with probability $O(\exp(-n^c))$. In particular this holds for the case where A_n is a uniformly random symmetric matrix with entries in $\{-1, 1\}$. Nguyen, Tao and Vu conjectured that in this case, the bound can be improved to $O(\exp(-cn))$. Theorem 3.1.3 resolves this conjecture.

1.3 Hypergraph Container Method

The method of hypergraph containers is one of the most powerful and flexible techniques in probabilistic combinatorics. Since its introduction by Balogh, Morris and Samotij [15] and Saxton and Thomason [142], it has been used to resolve a large number of well-known conjectures

in extremal, probabilistic and additive combinatorics, as well as problems in areas such as discrete geometry. As a matter of fact, a variant of the hypergraph container lemma will be one of main tools used in Chapters 5, 6 and 7.

Roughly speaking, the idea is that many interesting combinatorial objects (such as H-free graphs, and sets with no arithmetic progression of a given length) can be encoded as the independent sets of uniform hypergraphs, and that the independent sets of these hypergraphs exhibit a certain subtle clustering phenomenon, which can be exploited when counting, or when bounding the probabilities of certain events. The survey [16] provides a gentle introduction to the area, and numerous applications of the method.

The original container lemma provides sharp bounds (up to a constant factor) when the uniformity of the hypergraph is bounded, and still provides useful bounds even when the uniformity is poly-logarithmic in the number of vertices of the hypergraph. For hypergraphs whose edges are larger than this, however, the statement becomes trivial. In order to remedy this shortcoming of the method, Balogh and Samotij [18] introduced a strengthening of the container lemma, which they called the 'efficient' container lemma, that provides useful information for uniformities as large as n^c, for some (moderate) constant $c > 0$.

The proof the efficient container lemma in [18] is quite long and complicated, and relies on some fairly intricate geometric lemmas. The rough idea is to control the codegrees of the hypergraph by controlling the norm of certain vectors. Then the container algorithm determines what to do next according to geometric properties of this set of vectors, attempting to minimize their norms.

The purpose of Chapter 4 is to provide a much simpler proof of a slightly stronger result. We remark that the statement of our new container lemma, Theorem 4.1.1, was inspired by the recent breakthrough results by Alweiss, Lovett, Wu and Zhang [5] on the Erdős–Rado sunflower conjecture, and by Frankston, Kahn, Narayanan and Park [67] and Park and Pham [119] on the Kahn–Kalai conjecture.

In order to state the main result of Chapter 4, we will need to introduce a couple of important notions, which we will use to measure the 'size' of our containers. Let $\mathcal{G}$ and $\mathcal{H}$ be hypergraphs, and write

$$\langle \mathcal{G} \rangle = \bigcup_{E \in \mathcal{G}} \big\{ F \subset V(\mathcal{G}) : E \subset F \big\}$$

for the up-set generated by $\mathcal{G}$. We say that $\mathcal{G}$ is a *cover* for $\mathcal{H}$ if $\mathcal{H} \subset \langle \mathcal{G} \rangle$. In other words, $\mathcal{G}$ is a cover for $\mathcal{H}$ if for every edge $F \in \mathcal{H}$ there exists an edge $E \in \mathcal{G}$ with $E \subset F$.

Next, for each $p > 0$, define the *p-weight* of $\mathcal{G}$ to be

$$w_p(\mathcal{G}) = \sum_{E \in \mathcal{G}} p^{|E|}.$$

Note that $w_p(\mathcal{G})$ is just the expected number of edges of $\mathcal{G}$ in a p-random subset of $V(\mathcal{G})$. Finally, let $\mathcal{I}(\mathcal{H})$ denote the family of independent sets of $\mathcal{H}$. We are now ready to state our new container theorem.

Theorem 1.3.1. *Let $\mathcal{H}$ be an r-uniform hypergraph with n vertices, and let $0 < p < 1/4r$. There exists a family $\mathcal{S}$ of subsets of $V(\mathcal{H})$, and functions*

$$g \colon \mathcal{I}(\mathcal{H}) \to \mathcal{S} \qquad \text{and} \qquad f \colon \mathcal{S} \to 2^{V(\mathcal{H})},$$

such that:

(a) *For each $I \in \mathcal{I}(\mathcal{H})$ we have $g(I) \subset I \subset f(g(I))$.*

(b) *For each $S \in \mathcal{S}$, we have $|S| \leqslant 16r^2pn$.*

(c) *If $X = f(S)$ for some $S \in \mathcal{S}$, then there exists a cover $\mathcal{G}$ for $\mathcal{H}[X]$ with*

$$w_p(\mathcal{G}) < p|X|$$

and $|E| \geqslant 2$ for all $E \in \mathcal{G}$.

The main novelty of Theorem 4.1.1 is property (c), which refines the usual measures used[6] to control the 'size' of a container. Note that the condition that all edges of the cover have size at least 2 is necessary to prevent the conclusion from being trivial, since every hypergraph on vertex set X has a cover (of singletons) of p-weight $p|X|$. We would also like to draw the reader's attention to the dependence on r in the bound on the size of the 'fingerprint' S, which is polynomial (as in the efficient container lemma of Balogh and Samotij [18]) as opposed to super-exponential (as in the original container lemmas). In Section 4.5 we will show that the main theorems of [18] follow easily from Theorem 4.1.1, with slightly improved bounds.

[6] In [15, 142], and also in [18], the 'size' of a container is measured either by the number of vertices, or by the number of edges they contain.

1.4 Counting in additive combinatorics

1.4.1 Counting sets with bounded sumset

One of the central objects of interest in additive combinatorics is the sumset

$$A + B := \{a + b : a \in A,\, b \in B\}$$

of two sets $A, B \subset \mathbb{Z}$. If $|A + A| = \lambda |A|$ we say A has *doubling constant* (or sometimes simply *doubling*) λ. A cornerstone of the theory is the celebrated theorem of Freĭman [68], which states that if $|A+A| \leqslant \lambda |A|$, then A is contained in a generalized arithmetic progression[7] of dimension $O_\lambda(1)$ and size $O_\lambda(|A|)$, where the implicit constants depend only on λ. For an overview of the area, see the book of Tao and Vu [151].

In Chapters 5 and 6 we will be interested in the number and typical structure of sets with sumset of a given size. The study of this problem was initiated in 2005 by Green [72], who was motivated by applications to random Cayley graphs, and in recent years there has been significant interest in related questions [2, 13, 14, 42, 73]. In particular Alon, Balogh, Morris and Samotij [2] proved a refinement of the Cameron-Erdős conjecture about the number of sum-free subsets of $[n]$, which was solved independently by Green [71] and Sapozhenko [140]. In [2] the author used an early form of the method of hypergraph containers and also needed to prove a bound on the number of k-sets $A \subset [n]$ with doubling constant λ. They moreover conjectured that the following stronger (and, if true, best possible) bound holds.

Conjecture 1.4.1 (Alon, Balogh, Morris and Samotij)**.** *For every $\delta > 0$, there exists $C > 0$ such that the following holds. If $k \geqslant C \log n$ and if $\lambda \leqslant k/C$, then there are at most*

$$2^{\delta k} \binom{\frac{1}{2}\lambda k}{k}$$

sets $J \subset [n]$ with $|J| = k$ and $|J + J| \leqslant \lambda |J|$.

The conjecture was later confirmed for λ constant by Green and Morris [73]; in fact they proved a slightly more general result: for each fixed λ and as $k \to \infty$, the number of sets $J \subset [n]$ with $|J| = k$ and $|J + J| \leqslant \lambda |J|$ is at most

$$2^{o(k)} \binom{\frac{1}{2}\lambda k}{k} n^{\lfloor \lambda + o(1) \rfloor}.$$

In Chapter 5 we prove Conjecture 1.4.1 for all $\lambda = o\left(k/(\log n)^3\right)$.

[7] That is, a set of the form $P = \{a + i_1 d_1 + \cdots + i_s d_s : i_j \in \{0, \ldots, k_j\}\}$ for some $a, d_1, \ldots, d_s, k_1, \ldots, k_s \in \mathbb{N}$.

Theorem 1.4.2. *Let k, n be integers and $2 \leqslant \lambda \leqslant o\big(\frac{s}{(\log n)^3}\big)$. The number of sets $J \subset [n]$ with $|J| = k$ such that $|J + J| \leqslant \lambda |J|$ is at most*

$$2^{o(k)} \binom{\frac{1}{2}\lambda k}{k}.$$

In order to understand such why results should be true, recall first that, by Freĭman's theorem, a set has bounded doubling if and only if it is a subset of positive density of a generalized arithmetic progression P of bounded dimension. Now, if A were a random subset of P of positive density, then $A+A$ would be unlikely to 'miss' many elements of $P+P$, and this suggests that most sets of bounded doubling should in fact be contained in an arithmetic progression of size roughly $|A + A|/2$. If this intuition was true we should expect to have about

$$\binom{\lambda k/2}{k}$$

choices for A, which is roughly what Conjecture 1.4.1 states. This intuition about the typical structure of A will be confirmed in a stronger sense in Chapter 6, which is joint work with Maurćio Collares, Robert Morris, Natasha Morrison and Victor Souza, where we prove the following theorem.

Theorem 1.4.3 (C., Collares, Morris, Morrison, Souza)**.** *Fix $\lambda \geqslant 3$ and $\varepsilon > 0$, let $n \in \mathbb{N}$ be sufficiently large, and let $k \geqslant (\log n)^4$. Define $c(\lambda, \varepsilon) := 2^{20}\lambda^2 \log(1/\varepsilon) + 2^{560}\lambda^{32}$. Let $A \subset [n]$ be chosen uniformly at random from the sets with $|A| = k$ and $|A + A| \leqslant \lambda k$. Then there exists an arithmetic progression P with*

$$A \subset P \qquad \text{and} \qquad |P| \leqslant \frac{\lambda k}{2} + c(\lambda, \varepsilon)$$

with probability at least $1 - \varepsilon$.

1.4.2 The number of sumsets of a given size

In Chapter 7 we will consider another natural counting problem in additive combinatorics: how many sumsets of a given size are there in $\mathbb{Z}_n$? Questions of this type were first considered by Green and Ruzsa [74], who proved in 2004 that if n is prime then there are $2^{n/3+o(n)}$ sets of the form $A + A$ in $\mathbb{Z}_n$. A few years later, Alon, Granville and Ubis [4] proved a corresponding result in the asymmetric setting, showing that there are $2^{n/2+o(n)}$ sets in $\mathbb{Z}_n$ of the form $A + B$ for some $A, B \subset \mathbb{Z}_n$ with $\min\{|A|, |B|\} \gg 1$.

A natural refinement of the problem studied by Green and Ruzsa is as follows: how many sets of size m in $\mathbb{Z}_n$ are of the form $A + A$ for some $A \subset \mathbb{Z}_n$? In Chapter 7 we will resolve this problem up to a factor of $2^{o(m)}$ for all $(\log n)^4 \leqslant m \leqslant 2n/3$. Our main result is the following

theorem, which provides a sharp bound on the number of sumsets of a given size and doubling constant.

Theorem 1.4.4. *Let n be a prime, and let $m, k \in \mathbb{N}$ with $m \geqslant (2+\sqrt{5})k$ and $k \geqslant (\log n)^4$. There are at most*

$$2^{o(m)} \binom{\frac{m-k}{2}}{k} \tag{1.16}$$

sets of the form $A + A$ for some $A \subset \mathbb{Z}_n$ with $|A| = k$ and $|A + A| = m$.

1.5 Hadwiger's Conjecture

Hadwiger's covering problem asks: how many translates of the interior of a convex body $K \subset \mathbb{R}^d$ are needed to cover K? That is, it asks for the value of

$$N(K) = \min\Big\{N \in \mathbb{N} : \exists\, x_1, \dots, x_N \in \mathbb{R}^d \text{ such that } K \subset \bigcup_{i=1}^{N} \big(x_i + \mathrm{int}(K)\big)\Big\}.$$

Hadwiger [78] conjectured in 1957 that $N(K) \leqslant 2^d$ for all convex $K \subset \mathbb{R}^d$. Note that this bound is attained by the cube $[0,1]^d$. The conjecture was proved when $d \leqslant 2$ by Levy [99] in 1955, but for over 60 years the best known bound for general d was

$$N(K) \;\leqslant\; \big(d \log d + d \log\log d + 5d\big)\binom{2d}{d} \;=\; O\big(4^d \sqrt{d} \log d\big),$$

which follows from the Rogers–Shephard inequality [126], together with a bound of Rogers [125] on the minimum density of a covering of $\mathbb{R}^d$ with translates of K. A few years ago, however, a breakthrough was made by Huang, Slomka, Tkocz and Vritsiou [83], who used a large deviation result of Guédon and Milman [76], which is related to the so-called 'thin-shell' phenomenon (see below), to obtain a bound of the form

$$N(K) \;\leqslant\; e^{-\Omega(\sqrt{d})} \cdot 4^d. \tag{1.17}$$

In Chapter 8, which is presents joint work with Peter van Hintum, Robert Morris and Marius Tiba, we will prove the following almost-exponential improvement of their bound.

Theorem 1.5.1. *[C., van Hintum, Morris, Tiba] If $K \subset \mathbb{R}^d$ is a convex body, then*

$$N(K) \leqslant \exp\left(-\Omega\left(\frac{d}{(\log d)^8}\right)\right) \cdot 4^d$$

as $d \to \infty$.

We will deduce Theorem 1.5.1 from recent results of Chen [39] and Klartag and Lehec [92] on the Bourgain slicing problem, which asks for the smallest number $L_d > 0$ such that for every convex body $K \subset \mathbb{R}^d$ of volume 1, there exists a hyperplane H such that $K \cap H$ has $(d-1)$-dimensional volume at least $1/L_d$. In particular, Bourgain [25, 26] asked whether or not L_d is bounded from above by an absolute constant. This problem is still open, and for many years the best known bound was of the form $L_d = O(d^{1/4})$, proved by Bourgain [27, 28] (with an extra log-factor) and Klartag [91]. However, just a couple of years ago, Chen [39] made a major breakthrough on the problem by proving a bound of the form $L_d = d^{o(1)}$. His bound was then improved further by Klartag and Lehec [92], who showed that $L_d = O(\log d)^4$.

The breakthroughs in [39] and [92] both used "stochastic localization", a powerful and beautiful technique that was introduced about ten years ago by Eldan [45], to bound the *thin-shell constant*, σ_d, which is defined to be

$$\sigma_d := \sup_K \mathbb{E}\big[\big(\|X\|_2 - \sqrt{d}\big)^2\big],$$

where the supremum is over convex bodies $K \subset \mathbb{R}^d$ in isotropic position[8], and $X \sim \mathcal{U}(K)$ is a uniformly chosen random point of K. The thin-shell conjecture [7, 21] states that $\sigma_d = O(1)$, and it was shown by Eldan and Klartag [46] that

$$L_d = O(\sigma_d),$$

so bounds on the thin-shell constant imply bounds for the Bourgain slicing problem. We remark that, by a deep result of Eldan [45], bounds on the thin-shell constant also imply bounds for the Kannan–Lovász–Simonovits isoperimetric conjecture [89], see e.g. [45, 98].

We will use an equivalent formulation of the Bourgain slicing problem, due to Milman and Pajor [109] (see also [93]). Given a convex body $K \subset \mathbb{R}^d$, define the *isotropic constant* of K to be

$$L_K = \left(\frac{\sqrt{\det(\Sigma_K)}}{\mathrm{Vol}_d(K)}\right)^{1/d},$$

where $\Sigma_K = \mathbb{E}[X \otimes X]$ is the covariance matrix of the random variable $X \sim \mathrm{Unif}(K)$, that is, X is a uniformly random point of K. Equivalently, there exists an affine transformation that maps K to a convex body K' of volume 1 with $\Sigma_{K'} = L_K^2 I_d$, where I_d is the identity matrix. By [109, Corollary 3.2] we have $L_K = \Theta(L_d)$ for every convex body $K \subset \mathbb{R}^d$, and hence $L_K = O(\log d)^4$, by the bound on L_d proved by Klartag and Lehec [92].

[8] This means that $\mathbb{E}[X] = 0$ and $\mathbb{E}[X \otimes X] = I_d$, where $X \sim \mathcal{U}(K)$ is a uniformly-chosen random point of $K \subset \mathbb{R}^d$, and I_d is the identity matrix. For any convex body K there exists a unique (up to rotations) affine transformation that maps K to isotropic position.

Our main result in Chapter 8 is the following bound on the covering number of a convex body. Since $L_K = O(\log d)^4$, it implies the bound in Theorem 1.5.1 for Hadwiger's conjecture.

Theorem 1.5.2 (C., van Hintum, Morris, Tiba)**.** *If $K \subset \mathbb{R}^d$ is a convex body, then*

$$N(K) \leqslant \exp\left(-\frac{\Omega(d)}{L_K^2}\right) \cdot 4^d$$

as $d \to \infty$.

Chapter 2

The singularity probability of a random symmetric matrix is exponentially small

This chapter presents joint work with Matthew Jenssen, Marcus Michelen and Julian Sahasrabudhe. It is adapted from the paper [33] which has been submitted for publication.

2.1 Introduction

Let A be a random $n \times n$ symmetric matrix, uniformly drawn from all symmetric matrices with entries in $\{-1, 1\}$. It is generally believed that $\mathbb{P}(\det A = 0) = \Theta(n^2 2^{-n})$ (see, e.g. [41, 169, 170]).

The goal of this chapter is to prove exponential upper-bound on the singularity probability of a symmetric random matrix.

Theorem 2.1.1. *Let A be uniformly drawn from all $n \times n$ symmetric matrices with entries in $\{-1, 1\}$. Then*

$$\mathbb{P}(\det(A) = 0) \leqslant e^{-cn}, \tag{2.1}$$

where $c > 0$ is an absolute constant.

We will also prove a new Inverse Littlewood Offord type theorem. Define, for $\alpha \in (0, 1)$, the *least common denominator* of a vector $v \in \mathbb{R}^d$ to be

$$D_\alpha(v) := \inf \left\{ \phi > 0 : \ \|\phi \cdot v\|_{\mathbb{T}} \leqslant \min \left\{ \phi \|v\|_2/2, \sqrt{\alpha d} \right\} \right\}, \tag{2.2}$$

where $\|x\|_{\mathbb{T}} := \inf\{\|x-y\|_2 : y \in \mathbb{Z}^d\}$, for $x \in \mathbb{R}^d$, denotes the minimum distance to an integer point. We say that a random vector with entries in $\{-1,0,1\}$ is μ-lazy if each entry is independent and is equal to 0 with probability $1-\mu$ and is equal to each of $-1,1$ with probability $\mu/2$.

Theorem 2.1.2. *There exist $R, c_1, c_2 > 0$, for which the following holds for every $d \in \mathbb{N}$, $\alpha \in (0,1)$, $0 \leqslant k \leqslant c_1 \alpha d$ and $t \geqslant \exp(-c_1 \alpha d)$. Let $v \in \mathbb{S}^{d-1}$, let $w_1, \dots, w_k \in \mathbb{S}^{d-1}$ be orthogonal and let W be the matrix with rows $w_1, \dots, w_k$.*

If $\tau \in \{-1,0,1\}^d$ is a $1/4$-lazy random vector and

$$\mathbb{P}\left(|\langle \tau, v\rangle| \leqslant t \text{ and } \|W\tau\|_2 \leqslant c_2\sqrt{k}\right) \geqslant Rte^{-c_1 k}, \tag{2.3}$$

then $D_\alpha(v) \leqslant 16/t$.

In the next subsection we sketch the proof strategy and present how the proof is organized.

2.1.1 Proof sketch and a new "inversion of randomness" technique

Here we briefly sketch how our inverse Littlewood-Offord result is used alongside a novel scheme for "reusing randomness" to prove Theorem 2.1.1. As hinted at before, we will be helped along by treating the discretized sphere as a probability space, which will allow us to "recover" some of the randomness lost due to the symmetry of A. We keep our discussion here loose and impressionistic and we will take up our careful study in the following section.

Our first move will be to "locally replace" A with a random matrix M that has many of the entries zeroed out. This will allow us to untangle some of the more subtle and complicated dependencies and has the advantage that various associated Fourier transforms are non-negative. Indeed let[1]

$$M = \begin{bmatrix} \mathbf{0}_{[d]\times[d]} & H_1^T \\ H_1 & \mathbf{0}_{[d+1,n]\times[d+1,n]} \end{bmatrix}, \tag{2.4}$$

where $d = cn$ and H_1 is a $(n-d)\times d$ random matrix with i.i.d. entries that are μ-lazy, meaning that $(H_1)_{i,j} = 0$ with probability $1-\mu$ and $(H_1)_{i,j} = \pm 1$ with probability $\mu/2$. We stress here that we cannot "globally" replace A with M, and we may need to permute coordinates, depending on what part of the sphere we are working on.

[1] Here we use the notation $[n] := \{1, \dots, n\}$; for a vector $v \in \mathbb{R}^n$ and $S \subseteq [n]$, we use the notation $v_S := (v)_{i\in S}$ and for a $n \times m$ matrix A, and $R \subseteq [m]$, we use the notation $A_{S\times R}$ for the $|S| \times |R|$ matrix $(A_{i,j})_{i\in S, j\in R}$.

We now follow the strategy of [129, 165] and partition the sphere $\mathbb{S}^{n-1}$ based on the anti-concentration properties of the various $v \in \mathbb{S}^{n-1}$. Indeed, for each $v \in \mathbb{S}^{n-1}$, we find a corresponding "scale" $\varepsilon \in (0,1)$ for which

$$\mathbb{P}(\|Mv\|_2 < \varepsilon\sqrt{n}) \approx (L\varepsilon)^n, \tag{2.5}$$

where L is a large constant. Notice here that we have defined this "scale" relative to the symmetric matrix M, rather than A or $\rho(v)$, and so we expect it to capture the anti-concentration properties of v, specific to the matrix M. This ε should be interpreted as "the scale at which the anti-concentration properties of v just start to be felt", as we imagine gradually decreasing ε from 1 to 0. For example, if v is a random point on the sphere, it is not hard to see that v will typically have $\varepsilon \leqslant e^{-cn}$, which is in fact *so* small that we can safely ignore v (due to previous work). On the other hand, the constant vector $n^{-1/2}(1,\dots,1)$ will have $\varepsilon \approx n^{-1/2}$. Interestingly, this latter fact is not easy to establish rigorously, but is heuristically not hard to guess in analogy with the modified setting where M has iid entries.

We now study all vectors $v \in \mathbb{S}^{n-1}$ at a given scale $\varepsilon \geqslant e^{-cn}$. While this is an uncountable set, we build an efficient ε-net for these vectors in two steps. We first discretize the whole sphere by taking an ε-net for $\mathbb{S}^{n-1}$, which we call Λ_ε. We can then say something like

$$\mathbb{P}(Av = 0 \text{ for some } v \text{ at scale } \varepsilon) \leqslant \mathbb{P}\left(\|Mv\|_2 \leqslant \varepsilon\sqrt{n} \text{ for some } v \in \Lambda_\varepsilon\right).$$

One's first instinct might be to simply union bound over all $v \in \Lambda_\varepsilon$; however it turns out that even the most efficient epsilon nets have $|\Lambda_\varepsilon| \approx (C/\varepsilon)^n$, which is too large to say anything.

The key insight here is that most of Λ_ε is not used when approximating $v \in \mathbb{S}^{n-1}$ *at scale* ε and so we can refine our net Λ_ε by discarding all vectors $w \in \Lambda_\varepsilon$ with $\mathbb{P}(\|Mv\|_2 \leqslant \varepsilon\sqrt{n}) \ll (L\varepsilon)^n$. So if we let $\mathcal{N}_\varepsilon \subseteq \Lambda_\varepsilon$ be the collection of vectors with $\mathbb{P}(\|Mv\|_2 \leqslant \varepsilon\sqrt{n}) \geqslant (L\varepsilon)^n$, our problem reduces to showing that

$$|\mathcal{N}_\varepsilon| \leqslant L^{-2n}|\Lambda_\varepsilon| \leqslant \left(\frac{C}{L^2\varepsilon}\right)^n, \tag{2.6}$$

which brings us to the technical heart of this chapter (see Theorem 2.7.1). We point out that the factor of L^{-2n}, rather than L^{-n}, in (2.6) is important for us as it allows us to drown out the L^n coming from (2.5) *and* the factor C^n in (2.6), when we union bound over $\mathcal{N}_\varepsilon$.

To prove (2.6) we take a probabilistic perspective inspired by [165]; although we stress that our methods are considerably different. To show (2.6) it is enough to show, for $v \in \Lambda_\varepsilon$ chosen uniformly at random

$$\mathbb{P}_{v\in\Lambda_\varepsilon}\left(v \in \mathcal{N}_\varepsilon\right) \approx \mathbb{P}_{v\in\Lambda_\varepsilon}\left(\mathbb{P}_M\left(\|Mv\|_2 \leqslant \varepsilon\sqrt{n}\right) \geqslant (L\varepsilon)^n\right) \leqslant L^{-2n}, \tag{2.7}$$

(see Lemma 2.7.3, for the rigorous statement). To get a feel for how we tackle this, let us consider the event $\|Mv\|_2 \leqslant \varepsilon n^{1/2}$. Indeed recalling (2.4), the definition of M, we have that

$$Mv = \begin{bmatrix} H_1 v_{[d]} \\ H_1^T v_{[d+1,n]} \end{bmatrix}$$

and so to control the event $\|Mv\|_2 \leqslant \varepsilon\sqrt{n}$, it is enough to control the intersection of events $\|H_1 v_{[d]}\|_2 \leqslant \varepsilon n^{1/2}$ and $\|H_1^T v_{[d+1,n]}\|_2 \leqslant \varepsilon n^{1/2}$. Note that if we simply ignore the second event and bound

$$\mathbb{P}(\|Mv\|_2 \leqslant \varepsilon n^{1/2}) \leqslant \mathbb{P}(\|H_1 v_{[d]}\|_2 \leqslant \varepsilon n^{1/2}),$$

we land in a situation very similar to previous works; where half of the matrix is neglected entirely and we are thus limited by the $(n \log n)^{1/2}$ obstruction, described above. So to overcome this barrier, we need to control these two events simultaneously.

The key idea here is to use the *randomness in* H_1 to control the event $\|H_1 v_{[d]}\|_2 \leqslant \varepsilon n^{1/2}$ and to use the *randomness in* $v \in \Lambda_\varepsilon$ to control the event $\|H_1^T v_{[d+1,n]}\|_2 \leqslant \varepsilon n^{1/2}$. To get this to work, we crucially partition the outcomes in H_1, based on a robust notion of rank. Indeed, let

$$\mathcal{E}_k = \left\{ H_1 : \sigma_{d-k}(H_1) \geqslant c\sqrt{n} \text{ and } \sigma_{d-k+1}(H_1) < c\sqrt{n} \right\},$$

where $\sigma_1(H_1) \geqslant \cdots \geqslant \sigma_d(H_1)$ denote the singular values of H_1. We may then bound $\mathbb{P}_M(\|Mv\|_2 \leqslant \varepsilon n^{1/2})$ above by (only using the randomness in H_1, for the moment)

$$\sum_{k=0}^{d} \mathbb{P}_{H_1}\left(\|H_1^T v_{[d+1,n]}\|_2 \leqslant \varepsilon n^{1/2} \,\middle|\, \|H_1 v_{[d]}\|_2 \leqslant \varepsilon n^{1/2}, \mathcal{E}_k \right) \cdot \mathbb{P}_{H_1}\left(\|H_1 v_{[d]}\|_2 \leqslant \varepsilon n^{1/2}, \mathcal{E}_k \right). \quad (2.8)$$

It is here that we can see the link with our inverse Littlewood-Offord theorem, Theorem 2.1.2, which we use (after a good deal of preparation) to bound the probabilities

$$\mathbb{P}_{H_1}(\|H_1 v_{[d]}\|_2 \leqslant \varepsilon\sqrt{n}, \mathcal{E}_k),$$

that appear in (2.8). The event $\|H_1 v_{[d]}\|_2 \leqslant \varepsilon n^{1/2}$ corresponds to the "hard" constraint $|\langle \tau, v \rangle| \leqslant t$ in Theorem 2.1.2, while the event $\mathcal{E}_k$ corresponds to the "soft" constraint $\|W\tau\|_2 \leqslant c_2\sqrt{k}$, where we think of τ as a single row of H_1. And so, after a certain amount of work with Theorem 2.1.2, we are able to conclude that

$$\mathbb{P}_{H_1}(\|H_1 v_{[d]}\|_2 \leqslant \varepsilon\sqrt{n}, \mathcal{E}_k) \leqslant (C\varepsilon e^{-ck})^{n-d} \quad (2.9)$$

unless $v_{[d]}$ is structured, in which case we do something different (and substantially easier). Thus, for all non-structured v, we have (2.8) is roughly at most

$$(C\varepsilon)^{n-d}\sum_{k=0}^{d} e^{-ck(n-d)}\mathbb{P}_{H_1}\Big(\|H_1^T v_{[d+1,n]}\|_2 \leqslant \varepsilon n^{1/2} \,\Big|\, \|H_1 v_{[d]}\|_2 \leqslant \varepsilon n^{1/2}, \mathcal{E}_k\Big). \tag{2.10}$$

Up to this point, we have not appealed to the randomness in the choice of $v \in \Lambda_\varepsilon$, beyond demanding that v is non-structured. To see how we might take advantage of the randomness in v, let us consider the first moment of the quantity $\mathbb{P}_M(\|Mv\|_2 \leqslant \varepsilon n^{1/2})$, which we view as a random variable in v. Indeed, for $v \in \Lambda_\varepsilon$ taken uniformly at random, we show that

$$\mathbb{E}_{v\in\Lambda_\varepsilon}\,\mathbb{P}_{H_1}\left(\|H_1^T v_{[d+1,n]}\|_2 \leqslant \varepsilon\sqrt{n} \mid \|H_1 v_{[d]}\|_2 \leqslant \varepsilon n^{1/2}, \mathcal{E}_k\right) \leqslant (C\varepsilon)^{d-k}\,. \tag{2.11}$$

We establish this bound by swapping expectations, and bounding the probabilities

$$\mathbb{P}_{v_{[d+1,n]}}(\|H_1^T v_{[d+1,n]}\|_2 \leqslant \varepsilon n^{1/2})\,, \tag{2.12}$$

where H_1 is a *fixed* matrix satisfying $\mathcal{E}_k \cap \{H_1 : \|H_1 v_{[d]}\|_2 \leqslant \varepsilon n^{1/2}\}$. The idea here is that since H_1 has $d-k$ singular values of size $\approx n^{1/2}$, we should expect

$$\mathbb{P}_{v_{[d+1,n]}}(\|H v_{[d+1,k]}\|_2 \leqslant \varepsilon n^{1/2}) \approx (C\varepsilon)^{d-k}, \tag{2.13}$$

which is suggested, for example, by a Gaussian heuristic. This then results in the bound at (2.11). See Section 2.7.2 for details on this step. Putting (2.11) and (2.10) together, and using that $\varepsilon > e^{-cn}$, we have

$$\mathbb{E}_v\,\mathbb{P}_M(\|Mv\|_2 \leqslant \varepsilon n^{1/2}) \leqslant (C\varepsilon)^n.$$

Observe that the loss from the rank at (2.13) is compensated by the gain afforded by the extra constraint added to our Littlewood-Offord step.

While this is a good bound on the expectation, this is *not* enough for our purposes, as the first moment only tells us, via Markov, that

$$\mathbb{P}_{v\in\Lambda_\varepsilon}\left(\mathbb{P}_M\left(\|Mv\|_2 \leqslant \varepsilon n^{1/2}\right) \geqslant (L\varepsilon)^n\right) \leqslant L^{-n}\,,$$

falling short of our desired L^{-2n} bound.

So to prove our result, we instead study[2] the *second moment* of $\mathbb{P}_M(\|Mv\|_2 \leqslant \varepsilon n^{1/2})$,

$$\mathbb{E}_v\left(\mathbb{P}_M(\|Mv\|_2 \leqslant \varepsilon n^{1/2})\right)^2,$$

[2] Actually, we need a slight variant, where we cut out structured vectors.

in much the same way, but with a few added technicalities.

To say a few words about how the second moment is different, we will see (Fact 2.7.7)

$$\left(\mathbb{P}_M\left(\|Mv\|_2 \leqslant \varepsilon n^{1/2}\right)\right)^2 \leqslant \mathbb{P}(\|H_1 v_{[d]}\|_2 \leqslant \varepsilon n^{1/2}, \|H_2 v_{[d]}\|_2 \leqslant \varepsilon n^{1/2} \text{ and } \|H^T v_{[d+1,n]}\|_2 \leqslant 2\varepsilon n^{1/2}),$$

where H_2 is an independent copy of H_1 and $H := [H_1, H_2]$ is the concatenation of these two matrices. We then proceed in much the same way as above, but treating H, in place of H_1, and carrying forward the two "hard" constraints resulting from the two copies of $v_{[d]}$. This explains the shape of our "real" inverse Littlewood-Offord theorem, Lemma 2.3.1, where we allow for these two hard constraints. Ultimately, we arrive at the bound

$$\mathbb{E}_v\left(\mathbb{P}_M(\|Mv\|_2 \leqslant \varepsilon n^{1/2})\right)^2 \leqslant (C\varepsilon)^{2n},$$

which implies the desired conclusion at (2.7).

2.1.2 A few remarks on presentation

This chapter is roughly divided into three parts. The first part consists of Sections 2.3-2.5 which are dedicated to proving our conditioned inverse Littlewood-Offord result, Lemma 2.3.1, which is the "real" version of Theorem 2.1.2. These sections lay the groundwork for Section 2.6, where we prove Theorem 2.6.1, which is the only result we carry forward into later sections.

The second part consists of Section 2.7 and Section 2.8. In Section 2.7, we obtain our crucial bound on the size of our net $\mathcal{N}_\varepsilon$ using our novel "inversion of randomness" technique, as outline above. On the other hand, Section 2.8 contains the less exciting proof that $\mathcal{N}_\varepsilon$ is in fact a net for Σ_ε.

In the final section, Section 2.9, we pull together the various elements of this chapter, state the reductions we will use from previous work and prove Theorem 2.1.1.

In most cases, we have highlighted the main results of each section at the start. So if one does not want to delve into the details of a particular element of the proof, one can simply inspect the top of the section to glean what is needed for going forward.

2.2 Central Definitions

We now turn to give a proper treatment of the proof, by laying out the key definitions that will concern us in this chapter. We begin by partitioning the sphere $\mathbb{S}^{n-1}$ into "structured" and

"unstructured" vectors. Formally, we set $\gamma = e^{-cn}$, for sufficiently small $c > 0$, and then define the "structured" vectors as

$$\Sigma := \left\{ v \in \mathbb{S}^{n-1} : \rho(v) \geqslant \gamma \right\}.$$

The invertibility of a random symmetric matrix on the set of "unstructured" vectors $v \in \mathbb{S}^{n-1} \setminus \Sigma$ is already well understood and so we can restrict our attention to this set of structured vectors. We refer the reader to Section 2.9 for the details here.

Following Rudelson and Vershynin [129], we make a further reduction to working with vectors that are reasonably "flat" on a large part of their support. For $D \subseteq [n]$, with $|D| = d$, we define

$$\mathcal{I}(D) := \left\{ v \in \mathbb{S}^{n-1} : (\kappa_0 + \kappa_0/2)n^{-1/2} \leqslant |v_i| \leqslant (\kappa_1 - \kappa_0/2)n^{-1/2} \text{ for all } i \in D \right\}, \tag{2.14}$$

where $0 < \kappa_0 < 1 < \kappa_1$ are absolute constants, fixed throughout this chapter and defined in Section 2.2.1. We will set $d := c_0^2 n/2$, where c_0 is defined below in Section 2.2.1.

Now set

$$\mathcal{I} := \bigcup_{D \subseteq [n], |D| = d} \mathcal{I}(D).$$

The case of non-flat v is already taken care of in the work of Vershynin [166] (see Section 2.9) and so it is enough to work with $\mathcal{I} \cap \Sigma$. Since we will ultimately union bound over D, it is enough to work with $\mathcal{I}(D) \cap \Sigma$, for *some* fixed set D, and so, by symmetry it is enough to restrict our attention to vectors $v \in \mathcal{I}([d]) \cap \Sigma$.

Now, with this in mind, we further partition the set $\mathcal{I}([d]) \cap \Sigma \subseteq \mathbb{S}^{n-1}$, but for this we need to introduce another distribution on symmetric matrices. Define the probability space $\mathcal{M}_n(\mu)$ by defining $M \sim \mathcal{M}_n(\mu)$ to be the random matrix

$$M = \begin{bmatrix} \mathbf{0}_{[d]\times[d]} & H_1^T \\ H_1 & \mathbf{0}_{[d+1,n]\times[d+1,n]} \end{bmatrix},$$

where H_1 is a $(n-d) \times d$ random matrix with i.i.d. entries that are μ-lazy (that is, $(H_1)_{i,j} = 0$ with probability $1 - \mu$ and $(H_1)_{i,j} = \pm 1$ with probability $\mu/2$). In fact, we will fix $\mu = 1/4$ throughout this chapter.

Now, given $v \in \mathcal{I}([d])$ and $L > 0$, in the spirit of [165], we define the *threshold*

$$\mathcal{T}_L(v) = \sup \left\{ t \in [0,1] : \mathbb{P}(\|Mv\|_2 \leqslant t\sqrt{n}) \geqslant (4Lt)^n \right\},$$

or the "scale" of v, as we called it in Section 2.1.1. Observe carefully here that we are defining $\mathcal{T}_L$ *relative to the matrix* M, rather than our original distribution A.

We may now define our partition of $\mathcal{I}([d]) \cap \Sigma$. For $\varepsilon \in (0,1)$, let

$$\Sigma_\varepsilon := \{v \in \mathcal{I}([d]) : \mathcal{T}_L(v) \in [\varepsilon, 2\varepsilon]\} .$$

We shall show (as it is not obvious) that indeed

$$\Sigma \cap \mathcal{I}([d]) \subseteq \bigcup_{\varepsilon > \gamma^4} \Sigma_\varepsilon.$$

With the definition of Σ_ε in hand, we are able to define $\mathcal{N}_\varepsilon$ which will be an efficient net for Σ_ε at scale ε. It turns out that *defining* this net is not hard, although showing that it satisfies the desired properties will be the main challenge of this chapter. For this, we first define the *trivial net at scale* ε to be[3]

$$\Lambda_\varepsilon := B_n(0,2) \cap \left(4\varepsilon n^{-1/2} \cdot \mathbb{Z}^n\right) \cap \mathcal{I}'([d]),$$

which is a natural net for $\mathcal{I}([d])$. Where $\mathcal{I}'(D)$ is similar to $\mathcal{I}(D)$ but with slightly looser constraints and relative to $\mathbb{R}^n$;

$$\mathcal{I}'(D) := \left\{v \in \mathbb{R}^n : \kappa_0 n^{-1/2} \leqslant |v_i| \leqslant \kappa_1 n^{-1/2} \text{ for all } i \in D\right\}.$$

Since we are only interested in approximating vectors in Σ_ε, we can get away with a significantly more efficient net. For this we introduce two more concentration functions. First, we define the *Lévy concentration function*: if X is a random vector taking values in $\mathbb{R}^n$, define

$$\mathcal{L}(X,t) := \max_{w \in \mathbb{R}^n} \mathbb{P}\left(\|X - w\|_2 \leqslant t\right).$$

Second, we define a variant of this concentration function for the uniform distribution on random symmetric matrices with capped operator norm[4].

$$\mathcal{L}_{A,op}(v,t) := \max_{w \in \mathbb{R}^n} \mathbb{P}\left(\{\|Av - w\|_2 \leqslant t\} \cap \{\|A\| \leqslant 4\sqrt{n}\}\right).$$

Here we are just cutting out the slightly irritating event that A has large operator norm. Intuitively this is an acceptable move as the probability that $\|A\| \geqslant 4\sqrt{n}$, is exponentially small (see Lemma 2.9.5), however some care is needed as we are mostly concerned with far less likely events.

We now introduce our nets $\mathcal{N}_\varepsilon$,

$$\mathcal{N}_\varepsilon := \left\{v \in \Lambda_\varepsilon : \mathbb{P}(\|Mv\|_2 \leqslant 4\varepsilon\sqrt{n}) \geqslant (L\varepsilon)^n \text{ and } \mathcal{L}_{A,op}(v, \varepsilon\sqrt{n}) \leqslant (2^8 L\varepsilon)^n\right\}.$$

[3]Here and throughout, $B_n(x,r)$ is the ℓ^2 ball centered at x with radius r.
[4]For a matrix A, we use the notation $\|A\| := \max_{x:\|x\|_2=1} \|A\|_2$ to denote the usual $2 \to 2$ operator norm.

The reader should view the lower bound $\mathbb{P}(\|Mv\|_2 \leqslant 4\varepsilon\sqrt{n}) \geqslant (L\varepsilon)^n$ as the real core of this definition, while the upper bound for $\mathcal{L}_{A,op}$ is less important. The two main tasks of this chapter will be to show that $\mathcal{N}_\varepsilon$ is indeed a net for Σ_ε (an easier task) and secondly that $|N_\varepsilon|/|\Lambda_\varepsilon|$ is smaller than $\approx L^{-2n}$, where L is a large constant.

2.2.1 Discussion of constants and parameters

We will treat the constants κ_0, κ_1 (seen at (2.14)) as absolute throughout the chapter, and we allow other absolute constants $C, C', \cdots$ to depend on these exact quantities. In particular, we set $\kappa_0 = \rho$ and $\kappa_1 = \delta^{-1/2}/2$, where δ, ρ are as in Lemma 2.9.2 (which is a lemma from [166]). While we have not computed these constants, it would not be too much work to do so.

We also note our treatment of c_0, which, for most of the chapter, will be presented as a parameter and dependencies involving c_0 will be explicitly noted. However, we will ultimately fix $c_0 = \min\{2^{-24}, \rho\delta^{1/2}\}$ where, again, δ, ρ are as in Lemma 2.9.2. Thus it is no harm for the reader to view c_0 as an absolute constant which is fixed throughout the chapter. The reason for the extra care with c_0 comes from its delicate relationship to d/n. Indeed, we will ultimately set $d := \lceil c_0^2 n/2 \rceil$.

Another point to note is our use of R, which represents related, but different constants throughout the chapter. Roughly speaking, these related values of R increase as we get deeper into the proof.

2.3 Inverse Littlewood-Offord for conditioned random walks I: Statement of result and setting up the proof

This section is the first of three sections where we lay out and prove our main inverse Littlewood-Offord type theorem, Lemma 2.3.1, which works in the presence of a large number ($k \approx n$) of relatively soft constraints on our random walk. As mentioned before, the conclusion of our Littlewood-Offord theorem will be similar to that of Rudelson and Vershynin [129], who showed that vectors v, for which the random walk $\langle v, \tau \rangle$ concentrates, admit non-trivial least common denominators. As we will see, the proof of Lemma 2.3.1 is rather involved and consists mainly of a geometric argument on the Fourier side to "decouple" the many soft constraints from the few hard constraints.

Given a $2d \times \ell$ matrix W (which encodes these soft constraints on our walk, as in Theorem 2.1.2) and a vector $Y \in \mathbb{R}^d$, we define the Y-augmented matrix W_Y as

$$W_Y = \left[W, \begin{bmatrix} \mathbf{0}_d \\ Y \end{bmatrix}, \begin{bmatrix} Y \\ \mathbf{0}_d \end{bmatrix} \right]. \tag{2.15}$$

Here $Y \approx v/t$ will be a re-scaled version of v from Theorem 2.1.2.

We let $\|A\|_{\mathrm{HS}}$ denote the Hilbert-Schmidt norm of a matrix A, that is, $\|A\|_{\mathrm{HS}}^2 := \sum_{i,j} |A_{i,j}|^2$ and for $\mu \in (0,1)$, $m \in \mathbb{N}$, define the m-dimensional μ-lazy random vector $\tau \sim \mathcal{Q}(m,\mu)$ to be the vector with independent entries $(\tau_i)_{i=1}^m$, satisfying

$$\mathbb{P}(\tau_i = -1) = \mathbb{P}(\tau_i = +1) = \mu/2 \qquad \text{and} \qquad \mathbb{P}(\tau_i = 0) = 1 - \mu.$$

We now state our main inverse Littlewood-Offord type theorem, which is our "real" (and strengthened) version of Theorem 2.1.2, from Section 1.1.1.

Lemma 2.3.1. *For $d \in \mathbb{N}$ and $\alpha \in (0,1)$, let $0 \leqslant k \leqslant 2^{-10}\alpha d$ and $t \geqslant \exp(-2^{-9}\alpha d)$. For $0 < c_0 \leqslant 2^{-24}$, let $Y \in \mathbb{R}^d$ satisfy $\|Y\|_2 \geqslant 2^{-10}c_0/t$, let W be a $2d \times k$ matrix with $\|W\| \leqslant 2$ and $\|W\|_{\mathrm{HS}} \geqslant \sqrt{k}/2$.*

If $\tau \sim \mathcal{Q}(2d, 1/4)$ and $D_\alpha(Y) > 16$ then

$$\mathcal{L}\left(W_Y^T \tau, c_0^{1/2}\sqrt{k+1}\right) \leqslant (Rt)^2 \exp(-c_0 k), \tag{2.16}$$

where $R = 2^{32}c_0^{-2}$.

Before we start working towards the proof of Lemma 2.3.1, we make a few informal remarks on its statement and its connection to Theorem 2.1.2. The main difference to note is that there are now two "hard" constraints encoded in the left-hand side of (2.16); these are, in the notation of Theorem 2.1.2,

$$|\langle (v, 0_{[d]}), \tau \rangle| < t \text{ and } |\langle (0_{[d]}, v), \tau \rangle| < t.$$

The "soft" constraints are now encoded as the columns $w_1, \ldots, w_k$ of W.

To combine the "hard" and "soft" constraints into a single matrix inequality, we rescale v, thinking of $|\langle (v, 0_{[d]}), \tau \rangle| < t$ as $|\langle c_0^{1/2} t^{-1} (v, 0_{[d]}), \tau \rangle| < c_0^{1/2}$. This explains the scaling on Y, which is unusually written as $\|Y\|_2 \geqslant 2^{-10}c_0/t$, where t should be thought of a very small number $\approx e^{-cn}$.

The scaling of $D_\alpha(Y)$ in Lemma 2.3.1, in contrast with the statement of Theorem 2.1.2, is explained in a similar way. If $\phi \cdot Y \sim \mathbb{Z}^d$, where $\phi = O(1)$ then $(\phi/t) = O(1/t)$ satisfies $(\phi/t) \cdot v \sim \mathbb{Z}^d$, as we think of $Y \approx v/t$.

This also makes the numerology of Lemma 2.3.1 a little more transparent. If Y is a random vector with $\|Y\|_2 \approx 1/t$, we have $|Y_i| \approx t^{-1}n^{-1/2}$ and thus we expect the one dimensional random walk $\langle Y, \tau\rangle$ to have

$$\mathcal{L}\left(\langle Y, \tau\rangle, c_0^{1/2}\right) \approx t.$$

Thus we expect Y to have some special structure if $\mathcal{L}\left(\langle Y, \tau\rangle, c_0^{1/2}\right) \gg t$. On the other hand, for each w_i we expect that $|\langle w_i, \tau\rangle| \approx 1$ and, since the w_i must be "approximately orthogonal" (due to the assumption $\|W\| \leqslant 2$), we should expect

$$\mathcal{L}\left(W\tau, c_0^{1/2}\sqrt{k}\right) \approx e^{-ck},$$

being somewhat vague about this constant $c > 0$. Second, note that Lemma 2.3.1 is still interesting even in the case $k = 0$, where it is not hard to see that it reduces to

$$\mathcal{L}\left(\langle Y, \tau\rangle,\, c_0^{1/2}\right) \leqslant Rt,$$

whenever $D_\alpha(Y) \leqslant 16$, which is essentially the statement of the main inverse Littlewood-Offord theorem of Rudelson and Vershynin in [129].

Finally, we point out that the contrapositive of Lemma 2.3.1 is more conducive to the "inverse Littlewood-Offord" reading:

$$\text{if} \quad \mathcal{L}(W_Y^T\tau, c_0^{1/2}\sqrt{k+1}) \geqslant (Rt)^2\exp(-c_0k) \quad \text{then} \quad D_\alpha(Y) \leqslant 16.$$

For the remainder of this section, we take some first steps towards the proof of Lemma 2.3.1. We first pass to the Fourier side and set up our problem there, describing our goal in terms of a certain "level set". We then make a first reduction, by getting some basic control on the fibers of this level set. In the following section, Section 2.4, we make a more significant reduction about the geometry of our level set. In Section 2.5 we prove the key Lemma 2.5.1, the statement of which is very similar to that of Lemma 2.3.1, but with a more complicated quantity replacing the right-hand side of (2.16). Finally, with one further step, we conclude Section 2.5, with the proof of Lemma 2.3.1.

2.3.1 Passing to the Fourier side

To prove Lemma 2.3.1 we will prove the contrapositive; assume (2.16) fails and then obtain an upper bound on the least common denominator by finding a non-trivial $\phi > 0$ that satisfies $\phi = O(1)$ and $\|\phi \cdot Y\|_{\mathbb{T}} \leqslant \sqrt{\alpha d}$. Our first step in proving Lemma 2.3.1 is to use the lower

bound in the negation of (2.16) to obtain a lower bound on a level set of an appropriate Fourier transform. This manoeuvre was pioneered by Halász [79] and has been a key step in all of the Fourier approaches to inverse Littlewood-Offord theory.

For a $2d \times \ell$ matrix W, we define the *W-level set*, for $t \geqslant 0$, to be

$$S_W(t) := \left\{\theta \in \mathbb{R}^\ell : \|W\theta\|_{\mathbb{T}} \leqslant \sqrt{t}\right\}$$

and we define γ_ℓ to be the ℓ dimensional Gaussian measure defined by $\gamma_\ell(S) = \mathbb{P}(g \in S)$, where $g \sim \mathcal{N}(0, (2\pi)^{-1} I_\ell)$ and I_ℓ denotes the $\ell \times \ell$ identity matrix.

The following Esseen-type lemma, allows us relate the quantity seen at the left-hand side of (2.16) with the Gaussian volume of a level-set.

Lemma 2.3.2. *Let $\beta > 0$, $\nu \in (0, 1/4]$, let W be a $2d \times \ell$ matrix and let $\tau \sim \mathcal{Q}(2d, \nu)$. Then there exists $m > 0$ so that*

$$\mathcal{L}(W^T\tau, \beta\sqrt{\ell}) \leqslant 2\exp\left(2\beta^2\ell - \nu m/2\right)\gamma_\ell(S_W(m)).$$

The proof of this Lemma is a straightforward exercise with the characteristic function of $W^T\tau$ and is postponed to Appendix A.

We can now describe how our least common denominator can be spotted in Fourier space. From Lemma 2.3.2 along with the negation of (2.16), we obtain $m > 0$ and a set $S_{W_Y}(m) \subseteq \mathbb{R}^{k+2}$ with Gaussian volume bounded below by $(Rt)^2\exp(c_1 m - c_2 k)$. Now, for reasons that we will not explain here (since it is just a consequence of the Fourier transform), the first k-coordinates of the space, correspond to the k "soft" constraints while the final two coordinates correspond to the two "hard" constraints.

With this in mind, the idea is to find an element $\psi \in S_{W_Y}(m)$ for which $\|\psi_{[k]}\|_2 = O(\sqrt{k})$, and one of ψ_{k+1}, ψ_{k+2} is $O(1)$ and "non-trivial". It will turn out that one of ψ_{k+1}, ψ_{k+2} is a good candidate for our desired least common denominator. The condition on the $\psi_{[k]}$ should be thought of as just getting these coordinates "out of the way".

To find this desired $\psi \in S_{W_Y}(m)$, for $r, s > 0$, we define the *cylinder*

$$\Gamma_{r,s} := \left\{\theta \in \mathbb{R}^{k+2} : \left\|\theta_{[k]}\right\|_2 \leqslant r, |\theta_{k+1}| \leqslant s \text{ and } |\theta_{k+2}| \leqslant s\right\}. \tag{2.17}$$

We now restate our condition on ψ in terms of $\Gamma_{r,s}$: we want to show that there exists an $x \in S_{W_Y}(m)$ for which

$$(\Gamma_{2\sqrt{k},16} \setminus \Gamma_{2\sqrt{k},s} + x) \cap S_{W_Y}(m) \neq \emptyset, \tag{2.18}$$

where s is chosen depending on the non-triviality condition we need. We shall then ultimately see that if $y \in (\Gamma_{2\sqrt{k},16} \setminus \Gamma_{2\sqrt{k},s} + x)$, where $x \in S_{W_Y}(m)$, then $(x - y)$ is a good candidate for ψ (see Claims 2.5.4-2.5.6). In what remains in this section, we warm up by making a first easy reduction on the structure of $S_{W_Y}(m)$ under the assumption that (2.18) fails.

2.3.2 A first reduction: controlling the density on fibers

For our first reduction, we first record the following easy fact.

Fact 2.3.3. *For $s > 0$, let $S \subseteq \mathbb{R}^2$ be such that $\gamma_2(S) \geqslant 8s^2$, then there exists $x, y \in S$ so that $s < \|x - y\|_\infty \leqslant 16$.*

Proof. We prove the contrapositive and assume there is no pair $x, y \in S$ with $s < \|x-y\|_\infty \leqslant 16$. We cover $\mathbb{R}^2 = \bigcup_{p \in 16\cdot\mathbb{Z}^2} Q_p$ where $Q_p := p + [-8, 8]^2$. Thus $\gamma_2(S) \leqslant \sum_{p\in 16\cdot\mathbb{Z}^2} \gamma_2(S \cap Q_p)$. Since there is no $x, y \in S$ so that $s < \|x - y\|_\infty \leqslant 16$, then for each Q_p there is $x = x(p) \in Q_p$ so that

$$\gamma_2(S \cap Q_p) \leqslant \gamma_2(S \cap Q_p \cap (x(p) + [-s, s]^2)) \leqslant \gamma_2(x(p) + [-s, s]^2).$$

Letting $g \sim \mathcal{N}(0, (2\pi)^{-1})$, we have

$$\gamma_2(x + [-s, s]^2) \leqslant \mathbb{P}(x_1 - s \leqslant g \leqslant x_1 + s)\mathbb{P}(x_2 - s \leqslant g \leqslant x_2 + s) \leqslant 4s^2 \exp(-\pi\|p\|_2^2/16),$$

where we have used that $(x_i - s)^2 \geqslant p_i^2/8$, which holds since we may assume that $s \leqslant 1$ (else the statement holds trivially). Now we may bound

$$\gamma_2(S) \leqslant \sum_{p\in 16\cdot\mathbb{Z}^2} \gamma_2(S \cap Q_p) \leqslant 4s^2 \sum_{p\in 16\cdot\mathbb{Z}^2} \exp(-\pi\|p\|_2^2/16) < 8s^2,$$

which completes the proof. □

Now for $S \subseteq \mathbb{R}^{k+2}$, and $\theta_{[k]} \in \mathbb{R}^k$, we define the "vertical fiber"

$$S(\theta_{[k]}) := \left\{(\theta_{k+1}, \theta_{k+2}) \in \mathbb{R}^2 : (\theta_{[k]}, \theta_{k+1}, \theta_{k+2}) \in S\right\}. \tag{2.19}$$

The following lemma tells us that if we are unable to find a point in our desired intersection $(\Gamma_{r,16} \setminus \Gamma_{r,s} + x) \cap S$, for all $x \in S$, we can obtain good control on the measure of the vertical fibers of S.

Lemma 2.3.4. *For $k \in \mathbb{N}$, $r > 0$ and $s > 0$, let $S \subset \mathbb{R}^{k+2}$ be such that for all $x \in S$ we have*

$$(\Gamma_{r,16} \setminus \Gamma_{r,s} + x) \cap S = \emptyset.$$

Then

$$\max_{\theta_{[k]}\in\mathbb{R}^k} \gamma_2(S(\theta_{[k]})) \leqslant 8s^2\,.$$

Proof. We prove the contrapositive; let $\psi_{[k]}$ be such that $\gamma_2\left(S(\psi_{[k]})\right) > 8s^2$. This implies (Fact 2.3.3) that there exists $(\theta_{k+1},\theta_{k+2}),(\theta'_{k+1},\theta'_{k+2}) \in S(\psi_{[k]})$ with

$$s \leqslant \max\{|\theta_{k+1}-\theta'_{k+1}|,|\theta_{k+2}-\theta'_{k+2}|\} \leqslant 16\,.$$

Unpacking what this means in the full space $\mathbb{R}^{k+2}$: we have $\theta,\theta' \in S$ so that $\theta_{[k]},\theta'_{[k]} = \psi_{[k]}$, and $s \leqslant \max\{|\theta_{k+1}-\theta'_{k+1}|,|\theta_{k+2}-\theta'_{k+2}|\} \leqslant 16$. Thus

$$\theta \in (\theta' + \Gamma_{r,16} \setminus \Gamma_{r,s}),$$

as desired. □

In the next section we go on to obtain a more complicated reduction of this form, that will ultimately be key in proving Lemma 2.3.1.

2.4 Inverse Littlewood Offord II: A geometric inequality

We now turn to make a more intricate and subtle reduction from that seen in Section 2.3.2, that will be key in finding our least common denominator. The lemma we prove here is purely geometric, but one should always think of it as being applied to an appropriate level set $S = S_{W_Y}(m)$, as seen in Lemma 2.3.2.

Given a set $S \subset \mathbb{R}^{k+2}$ and $y \in \mathbb{R}^{k+2}$, define the "translated horizontal fiber",

$$F_y(S;a,b) := \{\theta_{[k]} = (\theta_1,\dots,\theta_k) \in \mathbb{R}^k : (\theta_1,\dots,\theta_k,a,b) \in S - y\}\,.$$

Our main goal of this section tells us that under the assumption

$$(\Gamma_{2\sqrt{k},16} \setminus \Gamma_{2\sqrt{k},s} + x) \cap S = \emptyset,$$

for all $x \in S$, the total measure of S can be controlled by the measure of the k-dimensional fibers $F_y(S;a,b)$. We state it in the contrapositive form to make the application (in Section 2.5) a little easier to spot.

Lemma 2.4.1. *For $k \in \mathbb{N}$ and $s > 0$, let $S \subset \mathbb{R}^{k+2}$ be a measurable set which satisfies*

$$8s^2e^{-k/8} + 64s^2 \max_{a,b,y}\left(\gamma_k(F_y(S;a,b) - F_y(S;a,b))\right)^{1/4} < \gamma_{k+2}(S)\,. \tag{2.20}$$

Then there is an $x \in S$ so that[5]

$$(\Gamma_{2\sqrt{k},16} \setminus \Gamma_{2\sqrt{k},s} + x) \cap S \neq \emptyset\,. \tag{2.21}$$

To prove this lemma, we will need a few facts about Gaussian space, which we collect in Sections 2.4.1 and 2.4.2, before moving on to prove Lemma 2.4.1 in Section 2.4.3.

2.4.1 A few facts about Gaussian space

Recall that for $\ell \in \mathbb{N}$, γ_ℓ is the ℓ dimensional Gaussian measure defined by $\gamma_\ell(S) = \mathbb{P}(g \in S)$, where $g \sim \mathcal{N}(0, (2\pi)^{-1} I_\ell)$.

Lemma 2.4.2. *Let $k \geqslant 0$, $r > 0$ and $S \subset \mathbb{R}^{k+2}$ be measurable. Then there exists $x \in S$, and $h \in \Gamma_{r,8}$ so that*

$$\gamma_{k+2}(S \cap B) \leqslant 8\gamma_{k+2}((S - x) \cap \Gamma_{2r,16} + h)\,,$$

where $B := \{\theta \in \mathbb{R}^{k+2} : \|\theta_{[k]}\|_2 \leqslant r\}$.

Proof. Consider translates $\Gamma_{r,8} + y$ where $y_{k+1}, y_{k+2} \in 16\mathbb{Z}^2$ to write

$$\gamma_{k+2}(S \cap B) = \sum_{y \in \{0\}^k \times 16\mathbb{Z}^2} \gamma_{k+2}(S \cap (\Gamma_{r,8} + y))\,. \tag{2.22}$$

We express $\gamma_{k+2}(S \cap (\Gamma_{r,8} + y))$ as

$$\int_{\mathbb{R}^{k+2}} \mathbb{1}\big[\theta \in S \cap (\Gamma_{r,8} + y)\big] e^{-\pi\|\theta\|_2^2/2}\, d\theta = \int_{\mathbb{R}^{k+2}} \mathbb{1}\big[\phi \in (S - y) \cap \Gamma_{r,8}\big] e^{-\pi\|\phi + y\|_2^2/2}\, d\phi. \tag{2.23}$$

Rewriting the exponent in the integrand at (2.23)

$$-\|\phi + y\|_2^2 = -\|\phi\|_2^2 - 2\phi_{k+1}y_{k+1} - 2\phi_{k+2}y_{k+2} - y_{k+1}^2 - y_{k+2}^2,$$

we use that $|\phi_{k+1}|, |\phi_{k+2}| \leqslant 8$ whenever $\mathbb{1}[\phi \in (S - y) \cap \Gamma_{r,8}] \neq 0$, to see

$$\gamma_{k+2}(S \cap (\Gamma_{8,r} + y)) \leqslant \exp\Big(-\frac{\pi}{2}y_{k+1}^2 - \frac{\pi}{2}y_{k+2}^2 + 8\pi|y_{k+1}| + 8\pi|y_{k+2}|\Big)\, \gamma_{k+2}((S - y) \cap \Gamma_{r,8})\,. \tag{2.24}$$

So, apply (2.24) to (2.22) to get

$$\begin{aligned}
\gamma_{k+2}(S \cap B) &\leqslant \sum_{y \in \{0\}^k \times 16\mathbb{Z}^2} \gamma_{k+2}((S - y) \cap \Gamma_{r,8}) e^{-\frac{\pi}{2}y_{k+1}^2 - \frac{\pi}{2}y_{k+2}^2 + 8\pi|y_{k+1}| + 8\pi|y_{k+2}|} \\
&\leqslant \max_y \gamma_{k+2}((S - y) \cap \Gamma_{r,8}) \sum_{y_{k+1}, y_{k+2} \in 16\mathbb{Z}} e^{-\frac{\pi}{2}y_{k+1}^2 - \frac{\pi}{2}y_{k+2}^2 + 8\pi|y_{k+1}| + 8\pi|y_{k+2}|} \\
&\leqslant 16 \max_y \gamma_{k+2}((S - y) \cap \Gamma_{r,8})\,.
\end{aligned}$$

[5] Note, in particular, that Lemma 2.4.1 says that *if* (2.20) is satisfied then we must have $s < 16$.

Let y be a vector at which the above maximum is attained. Now observe that if $S \cap (\Gamma_{r,8}+y) = \emptyset$ then $(S-y)\cap\Gamma_{r,8} = \emptyset$ and thus $\gamma_{k+2}(S\cap B) = 0$; so there is nothing to prove. Thus we may assume $S \cap (\Gamma_{r,8}+y) \neq \emptyset$ and let $x \in S \cap (\Gamma_{8,r}+y)$. Define $h := x - y \in \Gamma_{r,8}$ and notice that

$$(S-y)\cap\Gamma_{r,8} - h = (S-y-h)\cap(\Gamma_{r,8}-h) \subseteq (S-x)\cap\Gamma_{2r,16},$$

where the inclusion holds since $h \in \Gamma_{r,8}$. Therefore $(S-y)\cap\Gamma_{r,8} \subseteq (S-x)\cap\Gamma_{2r,16}+h$, allowing us to conclude that

$$\gamma_{k+2}(S\cap B) \leqslant 16\gamma_{k+2}((S-y)\cap\Gamma_{r,8}) \leqslant 16\gamma_{k+2}((S-x)\cap\Gamma_{2r,16}+h),$$

as desired. □

We also need the following standard tail estimate on a k-dimensional Gaussian.

Fact 2.4.3. $\gamma_k\big(\{x\in\mathbb{R}^k : \|x\|_2^2 \geqslant k\}\big) \leqslant \exp(-k/8)$.

Proof. For any $\varepsilon \in (0,1)$ the *standard* Gaussian measure of the set $\{x\in\mathbb{R}^k : \|x\|_2^2 \geqslant k/(1-\varepsilon)\}$ is at most $\exp(-\varepsilon^2 k/4)$. Recalling that γ_k has standard deviation $(2\pi)^{-1/2}$ and taking $\varepsilon = 1-(2\pi)^{-1}$, gives the desired bound. □

2.4.2 A Gaussian Brunn-Minkowski type theorem

We now lay out a useful tool which gives us some control of the Gaussian measure of the sum set $A+B$, relative to the Gaussian measures of A and B. Indeed, the following theorem due to Borell [23], can be viewed as a Brunn-Minkowski-type theorem for Gaussian space.

For this, let $\Phi(x)$ be the cumulative probability function $\Phi(x) := \mathbb{P}(Z \leqslant x)$, for the *standard* one dimensional Gaussian $Z \sim \mathcal{N}(0,1)$, while γ_k is (still) the k-dimensional Gaussian with covariance matrix $(2\pi)^{-1}I_k$.

Theorem 2.4.4 (Borell)**.** *Let $A, B \subseteq \mathbb{R}^k$ be Borel. Then*

$$\gamma_k(A+B) \geqslant \Phi\Big(\Phi^{-1}(\gamma_k(A)) + \Phi^{-1}(\gamma_k(B))\Big).$$

Proof. In [23] Theorem 2.4.4 is proved for the standard Gaussian measure rather than γ_k. However we can change the standard deviation of the measure by taking dilates of the sets A and B. □

We will use the following simple consequence of Theorem 2.4.4.

Lemma 2.4.5. *Let $A \subseteq \mathbb{R}^k$ be Borel. Then*

$$\gamma_k(A - A) \geqslant \gamma_k(A)^4 .$$

Proof. By Theorem 2.4.4, we have

$$\gamma_k(A - A) \geqslant \Phi(2\Phi^{-1}(\gamma_k(A))) = \Phi(2x), \tag{2.25}$$

where we have set $x = \Phi^{-1}(\gamma_k(A))$. Note that

$$\Phi(2x) = \mathbb{P}(Z \leqslant 2x) = \mathbb{P}\left(Z_1 + Z_2 + Z_3 + Z_4 \leqslant 4x\right) \geqslant \mathbb{P}(Z \leqslant x)^4 = \Phi(x)^4 \tag{2.26}$$

where Z_j are i.i.d. copies of $Z \sim \mathcal{N}(0,1)$. Combining (2.25) and (2.26) completes the proof. □

2.4.3 Proof of Lemma 2.4.1

With these pieces now in place, we can move on to prove Lemma 2.4.1, our key geometric lemma on the Fourier side.

Proof of Lemma 2.4.1. Write $r = \sqrt{k}$ for simplicity. We prove the contrapositive and assume for every $x \in S$ we have

$$(\Gamma_{2r,16} \setminus \Gamma_{2r,s} + x) \cap S = \emptyset. \tag{2.27}$$

We define

$$B := \{\theta \in \mathbb{R}^{k+2} : \|\theta_{[k]}\|_2 \leqslant r\}.$$

and proceed to bound $\gamma_{k+2}(S)$ from above by first bounding $\gamma_{k+2}(S \setminus B)$ and then bounding $\gamma_{k+2}(S \cap B)$.

Step 1: Upper bound for $\gamma_{k+2}(S \setminus B)$. For $\theta_{[k]} \in \mathbb{R}^k$, let $S(\theta_{[k]})$ be as defined at (2.19)

$$S(\theta_{[k]}) = \left\{(\theta_{k+1}, \theta_{k+2}) \in \mathbb{R}^2 : (\theta_{[k]}, \theta_{k+1}, \theta_{k+2}) \in S\right\}.$$

We may write

$$\gamma_{k+2}(S \setminus B) = \int_{\|\theta_{[k]}\|_2 \geqslant r} \gamma_2\left(S(\theta_{[k]})\right) \, d\gamma_k \tag{2.28}$$

and thus

$$\gamma_{k+2}(S \setminus B) \leqslant \left(\max_{\theta_{[k]} \in \mathbb{R}^k} \gamma_2\left(S(\theta_{[k]})\right)\right) \gamma_k\left(\{\|\theta_{[k]}\|_2 \geqslant r\}\right). \tag{2.29}$$

Lemma 2.3.4 and (2.27) shows

$$\max_{\theta_{[k]} \in \mathbb{R}^k} \gamma_2\left(S(\theta_{[k]})\right) \leqslant 8s^2. \tag{2.30}$$

Fact 2.4.3 bounds

$$\gamma_k\big(\{\|\theta_{[k]}\|_2 \geqslant r\}\big) \leqslant \exp(-k/8) \tag{2.31}$$

and so from (2.29), (2.30) and (2.31) we learn

$$\gamma_{k+2}(S \setminus B) \leqslant 8s^2 e^{-k/8}. \tag{2.32}$$

Step 2: Upper bound for $\gamma_{k+2}(S \cap B)$. By Lemma 2.4.2, there exists $x \in S$ and $h \in \Gamma_{r,8}$ such that

$$\gamma_{k+2}(S \cap B) \leqslant 16\gamma_{k+2}((S-x) \cap \Gamma_{2r,16} + h). \tag{2.33}$$

Now since we are assuming the claim is false, and $x \in S$, we use (2.27) to deduce that

$$(S-x) \cap \Gamma_{2r,16} \subseteq (S-x) \cap \Gamma_{2r,s} \tag{2.34}$$

and so letting $y = x - h$, we see

$$(S-x) \cap \Gamma_{2r,s} + h = (S - x + h) \cap (\Gamma_{2r,s} + h) = (S-y) \cap (\Gamma_{2r,s} + h). \tag{2.35}$$

Thus by (2.33), (2.34) and (2.35), we have

$$\gamma_{k+2}(S \cap B) \leqslant 16\gamma_{k+2}((S-y) \cap (\Gamma_{2r,s} + h)). \tag{2.36}$$

Bound

$$\gamma_{k+2}((S-y) \cap (\Gamma_{2r,s} + h)) \leqslant \int_{|a-h_{k+1}|,|b-h_{k+2}| \leqslant s} \gamma_k\big(F_y(S;a,b)\big)\, d\gamma_2 \tag{2.37}$$

and apply Lemma 2.4.5 to obtain

$$\gamma_{k+2}((S-y) \cap (\Gamma_{2r,s} + h)) \leqslant 4s^2 \max_{a,b,y} (\gamma_k(F_y(S;a,b) - F_y(S;a,b)))^{1/4}. \tag{2.38}$$

Combining (2.36) and (2.38) gives

$$\gamma_{k+2}(S \cap B) \leqslant 64s^2 \max_{a,b,y} \big(\gamma_k\left(F_y(S;a,b) - F_y(S;a,b)\right)\big)^{1/4} \tag{2.39}$$

Putting Step 1 and Step 2 together : (2.39) together with (2.32) implies

$$\gamma_{k+2}(S) \leqslant 8s^2 e^{-k/8} + 64s^2 \max_{a,b,y} (\gamma_k(F_y(S;a,b) - F_y(S;a,b)))^{1/4},$$

completing the proof of the contrapositive. □

2.5 Inverse Littlewood-Offord III: Comparison to a lazier walk and Proof of Lemma 2.3.1

In Section 2.4 we proved our key geometric ingredient, Lemma 2.4.1, to deal with the geometry of our level set (as seen in Section 2.3.1). We now use this lemma to take the following big step towards Lemma 2.3.1.

Lemma 2.5.1. *For $d \in \mathbb{N}$ and $\alpha \in (0,1)$, let $0 \leqslant k \leqslant 2^{-10}\alpha d$ and $t \geqslant \exp(-2^{-10}\alpha d)$. For $0 < c_0 \leqslant 2^{-24}$, let $Y \in \mathbb{R}^d$ satisfy $\|Y\| \geqslant 2^{-10}c_0/t$ and let W be a $2d \times k$ matrix with $\|W\| \leqslant 2$. Also let $\tau \sim \mathcal{Q}(2d, 1/4)$ and $\tau' \sim \mathcal{Q}(2d, 2^{-9})$ and $\beta \in [c_0/2^{10}, \sqrt{c_0}]$, $\beta' \in (0, 1/2)$.*

If

$$\mathcal{L}(W_Y^T\tau, \beta\sqrt{k+1}) \geqslant (Rt)^2 \exp(4\beta^2 k)\left(\mathbb{P}(\|W^T\tau'\|_2 \leqslant \beta'\sqrt{k}) + \exp(-\beta'^2 k)\right)^{1/4} \tag{2.40}$$

then $D_\alpha(Y) \leqslant 16$. Here we have set $R = 2^{31}/c_0^2$.

Of course, Lemma 2.5.1 looks quite a bit like Lemma 2.3.1 save for quantity

$$\mathbb{P}(\|W^T\tau'\|_2 \leqslant \beta'\sqrt{k}) + \exp(-\beta'^2 k), \tag{2.41}$$

on the right-hand side of (2.40). One should view this quantity as an approximation of the contribution that the "soft" constraints make. Indeed, if one reads this lemma in the contrapositive, it says that we can successfully "decouple" the "soft" constraints from the "hard" constraints, provided Y is sufficiently "unstructured", meaning $D_\alpha(Y) > 16$. Of course, this story is not quite an honest one; we have to use the lazier vector τ', rather than τ, to get things to work out, and we also take a loss in the exponent of $1/4$. The key here is that we obtain the correct power of t in our bound, which is deeply important for our application. We also note that our use of "decoupling" should not be confused with the "decoupling" step in Costello, Tao and Vu [41], which is used to deal with very unstructured vectors.

We prove this lemma in Section 2.5.2 after laying out a few facts on level sets in Section 2.5.1. We will then conclude this section in Section 2.5.3 with a proof of Lemma 2.3.1, by combining Lemma 2.5.1 with one further ingredient to bound (2.41).

2.5.1 Working with level sets

To prepare for the proof of Lemma 2.5.1, we record two basic facts about level sets. First off, we note a sort of converse to the Esseen-type inequality that we saw in Section 2.3, Lemma 2.3.2. Again, we will postpone the straightforward proof of this lemma to Appendix A. Recall that

we defined, for a $2d \times \ell$ matrix W, the *W-level set*, for $t \geqslant 0$, to be

$$S_W(t) := \left\{\theta \in \mathbb{R}^\ell : \|W\theta\|_{\mathbb{T}} \leqslant \sqrt{t}\right\}.$$

Lemma 2.5.2. *Let $\beta > 0, \mu \in (0, 1/4]$, let W be a $2d \times \ell$ matrix, and let $\tau \sim \mathcal{Q}(2d, \mu)$. Then for all $t \geqslant 0$, we have*

$$\gamma_\ell(S_W(t))e^{-32\mu t} \leqslant \mathbb{P}_\tau\left(\|W^T \cdot \tau\|_2 \leqslant \beta\sqrt{\ell}\right) + \exp\left(-\beta^2\ell\right).$$

We need also need the following basic fact about level sets. Recall that, for a set $S \subset \mathbb{R}^{k+2}$ and $y \in \mathbb{R}^{k+2}$, we defined the "translated horizontal fiber",

$$F_y(S; a, b) := \{\theta_{[k]} = (\theta_1, \ldots, \theta_k) \in \mathbb{R}^k : (\theta_1, \ldots, \theta_k, a, b) \in S - y\}.$$

Fact 2.5.3. *For any $2d \times (k+2)$ matrix W. If $m > 0$ we have*

$$S_W(m) - S_W(m) \subseteq S_W(4m).$$

Similarly, for any $y \in \mathbb{R}^{k+2}$ and $a, b \in \mathbb{R}$ we have

$$F_y(S_W(m); a, b) - F_y(S_W(m); a, b) \subseteq F_0(S_W(4m); 0, 0). \tag{2.42}$$

Proof. Notice that if $x, y \in S_W(m)$ then by definition $\|Wx\|_{\mathbb{T}}, \|Wy\|_{\mathbb{T}} \leqslant \sqrt{m}$. Thus, by the triangle inequality,

$$\|W(x-y)\|_{\mathbb{T}} \leqslant \|Wx\|_{\mathbb{T}} + \|Wy\|_{\mathbb{T}} \leqslant 2\sqrt{m}.$$

For (2.42), let $\theta_{[k]}, \theta'_{[k]} \in F_y(S; a, b)$. We have that

$$(\theta_1, \ldots, \theta_k, a, b), (\theta'_1, \ldots, \theta'_k, a, b) \in S_W(m) - y$$

and so $\theta'' := (\theta_1 - \theta'_1, \ldots, \theta_k - \theta'_k, 0, 0) \in S_W(4m)$. Thus $\theta_{[k]} - \theta'_{[k]} \in F_0(S_W(4m); 0, 0)$, implying (2.42). □

2.5.2 Proof of 2.5.1

We may now turn to prove Lemma 2.5.1, our big step towards Lemma 2.3.1.

Proof of Lemma 2.5.1. Apply Lemma 2.3.2 to find $m > 0$ such that the level set

$$S := S_{W_Y}(m) = \{\theta \in \mathbb{R}^{k+2} : \|W_Y\theta\|_{\mathbb{T}} \leqslant \sqrt{m}\},$$

satisfies

$$e^{-\frac{1}{8}m + 2\beta^2 k}\gamma_{k+2}(S) \geqslant \mathcal{L}(W_Y^T\tau, \beta\sqrt{k+1}). \tag{2.43}$$

Thus (2.43) together with our hypothesis (2.40) gives a lower bound

$$\gamma_{k+2}(S) \geqslant \frac{1}{4} e^{\frac{1}{8}m - 2\beta^2 k} \, (Rt)^2 \, T^{1/4}, \tag{2.44}$$

where we have set

$$T := \mathbb{P}(\|W^T \tau'\|_2 \leqslant \beta' \sqrt{k}) + \exp(-\beta'^2 k).$$

We now make the following important designations,

$$r_0 := \sqrt{k} \qquad \text{and} \qquad s_0 := 2^{16} c_0^{-1} (\sqrt{m} + \sqrt{k}) t. \tag{2.45}$$

Recall from (2.17) that for $r, s > 0$ we defined the *cylinder*

$$\Gamma_{r,s} := \left\{ \theta \in \mathbb{R}^{k+2} : \ \left\|\theta_{[k]}\right\|_2 \leqslant r \ \text{ and } |\theta_{k+1}| \leqslant s, \ |\theta_{k+2}| \leqslant s, \right\}.$$

Claim 2.5.4. *There exists $x \in S \subseteq \mathbb{R}^{k+2}$ so that*[6]

$$\left(\Gamma_{2r_0,16} \setminus \Gamma_{2r_0,s_0} + x\right) \cap S \neq \emptyset. \tag{2.46}$$

Proof of Claim 2.5.4. We look to apply Lemma 2.4.1 with $s = s_0$. For this, we bound

$$M := \max_{a,b,y} \left\{ \gamma_k \Big(F_y(S; a, b) - F_y(S; a, b) \Big) \right\},$$

above by $e^{m/4} T$, thus giving a lower bound on $\gamma_{k+2}(S)$ and allowing us to apply Lemma 2.4.1. Use Fact 2.5.3 to see that for any y, a, b, we have

$$F_y(S; a, b) - F_y(S; a, b) \subseteq F_0(S_{W_Y}(4m); 0, 0) \,. \tag{2.47}$$

Now carefully observe that

$$F_0(S_{W_Y}(4m); 0, 0) = \left\{ \theta_{[k]} \in \mathbb{R}^k : \|W\theta_{[k]}\|_{\mathbb{T}} \leqslant \sqrt{4m} \right\} = S_W(4m),$$

which is a level-set corresponding to the ("decoupled") event $\mathbb{P}_{\tau'}(\|W^T\tau'\|_2 \leqslant \beta'\sqrt{k})$, where $\tau' \sim \mathcal{Q}(2d, 2^{-9})$ and $\beta' \in (0, 1/2)$ is as in the hypothesis. Thus we may apply Lemma 2.5.2 along with (2.47) to obtain

$$M \leqslant \gamma_k(F_0(S_{W_Y}(4m), 0, 0)) = \gamma_k(S_W(4m)) \leqslant e^{m/4} T \,. \tag{2.48}$$

So combining (2.48) with (2.44), gives

$$\gamma_{k+2}(S) \geqslant (1/4) e^{m/16 + 2\beta^2 k} (Rt)^2 M^{1/4} \geqslant 8 s_0^2 e^{-k/8} + 64 s_0^2 M^{1/4}, \tag{2.49}$$

allowing us to apply Lemma 2.4.1 and complete the proof of the claim. The last inequality at (2.49) follows from a simple check: each term on the right-hand side of (2.49) is at most half of

[6] Note that this claim shows, in particular, that $s_0 < 16$.

the left-hand side. First note that

$$s_0^2 = 2^{32} c_0^{-2} (\sqrt{m} + \sqrt{k})^2 t^2 < 2^{33} (k+m)(t/c_0)^2 \tag{2.50}$$

and so

$$8 s_0^2 e^{-k/8} \leqslant \frac{1}{8} e^{m/8 - 2\beta^2 k} (Rt)^2 e^{-\beta'^2 k/4}$$

follows from $\beta' \leqslant 1/2$ and the definition of R. On the other hand, use (2.50) to bound

$$64 s_0^2 e^{m/16} \leqslant 2^{39} t^2 c_0^{-2} (2^{20} c_0^{-2} \beta^2 k + 8(m/8)) \leqslant \frac{1}{8} (Rt)^2 e^{m/8 + \beta^2 k}$$

thus showing the second inequality at (2.49) and finishing the proof of the claim. □

We now observe the simple consequence of Claim 2.5.4.

Claim 2.5.5. *We have that* $S_{W_Y}(4m) \cap (\Gamma_{2r_0,16} \setminus \Gamma_{2r_0,s_0}) \neq \emptyset$.

Proof of Claim 2.5.5. By Claim 2.5.4, there exists $x, y \in S = S_{W_Y}(m)$ so that $y \in (\Gamma_{2r_0,16} \setminus \Gamma_{2r_0,s_0} + x) \cap S$. Set $\phi := y - x$ and observe that $\phi \in S_{W_Y}(4m) \cap (\Gamma_{2r_0,16} \setminus \Gamma_{2r_0,s_0})$, by Fact 2.5.3. □

We now conclude the proof of Lemma 2.5.1 with the following claim.

Claim 2.5.6. *If* $\psi \in S_{W_Y}(4m) \cap (\Gamma_{2r_0,16} \setminus \Gamma_{2r_0,s_0})$ *then there exists* $i \in \{k+1, k+2\}$ *so that*

$$\|\psi_i Y\|_{\mathbb{T}} < \min\{\psi_i \|Y\|_2 / 2, \sqrt{\alpha d}\}\,.$$

Proof of Claim 2.5.6. Note that since $\psi \in S_{W_Y}(4m)$ there is a $p \in \mathbb{Z}^{2d}$ so that $W_Y \psi \in B_{2d}(p, 2\sqrt{m})$. So if we express

$$W_Y \psi = W \psi_{[k]} + \psi_{k+1} \begin{bmatrix} Y \\ \mathbf{0}_d \end{bmatrix} + \psi_{k+2} \begin{bmatrix} \mathbf{0}_d \\ Y \end{bmatrix},$$

we have that

$$\psi_{k+1} \begin{bmatrix} Y \\ \mathbf{0}_d \end{bmatrix} + \psi_{k+2} \begin{bmatrix} \mathbf{0}_d \\ Y \end{bmatrix} \in B_{2d}(p, 2\sqrt{m}) - W \psi_{[k]} \subseteq B_{2d}(p, 2\sqrt{m} + 4\sqrt{k}), \tag{2.51}$$

where the last inclusion holds because $\psi \in \Gamma_{2r_0,16}$ and so $\|\psi_{[k]}\|_2 \leqslant 2r_0 \leqslant 2\sqrt{k}$ and $\|W\| \leqslant 2$.

Since $\psi \notin \Gamma_{2r_0,s_0}$ we have that at least one of $|\psi_{k+1}|, |\psi_{k+2}|$ are $> s_0$. So, assume without loss that $|\psi_{k+1}| > s_0$ and that $\psi_{k+1} > 0$ (otherwise replace ψ with $-\psi$). Now project (2.51) onto the first d coordinates, to obtain

$$\psi_{k+1} Y \in B_d(p_{[d]}, 2\sqrt{m} + 4\sqrt{k}). \tag{2.52}$$

We now observe that $\|\psi_{k+1}Y\|_{\mathbb{T}} < \frac{\psi_{k+1}\|Y\|_2}{2}$. Indeed,

$$\frac{\psi_{k+1}\|Y\|_2}{2} > \frac{s_0\|Y\|_2}{2} \geqslant \left(\frac{2^{15}(\sqrt{m}+\sqrt{k})t}{c_0}\right)\left(2^{-10}\frac{c_0}{t}\right) > (2\sqrt{m}+4\sqrt{k}), \tag{2.53}$$

where we have used the definition of s_0 and that $\|Y\|_2 > 2^{-10}c_0/t$.

Finally, we note that $m \leqslant 2^{-4}\alpha d$. To see this, we use (2.44), $\gamma_{k+2}(S) \leqslant 1$ and our lower bound $t \geqslant \exp(-2^{-9}\alpha d)$ to see

$$e^{-m/8} \geqslant \gamma_{k+2}(S)e^{-m/8} \geqslant (Rt)^2 e^{-2\beta'^2 k} \geqslant \exp(-2^{-7}\alpha d),$$

where we have used $k \leqslant 2^{-9}\alpha d$ and $\beta' < 1$ for the last inequality, thus $m \leqslant 2^{-4}\alpha d$. Therefore from (2.52) and (2.53) we have

$$\|\psi_{k+1}Y\|_{\mathbb{T}} \leqslant 2\sqrt{m} + 4\sqrt{k} \leqslant \sqrt{\alpha d},$$

as desired. This completes the proof of the Claim 2.5.6. □

Let ψ and $i \in \{k+1, k+2\}$ be as guaranteed by Claim 2.5.6. Then $\psi_i \leqslant 16$, and

$$\|\psi_i Y\|_{\mathbb{T}} < \min\{\|\psi_i Y\|_2/2, \sqrt{\alpha d}\},$$

and so $D_\alpha(Y) \leqslant 16$ thus completing the proof of Lemma 2.5.1. □

2.5.3 Proof of Lemma 2.3.1

Before turning to prove Lemma 2.3.1, we require one further result which tells us that $\|W\sigma\|_2$ is anti-concentrated when σ is a random vector and W is a fixed matrix. While there are several interesting results of this type in the literature [64, 79, 132] (and we will encounter another in Subsection 2.7.2), we state here a variant of the Hanson-Wright inequality with an explicit constant. A proof can be found in Appendix D in [33], the arXiv version of this paper.

Lemma 2.5.7. *For $d \in \mathbb{N}$, $\nu \in (0,1)$, let $\delta \in (0, \sqrt{\nu}/16)$, let $\sigma \sim \mathcal{Q}(2d, \nu)$, and let W be a $2d \times k$ matrix satisfying $\|W\|_{\mathrm{HS}} \geqslant \sqrt{k}/2$ and $\|W\| \leqslant 2$. Then*

$$\mathbb{P}(\|W^T\sigma\|_2 \leqslant \delta\sqrt{k}) \leqslant 4\exp(-2^{-12}\nu k) \tag{2.54}$$

We now turn to prove Lemma 2.3.1.

Proof of Lemma 2.3.1. Setting $\beta' := 4\sqrt{c_0}$, we look to apply Lemma 2.5.1. For this, note that the hypotheses in Lemma 2.3.1 imply the hypotheses in Lemma 2.5.1 with respect to

c_0, d, α, k, Y, W and τ (and we have the extra condition on $\|W\|_{\mathrm{HS}}$). So if we additionally assume $D_\alpha(Y) > 16$, we may apply Lemma 2.5.1 (in the contrapositive) to obtain

$$\mathcal{L}\left(W_Y^T\tau, \beta\sqrt{k+1}\right) \leqslant (2^{31}c_0^{-2}t/2)^2 e^{4\beta^2 k}\left(\mathbb{P}(\|W^T\tau'\|_2 \leqslant \beta'\sqrt{k}) + e^{-\beta'^2 k}\right)^{1/4}. \tag{2.55}$$

To deal with the right-hand side, we apply Lemma 2.5.7 to take care of the quantity involving $\tau' \in \{-1, 0, 1\}^{2d}$, our $\nu = 2^{-9}$ lazy random vector. Note that $4\sqrt{c_0} \leqslant 2^{-10} \leqslant \sqrt{\nu}/16$, and that our given W satisfies $\|W\|_{\mathrm{HS}} \geqslant \sqrt{k}/2$ and $\|W\| \leqslant 2$. Thus we may apply Lemma 2.5.7, with $\delta = \beta'$ and $\sigma = \tau'$, to see

$$\mathbb{P}(\|W^T\tau'\|_2 \leqslant \beta'\sqrt{k}) \leqslant 4\exp(-2^{-12}\nu k). \tag{2.56}$$

Plugging this into the right-hand side of (2.55) yields

$$\begin{aligned}\exp(4\beta^2 k)\left(\mathbb{P}(\|W^T\tau'\|_2 \leqslant \beta'\sqrt{k}) + \exp(-\beta'^2 k)\right)^{1/4} &\leqslant 2\exp(4c_0 k - 2^{-21}k) + 2\exp(2c_0 k - 4c_0 k)\\ &\leqslant 4\exp(-c_0 k).\end{aligned}$$

Putting this together with (2.55), yields

$$\mathcal{L}\left(W_Y^T\tau, \beta\sqrt{k+1}\right) \leqslant (Rt)^2\exp(-c_0 k),$$

as desired. □

2.6 Inverse Littlewood-Offord for conditioned random matrices

In this section we lift the main result of the previous sections (Lemma 2.3.1) to study the concentration of the vector H_1X, where H_1 is a random $(n-d)\times d$ matrix, conditioned on having k singular values which are much smaller than "typical" and X is a fixed vector for which $|X_i| \approx N$ for each i.

Here N should be thought of as $\approx 1/\varepsilon$, in the context of the proof (see Section 2.1.1) and H_1 comes from its appearance in our matrix M,

$$M = \begin{bmatrix} \mathbf{0}_{[d]\times[d]} & H_1^T \\ H_1 & \mathbf{0}_{[d+1,n]\times[d+1,n]} \end{bmatrix}.$$

The main result of this section is the following theorem[7].

[7] For convenience, we define $\sigma_j(H) = 0$ for $j > \mathrm{rk}(H)$.

Theorem 2.6.1. *For $n \in \mathbb{N}$ and $0 < c_0 \leqslant 2^{-24}$, let $d \leqslant c_0^2 n$, and for $\alpha \in (0,1)$, let $0 \leqslant k \leqslant 2^{-10}\alpha d$ and $N \leqslant \exp(2^{-10}\alpha d)$. Let $X \in \mathbb{R}^d$ satisfy $\|X\|_2 \geqslant c_0 2^{-10} n^{1/2} N$, and let H be a random $(n-d) \times 2d$ matrix with i.i.d. $(1/4)$-lazy entries in $\{-1, 0, 1\}$.*

If $D_\alpha(r_n \cdot X) > 16$ then

$$\mathbb{P}_H\left(\sigma_{2d-k+1}(H) \leqslant c_0 2^{-4}\sqrt{n} \text{ and } \|H_1 X\|_2, \|H_2 X\|_2 \leqslant n\right) \leqslant e^{-c_0 nk/4}\left(\frac{R}{N}\right)^{2n-2d}, \tag{2.57}$$

where we have set $H_1 := H_{[n-d]\times[d]}$, $H_2 := H_{[n-d]\times[d+1,2d]}$, $r_n := \frac{c_0}{16\sqrt{n}}$ and $R := 2^{39} c_0^{-3}$.

To understand the numerology in Theorem 2.6.1, notice that if we only consider the "soft" constraints on the singular values (without the constraints imposed by X) we would expect something like

$$\mathbb{P}_H\left(\sigma_{2d-k+1}(H) \leqslant c_0 2^{-4}\sqrt{n}\right) \approx c^{nk}, \tag{2.58}$$

for some absolute $c \in (0,1)$, which depends on the value of c_0. Here we are using, crucially, that H is a *rectangular* matrix with aspect ratio bounded away from 1. Indeed, if H were a square matrix then $\sigma_{\min}(H) \approx n^{-1/2}$, with high probability[8].

On the other hand, the inverse Littlewood-Offord theorem of Rudelson and Vershynin [129] (with a bit of extra work) tells us that if X is such that $|X_i| \approx N$ for all $i \in [d]$, and

$$\mathbb{P}(\|H_1 X\|_2, \|H_2 X\|_2 \leqslant n) \geqslant \left(\frac{R}{N}\right)^{2n-2d},$$

then $D_\alpha(n^{-1/2}X) = O(1)$. Thus Theorem 2.6.1 is telling us that we maintain an inverse Littlewood-Offord type theorem even in the presence of many additional constraints imposed by the condition on the least singular values.

2.6.1 A tensorization step

We need the following basic fact.

Fact 2.6.2. *If $r \geqslant t > 0$ and X is a random variable taking values in $\mathbb{R}^{k+2}$, then*

$$\mathcal{L}(X,t) \leqslant \mathcal{L}(X,r) \leqslant (1 + 2r/t)^{k+2}\mathcal{L}(X,t).$$

Proof. The lower bound is trivial. The upper bound follows from the fact that a ball of radius r in $\mathbb{R}^{k+2}$ can be covered by $(1+2r/t)^{k+2}$ balls of radius t. □

[8] While we can refer the reader to [130, 131] for more on the singular values of rectangular random matrices, we were not able to find any result such as (2.58) in the literature. However, it is not so hard to deduce (2.58) from the Hanson-Wright inequality [132] along with a "random rounding" step similar to that in Appendix E in [33] .

We now prove a "tensorization" lemma which shows that anti-concentration of a single row in a random matrix H (with iid rows) implies the anti-concentration of matrix products involving H.

Lemma 2.6.3. *For $d < n$ and $k \geqslant 0$, let W be a $2d \times (k+2)$ matrix and let H be a $(n-d) \times 2d$ random matrix with i.i.d. rows. Let $\tau \in \mathbb{R}^{2d}$ be a random vector with the same distribution as the rows of H. If $\beta \in (0, 1/8)$ then*

$$\mathbb{P}_H\big(\|HW\|_{\mathrm{HS}} \leqslant \beta^2\sqrt{(k+1)(n-d)}\big) \leqslant \Big(2^5 e^{2\beta^2 k}\mathcal{L}\big(W^T\tau, \beta\sqrt{k+1}\big)\Big)^{n-d}.$$

Proof. Apply Markov's inequality to see that

$$\mathbb{P}\big(\|HW\|_{\mathrm{HS}} \leqslant \beta^2\sqrt{(k+1)(n-d)}\big) \leqslant \exp\big(2\beta^2(k+1)(n-d)\big)\,\mathbb{E}_H e^{-2\|HW\|_{\mathrm{HS}}^2/\beta^2}. \tag{2.59}$$

Letting $\tau_1, \dots, \tau_{n-d}$ denote the i.i.d. rows of H, we may rewrite

$$\mathbb{E}_H\, e^{-2\|HW\|_{\mathrm{HS}}^2/\beta^2} = \prod_{i=1}^{n-d} \mathbb{E}_{\tau_i}\, e^{-2\|W^T\tau_i\|^2/\beta^2} = \Big(\mathbb{E}_\tau\, e^{-2\|W^T\tau\|^2/\beta^2}\Big)^{n-d}. \tag{2.60}$$

Observe now that

$$\mathbb{E}_\tau\, e^{-2\|W^T\tau\|^2/\beta^2} = \int_0^\infty \mathbb{P}\left(e^{-2\|W^T\tau\|^2/\beta^2} > u\right) du = \int_0^\infty 4ue^{-2u^2}\mathbb{P}\big(\|W^T\tau\|_2/\beta \leqslant u\big)\, du.$$

Splitting the integral on the right-hand side gives

$$\mathbb{E}_\tau\, e^{-2\|W^T\tau\|^2/\beta^2} = \int_0^{\sqrt{k+1}} 4ue^{-2u^2}\mathbb{P}\big(\|W^T\tau\|_2 \leqslant \beta u\big) + \int_{\sqrt{k+1}}^\infty 4ue^{-2u^2}\mathbb{P}\big(\|W^T\tau\|_2 \leqslant \beta u\big).$$

We then appeal to Fact 2.6.2 to write

$$\mathbb{E}_\tau\, e^{-2\|W^T\tau\|^2/\beta^2} \leqslant \mathcal{L}\big(W^T\tau, \beta\sqrt{k+1}\big)\left(\int_0^{\sqrt{k+1}} 4ue^{-2u^2}\, du + \int_{\sqrt{k+1}}^\infty \left(1 + \frac{2u}{\sqrt{k+1}}\right)^{k+2} 4ue^{-2u^2}\, du\right).$$

Here the first integral is $\leqslant 1$, while the second integral is $\leqslant 8$ and thus

$$\mathbb{E}_\tau\, e^{-2\|W^T\tau\|^2/\beta^2} \leqslant 9\mathcal{L}\big(W^T\tau, \beta\sqrt{k+1}\big). \tag{2.61}$$

Combining lines (2.61) with (2.60) and (2.59) gives

$$\mathbb{P}_H(\|HW\|_{\mathrm{HS}} \leqslant \beta^2\sqrt{(k+1)(n-d)}) \leqslant \Big(9\exp(2\beta^2(k+1))\mathcal{L}\big(W^T\tau, \beta\sqrt{k+1}\big)\Big)^{n-d},$$

and the result follows. □

2.6.2 Approximating matrices W with nets

Note that in Theorem 2.6.1, the least singular values of the matrix H could, a priori, correspond to any of a huge number of possible directions. To limit the number of directions we need to consider, we build nets for k-tuples of these directions. Luckily, the construction of these nets is rendered relatively simple (unlike the nets $\mathcal{N}_\varepsilon$) by appealing to a randomized-rounding technique pioneered in the context of random matrices by Livshyts [103] (also see Section 3 of [104]).

With this in mind, let $\mathcal{U}_{2d,k}$ be the set of all $2d \times k$ matrices with orthonormal columns. The following theorem provides a net for $\mathcal{U}_{2d,k}$, when viewed as a subset of $\mathbb{R}^{[2d]\times[k]}$. A proof can be found in Appendix E of [33], the arXiv version of this paper.

Lemma 2.6.4. *For $k \leqslant d$ and $\delta \in (0, 1/2)$, there exists $\mathcal{N} = \mathcal{N}_{2d,k} \subset \mathbb{R}^{[2d]\times[k]}$ with $|\mathcal{N}| \leqslant (2^6/\delta)^{2dk}$ so that for any $U \in \mathcal{U}_{2d,k}$, any $r \in \mathbb{N}$ and $r \times 2d$ matrix A there exists $W \in \mathcal{N}$ so that*

1. $\|A(W-U)\|_{\mathrm{HS}} \leqslant \delta(k/2d)^{1/2}\|A\|_{\mathrm{HS}}$,
2. $\|W-U\|_{\mathrm{HS}} \leqslant \delta\sqrt{k}$ *and*
3. $\|W-U\| \leqslant 8\delta$.

Recall, for a $2d \times k$ matrix W and $Y \in \mathbb{R}^d$, we defined (at (2.15)) the augmented matrix

$$W_Y = \left[W, \begin{bmatrix}\mathbf{0}_d \\ Y\end{bmatrix}, \begin{bmatrix}Y \\ \mathbf{0}_d\end{bmatrix}\right].$$

2.6.3 Proof of Theorem 2.6.1

We recall a standard fact from linear algebra, reworded to suit our context.

Fact 2.6.5. *For $3d < n$, let H be a $(n-d) \times 2d$ matrix. If $\sigma_{2d-k+1}(H) \leqslant x$ then there exist k orthogonal unit vectors $w_1, \ldots, w_k \in \mathbb{R}^{2d}$ so that $\|Hw_i\|_2 \leqslant x$. In particular, there exists $W \in \mathcal{U}_{2d,k}$ so that $\|HW\|_{\mathrm{HS}} \leqslant x\sqrt{k}$.*

We also note that if H is a $(n-d) \times 2d$ matrix with entries in $\{-1, 0, 1\}$ then we immediately have $\|H\|_{\mathrm{HS}} \leqslant \sqrt{2d(n-d)}$.

Proof of Theorem 2.6.1. Write $Y := \frac{c_0}{16\sqrt{n}} \cdot X$. We use Fact 2.6.5 to upper bound the left-hand-side of (2.57) as

$$\begin{aligned}\mathbb{P}(\sigma_{2d-k+1}(H) \leqslant c_0 2^{-4}\sqrt{n} \text{ and } \|H_1X\|_2, \|H_2X\|_2 \leqslant n) \\ \leqslant \mathbb{P}(\exists U \in \mathcal{U}_{2d,k} : \|HU_Y\|_{\mathrm{HS}} \leqslant 3c_0\sqrt{n(k+1)}/16).\end{aligned}$$

Set $\delta := c_0/16$, and let $\mathcal{W}$ be the δ-net for $\mathcal{U}_{2d,k}$, given by Lemma 2.6.4.

We fix a matrix H for a moment. If there exists a matrix $U \in \mathcal{U}_{2d,k}$ so that $\|HU_Y\|_{\mathrm{HS}} \leqslant 3c_0\sqrt{n(k+1)}/16$, apply Lemma 2.6.4 to find $W \in \mathcal{W}$ so that

$$\|HW_Y\|_{\mathrm{HS}} \leqslant \|H(W_Y - U_Y)\|_{\mathrm{HS}} + \|HU_Y\|_{\mathrm{HS}} \leqslant \delta(k/2d)^{1/2}\|H\|_{\mathrm{HS}} + 3c_0\sqrt{n(k+1)}/16$$

which is at most $c_0\sqrt{n(k+1)}/4$, since $\|H\|_{\mathrm{HS}} \leqslant \sqrt{2nd}$. Thus

$$\mathbb{P}\left(\exists U \in \mathcal{U}_{2d,k} : \|HU_Y\|_{\mathrm{HS}} \leqslant \frac{c_0}{16}\sqrt{n(k+1)}\right) \leqslant \mathbb{P}\left(\exists W \in \mathcal{W} : \|HW_Y\|_{\mathrm{HS}} \leqslant \frac{c_0}{4}\sqrt{n(k+1)}\right).$$

So by the union bound, we have

$$\mathbb{P}\left(\exists W \in \mathcal{W} : \ \|HW_Y\|_{\mathrm{HS}} \leqslant (c_0/4)\sqrt{n(k+1)}\right) \leqslant \sum_{W\in\mathcal{W}} \mathbb{P}\left(\|HW_Y\|_{\mathrm{HS}} \leqslant (c_0/4)\sqrt{n(k+1)}\right).$$

Now

$$|\mathcal{W}| \leqslant (2^6/\delta)^{2dk} \leqslant \exp(32dk\log c_0^{-1}) \leqslant \exp(c_0k(n-d)/4),$$

where the last inequality holds since $d \leqslant c_0^2 n$, and so

$$\sum_{W\in\mathcal{W}} \mathbb{P}\left(\|HW_Y\|_{\mathrm{HS}} \leqslant \frac{c_0}{4}\sqrt{n(k+1)}\right) \leqslant e^{c_0k(n-d)/4} \max_{W\in\mathcal{W}} \mathbb{P}\left(\|HW_Y\|_{\mathrm{HS}} \leqslant \frac{c_0}{4}\sqrt{n(k+1)}\right). \tag{2.62}$$

Let $W \in \mathcal{W}$ be such that the maximum in (2.62) is attained, apply Lemma 2.6.3 with $\beta := \sqrt{c_0}/2$ to obtain

$$\mathbb{P}(\|HW_Y\|_{\mathrm{HS}} \leqslant (c_0/4)\sqrt{n(k+1)}) \leqslant \left(2^5 e^{c_0k/2}\mathcal{L}\big(W_Y^T\tau, c_0^{1/2}\sqrt{k+1}\big)\right)^{n-d}. \tag{2.63}$$

We now look to apply Lemma 2.3.1. We define $t := 16/(c_0N) \geqslant \exp(-2^{-9}\alpha d)$ and $R_0 := 2^{-7}c_0R = 2^{-7}c_0(2^{39}c_0^{-3}) = 2^{32}c_0^{-2}$ so that we have

$$\|Y\|_2 = c_0\|X\|_2/(16n^{1/2}) \geqslant 2^{-14}c_0^2N = 2^{-10}c_0/t.$$

By the construction of $\mathcal{W}$ in Lemma 2.6.4 we have $\|W\| \leqslant 2$ and $\|W\|_{\mathrm{HS}} \geqslant \sqrt{k}/2$. We also have $k \leqslant 2^{-10}\alpha d$ and $D_\alpha(\frac{c_0}{16\sqrt{n}}X) = D(Y) > 16$, therefore we may apply Lemma 2.3.1 to see that

$$\mathcal{L}\big(W_Y^T\tau, c_0^{1/2}\sqrt{k+1}\big) \leqslant (R_0t)^2\exp(-c_0k) \leqslant \left(\frac{R}{8N}\right)^2\exp(-c_0k).$$

Substituting this bound in (2.63) we get

$$\max_{W\in\mathcal{W}} \mathbb{P}_H(\|HW_Y\|_2 \leqslant (c_0/4)\sqrt{n(k+1)}) \leqslant \left(\frac{R}{N}\right)^{2n-2d}\exp(-c_0k(n-d)/2)$$

and finally combining it with the previous bounds gives

$$\mathbb{P}(\sigma_{2d-k+1}(H) \leqslant c_0\sqrt{n}/16 \text{ and } \|H_1X\|_2, \|H_2X\|_2 \leqslant n) \leqslant \left(\frac{R}{N}\right)^{2n-2d} \exp(-c_0k(n-d)/4).$$

This completes the proof of Theorem 2.6.1. □

2.7 Nets for structured vectors: Size of the Net

In this section we take a important step towards Theorem 2.1.1 by bounding the size of our net

$$\mathcal{N}_\varepsilon := \left\{v \in \Lambda_\varepsilon : (L\varepsilon)^n \leqslant \mathbb{P}(\|Mv\|_2 \leqslant 4\varepsilon\sqrt{n}) \text{ and } \mathcal{L}_{A,op}(v, \varepsilon\sqrt{n}) \leqslant (2^8L\varepsilon)^n\right\},$$

where we recall that

$$\Lambda_\varepsilon := B_n(0,2) \cap \left(4\varepsilon n^{-1/2} \cdot \mathbb{Z}^n\right) \cap \mathcal{I}'([d]).$$

In particular, our main goal of this section will be to prove the following theorem on the size of $\mathcal{N}_\varepsilon$.

Theorem 2.7.1. *For $L \geqslant 2$ and $0 < c_0 \leqslant 2^{-24}$, let $n \geqslant L^{64/c_0^2}$, let $d \in [c_0^2n/4, c_0^2n]$ and let $\varepsilon > 0$ be such that $\log \varepsilon^{-1} \leqslant nL^{-32/c_0^2}$. Then*

$$|\mathcal{N}_\varepsilon| \leqslant \left(\frac{C}{c_0^6L^2\varepsilon}\right)^n,$$

where $C > 0$ is an absolute constant.

As the geometry of the set Λ_ε is a bit complicated, we follow an idea of Tikhomirov [165], by working with the intersection of $\mathcal{N}_\varepsilon$ with a selection of "boxes" which cover (an appropriately re-scaled) Λ_ε.

Definition 2.7.2. Define a (N, κ, d)-*box* to be a set of the form $\mathcal{B} = B_1 \times \ldots \times B_n \subset \mathbb{Z}^n$ where $|B_i| \geqslant N$ for all $i \geqslant 1$; $B_i = [-\kappa N, -N] \cup [N, \kappa N]$, for $i \in [d]$; and $|\mathcal{B}| \leqslant (\kappa N)^n$.

The advantage of working with these boxes is that they lend themselves naturally to a probabilistic interpretation, which we now adopt. We ask "what is the probability that

$$\mathbb{P}_M(\|MX\|_2 \leqslant n) \geqslant \left(\frac{L}{N}\right)^n,$$

where X is chosen uniformly at random from $\mathcal{B}$?". This interpretation was used to ingenious effect in the work of Tikhomirov, who called this the "inversion of randomness". While we do take this vantage point, our path forward is considerably different from that of Tikhomirov.

We now state our key "box" version of Theorem 2.7.1, in this probabilistic framework. Indeed, almost all of the work in proving Theorem 2.7.1 goes into proving the following variant for boxes.

Lemma 2.7.3. *For $L \geqslant 2$ and $0 < c_0 \leqslant 2^{-24}$, let $n > L^{64/c_0^2}$ and let $\frac{1}{4}c_0^2 n \leqslant d \leqslant c_0^2 n$. For $N \geqslant 2$, satisfying $\log N \leqslant c_0 L^{-8n/d} d$, and $\kappa \geqslant 2$, let $\mathcal{B}$ be a (N, κ, d)-box and let X be chosen uniformly at random from $\mathcal{B}$. Then*

$$\mathbb{P}_X\left(\mathbb{P}_M(\|MX\|_2 \leqslant n) \geqslant \left(\frac{L}{N}\right)^n\right) \leqslant \left(\frac{R}{L}\right)^{2n},$$

where $R := Cc_0^{-3}$ and $C > 0$ is an absolute constant.

2.7.1 Counting with the least common denominator

In this subsection, we prove the following simple lemma, which says that the probability of choosing $X \in \mathcal{B}$ with "large" least common denominator is super-exponentially small. This will ultimately allow us to apply Theorem 2.6.1, which requires an upper-bound on the $D_\alpha(X)$ for application.

We point out that in Lemma 2.7.4, we rescale by a factor of $r_n = c_0 2^{-4} n^{-1/2}$, despite the fact we are working in $d < n$ dimensions. This is just a trace of the fact that $\mathbb{R}^n$ is our true point of reference. Additionally we will only need Lemma 2.7.4 when $K = 16$.

Lemma 2.7.4. *For $\alpha \in (0,1), K \geqslant 1$ and $\kappa \geqslant 2$, let $n \geqslant d \geqslant K^2/\alpha$ and let $N \geqslant 2$ be so that $KN < 2^d$. Let $\mathcal{B} = ([-\kappa N, -N] \cup [N, \kappa N])^d$ and let X be chosen uniformly at random from $\mathcal{B}$. Then*

$$\mathbb{P}_X\left(D_\alpha\big(r_n \cdot X\big) \leqslant K\right) \leqslant (2^{20}\alpha)^{d/4}, \tag{2.64}$$

where we have set $r_n := c_0 2^{-4} n^{-1/2}$.

Proof. If $D_\alpha\big(r_n \cdot X\big) \leqslant K$ then let $\psi \in (0, K]$ be the minimum[9] in the definition of least common denominator. Set $\phi := r_n \psi$ and observe that ϕ satisfies

$$\|\phi X\|_{\mathbb{T}} \leqslant \sqrt{\alpha d} \quad \text{and} \quad \phi \in [(2\kappa N)^{-1}, r_n K]. \tag{2.65}$$

To see the bound $\phi \geqslant (2\kappa N)^{-1}$, note that if $\phi < (2\kappa N)^{-1}$ then each coordinate of $\phi \cdot X$ lies in $(-1/2, 1/2)$ which would imply $\|\phi X\|_{\mathbb{T}} = \|\phi X\|_2 = \phi \|X\|_2$. Using the non-triviality condition in the definition of least common denominator (2.2), this would imply

$$\phi\|X\|_2 = \|\phi \cdot X\|_{\mathbb{T}} = \|\psi(r_n \cdot X)\|_{\mathbb{T}} \leqslant \psi \|r_n \cdot X\|_2/2 = \phi\|X\|_2/2,$$

[9] Technically the least common denominator is defined in terms of an infimum, however the minimum is always attained for non-zero vectors.

which is a contradiction. Thus the bounds in (2.65) hold.

Now to calculate the probability in (2.64), we discretize the range of possible ϕ. For each integer $i \in [1/\alpha, 2KN/\alpha] =: I$ we define $\phi_i := i\alpha/(2\kappa N)$ and note that if X, ϕ satisfy (2.65) then there exists ϕ_i for which

$$\|\phi_i X\|_{\mathbb{T}} \leqslant 2\sqrt{\alpha d} \quad \text{and } \phi_i \in [(2\kappa N)^{-1}, r_n K],$$

by simply choosing ϕ_i for which $|\phi_i - \phi| \leqslant \alpha/(\kappa N)$ and using triangle inequality

$$\|\phi_i X\|_{\mathbb{T}} \leqslant \|\phi X\|_{\mathbb{T}} + \|(\phi_i - \phi)X\|_2 \leqslant \sqrt{\alpha d} + |\phi_i - \phi| \cdot \sqrt{d}(\kappa N) \leqslant 2\sqrt{\alpha d}. \tag{2.66}$$

Thus we have that

$$\mathbb{P}_X(D_\alpha(r_n \cdot X) \leqslant K) \leqslant \sum_{i \in I} \mathbb{P}_X\left(\|\phi_i X\|_{\mathbb{T}} \leqslant 2\sqrt{\alpha d}\right). \tag{2.67}$$

To bound the terms on the right-hand side, note that if $\|\phi_i X\|_{\mathbb{T}} \leqslant 2\sqrt{\alpha d}$ then

$$\frac{1}{d}\sum_{j=1}^{d} \|\phi_i X_j\|_{\mathbb{T}}^2 \leqslant 4\alpha\,.$$

By averaging, there is a set $S(X,i) \subset [d]$ with $|S(X,i)| \geqslant d/2$ for which $\|\phi_i X_j\|_{\mathbb{T}} \leqslant 4\sqrt{\alpha}$ for all $j \in S(X,i)$. Union bounding over all sets $S \subseteq [d]$ and using the independence of the coordinates X_j we have

$$\mathbb{P}_X(D_\alpha(r_n \cdot X) \leqslant K) \leqslant 2^d \sum_{i \in I} \prod_{j=1}^{d/2} \mathbb{P}_{X_j}\left(\|\phi_i X_j\|_{\mathbb{T}} \leqslant 4\sqrt{\alpha}\right). \tag{2.68}$$

We now claim that

$$\mathbb{P}_{X_j}\left(\|\phi_i X_j\|_{\mathbb{T}} \leqslant 4\sqrt{\alpha}\right) \leqslant 32\sqrt{\alpha}. \tag{2.69}$$

For this, note that if $\|\phi_i X_j\|_{\mathbb{T}} \leqslant 4\sqrt{\alpha}$, then $|\phi_i X_j - p| \leqslant 4\sqrt{\alpha}$, where $p \in \mathbb{Z}$ satisfies $|p| \leqslant |\phi_i X_j| + 1 \leqslant \phi_i \kappa N + 1 =: T_i$. And so

$$\mathbb{P}_{X_j}(\|\phi_i X_j\|_{\mathbb{T}} \leqslant 4\sqrt{\alpha}) \leqslant \sum_{p=-T_i}^{T_i} \mathbb{P}_{X_j}(|X_j - p\phi_i^{-1}| \leqslant 4\sqrt{\alpha}\phi_i^{-1}) \leqslant \frac{(2T_i+1)(8\alpha^{1/2}\phi_i^{-1}+1)}{2(\kappa-1)N}.$$

where we have used that X_j is uniform on $[-\kappa N, -N] \cup [N, \kappa N]$ and the lower bound $\kappa N \phi_i \geqslant 1/2$ from (2.66) along with the assumption $\kappa \geqslant 2$. Also note that $8\alpha^{1/2}\phi_i^{-1} \geqslant 1$ since $\phi \leqslant r_n K \leqslant d^{-1/2}K$, allowing us to conclude (2.69).

Now, plugging (2.69) into (2.68) and bounding $|I| \leqslant (2KN/\alpha + 1) \leqslant 3^d$ completes the proof of Lemma 2.7.4. □

2.7.2 Anti-concentration for linear projections of random vectors

In this subsection we prove the following anti-concentration result for random variables HX, where H is a *fixed* matrix and X is a random vector with independent entries. One small remark regarding notation: H as stated in Lemma 2.7.5 will actually be H^T in Section 2.7.3.

Lemma 2.7.5. *Let $N \in \mathbb{N}$, $n, d, k \in \mathbb{N}$ be such that $n - d \geqslant 2d > 2k$, H be a $2d \times (n-d)$ matrix with $\sigma_{2d-k}(H) \geqslant c_0\sqrt{n}/16$ and $B_1, \ldots, B_{n-d} \subset \mathbb{Z}$ with $|B_i| \geqslant N$. If X is taken uniformly at random from $\mathcal{B} := B_1 \times \ldots \times B_{n-d}$, then*

$$\mathbb{P}_X(\|HX\|_2 \leqslant n) \leqslant \left(\frac{Cn}{dc_0 N}\right)^{2d-k},$$

where $C > 0$ is an absolute constant.

We derive this from the following anti-concentration result of Rudelson and Vershynin. This is essentially Corollary 1.4 along with Remark 2.3 in their paper [133], but we have restated their result slightly to better suit our context.

Theorem 2.7.6. *Let $N \in \mathbb{N}$ and let $n, d, k \in \mathbb{N}$ be such that $n - d \geqslant 2d > k$. Let P be an orthogonal projection of $\mathbb{R}^{n-d}$ onto a $(2d-k)$-dimensional subspace and let $X = (X_1, \ldots, X_{n-d})$ be a random vector with independent entries for which*

$$\mathcal{L}\big(X_i, 1/2\big) \leqslant N^{-1},$$

for all $i \in [n-d]$. Then, for all $K \geqslant 1$,

$$\max_{y \in \mathbb{R}^{n-d}} \mathbb{P}_X\big(\|PX - y\|_2 \leqslant K\sqrt{2d-k}\big) \leqslant \left(\frac{CK}{N}\right)^{2d-k},$$

where $C > 0$ is a absolute constant.

We can now deduce Lemma 2.7.5.

Proof of Lemma 2.7.5. Since $H^T H$ is a symmetric $(n-d) \times (n-d)$ matrix with $\mathrm{rk}(H) \leqslant 2d$, by the spectral theorem we have $H^T H = \sum_{i=1}^{2d} \sigma_i(H)^2 v_i v_i^T$, where $v_1, \ldots, v_{2d} \in \mathbb{R}^{n-d}$ are orthonormal. Define the orthogonal projection $P := \sum_{i=1}^{2d-k} v_i v_i^T$. Then we have

$$\|HX\|_2^2 = \langle X, H^T HX\rangle = \sum_{j=1}^{2d} \sigma_j(H)^2 \langle X, v_j\rangle^2 \geqslant \sigma_{2d-k}(H)^2 \sum_{j=1}^{2d-k} \langle X, v_j\rangle^2 \geqslant 2^{-8} c_0^2 n \|PX\|_2^2.$$

Therefore

$$\mathbb{P}_X(\|HX\|_2 \leqslant n) \leqslant \mathbb{P}_X(\|PX\|_2 \leqslant 16 c_0^{-1}\sqrt{n}). \tag{2.70}$$

We now apply Theorem 2.7.6 to the orthogonal projection P with $K = 16c_0^{-1}\sqrt{n/(2d-k)}$ to see

$$\mathbb{P}_X(\|PX\|_2 \leqslant K\sqrt{2d-k}) \leqslant \left(\frac{Cn}{dc_0N}\right)^{2d-k}, \tag{2.71}$$

which together with (2.70) completes the proof of Lemma 2.7.5. □

2.7.3 Proof of Theorem 2.7.3

We take a moment to prepare the ground for the proof of Theorem 2.7.3. We express our random matrix M, as in the statement of Theorem 2.7.3, as

$$M = \begin{bmatrix} \mathbf{0}_{[d]\times[d]} & H_1^T \\ H_1 & \mathbf{0}_{[n-d]\times[n-d]}, \end{bmatrix}$$

Where H_1 is a $(n-d)\times d$ random matrix with iid 1/4-lazy entries in $\{-1,0,1\}$. We shall also let H_2 be an independent copy of H_1 and define H to be the $(n-d)\times 2d$ matrix

$$H := \begin{bmatrix} H_1 & H_2 \end{bmatrix}.$$

For a vector $X \in \mathbb{R}^n$, we define the event $\mathcal{A}_1 = \mathcal{A}_1(X)$ by

$$\mathcal{A}_1 := \left\{H : \|H_1X_{[d]}\|_2 \leqslant n \text{ and } \|H_2X_{[d]}\|_2 \leqslant n\right\}$$

and let $\mathcal{A}_2 = \mathcal{A}_2(X)$ be the event

$$\mathcal{A}_2 := \left\{H : \|H^TX_{[d+1,n]}\|_2 \leqslant 2n\right\}.$$

We now note a simple inequality linking H, $\mathcal{A}_1$ and $\mathcal{A}_2$ with the event $\{\|MX\|_2 \leqslant n\}$.

Fact 2.7.7. *For $X \in \mathbb{R}^n$, let $\mathcal{A}_1 = \mathcal{A}_1(X)$, $\mathcal{A}_2 = \mathcal{A}_2(X)$ be as above. We have*

$$(\mathbb{P}_M(\|MX\|_2 \leqslant n))^2 \leqslant \mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{A}_2).$$

Proof. Let M' be an independent copy of M. Expand $\mathbb{1}(\|MX\|_2 \leqslant n)$ as a sum of indicators, apply $\mathbb{E}_M$ and square to see

$$(\mathbb{P}_M(\|MX\|_2 \leqslant n))^2 = \sum_{M,M'} \mathbb{P}(M')\mathbb{P}(M)\mathbb{1}(\|MX\|_2, \|M'X\|_2 \leqslant n),$$

which is at most

$$\sum_{H_1,H_2} \mathbb{P}(H_1)\mathbb{P}(H_2)\mathbb{1}\big(\|H_1X_{[d]}\|_2 \leqslant n, \|H_2X_{[d]}\|_2 \leqslant n \text{ and } \|H^TX_{[d+1,n]}\|_2 \leqslant 2n\big),$$

which is exactly $\mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{A}_2)$. $\square$

We shall also need a "robust" notion of the rank of the matrix H: Define $\mathcal{E}_k$ to be the event

$$\mathcal{E}_k := \left\{H : \sigma_{2d-k}(H) \geqslant c_0\sqrt{n}/16 \text{ and } \sigma_{2d-k+1}(H) < c_0\sqrt{n}/16\right\}$$

and note that always exactly one of the events $\mathcal{E}_0, \ldots, \mathcal{E}_{2d}$ holds.

We now set

$$\alpha := 2^{13}L^{-8n/d}, \tag{2.72}$$

and, given a box $\mathcal{B}$, we define the set of *typical* vectors $T(\mathcal{B}) \subseteq \mathcal{B}$ to be

$$T = T(\mathcal{B}) := \left\{X \in \mathcal{B} : D_\alpha(c_0 2^{-4} n^{-1/2} X_{[d]}) > 16\right\}. \tag{2.73}$$

Now set $K := 16$ and note that Lemma 2.7.4 implies that if X is chosen uniformly from $\mathcal{B}$ and $n \geqslant L^{64/c_0^2} \geqslant 2^8/\alpha$ we have

$$\mathbb{P}_X(X \notin T) = \mathbb{P}_X\left(D_\alpha(c_0 2^{-4} n^{-1/2} X_{[d]}) \leqslant 16\right) \leqslant \left(2^{33}L^{-8n/d}\right)^{d/4} \leqslant \left(\frac{2}{L}\right)^{2n}. \tag{2.74}$$

Proof of Lemma 2.7.3. Let M, H_1, H_2, H, $\mathcal{A}_1, \mathcal{A}_2$, $\mathcal{E}_k$, α and $T := T(\mathcal{B})$ be as above. We denote

$$\mathcal{E} := \left\{X \in \mathcal{B} : \mathbb{P}_M(\|MX\|_2 \leqslant n) \geqslant (L/N)^n\right\}$$

and write

$$\mathbb{P}_X(\mathcal{E}) \leqslant \mathbb{P}_X(\mathcal{E} \cap \{X \in T\}) + \mathbb{P}_X(X \notin T).$$

Now define

$$f(X) := \mathbb{P}_M(\|MX\|_2 \leqslant n)\mathbb{1}(X \in T)$$

and apply (2.74), the bound on $\mathbb{P}_X(X \notin T)$, to obtain

$$\mathbb{P}_X(\mathcal{E}) \leqslant \mathbb{P}_X\left(f(X) \geqslant (L/N)^n\right) + (2/L)^{2n} \leqslant (N/L)^{2n}\mathbb{E}_X f(X)^2 + (2/L)^{2n}, \tag{2.75}$$

where the last inequality follows from Markov's inequality. So to prove Lemma 2.7.3, it is enough to prove $\mathbb{E}_X f(X)^2 \leqslant 2(R/N)^{2n}$.

From Fact 2.7.7 we may write

$$\mathbb{P}_M(\|MX\|_2 \leqslant n)^2 \leqslant \mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{A}_2) = \sum_{k=0}^{d} \mathbb{P}_H(\mathcal{A}_2 | \mathcal{A}_1 \cap \mathcal{E}_k)\mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{E}_k) \tag{2.76}$$

and so

$$f(X)^2 \leqslant \sum_{k=0}^{d} \mathbb{P}_H(\mathcal{A}_2 | \mathcal{A}_1 \cap \mathcal{E}_k)\mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{E}_k)\mathbb{1}(X \in T). \tag{2.77}$$

We now look to apply Lemma 2.6.1 to obtain upper bounds for the quantities $\mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{E}_k)$, when $X \in T$. For this, note that $d \leqslant c_0^2 n$, $N \leqslant \exp(L^{-8n/d} d) \leqslant \exp(2^{-10}\alpha n)$ and set $R_0 := 2^{39} c_0^{-3}$ (This is the "R" in Theorem 2.6.1). Also note that, by the definition of a (N, κ, d)-box and the fact that $d \geqslant \frac{1}{4} c_0^2 n$, we have that $\|X_{[d]}\|_2 \geqslant d^{1/2} N \geqslant c_0 2^{-10} \sqrt{n} N$. Now set $\alpha' := 2^{-10}\alpha$ to see that for $X \in T$ and $0 \leqslant k \leqslant \alpha' d$,

$$\mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{E}_k) \leqslant \exp(-c_0 n k/4) \left(\frac{R_0}{N}\right)^{2n-2d}.$$

Moreover by Theorem 2.6.1,

$$\sum_{k \geqslant \alpha' d} \mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{E}_k) \leqslant \mathbb{P}_H\big(\{\sigma_{2d-\alpha' d}(H) \leqslant c_0\sqrt{n}/16\} \cap \mathcal{A}_1\big) \leqslant \exp(-c_0 \alpha' d n/4).$$

Thus, for all $X \in \mathcal{B}$, we have

$$f(X)^2 \leqslant \sum_{k=0}^{\alpha' d} \mathbb{P}_H(\mathcal{A}_2 \,|\, \mathcal{A}_1 \cap \mathcal{E}_k) \exp(-c_0 n k/4) \left(\frac{R_0}{N}\right)^{2n-2d} + \exp(-c_0 \alpha' d n/4)\,. \tag{2.78}$$

We now consider the quantities $g_k(X) := \mathbb{P}_H(\mathcal{A}_2 \,|\, \mathcal{A}_1 \cap \mathcal{E}_k)$ appearing in (2.78). Indeed,

$$\mathbb{E}_X[g_k(X)] = \mathbb{E}_X \mathbb{E}_H\big[\mathcal{A}_2 \,|\, \mathcal{A}_1 \cap \mathcal{E}_k\big] = \mathbb{E}_{X_{[d]}} \mathbb{E}_H\left[\mathbb{E}_{X_{[d+1,n]}} \mathbb{1}[\mathcal{A}_2] \,\middle|\, \mathcal{A}_1 \cap \mathcal{E}_k\right].$$

We now consider a fixed $H \in \mathcal{A}_1 \cap \mathcal{E}_k$ for $k \leqslant \alpha' d$. Each such H has $\sigma_{2d-k}(H) \geqslant c_0\sqrt{n}/16$ and thus we may apply Lemma 2.7.5 to see that

$$\mathbb{E}_{X_{[d+1,n]}} \mathbb{1}[\mathcal{A}_2] = \mathbb{P}_{X_{[d+1,n]}}(\|H^T X_{[d+1,n]}\|_2 \leqslant n) \leqslant \left(\frac{C' n}{c_0 d N}\right)^{2d-k} \leqslant \left(\frac{4C'}{c_0^3 N}\right)^{2d-k},$$

for an absolute constant $C' > 0$, using that $d \geqslant \frac{1}{4} c_0^2 n$. And so for each $0 \leqslant k \leqslant \alpha' d$, taking $R := \max\{8C' c_0^{-3}, 2R_0\}$, we have

$$\mathbb{E}_X[g_k(X)] \leqslant \left(\frac{R}{2N}\right)^{2d-k}. \tag{2.79}$$

We apply $\mathbb{E}_X$ to (2.78) and then use (2.79) to obtain

$$\mathbb{E}_X f(X)^2 \leqslant \left(\frac{R}{2N}\right)^{2n} \sum_{k=0}^{\alpha' d} \left(\frac{2N}{R}\right)^k \exp(-c_0 n k/4) + \exp(-c_0 \alpha' d n/4).$$

Using that $N \leqslant \exp(c_0 n/4)$ and $N \leqslant \exp(c_0 L^{-8n/d} d) = \exp(c_0 \alpha' d/8)$ gives

$$\mathbb{E}_X\, f(X)^2 \leqslant 2\left(\frac{R}{2N}\right)^{2n}. \tag{2.80}$$

Combining (2.80) with (2.75) completes the proof of Lemma 2.7.3. □

2.7.4 Proof of Theorem 2.7.1

The main work of this section is now complete with the proof of Lemma 2.7.3. We now just need to go from X in a "box" to X in a "sphere" Λ_ε. To accomplish this step, we simply cover the sphere with boxes. Recall that

$$\mathcal{I}'([d]) := \left\{v \in \mathbb{R}^n : \kappa_0 n^{-1/2} \leqslant |v_i| \leqslant \kappa_1 n^{-1/2} \text{ for all } i \in [d]\right\},$$

$$\Lambda_\varepsilon := B_n(0,2) \cap \left(4\varepsilon n^{-1/2} \cdot \mathbb{Z}^n\right) \cap \mathcal{I}'([d]),$$

and that $0 < \kappa_0 < 1 < \kappa_1$ are absolute constants defined in Section 2.2.

Lemma 2.7.8. *For all $\varepsilon \in [0,1]$, $\kappa \geqslant \max\{\kappa_1/\kappa_0, 2^8\kappa_0^{-4}\}$, there exists a family $\mathcal{F}$ of (N,κ,d)-boxes with $|\mathcal{F}| \leqslant \kappa^n$ so that*

$$\Lambda_\varepsilon \subseteq \bigcup_{\mathcal{B}\in\mathcal{F}} (4\varepsilon n^{-1/2}) \cdot \mathcal{B}, \tag{2.81}$$

where $N = \kappa_0/(4\varepsilon)$.

Proof. For $\ell \geqslant 1$ define the interval of integers $I_\ell := \left[-2^\ell N, 2^\ell N\right] \setminus \left[-2^{\ell-1}N, 2^{\ell-1}N\right]$ and $I_0 := [-N, N]$. Also take $J := [-\kappa N, \kappa N]\backslash[-N, N]$. For $(\ell_{d+1}, \dots, \ell_n) \in \mathbb{Z}^n_{\geqslant 0}$ we define the box $B(\ell_{d+1}, \dots, \ell_n) := J^d \times \prod_{j=d+1}^n I_{\ell_j}$ and the family of boxes

$$\mathcal{F} := \left\{ B(\ell_{d+1}, \dots, \ell_n) : \sum_{j:\ell_j>0} 2^{2\ell_j} \leqslant 8n/\kappa_0^2 \right\}.$$

We claim that $\mathcal{F}$ is the desired family. For this, we first show the inclusion at (2.81). Let $v \in \Lambda_\varepsilon$. Since $v \in 4\varepsilon n^{-1/2}\mathbb{Z}^n$, $X := vn^{1/2}/(4\varepsilon) \in \mathbb{Z}^n$. For $i \in [d+1, n]$, define ℓ_i so that $X_i \in I(\ell_i)$. We claim $X \in B(\ell_{d+1}, \dots, \ell_n)$. For this, observe that $X_i \in J$ for $i \in [d]$: since $v \in \mathcal{I}'([d])$, we have $\kappa_0 \leqslant |v_i|n^{1/2} \leqslant \kappa_1$, for $i \in [d]$. So $\kappa_0/(4\varepsilon) \leqslant |X_i| \leqslant \kappa_1/(4\varepsilon)$, for $i \in [d]$. Thus $X_i \in J$ since $N = \kappa_0/(4\varepsilon)$ and $\kappa \geqslant \kappa_1/\kappa_0$. Thus $v \in B(\ell_{d+1}, \dots, \ell_n)$. We now observe that $B(\ell_{d+1}, \dots, \ell_n) \in \mathcal{F}$, since

$$\sum_{j:\ell_j>0} 2^{2(\ell_j-1)}N^2 \leqslant \sum_{j=1}^n X_j^2 \leqslant n/(4\varepsilon)^2 \left(\sum_i v_i^2\right) \leqslant 4nN^2/\kappa_0^2.$$

Thus we have (2.81).

We now show $|\mathcal{F}| \leqslant \kappa^n$. For this we only need to count the number of sequences $(\ell_{d+1}, \dots, \ell_n)$ of non-negative integers for which $\sum_{\ell_i>0} 4^{\ell_i} \leqslant 8n/\kappa_0^2$. For each $t \geqslant 0$ are at most $8n/(4^t\kappa_0^2)$ values of $i \in [d+1, n]$ for which $\ell_i = t$ and there are at most $\binom{n}{\leqslant 8n/(4^t\kappa_0^2)}$ choices for these values of i. Hence, there are at most

$$\prod_{t\geqslant 0} \binom{n}{\leqslant 8n/(\kappa_0^2 4^t)} \leqslant (\kappa_0/4)^{-4n} < \kappa^n$$

such tuples.

It only remains to show an upper bound on the size of $B(\ell_{d+1}, \dots, \ell_n) \in \mathcal{F}$. We have

$$|B(\ell_{d+1}, \dots, \ell_n)| \leqslant N^n \kappa^d 2^{n+\sum_j \ell_j} \leqslant \kappa^d (16/\kappa_0^2)^n N^n \leqslant (\kappa N)^n$$

where the second inequality holds due to the fact $\prod_j 2^{\ell_j} \leqslant \left(\frac{1}{n}\sum_j 2^{2\ell_j}\right)^n \leqslant (8/\kappa_0^2)^n$ and the last inequality holds due to the choice of κ. □

We may now use our covering Lemma 2.7.8 to apply Theorem 2.7.3 to deduce Theorem 2.7.1, the main result of this section.

Proof of Theorem 2.7.1. Apply Lemma 2.7.8 with $\kappa = \max\{\kappa_1/\kappa_0, 2^8\kappa_0^{-4}\}$ and use the fact that $\mathcal{N}_\varepsilon \subseteq \Lambda_\varepsilon$ to write

$$\mathcal{N}_\varepsilon \subseteq \bigcup_{\mathcal{B}\in\mathcal{F}} \left((4\varepsilon n^{-1/2}) \cdot \mathcal{B}\right) \cap \mathcal{N}_\varepsilon$$

and so

$$|\mathcal{N}_\varepsilon| \leqslant \sum_{\mathcal{B}\in\mathcal{F}} |(4\varepsilon n^{-1/2} \cdot \mathcal{B}) \cap \mathcal{N}_\varepsilon| \leqslant |\mathcal{F}| \cdot \max_{\mathcal{B}\in\mathcal{F}} |(4\varepsilon n^{-1/2} \cdot \mathcal{B}) \cap \mathcal{N}_\varepsilon|.$$

By rescaling by $\sqrt{n}/(4\varepsilon)$ and applying Lemma 2.7.3, we have

$$|(4\varepsilon n^{-1/2} \cdot \mathcal{B}) \cap \mathcal{N}_\varepsilon| \leqslant \left|\left\{X \in \mathcal{B} : \mathbb{P}_M(\|MX\|_2 \leqslant n) \geqslant (L\varepsilon)^n\right\}\right| \leqslant \left(\frac{R}{L}\right)^{2n} |\mathcal{B}|.$$

Here the application of Lemma 2.7.3 is justified as $0 < c_0 \leqslant 2^{-24}$, $c_0^2 n/2 \leqslant d \leqslant c_0^2 n$; $\kappa \geqslant 2$; we have $\log 1/\varepsilon \leqslant n/L^{32/c_0^2}$ and therefore

$$\log N = \log \kappa_0/(4\varepsilon) \leqslant n/L^{32/c_0^2} \leqslant c_0 L^{-8n/d} d,$$

as specified in Lemma 2.7.3, since $\kappa_0 < 1$, $d \geqslant L^{-1/c_0^2} n$, $c_0 \geqslant L^{-1/c_0^2}$ and $8n/d \leqslant 16/c_0^2$. So, using that $|\mathcal{F}| \leqslant \kappa^n$ and $|\mathcal{B}| \leqslant (\kappa N)^n$ for each $\mathcal{B} \in \mathcal{F}$, we have

$$|\mathcal{N}_\varepsilon| \leqslant \kappa^n \left(\frac{R}{L}\right)^{2n} |\mathcal{B}| \leqslant \kappa^n \left(\frac{R}{L}\right)^{2n} (\kappa N)^n \leqslant \left(\frac{C}{c_0^6 L^2 \varepsilon}\right)^n,$$

where $C = \kappa^2 R^2 c_0^6$, thus completing the proof of Theorem 2.7.1. □

2.8 Nets for structured vectors: approximating with the net

While we have spent considerable energy up to this point showing that $\mathcal{N}_\varepsilon$ is small, we have so far not shown that it is in fact a *net*. We now show just this, by showing that vectors in Σ_ε are approximated by elements of $\mathcal{N}_\varepsilon$. As we will see, this is considerably easier and is taken care of

in Lemma 2.8.2, which, in a similar spirit to Lemma 2.6.4, is based on randomized rounding. For this, we recall that we defined

$$\Sigma_\varepsilon = \{v \in \mathcal{I}([d]) : \mathcal{T}_L(v) \in [\varepsilon, 2\varepsilon]\} \subset \mathbb{S}^{n-1}\,, \tag{2.82}$$

where $\mathcal{T}_L(v) = \sup\{t \in [0,1] : \mathbb{P}(\|Mv\|_2 \leqslant t\sqrt{n}) \geqslant (4Lt)^n\}$, and $d = c_0^2 n < 2^{-32}n$. Also recall the definition of our net

$$\mathcal{N}_\varepsilon = \left\{v \in \Lambda_\varepsilon : \mathbb{P}(\|Mv\|_2 \leqslant 4\varepsilon\sqrt{n}) \geqslant (L\varepsilon)^n \text{ and } \mathcal{L}_{A,op}(v, \varepsilon\sqrt{n}) \leqslant (2^8 L\varepsilon)^n\right\}.$$

We also make the basic observation that if $\mathcal{T}_L(v) = s$, then

$$(2sL)^n \leqslant \mathbb{P}(\|Mv\|_2 \leqslant s\sqrt{n}) \leqslant (8sL)^n\,.$$

Until now, we have almost entirely been working with the matrix M. The following lemma allows us to make a comparison between M and our central object of study: A, a uniform $n \times n$ symmetric matrix with entries in $\{-1, 1\}$. The proof of the lemma is based on a comparison of Fourier transforms and is deferred to Appendix B. This is similar to the replacement step in the work of Kahn Komlós and Szemerédi [88] and subsequent works [30, 153]. However, here we only need to "break even", whereas they are looking for a substantial gain at this step.

Lemma 2.8.1. *For $v \in \mathbb{R}^n$ and $t \geqslant \mathcal{T}_L(v)$ we have*

$$\mathcal{L}(Av, t\sqrt{n}) \leqslant (50Lt)^n\,.$$

We now prove Lemma 2.8.2 which tells us that $\mathcal{N}_\varepsilon$ is a net for Σ_ε.

Lemma 2.8.2. *Let $\varepsilon \in (0, \kappa_0/8)$, $d \leqslant n/32$. If $v \in \Sigma_\varepsilon$ then there is $u \in \mathcal{N}_\varepsilon$ with $\|u - v\|_\infty \leqslant 4\varepsilon n^{-1/2}$.*

Proof. Given $v \in \Sigma_\varepsilon$, we define a random variable $r = (r_1, \dots, r_n)$ where the r_i are independent, $\mathbb{E}\, r_i = 0$, $|r_i| \leqslant 4\varepsilon n^{-1/2}$ and such that $v - r \in 4\varepsilon n^{-1/2}\mathbb{Z}^n$, for all r. We then define the random variable $u := v - r$. We will show that with positive probability there is a choice of $u \in \mathcal{N}_\varepsilon$.

Note that $\|r\|_\infty = \|u - v\|_\infty \leqslant 4\varepsilon n^{-1/2}$ for all u. Also, $u \in \mathcal{I}'([d])$ for all u, since $v \in \mathcal{I}([d])$ and $\|u - v\|_\infty \leqslant 4\varepsilon/\sqrt{n} \leqslant \kappa_0/(2\sqrt{n})$. So, from the definition of $\mathcal{N}_\varepsilon$, we need only show that there exists such a u satisfying

$$\mathbb{P}(\|Mu\|_2 \leqslant 4\varepsilon\sqrt{n}) \geqslant (L\varepsilon)^n \text{ and } \mathcal{L}_{A,op}(u, \varepsilon\sqrt{n}) \leqslant (2^8 L\varepsilon)^n. \tag{2.83}$$

We first show that *all* u satisfy the upper bound at (2.83). To see this, write $\mathcal{E} = \{\|A\| \leqslant 4\sqrt{n}\}$ and let $w(u) \in \mathbb{R}^n$, be such that

$$\begin{aligned}\mathcal{L}_{A,op}(u, \varepsilon\sqrt{n}) &= \mathbb{P}\left(\|Av - Ar - w(u)\| \leqslant \varepsilon\sqrt{n} \text{ and } \mathcal{E}\right) \\ &\leqslant \mathbb{P}\left(\|Av - w(u)\| \leqslant 5\varepsilon\sqrt{n} \text{ and } \mathcal{E}\right) \\ &\leqslant \mathcal{L}_{A,op}(v, 5\varepsilon\sqrt{n}) \leqslant \mathcal{L}(Av, 5\varepsilon\sqrt{n}).\end{aligned}$$

Since $v \in \Sigma_\varepsilon$, Lemma 2.8.1 bounds

$$\mathcal{L}(Av, 5\varepsilon\sqrt{n}) \leqslant (2^8 L\varepsilon)^n \,. \tag{2.84}$$

We now show that

$$\mathbb{E}_u\, \mathbb{P}_M(\|Mu\|_2 \leqslant 4\varepsilon\sqrt{n}) \geqslant (1/2)\mathbb{P}_M(\|Mv\|_2 \leqslant 2\varepsilon\sqrt{n}) \geqslant (1/2)(2\varepsilon L)^n \,, \tag{2.85}$$

where the last inequality holds by the fact $v \in \Sigma_\varepsilon$. From (2.85), it follows that there exists $u \in \Lambda_\varepsilon$ satisfying (2.83).

So to prove the first inequality in (2.83), we define the event $\mathcal{E} := \{M : \|Mv\|_2 \leqslant 2\varepsilon\sqrt{n}\}$. For all u, we have

$$\mathbb{P}_M(\|Mu\|_2 \leqslant 4\varepsilon\sqrt{n}) = \mathbb{P}_M(\|Mv - Mr\|_2 \leqslant 4\varepsilon\sqrt{n}) \geqslant \mathbb{P}_M(\|Mr\|_2 \leqslant 2\varepsilon\sqrt{n} \text{ and } \mathcal{E});$$

Thus

$$\begin{aligned}\mathbb{P}_M(\|Mu\|_2 \leqslant 4\varepsilon\sqrt{n}) &\geqslant \mathbb{P}_M(\|Mr\|_2 \leqslant 2\varepsilon\sqrt{n}\,|\mathcal{E})\mathbb{P}(\mathcal{E}) \\ &\geqslant \left(1 - \mathbb{P}_M(\|Mr\|_2 > 2\varepsilon\sqrt{n}\,|\mathcal{E})\right)\mathbb{P}_M(\|Mv\|_2 \leqslant 2\varepsilon\sqrt{n})\,.\end{aligned}$$

Taking expectations with respect to u gives,

$$\mathbb{E}_u\mathbb{P}_M(\|Mu\|_2 \leqslant 4\varepsilon\sqrt{n}) \geqslant \left(1 - \mathbb{E}_u\mathbb{P}_M(\|Mr\|_2 > 2\varepsilon\sqrt{n}\,|\mathcal{E})\right)\mathbb{P}_M(\|Mv\|_2 \leqslant 2\varepsilon\sqrt{n}) \tag{2.86}$$

and exchanging the expectations reveals that it is enough to show

$$\mathbb{E}_M\left[\mathbb{P}_r(\|Mr\|_2 > 2\varepsilon\sqrt{n}) \,\middle|\, \mathcal{E}\right] \leqslant 1/2.$$

We will show that $\mathbb{P}_r(\|Mr\|_2 > 2\varepsilon\sqrt{n}) \leqslant 1/4$ for all $M \in \mathcal{E}$, by Markov's inequality. For this, fix a $n \times n$ matrix M with entries $|M_{i,j}| \leqslant 1$ and $M_{i,j} = 0$, if $(i,j) \in [d+1, n] \times [d+1, n]$, and note that

$$\mathbb{E}_r\, \|Mr\|_2^2 = \sum_{i,j} \mathbb{E}\,(M_{i,j} r_i)^2 = \sum_i \mathbb{E}\, r_i^2 \sum_j M_{i,j}^2 \leqslant 32\varepsilon^2 d \leqslant \varepsilon^2 n,$$

where, for the second equality, we have used that the r_i are mutually independent and $\mathbb{E}\, r_i = 0$, for the third inequality, we used $\|r\|_\infty \leqslant 4\varepsilon/\sqrt{n}$ and for the final inequality we used $d \leqslant n/32$.

Thus by Markov, we have

$$\mathbb{P}_r(\|Mr\|_2 \geqslant 2\varepsilon\sqrt{n}) \leqslant (2\varepsilon\sqrt{n})^{-2}\mathbb{E}_r \|Mr\|_2^2 \leqslant 1/4. \tag{2.87}$$

Putting (2.87) together with (2.86) proves (2.85), completing the proof of (2.83). □

2.9 Proof of Theorem 2.1.1

In this section we put together our results to prove Theorem 2.1.1. But before we get to this, we note a few reductions afforded by previous work. Let us define

$$q_n(\gamma) := \max_{w\in\mathbb{R}^n} \mathbb{P}_A(\exists v \in \mathbb{R}^n \setminus \{0\} : Av = w,\ \rho(v) \geqslant \gamma), \tag{2.88}$$

where

$$\rho(v) = \max_{w\in\mathbb{R}} \mathbb{P}\left(\sum_{i=1}^n \varepsilon_i v_i = w\right)$$

and $\varepsilon_1, \dots, \varepsilon_n \in \{-1, 1\}$ are i.i.d. and uniform. One slightly irritating aspect of the definition (2.88) is that the existential quantifies over *all non-zero* $v \in \mathbb{R}^n$, rather than all $v \in \mathbb{S}^{n-1}$, as we have been working with. So, as we will shortly see, we will need to approximate this extra dimension of freedom with a net.

These small issues aside, we will use the following inequality, which effectively allows us to remove very unstructured vectors from consideration.

Lemma 2.9.1. *Let A be a random $n \times n$ symmetric $\{-1, 1\}$-matrix. For all $\gamma > 0$ we have*

$$\mathbb{P}(\det(A) = 0) \leqslant 16n \sum_{m=n}^{2n-2} \left(\gamma^{1/8} + \frac{q_{m-1}(\gamma)}{\gamma}\right)$$

We record the details of this lemma in Appendix C of the arXiv version of this paper [33], although an almost identical lemma can be found in [37], which collected elements from [41, 62, 113].

2.9.1 Non-flat vectors

Here we note a lemma due to Vershynin [166] which tells us that it is enough for us to consider vectors $v \in \mathcal{I}$. For this, we reiterate the important notion of *compressible vectors*, introduced by Rudelson and Vershynin [129]. Say a vector in $\mathbb{S}^{n-1}$ is (δ, ρ)-compressible if it has distance $\leqslant \rho$ from a vector with support $\leqslant \delta n$. Let $\mathrm{Comp}\,(\delta, \rho)$ denote the set of such compressible vectors.

In [166, Proposition 4.2], Vershynin provides the following lemma which allows us to disregard all compressible vectors.

Lemma 2.9.2. *There exist $\delta, \rho, c \in (0,1)$ so that for all $n \in \mathbb{N}$,*

$$\max_{w \in \mathbb{R}^n} \mathbb{P}_A \left(\bigcup_{v \in \mathbb{S}^{n-1} \setminus \mathrm{Comp}\,(\delta,\rho)} \left\{ \|Av - w\|_2 \leqslant c\sqrt{n} \right\} \right) \leqslant 2e^{-cn},$$

where A is a random $n \times n$ symmetric $\{-1,1\}$-matrix.

The following lemma of Rudelson and Vershynin [129, Lemma 3.4] tells us that incompressible vectors are "flat" for a constant proportion of coordinates.

Lemma 2.9.3. *For $\delta, \rho \in (0,1)$, let $v \in \mathrm{Incomp}\,(\delta,\rho)$. Then*

$$(\rho/2)n^{-1/2} \leqslant |v_i| \leqslant \delta^{-1/2} n^{-1/2}$$

for at least $\rho^2 \delta n/2$ values of $i \in [n]$.

Now recall that we defined

$$\mathcal{I}(D) = \left\{ v \in \mathbb{S}^{n-1} : (\kappa_0 + \kappa_0/2)n^{-1/2} \leqslant |v_i| \leqslant (\kappa_1 - \kappa_0/2)n^{-1/2} \text{ for all } i \in D \right\}$$

and $\mathcal{I} = \bigcup_{D \subseteq [n], |D| = d} \mathcal{I}(D)$. Here we fix $\kappa_0 = \rho/3$ and $\kappa_1 = \delta^{-1/2} + \rho/6$, where δ, ρ are as in Lemma 2.9.2. We also fix $c_0 = \min\{2^{-24}, \rho\delta^{1/2}/2\}$.

The following lemma is what we will apply in the proof of Theorem 2.1.1.

Lemma 2.9.4. *For $n \in \mathbb{N}$, let $d < c_0^2 n$. Then*

$$\max_{w \in \mathbb{R}^n} \mathbb{P}_A \left(\bigcup_{v \in \mathbb{S}^{n-1} \setminus \mathcal{I}} \left\{ Av \in \{t \cdot w\}_{t>0},\ \|A\| \leqslant 4\sqrt{n} \right\} \right) \leqslant 16 c^{-1} e^{-cn}.$$

Proof. Apply Lemma 2.9.3 along with the definitions of κ_1, κ_2 and $\mathcal{I}$ to see $\mathbb{S}^{n-1} \setminus \mathcal{I} \subseteq \mathrm{Comp}\,(\delta,\rho)$. Now take a $c\sqrt{n}$-net $\mathcal{X}$ for $\{t \cdot w\}_{0 < t \leqslant 4\sqrt{n}}$ of size $8c^{-1}$. Then

$$\left\{ A : Av \in \{t \cdot w\}_{t>0},\ \|A\| \leqslant 4\sqrt{n} \right\} \subset \bigcup_{w' \in \mathcal{X}} \left\{ A : \|Av - w'\|_2 \leqslant c\sqrt{n} \right\}.$$

Union bounding over $\mathcal{X}$ and applying Lemma 2.9.2 completes the lemma. □

2.9.2 Proof of Theorem 2.1.1

As we noted in Section 2.2, matrices A with $\|A\| \geqslant 4\sqrt{n}$ will be a slight nuisance for us. The following concentration inequality for the operator norm of a random matrix will allow us to remove all such matrices A from consideration.

Lemma 2.9.5. *Let A be uniformly drawn from all $n \times n$ symmetric matrices with entries in $\{-1, 1\}$. Then for n sufficiently large,*

$$\mathbb{P}\left(\|A\| \geqslant 4\sqrt{n}\right) \leqslant 4e^{-n/32}.$$

This follows from a classical result of Bai and Yin [11] (see also [149, Theorem 2.3.23]) which implies that the median of $\|A\|$ is equal to $(2+o(1))\sqrt{n}$, combined with a concentration inequality due to Meckes [106, Theorem 2]. A version of Lemma 2.9.5 without explicit constants, is well-known and straightforward, though we have included a version with explicit constants for concreteness.

We will also need the following, rather weak, relationship between the threshold $\mathcal{T}_L$, defined in terms of the matrix M, and $\rho(v)$, the "one-dimensional" concentration function of v. For this we define one more bit of (standard) notation

$$\rho_\varepsilon(v) := \max_{b \in \mathbb{R}^n} \mathbb{P}\left(\sum_i v_i \varepsilon_i \in (b-\varepsilon, b+\varepsilon)\right).$$

Lemma 2.9.6. *Let $v \in \mathbb{S}^{n-1}$ and $\varepsilon = \mathcal{T}_L(v)$. Then $\rho_\varepsilon(v)^4 \leqslant 2^{12} L\varepsilon$.*

We postpone the proof of this lemma to Appendix B and move on to the proof of Theorem 2.1.1.

Proof of Theorem 2.1.1. It is not hard to see that $\mathbb{P}(\det(A) = 0) < 1$ for all n, and therefore it is enough to prove Theorem 2.1.1 for all sufficiently large n.

Now, as in Section 2.2, we set $\gamma = e^{-cn}$, where we now define, $c := L^{-32/c_0^2}/8$, $L := \max\{2^{26}C_1, 16/\kappa_0\}$, where $C_1 = C/c_0^6$ is the constant appearing in Theorem 2.7.1. We also let $c_0 > 0$ be as defined above and $d := \lceil c_0^2 n/2 \rceil$.

From Lemma 2.9.1 we have

$$\mathbb{P}(\det(A) = 0) \leqslant 16n \sum_{m=n}^{2n-2} \left(\gamma^{1/8} + \frac{q_{m-1}(\gamma)}{\gamma}\right)$$

and so it is enough to bound $q_n(\gamma)$ for all large n. Let $\Sigma = \{v \in \mathbb{S}^{n-1} : \rho(v) \geqslant \gamma\}$, as defined in Section 2.2, and note that

$$\{A : \exists v \in \mathbb{R}^n,\ Av = w,\ \rho(v) \geqslant \gamma\} \subset \{A : \exists v \in \Sigma,\ Av \in \{t \cdot w\}_{t>0}\}.$$

Since $d = \lceil c_0^2 n/2 \rceil$, by Lemma 2.9.4 and Lemma 2.9.5, we have

$$q_n(\gamma) \leqslant \max_{w\in\mathbb{R}^n} \mathbb{P}_A\left(\{\exists v \in \mathcal{I}\cap\Sigma:\ Av \in \{t\cdot w\}_{t>0}\}\cap\{\|A\|\leqslant 4\sqrt{n}\}\right) + 32c^{-1}e^{-cn} \tag{2.89}$$

and so it is enough to show the first term on the right-hand-side is $\leqslant 2^{-n}$. Using that $\mathcal{I} = \bigcup_D \mathcal{I}(D)$, we have the first term of (2.89) is

$$\leqslant 2^n \max_{D\in[n]^{(d)}} \max_{w\in\mathbb{R}^n} \mathbb{P}_A\left(\{\exists v \in \mathcal{I}(D)\cap\Sigma:\ Av \in \{t\cdot w\}_{t>0}\}\cap\{\|A\|\leqslant 4\sqrt{n}\}\right) \tag{2.90}$$

$$= 2^n \max_{w\in\mathbb{R}^n} \mathbb{P}_A\left(\{\exists v \in \mathcal{I}([d])\cap\Sigma:\ Av \in \{t\cdot w\}_{t>0}\}\cap\{\|A\|\leqslant 4\sqrt{n}\}\right), \tag{2.91}$$

where the last line holds by symmetry of the coordinates. Thus it is enough to show that the maximum at (2.91) is at most 4^{-n}.

Now, for $v\in\Sigma$ we have $\rho(v)\geqslant\gamma$ and so, by Lemma 2.9.6, we have that

$$\gamma^4 \leqslant \rho(v)^4 \leqslant \rho_{\mathcal{T}_L(v)}(v)^4 \leqslant 2^{12} L\mathcal{T}_L(v).$$

Define $\eta := \gamma^4/(2^{12}L) \leqslant \mathcal{T}_L(v)$. Also note that by definition, $\mathcal{T}_L(v)\leqslant 1/L\leqslant\kappa_0/8$.

Now, recalling definition (2.82) of $\Sigma_\varepsilon = \Sigma_\varepsilon([d])$ from Section 2.2, we may write

$$\mathcal{I}([d])\cap\Sigma \subseteq \bigcup_{i=1}^{n}\left\{v\in\mathcal{I}:\mathcal{T}_L(v)\in[2^{j-1}\eta, 2^j\eta]\right\} = \bigcup_{j=0}^{\log_2(\kappa_0/16\eta)} \Sigma_{2^j\eta}$$

and so by the union bound, it is enough to show

$$\max_{w\in\mathbb{R}^n} \mathbb{P}_A\left(\{\exists v\in\Sigma_\varepsilon:\ Av\in\{t\cdot w\}_{t>0}\}\cap\{\|A\|\leqslant 4\sqrt{n}\}\right)\leqslant 8^{-n},$$

for all $\varepsilon\in[\eta,\kappa_0/16]$. Fix an $\varepsilon\sqrt{n}$-net $\mathcal{X}$ for $\{t\cdot w\}_{0<t\leqslant 4\sqrt{n}}$ of size $8/\varepsilon\leqslant 2^n$ to get

$$\{A:\ Av\in\{t\cdot w\}_{t>0},\ \|A\|\leqslant 4\sqrt{n}\}\subset\bigcup_{w'\in\mathcal{X}}\{A:\ \|Av-w'\|_2\leqslant\varepsilon\sqrt{n},\ \|A\|\leqslant 4\sqrt{n}\}.$$

So by taking the union bound over $\mathcal{X}$ it is enough to prove that

$$Q_\varepsilon := \max_{w\in\mathbb{R}^n}\mathbb{P}_A\left(\{\exists v\in\Sigma_\varepsilon:\ \|Av-w\|_2\leqslant\varepsilon\sqrt{n}\}\cap\{\|A\|\leqslant 4\sqrt{n}\}\right)\leqslant 2^{-4n}. \tag{2.92}$$

Let $w\in\mathbb{R}^n$ be such that the maximum at (2.92) is attained. Now, since $\varepsilon<\kappa_0/8$ for $v\in\Sigma_\varepsilon$, we apply Lemma 2.8.2, to find a $u\in\mathcal{N}_\varepsilon=\mathcal{N}_\varepsilon([d])$ so that $\|v-u\|_2\leqslant 4\varepsilon$. So if $\|A\|\leqslant 4\sqrt{n}$ and $\|Av-w\|\leqslant\varepsilon\sqrt{n}$, we see that

$$\|Au-w\|_2\leqslant\|Av-w\|_2+\|A(v-u)\|_2\leqslant\|Av-w\|_2+\|A\|\|(v-u)\|_2\leqslant 32\varepsilon\sqrt{n}$$

and thus

$$\{A : \exists v \in \Sigma_\varepsilon : \ \|Av-w\| \leqslant \varepsilon\sqrt{n}\} \cap \{\|A\| \leqslant 4\sqrt{n}\} \subseteq \{A : \exists u \in \mathcal{N}_\varepsilon : \|Au-w\| \leqslant 32\varepsilon\sqrt{n}, \|A\| \leqslant 4\sqrt{n}\}.$$

So, by union bounding over our net $\mathcal{N}_\varepsilon$, we see that

$$Q_\varepsilon \leqslant \mathbb{P}_A \left(\exists u \in \mathcal{N}_\varepsilon : \|Au - w\| \leqslant 32\varepsilon\sqrt{n} \text{ and } \|A\| \leqslant 4\sqrt{n}\right) \leqslant \sum_{u \in \mathcal{N}_\varepsilon} \mathcal{L}_{A,op}\left(u, 32\varepsilon\sqrt{n}\right).$$

Now note that if $u \in \mathcal{N}_\varepsilon$, then $\mathcal{L}_{A,op}(u, \varepsilon\sqrt{n}) \leqslant (2^8 L\varepsilon)^n$ and so by Fact 2.6.2 we have that $\mathcal{L}_{A,op}\left(u, 32\varepsilon\sqrt{n}\right) \leqslant (2^{16}L\varepsilon)^n$. As a result,

$$Q_\varepsilon \leqslant |\mathcal{N}_\varepsilon|(2^{16}L\varepsilon)^n \leqslant \left(\frac{C}{L^2\varepsilon}\right)^n (2^{16}L\varepsilon)^n \leqslant 2^{-4n}.$$

where the second to last inequality follows from our Theorem 2.7.1 and the last inequality holds for our choice of $L = \max\{2^{26}C_1, 16/\kappa_0\}$. To see that the application of Theorem 2.7.1 is valid, note that

$$\log 1/\varepsilon \leqslant \log 1/\eta = \log 2^{12}L/\gamma^4 \leqslant nL^{-32/c_0^2}/2 + \log 2^{12}L \leqslant nL^{-32/c_0^2},$$

where the last inequality hold for all sufficiently large n. This completes the proof. □

Chapter 3

The least singular value of a random symmetric matrix

This chapter presents joint work with Matthew Jenssen, Marcus Michelen and Julian Sahasrabudhe. It is adapted from the paper [34] which has been submitted for publication.

3.1 Introduction

Let A be a $n \times n$ random symmetric matrix whose entries on and above the diagonal $(A_{i,j})_{i \leqslant j}$ are i.i.d. with mean 0 and variance 1.

In this chapter we study the extreme behavior of the *least singular value* of A, which we denote by $\sigma_{\min}(A)$. We prove a bound on this quantity which is optimal up to constants, for all random symmetric matrices with i.i.d. *subgaussian* entries. This confirms the folklore conjecture, explicitly stated by Vershynin in [166].

Theorem 3.1.1. *Let ζ be a subgaussian random variable with mean 0 and variance 1 and let A be a $n \times n$ random symmetric matrix whose entries above the diagonal $(A_{i,j})_{i \leqslant j}$ are independent and distributed according to ζ. Then for every $\varepsilon \geqslant 0$,*

$$\mathbb{P}_A(\sigma_{\min}(A) \leqslant \varepsilon n^{-1/2}) \leqslant C\varepsilon + e^{-cn}, \tag{3.1}$$

where $C, c > 0$ depend only on ζ.

We also prove a conjecture of Nguyen, Tao and Vu [116] on repeated eigenvalues.

Theorem 3.1.2. *Let ζ be a subgaussian random variable with mean 0 and variance 1 and let A be a $n \times n$ random symmetric matrix where $(A_{i,j})_{i \leqslant j}$ are independent and distributed according*

to ζ. Then A has no repeated eigenvalues with probability $1 - e^{-cn}$, where $c > 0$ is a constant depending only on ζ.

We actually prove the following stronger result, that is necessary in the proof of 3.1.1.

Theorem 3.1.3. *Let ζ be a subgaussian random variable with mean 0 and variance 1 and let A be a $n \times n$ random symmetric matrix where $(A_{i,j})_{i \leqslant j}$ are independent and distributed according to ζ. Then for each $\ell < cn$ and all $\varepsilon \geqslant 0$ we have*

$$\max_{k \leqslant n-\ell} \mathbb{P}(|\lambda_{k+\ell}(A) - \lambda_k(A)| \leqslant \varepsilon n^{-1/2}) \leqslant (C\varepsilon)^{\ell} + 2e^{-cn},$$

where $C, c > 0$ are constants, depending only on ζ.

3.1.1 Approximate negative correlation

Before we sketch the proof of Theorem 3.1.1, we highlight a technical theme of this chapter: the approximate negative correlation of certain "linear events". While this is only one of several new ingredients in this chapter, we isolate these ideas here, as they seem to be particularly amenable to wider application. We refer the reader to Section 3.2 for a more complete overview of the new ideas in this chapter.

We say that two events A, B in a probability space are *negatively correlated* if

$$\mathbb{P}(A \cap B) \leqslant \mathbb{P}(A)\mathbb{P}(B).$$

Here we state and discuss two *approximate* negative correlation results: one of which is a variant of Theorem 2.1.2, but is used in a entirely different context, and one of which is new.

We start by describing the latter result, which says that a "small ball" event is approximately negatively correlated with a large deviation event. This complements Theorem 2.1.2 which says that two "small ball events", of different types, are negatively correlated. In particular, we prove something in the spirit of the following inequality, though in a slightly more technical form, which will be sufficient for our purposes:

$$\mathbb{P}_X\big(|\langle X, v\rangle| \leqslant \varepsilon \text{ and } \langle X, u\rangle > t\big) \leqslant \mathbb{P}_X(|\langle X, v\rangle| \leqslant \varepsilon)\mathbb{P}_X(\langle X, u\rangle > t), \tag{3.2}$$

where u, v are unit vectors and $t, \varepsilon > 0$ and $X = (X_1, \ldots, X_n)$ with i.i.d. subgausian random variables with mean 0 and variance 1.

To state and understand our result, it makes sense to first consider in isolation the two events present in (3.2). The easier of the two events is $\langle X, u\rangle > t$, which is a large deviation event for

which we may apply the essentially sharp and classical inequality (see Chapter 3.4 in [167])

$$\mathbb{P}_X(\langle X, u\rangle > t) \leqslant e^{-ct^2},$$

where $c > 0$ is a constant depending only on the distribution of X.

We now turn to understand the more complicated small-ball event $|\langle X, v\rangle| \leqslant \varepsilon$ appearing in (3.2). Here, we have a more subtle interaction between v and the distribution of X, and thus we first consider the simplest possible case: when X has i.i.d. standard *gaussian* entries. Here, one may calculate

$$\mathbb{P}_X(|\langle X, v\rangle| \leqslant \varepsilon) \leqslant C\varepsilon, \tag{3.3}$$

for all $\varepsilon > 0$, where $C > 0$ is an absolute constant. However, as we depart from the case when X is gaussian, a much richer behavior emerges when the vector v admits some "arithmetic structure". For example, if $v = n^{-1/2}(1, \dots, 1)$ and the X_i are uniform in $\{-1, 1\}$ then

$$\mathbb{P}_X(|\langle X, v\rangle| \leqslant \varepsilon) = \Theta(n^{-1/2}),$$

for any $0 < \varepsilon < n^{-1/2}$. This, of course, stands in contrast to (3.3) for all $\varepsilon \ll n^{-1/2}$ and suggests that we employ an appropriate measure of the arithmetic structure of v.

For this, we use the notion of the "least common denominator" of a vector, introduced by Rudelson and Vershynin [129]. For parameters $\alpha, \gamma \in (0, 1)$ define the *Least Common Denominator* (LCD) of $v \in \mathbb{R}^n$ to be

$$D_{\alpha,\gamma}(v) := \inf\left\{\phi > 0 : \ \|\phi v\|_{\mathbb{T}} \leqslant \min\left\{\gamma\phi\|v\|_2, \sqrt{\alpha n}\right\}\right\}, \tag{3.4}$$

where $\|v\|_{\mathbb{T}} := \operatorname{dist}(v, \mathbb{Z}^n)$, for all $v \in \mathbb{R}^n$. What makes this definition useful is the important "inverse Littlewood-Offord theorem" of Rudelson and Vershynin [129], which tells us (roughly speaking) that one has (3.3) whenever $D_{\alpha,\gamma}(v) = \Omega(\varepsilon^{-1})$.

This notion of Least Common Denomonator is inspired by Tao and Vu's introduction and development of "inverse Littlewood-Offord theory", which is a collection of results guided by the meta-hypothesis: "If $\mathbb{P}_X(\langle X, v\rangle = 0)$ is large *then* v must have structure". We refer the reader to the paper of Tao and Vu [155] and the survery of Nguyen and Vu [118] for more background and history on inverse Littlewood-Offord theory and its role in random matrix theory.

We may now state our approximate version of (3.2), which uses $D_{\alpha,\gamma}(v)^{-1}$ as a proxy for $\mathbb{P}(|\langle X, v\rangle| \leqslant \varepsilon)$.

Theorem 3.1.4. *For $n \in \mathbb{N}$, $\varepsilon, t > 0$ and $\alpha, \gamma \in (0, 1)$, let $v \in \mathbb{S}^{n-1}$ satisfy $D_{\alpha,\gamma}(v) > C/\varepsilon$ and let $u \in \mathbb{S}^{n-1}$. Let ζ be a subgaussian random variable and let $X \in \mathbb{R}^n$ be a random vector whose*

coordinates are i.i.d. copies of ζ. Then

$$\mathbb{P}_X\left(|\langle X, v\rangle| \leqslant \varepsilon \text{ and } \langle X, u\rangle > t\right) \leqslant C\varepsilon e^{-ct^2} + e^{-c(\alpha n + t^2)},$$

where $C, c > 0$ depend only on γ and the distribution of ζ.

In fact, we need a significantly more complicated version of this result (Lemma 3.5.2) where the small-ball event $|\langle X, v\rangle| \leqslant \varepsilon$ is replaced with a small-ball event of the form

$$|f(X_1, \dots, X_n)| \leqslant \varepsilon,$$

where f is a quadratic polynomial in variables $X_1, \dots, X_n$. The proof of this result is carried out in Section 3.5 and is an important aspect of this chapter. Theorem 3.1.4, on the other hand, is only stated here to illustrate the general flavor of this result and is not actually used in this chapter. We do provide a proof in Appendix A of [34] for completeness and to suggest further inquiry into inequalities of the form (3.2).

We now turn to discuss our second approximate negative dependence result, which deals with the intersection of two different small ball events. TA variant of this theorem was proved in Chapter 2, but is put to a different use here. This result tells us that the events

$$|\langle X, v\rangle| \leqslant \varepsilon \qquad \text{and} \qquad |\langle X, w_1\rangle| \ll 1, \dots, |\langle X, w_k\rangle| \ll 1, \tag{3.5}$$

are approximately negatively correlated, where $X = (X_1, \dots, X_n)$ is a vector with i.i.d. subgaussian entries and $w_1, \dots, w_k$ are orthonormal. That is, we prove something in the spirit of

$$\mathbb{P}_X\left(\{|\langle X, v\rangle| \leqslant \varepsilon\} \cap \bigcap_{i=1}^{k} \{|\langle X, w_i\rangle| \ll 1\}\right) \leqslant \mathbb{P}_X\big(|\langle X, v\rangle| \leqslant \varepsilon\big)\mathbb{P}_X\left(\bigcap_{i=1}^{k} \{|\langle X, w_i\rangle| \ll 1\}\right),$$

though in a more technical form.

To understand our result, again it makes sense to consider the two events in (3.5) in isolation. Since we have already discussed the subtle event $|\langle X, v\rangle| \leqslant \varepsilon$, it remains only to consider the event on the right of (3.5). Returning to the gaussian case, we note that if X has independent standard gaussian entries, then one may compute directly that

$$\mathbb{P}_X\left(|\langle X, w_1\rangle| \ll 1, \dots, |\langle X, w_k\rangle| \ll 1\right) = \mathbb{P}(|X_1| \ll 1, \dots |X_k| \ll 1) \leqslant e^{-\Omega(k)}, \tag{3.6}$$

by rotational invariance of the gaussian. Here the generalization to other random variables is not as subtle, and the well-known Hanson-Wright inequality tells us that (3.6) holds more generally when X has general i.i.d. subgaussian entries.

Our innovation in this line is our second "approximate negative correlation theorem", which allows us to control these two events *simultaneously*. Again we use $D_{\alpha,\gamma}(v)^{-1}$ as a proxy for $\mathbb{P}(|\langle X, v\rangle| \leqslant \varepsilon)$.

Here, for ease of exposition, we state a less general version for $X = (X_1, \dots, X_n) \in \{-1, 0, 1\}$ with i.i.d. c-lazy coordinates, meaning that $\mathbb{P}(X_i = 0) \geqslant 1 - c$. Our theorem is stated in full generality in Section 3.9, see Theorem 3.9.2.

Theorem 3.1.5. *Let $\gamma \in (0,1)$, $d \in \mathbb{N}$, $\alpha \in (0,1)$, $0 \leqslant k \leqslant c_1 \alpha d$ and $\varepsilon \geqslant \exp(-c_1 \alpha d)$. Let $v \in \mathbb{S}^{d-1}$, let $w_1, \dots, w_k \in \mathbb{S}^{d-1}$ be orthogonal and let W be the matrix with rows $w_1, \dots, w_k$.*

If $X \in \{-1, 0, 1\}^d$ is a $1/4$-lazy random vector and $D_{\alpha,\gamma}(v) > 16/\varepsilon$ then

$$\mathbb{P}_X\left(|\langle X, v\rangle| \leqslant \varepsilon \text{ and } \|WX\|_2 \leqslant c_2\sqrt{k}\right) \leqslant C\varepsilon e^{-c_1 k},$$

where $C, c_1, c_2 > 0$ are constants, depending only on γ.

In this chapter we will put Theorem 3.1.5 to a very different use to that in Chapter 2, where we used it to prove a version of the following statement.

> Let $v \in \mathbb{S}^{d-1}$ be a vector on the sphere and let H be a $n \times d$ random $\{-1, 0, 1\}$-matrix *conditioned on the event* $\|Hv\|_2 \leqslant \varepsilon n^{1/2}$, for some $\varepsilon > e^{-cn}$. Here $d = cn$ and $c > 0$ is a sufficiently small constant. Then the probability that the rank of H is $n - k$ is $\leqslant e^{-ckn}$.

In this chapter we use (the generalization of) Theorem 3.1.5 to obtain good bounds on quantities of the form

$$\mathbb{P}_X(\|BX\|_2 \leqslant \varepsilon n^{1/2}),$$

where B is a fixed matrix with an exceptionally large eigenvalue (possibly as large as e^{cn}), but is otherwise pseudo-random, meaning (among other things) that the rest of the spectrum does not deviate too much from that of a random matrix. We use Theorem 3.1.5 to decouple the interaction of X with the largest eigenvector of B, from the interaction of X with the rest of B. We refer the reader to (3.16) in the sketch in Section 3.2 and to Section 3.9 for more details.

The proof of Theorem 3.9.2 follows closely along the lines of the proof of Theorem 2.1.2, requiring only technical modifications and adjustments. So as not to distract from the new ideas presented in this chapter, we have sidelined this proof to the supplementary paper [36].

Finally we note that it may be interesting to investigate to what extent one may sharpen these approximate negative correlation theorems in the direction of their idealized forms.

3.2 Proof sketch

Here we sketch the proof of Theorem 3.1.1. We begin by giving the rough "shape" of the proof, while making a few simplifying assumptions, (3.8) and (3.9). We shall then come to discuss the substantial new ideas of this chapter in Section 3.2.2 where we describe the considerable lengths we must go to, in order to remove our simplifying assumptions. Indeed, if one were to only tackle these assumptions using standard tools, one cannot hope for a bound much better than $\varepsilon^{1/3}$ in Theorem 3.1.1 (see Section 3.2.2.2).

3.2.1 The shape of the proof

Recall that A_{n+1} is a $(n+1) \times (n+1)$ random symmetric matrix with subgaussian entries. Let $X := X_1, \dots, X_{n+1}$ be the columns of A_{n+1}, let

$$V = \mathrm{Span}\{X_2, \dots, X_{n+1}\}$$

and let A_n be the matrix A_{n+1} with the first row and column removed. We now use an important observation from Rudelson and Vershynin [129] that allows for a geometric perspective on the least singular value problem[1]

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2}) \lesssim \mathbb{P}(\mathrm{dist}(X, V) \leqslant \varepsilon).$$

Here our first significant challenge presents itself: X and V are not independent and thus the event $\mathrm{dist}(X, V) \leqslant \varepsilon$ is hard to understand directly. However, one can establish a formula for $\mathrm{dist}(X, V)$ that is a rational function in the vector X with coefficients that depend only on V. This brings us to the useful inequality[2] due to Vershynin [166],

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2}) \lesssim \sup_{r \in \mathbb{R}} \mathbb{P}_{A_n, X}\big(|\langle A_n^{-1} X, X\rangle - r| \leqslant \varepsilon \|A_n^{-1} X\|_2\big), \tag{3.7}$$

where we are ignoring the possibility of A_n being singular for now. We thus arrive at our main technical focus of this chapter, which is bounding the quantity on the right-hand-side of (3.7).

We now make our two simplifying assumptions that shall allow us to give the overall shape of our proof without any added complexity. We then layer-on further complexities as we discuss how to remove these assumptions.

[1] Here and throughout we understand $A \lesssim B$ to mean that there exists an absolute constant $C > 0$ for which $A \leqslant CB$.

[2] In this sketch we will be ignoring a few exponentially rare events, and so the inequalities listed here should be understood as "up to an additive error of e^{-cn}."

As a first simplifying assumption, let us assume that the collection of X that dominates the probability at (3.7) satisfies

$$\|A_n^{-1}X\|_2 \approx \|A_n^{-1}\|_{\mathrm{HS}}, \tag{3.8}$$

where we point out that $\|A_n^{-1}\|_{\mathrm{HS}}^2 = \mathbb{E}_X \|A_n^{-1}X\|_2^2$. This seems to be a fairly innocent assumption at first, as the Hanson-Wright inequality tells us that $\|A_n^{-1}X\|_2$ is concentrated about its mean, for all reasonable A^{-1}. However, as we will see, mere concentration is not enough for us here.

As a second assumption, let us assume that the relevant A_n in the right-hand-side of (3.7) satisfies

$$\|A_n^{-1}\|_{\mathrm{HS}} \approx cn^{1/2}. \tag{3.9}$$

This turns out to be a *very* delicate assumption, as we will soon see, but is not entirely unreasonable to make for the moment: for example we have $\|A_n^{-1}\|_{\mathrm{HS}} = \Theta_\delta(n^{1/2})$ with probability $1-\delta$. This, for example, follows from Vershynin's theorem [166] along with Corollary 3.8.5, which is based on the work of [57].

With these assumptions, our task reduces to proving

$$\min_r \mathbb{P}_X\big(|\langle A^{-1}X, X\rangle - r| \leqslant \varepsilon n^{1/2}\big) \lesssim \varepsilon, \tag{3.10}$$

for all $\varepsilon > e^{-cn}$, where we have written $A^{-1} = A_n^{-1}$ and think of A^{-1} as a fixed (pseudo-random) matrix.

We observe, for a general fixed matrix A^{-1} there is no hope in proving such an inequality: indeed if $A^{-1} = n^{-1/2}J$, where J is the all-ones matrix, then the left-hand-side of (3.10) is $\geqslant cn^{-1/2}$ for *all* $\varepsilon > 0$, falling vastly short of our desired (3.10).

Thus, we need to introduce a collection of fairly strong "quasi-randomness properties" of A that hold with probably $1-e^{-cn}$. These will ensure that A^{-1} is sufficiently "non-structured" to make our goal (3.10) possible. The most important and of these quasi-randomness conditions is to show that all of the eigenvectors v of A satisfy

$$D_{\alpha,\gamma}(v) > e^{cn},$$

for some appropriate α, γ, where $D_{\alpha,\gamma}(v)$ is the *least common denominator* of v defined at (3.4). Roughly this means that none of the eigenvectors of A "correlate" with a re-scaled copy of the integer lattice $t\mathbb{Z}^n$, for any $e^{-cn} \leqslant t \leqslant 1$.

To prove that these quasi-randomness properties hold with probability $1-e^{-cn}$ is a difficult task and depends fundamentally on the ideas presented in Chapter 2. The details are carried out in a supplementary paper [36].

With these quasi-randomness conditions in tow, we can return to (3.10) and apply Esseen's inequality to bound the left-hand-side of (3.10) in terms of the characteristic function $\varphi(\theta)$ of the random variable $\langle A^{-1}X, X\rangle$,

$$\min_r \mathbb{P}_X\left(|\langle A^{-1}X, X\rangle - r| \leqslant \varepsilon n^{1/2}\right) \lesssim \varepsilon \int_{-1/\varepsilon}^{1/\varepsilon} |\varphi(\theta)|\, d\theta.$$

While this maneuver has been quite successful in work on characteristic functions for (linear) sums of independent random variables, the characteristic function of such quadratic functions has proved to be a more elusive object. For example, even the analogue of the Littlewood-Offord theorem is not fully understood in the quadratic case [40, 108]. Here, we appeal to our quasi-random conditions to avoid some of the traditional difficulties: we use an application of Jensen's inequality to *decouple* the quadratic form and bound $\varphi(\theta)$ point-wise in terms of an average over a related collection of characteristic functions of *linear* sums of independent random variables

$$|\varphi(\theta)|^2 \leqslant \mathbb{E}_Y |\varphi(A^{-1}Y; \theta)|,$$

where Y is a random vector with i.i.d. entries and $\varphi(v;\theta)$ denotes the characteristic function of the sum $\sum_i v_i X_i$, where X_i are i.i.d. distributed according to the original distribution ζ. We can then use our pseudo-random conditions on A to bound

$$|\varphi(A^{-1}Y;\theta)| \lesssim \exp\left(-c\theta^2\right),$$

for all but exponentially few Y, allowing us to show

$$\int_{-1/\varepsilon}^{1/\varepsilon} |\varphi(\theta)|\, d\theta \leqslant \int_{-1/\varepsilon}^{1/\varepsilon} \left[\mathbb{E}_Y |\varphi(A^{-1}Y;\theta)|\right]^{1/2} \leqslant \int_{-1/\varepsilon}^{1/\varepsilon} \left(\exp\left(-c\theta^2\right) + e^{-cn}\right)\, d\theta = O(1)$$

and thus completing the proof, up to our simplifying assumptions.

3.2.2 Removing the simplifying assumptions

While this is a good story to work with, the challenge starts when we turn to remove our simplifying assumptions (3.8), (3.9). We also note that if one only applies standard methods to remove these assumptions, then one would get stuck at the "base case" outlined below. We start by discussing how to remove the simplifying assumption (3.9), whose resolution governs the overall structure of the chapter.

3.2.2.1 Removing the assumption (3.9)

What is most concerning about making the assumption $\|A_n^{-1}\|_{\mathrm{HS}} \approx n^{-1/2}$ is that it is, in a sense, *circular*: If we assume the modest-looking hypothesis $\mathbb{E}\|A^{-1}\|_{\mathrm{HS}} \lesssim n^{1/2}$, we would be able to deduce

$$\mathbb{P}(\sigma_{\min}(A_n) \leqslant \varepsilon/n^{1/2}) = \mathbb{P}(\sigma_{\max}(A_n^{-1}) \geqslant n^{1/2}/\varepsilon) \leqslant \mathbb{P}(\|A_n^{-1}\|_{\mathrm{HS}} \geqslant n^{1/2}/\varepsilon) \lesssim \varepsilon,$$

by Markov. In other words, showing that $\|A^{-1}\|_{\mathrm{HS}}$ is concentrated about $n^{-1/2}$ (in the above sense) actually *implies* Theorem 3.1.1. However this is not as worrisome as one might first suspect: if we are trying to prove Theorem 3.1.1 for $(n+1)\times(n+1)$ matrices using the above outline, we only need to control the Hilbert-Schmidt norm of the inverse of the *minor* A_n^{-1}. This suggests an inductive or (as we use) an *iterative* "bootstrapping argument" to successively improve the bound. Thus, in effect, we look to prove

$$\mathbb{E}_A \|A^{-1}\|_{\mathrm{HS}}^{\alpha} \mathbb{1}(\sigma_{\min}(A_n) \geqslant e^{-cn}) \approx n^{\alpha/2},$$

for successively larger $\alpha \in (0,1)$. Note we have to cut out the event of A being singular from our expectation, as this event has non-zero probability.

3.2.2.2 Base case

In the first step of our iteration, we prove a "base case" of

$$\mathbb{P}(\sigma_{\min}(A_n) \leqslant \varepsilon/\sqrt{n}) \lesssim \varepsilon^{1/4} + e^{-cn} \tag{3.11}$$

without the assumption (3.9) which is equivalent to

$$\mathbb{E}_{A_n} \|A_n^{-1}\|_{\mathrm{HS}}^{1/4} \mathbb{1}(\sigma_{\min}(A_n) \geqslant e^{-cn}) \approx n^{1/8}.$$

To prove this "base case" we upgrade (3.7) to

$$\mathbb{P}\left(\sigma_{\min}(A_{n+1}) \leqslant \frac{\varepsilon}{\sqrt{n}}\right) \lesssim \varepsilon + \sup_{r\in\mathbb{R}} \mathbb{P}\left(\frac{|\langle A_n^{-1}X, X\rangle - r|}{\|A_n^{-1}X\|_2} \leqslant C\varepsilon, \|A_n^{-1}\|_{\mathrm{HS}} \leqslant \frac{n^{1/2}}{\varepsilon}\right). \tag{3.12}$$

In other words, we can intersect with the event

$$\|A_n^{-1}\|_{\mathrm{HS}} \leqslant n^{1/2}/\varepsilon \tag{3.13}$$

at a loss of only $C\varepsilon$ in probability.

We then push through the proof outlined in Section 3.2.1 to obtain our initial weak bound of (3.11). For this, we first use the Hanson-Wright inequality to give a weak version of (3.8), and then use (3.13) as a weak version of our assumption(3.9). We note that this base step (3.11) already improves the best known bounds on the least singular value problem for random symmetric matrices.

3.2.2.3 Bootstrapping

To improve on this bound we use a "bootstrapping" lemma which, after applying it three times, allows us to improve (3.11) to the near-optimal result

$$\mathbb{P}(\sigma_{\min}(A_n) \leqslant \varepsilon/\sqrt{n}) \lesssim \varepsilon\sqrt{\log 1/\varepsilon} + e^{-cn}\,. \tag{3.14}$$

Proving this bootstrapping lemma essentially reduces to the problem of getting good estimates on

$$\mathbb{P}_X\left(\|A^{-1}X\|_2 \leqslant s\right), \tag{3.15}$$

for $s \in (\varepsilon, n^{-1/2})$, where A is a matrix with $\|A^{-1}\|_{op} = \delta^{-1}$ and $\delta \in (\varepsilon, cn^{-1/2})$ but is "otherwise pseudo-random". Here we require two additional ingredients.

To start unpacking (3.15), we use that $\|A^{-1}\|_{op} = \delta^{-1}$ to see that if v is a unit eigenvector corresponding to the largest eigenvalue of A^{-1} then

$$\|A^{-1}X\|_2 \leqslant s \text{ implies that } |\langle X, v\rangle| < \delta s.$$

While this leads to a decent first bound of $O(\delta s)$ on the probability (3.15) (after using the quasi-randomness properties of A), it is not enough for our purposes and we have to use that X must *also* have small inner product with many other eigenvectors of A (assuming s is sufficiently small). Working along these lines, we show that (3.15) is bounded above by

$$\mathbb{P}_X\Big(|\langle X, v_1\rangle| \leqslant s\delta \text{ and } |\langle X, v_i\rangle| \leqslant \sigma_i s \text{ for all } i = 2, \dots, n-1\Big), \tag{3.16}$$

where w_i is a unit eigenvector of A corresponding to the singular value $\sigma_i = \sigma_i(A)$. Now, appealing to the quasi-random properties of the eigenvectors of A^{-1}, we may apply our approximate negative correlation theorem (Theorem 3.1.5) to see that (3.16) is at most

$$O(\delta s)\exp(-cN_A(-c/s, c/s)) \tag{3.17}$$

where $c > 0$ is a constant and $N_A(a, b)$ denotes the number of eigenvalues of the matrix A in the interval (a, b). The first $O(\delta s)$ factor comes from the event $|\langle X, v_1\rangle| \leqslant s\delta$ and the second

factor comes from approximating

$$\mathbb{P}_X\left(|\langle X, w_i\rangle| < c \text{ for all } i \text{ s.t. } s\sigma_i < c\right) \approx \exp(-cN_A(-c/s, c/s))\,. \tag{3.18}$$

This bound is now sufficiently strong for our purposes, *provided* the spectrum of A adheres sufficiently closely to the typical spectrum of A.

This now leads us to understand the rest of the spectrum of A_n and, in particular, the next smallest singular values $\sigma_{n-1}, \sigma_{n-2}, \ldots$. This might seem like a step backwards as we are now forced to understand the behavior of *many* singular values and not just the smallest. However, this "loss" is outweighed by the fact that we need only to understand these eigenvalues (for the most part) on scales of size $\Omega(n^{-1/2})$, which is now well understood due to the important work of Erdős, Schlein and Yau [57]. Although some additional information is needed about how the eigenvalues cluster on much smaller scales. Here we can use the work of Nguyen [115] and our own Theorem 3.1.3 on the crowding of the spectrum.

These results ultimately allow us to derive sufficiently strong results on quantities of the form (3.15), which in-turn allow us to prove our "bootstrapping lemma". We then use this lemma to prove the near-optimal result

$$\mathbb{P}(\sigma_{\min}(A_n) \leqslant \varepsilon/\sqrt{n}) \lesssim \varepsilon\sqrt{\log 1/\varepsilon} + e^{-cn}\,. \tag{3.19}$$

3.2.2.4 Removing the assumption (3.8) and the last jump to Theorem 3.1.1

We now turn to discuss how to remove our simplifying assumption (3.8), made above, which will allow us to close the gap between (3.19) and Theorem 3.1.1.

To achieve this, we need to consider how $\|A^{-1}X\|_2$ varies about $\|A^{-1}\|_{\mathrm{HS}}$. Now, the Hanson-Wright inequality tells us that indeed $\|A^{-1}X\|_2$ is concentrated about $\|A^{-1}\|_{\mathrm{HS}}$, on the scale of $\lesssim \|A^{-1}\|_{op}$. While this is certainly useful for us, it is far from enough to prove Theorem 3.1.1. For this, we need to rule out any "macroscopic" correlation between the events

$$\{|\langle A^{-1}X, X\rangle - r| < K\varepsilon\|A^{-1}\|_{\mathrm{HS}}\} \text{ and } \{\|A^{-1}X\|_2 > K\|A^{-1}\|_{\mathrm{HS}}\} \tag{3.20}$$

for all $K > 0$. Our first step towards understanding (3.20) is to replace the *quadratic* large deviation event $\|A^{-1}X\|_2 > K\|A^{-1}\|_{\mathrm{HS}}$ with a collection of *linear* large deviation events:

$$\langle X, w_i\rangle > K\log(i+1),$$

where $w_n, w_{n-1}, \dots, w_1$ are the eigenvectors of A corresponding to singular values $\sigma_n \leqslant \sigma_{n-1} \leqslant \dots \leqslant \sigma_1$ respectively and the $\log(i+1)$ factor should be seen as a weight function that assigns more weight to the smaller singular values.

Interestingly, we run into a similar obstacle as before (although we don't go into details here): if the "bulk" spectrum of A^{-1} is behaving erratically this replacement step will be too lossy for our purposes. Thus we are lead to prove another result showing that the spectrum of A^{-1} adheres sufficiently to its typical spectrum. This reduces to proving

$$\mathbb{E}_{A_n}\left[\frac{\sum_{i=1}^n \sigma_{n-i-1}^{-2}(\log i)^2}{\sum_{i=1}^n \sigma_{n-i-1}^{-2}}\right] = O(1),$$

where the left-hand-side is a statistic which measures the degree of distortion of the smallest singular values of A_n. To prove this we again lean on the works of Erdős, Schlein and Yau [57], Nguyen [115] and our own Theorem 3.1.3 on the crowding of the spectrum.

Thus we have reduced the task of proving the approximate independence of the events at (3.20) to proving the approximate independence of the collection of events

$$\{|\langle A^{-1}X, X\rangle - r| < K\varepsilon\|A^{-1}\|_{\mathrm{HS}}\} \text{ and } \{\langle v_i, X\rangle > K\log(i+1)\}.$$

This is something, it turns out, that we can handle on the Fourier side by using a quadratic analogue of our negative correlation inequality, Theorem 3.1.4. The idea here is to prove an Esseen-type bound of the form

$$\mathbb{P}(|\langle A^{-1}X, X\rangle - t| < \delta, \langle X, u\rangle \geqslant s) \lesssim \delta e^{-s}\int_{-1/\delta}^{1/\delta}\left|\mathbb{E}e^{2\pi i\theta\langle A^{-1}X,X\rangle + \langle X,u\rangle}\right| d\theta\,. \tag{3.21}$$

Which introduces this extra "exponential tilt" to the characteristic function. From here one can carry out the plan sketched in Section 3.2.1 with this more complicated version of Esseen, then integrate over s to upgrade (3.19) to Theorem 3.1.1.

3.2.3 Outline of the rest of the chapter

In the next short section we introduce some key definitions, notation, and preliminaries that we use throughout this chapter. In Section 3.4 we establish a collection of crucial quasi-randomness properties that hold for the random symmetric matrix A_n with probability $1-e^{-\Omega(n)}$ and that we condition on for most of the chapter. In Section 3.5 we detail our Fourier decoupling argument and establish an inequality of the form (3.21). This allows us to prove our new approximate negative correlation result Lemma 3.5.2. In Section 3.6 we prepare the ground for our iterative argument by establishing (3.12), thereby switching our focus to the study of the quadratic

form $\langle A_n^{-1}X, X\rangle$. In Section 3.7 we prove Theorem 3.1.2 and Theorem 3.1.3, which tell us that the eigenvalues of A cannot 'crowd' small intervals. In Section 3.8 we establish regularity properties for the bulk of the spectrum of A^{-1}. In Section 3.9 we deploy the approximate negative correlation result (Theorem 3.1.5) in order to carry out the portion of the proof sketched between (3.15) and (6.9). In Section 3.10 we establish our base step (3.11) and bootstrap this to prove the near optimal bound (3.19). In the final section, Section 3.11, we complete the proof of our main Theorem 3.1.1.

3.3 Key Definitions and Preliminaries

We first need a few notions out of the way which are related to Chapter 2.

3.3.1 Subgaussian and matrix definitions

Throughout, ζ will be a mean-zero, variance 1 random variable. We define the *subgaussian moment* of ζ to be

$$\|\zeta\|_{\psi_2} := \sup_{p\geqslant 1} \frac{1}{\sqrt{p}}(\mathbb{E}|\zeta|^p)^{1/p}.$$

A mean 0, variance 1 random variable is said to be *subgaussian* if $\|\zeta\|_{\psi_2}$ is finite. We define Γ be the set of subgaussian random variables and, for $B > 0$, we define $\Gamma_B \subseteq \Gamma$ to be subset of ζ with $\|\zeta\|_{\psi_2} \leqslant B$.

For $\zeta \in \Gamma$, define $\mathrm{Sym}_n(\zeta)$ to be the probability space on $n \times n$ symmetric matrices A for which $(A_{i,j})_{i\geqslant j}$ are independent and distributed according to ζ. Similarly, we write $X \sim \mathrm{Col}_n(\zeta)$ if $X \in \mathbb{R}^n$ is a random vector whose coordinates are i.i.d. copies of ζ.

We shall think of the spaces $\{\mathrm{Sym}_n(\zeta)\}_n$ as coupled in the natural way: the matrix $A_{n+1} \sim \mathrm{Sym}_{n+1}(\zeta)$ can be sampled by first sampling $A_n \sim \mathrm{Sym}_n(\zeta)$, which we think of as the principle minor $(A_{n+1})_{[2,n+1]\times[2,n+1]}$, and then generating the first row and column of A_{n+1} by generating a random column $X \sim \mathrm{Col}_n(\zeta)$. In fact it will make sense to work with a random $(n+1)\times(n+1)$ matrix, which we call A_{n+1} throughout. This is justified as much of the work is done with the principle minor A_n of A_{n+1}, due to the bound (3.7) as well as Lemma 3.6.1.

3.3.2 Compressible vectors

We shall require the now-standard notions of *compressible vectors* as defined by Rudelson and Vershynin [129].

For parameters $\rho, \delta \in (0,1)$, we define the set of compressible vectors $\mathrm{Comp}(\delta, \rho)$ to be the set of vectors in $\mathbb{S}^{n-1}$ that are distance at most ρ from a vector supported on at most δn coordinates. We then define the set of *incompressible* vectors to be all other unit vectors, i.e. $\mathrm{Incomp}(\delta, \rho) := \mathbb{S}^{n-1} \setminus \mathrm{Comp}(\delta, \rho)$. The following basic fact about incompressible vectors from [129] will be useful throughout:

Fact 3.3.1. *For each $\delta, \rho \in (0,1)$ there is a constant $c_{\rho,\delta} \in (0,1)$ so that for all $v \in \mathrm{Incomp}(\delta, \rho)$ we have that $|v_j|\sqrt{n} \in [c_{\rho,\delta}, c_{\rho,\delta}^{-1}]$ for at least $c_{\rho,\delta} n$ values of j.*

Fact 3.3.1 assures us that for each incompressible vector we can find a large subvector that is "flat." Using the work of Vershynin [166], we will safely be able to ignore compressible vectors. In particular, [166, Proposition 4.2] implies the following Lemma. We refer the reader to Appendix B in [34] for details.

Lemma 3.3.2. *For $B > 0$ and $\zeta \in \Gamma_B$, let $A \sim \mathrm{Sym}_n(\zeta)$. Then there exist constants $\rho, \delta, c \in (0,1)$, depending only on B, so that*

$$\sup_{u \in \mathbb{R}^n} \mathbb{P}\big(\exists x \in \mathrm{Comp}(\delta, \rho), \exists t \in \mathbb{R} : Ax = tu\big) \leqslant 2e^{-cn}$$

and

$$\mathbb{P}\big(\exists u \in \mathrm{Comp}(\delta, \rho), \exists t \in \mathbb{R} : Au = tu\big) \leqslant 2e^{-cn}.$$

The first statement says, roughly, that $A^{-1}u$ is incompressible for each fixed u; the second states that all unit eigenvectors are incompressible.

Remark 3.3.3 (Choice of constants, $\rho, \delta, c_{\rho,\delta}$). Throughout, we let ρ, δ denote the constants guaranteed by Lemma 3.3.2 and $c_{\rho,\delta}$ the corresponding constant from Fact 3.3.1. These constants shall appear throughout this chapter and shall always be considered as fixed.

Lemma 3.3.2 follows easily from [166, Proposition 4.2] with a simple net argument.

3.3.3 Notation

We quickly define some notation. For a random variable X, we use the notation $\mathbb{E}_X$ for the expectation with respect to X and we use the notation $\mathbb{P}_X$ analogously. For an event $\mathcal{E}$, we write $\mathbf{1}_{\mathcal{E}}$ or $\mathbf{1}\{\mathcal{E}\}$ for the indicator function of the event $\mathcal{E}$. We write $\mathbb{E}^{\mathcal{E}}$ to be the expectation defined by $\mathbb{E}^{\mathcal{E}}[\,\cdot\,] = \mathbb{E}[\,\cdot\,\mathbf{1}_{\mathcal{E}}]$. For a vector $v \in \mathbb{R}^n$ and $J \subset [n]$, we write v_J for the vector whose ith coordinate is v_i if $i \in J$ and 0 otherwise.

We shall use the notation $X \lesssim Y$ to indicate that there exists a constant $C > 0$ for which $X \leqslant CY$. In a slight departure from convention, *we will always allow this constant to depend on the subgaussian constant B*, if present. We shall also let our constants implicit in big-O

notation to depend on B, if this constant is relevant in the context. We hope that we have been clear as to where the subgaussian constant is relevant, and so this convention is to just reduce added clutter.

3.4 Quasirandomness properties

In this technical section, we define a list of "quasi-random" properties of A_n that hold with probability $1-e^{-\Omega(n)}$. This probability is large enough that we can assume that these properties hold for all the principle minors of A_{n+1}. Showing that several of these quasi-random properties hold with probability $1-e^{-\Omega(n)}$ will prove to be a challenging task and our proof will depend deeply on ideas presented in Chapter 2. Most of this work is not present in this chapter, but is available in a supplementary paper [36].

3.4.1 Defining the properties

It will be convenient to assume throughout that every minor of A_{n+1} is invertible and so we will perturb the matrix slightly so that we may assume this. If we add to A_{n+1} an independent random symmetric matrix whose upper triangular entries are independent gaussian random variables with mean 0 and variance n^{-n}, then with probability $1-e^{-\Omega(n)}$ the singular values of A_{n+1} move by at most, say, $n^{-n/3}$. Further, after adding this random gaussian matrix, every minor of the resulting matrix is invertible with probability 1. Thus, we will assume without loss of generality throughout that every minor of A_{n+1} is invertible.

In what follows, we let $A = A_n \sim \mathrm{Sym}_n(\zeta)$ and let $X \sim \mathrm{Col}_n(\zeta)$ be a random vector, independent of A. Our first quasi-random property is standard from the concentration of the operator norm of a random symmetric matrix. We define $\mathcal{E}_1$ by

$$\mathcal{E}_1 = \{\|A\|_{op} \leqslant 4\sqrt{n}\}. \tag{3.22}$$

For the next property we need a definition. Let $X, X' \sim \mathrm{Col}_n(\zeta)$ and define the random vector in $\mathbb{R}^n$ as $\tilde{X} := X_J - X'_J$, where $J \subseteq [n]$ is a μ-random subset, i.e. for each $j \in [n]$ we have $j \in J$ independently with probability μ. The reason behind this definition is slightly opaque at present, but will be clear in the context of Lemma 3.5.2 in Section 3.5. Until we get there it is reasonable to think of $\tilde{X}$ as being essentially X; in particular, it is a random vector with i.i.d. subgaussian entries with mean 0 and variance μ. We now define $\mathcal{E}_2$ to be the event in A defined by

$$\mathcal{E}_2 = \left\{\mathbb{P}_{\tilde{X}}\left(A^{-1}\tilde{X}/\|A^{-1}\tilde{X}\|_2 \in \mathrm{Comp}\,(\delta,\rho)\right) \leqslant e^{-c_2 n}\right\}. \tag{3.23}$$

We remind the reader that $\mathrm{Comp}\,(\delta,\rho)$ is defined in Section 3.3.2, and $\delta,\rho\in(0,1)$ are constants, fixed throughout this chapter, and chosen according to Lemma 3.3.2. In the (rare) case that $\widetilde{X}=0$, we interpret $\mathbb{P}_{\widetilde{X}}(A^{-1}\widetilde{X}/\|A^{-1}\widetilde{X}\|_2\in\mathrm{Comp}\,(\delta,\rho))=1$

Recalling the least common denominator defined at (3.4), we now define the event $\mathcal{E}_3$ by

$$\mathcal{E}_3=\{D_{\alpha,\gamma}(u)\geqslant e^{c_3 n}\text{ for every unit eigenvector }u\text{ of }A\}\,. \tag{3.24}$$

The next condition tells us that the random vector $A^{-1}\widetilde{X}$ is typically unstructured. We will need a slightly stronger notion of structure than just looking at the LCD, in that we will need all sufficiently large subvectors to be unstructured. For $\mu\in(0,1)$, define the *subvector least common denominator* as

$$\hat{D}_{\alpha,\gamma,\mu}(v):=\min_{\substack{I\subset[n]\\|I|\geqslant(1-2\mu)n}} D_{\alpha,\gamma}\left(v_I/\|v_I\|_2\right)\,.$$

If we define the random vector $v=v(\widetilde{X}):=A^{-1}\widetilde{X}$, then we define $\mathcal{E}_4$ to be the event that A satisfies

$$\mathcal{E}_4=\left\{\mathbb{P}_{\widetilde{X}}\left(\hat{D}_{\alpha,\gamma,\mu}\left(v\right)<e^{c_4 n}\right)\leqslant e^{-c_4 n}\right\}\,. \tag{3.25}$$

As is the case for $\mathcal{E}_2$, under the event that $\widetilde{X}=0$, we interpret $\mathbb{P}_{\widetilde{X}}(\hat{D}_{\alpha,\gamma,\mu}(v)<e^{c_4 n})=1$.

We now define our main quasirandomness event $\mathcal{E}$ to be the intersection of these events:

$$\mathcal{E}:=\mathcal{E}_1\cap\mathcal{E}_2\cap\mathcal{E}_3\cap\mathcal{E}_4\,. \tag{3.26}$$

The following lemma essentially allows us to assume that $\mathcal{E}$ holds in what follows.

Lemma 3.4.1. *For $B>0$, $\zeta\in\Gamma_B$, and all sufficiently small $\alpha,\gamma,\mu\in(0,1)$, there exist constants $c_2,c_3,c_4\in(0,1)$ appearing in (3.23), (3.24) and (3.25) so that*

$$\mathbb{P}_A(\mathcal{E}^c)\leqslant 2e^{-\Omega(n)}. \tag{3.27}$$

Remark 3.4.2 (Choice of constants, α,γ,μ). We take $\alpha,\gamma\in(0,1)$ to be sufficient small so that Lemma 3.4.1 holds. For μ we will choose it to be sufficiently small so that (1) Lemma 3.4.1 holds; (2) we have $\mu\in(0,2^{-15})$ and so that; (3) $\mu>0$ is small enough to guarantee that every set $I\subseteq[n]$ with $|I|\geqslant(1-2\mu)n$ satisfies

$$\|w\|_2\leqslant c_{\rho,\delta}^{-2}\|w_I\|_2, \tag{3.28}$$

for every $w\in\mathrm{Incomp}\,(\delta,\rho)$. This is possible by Fact 3.3.1. These constants α,γ,μ will appear throughout this chapter and will always be thought of as fixed according to this choice.

3.4.2 Statement of our master quasi-randomness theorem and the deduction of Lemma 3.4.1

We will deduce Lemma 3.4.1 from a "master quasi-randomness theorem" together with a handful of now-standard results in the area.

For the purposes of the following sections, we shall informally consider a vector as "structured" if

$$\hat{D}_{\alpha,\gamma,\mu}(v) \leqslant e^{c_\Sigma n}$$

where $c_\Sigma \in (0,1)$ is a small constant, to be chosen shortly. Thus it makes sense to define the set of "structured directions" on the sphere

$$\Sigma = \Sigma_{\alpha,\gamma,\mu} := \{v \in \mathbb{S}^{n-1} : \hat{D}_{\alpha,\gamma,\mu}(v) \leqslant e^{c_\Sigma n}\}\,. \tag{3.29}$$

We now introduce our essential quasi-randomness measure of a random matrix. For $\zeta \in \Gamma$, $A \sim \mathrm{Sym}_n(\zeta)$, and a given vector $w \in \mathbb{R}^n$, define

$$q_n(w) = q_n(w;\alpha,\gamma,\mu) := \mathbb{P}_A\left(\exists v \in \Sigma \text{ and } \exists s,t \in [-4\sqrt{n}, 4\sqrt{n}] :\ Av = sv + tw\right) \tag{3.30}$$

and set

$$q_n = q_n(\alpha,\gamma,\mu) := \sup_{w \in \mathbb{S}^{n-1}} q_n(w)\,. \tag{3.31}$$

We now state our "master quasi-randomness theorem", from which we deduce Lemma 3.4.1.

Theorem 3.4.3 (Master quasi-randomness theorem)**.** *For $B > 0$ and $\zeta \in \Gamma_B$, there exist constants $\alpha, \gamma, \mu, c_\Sigma, c \in (0,1)$ depending only on B so that*

$$q_n(\alpha,\gamma,\mu) \leqslant 2e^{-cn}\,.$$

The proof of Theorem 3.4.3 is quite similar to the proof of Theorem 2.1.1, albeit with a few technical adaptations, and a complete proof is available in the supplementary paper [36]. Note that $q_n(\alpha,\gamma,\mu)$ is monotone decreasing as α, γ and μ decrease. As such, Theorem 3.4.3 implies that its conclusion holds for all sufficiently small α, γ, μ as well.

We now prove that our pseudorandom event $\mathcal{E} = \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3 \cap \mathcal{E}_4$ holds with probability $1 - e^{-\Omega(n)}$.

Proof of Lemma 3.4.1. The event $\mathcal{E}_1$: From [61] we may deduce[3] the following concentration bound

$$\mathbb{P}(\|A\|_{op} \geqslant (3+t)\sqrt{n}) \lesssim e^{-ct^{3/2}n}, \tag{3.32}$$

which holds for all $t \geqslant 0$. Thus, by (3.32), the event $\mathcal{E}_1$ at (3.22) fails with probability $\lesssim e^{-\Omega(n)}$.

The event $\mathcal{E}_2$: By Lemma 3.3.2 there is a $c > 0$ so that for each $u \neq 0$ we have

$$\mathbb{P}_A(A^{-1}u/\|A^{-1}u\|_2 \in \mathrm{Comp}\,(\delta,\rho)) \leqslant e^{-cn}\,.$$

Applying Markov's inequality shows

$$\mathbb{P}_A\left(\mathbb{P}_{\widetilde{X}}\left(A^{-1}\widetilde{X}/\|A^{-1}\widetilde{X}\|_2 \in \mathrm{Comp}\,(\delta,\rho), \widetilde{X} \neq 0\right) > e^{-cn/2}\right) \leqslant e^{-cn/2}\,,$$

and so the event in (3.23) fails with probability at most $O\left(e^{-\Omega(n)}\right)$, under the event $\widetilde{X} \neq 0$. By Theorem 3.1.1 in [167] we have that

$$\mathbb{P}_{\widetilde{X}}(\widetilde{X} = 0) \leqslant e^{-\Omega(\mu n)}\,. \tag{3.33}$$

Choosing c_2 small enough shows an exponential bound on $\mathbb{P}(\mathcal{E}_2^c)$.

The event $\mathcal{E}_3$: If $D_{\alpha,\gamma}(u) \leqslant e^{c_3 n}$, for an u an eigenvector $Au = \lambda v$, we have that

$$\hat{D}_{\alpha,\gamma,\mu}(u) \leqslant D_{\alpha,\gamma}(u) \leqslant e^{c_3 n},$$

where the first inequality is immediate from the definition. Now note that if $\mathcal{E}_1$ holds then $\lambda \in [-4\sqrt{n}, 4\sqrt{n}]$ and so

$$\mathbb{P}(\mathcal{E}_3^c) \leqslant \mathbb{P}\big(\exists u \in \Sigma, \lambda \in [-4\sqrt{n}, 4\sqrt{n}] : Au = \lambda u\big) + \mathbb{P}(\mathcal{E}_1^c) \leqslant q_n(0) + e^{-\Omega(n)},$$

where the first inequality holds if we choose $c_3 \leqslant c_\Sigma$. We now apply Theorem 3.4.3 to see $q_n(0) \leqslant q_n \lesssim e^{-\Omega(n)}$, yielding the desired result.

The event $\mathcal{E}_4$: Note first that by (3.33), we may assume $\widetilde{X} \neq 0$. For a fixed instance of $\widetilde{X} \neq 0$, we have

$$\mathbb{P}_A\left(\hat{D}_{\alpha,\gamma,\mu}\left(A^{-1}\tilde{X}/\|\tilde{X}\|_2\right) < e^{c_4 n}\right) \leqslant \mathbb{P}_A\big(\exists v \in \Sigma : Av = \tilde{X}/\|\tilde{X}\|_2\big) \leqslant q_n\left(\tilde{X}/\|\tilde{X}\|_2\right), \tag{3.34}$$

which is at most $e^{-\Omega(n)}$, by Theorem 3.4.3. Here the first inequality holds when $c_4 \leqslant c_\Sigma$.

[3] Technically, the result of [61] is sharper and for random matrices whose entries are symmetric random variables. However (3.32) follows from [61] along with a "symmetrization trick". The details are written out in Appendix C of [34].

We now write $v = A^{-1}\tilde{X}/\|\tilde{X}\|_2$ and apply Markov's inequality

$$\mathbb{P}(\mathcal{E}_4^c) = \mathbb{P}_A\left(\mathbb{P}_{\tilde{X}}\left(\hat{D}_{\alpha,\gamma,\mu}(v) < e^{c_4 n}\right) \geqslant e^{-c_4 n}\right) \leqslant e^{c_4 n}\mathbb{E}_{\tilde{X}}\mathbb{P}_A(\hat{D}_{\alpha,\gamma,\mu}(v) < e^{c_4} n) = e^{-\Omega(n)},$$

where the last line follows when c_4 is taken small relative to the implicit constant in the bound on the right-hand-side of (3.34).

Since we have shown that each of $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_4$ holds with probability $1 - e^{-\Omega(n)}$, the intersection fails with exponentially small probability. □

3.5 Decoupling Quadratic Forms

In this section we will prove our Esseen-type inequality that will allow us to deal with a small ball event and a large deviation event simultaneously.

Lemma 3.5.1. *For $B > 0$, let $\zeta \in \Gamma_B$ and $X \sim \mathrm{Col}_n(\zeta)$. Let M be an $n \times n$ symmetric matrix, $u \in \mathbb{R}^n$, $t \in \mathbb{R}$ and $s, \delta \geqslant 0$. Then*

$$\mathbb{P}(|\langle MX, X\rangle - t| < \delta, \langle X, u\rangle \geqslant s) \lesssim \delta e^{-s} \int_{-1/\delta}^{1/\delta} \left|\mathbb{E}\, e^{2\pi i\theta\langle MX,X\rangle + \langle X,u\rangle}\right| d\theta\,. \tag{3.35}$$

We will then bound the integrand (our so-called "titled" characteristic function) with a *decoupling* maneuver, somewhat similar to a "van der Corput trick" in classical Fourier analysis. This amounts to a clever application of Cauchy-Schwarz inspired by Kwan and Sauermann's work on Costello's conjecture [97] (a similar technique appears in [19]). We shall then be able to mix in our quasi-random conditions on our matrix A to ultimately obtain Lemma 3.5.2, which gives us a rather tractable bound on the left-hand-side of (3.35). To state this lemma, let us recall that $\mathcal{E}$ (defined at (3.26)) is the set of symmetric matrices satisfying the quasi-randomness conditions in the previous section, Section 3.4. Also recall that the constant $\mu \in (0, 2^{-15})$ is defined in Section 3.4 so that Lemma 3.4.1 holds and is treated as fixed constant throughout this chapter.

Lemma 3.5.2. *For $B > 0$, let $\zeta \in \Gamma_B$, $X \sim \mathrm{Col}_n(\zeta)$ and let A be a real symmetric $n \times n$ matrix with $A \in \mathcal{E}$ and set $\mu_1 := \sigma_{\max}(A^{-1})$. Also let $s \geqslant 0, \delta > e^{-cn}$ and $u \in \mathbb{S}^{n-1}$. Then*

$$\mathbb{P}_X\left(\left|\langle A^{-1}X, X\rangle - t\right| \leqslant \delta\mu_1, \langle X, u\rangle \geqslant s\right) \lesssim \delta e^{-s} \int_{-1/\delta}^{1/\delta} I(\theta)^{1/2}\, d\theta + e^{-\Omega(n)}\,,$$

where

$$I(\theta) := \mathbb{E}_{J, X_J, X'_J} \exp\left(\langle (X + X')_J, u\rangle - c\theta^2\mu_1^{-2}\|A^{-1}(X - X')_J\|_2^2\right),$$

$X' \sim \mathrm{Col}_n(\zeta)$ is independent of X, and $J \subseteq [n]$ is a μ-random set. Here $c > 0$ is a constant depending only on B.

While the definition of $I(\theta)$ (and therefore the conclusion of the lemma) is a bit mysterious at this point, we assure the reader that this is a step in right direction.

All works bounding the singularity probability for random symmetric matrices contain a related decoupling step [35, 37, 62, 87, 113, 166], starting with Costello, Tao and Vu's breakthrough [41] building off of Costello's earlier work [40] on anticoncentration of bilinear and quadratic forms. A subtle difference in the decoupling approach from [97] used here is that the quadratic form is decoupled *after* bounding a small ball probability in terms of the integral of a characteristic function rather than on the probability itself; the effect of this approach is that we do not lose a power of δ, but only lose by a square root "under the integral" on the integrand $I(\theta)$.

3.5.1 Proofs

We now dive in and prove our Esseen-type inequality. For this we shall appeal to the classical Esseen inequality [59]: if Z is a random variable taking values in $\mathbb{R}$ with characteristic function $\varphi_Z(\theta) := \mathbb{E}_Z\, e^{2\pi i\theta Z}$, then for all $t \in \mathbb{R}$ we have

$$\mathbb{P}_X(|Z - t| \leqslant \delta) \lesssim \delta \int_{-1/\delta}^{1/\delta} |\varphi_Z(\theta)|\, d\theta.$$

We shall also use the following basic fact about subgaussian random vectors (see, for example, [167, Prop. 2.6.1]): If $\zeta \in \Gamma_B$ and $Y \sim \mathrm{Col}_n(\zeta)$ then for every vector $u \in \mathbb{R}^n$ we have

$$\mathbb{E}_Y e^{\langle Y,u\rangle} \leqslant \exp(2B^2\|u\|_2^2)\,. \tag{3.36}$$

Proof of Lemma 3.5.1. Since $\mathbf{1}\{x \geqslant s\} \leqslant e^{x-s}$, we may bound

$$\mathbb{P}_X(|\langle MX, X\rangle - t| < \delta, \langle X, u\rangle \geqslant s) \leqslant e^{-s}\mathbb{E}\left[\mathbf{1}\{|\langle MX, X\rangle - t| < \delta\}e^{\langle X,u\rangle}\right]. \tag{3.37}$$

Define the random variable $Y \in \mathbb{R}^n$ by

$$\mathbb{P}(Y \in U) = (\mathbb{E}\, e^{\langle X,u\rangle})^{-1}\mathbb{E}[\mathbf{1}_U e^{\langle X,u\rangle}], \tag{3.38}$$

for all open $U \subseteq \mathbb{R}^n$. Note that the expectation $\mathbb{E}e^{\langle X,u\rangle}$ is finite by (3.36). We now use this definition to rewrite the expectation on the right-hand-side of (3.37),

$$\mathbb{E}_X\left[\mathbf{1}\{|\langle MX, X\rangle - t| < \delta\}e^{\langle X,u\rangle}\right] = \left(\mathbb{E}\, e^{\langle X,u\rangle}\right)\mathbb{P}_Y(|\langle MY, Y\rangle - t| \leqslant \delta)\,.$$

Thus, we may apply Esseen's Lemma to the random variable Y to obtain

$$\mathbb{P}_Y(|\langle MY, Y\rangle - t| \leqslant \delta) \lesssim \delta \int_{-1/\delta}^{1/\delta} |\mathbb{E}_Y\, e^{2\pi i\theta\langle MY,Y\rangle}|\, d\theta\,.$$

By the definition of Y we have

$$\mathbb{E}_Y\, e^{2\pi i\theta\langle MY,Y\rangle} = \left(\mathbb{E}_X\, e^{\langle X,u\rangle}\right)^{-1} \mathbb{E}\, e^{2\pi i\theta\langle MX,X\rangle + \langle X,u\rangle},$$

completing the lemma. □

To control the integral on the right-hand-side of Lemma 3.5.1, we will appeal to the following decoupling lemma, which is adapted from Lemma 3.3 from [97].

Lemma 3.5.3 (Decoupling with an exponential tilt)**.** *Let $\zeta \in \Gamma$, let $X, X' \sim \mathrm{Col}_n(\zeta)$ be independent and let $J \cup I = [n]$ be a partition of $[n]$. Let M be a $n \times n$ symmetric matrix and let $u \in \mathbb{R}^n$. Then*

$$\left|\mathbb{E}_X\, e^{2\pi i\theta\langle MX,X\rangle + \langle X,u\rangle}\right|^2 \leqslant \mathbb{E}_{X_J,X'_J}\, e^{\langle (X+X')_J,u\rangle} \cdot \left|\mathbb{E}_{X_I} e^{4\pi i\theta\langle M(X-X')_J,X_I\rangle + 2\langle X_I,u\rangle}\right|.$$

Proof. After partitioning the coordinates of X according to J and writing $\mathbb{E}_X = \mathbb{E}_{X_I}\mathbb{E}_{X_J}$, we apply Jensen's inequality to obtain

$$E := \left|\mathbb{E}_X\, e^{2\pi i\theta\langle MX,X\rangle + \langle X,u\rangle}\right|^2 = \left|\mathbb{E}_{X_I}\mathbb{E}_{X_J}\, e^{2\pi i\theta\langle MX,X\rangle + \langle X,u\rangle}\right|^2 \leqslant \mathbb{E}_{X_I}\left|\mathbb{E}_{X_J} e^{2\pi i\theta\langle MX,X\rangle + \langle X,u\rangle}\right|^2.$$

We now expand the square $\left|\mathbb{E}_{X_J} e^{2\pi i\theta\langle MX,X\rangle + \langle X,u\rangle}\right|^2$ as

$$\begin{aligned}
&\mathbb{E}_{X_J,X'_J} e^{2\pi i\theta\langle M(X_I+X_J),(X_I+X_J)\rangle + \langle (X_I+X_J),u\rangle - 2\pi i\theta\langle M(X_I+X'_J),(X_I+X'_J)\rangle + \langle (X_I+X'_J),u\rangle} \\
&= \mathbb{E}_{X_J,X'_J} e^{4\pi i\theta\langle M(X_J-X'_J),X_I\rangle + \langle X_J+X'_J,u\rangle + 2\langle X_I,u\rangle + 2\pi i\langle MX_J,X_J\rangle - 2\pi i\langle MX'_J,X'_J\rangle},
\end{aligned}$$

where we used the fact that M is symmetric. Thus, swapping expectations yields

$$\begin{aligned}
E &\leqslant \mathbb{E}_{X_J,X'_J}\mathbb{E}_{X_I} e^{4\pi i\theta\langle M(X_J-X'_J),X_I\rangle + \langle X_J+X'_J,u\rangle + 2\langle X_I,u\rangle + 2\pi i\langle MX_J,X_J\rangle - 2\pi i\langle MX'_J,X'_J\rangle} \\
&\leqslant \mathbb{E}_{X_J,X'_J}\left|\mathbb{E}_{X_I} e^{4\pi i\theta\langle M(X_J-X'_J),X_I\rangle + \langle X_J+X'_J,u\rangle + 2\langle X_I,u\rangle + 2\pi i\langle MX_J,X_J\rangle - 2\pi i\langle MX'_J,X'_J\rangle}\right| \\
&= \mathbb{E}_{X_J,X'_J}\, e^{\langle X_J+X'_J,u\rangle}\left|\mathbb{E}_{X_I} e^{4\pi i\theta\langle M(X-X')_J,X_I\rangle + 2\langle X_I,u\rangle}\right|,
\end{aligned}$$

as desired. Here we could swap expectations since all expectations are finite, due to the sub-gaussian assumption on ζ. □

We need a basic bound that will be useful for bounding our tilted characteristic function. This bound appears in the proof of Theorem 6.3 in Vershynin's work [166].

Fact 3.5.4. *For $B > 0$, let $\zeta \in \Gamma_B$, let ζ' be an independent copy of ζ and set $\xi = \zeta - \zeta'$. Then for all $a \in \mathbb{R}^n$ we have*

$$\prod_j \mathbb{E}_\xi |\cos(2\pi\xi a_j)| \leqslant \exp\left(-c \min_{r\in[1,c^{-1}]} \|ra\|_{\mathbb{T}}^2\right),$$

where $c > 0$ depends only on B.

A simple symmetrization trick along with Cauchy-Schwarz will allow us to prove a similar bound for the tilted characteristic function.

Lemma 3.5.5. *For $B > 0$, let $\zeta \in \Gamma_B$, $X \sim \mathrm{Col}_n(\zeta)$ and let $u, v \in \mathbb{R}^n$. Then*

$$\left|\mathbb{E}_X e^{2\pi i\langle X,v\rangle + \langle X,u\rangle}\right| \leqslant \exp\left(-c \min_{r\in[1,c^{-1}]} \|rv\|_{\mathbb{T}}^2 + c^{-1}\|u\|_2^2\right), \tag{3.39}$$

where $c \in (0,1)$ depends only on B.

Proof. Let ζ' be an independent copy of ζ and note that

$$\left|\mathbb{E}_\zeta\, e^{2\pi i\zeta v_j + \zeta u_j}\right|^2 = \mathbb{E}_{\zeta,\zeta'}\, e^{2\pi i(\zeta-\zeta')v_j + (\zeta+\zeta')u_j} = \mathbb{E}_{\zeta,\zeta'}\left[e^{(\zeta+\zeta')u_j}\cos(2\pi(\zeta-\zeta')v_j)\right].$$

Let $\widetilde{X} = (\widetilde{X}_i)_{i=1}^n$, $\widetilde{Y} = (Y_i)_{i=1}^n$ denote vectors with i.i.d. coordinates distributed as $\xi := \zeta - \zeta'$ and $\zeta + \zeta'$, respectively. We have

$$\left|\mathbb{E}_X e^{2\pi i\langle X,v\rangle + \langle X,u\rangle}\right|^2 \leqslant \mathbb{E}\, e^{\langle \widetilde{Y},u\rangle} \prod_j \cos(2\pi\widetilde{X}_j v_j) \leqslant \left(\mathbb{E}_{\widetilde{Y}} e^{2\langle \widetilde{Y},u\rangle}\right)^{1/2} \left(\prod_j \mathbb{E}_\xi |\cos(2\pi\xi v_j)|\right)^{1/2}, \tag{3.40}$$

where we have applied the Cauchy-Schwarz inequality along with the bound $|\cos(x)|^2 \leqslant |\cos(x)|$ to obtain the last inequality. By (3.36), the first expectation on the right-hand-side of (3.40) is at most $\exp(O(\|u\|_2^2))$. Applying Fact 3.5.4 completes the Lemma. □

3.5.2 Quasi-random properties for triples (J, X_J, X'_J)

We now prepare for the proof of Lemma 3.5.2 by introducing a quasi-randomness notion on triples (J, X_J, X'_J). Here $J \subseteq [n]$ and $X, X' \in \mathbb{R}^n$. For this we fix a $n \times n$ real symmetric matrix $A \in \mathcal{E}$ and define the event $\mathcal{F} = \mathcal{F}(A)$ as the intersection of the events $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ and $\mathcal{F}_4$, which are defined as follows. Given a triple (J, X_J, X'_J), we write $\tilde{X} := X_J - X'_J$.

Define events $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3(A)$ by

$$\mathcal{F}_1 := \{|J| \in [\mu n/2, 2\mu n]\} \tag{3.41}$$

$$\mathcal{F}_2 := \{\|\widetilde{X}\|_2 n^{-1/2} \in [c, c^{-1}]\} \tag{3.42}$$

$$\mathcal{F}_3(A) := \{A^{-1}\widetilde{X}/\|A^{-1}\widetilde{X}\|_2 \in \mathrm{Incomp}\,(\delta, \rho)\}\,. \tag{3.43}$$

Finally, we write $v = v(\tilde{X}) := A^{-1}\widetilde{X}$ and $I := [n] \setminus J$ and then define $\mathcal{F}_4(A)$ by

$$\mathcal{F}_4(A) := \left\{ D_{\alpha,\gamma}\left(\frac{v_I}{\|v_I\|}\right) > e^{cn} \right\}. \tag{3.44}$$

We now define $\mathcal{F}(A) := \mathcal{F}_1 \cap \mathcal{F}_2 \cap \mathcal{F}_3(A) \cap \mathcal{F}_4(A)$ and prove the following basic lemma that will allow us to essentially assume that (3.41),(3.42),(3.43),(3.44) hold in all that follows. We recall that the constants $\delta, \rho, \mu, \alpha, \gamma$ were chosen in Lemma 3.3.2 and Lemma 3.4.1 as a function of the subgaussian moment B. Thus the only new parameter in $\mathcal{F}$ is the constant c in lines (3.42) and (3.44).

Lemma 3.5.6. *For $B > 0$, let $\zeta \in \Gamma_B$, let $X, X' \sim \mathrm{Col}_n(\zeta)$ be independent and let $J \subseteq [n]$ be a μ-random subset. Let A be a $n \times n$ real symmetric matrix with $A \in \mathcal{E}$. We may choose the constant $c \in (0,1)$ appearing in (3.42) and (3.44) as a function of B and μ so that*

$$\mathbb{P}_{J,X_J,X'_J}(\mathcal{F}^c) \lesssim e^{-cn}\,.$$

Proof. For $\mathcal{F}_1$, we use Hoeffding's inequality to see $\mathbb{P}(\mathcal{F}_1^c) \lesssim e^{-\Omega(n)}$. To bound $\mathbb{P}(\mathcal{F}_2^c)$, we note that the entries of $\widetilde{X}$ are independent, subgaussian, and have variance 2μ, and so $\widetilde{X}/(\sqrt{2\mu})$ has i.i.d. entries with mean zero, variance 1 and subgaussian moment bounded by $B/\sqrt{2\mu}$. Thus from Theorem 3.1.1 in [167] we have

$$\mathbb{P}(|\|\widetilde{X}\|_2 - \sqrt{2n\mu}| > t) < \exp(-c\mu t^2/B^4).$$

For $\mathcal{F}_3(A), \mathcal{F}_4(A)$, recall that $A \in \mathcal{E}$ means that (3.23) and (3.25) hold, thus exponential bounds on $\mathbb{P}(\mathcal{F}_3^c)$ and $\mathbb{P}(\mathcal{F}_4^c)$ follow from Markov's inequality. □

3.5.3 Proof of Lemma 3.5.2

We now prove Lemma 3.5.2 by applying the previous three lemmas in sequence.

Proof of Lemma 3.5.2. Let $\delta \geqslant e^{-c_1 n}$ where we will choose $c_1 > 0$ to be sufficiently small later in the proof. Apply Lemma 3.5.1 to write

$$\mathbb{P}_X\left(\left|\langle A^{-1}X, X\rangle - t\right| \leqslant \delta\mu_1, \langle X, u\rangle \geqslant s\right) \lesssim \delta e^{-s}\int_{-1/\delta}^{1/\delta}\left|\mathbb{E}_X\, e^{2\pi i\theta\frac{\langle A^{-1}X,X\rangle}{\mu_1}+\langle X,u\rangle}\right| d\theta\,, \tag{3.45}$$

where we recall that $\mu_1 = \sigma_{\max}(A^{-1})$. We now look to apply our decoupling lemma, Lemma 3.5.3. Let J be a μ-random subset of $[n]$, define $I := [n]\setminus J$ and let X' be an independent copy of X. By Lemma 3.5.3 we have

$$\left|\mathbb{E}_X\, e^{2\pi i\theta\frac{\langle A^{-1}X,X\rangle}{\mu_1}+\langle X,u\rangle}\right|^2 \leqslant \mathbb{E}_J\mathbb{E}_{X_J,X'_J}\, e^{\langle (X+X')_J,u\rangle}\cdot\left|\mathbb{E}_{X_I}\, e^{4\pi i\theta\left\langle\frac{A^{-1}\widetilde{X}}{\mu_1},X_I\right\rangle+2\langle X_I,u\rangle}\right|, \tag{3.46}$$

where we recall that $\widetilde{X} = (X - X')_J$.

We first consider the contribution to the expectation on the right-hand-side of (3.46) from triples $(J, X_J, X'_J) \notin \mathcal{F}$. For this let Y be a random vector such that $Y_j = X_j + X'_j$, if $j \in J$, and $Y_j = 2X_j$, if $j \in I$. Applying the triangle inequality, we have

$$\mathbb{E}^{\mathcal{F}^c}_{J,X_J,X'_J}\, e^{\langle (X+X')_J,u\rangle}\cdot\left|\mathbb{E}_{X_I}\, e^{4\pi i\theta\langle\frac{A^{-1}\widetilde{X}}{\mu_1},X_I\rangle+2\langle X_I,u\rangle}\right| \leqslant \mathbb{E}^{\mathcal{F}^c}_{J,X_J,X'_J}\, e^{\langle (X+X')_J,u\rangle}\mathbb{E}_{X_I}\, e^{2\langle X_I,u\rangle} = \mathbb{E}^{\mathcal{F}^c}_{J,X,X'}e^{\langle Y,u\rangle}.$$

By Cauchy-Schwarz, (3.36) and Lemma 3.5.6, we have

$$\mathbb{E}^{\mathcal{F}^c}_{J,X,X'}\, e^{\langle Y,u\rangle} \leqslant \mathbb{E}_{J,X,X'}\left[e^{\langle Y,2u\rangle}\right]^{1/2}\mathbb{P}_{J,X_J,X'_J}(\mathcal{F}^c)^{1/2} \lesssim e^{-\Omega(n)}\,. \tag{3.47}$$

We now consider the contribution to the expectation on the right-hand-side of (3.46) from triples $(J, X_J, X'_J) \in \mathcal{F}$. For this, let $w = w(X) := \frac{A^{-1}\widetilde{X}}{\mu_1}$ and assume $(J, X_J, X'_J) \in \mathcal{F}$. By Lemma 3.5.5, we have

$$\left|\mathbb{E}_{X_I}e^{4\pi i\theta\langle X_I,w\rangle+\langle X_I,2u\rangle}\right| \lesssim \exp\left(-c\min_{r\in[1,c^{-1}]}\|2r\theta w_I\|_{\mathbb{T}}^2\right). \tag{3.48}$$

Note that $\|w_I\|_2 \leqslant \|\widetilde{X}\|_2 \leqslant c^{-1}\sqrt{n}$, by the definition of $\mu_1 = \sigma_{\max}(A^{-1})$ and line (3.42) in the definition of $\mathcal{F}(A)$.

Now, from property (3.44) in that definition and by the hypothesis $\delta > e^{-c_1 n}$, we may choose $c_1 > 0$ small enough so that

$$D_{\alpha,\gamma}(w_I/\|w_I\|_2) \geqslant 2c^{-2}n^{1/2}/\delta \geqslant 2c^{-1}\|w_I\|_2/\delta.$$

By the definition of the least common denominator, for $|\theta| \leqslant 1/\delta$ we have

$$\min_{r\in[1,c^{-1}]}\|2r\theta w_I\|_{\mathbb{T}} = \min_{r\in[1,c^{-1}]}\left\|2r\theta\|w_I\|_2\cdot\frac{w_I}{\|w_I\|_2}\right\|_{\mathbb{T}} \geqslant \min\left\{\gamma\theta\|w_I\|_2, \sqrt{\alpha|I|}\right\}. \tag{3.49}$$

So for $|\theta| \leqslant 1/\delta$ we use (3.49) in (3.48) to bound the right-hand-side of (3.46) as

$$\mathbb{E}^{\mathcal{F}}_{J,X_J,X'_J} e^{\langle (X+X')_J,u\rangle} \cdot \left| \mathbb{E}_{X_I} e^{4\pi i\theta\langle w,X_I\rangle + 2\langle X_I,u\rangle} \right| \lesssim \mathbb{E}^{\mathcal{F}}_{J,X_J,X'_J} e^{\langle (X+X')_J,u\rangle} e^{-c\min\{\gamma^2\theta^2\|w_I\|_2^2,\alpha|I|\}}. \tag{3.50}$$

We now use that $(J, X_J, X'_J) \in \mathcal{F}$ to see that $w \in \mathrm{Incomp}\,(\delta,\rho)$ and that we chose μ to be sufficiently small, compared to ρ, δ, to guarantee that

$$\|w\|_2 \leqslant C\|w_I\|_2,$$

for some $C > 0$ (see (3.28)). Thus the right-hand-side of (3.50) is

$$\lesssim \mathbb{E}^{\mathcal{F}}_{J,X_J,X'_J} e^{\langle (X+X')_J,u\rangle} e^{-c'\theta^2\|w\|_2^2} + e^{-\Omega(n)}.$$

Combining this with (3.50), (3.46) obtains the desired bound in the case in the case $(J, X_J, X'_J) \in \mathcal{F}$. Combining this with (3.47) completes the proof of Lemma 3.5.2.

□

3.6 Preparation for the "Base step" of the iteration

As we mentioned at (3.7), Vershynin [166], gave a natural way of bounding the least singular value of a random symmetric matrix:

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon/n^{1/2}) \lesssim \sup_{r\in\mathbb{R}} \mathbb{P}_{A_n,X}\left(|\langle A_n^{-1}X, X\rangle - r| \leqslant \varepsilon\|A_n^{-1}X\|_2\right),$$

where we recall that A_n is obtained from A_{n+1} by deleting its first row and column. The main goal of this section is to prove the following lemma which tells us that we may intersect with the event $\sigma_{\min}(A_n) \geqslant \varepsilon/n^{1/2}$ in the probability on the right-hand-side at a loss of only $C\varepsilon$. This will be crucial for the base step in our iteration, since the bound we obtain on $\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon/n^{1/2})$ deteriorates as $\sigma_{\min}(A_n)$ decreases.

Lemma 3.6.1. *For $B > 0$, $\zeta \in \Gamma_B$, let $A_{n+1} \sim \mathrm{Sym}_{n+1}(\zeta)$ and let $X \sim \mathrm{Col}_n(\zeta)$. Then for all $\varepsilon > 0$,*

$$\mathbb{P}\left(\sigma_{\min}(A_{n+1}) \leqslant \frac{\varepsilon}{\sqrt{n}}\right) \lesssim \varepsilon + \sup_{r\in\mathbb{R}} \mathbb{P}\left(\frac{|\langle A_n^{-1}X, X\rangle - r|}{\|A_n^{-1}X\|_2} \leqslant C\varepsilon, \sigma_{\min}(A_n) \geqslant \frac{\varepsilon}{\sqrt{n}}\right) + e^{-\Omega(n)},$$

where $C > 0$ depends only on B.

We deduce this lemma from a geometric form of the lemma. For this, we let X_j denote the jth column of A_{n+1}, let H_j be the linear span of $X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_{n+1}$, and let $d_j(A_{n+1}) := \mathrm{dist}(X_j, H_j)$.

Lemma 3.6.2. *For $B > 0$, $\zeta \in \Gamma_B$, let $A_{n+1} \sim \mathrm{Sym}_{n+1}(\zeta)$. Then for all $\varepsilon > 0$,*

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon/\sqrt{n}) \lesssim \varepsilon + \mathbb{P}\left(d_1(A_{n+1}) \leqslant C\varepsilon \text{ and } \sigma_{\min}(A_n) \geqslant \varepsilon/\sqrt{n}\right) + e^{-\Omega(n)},$$

where $C > 0$ depends only on B.

3.6.1 Preparations

We require an elementary, but extremely useful, fact from linear algebra. This fact is actually a key step in the work of Nguyen, Tao and Vu on eigenvalue repulsion in random matrices (see [116, Section 4]) and we reproduce their short proof here for completeness. If M is a $n \times n$ matrix and $j \in [n]$, let $M^{(j)}$ denote the jth principle minor of M, i.e. M with the jth row and column removed.

Fact 3.6.3. *Let M be a $n \times n$ real symmetric matrix and let λ be an eigenvalue of M with corresponding unit eigenvector u. Let $j \in [n]$ and let λ' be an eigenvector of the minor $M^{(j)}$ with corresponding unit eigenvector v. Then*

$$|\langle v, X^{(j)}\rangle| \leqslant |\lambda - \lambda'|/|u_j|,$$

where $X^{(j)}$ is the jth column of M with the jth entry removed.

Proof. Without loss of generality, take $j = n$ and express $u = (w, u_n)$ where $w \in \mathbb{R}^{n-1}$. Then we have $(M^{(n)} - \lambda I)w + X^{(n)}u_n = 0$. Multiplying on the left by v^T yields

$$|u_n\langle v, X^{(n)}\rangle| = |\lambda - \lambda'||\langle v, w\rangle| \leqslant |\lambda - \lambda'|.$$

□

We will apply Fact 3.6.3 to see that when both $\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2}$ and $\sigma_{\min}(A^{(j)}) \leqslant \varepsilon n^{-1/2}$ hold we have $|\langle v, X^{(j)}\rangle| \lesssim \varepsilon$, assuming that $|u_j| \approx n^{-1/2}$. We then show that this latter event holds, subject to appropriate pseudo-random conditions, with probably $O(\varepsilon)$. The only wrinkle in this line of thinking is that we cannot rule out the possibility that for a *given* j we have $|u_j| \ll n^{-1/2}$. We *can*, however, rule out the possibility that many such $|u_j|$ are small, which will be enough for us. For this, we use a theorem of Rudelson and Vershynin [134] which we state here in a specialized form.

Theorem 3.6.4 (Theorem 1.5 of [134])**.** *For $B > 0$, $\zeta \in \Gamma_B$, let $A \sim \mathrm{Sym}_n(\zeta)$ and let v denote the unit eigenvector of A corresponding to the least singular value of A. Then there exists $c_2 > 0$ such that for all sufficiently small $c_1 > 0$ we have*

$$\mathbb{P}\left(|\{j : |v_j| \leqslant (c_2c_1)^6 n^{-1/2}\}| \geqslant c_1 n\right) \leqslant e^{-c_1 n},$$

for n sufficiently large.

To understand the event that $|\langle v, X^{(j)}\rangle| \lesssim \varepsilon$ (mentioned above), we need the Littlewood-Offord theorem of Rudelson and Vershynin [129], which we state here in a specialized form. Recall that $D_{\alpha,\gamma}(v)$ is the least common denominator of the vector v, as defined at (3.4).

Theorem 3.6.5. *For $n \in \mathbb{N}$, $B > 0$, $\gamma, \alpha \in (0,1)$ and $\varepsilon > 0$, let $v \in \mathbb{S}^{n-1}$ satisfy $D_{\alpha,\gamma}(v) > c\varepsilon^{-1}$ and let $X \sim \mathrm{Col}_n(\zeta)$, where $\zeta \in \Gamma_B$. Then*

$$\mathbb{P}(|\langle X, v\rangle| \leqslant \varepsilon) \lesssim \varepsilon + e^{-c\alpha n}.$$

Here $c > 0$ depends only on B.

The final ingredient in the proof of Theorem 3.6.1 is the observation that the event

$$\{\sigma_{\min}(A_{n+1}^{(1)}) \leqslant \varepsilon n^{-1/2}\} \cap \{\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2}\}$$

(as in Lemma 3.6.1) is roughly equivalent to the event

$$\{\text{there exist } \geqslant cn \text{ values of } j \text{ for which } \sigma_{\min}(A_{n+1}^{(j)}) \leqslant \varepsilon n^{-1/2}\} \cap \{\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2}\}.$$

Before we make this rigorous we prove this latter event has probability $\leqslant \varepsilon + e^{-\Omega(n)}$.

Lemma 3.6.6. *For $B > 0$, $\zeta \in \Gamma_B$, let $A_{n+1} \sim \mathrm{Sym}_{n+1}(\zeta)$. Then, for $\varepsilon > 0$, we have*

$$\mathbb{P}\left(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2} \text{ and } |\{j : \sigma_{\min}(A_{n+1}^{(j)}) \leqslant \varepsilon n^{-1/2}\}| \geqslant cn\right) \lesssim \varepsilon + e^{-\Omega(n)}, \tag{3.51}$$

where $c > 0$ depends only on B.

Proof. Let $\mathcal{A}_1$ denote the event on left-hand-side of (3.51). Let v be a unit eigenvector corresponding to the least singular value of A_{n+1}. We first show that if $\mathcal{A}_1$ holds then with probability $1 - e^{-\Omega(n)}$, we can find $\geqslant cn/2$ values of $j \in [n]$ so that $\sigma_{\min}(A_{n+1}^{(j)}) < \varepsilon n^{-1/2}$ and $|v_j| \gtrsim n^{-1/2}$. With this in mind we let

$$S_1 := \{j : \sigma_n(A_{n+1}^{(j)}) \leqslant \varepsilon n^{-1/2}\} \text{ and } S_2 := \{j : |v_j| \leqslant (cc_2/2)^6 n^{-1/2}\}$$

where c_2 is the constant from Theorem 3.6.4. We let $\mathcal{A}_2$ denote the event that $|S_2| < cn/2$ and apply Theorem 3.6.4 with $c_1 = c/2$ to see that $\mathbb{P}(\mathcal{A}_2^c) \leqslant e^{-cn/2}$. Now set $S := S_1 \cap ([n] \setminus S_2)$ and note that if $\mathcal{A}_1 \cap \mathcal{A}_2$ holds then $|S| \geqslant cn/2$.

Now, for $j \in [n]$, let $w_j = w(A_{n+1}^{(j)})$ denote a unit eigenvector of $A_{n+1}^{(j)}$ corresponding to the least singular value of $A_{n+1}^{(j)}$. Note that if $j \in S_1 \cap ([n] \setminus S_2)$ then, by Fact 3.6.3, we have

$$|\langle w_j, X^{(j)}\rangle| \leqslant 2\varepsilon/(c_2 c/2)^6 =: C\varepsilon. \tag{3.52}$$

Now let $\mathcal{Q}_j$ be the event that w_j satisfies $D_{\alpha,\gamma}(w_j) \geqslant e^{c_3 n}$ where α, γ, c_3 are chosen according to Lemma 3.4.1 and set $\mathcal{Q} = \cap_j \mathcal{Q}_j$. By Lemma 3.4.1 we have $\mathbb{P}(\mathcal{Q}^c) \lesssim e^{-\Omega(n)}$.

Putting this all together, we define the random variable

$$R := n^{-1} \sum_{j=1}^{n} \mathbb{1}\big(|\langle w_j, X^{(j)} \rangle| \leqslant C\varepsilon \text{ and } \mathcal{Q}_j\big),$$

and then observe that

$$\mathbb{P}(\mathcal{A}_1) \leqslant \mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{Q}) + e^{-\Omega(n)} \leqslant \mathbb{P}(R \geqslant c/2) + e^{-\Omega(n)}.$$

We now apply Markov and expand the definition of R to bound

$$\mathbb{P}(R \geqslant c/2) \lesssim n^{-1} \sum_{i=1}^{n} \mathbb{E}_{A_{n+1}^{(j)}} \mathbb{P}_{X^{(j)}} \left(|\langle w_j, X^{(j)} \rangle| \leqslant C\varepsilon \cap \mathcal{Q}_j\right) \lesssim \varepsilon$$

where the last inequality follows from the fact that $X^{(j)}$ is independent of the event Q_j and w_j and therefore we may put the property $\mathcal{Q}_j$ to use by applying Theorem 3.6.5. □

To prove Lemma 3.6.2, we will also use a basic fact which is at the heart of the geometric approach of Rudelson and Vershynin (see, e.g., [129, Lemma 3.5]).

Fact 3.6.7. *Let M be an $n \times n$ matrix and v be a unit vector satisfying $\|Mv\|_2 = \sigma_{\min}(M)$. Then*

$$\sigma_{\min}(M) \geqslant |v_j| \cdot d_j(M) \quad \text{for each } j \in [n].$$

Proof. Let X_j denote the jth column of M and let H_j denote the span of the remaining columns. Then

$$\sigma_{\min}(M) = \|Mv\|_2 \geqslant \operatorname{dist}(Mv, H_j) = \operatorname{dist}(v_j X_j, H_j) = |v_j| d_j(M).$$

□

3.6.2 Proofs of Lemma 3.6.2 and Lemma 3.6.1

With these preliminaries in-hand, we are now in a position to prove Lemma 3.6.2.

Proof of Lemma 3.6.2. We look to bound the quantity

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2}).$$

Let v denote a unit eigenvector corresponding to the least singular value of A_{n+1}. Let $\mathcal{A}$ denote the event that $v \in \operatorname{Incomp}(\delta, \rho)$: at least $c_{\rho,\delta} n$ coordinates of v have absolute value at least

$c_{\rho,\delta}n^{-1/2}$. By Lemma 3.3.2, $\mathbb{P}(\mathcal{A}^c) \lesssim e^{-\Omega(n)}$ and so

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2}) \leqslant \mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2} \text{ and } \mathcal{A}) + e^{-\Omega(n)}.$$

Now let $c > 0$ denote the constant from Lemma 3.6.6 and let $\mathcal{B}$ denote the event that at most cn principal minors of A_{n+1} satisfy $\sigma_{\min}(A_{n+1}^{(j)}) \leqslant \varepsilon n^{-1/2}$. Also note we may assume $c \leqslant c_{\rho,\delta}/2$. By Lemma 3.6.6 we have

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2} \text{ and } \mathcal{B}^c) \lesssim \varepsilon + e^{-\Omega(n)}$$

and so

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2}) \leqslant \mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2} \text{ and } \mathcal{A} \cap \mathcal{B}) + C\varepsilon + e^{-\Omega(n)}.$$

Now let

$$S := \{j : d_j(A_{n+1}) \leqslant \varepsilon/c_{\rho,\delta} \text{ and } \sigma_{\min}(A_{n+1}^{(j)}) \geqslant \varepsilon n^{-1/2}\}.$$

Observe that if $\sigma_{\min}(A_{n+1}) \leqslant \varepsilon/\sqrt{n}$ and $j \in [n]$ is such that $|v_j| \geqslant c_{\rho,\delta}n^{-1/2}$, then $d_j(A_{n+1}) \leqslant \varepsilon/c_{\rho,\delta}$, by Fact 3.6.7. Thus if $\sigma_{\min}(A_{n+1}) \leqslant \varepsilon/\sqrt{n}$ and $\mathcal{A}$ hold, then then there are $\geqslant c_{\rho,\delta}n$ values of j for which $d_j(A_{n+1}) < \varepsilon/c_{\rho,\delta}$. If $\mathcal{B}$ holds in addition to $\sigma_{n+1}(A_{n+1}) \leqslant \varepsilon/\sqrt{n}$ and $\mathcal{A}$, then at most $c_{\rho,\delta}n/2$ of these values of j have $\sigma_{\min}(A_{n+1}^{(j)}) < \varepsilon n^{-1/2}$. In other words,

$$\mathcal{A} \cap \mathcal{B} \cap \{\sigma_{n+1}(A_{n+1}) \leqslant \varepsilon/\sqrt{n}\} \subseteq \{|S| \geqslant c_{\rho,\delta}n/2\}. \tag{3.53}$$

Using (3.53) along with Markov's inequality tells us that

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon/\sqrt{n} \text{ and } \mathcal{A} \cap \mathcal{B}) \leqslant \mathbb{P}(|S| \geqslant c_{\rho,\delta}n/2) \leqslant \frac{2}{c_{\rho,\delta}n}\mathbb{E}|S|. \tag{3.54}$$

If we write

$$|S| = \sum_j \mathbb{1}(d_j(A_{n+1}) \leqslant \varepsilon/c_{\rho,\delta}, \sigma_{\min}(A_{n+1}^{(j)}) \geqslant \varepsilon n^{-1/2}),$$

then we see that

$$\mathbb{E}|S| = \mathbb{P}\left(d_1(A_{n+1}) \leqslant \varepsilon/c_{\rho,\delta},\ \sigma_{\min}(A_{n+1}^{(1)}) \geqslant \varepsilon/\sqrt{n}\right), \tag{3.55}$$

by symmetry. Putting (3.53),(3.54),(3.55) together gives us our desired conclusion. □

Lemma 3.6.1 now follows.

Proof of Lemma 3.6.1. If we set $a_{1,1}$ to be the first entry of $A = A_{n+1}$ then, by [166, Prop. 5.1], we have that

$$d_1(A_{n+1}) = \frac{|\langle A^{-1}X, X\rangle - a_{1,1}|}{\sqrt{1 + \|A^{-1}X\|_2^2}}.$$

Additionally, by [166, Prop. 8.2], we have $\|A^{-1}X\|_2 > 1/15$ with probability at least $1 - e^{-\Omega(n)}$. Replacing $a_{1,1}$ with r and taking a supremum completes the proof of Lemma 3.6.1. □

3.7 Eigenvalue crowding (and the proofs of Theorem 3.1.2 and Theorem 3.1.3)

The main purpose of this section is to prove the following theorem which gives an upper-bound on the probability that $k \geqslant 2$ eigenvalues of a random matrix fall in an interval of length ε. This will be key in our work on the "bulk" of the spectrum of A^{-1} in Section 3.8. This result is of independent interest as the $\varepsilon = 0$ case of this theorem tells us that the probability that a random symmetric matrix has *simple* spectrum (that is, has no repeated eigenvalue) is $1 - e^{-\Omega(n)}$, which is sharp and confirms a conjecture of Nguyen, Tao and Vu [116].

Given an $n \times n$ real symmetric matrix M, we let $\lambda_1(M) \geqslant \ldots \geqslant \lambda_n(M)$ denote its eigenvalues.

Theorem 3.7.1. *For $B > 0$, $\zeta \in \Gamma_B$, let $A_{n+1} \sim \mathrm{Sym}_{n+1}(\zeta)$. Then for each $j \leqslant cn$ and all $\varepsilon \geqslant 0$ we have*

$$\max_{k \leqslant n-j} \mathbb{P}(|\lambda_{k+j}(A_n) - \lambda_k(A_n)| \leqslant \varepsilon/\sqrt{n}) \leqslant (C\varepsilon)^j + 2e^{-cn},$$

where $C, c > 0$ are constants depending on B.

We suspect that the bound in Lemma 3.1.3 is actually *far* from the truth, for $\varepsilon > e^{-cn}$ and $j \geqslant 1$. In fact, one expects *quadratic* dependence on j in the exponent of ε. This type of dependence was recently confirmed by Nguyen [115] for $\varepsilon > e^{-n^c}$. As we shall also need Nguyen's result, we discuss it further in Section 3.8.

For the proof of Lemma 3.1.3, we remind the reader that if $u \in \mathbb{R}^n \cap \mathrm{Incomp}\,(\rho, \delta)$ then at least $c_{\rho,\delta} n$ coordinates of u have absolute value at least $c_{\rho,\delta} n^{-1/2}$.

In what follows, for a $n \times n$ symmetric matrix A, we use the notation $A^{(i_1,\ldots,i_r)}$ to refer to the minor of A for which the rows and columns indexed by $i_1, \ldots, i_r$ have been deleted. We also use the notation $A_{S \times T}$ to refer to the $|S| \times |T|$ submatrix of A defined by $(A_{i,j})_{i \in S, j \in T}$.

The following fact contains the key linear algebra required for the proof of Theorem 3.1.3.

Fact 3.7.2. *For $1 \leqslant k + j < n$, let A be a $n \times n$ symmetric matrix for which*

$$|\lambda_{k+j}(A) - \lambda_k(A)| \leqslant \varepsilon n^{-1/2}.$$

Let $(i_1, \ldots, i_k) \in [n]^k$ be such that $i_1, \ldots, i_k$ are distinct. Then there exist unit vectors $w^{(1)}, \ldots, w^{(k)}$ for which

$$\langle w^{(r)}, X_r \rangle \leqslant (\varepsilon n^{-1/2}) \cdot (1/|w^{(r-1)}_{i_r}|),$$

where $X_r \in \mathbb{R}^{n-r}$ is the i_rth column of A with coordinates indexed by $i_1, \dots, i_r$ removed. That is, $X_r := A_{[n]\setminus\{i_1,\dots,i_r\}\times\{i_r\}}$ and $w^{(r)}$ is a unit eigenvector corresponding to $\lambda_k(A^{(i_1,\dots,i_r)})$.

Proof. For $(i_1, \dots, i_j) \in [n]^j$, define the matrices $M_0, M_1, \dots, M_j$ by setting $M_r = A^{(i_1,\dots,i_r)}$ for $r = 1, \dots, j$ and then $M_0 := A$. Now if

$$|\lambda_{k+j}(A) - \lambda_k(A)| \leqslant \varepsilon n^{-1/2},$$

then Cauchy's interlacing theorem implies

$$|\lambda_k(M_r) - \lambda_k(M_{r-1})| \leqslant \varepsilon n^{-1/2},$$

for all $r = 1, \dots, j$. So let $w^{(r)}$ denote a unit eigenvector of M_r corresponding to eigenvalue $\lambda_k(M_r)$. Thus, by Fact 3.6.3, we see that

$$|\langle w^{(r)}, X_r\rangle| \leqslant (\varepsilon n^{-1/2}) \cdot (1/|w_{i_r}^{(r-1)}|),$$

for $r = 1, \dots, j$, where $X_r \in \mathbb{R}^{n-r}$ is the i_rth column of M_{r-1}, with the diagonal entry removed. In other words, $X_r \in \mathbb{R}^{n-r}$ is the i_rth column of A with coordinates indexed by $i_1, \dots, i_r$ removed. This completes the proof of Fact 3.7.2. □

Proof of Theorem 3.1.3. Note may assume that $\varepsilon > e^{-cn}$; the general case follows by taking c sufficiently small. Now, define $\mathcal{A}$ to be the event that *all* unit eigenvectors v of *all* $\binom{n}{j}$ of the minors $A_n^{(i_1,\dots,i_j)}$ lie in $\mathrm{Incomp}\,(\rho, \delta)$ and satisfy $D_{\alpha,\gamma}(v) > e^{c_3 n}$, where α, γ, c_3 are chosen according to Lemma 3.4.1. Note that by Lemma 3.4.1 and Lemma 3.3.2, we have

$$\mathbb{P}(\mathcal{A}^c) \leqslant \binom{n}{j+1} e^{-\Omega(n)} \leqslant n \left(\frac{en}{j}\right)^j e^{-\Omega(n)} \lesssim e^{-cn},$$

by taking c small enough, so that $j \log(en/j) < cn$ is smaller than the $\Omega(n)$ term.

With Fact 3.7.2 in mind, we define the event, $\mathcal{E}_{i_1,\dots,i_j}$, for each $(i_1, \dots, i_j) \in [n]^j$, to be the event that

$$|\langle w^{(r)}, X_r\rangle| \leqslant \varepsilon/c_{\rho,\delta} \quad \text{for all } r \in [j]\,,$$

where $X_r \in \mathbb{R}^{n-r}$ is the i_rth column of A with coordinates indexed by $i_1, \dots, i_r$ removed and $w^{(r)}$ is a unit eigenvector corresponding to $\lambda_k(A^{(i_1,\dots,i_r)})$.

If $\mathcal{A}$ holds then each $w^{(r)}$ has at least $c_{\rho,\delta} n$ coordinates with absolute value at least $c_{\rho,\delta} n^{-1/2}$. Thus, if additionally we have

$$|\lambda_{k+j}(A_n) - \lambda_k(A_n)| \leqslant \varepsilon n^{-1/2},$$

Fact 3.7.2 tells us that $\mathcal{E}_{i_1,\dots,i_j}$ occurs for at least $(c_{\rho,\delta} n/2)^j$ tuples $(i_1, \dots, i_j)$.

Define N to be the number of indices $(i_1, \dots, i_j)$ for which $\mathcal{E}_{i_1,\dots,i_j}$ occurs, and note

$$\mathbb{P}(|\lambda_{k+j}(A_n) - \lambda_k(A_n)| \leqslant \varepsilon/\sqrt{n}) \leqslant \mathbb{P}(N \geqslant (c_{\rho,\delta} n/2)^j \text{ and } \mathcal{A}) + O(e^{-cn}) \tag{3.56}$$

$$\leqslant \left(\frac{2}{c_{\rho,\delta}}\right)^j \mathbb{P}(\mathcal{E}_{1,\dots,j} \cap \mathcal{A}) + O(e^{-cn}) \tag{3.57}$$

where, for the second inequality, we applied Markov's inequality and used the symmetry of the events $\mathcal{E}_{i_1,\dots,i_j}$.

Thus we need only show that there exists $C > 0$ such that $\mathbb{P}(\mathcal{E}_{1,\dots,j} \cap \mathcal{A}) \leqslant (C\varepsilon)^j$. To use independence, we replace each of $w^{(r)}$ with the worst case vector, under $\mathcal{A}$

$$\mathbb{P}(\mathcal{E}_{1,\dots,j} \cap \mathcal{A}) \leqslant \max_{w_1,\dots,w_j : D_{\alpha,\gamma}(w_i) > e^{c_3 n}} \mathbb{P}_{X_1,\dots,X_r}\Big(|\langle w_r, X_r\rangle| \leqslant \varepsilon/c_{\rho,\delta} \quad \text{for all } r \in [j]\Big) \tag{3.58}$$

$$\leqslant \max_{w_1,\dots,w_j : D_{\alpha,\gamma}(w_i) > e^{c_3 n}} \prod_{r=1}^{j} \mathbb{P}_{X_r}\Big(|\langle w_r, X_r\rangle| \leqslant \varepsilon/c_{\rho,\delta}\Big) \leqslant (C\varepsilon)^j, \tag{3.59}$$

where the penultimate inequality follows from the independence of the X_r and the last inequality follows from the fact that $D_{\alpha,\gamma}(w_r) > e^{c_3 n} \gtrsim 1/\varepsilon$ (by choosing $c > 0$ small enough relative to c_3), and the Littlewood-Offord theorem of Rudelson and Vershynin, Lemma 3.6.5.

Putting (3.57) and (3.59) together completes the proof of Theorem 3.1.3. □

Of course, the proof of Theorem 3.1.2 follows immediately.

Proof of Theorem 3.1.2. Simply take $\varepsilon = 0$ in Theorem 3.1.3. □

3.8 Properties of the spectrum

In this section we describe and deduce Lemma 3.8.1 and Lemma 3.8.2, which are the tools we will use to control the "bulk " of the eigenvalues of A^{-1}. Here we understand "bulk" relative to the spectral measure of A^{-1}: our interest in an eigenvalue λ of A^{-1} is proportional to its contribution to $\|A^{-1}\|_{\mathrm{HS}}$. Thus the most delicate and important aspect of our analysis amounts to studying the *smallest* singular values of A.

For this we let $\sigma_n \leqslant \sigma_{n-1} \leqslant \cdots \leqslant \sigma_1$ be the singular values of A and let $\mu_1 \geqslant \ldots \geqslant \mu_n$ be the singular values of A^{-1}. Of course, we have $\mu_k = 1/\sigma_{n-k+1}$ for $1 \leqslant k \leqslant n$.

In short, these two lemmas, when taken together, tell us that

$$\sigma_{n-k+1} \approx k/\sqrt{n}, \tag{3.60}$$

for all $n \geqslant k \gg 1$ in some appropriate sense.

Lemma 3.8.1. *For $p > 1$, $B > 0$ and $\zeta \in \Gamma_B$, let $A \sim \mathrm{Sym}_n(\zeta)$. There is a constant C_p depending on B, p so that*

$$\mathbb{E}\left(\frac{\sqrt{n}}{\mu_k k}\right)^p \leqslant C_p\,,$$

for all k.

We shall deduce Lemma 3.8.1 from the "local semicircular law" of Erdős, Schlein and Yau [57], which gives us good control of the bulk of the spectrum at "scales" of size $\gg n^{-1/2}$. The next result is a type of "reverse" of Lemma 3.8.1 and will follow from a result of Nguyen [115] along with our Theorem 3.1.3.

Lemma 3.8.2. *For $p > 1$, $B > 0$ and $\zeta \in \Gamma_B$, let $A \sim \mathrm{Sym}_n(\zeta)$. There exist constants $C_p, c_p > 0$ depending on B, p so that for all $k \in [C_p, n]$ we have*

$$\mathbb{E}\left[\left(\frac{\mu_k k}{\sqrt{n}}\right)^p \mathbf{1}\{\mu_1 \leqslant e^{c_p n}\}\right] \leqslant C_p\,. \tag{3.61}$$

We point out that the condition $k \geqslant C_p$ is a very important assumption in Lemma 3.8.2 and the above statement with $k = 1$ and $p = 1$ would imply our main theorem. Thus one might think of Lemma 3.8.2 as a weaker relative of our main Theorem.

We also record a useful corollary of these two lemmas. For this, we define the function $\|\cdot\|_*$ for a $n \times n$ symmetric matrix M to be

$$\|M\|_*^2 = \sum_{k=1}^{n} \sigma_k(M)^2 (\log(1+k))^2. \tag{3.62}$$

The point of this definition is to give some measure to how the spectrum of A^{-1} is "distorted" from what it "should be", according to the heuristic at (3.60). Indeed if we have $\sigma_{n-k+1} = \Theta(k/\sqrt{n})$ for all k, say, then we have that

$$\|A^{-1}\|_* = \Theta(\mu_1).$$

Conversely, any deviation from this captures some macroscopic misbehavior on the part of the spectrum. In particular, the "weight function" $k \mapsto (\log(1+k))^2$ is designed to bias the smallest singular values, and thus we are primarily looking at this range for any poor behavior.

Corollary 3.8.3. *For $p > 1$, $B > 0$ and $\zeta \in \Gamma_B$, let $A \sim \mathrm{Sym}_n(\zeta)$. Then there exists constants $C_p, c_p > 0$ depending on B, p such that*

$$\mathbb{E}\left[\left(\frac{\|A^{-1}\|_*}{\mu_1}\right)^p \mathbf{1}\{\mu_1 \leqslant e^{c_p n}\}\right] \leqslant C_p\,.$$

In the remainder of this section we describe the results of Erdős, Schlein and Yau [57] and of Nguyen [115] and show how to use them to deduce Lemma 3.8.1 and Lemma 3.8.2 respectively. We then deduce Corollary 3.8.3.

3.8.1 The local semi-circular law and Lemma 3.8.1

For $a < b$ we define $N_A(a,b)$ to be the number of eigenvalues of A in the interval (a,b). One of the most fundamental results in the theory of random symmetric matrices is the *semi-circular law* which says that

$$\lim_{n\to\infty} \frac{N_A(a\sqrt{n}, b\sqrt{n})}{n} = \frac{1}{2\pi}\int_a^b (4-x^2)_+^{1/2}\,dx,$$

almost surely, where $A \sim \mathrm{Sym}_n(\zeta)$.

We use a powerful "local" version of the semi-circle law developed by Erdős, Schlein and Yau in a series of important papers [50, 51, 57]. Their results show that the spectrum of a random symmetric matrix actually adheres surprisingly closely to the semi-circular law. In this chapter, we need control on the number of eigenvalues in intervals of the form $[-t,t]$, where $1/n^{1/2} \ll t \ll n^{1/2}$. The semi-circular law predicts that

$$N_A(-t,t) \approx \frac{n}{2\pi}\int_{-t/n^{1/2}}^{t/n^{1/2}} (4-x^2)_+^{1/2}\,dx = \frac{2tn^{1/2}}{\pi}(1+o(1)).$$

Theorem 1.11 of [56] makes this prediction rigorous.

Theorem 3.8.4. *Let $B > 0$, $\zeta \in \Gamma_B$, and let $A \sim \mathrm{Sym}_n(\zeta)$. Then for all $t \in [Cn^{-1/2}, n^{1/2}]$ we have*

$$\mathbb{P}\left(\left|N_A(-t,t)/(n^{1/2}t) - 2\pi^{-1}\right| > \pi\right) \lesssim \exp\left(-c_1(t^2n)^{1/4}\right) \tag{3.63}$$

where $C, c_1 > 0$ are absolute constants.

Lemma 3.8.1 follows quickly from Theorem 3.8.4. In fact we shall only use the follow corollary.

Corollary 3.8.5. *Let $B > 0$, $\zeta \in \Gamma_B$, and let $A \sim \mathrm{Sym}_n(\zeta)$. Then for all $s \geqslant C$ and $k \in \mathbb{N}$ satisfying $sk \leqslant n$ we have*

$$\mathbb{P}\left(\frac{\sqrt{n}}{\mu_k k} \geqslant s\right) \lesssim \exp\left(-c(sk)^{1/2}\right),$$

where $C, c > 0$ are absolute constants.

Proof. Let C be the maximum of the constant C from Lemma 3.8.4 and π. If $\frac{\sqrt{n}}{\mu_k k} \geqslant s$ then $N_A(-skn^{-1/2}, skn^{-1/2}) \leqslant k$. We now apply Lemma 3.8.4 with $t = skn^{-1/2} \geqslant sn^{-1/2} \geqslant Cn^{-1/2}$ to see that this event occurs with probability $\lesssim e^{-c\sqrt{sk}}$. □

Proof of Lemma 3.8.1. Let C be the constant from Corollary 3.8.5. From bounds on the upper tail of $\|A\|_{op}$ (like (3.32)), we immediately see that for all $k \geqslant n/C$ we have

$$\mathbb{E}\left(\frac{\sqrt{n}}{\mu_k k}\right)^p \leqslant \mathbb{E}_A\left(\frac{\sigma_1(A)\sqrt{n}}{k}\right)^p = O_p((n/k)^p) = O_p(1).$$

Thus we can restrict our attention to the case when $k \leqslant n/C$. Define the events

$$E_1 = \left\{\frac{\sqrt{n}}{\mu_k k} \leqslant C\right\}, \quad E_2 = \left\{\frac{\sqrt{n}}{\mu_k k} \in [C, n/k]\right\}, \quad E_3 = \left\{\frac{\sqrt{n}}{\mu_k k} \geqslant \frac{n}{k}\right\}.$$

We may bound

$$\mathbb{E}\left(\frac{\sqrt{n}}{\mu_k k}\right)^p \leqslant C^p + \mathbb{E}\left(\frac{\sqrt{n}}{\mu_k k}\right)^p \mathbb{1}_{E_2} + \mathbb{E}\left(\frac{\sqrt{n}}{\mu_k k}\right)^p \mathbb{1}_{E_3}. \tag{3.64}$$

To deal with the second term in (3.64), we use Corollary 3.8.5 to see that

$$\mathbb{E}\left(\frac{\sqrt{n}}{\mu_k k}\right)^p \mathbb{1}_{E_2} \lesssim \int_C^{n/k} p s^{p-1} e^{-c\sqrt{sk}} ds = O_p(1).$$

To deal with the third term in (3.64), we note that since $n/k \geqslant C$ we may apply Corollary 3.8.5, with $s = n/k$, to conclude that $\mathbb{P}(E_3) \lesssim e^{-c\sqrt{n}}$. Thus, by Cauchy-Schwarz, we have

$$\mathbb{E}\left(\frac{\sqrt{n}}{\mu_k k}\right)^p \mathbb{1}_{E_3} \leqslant \left(\mathbb{E}\left(\frac{\sigma_1\sqrt{n}}{k}\right)^{2p}\right)^{1/2} \mathbb{P}(E_3)^{1/2} \leqslant O_p(1) \cdot n^p e^{-c\sqrt{n}} = O_p(1),$$

where we have used the upper tail estimate at (3.32) to see $\mathbb{E}\,\sigma_1^{2p} = O_p(n^p)$. □

3.8.2 Eigenvalue crowding and Lemma 3.8.2

In [115], Nguyen proved the following result which gives good estimates on the probability that kth smallest singular value is much smaller than typical, where $k \gg 1$. In fact he proved a more general result which bounds the probability that k eigenvalues fall into any interval of length ε. Here we need only the following less general result, which comes from Theorem 1.12 of [115].

Theorem 3.8.6. *For $B > 0$, $\zeta \in \Gamma_B$, let $A \sim \mathrm{Sym}_n(\zeta)$. Then for all $k \in [b_1^{-1}, b_1 n]$ and $\varepsilon > 0$ we have*

$$\mathbb{P}\left(\sigma_{n-k+1}(A) \leqslant \varepsilon n^{-1/2}\right) \leqslant (C\varepsilon/k)^{k^2/4} + O(e^{-n^{b_2}}),$$

where $C, b_1, b_2 > 0$ are absolute constants.

With Nguyen's result in hand, we quickly take care of the proof of Lemma 3.8.2.

Proof of Lemma 3.8.2. Let b_1, b_2 denote the constants b_1, b_2 from Theorem 3.8.6 and note that we may assume $k \leqslant b_1 n$ since for $k > b_1 n$ we may bound $\mu_k \leqslant \mu_{b_1 n}$. We now may assume

that C_p, the constant in the statement of Theorem 3.8.2, satisfies $C_p \geqslant \max\{b_1^{-1}, 3p\}$. Thus we may restrict our attention to k for which $k \geqslant \max\{b_1^{-1}, 3p\}$. We let $c_p > 0$ be a constant to be determined later. We set $R = e^{n^{b_2}}$ and integrate to see that

$$I := \mathbb{E}\left(\frac{\mu_k k}{\sqrt{n}}\right)^p \mathbf{1}\{\mu_1 \leqslant e^{c_p n}\}$$

is at most

$$\int_0^R k^p p s^{p-1} \mathbb{P}\left(\sigma_{n-k+1} \leqslant n^{-1/2}/s\right) ds + \int_R^{e^{c_p n}} p k^p s^{p-1} \mathbb{P}\left(\sigma_{n-k+1} \leqslant n^{-1/2}/s\right) ds =: I_0 + I_1.$$

Here we could truncate the integral I_1 at $e^{c_p n}$ since $\sigma_{n-k+1} \geqslant \sigma_n = 1/\mu_n \geqslant e^{-c_p n}$.

We bound these two ranges by applying different results. To bound I_0 we use Theorem 3.8.6 with $\varepsilon = 1/s$ to see

$$\mathbb{P}\left(\sigma_{n-k+1} \leqslant n^{-1/2}/s\right) \leqslant (C/ks)^{2p} + O(e^{-n^{c_2}}),$$

since $k \geqslant 2\sqrt{p}$. Thus integrating gives $I_0 = O_p(1)$.

To bound I_1, we use our Theorem 3.1.3 with $j = k \geqslant 3p$ and $\varepsilon = 1/s$ to see that

$$\mathbb{P}\left(\sigma_{n-k+1} \leqslant n^{-1/2}/s\right) \leqslant \sup_i \mathbb{P}\left(|\lambda_{i+j} - \lambda_i| \leqslant 2n^{-1/2}/s\right) \lesssim (C/s)^{3p} + \exp(-b_3 n)$$

for some $c_3 > 0$. Assuming that $c_p > 0$ is small enough relative to b_3 allows us to bound $I_1 = O_p(1)$. Thus we see $I = I_0 + I_1 = O_p(1)$, completing the proof of Lemma 3.8.2. □

3.8.3 Deduction of Corollary 3.8.3

We now conclude this section by deducing Corollary 3.8.3 from Lemma 3.8.1 and Lemma 3.8.2.

Proof of Corollary 3.8.3. Let c be the constant from Lemma 3.8.2 and $C = C_{2p}$ be the maximum of the two constants from Lemmas 3.8.1 and 3.8.2. Now define the event $\mathcal{E}_0 = \{\mu_1 \leqslant e^{c_p n}\}$ and express

$$\|A^{-1}\|_*^2 = \sum_{k=1}^n \mu_k^2 (\log(1+k))^2.$$

Note that we may omit the first C terms in this sum, as we can (trivially) bound $\mu_k^2 \leqslant \mu_1^2$. Further, by Hölder's inequality we may assume without loss of generality that $p \geqslant 2$. Applying the triangle inequality for the $L^{p/2}$ norm gives

$$\left[\mathbb{E}^{\mathcal{E}_0}\left(\sum_{k>C} \frac{\mu_k^2 (\log 2(1+k))^2}{\mu_1^2}\right)^{p/2}\right]^{2/p} \leqslant \sum_{k>C} (\log(1+k))^2 \mathbb{E}^{\mathcal{E}_0}\left[\frac{\mu_k^p}{\mu_1^p}\right]^{2/p}$$

which is

$$\sum_{k>C} \frac{(\log(1+k))^2}{k^2} \left(\mathbb{E}^{\mathcal{E}_0} \left(\frac{\mu_k k}{\sqrt{n}} \right)^p \left(\frac{\sqrt{n}}{\mu_1} \right)^p \right)^{2/p} \leqslant \sum_{k>C} \frac{(\log(1+k))^2}{k^2} \left(\mathbb{E}^{\mathcal{E}_0} \left(\frac{\mu_k k}{\sqrt{n}} \right)^{2p} \right)^{1/p} \left(\mathbb{E} \left(\frac{\sqrt{n}}{\mu_1} \right)^{2p} \right)^{1/p},$$

by Cauchy-Schwarz. Thus Lemmas 3.8.1 and 3.8.2 tell us that this is $O_p(1)$, completing the proof of Corollary 3.8.3. □

3.9 Controlling small balls and large deviations

The goal of this section is to prove the following lemma, which will be a main ingredient in our iteration in Section 3.10. We shall then use it again in the final step and proof of Theorem 3.1.1, in Section 3.11.

Lemma 3.9.1. *For $B > 0$ and $\zeta \in \Gamma_B$, let $A = A_n \sim \mathrm{Sym}_n(\zeta)$ and let $X \sim \mathrm{Col}_n(\zeta)$. Let $u \in \mathbb{R}^{n-1}$ be a random vector with $\|u\|_2 \leqslant 1$ that depends only on A. Then, for $\delta, \varepsilon > e^{-cn}$ and $s \geqslant 0$, we have*

$$\begin{aligned} \mathbb{E}_A \sup_r \mathbb{P}_X \left(\frac{|\langle A^{-1}X, X\rangle - r|}{\|A^{-1}\|_*} \leqslant \delta,\ \langle X, u\rangle \geqslant s,\ \frac{\mu_1}{\sqrt{n}} \leqslant \varepsilon^{-1} \right) \\ \lesssim \delta e^{-s} \left[\mathbb{E}_A \left(\frac{\mu_1}{\sqrt{n}} \right)^{7/9} \mathbf{1}\left\{ \frac{\mu_1}{\sqrt{n}} \leqslant \varepsilon^{-1} \right\} \right]^{6/7} + e^{-cn}, \end{aligned} \tag{3.65}$$

where $c > 0$ depends only on $B > 0$.

Note that with this lemma we have eliminated all "fine-grained" information about the spectrum of A^{-1} and all that remains is μ_1, which is the reciprocal of the least singular value of the matrix A. We also note that we will only need the full power of Lemma 3.9.1 in Section 3.11; until then, we will apply it with $s = 0, u = 0$.

We now turn our attention to proving Lemma 3.9.1. We start with an application of Theorem 3.1.5, our negative correlation theorem, which we restate here in its full-fledged form.

Theorem 3.9.2. *For $n \in \mathbb{N}$, $\alpha, \gamma \in (0,1), B > 0$ and $\mu \in (0, 2^{-15})$, there are constants $c, R > 0$ depending only on α, γ, μ, B so that the following holds. Let $0 \leqslant k \leqslant c\alpha n$ and $\varepsilon \geqslant \exp(-c\alpha n)$, let $v \in \mathbb{S}^{n-1}$, and let $w_1, \dots, w_k \in \mathbb{S}^{n-1}$ be orthogonal. For $\zeta \in \Gamma_B$, let ζ' be an independent copy of ζ and Z_μ a Bernoulli variable with parameter μ; let $\widetilde{X} \in \mathbb{R}^n$ be a random vector whose coordinates are i.i.d. copies of the random variable $(\zeta - \zeta')Z_\mu$.*

If $D_{\alpha,\gamma}(v) > 1/\varepsilon$ then

$$\mathbb{P}_X \left(|\langle \widetilde{X}, v\rangle| \leqslant \varepsilon \ \text{ and } \ \sum_{j=1}^{k} \langle w_j, \widetilde{X}\rangle^2 \leqslant ck \right) \leqslant R\varepsilon \cdot e^{-ck}. \tag{3.66}$$

The proof of Theorem 3.9.2 is provided in [36]. We now prove Lemma 3.9.3.

Lemma 3.9.3. *Let A be a $n \times n$ real symmetric matrix with $A \in \mathcal{E}$ and set $\mu_i := \sigma_i(A^{-1})$, for all $i \in [n]$. For $B > 0$, $\zeta \in \Gamma_B$, let $X, X' \sim \mathrm{Col}_n(\zeta)$ be independent, let $J \subseteq [n]$ be a μ-random subset with $\mu \in (0, 2^{-15})$, and set $\widetilde{X} := (X - X')_J$. If $k \in [1, cn]$ is such that $s \in (e^{-cn}, \mu_k/\mu_1)$ then*

$$\mathbb{P}_{\widetilde{X}}\left(\|A^{-1}\widetilde{X}\|_2 \leqslant s\mu_1\right) \lesssim se^{-ck}\,, \tag{3.67}$$

where $c > 0$ depends only on B.

Proof. For each $j \in [n]$ we let v_j denote a unit eigenvector of A^{-1} corresponding to μ_j. Using the resulting singular value decomposition of A^{-1}, we may express

$$\|A^{-1}\widetilde{X}\|_2^2 = \langle A^{-1}\widetilde{X}, A^{-1}\widetilde{X}\rangle = \sum_{j=1}^{n} \mu_j^2 \langle \widetilde{X}, v_j\rangle^2$$

and thus

$$\mathbb{P}_{\widetilde{X}}\left(\|A^{-1}\widetilde{X}\|_2\mu_1^{-1} \leqslant s\right) \leqslant \mathbb{P}_{\widetilde{X}}\left(|\langle v_1, \widetilde{X}\rangle| \leqslant s \text{ and } \sum_{j=2}^{k} \frac{\mu_j^2}{\mu_1^2}\langle v_j, \widetilde{X}\rangle^2 \leqslant s^2\right). \tag{3.68}$$

We now use that $s \leqslant 1$ and $\mu_k/\mu_1 \leqslant 1$ in (3.68) to obtain

$$\mathbb{P}_{\widetilde{X}}\left(\|A^{-1}\widetilde{X}\|_2\mu_1^{-1} \leqslant s\right) \leqslant \mathbb{P}_{\widetilde{X}}\left(|\langle v_1, \widetilde{X}\rangle| \leqslant s \text{ and } \sum_{j=2}^{k} \langle v_j, \widetilde{X}\rangle^2 \leqslant 1\right). \tag{3.69}$$

We now carefully observe that we are in a position to apply Theorem 3.1.5 to the right-hand-side of (3.69). The coordinates of $\widetilde{X}$ are of the form $(\zeta - \zeta')Z_\mu$, where Z_μ is a Bernoulli random variable taking 1 with probability $\mu \in (0, 2^{-15})$ and 0 otherwise. Also, the $v_2, \ldots, v_k$ are orthogonal and, importantly, we use that $A \in \mathcal{E}$ to learn that[4] $D_{\alpha,\gamma}(v_1) > 1/s$ by property (3.24), provided we choose the constant $c > 0$ (in the statement of Lemma 3.9.3) to be sufficiently small, depending on μ, B. Thus we may apply Theorem 3.1.5 and complete the proof of the Lemma 3.9.3. □

With this lemma in hand, we establish the following corollary of Lemma 3.5.2.

Lemma 3.9.4. *For $B > 0$ and $\zeta \in \Gamma_B$, let $X \sim \mathrm{Col}_n(\zeta)$ and let A be a $n \times n$ real symmetric matrix with $A \in \mathcal{E}$. If $s > 0$, $\delta \in (e^{-cn}, 1)$ and $u \in \mathbb{S}^{n-1}$ then*

$$\sup_r \mathbb{P}_X\big(\big|\langle A^{-1}X, X\rangle - r\big| \leqslant \delta\mu_1, \langle X, u\rangle \geqslant s\big) \lesssim \delta e^{-s}\sum_{k=2}^{cn} e^{-ck}\left(\frac{\mu_1}{\mu_k}\right)^{2/3} + e^{-cn}\,, \tag{3.70}$$

where $c > 0$ is a constant depending only on B.

[4] Recall here that the constants $\alpha, \gamma > 0$ are implicit in the definition of $\mathcal{E}$ and are chosen so that Lemma 3.4.1 holds.

Proof. We apply Lemma 3.5.2 to the left-hand-side of (3.70) to get

$$\sup_r \mathbb{P}_X\big(\big|\langle A^{-1}X, X\rangle - r\big| \leqslant \delta\mu_1, \langle X, u\rangle \geqslant s\big) \lesssim \delta e^{-s}\int_{-1/\delta}^{1/\delta} I(\theta)^{1/2}\,d\theta + e^{-\Omega(n)}\,, \tag{3.71}$$

where

$$I(\theta) := \mathbb{E}_{J,X_J,X'_J} \exp\big(\langle (X+X')_J, u\rangle - c'\theta^2\mu_1^{-2}\|A^{-1}(X-X')_J\|_2^2\big)\,,$$

and $c' = c'(B) > 0$ is a constant depending only on B and $J \subseteq [n]$ is a μ-random subset. Set

$$\widetilde{X} = (X - X')_J \text{ and } v = A^{-1}\widetilde{X},$$

and apply Hölder's inequality

$$I(\theta) = \mathbb{E}_{J,X_J,X'_J}\Big[e^{\langle (X+X')_J,u\rangle}e^{-c'\theta^2\|v\|_2^2/\mu_1^2}\Big] \lesssim \Big(\mathbb{E}_{\widetilde{X}}e^{-c''\theta^2\|v\|_2^2/\mu_1^2}\Big)^{8/9}\Big(\mathbb{E}_{J,X_J,X'_J}\,e^{9\langle (X+X')_J,u\rangle}\Big)^{1/9}. \tag{3.72}$$

Thus we apply (3.36) to see that the second term on the right-hand-side of (3.72) is $O(1)$. Thus, for each $\theta > 0$ we have

$$I(\theta)^{9/8} \lesssim_B \mathbb{E}_{\widetilde{X}}e^{-c''\theta^2\|v\|_2^2/\mu_1^2} \leqslant e^{-c''\theta^{1/5}} + \mathbb{P}_{\widetilde{X}}(\|v\|_2 \leqslant \mu_1\theta^{-9/10})\,.$$

As a result, we have

$$\int_{-1/\delta}^{1/\delta} I(\theta)^{1/2}\,d\theta \lesssim 1 + \int_1^{1/\delta}\mathbb{P}_{\widetilde{X}}(\|v\|_2 \leqslant \mu_1\theta^{-9/10})^{4/9}\,d\theta \lesssim 1 + \int_\delta^1 s^{-19/9}\mathbb{P}_{\widetilde{X}}(\|v\|_2 \leqslant \mu_1 s)^{4/9}\,ds.$$

To bound this integral, we partition $[\delta, 1] = [\delta, \mu_{cn}/\mu_1] \cup \bigcup_{k=2}^{cn}[\mu_k/\mu_1, \mu_{k-1}/\mu_1]$ and apply Lemma 3.9.3 to bound the integrand depending on which interval s lies in. Note this lemma is applicable since $A \in \mathcal{E}$. We obtain

$$\int_{\mu_k/\mu_1}^{\mu_{k-1}/\mu_1} s^{-19/9}\mathbb{P}_{\widetilde{X}}(\|v\|_2 \leqslant \mu_1 s)^{4/9} \leqslant e^{-ck}\int_{\mu_k/\mu_1}^{\mu_{k-1}/\mu_1} s^{-15/9}\,ds \leqslant e^{-ck}(\mu_1/\mu_k)^{2/3},$$

while

$$\int_\delta^{\mu_{cn}/\mu_1} s^{-19/9}\mathbb{P}_{\widetilde{X}}(\|v\|_2 \leqslant \mu_1 s)^{4/9} \leqslant e^{-cn}\delta^{-3/2} \leqslant e^{-\Omega(n)}.$$

Summing over all k and plugging the result into (3.71) completes the lemma. □

We may now prove Lemma 3.9.1 by using the previous Lemma 3.9.4 along with the properties of the spectrum of A established in Section 3.8.

Proof of Lemma 3.9.1. Let $\mathcal{E}$ be our quasi-random event as defined in Section 3.4 and let

$$\mathcal{E}_0 = \mathcal{E} \cap \left\{\frac{\mu_1}{\sqrt{n}} \leqslant \varepsilon^{-1}\right\}.$$

For fixed $A \in \mathcal{E}_0$ and $u = u(A) \in \mathbb{R}^n$ with $\|u\|_2 \leqslant 1$, we may apply Lemma 3.9.4 with $\delta' = \delta \frac{\|A^{-1}\|_*}{\mu_1}$ to see that

$$\sup_{r\in\mathbb{R}} \mathbb{P}_X\big(\big|\langle A^{-1}X, X\rangle - r\big| \leqslant \delta\|A\|_*, \langle X, u\rangle \geqslant s\big) \lesssim \delta e^{-s}\left(\frac{\|A^{-1}\|_*}{\mu_1}\right)\sum_{k=2}^{cn} e^{-ck}\left(\frac{\mu_1}{\mu_k}\right)^{2/3} + e^{-cn}.$$

By Lemma 3.4.1, $\mathbb{P}_A(\mathcal{E}^c) \lesssim \exp(-\Omega(n))$. Therefore it is enough to show that

$$\mathbb{E}_A^{\mathcal{E}_0}\left(\frac{\|A^{-1}\|_*}{\mu_1}\right)\left(\frac{\mu_1}{\mu_k}\right)^{2/3} \lesssim k \cdot \mathbb{E}_A^{\mathcal{E}_0}\left[\left(\frac{\mu_1}{\sqrt{n}}\right)^{7/9}\right]^{6/7}, \tag{3.73}$$

for each $k \in [2, cn]$. For this, apply Hölder's inequality to the left-hand-side of (3.73) to get

$$\mathbb{E}_A^{\mathcal{E}_0}\left(\frac{\|A^{-1}\|_*}{\mu_1}\right)\left(\frac{\mu_1}{\mu_k}\right)^{2/3} \leqslant \mathbb{E}_A^{\mathcal{E}_0}\left[\left(\frac{\|A^{-1}\|_*}{\mu_1}\right)^{14}\right]^{1/14} \mathbb{E}_A^{\mathcal{E}_0}\left[\left(\frac{\sqrt{n}}{\mu_k}\right)^{28/3}\right]^{1/14} \mathbb{E}_A^{\mathcal{E}_0}\left[\left(\frac{\mu_1}{\sqrt{n}}\right)^{7/9}\right]^{6/7}.$$

We now apply Corollary 3.8.3 to see the first term is $O(1)$ and Lemma 3.8.1 to see that the second term is $O(k)$. This establishes (3.73) and thus Lemma 3.9.1. □

3.10 Intermediate bounds: Bootstrapping the lower tail

In this short section we will use the tools developed so far to prove an "up-to-logarithms" version of Theorem 3.1.1. In the next section, Section 3.11, we will bootstrap this result (once again) to prove Theorem 3.1.1.

Lemma 3.10.1. *For $B > 0$, let $\zeta \in \Gamma_B$, and let $A_n \sim \mathrm{Sym}_n(\zeta)$. Then for all $\varepsilon > 0$*

$$\mathbb{P}(\sigma_{\min}(A_n) \leqslant \varepsilon/\sqrt{n}) \lesssim \varepsilon\sqrt{\log 1/\varepsilon} + e^{-\Omega(n)}.$$

To prove Lemma 3.10.1, we first prove the following "base step" (Lemma 3.10.3) which we then improve upon in three steps, ultimately arriving at Lemma 3.10.1. This "base step" is an easy consequence of Lemma 3.6.2 and Lemma 3.9.1 and actually already improves upon the best known bounds on the least-singular value problem for random symmetric matrices. For this we will need the well-known theorem due to Hanson and Wright [81, 175] (see also [167, Theorem 6.2.1]).

Theorem 3.10.2 (Hanson-Wright)**.** *For $B > 0$, let $\zeta \in \Gamma_B$, let $X \sim \mathrm{Col}_n(\zeta)$ and let M be a $m \times n$ matrix. Then for any $t \geqslant 0$, we have*

$$\mathbb{P}_X\big(|\|MX\|_2 - \|M\|_{\mathrm{HS}}| > t\big) \leqslant 2\exp\left(-\frac{ct^2}{B^4\|M\|^2}\right),$$

where $c > 0$ is absolute constant.

We now prove the base step of our iteration.

Lemma 3.10.3 (Base step)**.** *For $B > 0$, let $\zeta \in \Gamma_B$ and let $A_{n+1} \sim \mathrm{Sym}_{n+1}(\zeta)$. Then for all $\varepsilon > 0$,*

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon/\sqrt{n}) \lesssim \varepsilon^{1/4} + e^{-\Omega(n)}\,.$$

Proof. As usual, we let $A := A_n$. By Lemma 3.6.1, it will be sufficient to show that for $r \in \mathbb{R}$,

$$\mathbb{P}_{A,X}\left(\frac{|\langle A^{-1}X, X\rangle - r|}{\|A^{-1}X\|_2} \leqslant C\varepsilon,\ \sigma_n(A) \geqslant \frac{\varepsilon}{\sqrt{n}}\right) \lesssim \varepsilon^{1/4} + e^{-\Omega(n)}\,. \tag{3.74}$$

By the Hanson-Wright inequality (Theorem 3.10.2), there exists $C' > 0$ so that

$$\mathbb{P}_X(\|A^{-1}X\|_2 \geqslant C'\sqrt{\log 1/\varepsilon}\|A^{-1}\|_{\mathrm{HS}}) \leqslant \varepsilon \tag{3.75}$$

and so the left-hand-side of (3.74) is bounded above by

$$\varepsilon + \mathbb{P}_{A,X}\left(\frac{|\langle A^{-1}X, X\rangle - r|}{\|A^{-1}\|_{\mathrm{HS}}} \leqslant \delta,\ \sigma_n(A) \geqslant \varepsilon/\sqrt{n}\right)\,,$$

where $\delta := C''\varepsilon\sqrt{\log 1/\varepsilon}$. Now, by Lemma 3.9.1 with the choice of $u = 0, s = 0$, we have

$$\mathbb{P}_{A,X}\left(\frac{|\langle A^{-1}X, X\rangle - r|}{\|A^{-1}\|_{\mathrm{HS}}} \leqslant \delta,\ \sigma_n(A) \geqslant \frac{\varepsilon}{\sqrt{n}}\right) \lesssim \delta\varepsilon^{-2/3} + e^{-\Omega(n)} \lesssim \varepsilon^{1/4} + e^{-\Omega(n)}\,, \tag{3.76}$$

where we have used that $\|A^{-1}\|_* \geqslant \|A^{-1}\|_{\mathrm{HS}}$. We also note that Lemma 3.9.1 actually gives an upper bound on $\mathbb{E}_A \sup_r \mathbb{P}_X(\mathcal{A})$, where $\mathcal{A}$ is the event on the left-hand-side of (3.80). Since $\sup_r \mathbb{P}_{A,X}(\mathcal{A}) \leqslant \mathbb{E}_A \sup_r \mathbb{P}_X(\mathcal{A})$, the bound (3.76), and thus Lemma 3.10.3, follows. □

The next lemma is our "bootstrapping step": given bounds of the form

$$\mathbb{P}(\sigma_{\min}(A_n) \leqslant \varepsilon/\sqrt{n}) \lesssim \varepsilon^{\kappa} + e^{-cn},$$

this lemma will produce better bounds for the same problem with A_{n+1} in place of A_n.

Lemma 3.10.4. *(Bootstrapping step) For $B > 0$, let $\zeta \in \Gamma_B$, let $A_{n+1} \sim \mathrm{Sym}_{n+1}(\zeta)$ and let $\kappa \in (0,1) \setminus \{7/10\}$. If for all $\varepsilon > 0$, and all n we have*

$$\mathbb{P}\big(\sigma_{\min}(A_n) \leqslant \varepsilon/\sqrt{n}\big) \lesssim \varepsilon^{\kappa} + e^{-\Omega(n)}\,, \tag{3.77}$$

then for all $\varepsilon > 0$ and all n we have

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon/\sqrt{n}) \lesssim \varepsilon^{\min\{1,6\kappa/7+1/3\}}\sqrt{\log 1/\varepsilon} + e^{-\Omega(n)}\,.$$

Proof. Let $c > 0$ denote the implicit constant in the exponent on the right-hand-side of (3.77). Note that if $0 < \varepsilon < e^{-cn}$, by the assumption of the lemma, then we have

$$\mathbb{P}(\sigma_{\min}(A_n) \leqslant \varepsilon/\sqrt{n}) \lesssim e^{-\Omega(n)},$$

for all n, in which case we are done. So we may assume $\varepsilon > e^{-cn}$.

As in the proof of the "base step", Lemma 3.10.3, we look to apply Lemma 3.6.2 and Lemma 3.9.1 in sequence. For this we write $A = A_n$ and bound (3.65) as in the conclusion of Lemma 3.9.1

$$\mathbb{E}_A \left(\frac{\mu_1}{\sqrt{n}}\right)^{7/9} \mathbf{1}\left\{\frac{\mu_1}{\sqrt{n}} \leqslant \varepsilon^{-1}\right\} \leqslant \int_0^{\varepsilon^{-7/9}} \mathbb{P}\left(\sigma_{\min}(A) \leqslant x^{-9/7} n^{-1/2}\right) dx, \tag{3.78}$$

where we used that $\sigma_{\min}(A) = 1/\mu_1(A)$. Now use assumption (3.77) to see the right-hand-side of (3.78) is

$$\lesssim 1 + \int_1^{\varepsilon^{-7/9}} (x^{-9\kappa/7} + e^{-cn})\, dx \lesssim \max\left\{1, \varepsilon^{\kappa - 7/9}\right\}. \tag{3.79}$$

Now we apply Lemma 3.9.1 with $\delta = C\varepsilon\sqrt{\log 1/\varepsilon}$, $s = 0$ and $u = 0$ to see that

$$\sup_r \mathbb{P}_{A,X}\left(\frac{|\langle A^{-1}X, X\rangle - r|}{\|A^{-1}\|_{\mathrm{HS}}} \leqslant \delta,\ \frac{\mu_1}{\sqrt{n}} \leqslant \varepsilon^{-1}\right) \lesssim \max\left\{\varepsilon, \varepsilon^{6\kappa/7+1/3}\right\}\sqrt{\log 1/\varepsilon} + e^{-\Omega(n)}, \tag{3.80}$$

where we have used that $\|A^{-1}\|_{\mathrm{HS}} \leqslant \|A^{-1}\|_*$.

Now, by Hanson-Wright (Theorem 3.10.2), there exists $C' > 0$ such that

$$\mathbb{P}_X\left(\|A^{-1}X\|_2 \geqslant C'\|A^{-1}\|_{\mathrm{HS}}\sqrt{\log 1/\varepsilon}\right) \leqslant \varepsilon.$$

Thus we choose C'' to be large enough, so that

$$\sup_r \mathbb{P}_{A,X}\left(\frac{|\langle A^{-1}X, X\rangle - r|}{\|A^{-1}X\|_2} \leqslant C''\varepsilon, \sigma_n(A) \geqslant \frac{\varepsilon}{\sqrt{n}}\right) \lesssim \max\left\{\varepsilon, \varepsilon^{6\kappa/7+1/3}\right\}\sqrt{\log 1/\varepsilon} + e^{-\Omega(n)}.$$

Lemma 3.6.1 now completes the proof of Lemma 3.10.4. □

Lemma 3.10.1 now follows by iterating Lemma 3.10.4 three times.

Proof of Lemma 3.10.1. By Lemma 3.10.3 and Lemma 3.10.4 we have

$$\mathbb{P}(\sigma_{\min}(A) \leqslant \varepsilon/\sqrt{n}) \lesssim \varepsilon^{13/21}\sqrt{\log 1/\varepsilon} + e^{-\Omega(n)} \lesssim \varepsilon^{13/21-\eta} + e^{-\Omega(n)},$$

for some small $\eta > 0$. Applying Lemma 3.10.4 twice more gives an exponent of $\frac{127}{147} - \frac{6}{7}\eta$ and then 1, for η small, thus completing the proof. □

3.11 Proof of Theorem 3.1.1

We are now ready to prove our main result, Theorem 3.1.1. We use Lemma 3.6.1 (as in the proof of Lemma 3.10.1)) and Lemma 3.26 to see that that it is enough to prove

$$\mathbb{P}^{\mathcal{E}}\left(\frac{|\langle A^{-1}X, X\rangle - r|}{\|A^{-1}X\|_2} \leqslant C\varepsilon, \text{ and } \sigma_n(A) \geqslant \varepsilon n^{-1/2}\right) \lesssim \varepsilon + e^{-\Omega(n)}\,, \tag{3.81}$$

where C is as in Lemma 3.6.1 and the implied constants do not depend on r. Recall that $\mathcal{E}$ is the quasi-random event defined in Section 3.4.

To prepare ourselves for what follows, we put $\mathcal{E}_0 := \mathcal{E} \cap \{\sigma_{\min}(A) \geqslant \varepsilon/\sqrt{n}\}$ and

$$Q(A,X) := \frac{|\langle A^{-1}X, X\rangle - r|}{\|A^{-1}X\|_2} \quad \text{and} \quad Q_*(A,X) := \frac{|\langle A^{-1}X, X\rangle - r|}{\|A^{-1}\|_*}$$

where

$$\|A^{-1}\|_*^2 = \sum_{k=1}^{n} \mu_k^2 (\log(1+k))^2\,,$$

as defined in Section 3.8. We now split the left-hand-side of (3.81) as

$$\mathbb{P}^{\mathcal{E}_0}\left(Q(A,X) \leqslant C\varepsilon\right) \leqslant \mathbb{P}^{\mathcal{E}_0}\left(Q_*(A,X) \leqslant 2C\varepsilon\right) + \mathbb{P}^{\mathcal{E}_0}\left(Q(A,X) \leqslant C\varepsilon, \frac{\|A^{-1}X\|_2}{\|A^{-1}\|_*} \geqslant 2\right). \tag{3.82}$$

We can take care of the first term easily by combining Lemma 3.9.1 and Lemma 3.10.1.

Lemma 3.11.1. *For $\varepsilon > 0$,*

$$\mathbb{P}^{\mathcal{E}_0}(Q_*(A,X) \leqslant 2C\varepsilon) \lesssim \varepsilon + e^{-\Omega(n)}\,.$$

Proof. Apply Lemma 3.9.1, with $\delta = 2C\varepsilon$, $u = 0$ and $s = 0$ to obtain

$$\mathbb{P}^{\mathcal{E}_0}(Q_*(A,X) \leqslant 2C\varepsilon) \lesssim \varepsilon \left(\mathbb{E}_A \left(\frac{\mu_1}{\sqrt{n}}\right)^{7/9} \mathbf{1}\left\{\frac{\mu_1}{\sqrt{n}} \leqslant \varepsilon^{-1}\right\}\right)^{6/7} + e^{-\Omega(n)}\,.$$

By Lemma 3.10.1 and the calculation at (3.79), the expectation on the right is bounded by a constant. □

We now focus on the latter term on the right-hand-side of (3.82). By considering the dyadic partition $2^j \leqslant \|A^{-1}X\|_2/\|A^{-1}\|_* \leqslant 2^{j+1}$ we have

$$\mathbb{P}^{\mathcal{E}_0}\left(Q(A,X) \leqslant C\varepsilon, \frac{\|A^{-1}X\|_2}{\|A^{-1}\|_*} \geqslant 2\right) \lesssim \sum_{j=1}^{\log n} \mathbb{P}^{\mathcal{E}_0}\left(Q_*(A,X) \leqslant 2^{j+1}C\varepsilon\,, \frac{\|A^{-1}X\|_2}{\|A^{-1}\|_*} \geqslant 2^j\right) + e^{-\Omega(n)}\,. \tag{3.83}$$

Here we have dealt with $j \geqslant \log n$ by using Hanson-Wright (Theorem 3.10.2) and the fact that $\|A^{-1}\|_* \geqslant \|A^{-1}\|_{\mathrm{HS}}$ to see

$$\mathbb{P}_X\big(\|A^{-1}X\|_2 \geqslant \sqrt{n}\|A^{-1}\|_*\big) \lesssim e^{-\Omega(n)}\,.$$

We now show that the event $\|A^{-1}X\|_2 \geqslant t\|A^{-1}\|_*$ implies that X must correlate with one of the eigenvectors of A.

Lemma 3.11.2. *For $t > 0$, we have*

$$\mathbb{P}_X\left(Q_*(A,X) \leqslant 2Ct\varepsilon, \frac{\|A^{-1}X\|_2}{\|A^{-1}\|_*} \geqslant t\right) \leqslant 2\sum_{k=1}^{n} \mathbb{P}_X\left(Q_*(A,X) \leqslant 2Ct\varepsilon, \langle X, v_k\rangle \geqslant t\log(1+k)\right)$$

where $\{v_k\}$ is an orthonormal basis of eigenvectors of A.

Proof. Assume that $\|A^{-1}X\|_2 \geqslant t\|A^{-1}\|_*$ and use the singular value decomposition associated with $\{v_k\}_k$ to write

$$t^2\sum_k \mu_i^2(\log(k+1))^2 = t^2\|A\|_*^2 \leqslant \|A^{-1}X\|_2^2 = \sum_k \mu_k^2\langle v_k, X\rangle^2.$$

Thus

$$\{\|A^{-1}X\|_2 \geqslant t\|A^{-1}\|_*\} \subset \bigcup_k \big\{|\langle X, v_k\rangle| \geqslant t\log(k+1)\big\}\,.$$

To finish the proof of Lemma 3.11.2, we union bound and treat the case of $-X$ the same as X (by possibly changing the sign of v_k) at the cost of a factor of 2. □

Proof of Theorem 3.1.1. Recall that it suffices to establish (3.81). Combining (3.82) with Lemma 3.11.2 and Lemma 3.11.1 tells us that

$$\mathbb{P}^{\mathcal{E}_0}\left(Q(A,X) \leqslant C\varepsilon\right) \lesssim \varepsilon + 2\sum_{j=1}^{\log n}\sum_{k=1}^{n} \mathbb{P}^{\mathcal{E}_0}\left(Q_*(A,X) \leqslant 2^{j+1}C\varepsilon, \langle X, v_k\rangle \geqslant 2^j\log(1+k)\right) + e^{-\Omega(n)}\,. \tag{3.84}$$

We now apply Lemma 3.9.1 for all $t > 0$, with $\delta = 2Ct\varepsilon$, $s = t\log(k+1)$ and $u = v_k$ to see that,

$$\mathbb{P}^{\mathcal{E}_0}\big(Q_*(A,X) \leqslant 2Ct\varepsilon, \langle X, v_k\rangle \geqslant t\log(1+k)\big) \lesssim \varepsilon t(k+1)^{-t}\cdot I^{6/7} + e^{-\Omega(n)}\,. \tag{3.85}$$

where

$$I := \mathbb{E}_A\left(\frac{\mu_1(A)}{\sqrt{n}}\right)^{7/9}\mathbf{1}\left\{\frac{\mu_1(A)}{\sqrt{n}} \leqslant \varepsilon^{-1}\right\}.$$

Using (3.85) in (3.84) yields

$$\mathbb{P}^{\mathcal{E}_0}(Q(A,X) \leqslant C\varepsilon) \lesssim \varepsilon I^{6/7}\sum_{j=1}^{\log n}\sum_{k=1}^{n} 2^j(k+1)^{-2^j} + e^{-\Omega(n)} \lesssim \varepsilon\cdot I^{6/7} + e^{-\Omega(n)},$$

since $\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} 2^j(k+1)^{-2^j} = O(1)$. Now we write

$$I = \mathbb{E}_A\left(\frac{\mu_1(A)}{\sqrt{n}}\right)^{7/9}\mathbf{1}\left\{\frac{\mu_1(A)}{\sqrt{n}} \leqslant \varepsilon^{-1}\right\} \leqslant \int_0^{\varepsilon^{-7/9}} \mathbb{P}\left(\sigma_{\min}(A) \leqslant x^{-9/7}n^{-1/2}\right)\,dx$$

and apply Lemma 3.10.1 to see

$$\int_0^{\varepsilon^{-7/9}} \mathbb{P}\left(\sigma_{\min}(A) \leqslant x^{-9/7}n^{-1/2}\right)\,dx \lesssim \int_1^{\infty} s^{-9/7}\,ds + 1 \lesssim 1.$$

Thus, Lemma 3.6.1 completes the proof of Theorem 3.1.1. □

Chapter 4

A new proof of the efficient container lemma

4.1 Introduction

In this chapter we will provide a much simpler proof of the efficient hypergraph container lemma, with slightly improved bounds.

In order to state the main result of the chapter, we will need to introduce a couple of important notions, which we will use to measure the 'size' of our containers. Let $\mathcal{G}$ and $\mathcal{H}$ be hypergraphs, and write

$$\langle \mathcal{G} \rangle = \bigcup_{E \in \mathcal{G}} \{F \subset V(\mathcal{G}) : E \subset F\}$$

for the up-set generated by $\mathcal{G}$. We say that $\mathcal{G}$ is a *cover* for $\mathcal{H}$ if $\mathcal{H} \subset \langle \mathcal{G} \rangle$. In other words, $\mathcal{G}$ is a cover for $\mathcal{H}$ if for every edge $F \in \mathcal{H}$ there exists an edge $E \in \mathcal{G}$ with $E \subset F$.

Next, for each $p > 0$, define the *p-weight* of $\mathcal{G}$ to be

$$w_p(\mathcal{G}) = \sum_{E \in \mathcal{G}} p^{|E|}.$$

Note that $w_p(\mathcal{G})$ is just the expected number of edges of $\mathcal{G}$ in a p-random subset of $V(\mathcal{G})$. Finally, let $\mathcal{I}(\mathcal{H})$ denote the family of independent sets of $\mathcal{H}$. We are now ready to state our new container theorem.

Theorem 4.1.1. *Let $\mathcal{H}$ be an r-uniform hypergraph with n vertices, and let $0 < p < 1/4r$. There exists a family $\mathcal{S}$ of subsets of $V(\mathcal{H})$, and functions*

$$g \colon \mathcal{I}(\mathcal{H}) \to \mathcal{S} \qquad \text{and} \qquad f \colon \mathcal{S} \to 2^{V(\mathcal{H})},$$

such that:

(a) *For each* $I \in \mathcal{I}(\mathcal{H})$ *we have* $g(I) \subset I \subset f(g(I))$.

(b) *For each* $S \in \mathcal{S}$, *we have* $|S| \leqslant 16r^2pn$.

(c) *If* $X = f(S)$ *for some* $S \in \mathcal{S}$, *then there exists a cover* $\mathcal{G}$ *for* $\mathcal{H}[X]$ *with*

$$w_p(\mathcal{G}) < p|X|$$

and $|E| \geqslant 2$ *for all* $E \in \mathcal{G}$.

4.2 The algorithm

As in the original proofs of the container theorem, we will define f and g using an algorithm. However, our algorithm will differ from previous ones in several important ways. We will first give an informal description of the algorithm, and then provide a precise definition.

The algorithm will receive as inputs an r-uniform hypergraph $\mathcal{H}$, and an independent set $I \in \mathcal{I}(\mathcal{H})$, and will output sets $S, X \subset V(\mathcal{H})$, and a cover $\mathcal{G}$ for $\mathcal{H}[X]$. The algorithm proceeds in rounds: in round i, the inputs will be an r-bounded antichain $\mathcal{H}^{(i)}$ and a set $S_i \subset V(\mathcal{H})$, and the output will be an r-bounded antichain $\mathcal{H}^{(i+1)}$ and a set $S_{i+1} \subset V(\mathcal{H})$. The hypergraph $\mathcal{H}^{(i+1)}$ will moreover be a cover for $\mathcal{H}^{(i)}$ (and thus for $\mathcal{H}$). In order to define $\mathcal{H}^{(i+1)}$, we will use the *p-degree*

$$d_{\mathcal{G}}(L,p) := \sum_{L \subsetneq E \in \mathcal{G}} p^{|E|}$$

of a set $L \subset V(\mathcal{G})$. Note that the set itself does not count towards the p-degree.

We begin with $\mathcal{H}^{(0)} = \mathcal{H}$ and $S_0 = \emptyset$, noting that since $\mathcal{H}$ is r-uniform it is an r-bounded antichain. To perform round i, we first define a hypergraph

$$\mathcal{H}^{(i)}_* = \mathcal{G}^{(i)}_2 \cup \ldots \cup \mathcal{G}^{(i)}_r \qquad \text{where} \qquad \mathcal{G}^{(i)}_j = \big\{E \in \mathcal{H}^{(i)} : |E| = j\big\} \tag{4.1}$$

for each $1 \leqslant j \leqslant r$. We also define a set $X_i \subset V(\mathcal{H})$ by removing every vertex $v \in V(\mathcal{H})$ such that $\{x\}$ is an edge of $\mathcal{H}^{(i)}$. If $w_p(\mathcal{H}^{(i)}_*) < p|X_i|$ then we will stop the algorithm and show that $g(I) = S_i$, $f(S_i) = X_i$ and $\mathcal{G} = \mathcal{H}^{(i)}_*$ have the required properties. On the other hand, if $w_p(\mathcal{H}^{(i)}_*) \geqslant p|X_i|$ then we define $\mathcal{H}^{(i+1)}$ and S_{i+1} as follows.

Let $s > 1$ be minimal such that there exists $L \subset X_i$ satisfying

$$d_{\mathcal{G}^{(i)}_s}(L,p) \geqslant \frac{p^{|L|}}{4r} \tag{4.2}$$

and let $L \subset X_i$ be a maximal set such that (4.2) holds. We now ask whether or not $L \subset I$, and update $\mathcal{H}^{(i)}$ and S_i accordingly. To be precise, if $L \subset I$, then we will add L to the fingerprint

S_i, add the link graph

$$\mathcal{G}_s^{(i)}(L) = \big\{E \setminus L \,:\, L \subset E \in \mathcal{G}_s^{(i)}\big\}$$

to $\mathcal{H}^{(i)}$, and remove all edges of $\mathcal{H}^{(i)}$ containing one of these added edges. That is, we set $S_{i+1} := S_i \cup L$ and

$$\mathcal{H}^{(i+1)} := \big(\mathcal{H}^{(i)} \setminus \big\langle \mathcal{G}_s^{(i)}(L)\big\rangle\big) \cup \mathcal{G}_s^{(i)}(L).$$

On the other hand, if $L \not\subset I$ then we set $S_{i+1} := S_i$ and define

$$\mathcal{H}^{(i+1)} := \big(\mathcal{H}^{(i)} \setminus \big\langle L\big\rangle\big) \cup \big\{L\big\},$$

that is, we replace all edges of $\mathcal{H}^{(i)}$ containing L with the single edge L.

We suspect that for most readers the above description of the algorithm will suffice; however, for the benefit of those readers who prefer a more compact description, the algorithm is defined as follows.

Definition 4.2.1. Let $\mathcal{H}$ be an r-uniform hypergraph, and let $I \in \mathcal{I}(\mathcal{H})$ be an independent set of $\mathcal{H}$. Set $\mathcal{H}^{(0)} := \mathcal{H}$ and $S_0 := \emptyset$, and set $i := 0$. Repeat the following steps until STOP:

1. Define hypergraphs $\mathcal{G}_j^{(i)}$ (for each $1 \leqslant j \leqslant r$) and $\mathcal{H}_*^{(i)}$ as in (4.1), and set
$$X_i := V(\mathcal{H}) \setminus \bigcup_{E \in \mathcal{G}_1^{(i)}} E. \tag{4.3}$$
2. If $w_p(\mathcal{H}_*^{(i)}) < p|X_i|$, then set $J := i$ and STOP.
3. Otherwise let $s > 1$ be minimal such that there exists $L \subset X_i$ satisfying (4.2).
4. Let $L_i \subset X_i$ be a maximal set[1] such that (4.2) holds with $L = L_i$.
5. If $L_i \subset I$, then set $S_{i+1} := S_i \cup L_i$ and $\mathcal{H}^{(i+1)} := \big(\mathcal{H}^{(i)} \setminus \big\langle \mathcal{G}_s^{(i)}(L_i)\big\rangle\big) \cup \mathcal{G}_s^{(i)}(L_i)$.
6. If $L_i \not\subset I$ then set $S_{i+1} := S_i$ and $\mathcal{H}^{(i+1)} := \big(\mathcal{H}^{(i)} \setminus \big\langle L_i\big\rangle\big) \cup \big\{L_i\big\}$.

Define $\mathcal{G} := \mathcal{H}_*^{(J)}$, $S := S_J$ and $X = X_J$. These are the outputs of the algorithm.

The motivation for the choice of L_i above is that we always want to push weight towards the smaller uniformities while also guaranteeing certain maximum co-degree conditions. This way most of the weight will eventually go to edges of size 1, which will imply $w_p(\mathcal{H}_*^{(i)}) < p|X_i|$. The condition that L_i satisfies (4.2) will imply that the algorithm is always pushing enough weight towards the smaller uniformities. On the other hand, the minimality of s guarantees

[1] If there is more than one maximal L_i, fix a canonical way to choose among the maximal sets with the desired property.

that the co-degrees of $\mathcal{G}_j^{(i)}$ will be appropriately bounded for all $j < r$ and the maximality of L_i guarantees the same for $\mathcal{G}_s^{(i)}(L_i)$.

We will show in the analysis that we will always be able to reconstruct the process from S_J, that $\mathcal{H}^{(i)}$ is a cover for $\mathcal{H}$ and, crucially, that $I \in \mathcal{I}(\mathcal{H}^{(i)})$ for all $0 \leqslant i \leqslant J$. This last property motivates the dichotomy between Steps 5 and 6, since if $L_i \subset I$ and $I \in \mathcal{I}(\mathcal{H}^{(i)})$ then $I \in \mathcal{I}(\mathcal{G}_s^{(i)}(L_i))$, as we will see in Lemma 4.3.4. Otherwise if $L_i \not\subset I$ then by definition we may add L_i as an edge and I will still be an independent set.

4.3 The analysis

We will next prove various simple properties of the algorithm described in the previous section. Throughout this section we fix the hypergraph $\mathcal{H}$ and the independent set $I \in \mathcal{I}(\mathcal{H})$ that were the inputs of the algorithm. We will later see that the algorithm always terminates; for the next few lemmas I will assume that it does in this case, and let $\mathcal{G}$, S and X be the output of the algorithm. The first step is to observe that $\mathcal{H}^{(i)}$ is an antichain.

Lemma 4.3.1. *For every $0 \leqslant i \leqslant J$, the hypergraph $\mathcal{H}^{(i)}$ is an antichain. In particular, every edge of the hypergraph $\mathcal{H}_*^{(i)}$ is contained in the set X_i.*

Proof. Note that $\mathcal{H}^{(0)} = \mathcal{H}$ is an antichain, since it is r-uniform. For the induction step, suppose that $\mathcal{H}^{(i)}$ is an antichain, and that $E, F \in \mathcal{H}^{(i+1)}$ with $E \subset F$. Note first that if $E, F \in \mathcal{H}^{(i)}$, then $E = F$, since $\mathcal{H}^{(i)}$ is an antichain. Next, if $E \in \mathcal{H}^{(i+1)} \setminus \mathcal{H}^{(i)}$, then $\mathcal{H}^{(i+1)} \cap \langle E \rangle = E$, by Steps 5 and 6 of the algorithm, so in this case we must also have $E = F$. Finally, if $E \in \mathcal{H}^{(i)}$ and $F \in \mathcal{H}^{(i+1)} \setminus \mathcal{H}^{(i)}$, then either $F \cup L_i \in \mathcal{H}^{(i)}$ or $F = L_i$. Since $E \subsetneq F \cup L_i$ and $\mathcal{H}^{(i)}$ is an antichain, we must have $F = L_i$. However, if $E \subset L_i$ and $E \in \mathcal{H}^{(i)}$, then since $\mathcal{H}^{(i)}$ is an antichain, L_i would have degree 0, contradicting our choice of L_i. This completes the induction step, and hence shows that $\mathcal{H}^{(i)}$ is an antichain.

To deduce that $E \subset X_i$ for every edge $E \in \mathcal{H}_*^{(i)}$, observe that for each vertex $x \in V(\mathcal{H}) \setminus X_i$ we have $\{x\} \in \mathcal{H}^{(i)}$. Since $\mathcal{H}^{(i)}$ is an antichain, it follows that no other edge of $\mathcal{H}^{(i)}$ contains x, and hence the only edges of $\mathcal{H}^{(i)}$ that intersect $V(\mathcal{H}) \setminus X_i$ have size 1, as claimed. □

Next, we need to observe that our lower bound on $w_p(\mathcal{H}_*^{(i)})$ implies that there exists $s > 1$ and a set $L \subset X_i$ such that (4.2) holds.

Lemma 4.3.2. *If $w_p(\mathcal{H}_*^{(i)}) \geqslant p|X_i|$, then there exists $s > 1$ and $L \subset X_i$ such that*

$$d_{\mathcal{G}_s^{(i)}}(L, p) \geqslant \frac{p^{|L|}}{4r}.$$

Proof. Observe first that there exists $2 \leqslant s \leqslant r$ such that $w_p(\mathcal{G}_s^{(i)}) \geqslant p|X_i|/r$, since

$$\sum_{j=2}^{r} w_p(\mathcal{G}_j^{(i)}) = w_p(\mathcal{H}_*^{(i)}) \geqslant p|X_i|.$$

By Lemma 4.3.1, it follows that there exists $v \in X_i$ such that $d_{\mathcal{G}_s^{(i)}}(v,p) \geqslant sp/r$, since

$$\sum_{v \in X_i} d_{\mathcal{G}_s^{(i)}}(v,p) = s \cdot w_p(\mathcal{G}_s^{(i)}) \geqslant \frac{sp|X_i|}{r}.$$

Taking $L = \{v\}$ we have

$$d_{\mathcal{G}_s^{(i)}}(L,p) \geqslant \frac{sp}{r} \geqslant \frac{p^{|L|}}{4r},$$

as claimed. □

We will next show that the output $\mathcal{G}$, S and X of the algorithm has the desired properties. We begin with property (c), which is straightforward to verify.

Lemma 4.3.3. *$\mathcal{G}$ is a cover for $\mathcal{H}[X]$. Moreover, $w_p(\mathcal{G}) < p|X|$ and $|E| \geqslant 2$ for all $E \in \mathcal{G}$.*

Proof. Since $\mathcal{G} = \mathcal{H}_*^{(J)}$, it follows immediately from the definition (4.1) of $\mathcal{H}_*^{(i)}$ that $|E| \geqslant 2$ for all $E \in \mathcal{G}$. Similarly, the bound $w_p(\mathcal{G}) < p|X|$ holds because the algorithm does not terminate until $w_p(\mathcal{H}_*^{(i)}) < p|X_i|$. Thus we only need to show that $\mathcal{G}$ is a cover for $\mathcal{H}[X]$.

To do so, we claim first that $\mathcal{H}^{(i+1)}$ is a cover for $\mathcal{H}^{(i)}$ for each $0 \leqslant i < J$. To show this, let $E \in \mathcal{H}^{(i)}$, and suppose first that $L_i \subset I$. By Step 5 of the algorithm, either

$$E \in \mathcal{H}^{(i+1)} \subset \big\langle \mathcal{H}^{(i+1)} \big\rangle \qquad \text{or} \qquad E \in \big\langle \mathcal{G}_s^{(i)}(L_i) \big\rangle \subset \big\langle \mathcal{H}^{(i+1)} \big\rangle,$$

since $\mathcal{H}^{(i)} \setminus \mathcal{H}^{(i+1)} \subset \big\langle \mathcal{G}_s^{(i)}(L_i) \big\rangle$ and $\mathcal{G}_s^{(i)}(L_i) \subset \mathcal{H}^{(i+1)}$. On the other hand, if $L_i \not\subset I$, then by Step 6 of the algorithm, either

$$E \in \mathcal{H}^{(i+1)} \subset \big\langle \mathcal{H}^{(i+1)} \big\rangle \qquad \text{or} \qquad E \in \big\langle L_i \big\rangle \subset \big\langle \mathcal{H}^{(i+1)} \big\rangle,$$

since $\mathcal{H}^{(i)} \setminus \mathcal{H}^{(i+1)} \subset \langle L_i \rangle$ and $L_i \in \mathcal{H}^{(i+1)}$. Since in either case we have $E \in \langle \mathcal{H}^{(i+1)} \rangle$, it follows that $\mathcal{H}^{(i+1)}$ is a cover for $\mathcal{H}^{(i)}$, as claimed.

Since being a cover is transitive[2] this implies $\mathcal{H}^{(J)}$ is a cover for $\mathcal{H}$. To deduce that $\mathcal{G} = \mathcal{H}_*^{(J)}$ is a cover for $\mathcal{H}[X]$, let $E \in \mathcal{H}[X]$ and (recalling that $\mathcal{H} \subset \langle \mathcal{H}^{(J)} \rangle$) let $E' \in \mathcal{H}^{(J)}$ be such that $E' \subset E$. Now, observe that if $|E'| = 1$, then $E' \cap X = \emptyset$, by the definition of $X = X_J$, a contradiction with the fact that $E' \subset E \subset X$. Therefore $|E'| \geqslant 2$, and so $E' \in \mathcal{H}_*^{(J)} = \mathcal{G}$. It follows that $\mathcal{G}$ is a cover for $\mathcal{H}[X]$, as required. □

We next show that S and X satisfy property (a) of Theorem 4.1.1.

[2]Indeed, if $\mathcal{H}_i \subset \langle \mathcal{H}_{i+1} \rangle$ and $\mathcal{H}_{i-1} \subset \langle \mathcal{H}_i \rangle$, then $\mathcal{H}_{i-1} \subset \langle \mathcal{H}_{i+1} \rangle$.

Lemma 4.3.4. $S \subset I \subset X$.

Proof. It follows immediately from Steps 5 and 6 of the algorithm that $S \subset I$, since $S = S_J$ and we only add L_i to S_i if $L_i \subset I$. To show that $I \subset X$, we will prove that I is an independent set in $\mathcal{H}^{(J)}$, and therefore does not contain any singleton edge of $\mathcal{H}^{(J)}$. Recall that $I \in \mathcal{I}(\mathcal{H}^{(0)})$; we will prove that $I \in \mathcal{I}(\mathcal{H}^{(i)})$ for every $0 \leqslant i \leqslant J$ by induction on i.

Let $0 \leqslant i < J$, and assume that $I \in \mathcal{I}(\mathcal{H}^{(i)})$. Suppose that $I \notin \mathcal{I}(\mathcal{H}^{(i+1)})$, and let $E \in \mathcal{H}^{(i+1)}$ with $E \subset I$. Note that $E \notin \mathcal{H}^{(i)}$, since $I \in \mathcal{I}(\mathcal{H}^{(i)})$. If $L_i \subset I$, then it follows from Step 5 of the algorithm that $E \in \mathcal{G}_s^{(i)}(L_i)$, and therefore $E \cup L_i \in \mathcal{H}^{(i)}$. Since $E \subset I$ and $L_i \subset I$, it follows that $I \notin \mathcal{I}(\mathcal{H}^{(i)})$, which is a contradiction. Similarly, if $L_i \not\subset I$ then it follows from Step 6 of the algorithm that $E = L_i$. But this is again a contradiction, since we assumed that $E \subset I$. Hence $I \in \mathcal{I}(\mathcal{H}^{(i+1)})$, and this completes the induction step.

The induction above implies that $I \in \mathcal{I}(\mathcal{H}^{(J)})$, and hence $\{x\} \notin \mathcal{G}_1^{(J)}$ for every $x \in I$. By the definition (4.3) of $X = X_J$, it follows that $I \subset X$, as required. □

It remains to show that the algorithm always terminates, and that $|S| \leqslant 16r^2pn$. To do so, we will show that $w_p(\mathcal{G}_{<r}^{(i)}) \leqslant 2pn$, and that

$$w_p(\mathcal{G}_{<r}^{(i+1)}) \geqslant w_p(\mathcal{G}_{<r}^{(i)}) + \frac{1}{8r} \tag{4.4}$$

whenever $L_i \subset I$, where $\mathcal{G}_{<r}^{(i)} = \mathcal{G}_1^{(i)} \cup \ldots \cup \mathcal{G}_{r-1}^{(i)}$. It follows from these bounds that there are at most $16rpn$ rounds of the algorithm in which $L_i \subset I$, and this implies the claimed bound on the size of the fingerprint S, since $|L_i| < r$. We will also show that

$$w_p(\mathcal{G}_{<r}^{(i+1)}) \geqslant w_p(\mathcal{G}_{<r}^{(i)}) + \frac{p^r}{2} \tag{4.5}$$

when $L_i \subset I$, which (together with the bounds above) implies that the algorithm terminates after a finite number of rounds.

These three bounds are all fairly straightforward consequences of the following upper bound on the co-degrees in the hypergraphs $\mathcal{G}_s^{(i)}$. Given a hypergraph $\mathcal{G}$ and $\ell \in \mathbb{N}$, we write

$$\Delta_{\ell,p}(\mathcal{G}) = \max\big\{d_{\mathcal{G}}(L,p) : L \subset V(\mathcal{G}),\, |L| = \ell\big\}$$

for the maximum degree of a set of size ℓ in $\mathcal{G}$. We will use the minimality of s and maximality of L_i in the proof of the next lemma, which is the crucial step in the proof.

Lemma 4.3.5. *Let $0 \leqslant i \leqslant J$. For every $2 \leqslant j \leqslant r-1$ and $1 \leqslant \ell \leqslant j-1$, we have*

$$\Delta_{\ell,p}\big(\mathcal{G}_j^{(i)}\big) \leqslant \frac{p^\ell}{2r}. \tag{4.6}$$

Proof. We prove the lemma by induction on i. Note first that the base case $i = 0$ is trivial, since $\mathcal{H}$ is r-uniform, so the hypergraph $\mathcal{G}_s^{(0)}$ is empty for all $1 \leqslant s \leqslant r-1$.

For the induction step, assume that the lemma holds for some $0 \leqslant i < J$. Let $s > 1$ be the integer chosen in Step 2 of the algorithm, and observe that, by the minimality of s,

$$d_{\mathcal{G}_j^{(i)}}(L,p) < \frac{p^{|L|}}{4r} \tag{4.7}$$

for every $j < s$ and $L \subset X_i$. Similarly, by the maximality of L_i, we have

$$d_{\mathcal{G}_s^{(i)}}(L_i \cup L,p) < \frac{p^{|L_i|+|L|}}{4r} \tag{4.8}$$

for every $L \subset X_i \setminus L_i$. We will again consider two cases, depending on whether or not $L_i \subset I$.

First, if $L_i \subset I$, then by Step 5 of the algorithm we have $\mathcal{H}^{(i+1)} \setminus \mathcal{H}^{(i)} \subset \mathcal{G}_s^{(i)}(L_i)$, and therefore every new edge of $\mathcal{H}^{(i+1)}$ has size exactly $s - |L_i|$. We therefore have

$$d_{\mathcal{G}_j^{(i+1)}}(L,p) \leqslant d_{\mathcal{G}_j^{(i)}}(L,p) \leqslant \frac{p^{|L|}}{2r},$$

for every $j \neq s - |L_i|$ and every $L \subset V(\mathcal{H})$, by the induction hypothesis. We are similarly done if $L \not\subset X_i \setminus L_i$, since every edge of $\mathcal{G}_s^{(i)}(L_i)$ is contained in X_i, by Observation 4.3.1, and disjoint from L_i, by Step 5 of the algorithm. On the other hand, if $j = s - |L_i|$ and $L \subset X_i \setminus L_i$, then we have

$$d_{\mathcal{G}_j^{(i+1)}}(L,p) \leqslant d_{\mathcal{G}_j^{(i)}}(L,p) + p^{-|L_i|} \cdot d_{\mathcal{G}_s^{(i)}}(L_i \cup L,p) \leqslant \frac{p^{|L|}}{2r},$$

as required, where the final inequality follows using (4.7) and (4.8), since $j = s - |L_i| < s$.

Now, if $L_i \not\subset I$ then, by Step 6 of the algorithm, the only edge of $\mathcal{H}^{(i+1)} \setminus \mathcal{H}^{(i)}$ is L_i. Note that $|L_i| < s$, since the set L_i is not counted in the degree of L_i, and L_i has non-zero degree in $\mathcal{G}_s^{(i)}$. Observe also that unless $j = |L_i|$ and $L \subsetneq L_i$, we have

$$d_{\mathcal{G}_j^{(i+1)}}(L,p) \leqslant d_{\mathcal{G}_j^{(i)}}(L,p) \leqslant \frac{p^{\ell}}{2r},$$

by the induction hypothesis. Finally, if $j = |L_i|$ and $L \subsetneq L_i$, then we have

$$d_{\mathcal{G}_j^{(i+1)}}(L,p) \leqslant d_{\mathcal{G}_j^{(i)}}(L,p) + p^{|L_i|} \leqslant \frac{p^{|L|}}{4r} + p^{|L|+1} \leqslant \frac{p^{|L|}}{2r}$$

again using (4.7) (since $j = |L_i| < s$), and the bounds $|L| < |L_i|$ and $p < 1/4r$. □

For each $0 \leqslant i \leqslant J$, define $\mathcal{G}_*^{(i)} = \mathcal{G}_2^{(i)} \cup \ldots \cup \mathcal{G}_{r-1}^{(i)}$. We will actually use the following immediate consequence of Lemma 4.3.5.

Lemma 4.3.6. *For each $0 \leqslant i \leqslant J$ and $1 \leqslant \ell \leqslant r-1$, we have*

$$\Delta_{\ell,p}(\mathcal{G}_*^{(i)}) \leqslant \frac{p^{\ell}}{2}.$$

We are now ready to prove the three inequalities, (4.4), (4.5), and $w_p(\mathcal{G}^{(i)}_{<r}) \leqslant 2pn$, which together will allow us to show that the algorithm terminates, and also to deduce the desired bound on the size of S. We begin with the upper bound on $w_p(\mathcal{G}^{(i)}_{<r})$.

Lemma 4.3.7. *For all $0 \leqslant i \leqslant J$, we have*

$$w_p(\mathcal{G}^{(i)}_{<r}) \leqslant 2pn.$$

Proof. Recall that $\mathcal{G}^{(i)}_{<r} = \mathcal{G}^{(i)}_1 \cup \ldots \cup \mathcal{G}^{(i)}_{r-1}$, and observe that

$$w_p(\mathcal{G}^{(i)}_{<r}) = w_p(\mathcal{G}^{(i)}_1) + \sum_{s=2}^{r-1} \frac{1}{s} \sum_{v \in V(\mathcal{H})} d_{\mathcal{G}^{(i)}_s}(v,p).$$

Since $d_{\mathcal{G}^{(i)}_s}(v,p) \leqslant p/2r$ for every $v \in V(\mathcal{H})$, by Lemma 4.3.5, it follows that

$$w_p(\mathcal{G}^{(i)}_{<r}) \leqslant w_p(\mathcal{G}^{(i)}_1) + \sum_{s=2}^{r-1} \frac{pn}{2rs} \leqslant pn + \frac{pn \log r}{r} \leqslant 2pn,$$

where in the second inequality we used the fact that $\mathcal{G}^{(i)}_1$ is a simple 1-uniform hypergraph, so has at most n edges. □

We next prove (4.4), which says that $w_p(\mathcal{G}^{(i)}_{<r})$ increases by at least $1/8r$ whenever $L_i \subset I$.

Lemma 4.3.8. *Let $0 \leqslant i \leqslant J$. If $L_i \subset I$, then*

$$w_p(\mathcal{G}^{(i+1)}_{<r}) \geqslant w_p(\mathcal{G}^{(i)}_{<r}) + \frac{1}{8r}.$$

Proof. Observe first that $\mathcal{G}^{(i)}_s(L_i) \cap \mathcal{H}^{(i)} = \emptyset$, since $\mathcal{H}^{(i)}$ is an antichain, by Observation 4.3.1, and $L_i \neq \emptyset$. By Step 5 of the algorithm, it follows that if $L_i \subset I$, then

$$w_p(\mathcal{G}^{(i+1)}_{<r}) - w_p(\mathcal{G}^{(i)}_{<r}) \geqslant w_p(\mathcal{G}^{(i)}_s(L_i)) - \sum_{E \in \mathcal{G}^{(i)}_s(L_i)} d_{\mathcal{G}^{(i)}_{<r}}(E,p).$$

Using Lemma 4.3.6 to bound $d_{\mathcal{G}^{(i)}_{<r}}(E,p) = d_{\mathcal{G}^{(i)}_*}(E,p)$, it follows that

$$w_p\big(\mathcal{G}^{(i+1)}_{<r}\big) - w_p\big(\mathcal{G}^{(i)}_{<r}\big) \geqslant \sum_{E \in \mathcal{G}^{(i)}_s(L_i)} \left(p^{|E|} - \frac{p^{|E|}}{2}\right) \geqslant \frac{w_p(\mathcal{G}^{(i)}_s(L_i))}{2} \geqslant \frac{1}{8r},$$

by our choice of L_i, as required. □

Finally, we prove (4.5), which implies that $w_p(\mathcal{G}^{(i)}_{<r})$ increases in every round.

Lemma 4.3.9. *Let $0 \leqslant i \leqslant J$. If $L_i \not\subset I$, then*

$$w_p(\mathcal{G}^{(i+1)}_{<r}) \geqslant w_p(\mathcal{G}^{(i)}_{<r}) + \frac{p^r}{2}.$$

Proof. Observe first that $L_i \notin \mathcal{H}^{(i)}$, since $\mathcal{H}^{(i)}$ is an antichain, by Observation 4.3.1, and L_i has non-zero degree in $\mathcal{H}^{(i)}$. By Step 6 of the algorithm, it follows that if $L_i \not\subset I$, then

$$w_p(\mathcal{G}^{(i+1)}_{<r}) - w_p(\mathcal{G}^{(i)}_{<r}) \geqslant p^{|L_i|} - d_{\mathcal{G}^{(i)}_{<r}}(L_i, p).$$

Using Lemma 4.3.6 to bound $d_{\mathcal{G}^{(i)}_{<r}}(L_i, p) = d_{\mathcal{G}^{(i)}_{*}}(L_i, p)$, it follows that

$$w_p\big(\mathcal{G}^{(i+1)}_{<r}\big) - w_p\big(\mathcal{G}^{(i)}_{<r}\big) \geqslant p^{|L_i|} - \frac{p^{|L_i|}}{2} \geqslant \frac{p^r}{2},$$

as required, since $|L_i| \leqslant r$. □

It follows immediately from Lemmas 4.3.7, 4.3.8 and 4.3.9 that the algorithm terminates after a bounded number (to be precise, at most $4p^{1-r}n$) steps. The bound on $|S|$ that is required for property (b) of Theorem 4.1.1 also follows easily.

Lemma 4.3.10. $|S| \leqslant 16r^2pn$

Proof. Suppose there are exactly m rounds of the algorithm in which $L_i \subset I$. By Lemmas 4.3.8 and 4.3.9, we have

$$\frac{m}{8r} \leqslant w_p(\mathcal{G}^{(J)}_{<r}) \leqslant 2pn,$$

and hence $m \leqslant 16rn$. Since $|S_i|$ increases by at most r in each of these rounds, and does not increase otherwise, the claimed bound follows immediately. □

4.4 The proof of the efficient container theorem

We are now almost ready to complete the proof of Theorem 4.1.1; the only missing observation is that the container X of an independent set $I \in \mathcal{I}(\mathcal{H})$ is determined by S, the fingerprint of I. This is shown in the following lemma.

Lemma 4.4.1. *Suppose the algorithm applied to $I, \tilde{I} \in \mathcal{I}(\mathcal{H})$ outputs $(S, X, \mathcal{G})$ and $(\tilde{S}, \tilde{X}, \tilde{\mathcal{G}})$ respectively. If $S = \tilde{S}$, then $X = \tilde{X}$.*

Proof. The proof is not difficult, but it is a little subtle. Observe first that the algorithm only depends on the hypergraph $\mathcal{H}$ and the decision (in each round) whether to perform Step 5 or Step 6. Thus, if the outputs of the algorithm applied to I and $\tilde{I}$ are different, then there must be a round i for which $L_i \subset \tilde{I}$ but $L_i \not\subset I$, or vice versa. Consider the first such round, and note that (by symmetry) we may assume that $L_i \subset \tilde{I}$ and $L_i \not\subset I$.

The crucial observation is now as follows. Note that by Steps 5 and 6 of the algorithm, we have $L_i \subset \tilde{S}$ and $L_i \in \mathcal{H}^{(i+1)}$. Since S is an independent set in $\mathcal{H}^{(i)}$ for every $0 \leqslant i \leqslant J$, it follows that $L_i \not\subset S$, giving the desired contradiction.

To spell out the details, recall from the proof of Lemma 4.3.4 that $I \in \mathcal{I}(\mathcal{H}^{(i)})$ for every $0 \leqslant i \leqslant J$. Since $S \subset I$ (again by Lemma 4.3.4, though in this case the proof was immediate), it follows that we also have $S \in \mathcal{I}(\mathcal{H}^{(i)})$ for every $0 \leqslant i \leqslant J$. Now, since $L_i \in \mathcal{H}^{(i+1)}$, it follows that $L_i \not\subset S$, as claimed. □

We are now ready to prove Theorem 4.1.1.

Proof of Theorem 4.1.1. For each independent set $I \in \mathcal{I}(\mathcal{H})$, define $g(I) := S$ and $f(S) := X$, where $(S, X, \mathcal{G})$ is the output of the algorithm with inputs $\mathcal{H}$ and I, and set $\mathcal{S} := \{g(I) : I \in \mathcal{I}(\mathcal{H})\}$. By Lemma 4.4.1, the function f is well-defined.

Now, by Lemma 4.3.4 we have $g(I) \subset I \subset f(g(I))$ for every $I \in \mathcal{I}(\mathcal{H})$, so property (a) holds. By Lemma 4.3.10 we have $|S| \leqslant 16r^2pn$ for every $S \in \mathcal{S}$, so property (b) holds. Finally, if $X = f(g(I))$ then, by Lemma 4.3.3, the hypergraph $\mathcal{G}$ is a cover for $\mathcal{H}[X]$ with $w_p(\mathcal{G}) < p|X|$, and moreover $|E| \geqslant 2$ for all $E \in \mathcal{G}$, so property (c) holds. This completes the proof of the efficient container theorem. □

4.5 Deducing the standard container theorems

In this section we will show that the usual formulations of the hypergraph container lemma can be easily deduced from Theorem 4.1.1. Given an r-uniform hypergraph $\mathcal{H}$ and a set $L \subset V(\mathcal{H})$, let

$$d_{\mathcal{H}}(L) := \big|\big\{E \in \mathcal{H} : L \subset E\big\}\big|$$

and for each $\ell \geqslant 1$ define

$$\Delta_\ell(\mathcal{H}) := \max\big\{d_{\mathcal{H}}(L) : |L| = \ell\big\}.$$

We will first deduce the following slight strengthening of the efficient container lemma of Balogh and Samotij [18, Theorem 1.1], which itself significantly strengthened the original container lemmas of Balogh, Morris and Samotij [15] and Saxton and Thomason [142].

Corollary 4.5.1. *Let $\mathcal{H}$ be an r-uniform hypergraph on n vertices. Suppose that $\tau \in (0,1)$ and $K > 0$ are such that*

$$\Delta_\ell(\mathcal{H}) \leqslant K \cdot \left(\frac{\tau}{2^5 K r^2}\right)^{\ell-1} \frac{e(\mathcal{H})}{v(\mathcal{H})}. \tag{4.9}$$

for every $\ell \in \{1, \dots, r\}$. Then there exists a family $\mathcal{S}$ of subsets of $V(\mathcal{H})$, and functions

$$g \colon \mathcal{I}(\mathcal{H}) \to \mathcal{S} \qquad \text{and} \qquad f \colon \mathcal{S} \to 2^{V(\mathcal{H})},$$

such that

(a) *For each $I \in \mathcal{I}(\mathcal{H})$ we have $g(I) \subset I \subset f(g(I))$,*

(*b*) *For each* $S \in \mathcal{S}$ *we have* $|S| \leqslant \tau n$,

(*c*) *For each* $S \in \mathcal{S}$ *we have* $|f(S)| \leqslant (1-\delta)n$,

where $\delta = (2K)^{-1}$.

For comparison, in [18] the statement above was proved with $2^5 Kr^2$ replaced by $10^6 r^5$, and with $\delta = (10^3 r^3 K)^{-1}$, while in [15, 142] the parameters had super-exponential dependence on r (see, for example, Section 3 of the survey [16]).

Proof of Corollary 4.5.1. We apply Theorem 4.1.1 with $p = \tau/16r^2$. We obtain a family $\mathcal{S}$ and functions f and g satisfying properties (a) and (b), so we only need to verify property (c).

To do so, let $S \in \mathcal{S}$ and set $X := f(S)$. By property (c) of Theorem 4.1.1, there exists a cover $\mathcal{G}$ for $\mathcal{H}[X]$ with $w_p(\mathcal{G}) < p|X|$ and $|E| \geqslant 2$ for all $E \in \mathcal{G}$. It follows that

$$e\big(\mathcal{H}[X]\big) \leqslant \sum_{E\in\mathcal{G}} d_{\mathcal{H}}(E) \leqslant \sum_{E\in\mathcal{G}} \left(\frac{p}{2}\right)^{|E|-1} \frac{e(\mathcal{H})}{v(\mathcal{H})} \leqslant w_p(\mathcal{G}) \cdot \frac{e(\mathcal{H})}{2pn} < \frac{e(\mathcal{H})}{2},$$

where the first inequality holds by the definition of a cover,[3] the second follows by (4.9) and our choice of p, the third follows from the definition of $w_p(\mathcal{G})$, and the fourth follows from the bound $w_p(\mathcal{G}) < p|X| \leqslant pn$. In the second and third steps we also used the fact that $|E| \geqslant 2$ for all $E \in \mathcal{G}$.

On the other hand, if $|X| > (1-\delta)n$, then

$$e\big(\mathcal{H}[X]\big) \geqslant e(\mathcal{H}) - \big(n - |X|\big)\Delta_1(\mathcal{H}) \geqslant \frac{e(\mathcal{H})}{2},$$

since $\Delta_1(\mathcal{H}) \leqslant K \cdot \frac{e(\mathcal{H})}{v(\mathcal{H})}$ and $\delta = (2K)^{-1}$. This contradiction shows that $|f(S)| \leqslant (1-\delta)n$ for every $S \in \mathcal{S}$, as required, and therefore completes the proof of the container lemma. □

Remark 4.5.2. A careful examination of the proof allows one to improve the bounds in Corollary 4.5.1 somewhat further. To be precise, we only needed the bounds

$$\Delta_1(\mathcal{H}) \leqslant K \cdot \frac{e(\mathcal{H})}{v(\mathcal{H})} \qquad \text{and} \qquad \Delta_\ell(\mathcal{H}) \leqslant \left(\frac{\tau}{2^5 r^2}\right)^{\ell-1} \frac{e(\mathcal{H})}{v(\mathcal{H})}$$

for $\ell \geqslant 2$. Moreover, if we halt the algorithm as soon as $|X| \leqslant (1-\delta)n$, then we gain a factor of roughly $r/\log r$ in Lemma 4.3.7 (since the bound on $\Delta_1(\mathcal{H})$ implies that $K \geqslant r$), and hence also in the size of the fingerprint S. In this case the bounds

$$\Delta_1(\mathcal{H}) \leqslant K \cdot \frac{e(\mathcal{H})}{v(\mathcal{H})} \qquad \text{and} \qquad \Delta_\ell(\mathcal{H}) \leqslant \left(\frac{\tau}{2^5 r \log r}\right)^{\ell-1} \frac{e(\mathcal{H})}{v(\mathcal{H})}$$

would suffice to obtain the conclusion of Corollary 4.5.1.

[3] Indeed, for every edge $F \in \mathcal{H}[X]$, there exists an edge $E \in \mathcal{G}$ with $E \subset F$.

In most applications of the method of hypergraph containers one needs to iterate the hypergraph container lemma, and doing gives a 'packaged' hypergraph container theorem, see e.g. [18, Theorem 1.6]. We can deduce the following much cleaner packaged version with better bounds by applying Theorem 4.1.1 directly.

Corollary 4.5.3. *Let $\mathcal{H}$ be an r-uniform hypergraph with n vertices. Suppose that $0 < \tau < 1/4r^2$ and $m \geqslant n$ are such that*

$$\Delta_\ell(\mathcal{H}) \leqslant \frac{\tau^{\ell-1}m}{n} \tag{4.10}$$

for every $\ell \in \{2, \dots, r\}$. Then there exists a family $\mathcal{C}$ of subsets of $V(\mathcal{H})$, with

$$|\mathcal{C}| \leqslant \exp\Big(16r^2 \log\big((r^2\tau)^{-1}\big)\tau n\Big),$$

such that

(i) For every $I \in \mathcal{I}(\mathcal{H})$, there exists $C \in \mathcal{C}$ such that $I \subset C$.

(ii) For every $C \in \mathcal{C}$, we have $e(\mathcal{H}[C]) < m$.

Proof. We apply Theorem 4.1.1 to $\mathcal{H}$ with $p = \tau$, and set

$$\mathcal{C} = \big\{f(S) : S \in \mathcal{S}\big\}.$$

Observe that

$$|\mathcal{C}| \leqslant |\mathcal{S}| \leqslant \sum_{s=0}^{16r^2\tau n} \binom{n}{s} \leqslant \exp\Big(16r^2 \log\big((r^2\tau)^{-1}\big)\tau n\Big),$$

since by property (b) of Theorem 4.1.1 we have $|S| \leqslant 16r^2\tau n$ for every $S \in \mathcal{S}$. Note also that for each $I \in \mathcal{I}(\mathcal{H})$ there exists $C = f(g(I)) \in \mathcal{C}$ such that $I \subset C$, by property (a).

Finally, let $C \in \mathcal{C}$, and observe that, by property (c) of Theorem 4.1.1, there exists a cover $\mathcal{G}$ for $\mathcal{H}[C]$ with $w_\tau(\mathcal{G}) < \tau|C|$. It follows that

$$e\big(\mathcal{H}[C]\big) \leqslant \sum_{E\in\mathcal{G}} d_{\mathcal{H}}(E) \leqslant \sum_{E\in\mathcal{G}} \frac{\tau^{|E|-1}m}{n} < m,$$

where the first inequality holds by the definition of a cover, the second follows[4] by (4.10), and the third follows from the definition of $w_\tau(\mathcal{G})$ and from the bound $w_\tau(\mathcal{G}) < \tau|C| \leqslant \tau n$. □

Finally, we conjecture that the dependence on r in the bound on the size of the fingerprint in Theorem 4.1.1 can be removed completely.

Conjecture 4.5.4. *There exists a constant $C > 0$ such that the following holds. Let $\mathcal{H}$ be an r-uniform hypergraph with n vertices, and let $p > 0$. There exists a family $\mathcal{S}$ of subsets of $V(\mathcal{H})$,*

[4] Note that (4.10) holds for $\ell \geqslant 2$, and that $|E| \geqslant 2$ for all $E \in \mathcal{G}$.

and functions

$$g\colon \mathcal{I}(\mathcal{H}) \to \mathcal{S} \qquad \text{and} \qquad f\colon \mathcal{S} \to 2^{V(\mathcal{H})},$$

such that:

(a) *For each* $I \in \mathcal{I}(\mathcal{H})$ *we have* $g(I) \subset I \subset f(g(I))$.

(b) *For each* $S \in \mathcal{S}$, *we have* $|S| \leqslant Cpn$.

(c) *If* $X = f(S)$ *for some* $S \in \mathcal{S}$, *then there exists a cover* $\mathcal{G}$ *for* $\mathcal{H}[X]$ *with*

$$w_p(\mathcal{G}) < p|X|$$

and $|E| \geqslant 2$ *for all* $E \in \mathcal{G}$.

Chapter 5

On the number of sets with a given doubling constant

This chapter is adapted from the paper [31] which has been published at Israel Journal of Mathematics.

5.1 Introduction

Our main theorem in this chapter confirms Conjecture 1.4.1 for all $K = o(s/(\log n)^3)$.

Theorem 5.1.1. *Let s, n be integers and $2 \leqslant K \leqslant o\big(\frac{s}{(\log n)^3}\big)$. The number of sets $J \subset [n]$ with $|J| = s$ such that $|J + J| \leqslant K|J|$ is at most*

$$2^{o(s)} \binom{\frac{1}{2}Ks}{s}.$$

We will in fact prove stronger bounds on the error term than those stated above, see Theorem 5.4.1. Nevertheless, we are unable to prove the conjecture in the range $K = \Omega(s/(\log n)^3)$, and actually the conjecture is false for a certain range of values of s and $K \gg s/\log n$. More precisely, for any integers n, s, and any positive numbers K, ϵ with $\min\{s, n^{1/2-\epsilon}\} \geqslant K \geqslant \frac{4\log(24C)s}{\epsilon \log n}$, there are at least

$$\binom{\frac{n}{2}}{\frac{K}{4}} \binom{\frac{Ks}{8}}{s - \frac{K}{4}} \geqslant \binom{CKs}{s}$$

sets $J \subset [n]$ with $|J| = s$ and $|J + J| \leqslant Ks$. The construction is very simple: let P be an arithmetic progression of size $Ks/8$ and set $J = J_0 \cup J_1$, where J_0 is any subset of P of size $s - K/4$, and J_1 is any subset of $[n] \setminus P$ of size $K/4$. For convenience we provide the details in the appendix of [31].

Our methods also allow us to characterize the typical structure of an s-set with doubling constant K, and obtain the following result.

Theorem 5.1.2. *Let s, n be integers and $2 \leqslant K \leqslant o\big(\frac{s}{(\log n)^3}\big)$. For almost all sets $J \subset [n]$ with $|J| = s$ such that $|J + J| \leqslant K|J|$, there is a set $T \subset J$ such that $J \setminus T$ is contained in an arithmetic progression of size $\frac{1+o(1)}{2}Ks$ and $|T| = o(s)$.*

In the case $s = \Omega(n)$ (and hence $K = O(1)$), this result was proved by Mazur [105]. We will provide better bounds for the error terms in Theorem 5.5.1, below.

5.1.1 Abelian Groups

Notice that the doubling constant is defined for finite subsets of any abelian group. So, given a finite subset Y of an abelian group, one might ask: how many subsets of Y of size s with doubling constant K there are? We are also able to provide an answer to this more general question. From now on, fix an arbitrary abelian group G throughout this chapter. To state our main result formally in the context of general abelian groups we define, for each positive real number t, the quantity $\beta(t)$ to be the size of the biggest subgroup of G of size at most t, that is,

$$\beta(t) = \max\big\{|H| : H \leqslant G,\ |H| \leqslant t\big\}. \tag{5.1}$$

Theorem 5.1.3. *Let s, n be integers, $2 \leqslant K \leqslant o(\frac{s}{(\log n)^3})$, and $Y \subset G$ with $|Y| = n$. The number of sets $J \subset Y$ with $|J| = s$ such that $|J + J| \leqslant K|J|$ is at most*

$$2^{o(s)}\binom{\frac{1}{2}(Ks+\beta)}{s},$$

where $\beta := \beta((1 + o(1))Ks)$.

Again we will actually prove somewhat stronger (although slightly more convoluted) bounds for Theorem 5.1.3, see Theorem 5.4.1. We remark that Theorem 5.1.3 implies Theorem 5.1.1, since the only finite subgroup of $\mathbb{Z}$ is the trivial one, so in this case $\beta(t) = 1$ for all t. Finally let us remark that Theorem 5.1.3 is best possible in many cases. Indeed suppose for some integers l, m, that the largest subgroup $H \leqslant G$ with $|H| \leqslant m \leqslant |G|$ is of size $\beta = \frac{m}{2l-1}$, then there are at least

$$\binom{\frac{m+\beta}{2}}{s}$$

sets $J \subset G$ of size s such that $|J + J| \leqslant m$. To see this, take an arithmetic progression $P \subset G/H$ of size l (there exists one because of the choice of H) and consider $B = P + H$. Since $|B + B| \leqslant |P + P||H| = m$, for every set $J \subset B$ of size s we have $|J + J| \leqslant |B + B| \leqslant m$. Therefore, there are at least

$$\binom{\frac{lm}{2l-1}}{s} = \binom{\frac{m+\beta}{2}}{s}$$

sets $J \subset B$ of size s with $|J + J| \leqslant m$.

5.2 The Asymmetric Container Lemma

In this section we will state our main tool and give a brief explanation of how we will apply it to our problem. Let $Y \subset G$, with $|Y| = n$, and observe that when trying to count sets $J \subset Y$ with $|J| = s$ and $|J+J| \leqslant Ks$, one may instead count sets $J \subset Y$ such that there is a set $I \subset Y$ with $J + J \subset I$ and $|I| \leqslant Ks$. Keeping this in mind, the following definition will be useful.

Definition 5.2.1. *Given disjoint copies of $Y+Y$ and Y, namely Y_0, Y_1 respectively, and $A \subset Y_0$ and $B \subset Y_1$, we define $\mathcal{H}(A, B)$ to be the hypergraph with vertex set $V(\mathcal{H}(A, B)) := (Y_0 \setminus A) \cup B$ and edge set*

$$E(\mathcal{H}(A, B)) := \big\{(\{c\}, \{a, b\}) : c \in Y_0 \setminus A,\ a, b \in B,\ a + b = c\big\}.$$

Sometimes when A and B are clear from the context we will denote $\mathcal{H}(A, B)$ simply by $\mathcal{H}$. Notice that $\mathcal{H}(A, B)$ is not uniform since there are edges $(\{c\}, \{a\})$ corresponding to $a + a = c$, but these will not be a problem. The usefulness of Definition 5.2.1 is that now for every pair of sets (I, J) with $J + J \subset I$ we know that $(Y_0 \setminus I) \cup J$ doesn't contain any edges of $\mathcal{H}(A, B)$, so $(Y_0 \setminus I) \cup J$ would usually be called an independent set, but instead we will call the pair (I, J) independent for convenience. Since we have a method for counting what are usually called independent sets in hypergraphs, and each of those is in correspondence to what we call an independent pair, we can obtain a theorem for counting independent pairs.

To state the main tool in this chapter we will need to go into some more slightly technical definitions. We first define a useful generalization of uniform hypergraphs, that includes the hypergraph presented in Definition 5.2.1. Given disjoint finite sets $V_0,\ V_1$ we define an (r_0, r_1)-bounded hypergraph $\mathcal{H}$ on the vertex set $V = V_0 \cup V_1$ to be a set of edges $E(\mathcal{H}) \subset \binom{V_0}{\leqslant r_0} \times \binom{V_1}{\leqslant r_1}$. Note that the hypergraph in Definition 5.2.1 is $(1, 2)$-bounded. Given a pair $(W_0, W_1) \in 2^{V_0} \times 2^{V_1}$, we say (W_0, W_1) *violates* $(e_0, e_1) \in E(\mathcal{H})$ if $e_0 \subset V_0 \setminus W_0$ and $e_1 \subset W_1$. If a set (W_0, W_1) doesn't violate any $(e_0, e_1) \in E(\mathcal{H})$ then we call (W_0, W_1) *independent* with respect to $\mathcal{H}$. Let $\mathcal{F}_{\leqslant m}(\mathcal{H}) \subset 2^{V(\mathcal{H})}$ be the family of independent pairs (W_0, W_1) such that $|W_0| \leqslant m$, and observe that for any pair of sets (I, J), with $|I| \leqslant m$ and $J + J \subset I$, we have $(I, J) \in \mathcal{F}_{\leqslant m}(\mathcal{H}(\emptyset, Y))$. We define the codegree $d_{(L_0,L_1)}(\mathcal{H})$ of $L_0 \subset V_0,\ L_1 \subset V_1$ to be the size of the set

$$\{(e_0, e_1) \in E(\mathcal{H}) :\ L_0 \subset e_0,\ L_1 \subset e_1\}$$

and we define the maximum (ℓ_0, ℓ_1)-codegree of $\mathcal{H}$ to be

$$\Delta_{(\ell_0,\ell_1)} := \max\{d_{(L_0,L_1)}(\mathcal{H}) :\ L_0 \subset V_0,\ L_1 \subset V_1,\ |L_0| = \ell_0,\ |L_1| = \ell_1\}.$$

With all of this in mind we introduce a variant of the asymmetric container lemma of Morris, Samotij and Saxton [111] that we can, once we have suitable supersaturation theorem to check the codegree condition, apply iteratively and prove Theorem 5.1.1.

Theorem 5.2.2. *For all non-negative integers r_0, r_1, not both zero, and each $R > 0$, the following holds. Suppose that $\mathcal{H}$ is a non-empty (r_0, r_1)-bounded hypergraph with $V(\mathcal{H}) = V_0 \cup V_1$, and b, m, and q are integers with $b \leqslant \min\{m, |V_1|\}$, satisfying*

$$\Delta_{(\ell_0,\ell_1)}(\mathcal{H}) \leqslant R\frac{b^{\ell_0+\ell_1-1}}{m^{\ell_0}|V_1|^{\ell_1}}e(\mathcal{H})\left(\frac{m}{q}\right)^{1[\ell_0>0]} \tag{5.2}$$

for every pair $(\ell_0, \ell_1) \in \{0, 1, \dots, r_0\} \times \{0, 1, \dots, r_1\} \setminus \{(0, 0)\}$. Then there exists a family $\mathcal{S} \subset \binom{V_0}{\leqslant r_0 b} \times \binom{V_1}{\leqslant r_1 b}$ and functions

$$f\colon \mathcal{S} \to 2^{V_0} \times 2^{V_1} \qquad \text{and} \qquad g\colon \mathcal{F}_{\leqslant m}(\mathcal{H}) \to \mathcal{S},$$

such that, letting $\delta = 2^{-(r_0+r_1+1)(r_0+r_1)}R^{-1}$:

(i) If $f(g(I, J)) = (A, B)$ with $A \subset V_0$ and $B \subset V_1$, then $A \subset I$ and $J \subset B$.

(ii) For every $(A, B) \in f(\mathcal{S})$ either $|A| \geqslant \delta q$ or $|B| \leqslant (1 - \delta)|V_1|$.

(iii) If $g(I, J) = (S_0, S_1)$ and $f(g(I, J)) = (A, B)$ then $S_0 \subset V_0 \setminus I$ and $S_1 \subset J$, and $|S_0| > 0$ only if $|A| \geqslant \delta q$.

The proof of this variant of the asymmetric container lemma is virtually identical to that in [111], but, for the sake of completeness, it is provided in the appendix of [31]. Let us remark that the main difference between this statement of the asymmetric container lemma and the one in [111] is that we partition the vertex set in two parts and treat them differently, which is essential in our application. More specifically, we will apply the container lemma iteratively in such a way that V_1 will shrink much more than V_0, and to account for this imbalance we must differentiate between the two sets of the partition. Another small difference is that the hypergraph $\mathcal{H}$ doesn't need to be uniform. Finally we observe that if S_0 is non-empty, where $g(I, J) = (S_0, S_1)$, then we must have $|A| \geqslant \delta q$, where $f(g(I, J)) = (A, B)$.

5.3 The Supersaturation Results

We would like to remind the reader that G will always be a fixed abelian group throughout this chapter. To apply Theorem 5.2.2 to our setting we will need, for sets $A, B \subset G$, bounds on the number of pairs $(b_1, b_2) \in B \times B$ such that $b_1 + b_2 \notin A$. In the case $G = \mathbb{Z}$, one such result is Pollard's Theorem [123], which tell us that if $|B| \geqslant (1/2 + \epsilon)|A|$ and $\epsilon < 1/2$ then at least an ϵ^2 proportion of all pairs $(b_1, b_2) \in B \times B$ are such that $b_1 + b_2 \notin A$. To prove similar results for

arbitrary abelian groups one has to have some control on the structure of the group. With this in mind, we define the following quantity.

Definition 5.3.1. *Given finite sets $U, V \subset G$, we define*

$$\alpha(U,V) = \max\big\{|V'| : V' \subset G,\ |V'| \leqslant |V|,\ |\langle V'\rangle| \leqslant |U| + |V| - |V'|\big\}.$$

Given $U, V \subset G$ and $x \in G$ we will use the notation $1_U * 1_V(x)$ to denote the number of pairs $(u,v) \in U \times V$ such that $u + v = x$. The following theorem is the generalization we want of Pollard's theorem for arbitrary abelian groups. It is a simple variant of a result of Hamidoune and Serra [80], but for completeness we provide a proof in the appendix of [31].

Theorem 5.3.2. *Let t be a positive integer and $U, V \subset G$ with $t \leqslant |V| \leqslant |U| < \infty$. Then*

$$\sum_{x\in G} \min(1_U * 1_V(x), t) \geqslant t\big(|U| + |V| - t - \alpha\big), \tag{5.3}$$

where $\alpha := \alpha(U,V)$

This implies the following corollary.

Corollary 5.3.3. *Let $A, B \subset G$ be finite and non-empty sets, let $0 < \epsilon < \frac{1}{2}$ and set $\beta := \beta((1+4\epsilon)|A|)$. If $|B| \geqslant (\frac{1}{2} + \epsilon)(|A| + \beta)$ then there are at least $\epsilon^2|B|^2$ pairs $(b_1, b_2) \in B^2$ such that $b_1 + b_2 \notin A$.*

Proof. Note first that if $|B| \geqslant (1+\epsilon)|A|$ then the result is trivial, since for each element $a \in A$ there are at most $|B|$ pairs $(b_1, b_2) \in B^2$ with $b_1 + b_2 = a$, and therefore there are at least $|B|^2 - |A||B| \geqslant \epsilon^2|B|^2$ pairs in B whose sum is not in A. When $|B| \leqslant (1+\epsilon)|A|$ we will apply Theorem 5.3.2 with $U = V = B$ and $t = \epsilon|B|$. We first observe that

$$\alpha(B,B) \leqslant \max\big(\beta, 2|B| - (1+4\epsilon)|A|\big).$$

Indeed, suppose that $B' \subset G$ satisfies $|\langle B'\rangle| \leqslant 2|B| - |B'|$. If $|\langle B'\rangle| > (1+4\epsilon)|A|$ then $|B'| \leqslant 2|B| - |\langle B'\rangle| \leqslant 2|B| - (1+4\epsilon)|A|$. Otherwise, if $|\langle B'\rangle| \leqslant (1+4\epsilon)|A|$, then by the definition (5.1) of β, we have $|B'| \leqslant |\langle B'\rangle| \leqslant \beta$.

Now by Theorem 5.3.2, we have

$$\sum_{x\in G} \min(1_B * 1_B(x), \epsilon|B|) \geqslant \epsilon|B|\Big((2-\epsilon)|B| - \max\big(\beta, 2|B| - (1+4\epsilon)|A|\big)\Big).$$

By subtracting from both sides the sum over $x \in A$, we obtain

$$\sum_{x\in G\setminus A} \min(1_B * 1_B(x), \epsilon|B|) \geqslant \epsilon|B|\Big((2-\epsilon)|B| - \max\big(\beta, 2|B| - (1+4\epsilon)|A|\big) - |A|\Big).$$

Now, if $2|B| - (1+4\epsilon)|A| \geqslant \beta$, then, using that $|B| \leqslant 2|A|$,

$$\sum_{x \in G \setminus A} 1_B * 1_B(x) \geqslant \epsilon|B|\big(4\epsilon|A| - \epsilon|B|\big) \geqslant \epsilon^2|B|^2$$

as required. Otherwise, if $\beta \geqslant 2|B| - (1+4\epsilon)|A|$, then

$$\sum_{x \in G \setminus A} 1_B * 1_B(x) \geqslant \epsilon|B|\big((2-\epsilon)|B| - \beta - |A|\big) \geqslant \epsilon^2|B|^2,$$

since $|B| \geqslant (\frac{1}{2}+\epsilon)(|A|+\beta)$ and $0 < \epsilon < \frac{1}{2}$, so $(2-\epsilon) - \frac{2}{1+2\epsilon} \geqslant \epsilon$. □

To prove a stability theorem for almost all sets with a given size and doubling constant we will also need the following result of Mazur [105].

Theorem 5.3.4. *Let l and t be positive integers, with $t \leqslant l/40$, and let $B \subset \mathbb{Z}$ be a set of size l. Suppose that*

$$\sum_{x \in \mathbb{Z}} \min(1_B * 1_B(x), t) \leqslant (2+\delta)lt,$$

for some $0 < \delta \leqslant 1/8$. Then there is an arithmetic progression P of length at most $(1+2\delta)l + 6t$ containing all but at most $3t$ points of B.

From Theorem 5.3.4 we can easily deduce the following corollary:

Corollary 5.3.5. *Let s be an integer, $K > 0$, and $0 < \epsilon < 2^{-10}$. If $A, B \subset \mathbb{Z}$, with $(1-\epsilon)\frac{Ks}{2} \leqslant |B| \leqslant (1+2\epsilon)\frac{Ks}{2}$ and $|A| \leqslant Ks$ then one of the following holds:*

- *(a) There are at least $4\epsilon^2K^2s^2$ pairs $(b_1, b_2) \in B^2$ such that $b_1 + b_2 \notin A$.*
- *(b) There is an arithmetic progression P of size at most $\frac{Ks}{2} + 32\epsilon Ks$ containing all but at most $8\epsilon Ks$ points of B.*

Proof. Suppose first that

$$\sum_{x \in \mathbb{Z}} \min(1_B * 1_B(x), t) \leqslant (2+8\epsilon)2\epsilon|B|Ks. \tag{5.4}$$

In this case we apply Theorem 5.3.4 with $l := |B|$, $\delta := 8\epsilon$, and $t = 2\epsilon Ks \leqslant l/40$, and deduce that (b) holds. Therefore suppose (5.4) doesn't hold, in this case

$$\sum_{x \in \mathbb{Z} \setminus A} \min(1_B * 1_B(x), t) \geqslant (2+8\epsilon)(1-\epsilon)\epsilon K^2s^2 - t|A|,$$

since $|B| \geqslant (1-\epsilon)\frac{1}{2}Ks$. Noting that $t|A| \leqslant 2\epsilon K^2s^2$ it follows that

$$\sum_{x \in \mathbb{Z} \setminus A} 1_B * 1_B(x) \geqslant \Big((2+8\epsilon)(1-\epsilon) - 2\Big)\epsilon K^2s^2 \geqslant 4\epsilon^2K^2s^2,$$

since $\epsilon < 2^{-10}$, so (a) holds as required. □

5.4 The Number of Sets with a given Doubling

In this section we prove the following statement which implies Theorems 5.1.1 and 5.1.3.

Theorem 5.4.1. *Let s, n be integers, let $2 \leqslant K < 2^{-36}\frac{s}{(\log n)^3}$, and let $Y \subset G$ with $|Y| = n$. The number of sets $J \subset Y$ with $|J| = s$ such that $|J + J| \leqslant K|J|$ is at most*

$$\exp\left(2^9 \lambda K^{1/6} s^{5/6} \sqrt{\log n}\right) \binom{\frac{1}{2}(Ks + \beta)}{s},$$

where $\beta := \beta\big(Ks + 2^6 K^{7/6} s^{5/6} \sqrt{\log n}\big)$ and $\lambda := \min\left\{\frac{K}{K-2}, \log s\right\}$.

Theorem 5.4.1 will follow easily from the following container theorem combined with Corollary 5.3.3. We will also use it together with Corollary 5.3.5 to prove Theorem 5.5.1.

Theorem 5.4.2. *Let m, n be integers with $m \geqslant (\log n)^2$, let $Y \subset G$ with $|Y| = n$, and let $0 < \epsilon < \frac{1}{4}$. There is a family $\mathcal{A} \subset 2^{Y+Y} \times 2^Y$ of pairs of sets (A, B), of size*

$$|\mathcal{A}| \leqslant \exp\left(2^{16} \frac{1}{\epsilon^2} \sqrt{m} (\log n)^{3/2}\right) \tag{5.5}$$

such that:

(i) For every pair of sets $J \subset Y$, $I \subset Y + Y$, with $J + J \subset I$ and $|I| \leqslant m$ there is $(A, B) \in \mathcal{A}$ such that $A \subset I$ and $J \subset B$.

(ii) For every $(A, B) \in \mathcal{A}$, $|A| \leqslant m$ and either $|B| \leqslant \frac{m}{\log n}$ or there are at most $\epsilon^2 |B|^2$ pairs $(b_1, b_2) \in B \times B$ such that $b_1 + b_2 \notin A$.

Proof that Theorem 5.4.2 implies Theorem 5.4.1. Let $\mathcal{A}$ be a family given by Theorem 5.4.2 applied with $m := Ks$ and $\epsilon > 0$ to be chosen later. Then by condition (i), for every s-set J with doubling constant K there is a pair $(A, B) \in \mathcal{A}$ such that $J \subset B$ and $A \subset J + J$. Define $\mathcal{B}$ to be the family of all sets B that are in some container pair, that is

$$\mathcal{B} = \{B \subset Y : \exists A \text{ such that } (A, B) \in \mathcal{A}\}.$$

Observe that, by Corollary 5.3.3 and condition (ii) on $\mathcal{A}$, for every $B \in \mathcal{B}$ we have $|B| \leqslant (\frac{1}{2} + \epsilon)(m + \beta)$, where $\beta := \beta((1 + 4\epsilon)m)$, since the number of pairs $(b_1, b_2) \in B^2$ such that $b_1 + b_2 \notin A$ is at most $\epsilon^2 |B|^2$ and $\frac{m}{\log n} \leqslant (\frac{1}{2} + \epsilon)(m + \beta)$. Therefore the number of sets of size s with doubling constant K is at most

$$|\mathcal{B}| \max_{B \in \mathcal{B}} \binom{|B|}{s} \leqslant \exp\left(2^{16} \frac{1}{\epsilon^2} \sqrt{Ks} (\log n)^{3/2}\right) \binom{(\frac{1+2\epsilon}{2})(Ks + \beta)}{s}. \tag{5.6}$$

Let $\lambda := \min\{\frac{K}{K-2}, \log s\}$, suppose first that $\frac{K}{K-2} \leqslant \log s$. By applying the inequality $\binom{cn}{k} \leqslant (\frac{cn-k}{n-k})^k \binom{n}{k}$ with $k = s$, $c = 1 + 2\epsilon$ and $n = \frac{Ks+\beta}{2}$, it follows that in this case (5.6) is at most

$$\exp\left(2^{16}\frac{1}{\epsilon^2}\sqrt{Ks}(\log n)^{3/2} + 2\epsilon\lambda s\right)\binom{\frac{Ks+\beta}{2}}{s}.$$

Now choosing $\epsilon := 2^4\big(\frac{K}{s}\big)^{1/6}\sqrt{\log n}$, by our restrictions on K we see that

$$\epsilon < 2^4\left(\frac{1}{2^{36}(\log n)^3}\right)^{1/6}\sqrt{\log n} = \frac{1}{4}.$$

It follows that there are at most $\exp\big(2^9\lambda K^{1/6}s^{5/6}\sqrt{\log n}\big)\binom{\frac{1}{2}(Ks+\beta)}{s}$ sets of size s with doubling constant K, when $\frac{K}{K-2} \leqslant \log s$. If $\log s \leqslant \frac{K}{K-2}$ we use the binomial estimate

$$\binom{(\frac{1+2\epsilon}{2})(Ks+\beta)}{s} \leqslant \exp\left(4\epsilon s \log\frac{1}{\epsilon}\right)\binom{\frac{Ks+\beta}{2}}{s}$$

and the result follows by a similar calculation. Since $\beta(m+4\epsilon m) = \beta(Ks + 2^6K^{7/6}s^{5/6}\sqrt{\log n})$, this proves the theorem. □

Before we proceed with the proof of Theorem 5.4.2, let us give a brief overview of how we will deduce it from Theorem 5.2.2. We fix from now on a finite subset $Y \subset G$ with $|Y| = n$, and recall that the $(1,2)$-bounded hypergraph $\mathcal{H}(A,B)$ in Definition 5.2.1 was defined to have as edges pairs $(\{c\},\{a,b\})$ where $a+b=c$, with $a,b \in B$ and $c \notin A$. Note that condition (ii) in Theorem 5.4.2 implies that $\mathcal{H}(A,B)$ has at most $\frac{\epsilon^2}{2}|B|^2$ edges, as long as $|B| > \frac{m}{\log n}$. We remind the reader that a pair of sets $I \subset Y+Y$ and $J \subset Y$ with $J+J \subset I$ correspond to an independent set in $\mathcal{H}(A,B)$ for any $A \subset Y+Y$ and $B \subset Y$, since there are no $c \notin I$ and $a,b \in J$ such that $a+b=c$. If we additionally assume that $(I,J) \in \mathcal{F}_{\leqslant m}(\mathcal{H})$, then we know that every J that is in such an independent pair satisfies $|J+J| \leqslant m$.

Our strategy will be to iteratively apply the container lemma until either there are few edges in the hypergraph $\mathcal{H}(A,B)$, or $|A| > m$, in which case the container doesn't contain any elements of $\mathcal{F}_{\leqslant m}(\mathcal{H})$. More precisely we will build a rooted tree $\mathcal{T}$ with root $\mathcal{H}(\emptyset, Y)$ whose vertices correspond to hypergraphs $\mathcal{H}(A,B)$ and whose leaves correspond to a family $\mathcal{A}$ satisfying the conclusion of Theorem 5.4.2. Given a vertex $\mathcal{H}(A,B)$ of the tree, such that $|A| \leqslant m$, $|B| > \frac{m}{\log n}$ and

$$e(\mathcal{H}(A,B)) > \frac{\epsilon^2}{2}|B|^2, \tag{5.7}$$

we will generate its children by applying the following procedure:

(a) Apply the asymmetric container lemma (Theorem 5.2.2) to $\mathcal{H} := \mathcal{H}(A,B)$ setting

$$R := \frac{2}{\epsilon^2}, \quad q := \frac{m}{\log n}, \quad b := \sqrt{\frac{m}{\log n}}.$$

Notice that the co-degrees of $\mathcal{H}$ satisfy

$$\max\left\{\Delta_{(1,0)}(\mathcal{H}),\Delta_{(0,1)}(\mathcal{H})\right\} \leqslant |B| = \frac{2}{\epsilon^2}\frac{\epsilon^2|B|^2}{2|B|} \leqslant R\frac{e(\mathcal{H})}{|B|}$$

and

$$\Delta_{(0,2)}(\mathcal{H}) = \Delta_{(1,1)}(\mathcal{H}) = \Delta_{(1,2)}(\mathcal{H}) = 1 = \frac{2}{\epsilon^2}\frac{b^2}{q|B|^2}\frac{\epsilon^2}{2}|B|^2 \leqslant R\frac{b^2}{q|B|^2}e(\mathcal{H}),$$

since (5.7) holds. Since $b < q < |B|$, it follows that

$$\Delta_{(0,2)}(\mathcal{H}) \leqslant R\frac{b^2}{q|B|^2}e(\mathcal{H}) \leqslant R\frac{b}{|B|^2}e(\mathcal{H}),$$

$$\Delta_{(1,1)}(\mathcal{H}) \leqslant R\frac{b^2}{q|B|^2}e(\mathcal{H}) \leqslant R\frac{b}{q|B|}e(\mathcal{H})$$

and

$$\Delta_{(1,0)}(\mathcal{H}) \leqslant R\frac{e(\mathcal{H})}{|B|} \leqslant R\frac{e(\mathcal{H})}{q},$$

as required.

(b) By Theorem 5.2.2, there exists a family $\mathcal{C} \subset 2^{(Y+Y)\setminus A} \times 2^B$ of at most

$$\binom{n^2}{b}\binom{|B|}{2b} \leqslant n^{4b} \leqslant e^{4\sqrt{m\log n}}, \tag{5.8}$$

pairs of sets (C,D) that satisfies the conditions of the container lemma. That is for each independent pair $(I,J) \in \mathcal{F}_{\leqslant m}(\mathcal{H})$, with $I \subset Y+Y$ and $J \subset Y$, there is $(C,D) \in \mathcal{C}$ such that $C \subset I$ and $J \subset D$, and either $|C| \geqslant \delta\frac{m}{\log n}$, or $D \leqslant (1-\delta)|B|$.

(c) For each $(C,D) \in \mathcal{C}$, let $\mathcal{H}(A\cup C, D)$ be a child of $\mathcal{H}(A,B)$ in the tree $\mathcal{T}$.

Now to count the number of leaves of $\mathcal{T}$ we will first bound its depth.

Lemma 5.4.3. *The tree $\mathcal{T}$ has depth at most $d = 2^{14}\epsilon^{-2}\log n$.*

Proof. We will prove that after d iterations either $|A| > m$, $|B| \leqslant \frac{m}{\log n}$ $e(\mathcal{H}(A,B)) \leqslant \frac{\epsilon^2}{2}|B|^2$. Notice that the δ provided by Theorem 5.2.2 in this application is $2^{-13}\epsilon^2$ and in each iteration either we increase the size of A by δq or we decrease the size of B by $\delta|B|$. After d iterations, either we would have increased the size of A more than $\frac{d}{2}$ times, in which case

$$|A| > \frac{d}{2}\delta q = \frac{2^{13}\log n}{\epsilon^2}2^{-13}\epsilon^2\frac{m}{\log n} = m,$$

or we would have reduced the size of B at least $\frac{d}{2}$ times, in which case

$$|B| \leqslant (1-\delta)^{\frac{d}{2}}n < e^{-\frac{\delta d}{2}}n \leqslant e^{-\log n}n = 1.$$

In either case, we would have stopped already by this point because we only generate children of $\mathcal{H}(A,B)$ if $|A| \leqslant m$, $|B| > \frac{m}{\log n}$ and (5.7) holds. □

Proof of Theorem 5.4.2. Let $\mathcal{L}$ be the set of leaves of the tree $\mathcal{T}$ constructed above, and define

$$\mathcal{A} := \{(A,B) : A \subset Y + Y,\ B \subset Y,\ \mathcal{H}(A,B) \in \mathcal{L},\ |A| \leqslant m\}.$$

Notice that for every $(A,B) \in \mathcal{A}$, we have either the bound $e(\mathcal{H}(A,B)) \leqslant \frac{\epsilon^2}{2}|B|^2$ or $|B| \leqslant \frac{m}{\log n}$, since they come from the leaves of $\mathcal{T}$ and $|A| \leqslant m$. Since the edges of $\mathcal{H}(A,B)$ correspond exactly to pairs $a, b \in B$ such that $a + b \notin A$, it follows that $\mathcal{A}$ has property (ii).

To bound the size of $\mathcal{A}$, notice that the number of leaves of the tree $\mathcal{T}$ is at most Z^d where Z denotes the maximum number of children of a vertex of the tree and d denotes its depth. Thus, by (5.8) and Lemma 5.4.3,

$$|\mathcal{A}| \leqslant |\mathcal{L}| \leqslant Z^d \leqslant \exp\left(2^{16}\frac{1}{\epsilon^2}\sqrt{m}(\log n)^{3/2}\right),$$

so $\mathcal{A}$ satisfies (7.4), as required.

Finally, observe that for every pair of sets $J \subset Y$, $I \subset Y + Y$ with $J + J \subset I$ and $|I| \leqslant m$, there is $(A,B) \in \mathcal{A}$ such that $A \subset I$ and $J \subset B$. Indeed $(I,J) \in \mathcal{F}_{\leqslant m}\big(\mathcal{H}(\emptyset, Y)\big)$ and therefore, by property (b) of our containers, there exists a path from the root to a leaf of $\mathcal{T}$ such that $A \subset I$ and $J \subset B$ for every vertex $\mathcal{H}(A,B)$ of the path, so (i) holds. □

5.5 Typical Structure Result

In this section we use Theorem 5.4.2 to determine the typical structure of a set $J \subset [n]$ of a given size with doubling constant K.

Theorem 5.5.1. *Let s, n be integers, let $2 \leqslant K \leqslant \frac{s}{2^{120}(\log n)^3}$, and let $J \subset [n]$ be a uniformly chosen random set with $|J| = s$ and $|J + J| \leqslant K|J|$. With probability at least $1 - \exp(-K^{1/6}s^{5/6}\sqrt{\log n})$ the following holds: there is a set $T \subset J$, of size $|T| \leqslant 2^{15}K^{1/6}s^{5/6}\sqrt{\log n}$, such that $J \setminus T$ is contained in an arithmetic progression of size*

$$\frac{Ks}{2} + 2^{17}K^{7/6}s^{5/6}\sqrt{\log n}.$$

The proof of Theorem 5.5.1 is similar to that of Theorem 5.4.1, but we use Corollary 5.3.5 as well as Corollary 5.3.3.

Proof of Theorem 5.5.1. Let $G := \mathbb{Z}$ and apply Theorem 5.4.2 to the set $Y := [n]$ with $m := Ks$ and $\epsilon > 0$ to be chosen later. We say $B \subset [n]$ is (ϵ, Ks)-*close to an arithmetic progression* if there is an arithmetic progression P with $|P| \leqslant \frac{Ks}{2} + 2^5\epsilon Ks$, and a set $T \subset B$ with $|T| \leqslant 2^5\epsilon|B|$

such that $B \setminus T \subset P$. We claim that if $\mathcal{A}$ is the family provided by Theorem 5.4.2, then for every pair $(A, B) \in \mathcal{A}$ either

(I) $|B| \leqslant (1-\epsilon)\frac{Ks}{2}$ or

(II) B is (ϵ, Ks)-close to an arithmetic progression.

To see this, note first that, by condition (ii) in Theorem 4.2, for every pair $(A, B) \in \mathcal{A}$ either there are at most $\epsilon^2|B|^2$ pairs $b_1, b_2 \in B$ with $b_1 + b_2 \notin A$ or $|B| \leqslant \frac{m}{\log n}$, and so, by Corollary 5.3.3, $|B| \leqslant (1+2\epsilon)\frac{Ks}{2}$. Now, if (I) doesn't hold, that is $|B| \geqslant (1-\epsilon)\frac{Ks}{2}$, then, by Corollary 5.3.5, (II) holds, since there are at most $\epsilon^2|B|^2 < 4\epsilon^2K^2s^2$ pairs $b_1, b_2 \in B$ such that $b_1 + b_2 \notin A$.

Now we will count the number of sets J of size s and doubling constant K such that J is not $(2^4\epsilon, Ks)$-close to an arithmetic progression. Recall from Theorem 5.4.2 (i) that, for any such set, there exists $(A, B) \in \mathcal{A}$ such that $J \subset B$. Now, observe that there are at most $|\mathcal{A}|\binom{(1-\epsilon)\frac{Ks}{2}}{s}$ sets J of size s that are contained in a set B such that $(A, B) \in \mathcal{A}$ and $|B| \leqslant (1-\epsilon)\frac{Ks}{2}$. Choosing $\epsilon := 2^6(\frac{K}{s})^{1/6}\sqrt{\log n} < 2^{-10}$ and using the bound (7.4) on the size of $\mathcal{A}$, we obtain

$$\begin{aligned} |\mathcal{A}|\binom{(1-\epsilon)\frac{Ks}{2}}{s} &\leqslant \exp\left(2^{16}\epsilon^{-2}\sqrt{Ks}(\log n)^{3/2} - \epsilon s\right)\binom{\frac{Ks}{2}}{s} \\ &\leqslant \exp\left(-2^5K^{1/6}s^{5/6}(\log n)^{1/2}\right)\binom{\frac{Ks}{2}}{s}. \end{aligned} \tag{5.9}$$

Finally we count the number of sets J of size s that are not $(2^4\epsilon, Ks)$-close to an arithmetic progression and are contained in a set B such that $(A, B) \in \mathcal{A}$ and B is (ϵ, Ks)-close to an arithmetic progression. For each such B, let P be an arithmetic progression with $|P| \leqslant \frac{Ks}{2} + 2^5\epsilon Ks$, and $T \subset B$ be a set with $|T| \leqslant 2^5\epsilon|B| \leqslant 2^5\epsilon Ks$, such that $B \setminus T \subset P$. Observe that, there at most

$$\sum_{s' \geqslant 2^9\epsilon s} \binom{(1+2\epsilon)\frac{Ks}{2}}{s-s'}\binom{2^5\epsilon Ks}{s'} \tag{5.10}$$

s-sets $J \subset B$ that are not $(2^4\epsilon, Ks)$-close to an arithmetic progression, since they must have $s - s'$ elements in $B \setminus T$ and s' elements in T for some $s' \geqslant 2^9\epsilon s$. Indeed, otherwise $J \setminus T \subset P$, with $|P| \leqslant Ks + 2^9\epsilon Ks$ and $|J \cap T| < 2^9\epsilon|J|$. To bound this we will use

$$\binom{a}{c-d}\binom{b}{d} \leqslant \binom{a}{c}\left(\frac{4bc}{ad}\right)^d,$$

valid for $d \leqslant c \leqslant a/4$. Note that, by our choice of ϵ, we have $|\mathcal{A}| \leqslant e^{\epsilon s}$ (cf. (5.9)). Hence summing (5.10) over $(A, B) \in \mathcal{A}$ we obtain[1]

$$\begin{aligned} |\mathcal{A}| \cdot s \max_{s' \geqslant 2^9 \epsilon s} (1+4\epsilon)^s \binom{\frac{Ks}{2}}{s-s'} \binom{2^5 \epsilon K s}{s'} &\leqslant |\mathcal{A}| \cdot s \max_{s' \geqslant 2^9 \epsilon s} (1+4\epsilon)^s \binom{\frac{Ks}{2}}{s} \left(\frac{2^8 \epsilon s}{s'}\right)^{s'} \\ &\leqslant \left(\frac{2^8 \epsilon s}{2^9 \epsilon s}\right)^{2^9 \epsilon s} 2^{6\epsilon s} \binom{\frac{Ks}{2}}{s} \leqslant \exp\left(-2^{11} K^{1/6} s^{5/6} \sqrt{\log n}\right) \binom{\frac{Ks}{2}}{s}. \end{aligned} \tag{5.11}$$

Finally observe that the bound (5.9) and (5.11) imply the probability we claimed in the statement since, by taking all subsets of size s of an arithmetic progression of length $\frac{Ks}{2}$, there are at least $\binom{\frac{Ks}{2}}{s}$ sets of size s and doubling constant K. □

[1] We remark that if $K < 16$ then $\binom{2^5 \epsilon K s}{s'} = 0$ for all $s' \geqslant 2^9 \epsilon s$, so we may suppose that $K \geqslant 16$.

Chapter 6

The typical structure of sets with small sumset

This chapter presents joint work with Maurício Collares, Robert Morris, Natasha Morrison and Victor Souza. It is adapted from the paper [32] which has been published in the journal International Mathematics Research Notices.

6.1 Introduction

Here we will obtain a significantly more precise structural description in the case $\lambda = O(1)$. For each $\lambda \geqslant 3$ and $\varepsilon > 0$, define

$$c(\lambda, \varepsilon) := 2^{20}\lambda^2 \log(1/\varepsilon) + 2^{560}\lambda^{32}. \tag{6.1}$$

Our main theorem, which determines (up to an additive constant) the length of the smallest arithmetic progression containing a typical set with bounded doubling[1], is as follows.

Theorem 6.1.1. *Fix $\lambda \geqslant 3$ and $\varepsilon > 0$, let $n \in \mathbb{N}$ be sufficiently large, and let $k \geqslant (\log n)^4$. Let $A \subset [n]$ be chosen uniformly at random from the sets with $|A| = k$ and $|A + A| \leqslant \lambda k$. Then there exists an arithmetic progression P with*

$$A \subset P \qquad \text{and} \qquad |P| \leqslant \frac{\lambda k}{2} + c(\lambda, \varepsilon)$$

with probability at least $1 - \varepsilon$.

[1] We (informally) call $|A + A|/|A|$ the *doubling* of A, so A has bounded doubling if $|A + A| = O(|A|)$.

When λ is large and ε is very small the constant $c(\lambda,\varepsilon)$ is not far from best possible. Indeed, a simple construction (see Section 6.10) shows that with probability at least ε the smallest arithmetic progression containing A has size $\lambda k/2 + \Omega\big(\lambda^2 \log(1/\varepsilon)\big)$.

We will use Theorem 6.1.1 to deduce the following counting result.

Corollary 6.1.2. *For every $\lambda \geqslant 3$, and every $n, k \in \mathbb{N}$ with $(\log n)^4 \leqslant k = o(n)$, we have*

$$\big|\big\{A \subset [n] : |A| = k,\ |A+A| \leqslant \lambda k\big\}\big| = \Theta_\lambda(1) \cdot \frac{n^2}{k}\binom{\lambda k/2}{k}. \tag{6.2}$$

The upper bound in Corollary 6.1.2 is an almost immediate consequence of Theorem 6.1.1, and our lower bound follows from a straightforward calculation (see Sections 6.9 and 6.10). For both bounds we obtain a constant of the form $\exp\big(\lambda^{\Theta(1)}\big)$ for λ large, and it would be interesting to determine the correct exponent of λ.

We remark that similar results can be deduced from our proof for all $2 < \lambda < k^{o(1)}$ (see Section 6.9), but the constant given by our method tends to infinity as $\lambda \to 2$. In order to keep the calculations as simple as possible, we have chosen to focus on the case $\lambda \geqslant 3$. Let us note here also that the bound $k \geqslant (\log n)^4$ can be improved somewhat (see Theorem 6.9.1); however, some polylogarithmic factor is necessary, since (as was observed in Chapter 5) the union of an arithmetic progression of length $k - \lambda + 2$ with $\lambda - 2$ arbitrary points satisfies $|A| = k$ and $|A + A| \leqslant \lambda k$, and there are at least $\Theta(n^\lambda)$ such sets, which is larger than (6.2) when $k = o(\log n)$. It seems plausible, however, that Theorem 6.1.1 and Corollary 6.1.2 could hold (for λ fixed) whenever $k/\log n \to \infty$.

In order to understand why Theorem 6.1.1 should be true, recall first that, by Freĭman's theorem, a set has bounded doubling if and only if it is a subset of positive density of a generalised arithmetic progression of bounded dimension. Now, there are $O(n^{d+1})$ generalised arithmetic progressions P of dimension d, and if A were a random subset of P of positive density, then $A + A$ would be unlikely to 'miss' many elements of $P + P$, which implies that (typically) $|A + A| \geqslant (d + 1 + o(1)) \cdot |P|$.[2] This suggests that the number of choices for A should be roughly $n^{d+1} \cdot \binom{\lambda k/(d+1)}{k}$, which (for $k \gg \log n$) is maximised by taking $d = 1$, and this leads to the intuition that most sets of bounded doubling should in fact be contained in an arithmetic progression of size roughly $|A + A|/2$. As explained in Chapter 5, this intuition was partially confirmed in previous works, which showed that there typically exists an arithmetic progression P of length $(1/2 + o(1))|A + A|$ such that $|A \setminus P| = o(|A|)$.

The main tool in the proof of Theorem 6.1.1 is Theorem 5.4.2 proved in Chapter 5. We will use this container theorem in three different ways: first, to control the rough structure of a

[2] The factor of $d + 1$ is attained by a union of d (unrelated) arithmetic progressions; for most d-dimensional generalised arithmetic progressions the doubling would be even larger.

set with bounded doubling (see Theorem 6.3.3 and Lemma 6.5.2); then to prove a variant of a probabilistic lemma of Green and Morris [76] (see Lemma 6.4.1); and finally to control the fine structure of the set near the ends of the progression containing it (see Section 6.8). We consider this last step to be the most interesting aspect of the proof, since we are not aware of any previous application of containers to the task of 'cleaning up' a set, that is, replacing a rough structural result with a precise one. We hope that our proof will inspire further applications of this type in other combinatorial settings.

6.2 An overview of the proof

In this section we will prepare the reader for the details of the proof by giving a rough outline of the main ideas. Let us fix $\lambda \geqslant 3$, and let $k \in \mathbb{N}$ be sufficiently large. We will mostly work with sets of integers that are 'close' to being a subset of the interval $[\lambda k/2]$, since Theorem 5.5.1 implies that almost all of the sets that we need to count are close to an arithmetic progression of length $\lambda k/2$, and any such progression can be mapped into $[\lambda k/2]$ (see Section 6.5 for the details).[3]

Given a set $A \subset \mathbb{Z}$, let us write

$$b(A) := |A \setminus [\lambda k/2]| \qquad \text{and} \qquad r(A) := \max(A) - \min(A) - \lambda k/2. \tag{6.3}$$

Let us also fix $\varepsilon > 0$ and set $\delta := 2^{-32}\lambda^{-3}$. By Lemma 6.5.1, below, the problem will reduce to bounding the size of the following family of sets.

Definition 6.2.1. *Let $\mathcal{I}$ denote the family of sets $A \subset \{-\lambda k/2, \dots, \lambda k\}$ with $|A| = k$ and $|A + A| \leqslant \lambda k$, such that*

$$b(A) \leqslant \delta k \qquad \text{and} \qquad r(A) \geqslant c(\lambda, \varepsilon),$$

and the sets $\big\{x \in A : x \leqslant 0\big\}$ and $\big\{x \in A : x > \lambda k/2\big\}$ are non-empty.

We will partition the family $\mathcal{I}$ according to the 'density' of the set $B := A \setminus [\lambda k/2]$. To be precise, set

$$f(\lambda) := 2^{10}\lambda^3, \tag{6.4}$$

and say that B is *sparse* if $r(A) > f(\lambda)b(A)$. The following lemma, which is proved in Section 6.6, bounds the number of sets $A \in \mathcal{I}$ such that B is sparse.

[3] For simplicity, we will assume throughout the paper that $\lambda k/2$ is an integer.

Lemma 6.2.2. *For every $\lambda \geqslant 3$ and $\varepsilon \in (0,1)$, and every $k \in \mathbb{N}$, we have*

$$\left|\left\{A \in \mathcal{I} : r(A) > f(\lambda)b(A)\right\}\right| \leqslant \frac{\varepsilon}{\lambda^3}\binom{\lambda k/2}{k}.$$

To bound the number of choices for A, we will bound separately the choices for B and $A' := A \setminus B$. Assume (for simplicity) that $\min(A) = 0$, so $\max(A) = \lambda k/2 + r$. The proof of Lemma 6.2.2 uses the following simple idea: the set $(A' + \max(A)) \setminus [\lambda k]$ typically contains about $2r/\lambda$ elements, and this restricts the size of the set $A' + A'$, and hence the number of choices for A'. More precisely, we will use a straightforward counting argument when $(A' + \max(A)) \setminus [\lambda k]$ is much smaller than r/λ (see Lemma 6.6.3), and an application of the container theorem when it is larger (see Lemma 6.6.4). Moreover, the assumption that B is sparse allows us to (trivially) bound the number of choices for B.

We remark that our application of the container theorem in the proof of Lemma 6.2.2 proceeds via a probabilistic lemma (Lemma 6.4.1), which is a generalisation of a result of Green and Morris [76]. This lemma gives a (close to tight) upper bound on the number of k-subsets of $[n]$ whose sumset missed many elements of $\{2, \dots, 2n\}$, and is proved in Section 6.4, using Theorem 5.4.2.

When $r(A) \leqslant f(\lambda)b(A)$, we will say that the set is *dense*. In Sections 6.7 and 6.8 we will prove the following lemma, which bounds the number of dense sets in $\mathcal{I}$.

Lemma 6.2.3. *For every $\lambda \geqslant 3$ and $\varepsilon \in (0,1)$, and every $k \in \mathbb{N}$, we have*

$$\left|\left\{A \in \mathcal{I} : r(A) \leqslant f(\lambda)b(A)\right\}\right| \leqslant \frac{\varepsilon}{\lambda^3}\binom{\lambda k/2}{k}.$$

The proof of Lemma 6.2.3 is significantly more difficult than that of Lemma 6.2.2, and is the most interesting and novel part of the argument, involving a surprising and unusual application of the container method. Set $A' := A \cap [\lambda k/2]$ and $B := A \setminus A'$, as before, and suppose that $|B| = b$ and $|(B + B) \setminus [\lambda k]| = \mu b$. The main difficulties arise when $r = O(\mu b)$ and $\mu = \Theta(\lambda)$, and we first take care of the remaining cases in Section 6.7.

For these 'easy' cases (see Lemmas 6.7.4 and 6.7.5) we use similar ideas to those used in the proof of Lemma 6.2.2 (see the sketch above), except that instead of using a trivial bound, we will need to apply the container theorem (via Theorem 6.3.2) in order to bound the number of choices for the set B (see Lemma 6.7.2), and the calculations are significantly more delicate. In particular, we will need to use our bounds on the size of both $(A' + \max(A)) \setminus [\lambda k]$ and $(B + B) \setminus [\lambda k]$ to bound the size of $A' + A'$, and thus the number of choices for A'.

Counting the sets with $r = O(\mu b)$ and $\mu = \Theta(\lambda)$ is the most interesting part of the proof. The key idea is to use the container theorem to obtain a collection of 'containers' (C, D) for the

'missing' set $M(A) := [\lambda k] \setminus (A + A)$, which is typically (see Lemma 6.8.2) contained in the set $Y + Y$, where Y is the set of points that are 'close' to the endpoints of $[\lambda k/2]$. The containers satisfy $M(A) \subset C$ and $A \cap Y \subset D$, and moreover D misses roughly $|C|/2$ points of Y (for the precise statement, see Corollary 6.8.1). The key step (Lemma 6.8.3) then uses these properties to bound the number of sets A corresponding to each container. Taking a union bound over containers, it follows that there are at most

$$\exp\left(-\frac{r}{2^{19}\lambda^2}\right)\binom{\lambda k/2}{k}$$

sets $A \in \mathcal{I}$ with $r(A) = r \leqslant f(\lambda)b(A)$, and this easily implies Lemma 6.2.3.

The rest of the paper is organised as follows. First, in Section 6.3, we recall the main results of Chapter 5, and deduce the container theorem we will use in the proof (Corollary 6.3.4). In Section 6.4 we use this container theorem to prove the probabilistic lemma mentioned above (Lemma 6.4.1), and in Section 6.5 we will use the results of Chapter 5 to reduce the problem to that of bounding the size of the set $\mathcal{I}$. In Section 6.6 we prove Lemma 6.2.2, in Sections 6.7 and 6.8 we prove Lemma 6.2.3, and in Section 6.9 we put the pieces together and prove Theorem 6.1.1. Finally, in Section 6.10, we provide two simple constructions that show that the upper bounds in Theorem 6.1.1 and Corollary 6.1.2 are not far from best possible.

6.3 The container theorem

In this section we will recall the main results of Chapter 5, which will play an important role in the proofs of our main theorems. We begin by restating the main container theorem from Chapter 5 in a slightly simpler form.

Theorem 6.3.1 (Theorem 5.4.2)**.** *Let $m \geqslant (\log n)^2$, let $Y \subset \mathbb{Z}$ with $|Y| = n$, and let $0 < \gamma < 1/4$. There is a family $\mathcal{A} \subset 2^{Y+Y} \times 2^Y$ of pairs of sets (A, B), of size*

$$|\mathcal{A}| \leqslant \exp\left(2^{16}\gamma^{-2}\sqrt{m}\,(\log n)^{3/2}\right), \tag{6.5}$$

such that:

(*i*) *For each $J \subset Y$ with $|J + J| \leqslant m$, there is $(A, B) \in \mathcal{A}$ with $A \subset J + J$ and $J \subset B$.*

(*ii*) *For every $(A, B) \in \mathcal{A}$, $|A| \leqslant m$ and either $|B| \leqslant \frac{m}{\log n}$ or there are at most $\gamma^2|B|^2$ pairs $(b_1, b_2) \in B \times B$ such that $b_1 + b_2 \notin A$.*

The reader may find it useful to imagine[4] the statement of Theorem 6.3.1 as saying that for each set $J \subset Y$, there exists a 'container' $(A, B) \in \mathcal{A}$ such that

$$J \subset B, \qquad B + B \approx A \qquad \text{and} \qquad A \subset J + J.$$

Moreover, and crucially, the number of containers is sub-exponential in m.

We will also use the following two consequences of Theorem 6.3.1, which were both proved in Chapter 5. The first determines the number of sets $A \subset [n]$ with $|A| = k$ and $|A + A| \leqslant \lambda k$ up to a factor of $2^{o(k)}$. We will use it in Section 6.7 to bound the number of choices for $A \setminus [\lambda k/2]$.

Theorem 6.3.2 (Theorem 5.4.1)**.** *Let $n, k \in \mathbb{N}$, and let $2 < \lambda < 2^{-36}\frac{k}{(\log n)^3}$. The number of sets $A \subset [n]$ with $|A| = k$ such that $|A + A| \leqslant \lambda k$ is at most*

$$\exp\left(2^9 \lambda^{1/6} k^{5/6} \log k \sqrt{\log n}\right) \binom{\lambda k/2}{k}.$$

The second gives structural information about a typical set with small doubling; we will use it in Section 6.5. The following is a slight generalisation of Theorem 5.5.1, but follows from the same proof; the details can be found in [32, Appendix A].

Theorem 6.3.3 (Theorem 5.5.1)**.** *Let $n, k \in \mathbb{N}$ and $2 \leqslant \lambda \leqslant 2^{-120}\frac{k}{(\log n)^3}$. Suppose that $2^8 \lambda^{1/6} k^{-1/6} \sqrt{\log n} \leqslant \gamma < 2^{-8}$. For all but at most*

$$e^{-\gamma k} \binom{\lambda k/2}{k}$$

sets $A \subset [n]$ with $|A| = k$ and $|A + A| \leqslant \lambda k$, the following holds: there exists $T \subset A$, with $|T| \leqslant 2^9 \gamma k$, such that $A \setminus T$ is contained in an arithmetic progression of size $\lambda k/2 + 2^7 \gamma \lambda k$.

The upper bounds on λ in Theorems 6.3.2 and 6.3.3 are the reason why we require the bound $k \geqslant (\log n)^4$ in Theorem 6.1.1 and Corollary 6.1.2. We remark that some log-factor is necessary here, since it was observed in Chapter 5 that the conclusions of the theorems fail to hold if $k = o\big(\lambda \log n\big)$.

We will apply Theorem 6.3.1 (in Sections 6.4 and 6.8) via the following corollary.

Corollary 6.3.4. *Let $0 < \gamma < 1/4$, let $S_1, S_2 \subset \mathbb{Z}$ be intervals, and set*

$$Y := S_1 \cup S_2 \qquad \text{and} \qquad X := (S_1 + S_1) \cup (S_2 + S_2). \tag{6.6}$$

Then there is a family $\mathcal{B} \subset 2^X \times 2^Y$ of size at most

$$\exp\left(2^{17} \gamma^{-2} \sqrt{|Y|} \left(\log |Y|\right)^{3/2}\right) \tag{6.7}$$

[4] This intuition will be sufficient in this chapter; however, in Chapter 7 we will need a more refined notion of $B + B \approx A$.

such that:

(a) *For every pair of sets $U \subset Y$ and $W \subset X \setminus (U+U)$, there exists $(C,D) \in \mathcal{B}$ such that $W \subset C$ and $U \subset D$.*

(b) *For every $(C,D) \in \mathcal{B}$,*

$$|D| \leqslant \max\left\{(1+4\gamma)|Y| - \frac{|C|}{2}, \frac{3|Y|}{\log|Y|}\right\}. \tag{6.8}$$

Note that replacing $W \subset X \setminus (U+U)$ by $W = X \setminus (U+U)$ in part (a) would give an equivalent statement; however, we will find this formulation convenient. To deduce Corollary 6.3.4 from Theorem 6.3.1, we will need the following easy lemma.

Lemma 6.3.5. *Let $\gamma > 0$, let $S_1, S_2 \subset \mathbb{Z}$ be intervals, and set*

$$Y := S_1 \cup S_2 \qquad \text{and} \qquad X := (S_1 + S_1) \cup (S_2 + S_2).$$

Let $C \subset X$ and $D \subset Y$. If

$$|D| \geqslant (1+4\gamma)|Y| - |C|/2$$

then there are at least $\gamma^2|D|^2$ pairs $(b_1, b_2) \in D \times D$ such that $b_1 + b_2 \in C$.

Proof. Suppose first that $S_1 \cap S_2$ is non-empty, so $X = Y + Y$, and let the elements of D be $d_1 < \cdots < d_\ell$. Then $D + D \subset X$ contains the $2\ell - 1$ elements

$$d_1 + d_1 < d_1 + d_2 < \cdots < d_1 + d_\ell < d_2 + d_\ell < \cdots < d_\ell + d_\ell,$$

and $2\ell - 1 \geqslant (2+8\gamma)|Y| - |C| - 1 = |X| - |C| + 8\gamma|Y|$, since $|X| = 2|Y| - 1$. Since $C \subset X$, it follows that there are at least $8\gamma|Y|$ pairs $(b_1, b_2) \in D \times D$ such that $b_1 + b_2 \in C$ and $\{b_1, b_2\} \cap \{d_1, d_\ell\}$ is non-empty. Removing d_1 and d_ℓ from D, and repeating the argument $\gamma|Y|$ times, we obtain $\gamma^2|Y|^2$ pairs $(b_1, b_2) \in D \times D$ such that $b_1 + b_2 \in C$.

When S_1 and S_2 are disjoint, we simply apply the argument above for the two sets $D_1 := D \cap S_1$ and $D_2 := D \cap S_2$. To spell out the details, for each $i \in \{1, 2\}$ there are $2|D_i| - 1$ pairs $(b_1, b_2) \in D_i \times D_i$ with distinct sums such that either $b_1 = \min(D_i)$ or $b_2 = \max(D_i)$. Moreover, $D_1 + D_1$ and $D_2 + D_2$ are disjoint subsets of X, and

$$2|D| - 2 \geqslant (2+8\gamma)|Y| - |C| - 2 = |X| - |C| + 8\gamma|Y|,$$

since $|X| = 2|Y| - 2$. As before, it follows that there are at least $8\gamma|Y|$ pairs $(b_1, b_2) \in D \times D$ such that $b_1 + b_2 \in C$ and either $b_1 \in \{\min(D_1), \min(D_2)\}$ or $b_2 \in \{\max(D_1), \max(D_2)\}$. Removing the minimum and maximum elements of D_1 and D_2, and repeating the argument $\gamma|Y|$ times, we obtain $\gamma^2|Y|^2$ pairs $(b_1, b_2) \in D \times D$ such that $b_1 + b_2 \in C$, as claimed. □

Proof of Corollary 6.3.4. Applying Theorem 6.3.1 with $n := |Y|$ and $m := 3|Y|$, we obtain a family $\mathcal{A} \subset 2^{Y+Y} \times 2^Y$, with

$$|\mathcal{A}| \leqslant \exp\Big(2^{17}\gamma^{-2}\sqrt{|Y|}\big(\log|Y|\big)^{3/2}\Big),$$

satisfying properties (i) and (ii) of the theorem. We claim that

$$\mathcal{B} := \big\{(X \setminus A, B) : (A, B) \in \mathcal{A}\big\} \subset 2^X \times 2^Y$$

satisfies properties (a) and (b) of Corollary 6.3.4.

To show that property (a) holds, let $U \subset Y$ and $W \subset X \setminus (U + U)$, and set $J := U$. Noting that $J \subset Y$, and that

$$|J + J| \leqslant |Y + Y| \leqslant 3|Y| = m,$$

it follows from Theorem 6.3.1(i) that there exists $(A, B) \in \mathcal{A}$ with $A \subset J + J$ and $J \subset B$, and hence there exists $(C, D) = (X \setminus A, B) \in \mathcal{B}$ such that $W \subset C$ and $U \subset D$.

For property (b), let $(C, D) \in \mathcal{B}$, and observe that, by Theorem 6.3.1(ii), either $|D| \leqslant \frac{3|Y|}{\log|Y|}$, or there are at most $\gamma^2|D|^2$ pairs $(b_1, b_2) \in D \times D$ such that $b_1 + b_2 \in C$. In the latter case, we have $|D| \leqslant (1 + 4\gamma)|Y| - |C|/2$, by Lemma 6.3.5. Since $|\mathcal{B}| \leqslant |\mathcal{A}|$, the corollary follows. □

6.4 A probabilistic lemma

Green and Morris [76, Theorem 1.3] used their bounds on the number of sets with small sumset to prove that if S is a random subset of $\mathbb{N}$, with each element included in S independently with probability $1/2$, then

$$\mathbb{P}\Big(\big|\mathbb{N} \setminus (S + S)\big| \geqslant m\Big) = 2^{-m/2+o(m)}.$$

We will use Corollary 6.3.4 to prove the following generalisation of their theorem. We remark that a similar result (with a slightly larger error term) for larger values of k can be deduced from exactly the same proof.

Lemma 6.4.1. *Let $n, k \in \mathbb{N}$, with $k \leqslant 2n/3$, and set $p := k/n$. If S is a uniformly-chosen random subset of $[n]$ of size k, then*

$$\mathbb{P}\Big(\big|\{2, \dots, 2n\} \setminus (S + S)\big| \geqslant m\Big) \leqslant \exp\big(2^{16}m^{7/8}\big) \cdot (1 - p)^{m/2}. \tag{6.9}$$

In the proof of Lemma 6.4.1 we will also use the following well-known inequality (see, e.g., [3, Lemma 5.2]).

Lemma 6.4.2 (Pittel's inequality)**.** *Let $n, k \in \mathbb{N}$ with $k \leqslant n$, and set $p := k/n$. If $\mathcal{I}$ is a monotone decreasing property on $[n]$, then*

$$\mathbb{P}\big(\mathcal{I} \text{ holds for a random } k\text{-subset of } [n]\big) \leqslant 2 \cdot \mathbb{P}\big(\mathcal{I} \text{ holds for a } p\text{-random subset of } [n]\big).$$

Proof. Following the proof in [3], recall that $\mathrm{Bin}(n,p) \leqslant \lceil pn \rceil = k$ holds with probability at least $1/2$. Since $\mathcal{I}$ is monotone decreasing, the claimed bound follows. □

We first prove a simple lemma that will also be useful in Section 6.8.

Lemma 6.4.3. *Let $n \in \mathbb{N}$ and $k \in [n]$, set $p := k/n$, and let $M \in \mathbb{N}$. If S is a uniformly-chosen random subset of $[n]$ of size k, then*

$$\mathbb{P}\Big(\big\{M+1, \dots, 2n-M+1\big\} \not\subset S+S\Big) \leqslant \frac{8}{p^2} \cdot \big(1-p^2\big)^{M/2}.$$

Proof. Observe that the left-hand side is at most

$$\sum_{x=M+1}^{2n-M+1} \mathbb{P}\big(x \notin S+S\big) \leqslant 2 \sum_{x=M+1}^{n+1} \mathbb{P}\big(x \notin S+S\big),$$

since, by symmetry, $\mathbb{P}\big(x \notin S+S\big) = \mathbb{P}\big(2n+2-x \notin S+S\big)$. Now, for $x \leqslant n+1$, we can use Pittel's inequality to bound

$$\mathbb{P}\big(x \notin S+S\big) = \mathbb{P}\Big(\bigcap_{i=1}^{\lfloor x/2 \rfloor} \big(\{i \notin S\} \cup \{x-i \notin S\}\big)\Big) \leqslant 2\big(1-p^2\big)^{(x-1)/2}.$$

It follows that

$$\mathbb{P}\Big(\big\{M+1, \dots, 2n-M+1\big\} \not\subset S+S\Big) \leqslant 4 \sum_{x=M+1}^{\infty} \big(1-p^2\big)^{(x-1)/2} \leqslant \frac{8}{p^2}\big(1-p^2\big)^{M/2},$$

as claimed. □

We are now ready to deduce Lemma 6.4.1 from Corollary 6.3.4.

Proof of Lemma 6.4.1. Observe first that, since $1-p \geqslant e^{-2p}$ for $0 \leqslant p \leqslant 2/3$, the claimed bound holds trivially if $pm \leqslant 2^{16} m^{7/8}$. We may therefore assume that $m \geqslant 2^{128} p^{-8}$.

We will use Lemma 6.4.3 to deal with the case that the 'middle' is not covered by $S+S$. To be precise, set $M := \lceil 4m/p \rceil$ and let us write $\mathcal{E}$ for the event that $\big\{2M+1, \dots, 2n-2M+1\big\} \subset S+S$. Note that if $\mathcal{E}$ holds, then

$$\{2, \dots, 2n\} \setminus (S+S) \subset X := \big\{2, \dots, 2M\big\} \cup \big\{2n-2M+2, \dots, 2n\big\}.$$

Setting $W := X \setminus (S+S)$, it follows that

$$\mathbb{P}\Big(\big|\{2,\dots,2n\} \setminus (S+S)\big| \geqslant m\Big) \leqslant \mathbb{P}\big(|W| \geqslant m\big) + \mathbb{P}(\mathcal{E}^c).$$

By Lemma 6.4.3, we have[5]

$$\mathbb{P}(\mathcal{E}^c) \leqslant \frac{8}{p^2}\big(1-p^2\big)^M \leqslant \frac{8}{p^2}\big(1-p\big)^m,$$

where the second inequality follows since $1-x^2 \leqslant (1-x)^{x/2}$ for all $0 \leqslant x \leqslant 1$.

To complete the proof, we will use Corollary 6.3.4 to bound the probability that $|W| \geqslant m$. Indeed, applying the corollary to the set

$$Y := \{1,\dots,M\} \cup \{n-M+1,\dots,n\},$$

and noting that the set X defined above is the same as that defined in (6.6), we obtain a family $\mathcal{B} \subset 2^X \times 2^Y$ of containers of size at most

$$\exp\left(2^{18}\gamma^{-2}\sqrt{M}(\log M)^{3/2}\right) = \big(1-p\big)^{-\gamma M}, \tag{6.10}$$

where $\gamma > 0$ is chosen so that the equality holds. In particular, if $(C,D) \in \mathcal{B}$, then

$$|D| \leqslant \max\left\{(1+4\gamma)|Y| - \frac{|C|}{2}, \frac{3|Y|}{\log|Y|}\right\}, \tag{6.11}$$

and if $U \subset Y$ and $W \subset X \setminus (U+U)$, then there exists $(C,D) \in \mathcal{B}$ with $W \subset C$ and $U \subset D$.

To apply Corollary 6.3.4, we need to check that $\gamma < 1/4$. Using the bounds $1-p \leqslant e^{-p}$ and $M \geqslant m/p$, and noting that the function $x \mapsto (\log x)^{3/2}/\sqrt{x}$ is decreasing for $x > 2^5$, it follows from (6.10) that

$$\gamma^3 \leqslant \frac{2^{18}(\log M)^{3/2}}{p\sqrt{M}} \leqslant \frac{2^{18}}{\sqrt{pm}}\left(\log\frac{m}{p}\right)^{3/2}.$$

Therefore, since $M \leqslant 8m/p$, we have

$$\gamma M \leqslant \frac{8\gamma m}{p} \leqslant \frac{2^9 m^{5/6}}{p^{7/6}}\left(\log\frac{m}{p}\right)^{1/2} < m, \tag{6.12}$$

where the final inequality follows from the assumption that $m \geqslant 2^{128}p^{-8}$. Since $M \geqslant 4m$, it follows from (6.12) that $\gamma < 1/4$, and so this is a valid choice of γ in Corollary 6.3.4.

We next claim that

$$\mathbb{P}\big(|W| \geqslant m\big) \leqslant \sum_{(C,D)\in\mathcal{B}} \mathbb{P}\Big(\big(W \subset C\big) \cap \big(S \cap Y \subset D\big)\Big). \tag{6.13}$$

[5]Note that if $m \geqslant pn/4$, then $M \geqslant n$, and so the event $\mathcal{E}$ holds trivially.

To see this, observe first that

$$W = X \setminus (S + S) \subset X \setminus \big((S \cap Y) + (S \cap Y)\big)$$

since $S \cap Y \subset S$. By the property of $\mathcal{B}$ guaranteed by Corollary 6.3.4(a), applied with $U := S \cap Y$, it follows that there exists a pair $(C, D) \in \mathcal{B}$ with $W \subset C$ and $S \cap Y \subset D$.

To bound the right-hand side of (6.13), observe first that

$$\mathbb{P}\big(S \cap Y \subset D\big) \leqslant \binom{n - |Y \setminus D|}{pn} \binom{n}{pn}^{-1} \tag{6.14}$$

for every $(C, D) \in \mathcal{B}$, since S is a uniformly-chosen set of size $k = pn$, and if $S \cap Y \subset D$ then $S \cap (Y \setminus D) = \emptyset$. Moreover, by (6.11), if $|W| \geqslant m$ then

$$|Y \setminus D| \geqslant |Y| - |D| \geqslant \frac{m}{2} - 8\gamma M \tag{6.15}$$

for every $(C, D) \in \mathcal{B}$ with $W \subset C$. It follows from (6.10), (6.13), (6.14) and (6.15) that

$$\mathbb{P}\big(|W| \geqslant m\big) \;\leqslant\; \big(1 - p\big)^{-\gamma M} \binom{n - m/2 + 8\gamma M}{pn} \binom{n}{pn}^{-1} \;\leqslant\; \big(1 - p\big)^{m/2 - 9\gamma M}, \tag{6.16}$$

where the second inequality follows from the standard binomial inequality

$$\binom{a - c}{b} \leqslant \left(\frac{a - b}{a}\right)^{c} \binom{a}{b}. \tag{6.17}$$

Combining (6.12) and (6.16), and noting that $1 - p \geqslant e^{-2p}$ for $0 \leqslant p \leqslant 2/3$, it follows that

$$\mathbb{P}\big(|W| \geqslant m\big) \;\leqslant\; \exp\Big(2^{14} m^{5/6} p^{-1/6} (\log m)^{1/2}\Big) \cdot \big(1 - p\big)^{m/2}.$$

Since $p^{-1/6} (\log m)^{1/2} \leqslant m^{1/24}$, by our lower bound on m, the claimed bound follows. □

We will usually apply Lemma 6.4.1 in the following form. Recall that $\delta = 2^{-32} \lambda^{-3}$.

Corollary 6.4.4. *Let $\lambda \geqslant 3$ and $k, m, b \in \mathbb{N}$, with $m \geqslant 2^{400} \lambda^{24}$ and $b \leqslant \delta k$. There are at most*

$$e^{2\delta m} \left(\frac{\lambda - 2}{\lambda}\right)^{m/2} \binom{\lambda k/2}{k - b}$$

sets $A' \subset [\lambda k/2]$ of size $k - b$ such that $\big|[\lambda k] \setminus (A' + A')\big| \geqslant m$.

Proof. We simply apply Lemma 6.4.1 with $p = 2(k - b)/\lambda k \leqslant 2/3$, and observe that

$$\exp\big(2^{16} m^{7/8}\big) \big(1 - p\big)^{m/2} \leqslant e^{2\delta m} \left(\frac{\lambda - 2}{\lambda}\right)^{m/2},$$

by our bounds on b and m. To spell out the details, note that

$$2^{16}m^{7/8} \leqslant \delta m,$$

since $\delta = 2^{-32}\lambda^{-3}$ and $m \geqslant 2^{400}\lambda^{24}$. Now, observe that

$$(1-p)^{m/2} \leqslant \left(\frac{\lambda - 2 + 2\delta}{\lambda}\right)^{m/2} \leqslant \exp\left(\frac{\delta m}{\lambda - 2}\right)\left(\frac{\lambda - 2}{\lambda}\right)^{m/2}.$$

Since $\lambda \geqslant 3$, the claimed bound follows. □

Since we will often only need a weaker bound, let us note here, for convenience, that

$$e^{2\delta m}\left(\frac{\lambda - 2}{\lambda}\right)^{m/2} \leqslant \left(\frac{\lambda - 1}{\lambda}\right)^{m/2}, \tag{6.18}$$

since $\delta = 2^{-32}\lambda^{-3}$.

6.4.1 Tools and inequalities

To finish this section, let us state some standard tools that we will use in the proof of Theorem 6.1.1. The first is known as Ruzsa's covering lemma (see, e.g., [151, Lemma 2.14]), and was first proved in [137]. For completeness, we give the proof.

Lemma 6.4.5 (Ruzsa's covering lemma)**.** *Let $A, B \subset \mathbb{Z}$ be non-empty sets of integers, and suppose that $|A + B| \leqslant \mu|A|$. Then there exists a set $X \subset B$ with $|X| \leqslant \mu$ such that*

$$B \subset A - A + X.$$

Proof. Let $X \subset B$ be maximal such that the sets $A + x$ for $x \in X$ are disjoint. Observe that $|A + B| \geqslant |A||X|$, and therefore $|X| \leqslant \mu$. Now, since X is maximal, $A + b$ intersects $A + X$ for every $b \in B \setminus X$, and hence $B \subset A - A + X$, as claimed. □

We will also use the following special case of the Plünnecke–Ruzsa inequalities [121, 122, 136], which is also an immediate consequence of Ruzsa's triangle inequality [135].

Lemma 6.4.6 (Plünnecke–Ruzsa inequality)**.** *If $|A + A| \leqslant \lambda|A|$, then $|A - A| \leqslant \lambda^2|A|$.*

Proof. To prove that $|A - A| \cdot |A| \leqslant |A + A|^2$, it suffices to construct an injective map $\varphi\colon (A - A) \times A \to (A + A)^2$. To do so, choose an arbitrary function $f\colon A - A \to A^2$ such that if $f(x) = (a, b)$ then $a - b = x$, and define $\varphi(x, c) \mapsto (a + c, b + c)$, where $(a, b) = f(x)$. To see that φ is injective, observe that $x = (a + c) - (b + c)$ and that $(a, b) = f(x)$. □

In Section 6.7 we will use a simple special case of the following result of Freĭman [68].

Lemma 6.4.7 (Freĭman's $3k-4$ theorem)**.** *If $|A+A| \leqslant 3|A|-4$, then $A \subset P$ for some arithmetic progression P of size $|A+A|-|A|+1$.*

We will also make frequent use of the following standard inequality in the calculations below:

$$\binom{a-c}{b-d} \leqslant \left(\frac{a-c}{a}\right)^{b-d} \left(\frac{b}{a-b}\right)^{d} \binom{a}{b}. \tag{6.19}$$

In particular, note that

$$\binom{\lambda k/2}{k-b} \leqslant \left(\frac{2}{\lambda-2}\right)^{b} \binom{\lambda k/2}{k}. \tag{6.20}$$

We will also use the following inequality once, in Section 6.7.

Observation 6.4.8.

$$\binom{ca}{a} \leqslant \left(\frac{c^c}{(c-1)^{c-1}}\right)^a,$$

for every $a \in \mathbb{N}$ and $1 < c \in \mathbb{R}$.

Proof. Set $y = (c-1)^{1/c}$, and note that $y/(c-1) = y^{1-c}$. It follows that

$$\begin{aligned}\left(\frac{c^c}{(c-1)^{c-1}}\right)^a &= \left(\left(1+\frac{1}{c-1}\right)(c-1)^{1/c}\right)^{ca} \\ &= \left(y+y^{1-c}\right)^{ca} = \sum_{i=0}^{ca} \binom{ca}{i} y^{ca-i} \cdot y^{(1-c)i} \geqslant \binom{ca}{a},\end{aligned}$$

where the last step follows by considering the term $i = a$. □

6.5 Reducing to an interval

Let us fix $\lambda \geqslant 3$, and for each $n, k \in \mathbb{N}$ define

$$\Lambda = \Lambda(n,k) := \big\{A \subset [n] \,:\, |A| = k \text{ and } |A+A| \leqslant \lambda k\big\}. \tag{6.21}$$

Let us also fix $\varepsilon \in (0,1)$ (since Theorem 6.1.1 holds trivially for $\varepsilon \geqslant 1$) and, writing $\ell(A)$ for the length of the smallest arithmetic progression containing A, define

$$\Lambda^* = \Lambda^*(n,k) := \big\{A \in \Lambda : \ell(A) \leqslant \lambda k/2 + c(\lambda,\varepsilon)\big\}. \tag{6.22}$$

In this section we will prove the following lemma, which reduces the problem of bounding $|\Lambda \setminus \Lambda^*|$ to that of bounding $|\mathcal{I}|$ (see Definition 6.2.1). Recall that $\delta = 2^{-32}\lambda^{-3}$.

Lemma 6.5.1. *Let $\lambda \geqslant 3$ and $n, k \in \mathbb{N}$, with $k \geqslant 2^{400}\lambda^{25}(\log n)^3$. We have*

$$|\Lambda \setminus \Lambda^*| \leqslant \frac{n^2}{k} \cdot |\mathcal{I}| + \exp\left(-\frac{\delta k}{2^{10}\lambda}\right)\binom{\lambda k/2}{k}.$$

To prove Lemma 6.5.1, we will successively refine $\Lambda \setminus \Lambda^*$, at each step showing that some subset with a particular property is small. The first step in the proof of Lemma 6.5.1 is the following stability lemma, which is an almost immediate consequence of Theorem 6.3.3.

Lemma 6.5.2. *Let $\lambda \geqslant 3$ and $n, k \in \mathbb{N}$, with $k \geqslant 2^{400}\lambda^{25}(\log n)^3$. There are at most*

$$\exp\left(-\frac{\delta k}{2^9\lambda}\right)\binom{\lambda k/2}{k}$$

sets $A \in \Lambda$ such that

$$|A \setminus P| \geqslant \delta k$$

for every arithmetic progression P of size $\lambda k/2$.

Proof. Set $\gamma = 2^{-9}\lambda^{-1}\delta = 2^{-41}\lambda^{-4}$, and observe that, since $k \geqslant 2^{400}\lambda^{25}(\log n)^3$, we have

$$2^8\lambda^{1/6}k^{-1/6}\sqrt{\log n} \leqslant \gamma < 2^{-8}.$$

Therefore, by Theorem 6.3.3, for all but at most

$$\exp\left(-\frac{\delta k}{2^9\lambda}\right)\binom{\lambda k/2}{k}$$

sets $A \in \Lambda$, there exists $T \subset A$, with $|T| \leqslant (2^9 + 2^7\lambda)\gamma k < \delta k$ (moving some elements of the progression given by the theorem into T), such that $A \setminus T$ is contained in an arithmetic progression of size $\lambda k/2$, as required. □

The next step is to show that almost all sets $A \in \Lambda$ are contained in an arithmetic progression of length $3\lambda k/2$. Let us write $\mathcal{F}$ for the family of sets $A \in \Lambda$ such that

$$A \subset \{a + jd : -\lambda k/2 \leqslant j \leqslant \lambda k\} \qquad \text{and} \qquad \big|A \setminus \{a + jd : 1 \leqslant j \leqslant \lambda k/2\}\big| \leqslant \delta k$$

for some $a, d \in \mathbb{Z}$. Recall that we assume, for simplicity, that $\lambda k/2$ is an integer.

Lemma 6.5.3. *Let $\lambda \geqslant 3$ and $n, k \in \mathbb{N}$, with $k \geqslant 2^{400}\lambda^{25}(\log n)^3$. Then*

$$|\Lambda \setminus \mathcal{F}| \leqslant \exp\left(-\frac{\delta k}{2^{10}\lambda}\right)\binom{\lambda k/2}{k}.$$

Proof. Fix an arithmetic progression $P = \{a + jd : 1 \leqslant j \leqslant \lambda k/2\}$. We will bound the number of sets $A \in \Lambda \setminus \mathcal{F}$ with $|A \setminus P| \leqslant \delta k$, and then sum over choices of P. We will then use Lemma 6.5.2 to count the remaining sets, and hence prove the lemma.

Note first that if $A \in \Lambda \setminus \mathcal{F}$ and $|A \setminus P| \leqslant \delta k$, then

$$A \not\subset P + P - P,$$

so let $Z := A \setminus (P + P - P)$ and choose an element $x \in Z$. We will first count the possible sets $A' := A \cap P$, and then (given A') the choices for $B := A \setminus P$. Observe that

$$\big(x + A'\big) \cap \big(A' + A'\big) = \emptyset,$$

since $A' \subset P$, and that if $|A \setminus P| \leqslant \delta k$, then $|x + A'| = |A'| \geqslant k - \delta k$. Since $A \in \Lambda$, it follows that

$$|A' + A'| \leqslant \lambda k - \big(k - \delta k\big) \leqslant \lambda k - k/2.$$

Hence, by Corollary 6.4.4 (applied with $m = k/2 \geqslant 2^{400}\lambda^{24}$), and using (6.18) and (6.20), it follows that, for each $b \leqslant \delta k$, there are at most

$$\left(\frac{\lambda - 1}{\lambda}\right)^{k/4} \binom{\lambda k/2}{k - b} \leqslant \exp\left(-\frac{k}{8\lambda}\right)\binom{\lambda k/2}{k}$$

choices for the set $A' = A \cap P$ such that $|A'| = k - b$.

To count the sets B (given A'), we apply Ruzsa's covering lemma (Lemma 6.4.5) to the pair (A', B) to obtain a set $X \subset B$, with $|X| \leqslant |A' + B|/|A'| \leqslant \lambda k/(k - b) \leqslant 2\lambda$, such that $B \subset A' - A' + X$. Moreover, by the Plünnecke–Ruzsa inequality (Lemma 6.4.6),

$$|A' - A' + X| \leqslant |X| \cdot |A' - A'| \leqslant 2\lambda^3 k.$$

Hence, choosing X first and then $B \setminus X$, and recalling that $b \leqslant \delta k = 2^{-32}\lambda^{-3}k$, and that $k \geqslant 2^{400}\lambda^{25}(\log n)^3$, it follows that there are at most

$$n^{2\lambda}\binom{2\lambda^3 k}{b - 2\lambda} \leqslant \exp\Big(\delta k \log\big(2e\lambda^3/\delta\big) + 2\lambda \log n\Big) \leqslant \exp\big(\delta^{1/2}k\big)$$

choices for the set B, given a set A' with $|A'| = k - b$.

Combining the bounds above on the number of choices for A' and B, it follows that the number of sets $A \in \Lambda$ with Z non-empty is at most

$$\sum_{b=1}^{\delta k} \exp\left(\delta^{1/2}k - \frac{k}{8\lambda}\right)\binom{\lambda k/2}{k} \leqslant \exp\left(-\frac{k}{2^4\lambda}\right)\binom{\lambda k/2}{k},$$

Summing over choices of P, and using Lemma 6.5.2 to bound the number of sets such that $|A \setminus P'| \geqslant \delta k$ for every arithmetic progression P' of size $\lambda k/2$, the lemma follows. □

Finally, to bound $|\Lambda \setminus \Lambda^*|$ in terms of $|\mathcal{I}|$, we need to map our arithmetic progression P into the interval $[\lambda k/2]$. Lemma 6.5.1 will follow from Lemma 6.5.3 and the following bound.

Lemma 6.5.4. *Let $\lambda \geqslant 3$ and $n, k \in \mathbb{N}$. Then*

$$|\mathcal{F} \setminus \Lambda^*| \leqslant \frac{n^2}{k} \cdot |\mathcal{I}|.$$

Proof. We will define a function $\varphi\colon \mathcal{F} \setminus \Lambda^* \to \mathcal{I}$ such that $|\varphi^{-1}(S)| \leqslant n^2/k$ for every $S \in \mathcal{I}$, which will suffice to prove the lemma. To do so, let $A \in \mathcal{F} \setminus \Lambda^*$, and choose $a, d \in \mathbb{N}$ such that

$$A \subset \big\{a + jd : -\lambda k/2 \leqslant j \leqslant \lambda k\big\}$$

and such that the sets

$$\big\{x \in A : x \leqslant a\big\} \qquad \text{and} \qquad \big\{x \in A : x > a + \lambda k d/2\big\} \tag{6.23}$$

are both non-empty and together contain at most δk elements. Indeed, to obtain such a pair, take the arithmetic progression given by the definition of $\mathcal{F}$, and (recalling the definition (6.22) of Λ^*) translate it if necessary so that the sets in (6.23) are both non-empty. Now define

$$\varphi(A) := \big\{j \in \mathbb{Z} \,:\, a + jd \in A\big\},$$

and observe that $\varphi(A) \subset \{-\lambda k/2, \dots, \lambda k\}$, and that

$$b\big(\varphi(A)\big) = \big|\big\{x \in \varphi(A) : x \leqslant 0\big\}\big| + \big|\big\{x \in \varphi(A) : x > \lambda k/2\big\}\big| \leqslant \delta k.$$

Moreover, we have

$$r\big(\varphi(A)\big) = \max\big(\varphi(A)\big) - \min\big(\varphi(A)\big) - \frac{\lambda k}{2} > c(\lambda, \varepsilon),$$

since $A \notin \Lambda^*$, and hence $\varphi(A) \in \mathcal{I}$, as required.

Finally, observe that $|\varphi^{-1}(S)|$ is bounded from above by the number of pairs $(a, d) \in \mathbb{Z}^2$ such that $A := \{a + jd : j \in S\} \subset [n]$. For each set S of size k there are at most

$$\sum_{a=1}^{n} \frac{n-a}{k-1} \leqslant \frac{n^2}{k}$$

such pairs (a, d). Hence $|\varphi^{-1}(S)| \leqslant n^2/k$, as claimed, and the lemma follows. □

We are now ready to prove Lemma 6.5.1.

Proof of Lemma 6.5.1. By Lemmas 6.5.3 and 6.5.4, we have

$$|\Lambda \setminus \Lambda^*| \leqslant |\Lambda \setminus \mathcal{F}| + |\mathcal{F} \setminus \Lambda^*| \leqslant \exp\left(-\frac{\delta k}{2^{10}\lambda}\right)\binom{\lambda k/2}{k} + \frac{n^2}{k} \cdot |\mathcal{I}|,$$

as claimed. □

6.6 Counting the sparse sets in $\mathcal{I}$

Recall that, for any $A \subset \mathbb{Z}$,

$$b(A) = |A \setminus [\lambda k/2]| \qquad \text{and} \qquad r(A) = \max(A) - \min(A) - \lambda k/2,$$

and that $f(\lambda) = 2^{10}\lambda^3$, and (recalling Definition 6.2.1) let us write

$$\mathcal{S} := \Big\{ A \in \mathcal{I} : r(A) > f(\lambda) b(A) \Big\}$$

for the family of 'sparse' sets in $\mathcal{I}$. In this section we will bound the size of $\mathcal{S}$, and hence prove the following quantitative version of Lemma 6.2.2.

Lemma 6.6.1. *Let $\lambda \geqslant 3$ and $\varepsilon \in (0,1)$, and let $k \in \mathbb{N}$. Then*

$$|\mathcal{S}| \leqslant \exp\left(- \frac{c(\lambda,\varepsilon)}{2^9\lambda^2} \right) \binom{\lambda k/2}{k}.$$

For each $B \subset \{-\lambda k/2, \dots, \lambda k\} \setminus [\lambda k/2]$, let us define[6]

$$\mathcal{G}(B) := \big\{ A \in \mathcal{I} : A \setminus [\lambda k/2] = B \big\}. \tag{6.24}$$

Recalling Definition 6.2.1, observe that $\mathcal{G}(B) = \emptyset$ if either $\min(B) > 0$ or $\max(B) \leqslant \lambda k/2$, and also if either $|B| > \delta k$ or $r(B) < c(\lambda,\varepsilon)$ (note that $r(A) = r(B)$ for every $A \in \mathcal{G}(B)$).

We will deduce Lemma 6.6.1 from the following bound on the size of $\mathcal{G}(B)$ by summing over $r \geqslant c(\lambda,\varepsilon)$ and sets B with $|B| < r/f(\lambda)$.

Lemma 6.6.2. *If $B \subset \{-\lambda k/2, \dots, \lambda k\} \setminus [\lambda k/2]$, then*

$$|\mathcal{G}(B)| \leqslant \exp\left(- \frac{r}{2^6\lambda^2} \right) \binom{\lambda k/2}{k-b}$$

where $b = |B|$ and $r = r(B)$.

For each $A \in \mathcal{G}(B)$, set $A' := A \setminus B$. The idea of the proof is simple: if A' contains many elements close to its ends, then we can add these to $\min(B)$ and $\max(B)$, and obtain many elements of $A+A$ outside $[\lambda k]$. Therefore, either $A'+A'$ misses many elements of $[\lambda k]$, in which case we can apply Corollary 6.4.4 to bound the number of choices, or it has few elements close to its ends, and it is straightforward to count sets A' with this property.

[6] Note that we include sets of $\mathcal{I} \setminus \mathcal{S}$ in $\mathcal{G}(B)$; we will not need to use the bound $r(A) > f(\lambda)b(A)$ when bounding the size of $\mathcal{G}(B)$ (we use it only when counting the choices for the set B), and we shall also want to reuse our bounds on $|\mathcal{G}(B)|$ in Section 6.7, below, where we will be dealing with dense sets.

To be precise, define

$$Y := \big\{x \leqslant 0 : x - \min(B) \in A'\big\} \cup \big\{x > \lambda k : x - \max(B) \in A'\big\}, \tag{6.25}$$

and set $m(B) := r(B)/8\lambda$. The following bound follows from some simple counting.

Lemma 6.6.3. *If $B \subset \{-\lambda k/2, \dots, \lambda k\} \setminus [\lambda k/2]$, then there are at most*

$$e^{-m(B)}\binom{\lambda k/2}{k-b}$$

sets $A \in \mathcal{G}(B)$ with $|Y| \leqslant m(B)$.

Proof. We claim first that if $r := r(B) \geqslant \lambda k/2$, then there are no such sets $A \in \mathcal{G}(B)$. Indeed, if $\max(B) - \min(B) \geqslant \lambda k$ then for each $y \in A'$ either $y + \min(B) \leqslant 0$, or $y + \max(B) > \lambda k$, and therefore $|Y| \geqslant |A'|$. It follows that if $A \in \mathcal{G}(B)$ with $|Y| \leqslant m := m(B)$, then $m \geqslant |Y| \geqslant |A'| = k - b \geqslant k/4$, since $b(A) \leqslant \delta k$ for every $A \in \mathcal{I}$. But this implies that $r = 8\lambda m > \lambda k$, which is impossible. Let us therefore assume that $r < \lambda k/2$.

Now, the number of sets $A \in \mathcal{G}(B)$ with $|Y| \leqslant m$ is at most

$$\sum_{\ell=0}^{m}\binom{r}{\ell}\binom{\lambda k/2 - r}{k-b-\ell} \leqslant \sum_{\ell=0}^{m}\left(\frac{er}{\ell}\right)^{\ell}\left(1 - \frac{2r}{\lambda k}\right)^{k-b-\ell}\left(\frac{2}{\lambda - 2}\right)^{\ell}\binom{\lambda k/2}{k-b}, \tag{6.26}$$

where the inequality holds by (6.19). Now, observe that

$$\left(1 - \frac{2r}{\lambda k}\right)^{k-b-\ell} \leqslant \left(1 - \frac{2r}{\lambda k}\right)^{k/2} \leqslant \exp\left(-\frac{r}{\lambda}\right) = e^{-8m},$$

since $b + \ell \leqslant k/2$ and $r = 8\lambda m$, and that

$$\sum_{\ell=0}^{m}\left(\frac{er}{\ell} \cdot \frac{2}{\lambda - 2}\right)^{\ell} \leqslant \sum_{\ell=0}^{m}\left(\frac{2^4 e\lambda}{\lambda - 2} \cdot \frac{m}{\ell}\right)^{\ell} \leqslant (m+1)\left(\frac{2^4 e\lambda}{\lambda - 2}\right)^{m} \leqslant \big(2^7 e\big)^m,$$

since $r = 8\lambda m$ and $\lambda \geqslant 3$, and since $(C/x)^x$ is increasing for $x < C/e$. It follows that the right-hand side of (6.26) (and hence the number of sets $A \in \mathcal{G}(B)$ with $|Y| \leqslant m$) is at most

$$\left(\frac{2}{e}\right)^{7m}\binom{\lambda k/2}{k-b} \leqslant e^{-m}\binom{\lambda k/2}{k-b},$$

as claimed. □

It remains to count sets $A \in \mathcal{G}(B)$ with $|Y| > m$. To do so, set $X := A' + A'$, and observe that X and Y are disjoint subsets of $A + A$. Since $|A + A| \leqslant \lambda k$, it follows that

$$\big|[\lambda k] \setminus X\big| \geqslant |Y| > m(B). \tag{6.27}$$

We will use Corollary 6.4.4 to count the sets with $|[\lambda k] \setminus X| \geqslant m(B)$.

Lemma 6.6.4. *If $B \subset \{-\lambda k/2, \dots, \lambda k\} \setminus [\lambda k/2]$, then there are at most*

$$\left(\frac{\lambda - 1}{\lambda}\right)^{m(B)/2} \binom{\lambda k/2}{k - b}$$

sets $A \in \mathcal{G}(B)$ with $|[\lambda k] \setminus X| \geqslant m(B)$.

Proof. We want to bound the number of sets $A' \subset [\lambda k/2]$, with $|A'| = k - b$, such that $|[\lambda k] \setminus (A' + A')| \geqslant m := m(B)$. Recall that $|B| \leqslant \delta k$ and $r(B) \geqslant c(\lambda, \varepsilon)$ (otherwise $\mathcal{G}(B)$ is empty), and note that therefore $m = r(B)/8\lambda \geqslant 2^{400}\lambda^{24}$. It follows, by Corollary 6.4.4 and (6.18), that there are at most

$$\left(\frac{\lambda - 1}{\lambda}\right)^{m/2} \binom{\lambda k/2}{k - b}$$

sets $A \in \mathcal{G}(B)$ such that $|[\lambda k] \setminus (A' + A')| \geqslant m$, as claimed. □

We can now easily deduce the claimed upper bound on the size of $\mathcal{G}(B)$.

Proof of Lemma 6.6.2. By (6.27), $|\mathcal{G}(B)|$ is at most the sum of the bounds in Lemmas 6.6.3 and 6.6.4. Recalling that $m(B) = r(B)/8\lambda$, this gives

$$|\mathcal{G}(B)| \leqslant \left(e^{-m(B)} + e^{-m(B)/2\lambda}\right)\binom{\lambda k/2}{k - b} \leqslant \exp\left(-\frac{r(B)}{2^5\lambda^2}\right)\binom{\lambda k/2}{k - b},$$

as required. □

Lemma 6.6.1 is a straightforward consequence.

Proof of Lemma 6.6.1. Fix b and r, and consider the sets $B \subset \{-\lambda k/2, \dots, \lambda k\} \setminus [\lambda k/2]$ with $|B| = b$ and $r(B) = r$. We may assume that $r > f(\lambda)b$ and $r \geqslant c(\lambda, \varepsilon)$, since otherwise $\mathcal{G}(B) \cap \mathcal{S} = \emptyset$. The number of choices for B (given b and r) is therefore at most

$$\binom{r}{b} \leqslant \left(2^{10}e\lambda^3\right)^{2^{-10}\lambda^{-3}r} \leqslant \exp\left(\frac{r}{2^7\lambda^2}\right)$$

since $r/b > f(\lambda) = 2^{10}\lambda^3$. By Lemma 6.6.2, it follows that

$$\left|\{A \in \mathcal{S} : b(A) = b,\, r(A) = r\}\right| \leqslant \exp\left(-\frac{r}{2^7\lambda^2}\right)\binom{\lambda k/2}{k - b} \leqslant \exp\left(-\frac{r}{2^8\lambda^2}\right)\binom{\lambda k/2}{k},$$

where the second inequality follows from (6.20), since $r/b > f(\lambda)$ and $\lambda \geqslant 3$.

Summing over choices of $r \geqslant c(\lambda, \varepsilon)$ and $b < r/f(\lambda)$, it follows that

$$|\mathcal{S}| \leqslant \sum_{r \geqslant c(\lambda,\varepsilon)} \frac{r}{f(\lambda)} \exp\left(-\frac{r}{2^8\lambda^2}\right)\binom{\lambda k/2}{k} \leqslant \exp\left(-\frac{c(\lambda, \varepsilon)}{2^9\lambda^2}\right)\binom{\lambda k/2}{k},$$

as required. □

6.7 Counting the moderately dense sets

Recall from Definition 6.2.1 and (6.3) the definitions of $b(A)$, $r(A)$ and $\mathcal{I}$, and let us write

$$\mathcal{D} := \Big\{ A \in \mathcal{I} \,:\, r(A) \leqslant f(\lambda) b(A) \Big\} \tag{6.28}$$

for the family of 'dense' sets in $\mathcal{I}$, where $f(\lambda) = 2^{10}\lambda^3$. In the next two sections we will prove the following quantitative version of Lemma 6.2.3.

Lemma 6.7.1. *Let $\lambda \geqslant 3$ and $\varepsilon \in (0,1)$, and let $k \in \mathbb{N}$. Then*

$$|\mathcal{D}| \leqslant \exp\bigg(- \frac{c(\lambda,\varepsilon)}{2^{20}\lambda^2} \bigg) \binom{\lambda k/2}{k}.$$

Let us fix $\lambda \geqslant 3$, $\varepsilon \in (0,1)$ and $k \in \mathbb{N}$ until the end of the proof of Lemma 6.7.1. In this section, we will deal with some relatively easy cases using the method of the previous section. Observe that

$$b(A) \geqslant \frac{c(\lambda,\varepsilon)}{f(\lambda)} \geqslant 2^{550}\lambda^{29} \tag{6.29}$$

for every $A \in \mathcal{D}$, since $r(A) \geqslant c(\lambda,\varepsilon)$ for every $A \in \mathcal{I}$, and by the definition (6.1) of $c(\lambda,\varepsilon)$.

For convenience, let us define, for each $b \in \mathbb{N}$ and $\mu \geqslant 1$,

$$\mathcal{D}(b,\mu) := \Big\{ A \in \mathcal{D} \,:\, |B| = b \text{ and } |(B+B) \setminus [\lambda k]| = \mu b, \text{ where } B = A \setminus [\lambda k/2] \Big\}. \tag{6.30}$$

The first step is to use Theorem 6.3.2 to bound the number of choices for $B = A \setminus [\lambda k/2]$. We will use the following lemma several times in the proof of Lemma 6.7.1.

Lemma 6.7.2. *Let $b \in \mathbb{N}$ and $\mu > 2$. There are at most*

$$e^{2\delta b} \bigg(\frac{\mu-2}{2} \bigg)^b \bigg(\frac{\mu}{\mu-2} \bigg)^{\mu b/2} \tag{6.31}$$

sets B such that $B = A \setminus [\lambda k/2]$ for some $A \in \mathcal{D}(b,\mu)$.

We will use the following observation in the proof of Lemma 6.7.2, and then again (several times) in the applications below.

Observation 6.7.3.

$$(x-2) \cdot \bigg(\frac{x}{x-2} \bigg)^{x/2} \leqslant (y-2) \bigg(\frac{y}{y-2} \bigg)^{x/2}$$

for every $x, y > 2$.

Proof. Set $q(x,y) := (x/y)^{x/2} \cdot \big((y-2)/(x-2)\big)^{(x-2)/2}$, and observe that

$$\log\big(q(x,y)^{2/x}\big) = \frac{2}{x} \cdot \log\frac{x}{y} + \frac{x-2}{x} \cdot \log\left(\frac{x(y-2)}{y(x-2)}\right) \leqslant \log\left(\frac{2}{x} \cdot \frac{x}{y} + \frac{x-2}{x} \cdot \frac{x(y-2)}{y(x-2)}\right) = 0,$$

using the concavity of the log function. □

Proof of Lemma 6.7.2. Set $B_1 := \{x \in B : x \leqslant 0\}$ and $B_2 := \{x \in B : x > \lambda k/2\}$, and recall from (6.29) that $b \geqslant 2^{550}\lambda^{29}$, and that $\delta = 2^{-32}\lambda^{-3}$. Observe first that, since $r(A) \leqslant f(\lambda)b$ for each $A \in \mathcal{D}(b,\mu)$, for each $i \in \{1,2\}$ there are at most

$$\binom{f(\lambda)b}{b^{3/4}} \leqslant \exp\big(b^{3/4}\log b\big) \leqslant e^{\delta b} \tag{6.32}$$

choices for the set B_i with $|B_i| \leqslant b^{3/4}$. Moreover, by Lemma 6.4.7, if $|B_i + B_i| \leqslant 2|B_i|$, then B_i is contained in an arithmetic progression of size $|B_i|+1$, and so in this case there are at most $r^3 \leqslant 2^{30}\lambda^9 b^3 \leqslant e^{\delta b}$ choices for B_i. Note that $\frac{\mu-2}{2}\big(\frac{\mu}{\mu-2}\big)^{\mu/2} \geqslant 1$ for every $\mu > 2$, so these bounds suffice when either $|B_i| \leqslant b^{3/4}$ or $|B_i + B_i| \leqslant 2|B_i|$.

Now, set $b_i = |B_i|$ and $\mu_i b_i = |B_i + B_i|$, and suppose that $b_i \geqslant b^{3/4}$, and $\mu_i > 2$. In this case we will use Theorem 6.3.2 to count the number of choices for B_i, and hence prove that, for each $i \in \{1,2\}$, the bound (6.31) holds with the pair (B,b) replaced by (B_i,b_i). Since $b = b_1 + b_2$, multiplying these two bounds will give (6.31) for the pair (B,b).

To check the condition on μ_i, observe that $B_i + B_i \subset [2\min(B), 2\max(B)] \setminus [\lambda k]$, and therefore

$$\mu_i b_i \leqslant 2 \cdot r(A) \leqslant 2f(\lambda)b, \tag{6.33}$$

for every $A \in \mathcal{D}(b,\mu)$, by (6.3) and (6.28). Since $b_i \geqslant b^{3/4}$, and recalling that $b \geqslant 2^{550}\lambda^{29}$, it follows that[7]

$$\mu_i \leqslant \frac{2f(\lambda)b}{b_i} \leqslant 2^{-36}\frac{b_i}{\big(\log(f(\lambda)b)\big)^3}.$$

Hence, by Theorem 6.3.2, the number of choices for B_i (given b_i and μ_i) is at most

$$\exp\left(2^9\mu_i^{1/6}b_i^{5/6}\log b_i\sqrt{\log(f(\lambda)b)}\right)\binom{\mu_i b_i/2}{b_i} \leqslant e^{\delta b}\binom{\mu_i b_i/2}{b_i}, \tag{6.34}$$

where the final inequality holds since $\mu_i^{1/6}b_i^{5/6} \leqslant (\mu_i b_i)^{1/6}b^{3/4} \leqslant 4\lambda \cdot b^{5/6}$, by (6.33), so

$$2^9\mu_i^{1/6}b_i^{5/6}\log b_i\sqrt{\log(f(\lambda)b)} \leqslant 2^{11}\lambda \cdot b^{5/6}(\log b)^2 \leqslant \delta b,$$

since $\delta = 2^{-32}\lambda^{-3}$ and $b \geqslant 2^{550}\lambda^{29}$.

[7] Using the bound $b_i \geqslant b^{3/4}$, the second inequality reduces to $b \geqslant 2^{74}f(\lambda)^2\big(\log(f(\lambda)b)\big)^6$, which follows (with room to spare) from $b \geqslant 2^{550}\lambda^{29}$, since $f(\lambda) = 2^{10}\lambda^3$.

Now, by Observations 6.4.8 and 6.7.3, it follows that

$$\binom{\mu_i b_i/2}{b_i} \leqslant \left(\frac{\mu_i - 2}{2} \cdot \left(\frac{\mu_i}{\mu_i - 2}\right)^{\mu_i/2}\right)^{b_i} \leqslant \left(\frac{\mu - 2}{2}\right)^{b_i} \left(\frac{\mu}{\mu - 2}\right)^{\mu_i b_i/2}. \tag{6.35}$$

Since $\mu b = \mu_1 b_1 + \mu_2 b_2$, the lemma follows from (6.32), (6.34) and (6.35). □

We can now bound the number of sets $A \in \mathcal{D}(b, \mu)$ such that $r(A) \geqslant 2^{11}\mu b$.

Lemma 6.7.4. *Let $b \in \mathbb{N}$ and $\mu \geqslant 1$. If $r \geqslant 2^{11}\mu b$, then there are at most*

$$\exp\left(-\frac{r}{2^7\lambda^2}\right)\binom{\lambda k/2}{k}$$

sets $A \in \mathcal{D}(b, \mu)$ with $r(A) = r$.

Proof. Observe first that if $\mu \leqslant 2$, then B is contained in two arithmetic progressions of combined size at most $|B| + 2$, by Lemma 6.4.7 (cf. the proof of Lemma 6.7.2), and so in this case there are at most r^6 choices for B. By Lemma 6.6.2 and (6.20), it follows that there are at most

$$\sum_B |\mathcal{G}(B)| \leqslant r^6 \exp\left(-\frac{r}{2^6\lambda^2}\right)\left(\frac{2}{\lambda - 2}\right)^b \binom{\lambda k/2}{k} \tag{6.36}$$

sets $A \in \mathcal{D}(b, \mu)$ with $r(A) = r$, where the sum is over sets with $|(B + B) \setminus [\lambda k]| \leqslant 2|B| = 2b$ and $r(B) = r$. Now, since $r \geqslant 2^{11}b \geqslant 2^{561}\lambda^{29}$, by (6.29), and $\lambda \geqslant 3$, we have[8]

$$r^6\left(\frac{2}{\lambda - 2}\right)^b \leqslant \exp\left(\frac{r}{2^7\lambda^2}\right),$$

and combining this with (6.36), we obtain the claimed bound.

Let us therefore assume from now on that $\mu > 2$. By Lemma 6.7.2 and Observation 6.7.3, it follows that there are at most

$$e^{2\delta b}\left(\frac{\lambda - 2}{2}\right)^b \left(\frac{\lambda}{\lambda - 2}\right)^{\mu b/2} \tag{6.37}$$

sets B such that $B = A \setminus [\lambda k/2]$ for some $A \in \mathcal{D}(b, \mu)$. In order to count the sets A for a given B, we will need to consider three cases. For each set B that is counted in (6.37), set $m := m(B) = r(B)/8\lambda$, and for each $A \in \mathcal{G}(B)$, let $Y = Y(A)$ be the set defined in (6.25).

Case 1: $|Y| \leqslant m$.

[8] Indeed, if $\lambda \geqslant 4$ then note that $r^6 \leqslant \exp\left(r/2^7\lambda^2\right)$, and if $\lambda \leqslant 4$ then note that $r^6 2^{r/2^{11}} \leqslant \exp\left(r/2^{11}\right)$.

By Lemma 6.6.3 and (6.20), for each set B there are at most

$$e^{-m}\binom{\lambda k/2}{k-b} \leqslant e^{-m}\left(\frac{2}{\lambda-2}\right)^b\binom{\lambda k/2}{k}$$

sets $A \in \mathcal{G}(B)$ with $|Y| \leqslant m$. Summing over sets B as in (6.37), noting that if $r(B) = r$ then $\mu b \leqslant 2^{-11}r \leqslant 2^{-8}\lambda m$, and recalling that $\lambda \geqslant 3$, it follows that there are at most[9]

$$e^{2\delta b}\left(\frac{\lambda}{\lambda-2}\right)^{\mu b/2} e^{-m}\binom{\lambda k/2}{k} \leqslant e^{-m/2}\binom{\lambda k/2}{k} \tag{6.38}$$

sets $A \in \mathcal{D}(b,\mu)$ with $r(A) = r$ and $|Y| \leqslant m$.

Counting the sets with larger Y is somewhat more delicate, and we will need to partition into two cases, depending on the intersection of Y with the set $B + B$.

Case 2: $|Y| \geqslant m$ and $|Y \cap (B+B)| \leqslant m/2$.

In this case we will apply Corollary 6.4.4. To do so, observe first that

$$|A' + A'| + |Y \cup (B+B) \setminus [\lambda k]| \leqslant |A + A| \leqslant \lambda k,$$

since $A' + A' \subset [\lambda k]$ and $Y \subset (A' + B) \setminus [\lambda k]$, by (6.25). Recall that $|(B+B) \setminus [\lambda k]| = \mu b$ for each $A \in \mathcal{D}(b,\mu)$, by (6.30). Therefore, if $|Y| \geqslant m$ and $|Y \cap (B+B)| \leqslant m/2$, then

$$|[\lambda k] \setminus (A' + A')| \geqslant \mu b + m/2.$$

Moreover, if $\mathcal{D}(b,\mu)$ is non-empty then $2^{400}\lambda^{24} \leqslant b \leqslant \delta k$, by (6.29) and Definition 6.2.1, and since $\mathcal{D}(b,\mu) \subset \mathcal{D} \subset \mathcal{I}$. Hence, by Corollary 6.4.4 and (6.20), it follows that for each set B counted in (6.37), there are at most

$$\exp\big(2\delta \cdot (\mu b + m/2)\big)\left(\frac{\lambda-2}{\lambda}\right)^{\mu b/2+m/4}\left(\frac{2}{\lambda-2}\right)^b\binom{\lambda k/2}{k}$$

sets $A \in \mathcal{G}(B)$ such that $|Y| \geqslant m$ and $|Y \cap (B+B)| \leqslant m/2$.

Summing over sets B, and using (6.37), it follows that there are at most

$$\exp\big(2\delta \cdot (b + \mu b + m/2)\big)\left(\frac{\lambda-2}{\lambda}\right)^{m/4}\binom{\lambda k/2}{k}$$

choices for A in this case. Now, since $\mu b \leqslant 2^{-8}\lambda m$ and $\delta = 2^{-32}\lambda^{-3}$, we have

$$\exp\big(2\delta \cdot (b + \mu b + m/2)\big)\left(\frac{\lambda-2}{\lambda}\right)^{m/4} \leqslant \exp\left(\delta\lambda m - \frac{m}{2\lambda}\right) \leqslant \exp\left(-\frac{m}{4\lambda}\right),$$

[9] Indeed, if $\lambda \leqslant 4$ then $3^{\mu b} \leqslant 3^{m/2^6} \leqslant e^{m/2^5}$, and otherwise $\lambda/(\lambda-2) \leqslant e^{2/(\lambda-2)} \leqslant e^{4/\lambda}$.

and hence the number of sets A with $|Y| \geqslant m$ and $|Y \cap (B+B)| \leqslant m/2$ is at most

$$\exp\Big(-\frac{m}{4\lambda}\Big)\binom{\lambda k/2}{k} = \exp\Big(-\frac{r}{2^5\lambda^2}\Big)\binom{\lambda k/2}{k}. \tag{6.39}$$

Finally, we count sets such that Y has large intersection with $B+B$.

Case 3: $|Y| \geqslant m$ and $|Y \cap (B+B)| > m/2$.

Let B be such that $B = A \setminus [\lambda k/2]$ for some $A \in \mathcal{D}(b,\mu)$, and consider the set

$$Z := \big\{x \in [\lambda k/2] : x + \min(B) \in (B+B)\setminus[\lambda k] \ \text{ or } \ x + \max(B) \in (B+B)\setminus[\lambda k]\big\}.$$

Observe that $|A' \cap Z| > m/2$ and $|Z| \leqslant |(B+B)\setminus[\lambda k]|$. It follows that the number of choices for A' is at most

$$\sum_{\ell > m/2}\binom{\mu b}{\ell}\binom{\lambda k/2}{k-b-\ell} \leqslant \sum_{\ell > m/2}\Big(\frac{e\mu b}{\ell}\cdot\frac{2}{\lambda-2}\Big)^{\ell}\binom{\lambda k/2}{k-b} \leqslant 2^{-m}\Big(\frac{2}{\lambda-2}\Big)^{b}\binom{\lambda k/2}{k},$$

where the inequalities follow from (6.20) and the bounds $\mu b \leqslant 2^{-8}\lambda m$ and $\lambda \geqslant 3$, which together imply that

$$\frac{2e\mu b}{m}\cdot\frac{2}{\lambda-2} \leqslant \frac{e\lambda}{2^5(\lambda-2)} \leqslant \frac{1}{4}.$$

By (6.37), and recalling again that $\mu b \leqslant 2^{-8}\lambda m$, it follows that there are at most

$$e^{2\delta b}\Big(\frac{\lambda}{\lambda-2}\Big)^{\mu b/2}2^{-m}\binom{\lambda k/2}{k} \leqslant 2^{-m/2}\binom{\lambda k/2}{k} \leqslant \exp\Big(-\frac{r}{2^5\lambda}\Big)\binom{\lambda k/2}{k} \tag{6.40}$$

choices for A in this case. Summing (6.38), (6.39) and (6.40) gives the required bound on the number of sets $A \in \mathcal{D}(b,\mu)$ with $r(A) = r$. □

It will be useful in the next section (which deals with the case $r \leqslant 2^{11}\mu b$) to be able to assume that $\mu = \Theta(\lambda)$. The next lemma, which follows from Corollary 6.4.4, provides a suitable bound on the size of $\mathcal{D}(b,\mu)$ when this is not the case.

Lemma 6.7.5. *Let $b \in \mathbb{N}$. If $r \leqslant 2^{11}\mu b$ and either $\mu \leqslant 2$ or $\mu \notin (\lambda/2, 2\lambda)$, then there are at most*

$$\exp\Big(-\frac{r}{2^{16}\lambda}\Big)\binom{\lambda k/2}{k}$$

sets $A \in \mathcal{D}(b,\mu)$ with $r(A) = r$.

Proof. For each $A \in \mathcal{D}(b,\mu)$, set $A' := A \cap [\lambda k/2]$ and $B := A \setminus [\lambda k/2]$, and observe that

$$\big|[\lambda k] \setminus (A' + A')\big| \ \geqslant \ \big|(B+B)\setminus[\lambda k]\big| \ = \ \mu b,$$

since $|A+A| \leqslant \lambda k$. Hence, by Corollary 6.4.4 applied with $m = \mu b \geqslant 2^{400}\lambda^{24}$, and using (6.20), there are at most

$$e^{2\delta\mu b}\left(\frac{\lambda-2}{\lambda}\right)^{\mu b/2}\left(\frac{2}{\lambda-2}\right)^{b}\binom{\lambda k/2}{k}, \tag{6.41}$$

choices for the set A'.

Suppose first that $\mu > 2$, and recall from Lemma 6.7.2 that in this case there are at most

$$e^{2\delta b}\left(\frac{\mu-2}{2}\right)^{b}\left(\frac{\mu}{\mu-2}\right)^{\mu b/2} \tag{6.42}$$

sets B with $B = A \setminus [\lambda k/2]$ for some $A \in \mathcal{D}(b,\mu)$. Moreover, applying Observation 6.7.3 with $x = \mu$ and $y = 2\lambda - 2$ gives

$$\left(\frac{\mu-2}{2}\right)^{b}\left(\frac{\mu}{\mu-2}\right)^{\mu b/2} \leqslant (\lambda-2)^{b}\left(\frac{\lambda-1}{\lambda-2}\right)^{\mu b/2}.$$

Thus, if $\mu \geqslant 2\lambda$ (and therefore $\mu \geqslant 6$), then the product of (6.41) and (6.42) is at most[10]

$$e^{3\delta\mu b}\cdot 2^{b}\left(\frac{\lambda-1}{\lambda}\right)^{\mu b/2}\binom{\lambda k/2}{k} \leqslant \exp\left(-\frac{\mu b}{2^5\lambda}\right)\binom{\lambda k/2}{k} \leqslant \exp\left(-\frac{r}{2^{16}\lambda}\right)\binom{\lambda k/2}{k},$$

since $r \leqslant 2^{11}\mu b$.

Next, if $\mu > 2$ and $\lambda > 4$, then applying Observation 6.7.3 with $x = \mu$ and $y = \lambda/2$ gives

$$\left(\frac{\mu-2}{2}\right)^{b}\left(\frac{\mu}{\mu-2}\right)^{\mu b/2} \leqslant \left(\frac{\lambda-4}{4}\right)^{b}\left(\frac{\lambda}{\lambda-4}\right)^{\mu b/2}.$$

Thus, if $2 < \mu \leqslant \lambda/2$ (and therefore $\lambda > 4$), then the product of (6.41) and (6.42) is at most

$$e^{3\delta\mu b}\cdot\left(\frac{\lambda-2}{\lambda-4}\right)^{\mu b/2}\left(\frac{\lambda-4}{2\lambda-4}\right)^{b}\binom{\lambda k/2}{k} \leqslant 2^{-b/4}\binom{\lambda k/2}{k},$$

where the final step holds since $\delta = 2^{-32}\lambda^{-3}$, $\mu \leqslant \lambda/2$, and

$$2^4\cdot\left(\frac{\lambda-2}{\lambda-4}\right)^{\lambda}\left(\frac{\lambda-4}{2\lambda-4}\right)^{4} = \left(1+\frac{2}{\lambda-4}\right)^{\lambda-4} \leqslant e^2.$$

Since $r \leqslant 2^{11}\mu b \leqslant 2^{10}\lambda b$, it follows that if $2 < \mu \leqslant \lambda/2$ then there are at most

$$\exp\left(-\frac{r}{2^{13}\lambda}\right)\binom{\lambda k/2}{k}$$

sets $A \in \mathcal{D}(b,\mu)$ with $r(A) = r$.

[10] For the penultimate step, recall that $\delta \leqslant 2^{-7}\lambda^{-1}$, and apply the inequality $2\cdot e^{-x/2} < e^{-x/16}$, which holds for all $x \geqslant 2$, with $x = \mu/\lambda$.

Finally, if $\mu \leqslant 2$ then B is contained in two arithmetic progressions of combined size at most $|B|+2$, by Lemma 6.4.7, and so in this case there are at most $r^6 \leqslant 2^{72}b^6 \leqslant e^{\delta b}$ choices for B, by (6.29). Noting that $\mu b \geqslant 2b-2$, it follows from (6.41) that there are at most

$$e^{5\delta b}\left(\frac{\lambda-2}{\lambda}\right)^{b-1}\left(\frac{2}{\lambda-2}\right)^{b}\binom{\lambda k/2}{k} \leqslant \exp\left(-\frac{r}{2^{14}}\right)\binom{\lambda k/2}{k}$$

choices for A, where the last inequality holds since $\lambda \geqslant 3$ and $r \leqslant 2^{11}\mu b \leqslant 2^{12}b$. □

6.8 Counting the very dense sets with containers

It remains to bound the size of the family[11]

$$\mathcal{D}^*(b,\mu) := \Big\{A \in \mathcal{D}(b,\mu) : r(A) \leqslant 2^{11}\mu b\Big\} \tag{6.43}$$

of *very dense* sets, for each $\mu > 2$ with $\lambda/2 \leqslant \mu \leqslant 2\lambda$. To do so, we will once again use Theorem 6.3.1, but this time our application of it will be rather different. Recall first, from Lemma 6.4.3, that the 'missing' set $M(A) := [\lambda k] \setminus (A+A)$ is typically contained near the ends of $[\lambda k]$ (see Lemma 6.8.2, below). We will use Corollary 6.3.4 to find a family of $2^{o(b)}$ containers (C,D) for the parts of A 'close' to the endpoints, and for $M(A)$ (see Corollary 6.8.1). We will then, in Lemma 6.8.3, bound the number of sets $A \in \mathcal{D}^*(b,\mu)$ corresponding to each container. Our bound decreases exponentially with b, and we will therefore be able to take a union bound over containers.

To state the version of Corollary 6.3.4 we will use, we need a little additional notation. First, for each $b \in \mathbb{N}$, set $Y(b) := Y_1 \cup Y_2$ and $X(b) := (Y_1+Y_1) \cup (Y_2+Y_2)$, where

$$Y_1 := \big\{0,\dots,2^{18}\lambda^2 b\big\}, \qquad \text{and} \qquad Y_2 := \big\{\lambda k/2 - 2^{18}\lambda^2 b,\dots,\lambda k/2\big\},$$

Moreover, define $M(A) := [\lambda k] \setminus (A+A)$ and

$$\mathcal{T}(b) := \big\{A \in \mathcal{I} \,:\, b(A) = b \text{ and } M(A) \subset X(b)\big\}.$$

As we will see below (see Lemma 6.8.2), this family contains almost all of $\mathcal{D}^*(b,\mu)$.

Our key tool in this section will be the following immediate consequence of Corollary 6.3.4.

Corollary 6.8.1. *For each $b \in \mathbb{N}$, there exists a family $\mathcal{B}(b) \subset 2^{X(b)} \times 2^{Y(b)}$ of size at most*

$$\exp\big(2^{50}\lambda^2 b^{7/8}\big)$$

[11] Recall that the family $\mathcal{D}(b,\mu)$ was defined in (6.30).

such that:

(a) *For each $A \in \mathcal{T}(b)$, there exists $(C,D) \in \mathcal{B}(b)$ with $M(A) \subset C$ and $A \cap Y(b) \subset D$.*

(b) *For every $(C,D) \in \mathcal{B}(b)$,*

$$|D| \leqslant \max\left\{|Y(b)| - \frac{|C|}{2} + |Y(b)|^{5/6}, \frac{3|Y(b)|}{\log|Y(b)|}\right\}.$$

Proof. We apply Corollary 6.3.4 with $S_1 = Y_1$, $S_2 = Y_2$ and $\gamma = |Y(b)|^{-1/6}/4$. The bound on the size of $\mathcal{B}(b)$ follows from (6.7), since $\log|Y(b)| \leqslant 2^6(\lambda^2 b)^{1/36}$ and

$$2^{17}\gamma^{-2}\sqrt{|Y(b)|} = 2^{21}|Y(b)|^{5/6} \leqslant 2^{21}\big(2^{20}\lambda^2 b\big)^{5/6} \leqslant 2^{41}\lambda^{5/3}b^{5/6},$$

where in both cases we used the bound $|Y(b)| \leqslant 2^{20}\lambda^2 b$.

The bound on $|D|$ for each $(C,D) \in \mathcal{B}(b)$ follows from (6.8). Finally, for each $A \in \mathcal{T}(b)$ we apply Corollary 6.3.4(a) with $U := A \cap Y(b)$ and $W := M(A) \subset X(b) \setminus (U+U)$. It follows that there exists $(C,D) \in \mathcal{B}(b)$ such that $M(A) \subset C$ and $A \cap Y(b) \subset D$, as claimed. □

Recall from (6.29) that $b(A) \geqslant 2^{550}\lambda^{29}$ for every $A \in \mathcal{D}^*(b,\mu) \subset \mathcal{D}$. Since $\delta = 2^{-32}\lambda^{-3}$, it follows that

$$|Y(b)|^{5/6} \leqslant \delta b. \tag{6.44}$$

In the calculations below, we will also need the inequalities

$$\binom{a+c}{b} \leqslant \left(1 + \frac{c}{a-b}\right)^b \binom{a}{b} \qquad \text{and} \qquad \binom{a-c}{b-c} \leqslant \left(\frac{b}{a}\right)^c \binom{a}{b} \tag{6.45}$$

Before bounding the number of sets in each container, let's first observe that, by our choice of $X(b)$, most members of $\mathcal{D}^*(b,\mu)$ are also in $\mathcal{T}(b)$.

Lemma 6.8.2. *For each $b \leqslant \delta k$ and $\mu \leqslant 2\lambda$, there are at most*

$$e^{-b}\binom{\lambda k/2}{k} \tag{6.46}$$

sets $A \in \mathcal{D}^*(b,\mu)$ such that $M(A) \not\subset X(b)$.

Proof. Recalling (6.43), let A be a uniformly random k-subset of $L := [-2^{11}\mu b, \lambda k/2 + 2^{11}\mu b]$, and observe that

$$\mathbb{P}\big(M(A) \not\subset X(b)\big) \leqslant \mathbb{P}\Big(\big\{M'+1, \dots, \lambda k - M' - 1\big\} \not\subset A + A\Big),$$

where $M' := 2^{19}\lambda^2 b$, by the definitions of $M(A) = [\lambda k] \setminus (A + A)$ and $X(b)$. By Lemma 6.4.3 (applied with $n = \lambda k/2 + 2^{12}\mu b + 1$ and $M = M' + 2^{12}\mu b + 2$), it follows that

$$\mathbb{P}\big(M(A) \not\subset X(b)\big) \leqslant \frac{8}{p^2} \cdot \big(1 - p^2\big)^{M/2},$$

where $p = k\big(\lambda k/2 + 2^{12}\mu b + 1\big)^{-1}$. Now, recall that $b \leqslant \delta k$ and $\mu \leqslant 2\lambda$, and observe that therefore $p \geqslant 1/\lambda$. Since $M \geqslant M' = 2^{19}\lambda^2 b$, it follows that

$$\mathbb{P}\big(M(A) \not\subset X(b)\big) \leqslant 8\lambda^2 \cdot e^{-M/2\lambda^2} \leqslant \exp\big(-2^{17}b\big),$$

since $b \geqslant 2^{550}\lambda^{29}$. In order to deduce a bound on the number of sets such that $M(A) \not\subset X(b)$, we simply need to multiply by the total number of k-subsets of L. There are at most

$$\binom{\lambda k/2 + 2^{12}\mu b + 1}{k} \leqslant \left(1 + \frac{2^{13}\mu b + 2}{(\lambda - 2)k}\right)^k \binom{\lambda k/2}{k} \leqslant \exp\big(2^{16}b\big)\binom{\lambda k/2}{k}$$

such sets, where the first inequality holds by (6.45), and the second because $\lambda \geqslant 3$ and $\mu \leqslant 2\lambda$. Hence, there are at most

$$\exp\big(-2^{17}b + 2^{16}b\big)\binom{\lambda k/2}{k} \leqslant e^{-b}\binom{\lambda k/2}{k}$$

sets $A \in \mathcal{D}^*(b, \mu)$ with $M(A) \not\subset X(b)$, as claimed. □

To deduce Lemma 6.7.1 from Corollary 6.8.1, we will need to bound the size of the containers in $\mathcal{B}(b)$. The following lemma provides the bound we need.

Lemma 6.8.3. *Let $b \leqslant \delta k$ and $\mu > 2$, with $\lambda/2 \leqslant \mu \leqslant 2\lambda$. For each $(C, D) \in \mathcal{B}(b)$, there are at most*

$$e^{-b/32\lambda}\binom{\lambda k/2}{k}$$

sets $A \in \mathcal{D}^*(b, \mu)$ such that $M(A) \subset C$ and $A \cap Y(b) \subset D$.

In the proof of Lemma 6.8.3, we will need the following binomial inequalities, whose (straightforward, but slightly tedious) proofs are given in [32, Appendix B]. Set $\alpha := 2^{25}\lambda^2\delta = 2^{-7}\lambda^{-1}$.

Observation 6.8.4. *Let $b \leqslant \delta k$ and $\mu > 2$, with $\lambda/2 \leqslant \mu \leqslant 2\lambda$, and let $s \leqslant t \leqslant 2^{22}\lambda^2 b$. Then*

$$\binom{\lambda k/2 - \mu b - s}{k - b - s} \leqslant e^{\alpha b}\left(\frac{\lambda - 2}{\lambda}\right)^{\mu b}\left(\frac{2}{\lambda - 2}\right)^b\binom{\lambda k/2 - s}{k - s},$$

and

$$\binom{\lambda k/2 - \mu b/2 - s/2 - t + \delta b}{k - b - s} \leqslant e^{\alpha b}\left(\frac{\lambda - 2}{\lambda}\right)^{\mu b/2}\left(\frac{2}{\lambda - 2}\right)^b\binom{\lambda k/2 - s/2 - t}{k - s}.$$

We are now ready to prove our key lemma.

Proof of Lemma 6.8.3. First, by Lemma 6.7.2 and Observation 6.7.3, there are at most

$$e^{2\delta b}\left(\frac{\mu-2}{2}\right)^b\left(\frac{\mu}{\mu-2}\right)^{\mu b/2} \leqslant e^{2\delta b}\left(\frac{\lambda-2}{2}\right)^b\left(\frac{\lambda}{\lambda-2}\right)^{\mu b/2} \tag{6.47}$$

sets B such that $B = A \setminus [\lambda k/2]$ for some $A \in \mathcal{D}^*(b,\mu) \subset \mathcal{D}(b,\mu)$. Fix such a set B, let $A \in \mathcal{D}^*(b,\mu)$ with $B = A\setminus[\lambda k/2]$, and recall that $M(A) = [\lambda k]\setminus(A+A)$. Note that $|M(A)| \geqslant \mu b$, since $|A+A| \leqslant \lambda k$ and $|(B+B)\setminus[\lambda k]| = \mu b$. Now, define[12]

$$\tilde{C} := C - \{\min(B), \max(B)\} \qquad \text{and} \qquad \tilde{D} := \big([\lambda k/2] \setminus Y(b)\big) \cup D, \tag{6.48}$$

and observe that if $A \cap Y(b) \subset D$, then $A' \subset \tilde{D}$, where (as usual) $A' = A \cap [\lambda k/2]$. Set

$$S := \tilde{C} \cap A' \qquad \text{and} \qquad T := \tilde{C} \cap \tilde{D},$$

and observe that $S \subset T$, and that for each $x \in S$, either $x + \max(B)$ or $x + \min(B)$ is contained in $C \cap (A+A)$. Moreover, the sets $S + \max(B)$ and $S + \min(B)$ are disjoint, since $S \subset [\lambda k/2]$ and $\max(B) - \min(B) > \lambda k/2$. It follows that if $M(A) \subset C$, then

$$|C| \geqslant |M(A)| + |S| \geqslant \mu b + |S|. \tag{6.49}$$

For each $s,t \in \mathbb{N}$, let us write $g(s,t)$ for the number of sets $A \in \mathcal{D}^*(b,\mu)$ such that $M(A) \subset C$ and $A \cap Y(b) \subset D$, and such that

$$|S| = s \qquad \text{and} \qquad |T| = t.$$

Since $S \subset T$, we have at most $\binom{t}{s}$ choices for S. We will bound $g(s,t)$ in two different ways, depending on the values of s and t.

Claim: If $s \leqslant b/16$ and $t \leqslant \lambda b$, then

$$g(s,t) \leqslant e^{-b/8}\binom{\lambda k/2}{k}. \tag{6.50}$$

Proof of claim. In this case we will use the bound

$$|\tilde{D} \setminus \tilde{C}| \leqslant \frac{\lambda k}{2} - |C| \leqslant \frac{\lambda k}{2} - \mu b - s. \tag{6.51}$$

The second inequality is (6.49), and therefore, recalling that $\tilde{D} \subset [\lambda k/2]$, to prove (6.51) it will suffice to show that

$$\big|\tilde{C} \cap [\lambda k/2]\big| \geqslant |C|. \tag{6.52}$$

[12] To avoid any possible confusion, we emphasize that $\tilde{C}$ is the union of two shifted copies of the set C.

To prove (6.52), observe first that $2^{19}\lambda^2 b \leqslant \lambda k/8$, since $b \leqslant \delta k$ and $\delta = 2^{-32}\lambda^{-3}$, and therefore

$$C \subset X(b) \subset \{0, \dots, \lambda k/8\} \cup \{7\lambda k/8, \dots, \lambda k\}. \tag{6.53}$$

Moreover, $r(A) \leqslant 2^{11}\mu b \leqslant 2^{12}\lambda \cdot \delta k \leqslant \lambda k/8$ for every $A \in \mathcal{D}^*(b, \mu)$, and therefore

$$-\lambda k/8 < \min(B) \leqslant 0 \qquad \text{and} \qquad \lambda k/2 < \max(B) < 5\lambda k/8. \tag{6.54}$$

It follows from (6.53) and (6.54) that $\big|\tilde{C} \cap [\lambda k/2]\big| \geqslant |C|$, as claimed.

Now, recalling that $A' \subset \tilde{D}$ and $S = \tilde{C} \cap A'$, it follows from (6.47) and (6.51) that

$$g(s,t) \leqslant e^{2\delta b}\left(\frac{\lambda-2}{2}\right)^b \left(\frac{\lambda}{\lambda-2}\right)^{\mu b/2} \binom{\lambda k/2 - \mu b - s}{k-b-s}\binom{t}{s}.$$

Observe that $s \leqslant t \leqslant |\tilde{C}| \leqslant 2|C| \leqslant 2|X(b)| \leqslant 2^{22}\lambda^2 b$. Thus, by Observation 6.8.4, we have

$$\binom{\lambda k/2 - \mu b - s}{k-b-s} \leqslant e^{\alpha b}\left(\frac{\lambda-2}{\lambda}\right)^{\mu b}\left(\frac{2}{\lambda-2}\right)^b \binom{\lambda k/2 - s}{k-s},$$

and therefore, by (6.45),

$$g(s,t) \leqslant e^{2\alpha b}\left(\frac{\lambda-2}{\lambda}\right)^{\mu b/2}\binom{\lambda k/2 - s}{k-s}\binom{t}{s} \leqslant e^{2\alpha b}\left(\frac{\lambda-2}{\lambda}\right)^{\mu b/2}\left(\frac{2}{\lambda}\cdot\frac{et}{s}\right)^s \binom{\lambda k/2}{k}.$$

Since $s \leqslant b/16$ and $t \leqslant \lambda b$, and recalling that $\mu \geqslant \lambda/2$, it follows that

$$g(s,t) \leqslant e^{2\alpha b}\cdot e^{-b/2}\cdot (32e)^{b/16}\binom{\lambda k/2}{k} \leqslant e^{-b/8}\binom{\lambda k/2}{k},$$

as claimed. □

We may therefore assume that either $s \geqslant b/16$ or $t \geqslant \lambda b$. In this case observe that $|\tilde{D}| = \lambda k/2 - |Y(b)| + |D|$, by (6.48) (and since $D \subset Y(b) \subset [\lambda k/2]$), and therefore

$$|\tilde{D}| \leqslant \frac{\lambda k}{2} - \frac{|C|}{2} + |Y(b)|^{5/6} \leqslant \frac{\lambda k - \mu b - s}{2} + \delta b, \tag{6.55}$$

by Corollary 6.8.1(b), together with (6.44) and (6.49). Since $A' \subset \tilde{D}$ and $S = \tilde{C} \cap A'$, it follows from (6.47) and (6.55) that

$$g(s,t) \leqslant e^{2\delta b}\left(\frac{\lambda-2}{2}\right)^b \left(\frac{\lambda}{\lambda-2}\right)^{\mu b/2}\binom{\lambda k/2 - \mu b/2 - s/2 - t + \delta b}{k-b-s}\binom{t}{s}.$$

Since $s \leqslant t \leqslant 2^{22}\lambda^2 b$, it follows by Observation 6.8.4 that

$$g(s,t) \leqslant e^{2\alpha b}\binom{\lambda k/2 - s/2 - t}{k-s}\binom{t}{s}. \tag{6.56}$$

Now, if $s \geqslant b/16$ then, by (6.17), we have

$$\binom{\lambda k/2 - s/2 - t}{k-s}\binom{t}{s} \leqslant \binom{\lambda k/2 - s/2}{k} \leqslant \left(\frac{\lambda - 2}{\lambda}\right)^{s/2}\binom{\lambda k/2}{k} \leqslant e^{-b/16\lambda}\binom{\lambda k/2}{k}.$$

On the other hand, if $s \leqslant b/16$ and $t \geqslant \lambda b$, then using (6.45) and (6.17), and noting that $t - s/2 \geqslant t/2$, we obtain

$$\binom{\lambda k/2 - s/2 - t}{k-s}\binom{t}{s} \leqslant \left(\frac{2}{\lambda}\cdot\frac{et}{s}\right)^s\binom{\lambda k/2 + s/2 - t}{k} \leqslant \left(\frac{2et}{\lambda s}\right)^s\left(\frac{\lambda-2}{\lambda}\right)^{t/2}\binom{\lambda k/2}{k}.$$

Now, observe that (for $s \leqslant b/16$ and $t \geqslant \lambda b$) the right-hand side is increasing in s and decreasing in t, since $2et/\lambda s \geqslant 32e$ (and by simple calculus). It follows that

$$\left(\frac{2et}{\lambda s}\right)^s\left(\frac{\lambda-2}{\lambda}\right)^{t/2} \leqslant (32e)^{b/16}\cdot e^{-b} \leqslant e^{-b/2}.$$

Combining these bounds, and recalling that $\alpha = 2^{-7}\lambda^{-1}$, we obtain

$$g(s,t) \leqslant e^{2\alpha b}\left(e^{-b/16\lambda} + e^{-b/2}\right)\binom{\lambda k/2}{k} \leqslant e^{-b/24\lambda}\binom{\lambda k/2}{k}. \tag{6.57}$$

Finally, summing the bounds (6.50) and (6.57) over $s \leqslant t \leqslant 2^{22}\lambda^2 b$, and recalling that $b \geqslant 2^{550}\lambda^{29}$, we obtain the claimed bound. □

We are finally ready to prove Lemma 6.7.1.

Proof of Lemma 6.7.1. Let us fix $b, r \in \mathbb{N}$ and $\mu \geqslant 1$, and bound the number of sets $A \in \mathcal{D}(b,\mu)$ with $r(A) = r$. Recall first that if $r \geqslant 2^{11}\mu b$ then, by Lemma 6.7.4, there are at most

$$\exp\left(-\frac{r}{2^7\lambda^2}\right)\binom{\lambda k/2}{k}$$

such sets, and if $r \leqslant 2^{11}\mu b$ and either $\mu \leqslant 2$, $\mu \leqslant \lambda/2$ or $\mu \geqslant 2\lambda$, then by Lemma 6.7.5 there are at most

$$\exp\left(-\frac{r}{2^{16}\lambda}\right)\binom{\lambda k/2}{k}$$

such sets. Now, if $r \leqslant 2^{11}\mu b$ and $\lambda/2 \leqslant \mu \leqslant 2\lambda$, then by Lemma 6.8.2 there are at most

$$e^{-b}\binom{\lambda k/2}{k} \leqslant \exp\left(-\frac{r}{2^{12}\lambda}\right)\binom{\lambda k/2}{k}$$

such sets that are not in $\mathcal{T}(b)$. Moreover, by Corollary 6.8.1, there exists a family $\mathcal{B}(b)$ of size at most

$$\exp\left(2^{50}\lambda^2 b^{7/8}\right)$$

such that for every $A \in \mathcal{T}(b)$, there exists $(C, D) \in \mathcal{B}(b)$ with $M(A) \subset C$ and $A \cap Y(b) \subset D$. Finally, by Lemma 6.8.3, for each $(C, D) \in \mathcal{B}(b)$ there are at most

$$e^{-b/32\lambda}\binom{\lambda k/2}{k} \leqslant \exp\left(-\frac{r}{2^{18}\lambda^2}\right)\binom{\lambda k/2}{k}$$

sets $A \in \mathcal{T}(b) \cap \mathcal{D}^*(b, \mu)$ such that $M(A) \subset C$ and $A \cap Y(b) \subset D$.

Combining these bounds, it follows that there are at most

$$\exp\left(2^{50}\lambda^2 b^{7/8}\right)\exp\left(-\frac{r}{2^{18}\lambda^2}\right)\binom{\lambda k/2}{k}$$

sets $A \in \mathcal{D}(b, \mu)$ with $r(A) = r$. Now, summing over choices of $b \leqslant r$ and $\mu \leqslant 2r/b$ such that $\mu b \in \mathbb{N}$, and recalling that $r \geqslant 2^{560}\lambda^{32}$, it follows that there are at most

$$\exp\left(-\frac{r}{2^{19}\lambda^2}\right)\binom{\lambda k/2}{k}$$

sets $A \in \mathcal{D}$ with $r(A) = r$.

Finally, summing over $r \geqslant c(\lambda, \varepsilon)$, we deduce that

$$|\mathcal{D}| \leqslant \exp\left(-\frac{c(\lambda, \varepsilon)}{2^{20}\lambda^2}\right)\binom{\lambda k/2}{k},$$

as claimed. □

6.9 The proof of Theorem 6.1.1

In this section we will prove the following quantitative version of Theorem 6.1.1, which allows us to control the typical structure of A when $\lambda = k^{o(1)}$. Recall that $\delta = 2^{-32}\lambda^{-3}$.

Theorem 6.9.1. *Let $\lambda \geqslant 3$ and $n, k \in \mathbb{N}$ be such that $k \geqslant 2^{400}\lambda^{25}(\log n)^3$, and let $\varepsilon > e^{-\delta^2 k}$. Let $A \subset [n]$ be chosen uniformly at random from the sets with $|A| = k$ and $|A + A| \leqslant \lambda k$. Then there exists an arithmetic progression P with*

$$A \subset P \qquad \text{and} \qquad |P| \leqslant \frac{\lambda k}{2} + c(\lambda, \varepsilon)$$

with probability at least $1 - \varepsilon$.

There is only one piece still missing in the proof of Theorem 6.9.1: a lower bound on the size of the set

$$\Lambda = \big\{A \subset [n] \,:\, |A| = k \text{ and } |A + A| \leqslant \lambda k\big\}.$$

The following very simple bound will suffice for our current purposes; a stronger lower bound (at least, for large λ) will be proved in Section 6.10.

Lemma 6.9.2. *Let $\lambda \geqslant 3$ and $n, k \in \mathbb{N}$, with $\lambda k \leqslant n$. Then*

$$\big|\big\{A \subset [n] : |A| = k,\, |A+A| \leqslant \lambda k\big\}\big| \geqslant \frac{1}{\lambda^3} \cdot \frac{n^2}{k}\binom{\lambda k/2}{k}.$$

Proof. We consider, for each arithmetic progression P of length $\lambda k/2$ in $[n]$, all subsets $A \subset P$ of size k containing both endpoints of P. All of these sets are distinct, and all satisfy $|A+A| \leqslant \lambda k$. There are at least $n^2/2\lambda k$ choices for the arithmetic progression, and therefore

$$|\Lambda| \geqslant \frac{n^2}{2\lambda k}\binom{\lambda k/2 - 2}{k-2} \geqslant \frac{n^2}{\lambda^3 k}\binom{\lambda k/2}{k},$$

as claimed, where the final step follows since $\binom{a}{b} = \frac{a(a-1)}{b(b-1)}\binom{a-2}{b-2}$. □

We can now deduce Theorem 6.9.1 from Lemmas 6.5.1, 6.6.1, 6.7.1 and 6.9.2.

Proof of Theorem 6.9.1. For simplicity, we will assume that $\lambda k \leqslant n$; the case $\lambda k > n$ is dealt with in [32, Appendix C]. Recall from (6.22) that Λ^* denotes the collection of sets $A \in \Lambda$ that are contained in an arithmetic progression of length $\lambda k/2 + c(\lambda, \varepsilon)$. Observe first that, by Lemma 6.5.1 and our assumption that $\varepsilon > e^{-\delta^2 k}$, we have

$$|\Lambda \setminus \Lambda^*| \leqslant \frac{n^2}{k} \cdot |\mathcal{I}| + \exp\left(-\frac{\delta k}{2^{10}\lambda}\right)\binom{\lambda k/2}{k} \leqslant \frac{n^2}{k} \cdot |\mathcal{I}| + \frac{\varepsilon}{2\lambda^3}\binom{\lambda k/2}{k}.$$

Now, by Lemmas 6.6.1 and 6.7.1, and recalling that $\mathcal{S} \cup \mathcal{D} = \mathcal{I}$, we have

$$|\mathcal{I}| = |\mathcal{S}| + |\mathcal{D}| \leqslant 2 \cdot \exp\left(-\frac{c(\lambda,\varepsilon)}{2^{20}\lambda^2}\right)\binom{\lambda k/2}{k} \leqslant \frac{\varepsilon}{2\lambda^3}\binom{\lambda k/2}{k}$$

since $c(\lambda, \varepsilon) = 2^{20}\lambda^2 \log(1/\varepsilon) + 2^{560}\lambda^{32}$. By Lemma 6.9.2, it follows that

$$|\Lambda \setminus \Lambda^*| \leqslant \frac{\varepsilon}{\lambda^3} \cdot \frac{n^2}{k}\binom{\lambda k/2}{k} \leqslant \varepsilon|\Lambda|,$$

as required. □

When $\lambda \in (2, 3)$, the proof of Theorem 6.9.1 implies the following weaker bound.

Theorem 6.9.3. *For each $\gamma > 0$, there exists a constant $C(\gamma) > 0$ such that the following holds. Let $2 + \gamma \leqslant \lambda \leqslant 3$ and $\varepsilon > 0$ be fixed, let n be sufficiently large, and let $k \geqslant (\log n)^4$. If $A \subset [n]$ is chosen uniformly at random from those sets with $|A| = k$ and $|A+A| \leqslant \lambda k$, then there exists an arithmetic progression P with*

$$A \subset P \qquad \text{and} \qquad |P| \leqslant \frac{\lambda k}{2} + C(\gamma)\log(1/\varepsilon)$$

with probability at least $1 - 2\varepsilon$.

Theorem 6.9.3 follows by repeating the (entire) proof of Theorem 6.9.1, replacing (everywhere) the condition $\lambda \geqslant 3$ by the condition $\lambda \geqslant 2 + \gamma$, and the conditions $r(A) \geqslant c(\lambda, \varepsilon)$ and $k \geqslant 2^{400}\lambda^{25}(\log n)^3$ by the conditions $r(A) \geqslant C(\gamma)\log(1/\varepsilon)$ and $k \geqslant (\log n)^4$. We leave the details to the reader.

To finish the section, let us quickly deduce Corollary 6.1.2.

Proof of Corollary 6.1.2. The lower bound follows from Lemma 6.9.2 (see also Proposition 6.10.2, below), so it remains to prove the upper bound. To do so, note that (by increasing the implicit constant in the upper bound if necessary) we may assume that $\log n \geqslant 2^{320}\lambda^{25}$, and hence we may apply Theorem 6.9.1 with $\varepsilon := 1/2$. Since there are at most n^2/k arithmetic progressions of length $\lambda k/2 + c(\lambda, \varepsilon)$, it follows that

$$|\Lambda| \leqslant \frac{2n^2}{k}\binom{\lambda k/2 + c(\lambda,\varepsilon)}{k} \leqslant \exp\left(\frac{2c(\lambda,\varepsilon)}{\lambda}\right)\frac{n^2}{k}\binom{\lambda k/2}{k} \leqslant \exp\big(c(\lambda,\varepsilon)\big)\cdot\frac{n^2}{k}\binom{\lambda k/2}{k},$$

as required. □

6.10 The lower bounds

In this section, we prove lower bounds for the size of Λ, and for the typical size of the smallest arithmetic progression containing a set $A \in \Lambda$. The bounds we obtain indicate that the upper bounds in Theorem 6.1.1 and Corollary 6.1.2 are not far from best possible. We begin with the construction for the typical structure, which is very simple.

Proposition 6.10.1. *Given $\lambda \geqslant 4$, let $\varepsilon > 0$ be sufficiently small, and let $n, k \in \mathbb{N}$ be sufficiently large. If $A \subset [n]$ is chosen uniformly at random from the sets with $|A| = k$ and $|A + A| \leqslant \lambda k$, then with probability at least ε,*

$$|P| \geqslant \frac{\lambda k}{2} + 2^{-6}\lambda^2\log(1/\varepsilon)$$

for every arithmetic progression P containing A.

Proof. Set $r := 2^{-6}\lambda^2\log(1/\varepsilon)$, and consider the family $\mathcal{A}(r)$ of sets $A = A' \cup \{0, v\}$, where $A' \subset [\lambda k/2 - 8r/\lambda]$ with $|A'| = k - 2$, and $v = \lambda k/2 + r$. We claim that most such sets satisfy $|A + A| \leqslant \lambda k$. Indeed, since $A' + A' \subset [\lambda k - 16r/\lambda]$, this holds as long as the set $\{x \in A' : x > \lambda k/2 - r - 16r/\lambda\}$ has at most $16r/\lambda$ elements. If $k \geqslant 16r/\lambda$, then the expected number of elements of this set is

$$\frac{k-2}{\lambda k/2 - 8r/\lambda}\cdot\left(r + \frac{8r}{\lambda}\right) < \frac{2(\lambda+8)}{\lambda - 1}\cdot\frac{r}{\lambda} \leqslant \frac{8r}{\lambda},$$

since $\lambda \geqslant 4$, and it follows by Markov's inequality that $|A+A| \leqslant \lambda k$ with probability at least $1/2$, as claimed. Now, observe that

$$|\mathcal{A}(r)| = \binom{\lambda k/2 - 8r/\lambda}{k-2} \geqslant \frac{2}{\lambda^2} \exp\left(-\frac{16r}{\lambda(\lambda-1)}\right)\binom{\lambda k/2}{k} \geqslant \frac{\sqrt{\varepsilon}}{\lambda^2}\binom{\lambda k/2}{k},$$

where the first inequality follows from the binomial inequalities

$$\binom{a}{b-2} \geqslant \frac{b^2}{2a^2}\binom{a}{b} \qquad \text{and} \qquad \binom{a-c}{b} \geqslant \left(1-\frac{b}{a-c}\right)^c \binom{a}{b},$$

again using the bound $k \geqslant 16r/\lambda$, and the second follows since $r \leqslant 2^{-5}\lambda(\lambda-1)\log(1/\varepsilon)$. Now, for each $a \in [n/\lambda k]$ and $b \in [n/4]$, and each set A as above, we apply the linear map $x \mapsto ax+b$ to A. We obtain at least

$$\left\lfloor \frac{n}{\lambda k} \right\rfloor \cdot \frac{n}{4} \cdot \frac{1}{2} \cdot \frac{\sqrt{\varepsilon}}{\lambda^2}\binom{\lambda k/2}{k} \geqslant \varepsilon^{2/3} \cdot \frac{n^2}{k}\binom{\lambda k/2}{k} \tag{6.58}$$

distinct sets $A \subset [n]$ with $|A| = k$ and $|A+A| \leqslant \lambda k$.

Finally, note that few of these sets A are contained in a shorter arithmetic progression, since such an arithmetic progression would have length at most $\lambda k/4 + r/2 < \lambda k/3$. Recalling the upper bound on $|\Lambda|$ given by Corollary 6.1.2, and that ε was chosen sufficiently small, it follows that the right-hand side of (6.58) is at least $\varepsilon|\Lambda|$, as required. □

Obtaining our lower bound on the size of $|\Lambda|$ will be slightly more delicate.

Proposition 6.10.2. *If $\lambda \geqslant 2^{30}$ and $n, k \in \mathbb{N}$ are sufficiently large, then*

$$\left|\left\{A \subset [n] : |A| = k,\, |A+A| \leqslant \lambda k\right\}\right| \geqslant \exp\left(2^{-8}\lambda^{1/2}\right)\frac{n^2}{k}\binom{\lambda k/2}{k}. \tag{6.59}$$

In the proof of Lemma 6.10.2, we will need the following simple bound on the number of independent sets of a given size in a graph.

Lemma 6.10.3. *Let G be a graph with n vertices, m edges and ℓ loops. Let R be a uniformly chosen random subset of k vertices, where $k \leqslant \lfloor n/2 \rfloor$. If $\mathcal{B}$ is the event that R is an independent set, then*

$$\mathbb{P}(\mathcal{B}) \geqslant \exp\left(-\frac{9mk^2}{2n^2} - \frac{3\ell k}{n}\right) - \exp\left(-\frac{k}{16}\right).$$

Lemma 6.10.3 is an almost immediate consequence of the FKG inequality for the hypergeometric distribution, see, e.g., [17, Lemma 3.2].

Lemma 6.10.4 (Hypergeometric FKG Inequality)**.** *Suppose that $\{B_i\}_{i\in I}$ is a family of subsets of an n-element set Ω. Let $t \in \{0, \dots, \lfloor n/2 \rfloor\}$, let R be the uniformly chosen random t-subset*

of Ω, and let $\mathcal{B}$ denote the event that $B_i \not\subseteq R$ for all $i \in I$. Then for every $\eta \in (0,1)$,

$$\mathbb{P}(\mathcal{B}) \geqslant \prod_{i\in I}\left(1 - \left(\frac{(1+\eta)t}{n}\right)^{|B_i|}\right) - \exp\big(-\eta^2 t/4\big).$$

Proof of Lemma 6.10.3. The claimed bound follows immediately from Lemma 6.10.4, applied with $t = k$ and $\eta = 1/2$, and with the sets B_i being the edges and loops of G, using the fact that $1 - x \geqslant e^{-2x}$ for $0 \leqslant x \leqslant 3/4$. □

Proof of Proposition 6.10.2. Set $c := 2^{-8}$ and $r := 2c\lambda^{3/2}$. We will first prove that there are at least $\exp\big(2c\lambda^{1/2}\big)\binom{\lambda k/2}{k}$ subsets $A \subset [\lambda k/2 + r]$ of size k with $|A + A| \leqslant \lambda k$, each containing the endpoints 1 and $\lambda k/2 + r$. Since this bound can be applied in each of the (at least) $n^2/4\lambda k$ arithmetic progressions of length $\lambda k/2 + r$ in $[n]$, and since the sets A obtained for different arithmetic progressions are distinct, it will follow that

$$|\Lambda| \geqslant \frac{n^2}{4\lambda k} \cdot \exp\big(2c\lambda^{1/2}\big)\binom{\lambda k/2}{k} \geqslant \exp\big(c\lambda^{1/2}\big)\frac{n^2}{k}\binom{\lambda k/2}{k},$$

as required.

To prove the claimed bound, let R be a uniformly chosen subset of $[2, \lambda k/2 + r - 1]$ with exactly $k - 2$ elements, and set $A := R \cup \{1, \lambda k/2 + r\}$. Observe first that

$$\binom{\lambda k/2 + r - 2}{k-2} \geqslant \frac{1}{\lambda^2}\left(\frac{\lambda k + 2r}{\lambda k}\right)^k \binom{\lambda k/2}{k} \geqslant \exp\big(3c\lambda^{1/2}\big)\binom{\lambda k/2}{k}, \tag{6.60}$$

where the first inequality holds since $\binom{a}{b} = \frac{a(a-1)}{b(b-1)}\binom{a-2}{b-2}$ and using (6.19), and the second follows since $r = 2c\lambda^{3/2}$, and because λ and k were chosen sufficiently large.

It will therefore suffice to prove that $|A + A| \leqslant \lambda k$ with probability at least $\exp\big(-c\lambda^{1/2}\big)$. To do so, define

$$A' := \big\{x \in A : x \leqslant \lambda k/2 - r\big\} \qquad \text{and} \qquad B := \big\{x \in A : x > \lambda k/2 - r\big\},$$

and set $b := 16c\lambda^{1/2}$. Observe that $\mathbb{E}[|B|] \leqslant 4r/\lambda = b/2$, and hence

$$\mathbb{P}\big(|B + B| \geqslant b^2\big) \leqslant \mathbb{P}\big(|B| \geqslant b\big) \leqslant \exp\big(-c\lambda^{1/2}\big), \tag{6.61}$$

by Hoeffding's inequality. We claim that, setting $X := [\lambda k - 2r + 1, \lambda k - 2r + b^2]$, we have

$$\mathbb{P}\big((A' + B) \cap X = \emptyset\big) \geqslant 2 \cdot \exp\big(-c\lambda^{1/2}\big). \tag{6.62}$$

Before proving (6.62), observe that, together with (6.60) and (6.61), it will suffice to deduce the proposition. Indeed, if $(A'+B)\cap X=\emptyset$ and $|B+B|\leqslant b^2=|X|$, then

$$|A+A|\leqslant \lambda k-2r+|(A'+B)\setminus[\lambda k-2r]|+|B+B|\leqslant \lambda k,$$

since $A'+A'\subset[\lambda k-2r]$ and $A'+B\subset[\lambda k]$, and noting that $b^2=2^8c^2\lambda\leqslant 4c\lambda^{3/2}=2r$.

To prove (6.62) we will use Lemma 6.10.3. To do so, we define a graph G with vertex set $[\lambda k/2+r]$ and edge set

$$E(G)=\big\{xy\,:\,x\leqslant \lambda k/2-r,\, y>\lambda k/2-r \text{ and } x+y\in X\big\}\cup\big\{x\,:\,x+\lambda k/2+r\in X\big\}.$$

Observe that if R is an independent set in G, then $(A'+B)\cap X=\emptyset$. Note that G has at most $2rb^2\leqslant 2^{10}c^3\lambda^{5/2}$ edges and at most $b^2=2^8c^2\lambda$ loops, and that

$$\frac{9\cdot 2^{10}c^3\lambda^{5/2}k^2}{2(\lambda k/2+r)^2}+\frac{3\cdot 2^8c^2\lambda k}{\lambda k/2+r}\leqslant 2^{15}c^3\lambda^{1/2}+2^{11}c^2\leqslant c\lambda^{1/2}-1,$$

since $c=2^{-8}$ and $\lambda\geqslant 2^{30}$. It follows by Lemma 6.10.3 that

$$\mathbb{P}\big((A'+B)\cap X=\emptyset\big)\geqslant \exp\big(-c\lambda^{1/2}+1\big)-\exp\big(-k/16\big)\geqslant 2\cdot\exp\big(-c\lambda^{1/2}\big)$$

as required, since k is sufficiently large. This completes the proof of Proposition 6.10.2. □

Chapter 7

The number of sumsets of a given size

7.1 Introduction

In this chapter we will consider another natural counting problem in additive combinatorics: how many sumsets of a given size are there in $\mathbb{Z}_n$? Our main result is the following theorem, which provides a sharp bound on the number of sumsets of a given size and doubling constant.

Theorem 7.1.1. *Let n be a prime, and let $m, k \in \mathbb{N}$ with $m \geqslant (2+\sqrt{5})k$ and $k \geqslant (\log n)^4$. There are at most*

$$2^{o(m)} \binom{\frac{m-k}{2}}{k} \tag{7.1}$$

sets of the form $A + A$ for some $A \subset \mathbb{Z}_n$ with $|A| = k$ and $|A + A| = m$.

Observe that (7.1) is maximised with $m/k = \Theta(1)$, and so our main focus will be on sets with bounded doubling. In fact, when $m/k \to \infty$ a stronger bound follows from Theorem 5.1.1, by simply counting all sets A of size k with $|A + A| \leqslant m$.

We will also prove the following (sharp) bound when $m \leqslant (2+\sqrt{5})k$.

Theorem 7.1.2. *Let n be a prime, and let $m, k \in \mathbb{N}$ with $k \geqslant (\log n)^4$ and $m = \lambda k$ for some fixed $2 < \lambda \leqslant 2 + \sqrt{5}$. There are*

$$2^{o(m)} \binom{\frac{1+\sqrt{5}}{2\sqrt{5}}(m-2k)}{\frac{1}{\sqrt{5}}(m-2k)}, \tag{7.2}$$

sets of the form $A + A$ for some $A \subset \mathbb{Z}_n$ with $|A| = k$ and $|A + A| = m$.

In order to deduce an upper bound on the number of sumsets of size m, we simply sum (7.1) and (7.2) over k, noting that the maximum occurs for some $k \in (5m, 6m)$.

The following construction provides a matching lower bound (in the critical range $(2+\sqrt{5})k \leqslant m = O(k)$) when $m - k + 1 \leqslant 2n/3$. Let

$$P = \left\{ x \in \mathbb{Z}_n : \frac{n}{3} < x \leqslant \frac{n}{3} + \frac{m-k+1}{2} \right\},$$

and note for any $\{0\} \subset A \subset \{0\} \cup P$ we have $(A+A) \cap P = A$. Moreover, since $m = O(k)$, there are roughly $\binom{\frac{m-k}{2}}{k}$ such sets A of size k for which $P+P \subset A+A$. Since $(P+P) \cap P = \emptyset$, for each such set we have $|A+A| = |P+P| + |A| = m$.

The construction showing that Theorem 7.1.2 is sharp is slightly more complicated. Set $s := \lfloor (\frac{1+\sqrt{5}}{2\sqrt{5}})(m-2k) \rfloor$ and $t := \lfloor \frac{1}{\sqrt{5}}(m-2k) \rfloor$, and consider k-sets of the form $A = A' \cup Q \cup \{-s\}$, where

$$Q = \big\{ x \in \mathbb{Z}_n : s < x < s + k - t \big\}$$

and $A' \subset P = \big\{ x \in \mathbb{Z}_n : 0 < x \leqslant s \big\}$. Note that there are $\binom{s}{t}$ choices for A', and that if $m \leqslant 2n/3$ then $(A+A) \cap (P-s) = A' - s$, since $2(s+k-t) < n - s$. Moreover, it is easy to check that if $A' + A' = P + P$ then $|A+A| \approx m$.

The rest of the chapter is organised as follows. In Section 7.2 we will deduce the container theorem we need from Theorem 5.4.2 using a removal lemma due to Shao [143], in Section 7.3 we will prove a lemma about counting neighborhoods in graphs, and in Section 7.4 we will put the pieces together and prove Theorems 7.1.1 and 7.1.2.

7.2 A container theorem for sumsets

In this section we prove our main container theorem for sumsets, Theorem 7.2.1, below. Since in this part of the argument we will not require any specific properties of $\mathbb{Z}_n$, we will work in an arbitrary subset X of a general abelian group G.

We will find a relatively small family of "containers" $\mathcal{C}$ such that for all $J \subset X$ with bounded doubling, there is a corresponding triple $(A, B, \Upsilon) \in \mathcal{C}$, where $A \subset X + X$ and $B, \Upsilon \subset X$, with certain useful properties. More precisely, we will have

$$J \subset B \cup \Upsilon \qquad \text{and} \qquad J + J = A \cup \big((J \setminus \Upsilon) + \Upsilon \big),$$

and also (less precisely) Υ is 'small' and $B + B \approx A$.

To see why this is a useful family to consider, it may help to consider the construction of sumsets described in the introduction, where we set P to be an interval of length $\frac{(\lambda-1)k}{2}$, letting $m = \lambda k$, and took all sumsets $J + J$ with $\{0\} \subset J \subset \{0\} \cup P$. In this case one should think of $B = P$ and $A = P + P$, so A is the 'structured' part of the sumset, and $\Upsilon = \{0\}$.

Theorem 7.2.1. *Fix $\lambda \geqslant 1$ and $\delta > 0$, and let $n \in \mathbb{N}$ be sufficiently large. Let $k \in \mathbb{N}$ with $k \geqslant (\log n)^4$, let G be an abelian group, and let $X \subset G$ with $|X| = n$. Then there exists a family $\mathcal{C}$ of triples (A, B, Υ), where $A \subset X + X$ and $B, \Upsilon \subset X$, with*

$$|\mathcal{C}| \leqslant \exp\big(3\delta k\big), \tag{7.3}$$

such that

(a) For all $J \subset X$ with $|J| = k$ and $|J + J| = \lambda k$, there exists $(A, B, \Upsilon) \in \mathcal{C}$ such that

$$J + J = A \cup \big((J \setminus \Upsilon) + \Upsilon\big) \qquad \text{and} \qquad \Upsilon \subset J \subset B \cup \Upsilon.$$

(b) For all $(A, B, \Upsilon) \in \mathcal{C}$ we have $|(B + B) \setminus A| \leqslant \delta k$ and $|\Upsilon| \leqslant \delta k$.

Roughly speaking, Property (b) says that $B + B$ is a good approximation for the structured part A of the sumset $J + J$. In particular, this will allow us to deduce that $J + J$ has a large structured part. Property (a) tells us that the part of $J + J$ that is not structured comes from the sum of J with a small fixed set Υ.

We will deduce Theorem 7.2.1 from Theorem 5.4.2 which we restate here for convenience.

Theorem 7.2.2 (Theorem 5.4.2)**.** *Let m, n be integers with $m \geqslant (\log n)^2$, let G be an abelian group, let $X \subset G$ with $|X| = n$, and let $0 < \epsilon < \frac{1}{4}$. There exists a family $\mathcal{A} \subset 2^{X+X} \times 2^X$ of size*

$$|\mathcal{A}| \leqslant \exp\left(2^{16}\frac{1}{\epsilon^2}\sqrt{m}(\log n)^{3/2}\right) \tag{7.4}$$

such that:

(i) For every set $J \subset X$ with $|J + J| \leqslant m$, there exists $(C, D) \in \mathcal{A}$ such that

$$C \subset J + J \qquad \text{and} \qquad J \subset D.$$

(ii) For every $(C, D) \in \mathcal{A}$ we have $|C| \leqslant m$, and moreover either $|D| \leqslant \frac{m}{\log n}$ or there are at most $\epsilon^2|D|^2$ pairs $(d_1, d_2) \in D \times D$ such that $d_1 + d_2 \notin C$.

One might initially think that $D + D \approx C$ for every $(C, D) \in \mathcal{A}$, since for almost all pairs $(d_1, d_2) \in D \times D$ we have $d_1 + d_2 \in C$. However, notice that by adding a few random elements to D, we can make $D + D$ much larger than C without having many bad pairs. The point of Theorem 7.2.1 is that we can remove these 'unstructured' elements and put them in a set Υ, so that for $B = D \setminus \Upsilon$ we have $B + B \approx C$. The following theorem of Shao [143] will be the fundamental tool that allows us to make this intuition precise.

Theorem 7.2.3. *Let G be an abelian group, and let $D \subset G$. Let $K \geqslant 1$ and $\epsilon > 0$, and let $\Gamma \subset D \times D$ with $|\Gamma| \geqslant (1-\epsilon^2)|D|^2$. If $|D +_\Gamma D| \leqslant K|D|$, then there exists $\delta := \delta(\epsilon, K) = o_{\epsilon \to 0}(1)$ and a subset $B \subset D$ such that*

$$|B| \geqslant (1 - \delta)|D| \qquad \text{and} \qquad |(B + B) \setminus (D +_\Gamma D)| \leqslant \delta|D|.$$

We are now ready to prove Theorem 7.2.1.

Proof of Theorem 7.2.1. Let $\mathcal{A}$ be the family given by Theorem 5.4.2 applied with $m = \lambda k$ and

$$\epsilon := \left(\frac{2^{32}\lambda(\log n)^3}{\delta^2 k}\right)^{1/4} \leqslant (\log n)^{-1/8},$$

by choosing n large enough with respect to λ and δ. Notice that, with this choice of parameters,

$$|\mathcal{A}| \leqslant \exp\left(2^{16}\frac{1}{\epsilon^2}\sqrt{m}(\log n)^{3/2}\right) = \exp(\delta k). \tag{7.5}$$

Now let $J \subset X$ with $|J| = k$ and $|J + J| = \lambda k$. By Property (i) of Theorem 5.4.2, there exists $(C, D) \in \mathcal{A}$ such that $C \subset J + J$ and $J \subset D$. Observe that

$$k = |J| \leqslant |D| \leqslant 2\lambda k, \tag{7.6}$$

since for each $d_1 \in D$ there are at least $|D| - |C|$ choices for $d_2 \in D$ such that $d_1 + d_2 \notin C$, and thus if $|D| \geqslant 2\lambda k = 2|J + J| \geqslant 2|C|$, then there would be at least $\frac{1}{2}|D|^2$ pairs $(d_1, d_2) \in D^2$ with $d_1 + d_2 \notin C$, contradicting Property (ii) of Theorem 5.4.2.

We apply Theorem 7.2.3 to D, with $K = \lambda$ and

$$\Gamma = \big\{(d_1, d_2) \in D \times D : d_1 + d_2 \in C\big\}.$$

Notice that, by (7.6) and Properties (i) and (ii) of Theorem 5.4.2, we have

$$|\Gamma| \geqslant (1 - \epsilon^2)|D|^2 \qquad \text{and} \qquad |D +_\Gamma D| \leqslant |C| \leqslant |J + J| = \lambda k \leqslant \lambda |D|,$$

so the conditions of the theorem hold. It follows that there exists a subset $B \subset D$ such that

$$|B| \geqslant \left(1 - \frac{\delta}{2\lambda}\right)|D| \qquad \text{and} \qquad |(B + B) \setminus C| \leqslant \frac{\delta}{2\lambda}|D|, \tag{7.7}$$

since $\epsilon \to 0$ as $n \to \infty$, and we chose n large with respect to δ and λ.

Now, let $\mathcal{A}' \subset \mathcal{A}$ denote the collection of $(C, D) \in \mathcal{A}$ as above, i.e., that are the container of some set $J \subset X$ with $|J| = k$ and $|J + J| = \lambda k$. For each $(C, D) \in \mathcal{A}'$ denote by $B(C, D)$ the set $B \subset D$ satisfying (7.7) given by Theorem 7.2.3, define

$$\mathcal{C}(C, D) := \Big\{(A, B, \Upsilon) \,:\, B = B(C, D),\ \Upsilon \subset D \setminus B,\ \text{and}\ A = T \cup (\Upsilon + \Upsilon)\ \text{for some}\ (B + B) \cap C \subset T \subset B + B\Big\},$$

and set

$$\mathcal{C} := \bigcup_{(C,D)\in\mathcal{A}'} \mathcal{C}(C, D).$$

In order to prove that $\mathcal{C}$ has the required properties, we will first verify Property (b), then deduce the bound (7.3) on the size of $\mathcal{C}$, and finally show that Property (a) holds.

To show that Property (b) holds, let $(C,D) \in \mathcal{A}'$ and $(A,B,\Upsilon) \in \mathcal{C}(C,D)$. Note first that, by the definition of $\mathcal{C}(C,D)$, there exists a set T such that $(B+B)\cap C \subset T \subset A$. Using (7.6) and (7.7), it follows that

$$|(B+B)\setminus A| \leqslant |(B+B)\setminus C| \leqslant \frac{\delta}{2\lambda}|D| \leqslant \delta k.$$

Similarly, recall that $\Upsilon \subset D\setminus B$, by the definition of $\mathcal{C}(C,D)$, and that $B \subset D$. Again using (7.6) and (7.7), it follows that

$$|\Upsilon| \leqslant |D\setminus B| = |D| - |B| \leqslant \frac{\delta}{2\lambda}|D| \leqslant \delta k.$$

Having verified Property (b), we can now bound the size of $\mathcal{C}$. To do so, recall that $\mathcal{A}' \subset \mathcal{A}$, and thus by (7.5) we have $|\mathcal{A}'| \leqslant |\mathcal{A}| \leqslant \exp(\delta k)$. Now let $(C,D) \in \mathcal{A}'$, and note that B is uniquely determined by (C,D), that there are at most $2^{\delta k}$ choices for Υ, since $\Upsilon \subset D\setminus B$ and $|D\setminus B| \leqslant \delta k$, and that there are at most $2^{\delta k}$ choices for T, since $(B+B)\cap C \subset T \subset B+B$ and $|(B+B)\setminus C| \leqslant \delta k$. It follows that

$$|\mathcal{C}| \leqslant \sum_{(C,D)\in\mathcal{A}'} |\mathcal{C}(C,D)| \leqslant \exp(\delta k)\cdot 2^{2\delta k} \leqslant \exp(3\delta k).$$

It remains to show Property (a) holds, so let $J \subset X$ with $|J| = k$ and $|J+J| = \lambda k$, let $(C,D) \in \mathcal{A}$ be the container of J (so $C \subset J+J$ and $J \subset D$), and set $B = B(C,D)$. We claim that if

$$\Upsilon = J\setminus B \qquad \text{and} \qquad A = \big((B+B)\cap(J+J)\big)\cup\big(\Upsilon+\Upsilon\big), \tag{7.8}$$

then $(A,B,\Upsilon) \in \mathcal{C}(C,D)$, and moreover that

$$J+J = A\cup\big((J\setminus\Upsilon)+\Upsilon\big) \qquad \text{and} \qquad \Upsilon \subset J \subset B\cup\Upsilon,$$

as required. First, observe that since $J \subset D$ and $C \subset J+J$, we have

$$J\setminus B \subset D\setminus B \qquad \text{and} \qquad (B+B)\cap C \subset (B+B)\cap(J+J) \subset B+B.$$

Hence, setting $T = (B+B)\cap(J+J)$, it follows that $(A,B,\Upsilon) \in \mathcal{C}(C,D)$. Next, note that $\Upsilon \subset J \subset B\cup\Upsilon$ follows immediately from the definition of $\Upsilon = J\setminus B$. Finally, observe that $J\setminus\Upsilon = B\cap J$, so

$$\begin{aligned} J+J &= \big((J\setminus\Upsilon)+(J\setminus\Upsilon)\big)\cup\big(\Upsilon+\Upsilon\big)\cup\big((J\setminus\Upsilon)+\Upsilon\big) \\ &\subset \big((B+B)\cap(J+J)\big)\cup\big(\Upsilon+\Upsilon\big)\cup\big((J\setminus\Upsilon)+\Upsilon\big) = A\cup\big((J\setminus\Upsilon)+\Upsilon\big). \end{aligned}$$

Since $\Upsilon \subset J$, it follows from (7.8) that in fact $J+J = A\cup\big((J\setminus\Upsilon)+\Upsilon\big)$. This proves Property (a), and hence completes the proof of the theorem. □

7.3 Counting Neighborhoods

For $k \in \mathbb{N}$ and $m \geqslant 2k$ the problem we are interested in is essentially to bound the size of the following family

$$\mathcal{S}_{m,k} = \left\{ I \in \binom{\mathbb{Z}_n}{m} : \exists\, J \text{ with } I = J + J,\ |J| = k \right\}. \tag{7.9}$$

Theorem 7.2.1 provides us with a collection of containers (A, B, Υ) for this family of sets. For any triple of sets $A, B, \Upsilon \subset \mathbb{Z}_n$ define the family

$$\mathcal{T}_{m,k}(A, B, \Upsilon) = \left\{ I \in \binom{\mathbb{Z}_n}{m} : \exists\, S \subset B \text{ with } I = A \cup (\Upsilon + S),\ |S| = k - |\Upsilon| \right\}, \tag{7.10}$$

and observe that if $\mathcal{C}$ is the family of containers given by Theorem 7.2.1, and $(A, B, \Upsilon) \in \mathcal{C}$ is the container corresponding to J, then $J + J \in \mathcal{T}_{\lambda k,k}(A, B, \Upsilon)$. Since the number of containers is $e^{o(k)}$, we are thus left with the task of bounding $|\mathcal{T}_{m,k}(A, B, \Upsilon)|$ for each container $(A, B, \Upsilon) \in \mathcal{C}$. The aim of this section is to provide suitable bounds.

We begin with a simple observation, which will suffice when $|B| \leqslant \frac{(\lambda-1)k}{2}$.

Lemma 7.3.1. *Let n be a prime, and let k and m be integers. For any $A, B, \Upsilon \subset \mathbb{Z}_n$,*

$$|\mathcal{T}_{m,k}(A, B, \Upsilon)| \leqslant \binom{|B|}{k - |\Upsilon|}$$

Proof. We only need to count the number of choices for $S \subset B$ with $|S| = k - |\Upsilon|$, since every element of $\mathcal{T}_{m,k}(A, B, \Upsilon)$ is of the form $A \cup (\Upsilon + S)$, where A and Υ are fixed. □

Proving a sufficiently strong bound on the size of $\mathcal{T}_{m,k}(A, B, \Upsilon)$ when $|B| \geqslant \frac{(\lambda-1)k}{2}$ will not be quite so easy. To do so, we'll prove a general lemma about counting neighborhoods in graphs (Lemma 7.3.2, below), and apply it to the auxiliary bipartite graph

$$F(A, B, \Upsilon) := \big\{ (x, y) \in (\mathbb{Z}_n \setminus A) \times B : x - y \in \Upsilon \big\}. \tag{7.11}$$

To motivate this definition, observe that each set $I' := (\Upsilon + S) \setminus A$ that we need to count is the neighborhood in the graph $F = F(A, B, \Upsilon)$ of some set $S \subset B$, i.e., a set of the form

$$N_F(S) := \bigcup_{x \in S} N_F(x).$$

Our main lemma for counting neighborhoods in graphs is as follows.

Lemma 7.3.2. *Let $t, \ell \in \mathbb{N}$, let F be a graph, and let $B \subset V(F)$ be an independent set. Then there are at most*

$$2 \cdot \binom{|B| + t - \ell}{t}$$

sets of the form $N_F(S)$ with $S \subset B$, $|S| = \ell$, and $|N_F(S)| \leqslant t$.

When $t \leqslant \ell$, Lemma 7.3.2 is tight up to a factor of 2, since if F is a matching with $|B| + t - \ell$ edges, and B contains one endpoint of each edge of F, then every set S as in the lemma contains exactly t vertices of degree 1, and all $\ell - t$ vertices in B of degree 0. Since each such set has a distinct neighborhood, it follows that in this case there are exactly $\binom{|B|+t-\ell}{t}$ sets of the form $N_F(S)$ with S as described. We remark also that the slightly weaker[1] bound

$$\sum_{s=0}^{t} \binom{|B|}{s} \tag{7.12}$$

follows easily from the observation that if $|N_F(S)| \leqslant t$, then there exists a subset $S' \subset S$ such that $|S'| \leqslant t$ and $N_F(S') = N_F(S)$. It follows that the number of sets $S' \subset B$ with $|S'| \leqslant t$ is an upper bound on the number of neighborhoods $N_F(S)$ with $S \subset B$ and $|N_F(S)| \leqslant t$. This simple bound is actually enough to prove Theorem 7.1.1 when $\lambda \geqslant 5$, but for $\lambda < 5$ we will need the stronger bound given by Lemma 7.3.2.

The basic idea of our proof of Lemma 7.3.2 is quite simple: for each set S as in the lemma, we will carefully choose a subset $S' \subset S$ of size at most t that 'encodes' the neighborhood of S, and then count the sets S' formed via this process. Our choice of S' is inspired by the proof of the graph container lemma: in each step we select a 'maximum-degree' vertex v_i and reveal whether or not it is in S. If $v_i \in S$, then we remove its neighborhood from the graph. Crucially, the choice of v_i will only depend on the set $S \cap \{v_1, \dots, v_{i-1}\}$.

The key observation about this process is given by the following simple lemma, which will allow us to bound the number of sets S' of a given size that can be produced by the algorithm in a given number of steps.

Lemma 7.3.3. *Let B be a finite set, and let $f \colon 2^B \times \mathbb{N} \to B$ be an arbitrary function. For each set $S \subset B$, define a sequence $(v_{i,S})_{i \in \mathbb{N}}$ of elements of B by setting*

$$S_{i-1} := S \cap \{v_{1,S}, \dots, v_{i-1,S}\} \qquad \text{and} \qquad v_{i,S} := f(S_{i-1}, i)$$

for each $i \in \mathbb{N}$. Then

$$\big|\{S_a : S \subset B, \, |S_a| = b\}\big| \leqslant \binom{a}{b}$$

for every $a, b \in \mathbb{N}$.

Proof. We claim that, for each $S \subset B$ and $i \in \mathbb{N}$, the set S_i is uniquely defined by the set of indices $\{j \in [i] : v_{j,S} \in S_i\}$. Since when $|S_a| = b$ the number of choices for the set of indices is exactly $\binom{a}{b}$, this will suffice to prove the lemma.

To prove the claim, simply note that it holds for $i = 0$ (since S_0 is always the empty set), and that the choice of $v_{i,S}$ depends only on S_{i-1}. We may therefore use induction on i, and deduce

[1] In fact, if $t - \ell$ is sufficiently large then this bound is slightly stronger than that given by the lemma; however, our main interest will be in the case $t \leqslant \ell$.

that if two sets S and T satisfy

$$\big\{j \in [i] : v_{j,S} \in S_i\big\} = \big\{j \in [i] : v_{j,T} \in T_i\big\},$$

then $S_{i-1} = T_{i-1}$ (by the induction hypothesis), $v_{i,S} \in S$ if and only if $v_{i,T} \in T$, and moreover $v_{i,S} = v_{i,T}$. Hence $S_i = T_i$, as claimed. □

We are now ready to prove Lemma 7.3.2.

Proof of Lemma 7.3.2. Fix a set $S \subset B$ with $|S| = \ell$ and $|N_F(S)| \leqslant t$. We will define a sequence $(v_1, \dots, v_{|B|})$ of elements of B, and an auxiliary sequence $(A_i, B_i, F_i)_{i=1}^{|B|+1}$, where $A_i \subset V(F)$, $B_i \subset B$ and $F_i \subset F$ for each i, as follows:

(*a*) Set $A_1 := V(F)$, $B_1 := B$ and $F_1 := F$.

Now, for each $1 \leqslant i \leqslant |B|$, if we have already defined $(v_1, \dots, v_{i-1})$ and (A_i, B_i, F_i), then:

(*b*) Choose $v_i \in B_i$ with $d_{F_i}(v_i)$ (the degree of v_i in F_i) maximal.[2]

(*c*) If $v_i \in S$, then set

$$A_{i+1} := A_i \setminus N_{F_i}(v_i), \quad B_{i+1} := B_i \setminus \{v_i\} \quad \text{and} \quad F_{i+1} := F[A_{i+1}].$$

If $v_i \notin S$, then set $A_{i+1} := A_i$, $B_{i+1} := B_i \setminus \{v_i\}$ and $F_{i+1} := F_i$.

Note that the choice of v_i is determined by (A_i, B_i, F_i), and that $(A_{i+1}, B_{i+1}, F_{i+1})$ is determined by (A_i, B_i, F_i) and the event $\{v_i \in S\}$. It follows that for each $1 \leqslant i \leqslant |B|$, there exists a function $f_i \colon 2^B \to B$ such that $v_i = f_i(S_{i-1})$, where $S_i = S \cap \{v_1, \dots, v_i\}$. Crucially, observe that these functions do not depend on S.

The next observation we will need is that

$$N_F(S_i) = \bigcup_{\substack{v_j \in S \\ j \leqslant i}} N_{F_j}(v_j) \tag{7.13}$$

for every $i \in \mathbb{N}$, where $S_i := S \cap \{v_1, \dots, v_i\}$. One direction is trivial, since $F_j \subset F$ for every $j \in \mathbb{N}$, and the other direction holds because $F_j = F[A_j]$, and any vertex in $V(F) \setminus A_j$ must be contained in $N_{F_{j'}}(v_{j'})$ for some $j' < j$ such that $v_{j'} \in S$. Note also that the union in (7.13) is in fact a disjoint union, that is

$$N_{F_j}(v_j) \cap N_{F_{j'}}(v_{j'}) = \emptyset \tag{7.14}$$

for every $1 \leqslant j < j' \leqslant i$ with $v_j, v_{j'} \in S$, because $v_j \in S$ implies that $N_{F_j}(v_j)$ is removed from $A_{j'}$ in the algorithm, and recalling that $F_{j'} = F[A_{j'}]$.

Now define

$$a = a(S) := \min\big\{i \in \mathbb{N} : N_F(S_i) = N_F(S)\big\},$$

[2] If there is more than one vertex with the maximum degree in B_i, then choose v_i according to some arbitrary (but fixed) rule.

and note that this is well-defined since B is independent, so no vertex of B is ever deleted from A_i, and thus every vertex of B appears in the sequence $(v_1, \dots, v_{|B|})$. We will use Lemma 7.3.3 to bound the number choices for S_a, and hence the number of sets $N_F(S)$. To do so we will need the following bounds on a and the size of S_a.

Claim 7.3.4. *$a \leqslant |B| + |S_a| - \ell$ and $|S_a| \leqslant t$.*

Proof. Observe first that

$$|N_{F_i}(v_i)| \geqslant |N_{F_{i+1}}(v_{i+1})| \tag{7.15}$$

for every $1 \leqslant i < |B|$, since the degree of v_i in F_i is maximal amongst vertices of B_i, and since $F_{i+1} \subset F_i$ and $B_{i+1} \subset B_i$. It follows that $|N_{F_i}(v_i)| \geqslant 1$ for all $i \leqslant a$, and hence, by (7.13) and (7.14), we deduce that $|S_a| \leqslant |N_F(S)| \leqslant t$. To bound a, simply note that

$$|B| - a = |B \setminus \{v_1, v_2 \dots, v_a\}| \geqslant |S \setminus \{v_1, v_2 \dots, v_a\}| = \ell - |S_a|,$$

because $S \subset B$ and $|S| = \ell$. It follows that $a \leqslant |B| + |S_a| - \ell$, as claimed. □

It moreover follows from the proof of Claim 7.3.4 that if $|S_a| = t$ then $d_F(v) \leqslant 1$ for every $v \in S$. It follows that if B has x vertices of degree 0 in F, then the number of choices for $N_F(S)$ with $|S_a| = t$ is at most

$$\sum_{s=\ell-x}^{t} \binom{|B| - x}{s} \leqslant \binom{|B| + t - \ell}{t},$$

with equality when $x \in \{\ell - t, \ell - t + 1\}$. Moreover, by Lemma 7.3.3, the number of sets S_a such that $S \subset B$ and $|S_a| \leqslant t - 1$ is at most

$$\sum_{s=0}^{t-1} \binom{|B| + s - \ell}{s} = \binom{|B| + t - \ell}{t}.$$

Since $N_F(S_a) = N_F(S)$, by the definition of $a = a(S)$, the lemma follows. □

Applying Lemma 7.3.2 to the graph $F(A, B, \Upsilon)$, defined in (7.11), we obtain the following bound on the size of the set $\mathcal{T}_{m,k}(A, B, \Upsilon)$, which was defined in (7.10).

Lemma 7.3.5. *Let n be a prime, and let k, m and t be integers. Then*

$$|\mathcal{T}_{m,k}(A, B, \Upsilon)| \leqslant 2 \cdot \binom{|B| + |\Upsilon| + t - k}{t}$$

for any sets $A, B, \Upsilon \subset \mathbb{Z}_n$ with $|A| \geqslant m - t$.

Proof. Fix $A, B, \Upsilon \subset \mathbb{Z}_n$, with $|A| = m - t + 1$. By the definition (7.10) of $\mathcal{T}_{m,k}(A, B, \Upsilon)$, our task is to bound the number of sets $I' \subset \mathbb{Z}_n$ with $|I'| = m - |A| \leqslant t$ such that $I' = (\Upsilon + S) \setminus A$ for some $S \subset B$ with $|S| = k - |\Upsilon|$. Set $F = F(A, B, \Upsilon)$, and observe that for each such pair (I', S) we have $I' = N_F(S)$. By Lemma 7.3.2, applied with $\ell = k - |\Upsilon|$, it follows that there are at most

$$2 \cdot \binom{|B| + |\Upsilon| + t - k}{t}$$

such sets I', as claimed. □

7.4 Putting together the pieces to count sumsets

Theorems 7.1.1 and 7.1.2 are both fairly straightforward consequences of Theorem 7.2.1 and Lemmas 7.3.1 and 7.3.5. We will also use the following simple fact.

Fact 7.4.1. *For each $x > 0$, the function $f(s) = \binom{\frac{x+s}{2}}{s}$ is maximised with $s = x/\sqrt{5} + O(1)$. Moreover,*

$$\max_s \binom{\frac{x+s}{2}}{s} = \big(1+o(1)\big)\binom{\frac{(1+\sqrt{5})x}{2\sqrt{5}}}{\frac{x}{\sqrt{5}}}$$

as $x \to \infty$.

Proof of Theorems 7.1.1 and 7.1.2. Fix $\lambda \geqslant 2$ and $\delta > 0$, and let n be a sufficiently large prime and $k \geqslant (\log n)^4$. Set $m := \lambda k$, and (recalling (7.9)) observe that in order to prove Theorem 7.1.1, it will suffice to show that

$$|\mathcal{S}_{m,k}| \leqslant (\lambda/\delta)^{4\delta k}\binom{\frac{(\lambda-1)k}{2}}{k}.$$

By Theorem 7.2.1, and recalling the definition (7.10) of $\mathcal{T}_{m,k}(A,B,\Upsilon)$, there exists a family $\mathcal{C}$ of triples (A,B,Υ), where $A \subset X+X$ and $B, \Upsilon \subset X$, such that

$$|\mathcal{S}_{m,k}| \leqslant e^{3\delta k} \max_{(A,B,\Upsilon)\in\mathcal{C}} |\mathcal{T}_{m,k}(A,B,\Upsilon)|.$$

Indeed, property (a) of Theorem 7.2.1 guarantees that for each sumset $J+J \in \mathcal{S}_{m,k}$, there exists $(A,B,\Upsilon) \in \mathcal{C}$ such that $J+J \in \mathcal{T}_{m,k}(A,B,\Upsilon)$.

In order to bound $|\mathcal{T}_{m,k}(A,B,\Upsilon)|$, we will split into two cases, depending on the size of the set B. When $|B| \leqslant \frac{(\lambda-1)k}{2}$, it suffices to use Lemma 7.3.1, which implies that

$$|\mathcal{T}_{m,k}(A,B,\Upsilon)| \leqslant \binom{\frac{(\lambda-1)k}{2}}{k-|\Upsilon|} \leqslant \binom{\lambda k}{\delta k}\binom{\frac{(\lambda-1)k}{2}}{k} \leqslant (\lambda/\delta)^{2\delta k}\binom{\frac{(\lambda-1)k}{2}}{k},$$

where the second inequality follows since $|\Upsilon| \leqslant \delta k$ and $\binom{a}{b-c} \leqslant \binom{a}{b}\binom{a}{c}$.

To deal with the case $|B| > \frac{(\lambda-1)k}{2}$, observe first that, by Lemma 7.3.5, we have

$$|\mathcal{T}_{m,k}(A,B,\Upsilon)| \leqslant 2\cdot\binom{|B|+|\Upsilon|+s-k}{s}, \tag{7.16}$$

where $s = m-|A|$. To bound the right-hand side of (7.16), recall that $|\Upsilon| \leqslant \delta k$, by property (b) of Theorem 7.2.1, and observe that

$$|A| \geqslant |B+B| - \delta k \geqslant 2|B| - 1 - \delta k = m - t, \tag{7.17}$$

where $t := (1+\delta)k - 1$. Indeed, the first inequality follows since $|(B+B)\setminus A| \leqslant \delta k$, the second by Cauchy–Davenport[3], and the third since $|B| > \frac{(\lambda-1)k}{2}$ and $m = \lambda k$.

[3] If $B+B = \mathbb{Z}_n$, then $m \geqslant |A| \geqslant n - \delta k$ (since $A \subset J+J$ and $|(B+B)\setminus A| \leqslant \delta k$), and so there are at most $\binom{n}{\delta k} \leqslant \binom{m/4}{k}$ sumsets $J+J$ of size m in $\mathbb{Z}_n$, as required.

It follows from (7.17) that $s \leqslant t$, and that $|B| \leqslant \frac{|A|+\delta k+1}{2} \leqslant \frac{m-s}{2} + \delta k$, and hence

$$\binom{|B|+|\Upsilon|+s-k}{s} \leqslant \binom{\frac{(\lambda-2)k+s}{2}+2\delta k}{s} \leqslant \binom{\lambda k}{2\delta k}\binom{\frac{(\lambda-2)k+s}{2}}{s} \leqslant (\lambda/\delta)^{3\delta k}\binom{\frac{(\lambda-2)k+s}{2}}{s},$$

since $|\Upsilon| \leqslant \delta k$, and using the inequality $\binom{a+c}{b} \leqslant \binom{a}{b}\binom{a+c}{c}$, which holds for all $a \geqslant b \geqslant c$.

Now, if $\lambda \geqslant 2+\sqrt{5}$ then, by Fact 7.4.1 applied with $x = (\lambda-2)k$, we have

$$\max_{s \leqslant t}\binom{\frac{(\lambda-2)k+s}{2}}{s} \leqslant \lambda^{\delta k}\binom{\frac{(\lambda-1)k+\delta k}{2}}{k} \leqslant (\lambda/\delta)^{2\delta k}\binom{\frac{(\lambda-1)k}{2}}{k},$$

as required. On the other hand, if $2 \leqslant \lambda \leqslant 2+\sqrt{5}$, then

$$\max_{s \leqslant t}\binom{\frac{(\lambda-2)k+s}{2}}{s} \leqslant 2 \cdot \binom{\frac{(1+\sqrt{5})\theta}{2}}{\theta},$$

where $\theta = \frac{(\lambda-2)k}{\sqrt{5}}$, again by Fact 7.4.1, and since k is sufficiently large. Combining the bounds above, we obtain Theorems 7.1.1 and 7.1.2. □

Chapter 8

Towards Hadwiger's conjecture via Bourgain Slicing

This chapter presents joint work with Peter van Hintum, Robert Morris and Marius Tiba. It is adapted from the paper [38] which has been submitted for publication.

8.1 Introduction

Define

$$N(K) = \min\Big\{N \in \mathbb{N} : \exists\, x_1, \dots, x_N \in \mathbb{R}^d \text{ such that } K \subset \bigcup_{i=1}^{N} \big(x_i + \mathrm{int}(K)\big)\Big\}.$$

Hadwiger [78] conjectured in 1957 that $N(K) \leqslant 2^d$ for all convex $K \subset \mathbb{R}^d$. Note that this bound is attained by the cube $[0,1]^d$. The conjecture was proved when $d \leqslant 2$ by Levy [99] in 1955, but for over 60 years the best known bound for general d was

$$N(K) \;\leqslant\; \big(d\log d + d\log\log d + 5d\big)\binom{2d}{d} \;=\; O\big(4^d\sqrt{d}\log d\big).$$

A few years ago, Huang, Slomka, Tkocz and Vritsiou [83] proved that

$$N(K) \;\leqslant\; e^{-\Omega(\sqrt{d})}\cdot 4^d. \tag{8.1}$$

Here we will prove the following almost-exponential improvement of their bound.

Theorem 8.1.1. *If $K \subset \mathbb{R}^d$ is a convex body, then*

$$N(K) \leqslant \exp\left(-\Omega\left(\frac{d}{(\log d)^8}\right)\right)\cdot 4^d$$

as $d \to \infty$.

Given a convex body $K \subset \mathbb{R}^d$, define the *isotropic constant* of K to be

$$L_K = \left(\frac{\sqrt{\det(\Sigma_K)}}{\mathrm{Vol}_d(K)} \right)^{1/d},$$

where $\Sigma_K = \mathbb{E}[X \otimes X]$ is the covariance matrix of the random variable $X \sim \mathrm{Unif}(K)$, that is, X is a uniformly random point of K. A consequence of a theorem of Klartag and Lehec [92] is that for every convex body $K \subset \mathbb{R}^d$ it holds that $L_K = O(\log d)^4$.

Our main result is the following bound on the covering number of a convex body. Since $L_K = O(\log d)^4$, it implies the bound in Theorem 8.1.1 for Hadwiger's conjecture.

Theorem 8.1.2. *If $K \subset \mathbb{R}^d$ is a convex body, then*

$$N(K) \leqslant \exp\left(- \frac{\Omega(d)}{L_K^2} \right) \cdot 4^d$$

as $d \to \infty$.

One of the key innovations of [83] was a method of deducing bounds on the covering number $N(K)$ from bounds on the Kövner–Besicovitch measure of symmetry

$$\Delta_{KB}(K) := \max_{x \in \mathbb{R}^d} \frac{|K \cap (x - K)|}{|K|}.$$

In particular, the authors of [83] improved the (straightforward, but until then best known) lower bound $\Delta_{KB}(K) \geqslant 2^{-d}$ by a factor of $e^{\Omega(\sqrt{d})}$, and used that bound to prove (8.1). We will similarly deduce Theorem 8.1.2 from the following lower bound on $\Delta_{KB}(K)$.

Theorem 8.1.3. *If $K \subset \mathbb{R}^d$ is a convex body, then*

$$\Delta_{KB}(K) \geqslant \exp\left(\frac{d}{2^{15} L_K^2} \right) \cdot 2^{-d}.$$

In addition to the application to Hadwiger's conjecture described above, our method also has an application to the geometry of numbers. To be precise, Ehrhart [44] conjectured in 1964 that a convex body in $\mathbb{R}^d$ centred at the origin[1] whose interior contains no lattice point other than the origin has volume at most $(d+1)^d/d!$ (this bound is attained by a simplex). The best-known upper bound for the volume of K is of the form $e^{-\Omega(\sqrt{d})} \cdot 4^d$, obtained by Huang, Slomka, Tkocz and Vritsiou [83]. We will use the bound on L_K proved by Klartag and Lehec [92] to deduce the following almost-exponential improvement of their bound.

[1] We say that a convex body $K \subset \mathbb{R}^d$ is *centred* at its centre of mass $\mathbb{E}[X]$, where $X \sim \mathcal{U}(K)$.

Theorem 8.1.4. *Let $K \subset \mathbb{R}^d$ be a convex body centred at the origin. If $K \cap \mathbb{Z}^d = \{0\}$, then*

$$|K| \leqslant \exp\left(-\Omega\left(\frac{d}{(\log d)^8}\right)\right) \cdot 4^d$$

as $d \to \infty$.

In order to prove Theorem 8.1.4, we will need a variant of Theorem 8.1.3 that provides a similar lower bound on the ratio $|K \cap (-K)|/|K|$ (see Theorem 8.4.1). The application of such bounds to Ehrhart's conjecture was first observed by Henk, Henze and Hernández Cifre [82], who used the bound $|K \cap (-K)|/|K| \geqslant 2^{-d}$, due to Milman and Pajor [110], together with Minkowski's theorem, to prove an upper bound of 4^d for Ehrhart's conjecture.

The rest of this note is organised as follows. In Section 8.2 we will prove Theorem 8.1.3, in Section 8.3 we will deduce Theorems 8.1.1 and 8.1.2, and in Section 8.4 we will prove Theorem 8.1.4.

8.2 Bounding the Kövner–Besicovitch measure

One of the key ideas introduced in [83] was that a lower bound on $\Delta_{KB}(K)$ can be obtained by considering the maximum density of the random variable $X+Y$, where X and Y are independent uniform elements of K. More precisely, they made the following observation. We write f_X for the probability density function of a random variable X.

Lemma 8.2.1. *Let $K \subset \mathbb{R}^d$ be a convex body of volume 1, and let X and Y be independent uniformly-chosen random elements of K. Then, for any $z \in K$,*

$$f_{\frac{X+Y}{2}}(z) = 2^d \cdot \big|K \cap \big(2z - K\big)\big|.$$

Proof. Observe first that

$$2^{-d} \cdot f_{\frac{X+Y}{2}}(z) \;=\; f_{X+Y}(2z) \;=\; \int_{x \in \mathbb{R}^d} f_X(x) f_Y(2z - x)\, dx.$$

Now simply note that

$$\begin{aligned}\int_{x \in \mathbb{R}^d} f_X(x) f_Y(2z - x)\, dx \;&=\; \int_{x \in \mathbb{R}^d} \mathbb{1}\big[x \in K\big]\, \mathbb{1}\big[2z - x \in K\big]\, dx \\ &=\; \int_{x \in \mathbb{R}^d} \mathbb{1}\big[x \in K \cap (2z - K)\big]\, dx \;=\; |K \cap (2z - K)|,\end{aligned}$$

as claimed. □

It follows immediately from Lemma 8.2.1 that if $|K| = 1$, then

$$\Delta_{KB}(K) \geqslant 2^{-d} \cdot \big\| f_{\frac{X+Y}{2}} \big\|_\infty \geqslant 2^{-d} \cdot \frac{\mathbb{P}\big(\frac{X+Y}{2} \in A\big)}{\mathbb{P}(X \in A)} \tag{8.2}$$

for any measurable set $A \subset \mathbb{R}^d$. In order to prove their lower bound on $\Delta_{KB}(K)$, the authors of [83] observed that the random variable $\big\|\frac{X+Y}{2}\big\|_2$ is typically about $\sqrt{2}$ times smaller than $\|X\|_2$, and applied the inequality (8.2) to a ball A with radius halfway between these two typical values. They then used a 'thin-shell' theorem of Guédon and Milman [76], which implies that if $K \subset \mathbb{R}^d$ is a convex body in isotropic position then, for any fixed $c > 0$,

$$\mathbb{P}\Big(\big|\|X\|_2 - \sqrt{d}\big| \geqslant c\sqrt{d}\Big) \leqslant \exp\Big(-\Omega\big(\sqrt{d}\big)\Big), \tag{8.3}$$

to deduce that $\mathbb{P}\big(\frac{X+Y}{2} \in A\big) \approx 1$ and $\mathbb{P}(X \in A) \leqslant e^{-\Omega(\sqrt{d})}$ for this set A, giving their bound

$$\Delta_{KB}(K) \geqslant e^{\Omega(\sqrt{d})} \cdot 2^{-d}.$$

The Guédon–Milman bound (8.3) is best possible (to see this, consider the simplex), so it may seem at first sight that there is not much hope of using the method of [83] to prove a significantly stronger lower bound on $\Delta_{KB}(K)$. In order to do so, we will replace the thin-shell estimate (8.3) by a 'small-ball' bound which depends on L_K, and the random variable $X + Y$ by a sum of arbitrarily many independent random variables.

To be more precise, let $X_1, X_2, \dots$ be a sequence of independent random variables, each chosen uniformly at random from the set K, and for each $k \in \mathbb{N}$, define

$$S_k := \frac{1}{2^k} \sum_{i=1}^{2^k} X_i. \tag{8.4}$$

Since K is convex, it follows from the Prékopa–Leindler inequality that f_{S_k} is log-concave.

The key step is the following lemma, which bounds $f_{S_k}(z)$ in terms of $f_{\frac{X+Y}{2}}(z)$.

Lemma 8.2.2. *For any convex body $K \subset \mathbb{R}^d$ with volume 1, we have*

$$f_{S_k}(z) \leqslant \Big(f_{\frac{X+Y}{2}}(z)\Big)^{2^k - 1}$$

for all $z \in \mathbb{R}^d$ and every $k \in \mathbb{N}$.

Proof. The proof is by induction on k. Note that the conclusion holds trivially in the case $k = 1$, so let $k \geqslant 1$ and assume that the inequality holds for k; we will prove that it holds for $k + 1$. Define $T_k := 2^{-k} \sum_{i=1}^{2^k} X_{2^k+i}$, and note that $S_{k+1} = \frac{S_k + T_k}{2}$, and that S_k and T_k are independent and identically distributed random variables with support K. It follows that

$$f_{S_{k+1}}(z) = f_{\frac{S_k+T_k}{2}}(z) = 2^d \int_{y \in K} f_{S_k}(y) f_{S_k}(2z - y)\, dy$$

for every $z \in K$. Moreover, since f_X and f_Y are indicator functions on K, and f_{S_k} is a log-concave function supported on K,

$$\int_{y\in K} f_{S_k}(y) f_{S_k}(2z-y)\,dy \;\leqslant\; f_{S_k}(z)^2 \int_{y\in K} f_X(y) f_Y(2z-y)\,dy.$$

Now, by the induction hypothesis, we have

$$f_{S_k}(z) \;\leqslant\; \left(f_{\frac{X+Y}{2}}(z)\right)^{2^k-1},$$

and therefore, noting again that

$$\int_{y\in K} f_X(y) f_Y(2z-y)\,dy \;=\; 2^{-d}\cdot f_{\frac{X+Y}{2}}(z),$$

we obtain

$$f_{S_{k+1}}(z) \;\leqslant\; \left(f_{\frac{X+Y}{2}}(z)\right)^{2(2^k-1)} f_{\frac{X+Y}{2}}(z) \;=\; \left(f_{\frac{X+Y}{2}}(z)\right)^{2^{k+1}-1},$$

for every $z \in K$, as required. □

We remark that in order to prove Theorem 8.1.3 (and hence also Theorems 8.1.1 and 8.1.2) we will only need the inequality

$$\|f_{S_k}\|_\infty \;\leqslant\; \left(\left\|f_{\frac{X+Y}{2}}\right\|_\infty\right)^{2^k-1}.$$

However, in the proof of Theorem 8.1.4 we shall require the full strength of Lemma 8.2.2.

Recall that, for any convex body $K \subset \mathbb{R}^d$, there exists an affine transformation that maps K to a convex body K' of volume 1 such that $\Sigma_{K'} = L_K^2 I_d$, where $\Sigma_{K'} = \mathbb{E}[X \otimes X]$ is the covariance matrix of the uniform random variable $X \sim \mathcal{U}(K')$, and I_d is the identity matrix. For such a convex body K', it is straightforward to calculate the covariance matrix of S_k.

Lemma 8.2.3. *Let K be a convex body, let $X \sim \mathcal{U}(K)$, and suppose that $\mathbb{E}[X \otimes X] = L_K^2 I_d$. Then*

$$\mathbb{E}\big[S_k \otimes S_k\big] = 2^{-k} L_K^2 I_d$$

for every $k \in \mathbb{N}$.

Proof. Since $S_k = 2^{-k}\sum_{i=1}^{2^k} X_i$ and the X_i are uniform and independent, it follows that

$$\mathbb{E}\big[S_k \otimes S_k\big] = \frac{1}{2^{2k}} \sum_{i,j=1}^{2^k} \mathbb{E}\big[X_i \otimes X_j\big] \;=\; \frac{1}{2^{2k}} \sum_{i=1}^{2^k} \mathbb{E}\big[X_i \otimes X_i\big] = 2^{-k} L_K^2 I_d,$$

as claimed. □

We are now ready to prove Theorem 8.1.3.

Proof of Theorem 8.1.3. By applying an affine transformation, we may assume that K has volume 1 and is centred at the origin, and that $\mathbb{E}[X \otimes X] = L_K^2 I_d$, where $X \sim \mathcal{U}(K)$. Fix $k \in \mathbb{N}$ such that

$$2^{15} L_K^2 \leqslant 2^k \leqslant 2^{16} L_K^2,$$

set $R := 2^{-7}\sqrt{d}$, and observe that, by Markov's inequality and Theorem 8.2.3, we have

$$\mathbb{P}\big(\|S_k\|_2 \geqslant R\big) \leqslant \frac{2^{14}}{d} \cdot \mathbb{E}\big[\|S_k\|_2^2\big] = \frac{2^{14}}{d} \cdot \sum_{i=1}^{d} 2^{-k} L_K^2 = \frac{2^{14} L_K^2}{2^k} \leqslant \frac{1}{2}.$$

Moreover, bounding $\mathbb{P}\big(\|X\|_2 \leqslant R\big)$ simply by the volume of the ball of radius R, we obtain

$$\mathbb{P}\big(\|X\|_2 \leqslant R\big) \leqslant \frac{\pi^{d/2} R^d}{\Gamma(\frac{d}{2}+1)} \leqslant \left(\frac{2e\pi R^2}{d}\right)^{d/2} \leqslant e^{-2d-1}.$$

Combining these two bounds, we deduce that

$$\|f_{S_k}\|_\infty \geqslant \frac{\mathbb{P}\big(\|S_k\|_2 \leqslant R\big)}{\mathbb{P}\big(\|X\|_2 \leqslant R\big)} \geqslant \frac{e^{2d+1}}{2} \geqslant e^{2d}. \tag{8.5}$$

Now, by Lemma 8.2.2, it follows that

$$\big\|f_{\frac{X+Y}{2}}\big\|_\infty \geqslant \big(\|f_{S_k}\|_\infty\big)^{1/(2^k-1)} \geqslant e^{d/2^{k-1}},$$

and hence, by Lemma 8.2.1 and since $2^k \leqslant 2^{16} L_K^2$, we obtain

$$\Delta_{KB}(K) \geqslant 2^{-d} \cdot \big\|f_{\frac{X+Y}{2}}\big\|_\infty \geqslant \exp\left(\frac{d}{2^{15} L_K^2}\right) \cdot 2^{-d},$$

as required. □

We remark that the constant 2^{-15} in Theorem 8.1.3 could be improved somewhat by taking R a little larger (and thus k a little smaller); however, we shall need (8.5) again in Section 8.4, and we chose the constants in the proof above with the application there in mind.

8.3 Hadwiger's conjecture

In this section we will deduce Theorems 8.1.1 and 8.1.2 from Theorem 8.1.3. We begin with the proof of Theorem 8.1.2, for which we will need the following asymmetric variant of $N(K)$: given convex bodies A and B in $\mathbb{R}^d$, define

$$N(A,B) = \min\Big\{N \in \mathbb{N} : \exists\, x_1, \dots, x_N \in \mathbb{R}^d \text{ such that } A \subset \bigcup_{i=1}^{N} \big(x_i + \mathrm{int}(B)\big)\Big\}.$$

We will use the following classical fact (see [127] or [112, Corollary 3.5]), which follows from Rogers' bound [125] on the density of coverings of $\mathbb{R}^d$ with translates of convex bodies.

Lemma 8.3.1. *If $A, B \subset \mathbb{R}^d$ are convex bodies, then*

$$N(A,B) \leqslant O\big(d \log d\big) \cdot \frac{|A-B|}{|B|}.$$

We are now ready to deduce Theorem 8.1.2 from Theorem 8.1.3.

Proof of Theorem 8.1.2. By Theorem 8.1.3, there exists $x \in \mathbb{R}^d$ such that

$$\frac{|K \cap (x-K)|}{|K|} \geqslant \exp\left(\frac{d}{2^{15} L_K^2}\right) \cdot 2^{-d}. \tag{8.6}$$

Set $S := K \cap (x-K)$, and note that

$$N(K) \leqslant N(K,S) \qquad \text{and} \qquad |K-S| \leqslant |K+K| = 2^d \cdot |K|,$$

since $S \subset K$ and $S \subset x-K$, respectively. It therefore follows from Theorem 8.3.1 that

$$N(K) \leqslant N(K,S) \leqslant O(d \log d) \cdot \frac{|K-S|}{|S|} \leqslant O(d \log d) \cdot 2^d \cdot \frac{|K|}{|S|},$$

and hence, by (8.6), we obtain

$$N(K) \leqslant O(d \log d) \cdot \exp\left(-\frac{d}{2^{15} L_K^2}\right) \cdot 4^d = \exp\left(-\frac{\Omega(d)}{L_K^2}\right) \cdot 4^d$$

as $d \to \infty$, as required. □

In order to deduce Theorem 8.1.1 and Theorem 8.1.4, we will need the following theorem of Klartag and Lehec [92].

Theorem 8.3.2. *If $K \subset \mathbb{R}^d$ is a convex body, then*

$$L_K = O(\log d)^4.$$

Theorem 8.1.1 now follows immediately.

Proof of Theorem 8.1.1. By Theorems 8.1.2 and 8.3.2, it follows that

$$N(K) \leqslant \exp\left(-\frac{\Omega(d)}{L_K^2}\right) \cdot 4^d \leqslant \exp\left(-\Omega\left(\frac{d}{(\log d)^8}\right)\right) \cdot 4^d$$

as $d \to \infty$, as required. □

8.4 Ehrhart's conjecture

In order to prove Theorem 8.1.4, we will need the following variant of Theorem 8.1.3.

Theorem 8.4.1. *If $K \subset \mathbb{R}^d$ is a convex body centred at the origin, then*

$$\Delta_{KB}(K) \geqslant \frac{|K \cap (-K)|}{|K|} \geqslant \exp\left(\frac{d}{2^{16}L_K^2}\right) \cdot 2^{-d}.$$

We will deduce Theorem 8.4.1 from the proof of Theorem 8.1.3, together with the following bound on the value of a log-concave function at its centre of mass [65, Theorem 4].

Theorem 8.4.2. *If $f\colon \mathbb{R}^d \to \mathbb{R}_+$ is a log-concave function, then*

$$f(y) \geqslant e^{-d} \cdot \|f\|_\infty,$$

where $y = \int_{x\in\mathbb{R}^d} f(x) \cdot x\,dx$ is the centre of mass of f.

Theorem 8.4.1 now follows from Lemma 8.2.2, as before.

Proof of Theorem 8.4.1. Recall that the function f_{S_k} is log-concave, where S_k is the random variable defined in (8.4), and note that, since K is centred at the origin, the centre of mass of f_{S_k} is also the origin. By Theorem 8.4.2 and (8.5), it follows that

$$f_{S_k}(0) \geqslant e^{-d} \cdot \|f_{S_k}\|_\infty \geqslant e^d.$$

Now, by Lemma 8.2.2, it follows that

$$f_{\frac{X+Y}{2}}(0) \geqslant \big(f_{S_k}(0)\big)^{1/(2^k-1)} \geqslant e^{d/2^k},$$

and hence, by Lemma 8.2.1 and since $2^k \leqslant 2^{16}L_K^2$, we obtain

$$\frac{|K \cap (-K)|}{|K|} = 2^{-d} \cdot f_{\frac{X+Y}{2}}(0) \geqslant \exp\left(\frac{d}{2^{16}L_K^2}\right) \cdot 2^{-d},$$

as claimed. □

Finally, to deduce Theorem 8.1.4, recall that, by Minkowski's theorem, every convex body $K \subset \mathbb{R}^d$ such that $K = -K$ and $K \cap \mathbb{Z}^d = \{0\}$ has volume at most 2^d.

Proof of Theorem 8.1.4. By Minkowski's inequality and Theorems 8.3.2 and 8.4.1, we have

$$\frac{2^d}{|K|} \geqslant \frac{|K \cap (-K)|}{|K|} \geqslant \exp\left(\frac{d}{2^{16}L_K^2}\right) \cdot 2^{-d} \geqslant \exp\left(\Omega\left(\frac{d}{(\log d)^8}\right)\right) \cdot 2^{-d}$$

as $d \to \infty$, as required. □

Appendix A

The Proofs of two Esseen-type lemmas

This appendix presents joint work with Matthew Jenssen, Marcus Michelen and Julian Sahasrabudhe. In this appendix we prove our two Esseen-type lemmas, Lemma 2.3.2 and Lemma 2.5.2, for random variables of the form $W^T\tau$, where τ is a μ-lazy random vector in $\{-1,0,1\}^{2d}$ and W is a (fixed) $2d\times\ell$ matrix for some $\ell\in\mathbb{N}$. Recall that for a vector $u\in\mathbb{R}^\ell$, we let $\|u\|_{\mathbb{T}}$ denote the Euclidean distance from u to the integer lattice $\mathbb{Z}^\ell$.

A.1 Basics of Fourier representation

As above, we let τ be a μ-lazy random vector in $\{-1,0,1\}^{2d}$ and let W be a $2d\times\ell$ matrix. Recall the characteristic function φ_X of a vector valued random variable X is defined as

$$\varphi_X(\theta)=\mathbb{E}\exp(2\pi i\langle X,\theta\rangle),$$

and so we may express characteristic function of $W^T\tau$ as

$$\varphi(\theta)=\mathbb{E}\exp(2\pi i\langle\tau,W\theta\rangle)=\prod_{j=1}^{2d}\left((1-\mu)+\mu\cos(2\pi(W\theta)_j)\right).$$

We note the elementary fact that for $\mu\in[0,1/4]$ we have

$$\mu\|x\|_{\mathbb{T}}^2\leqslant-\log\left(1-\mu+\mu\cos(2\pi x)\right)\leqslant 32\mu\|x\|_{\mathbb{T}}^2, \tag{A.1}$$

from which we deduce

$$\exp\left(-32\mu\left\|W\theta\right\|_{\mathbb{T}}^2\right)\leqslant\varphi(\theta)\leqslant\exp\left(-\mu\left\|W\theta\right\|_{\mathbb{T}}^2\right). \tag{A.2}$$

We now note a standard fact regarding Fourier inversion (see [151] p.290).

Fact A.1.1 (Fourier inversion)**.** *Let X be a random vector in $\mathbb{R}^\ell$, then for $w \in \mathbb{R}^\ell$ we have*

$$\mathbb{E}\exp\left(-\frac{\pi\|X-w\|_2^2}{2}\right) = \int_{\mathbb{R}^\ell} e^{-\pi\|\theta\|_2^2}\cdot e^{-2\pi i\langle w,\theta\rangle}\varphi_X(\theta)\,d\theta\,.$$

In particular, letting $g \sim \mathcal{N}(0,(2\pi)^{-1}I_\ell)$, we have

$$\mathbb{E}\exp\left(-\frac{\pi\|X-w\|_2^2}{2}\right) = \mathbb{E}_g(e^{-2\pi i\langle w,g\rangle}\varphi_X(g))\,.$$

A.2 Proof of Lemma 2.3.2 and Lemma 2.5.2

Recall that for $\ell \in \mathbb{N}$, γ_ℓ denotes the ℓ dimensional Gaussian measure defined by $\gamma_\ell(S) = \mathbb{P}(g \in S)$, where $g \sim \mathcal{N}(0,(2\pi)^{-1}I_\ell)$. We begin with the proof of Lemma 2.3.2.

Proof of Lemma 2.3.2. Let $w \in \mathbb{R}^\ell$. We apply Markov's inequality to obtain

$$\mathbb{P}_\tau\big(\|W^T\tau - w\|_2 \leqslant \beta\sqrt{\ell}\big) \leqslant \exp\left(\frac{\pi}{2}\beta^2\ell\right)\mathbb{E}_\tau\exp\left(-\frac{\pi\|W^T\cdot\tau-w\|_2^2}{2}\right)\,.$$

As above, let φ be the characteristic function of $W^T\tau$. We apply Fact A.1.1 and (A.2) to obtain

$$\mathbb{E}_\tau\exp\left(-\frac{\pi\|W^T\cdot\tau-w\|_2^2}{2}\right) = \mathbb{E}_g[e^{-2\pi i\langle w,g\rangle}\varphi(g)] \leqslant \mathbb{E}_g[\exp(-\nu\|Wg\|_{\mathbb{T}}^2)].$$

The right-hand-side of the above may be rewritten as

$$\int_0^1 \mathbb{P}_g(\exp(-\nu\|Wg\|_{\mathbb{T}}^2) \geqslant t)\,dt = \nu\int_0^\infty \mathbb{P}_g(\|Wg\|_{\mathbb{T}}^2 \leqslant u)e^{-\nu u}\,du = \nu\int_0^\infty \gamma_\ell(S_W(u))e^{-\nu u}\,du,$$

where for the first equality we made the change of variable $t = e^{-\nu u}$.

Choosing m to maximize $\gamma_\ell(S_W(u))e^{-\nu u/2}$ (as a function of u), we may bound

$$\nu\int_0^\infty \gamma_\ell(S_W(u))e^{-\nu u}du \leqslant \nu\gamma_\ell(S_W(m))e^{-\nu m/2}\int_0^\infty e^{-\nu u/2}du = 2\gamma_\ell(S_W(m))e^{-\nu m/2}\,.$$

Putting everything together we obtain

$$\mathbb{P}_\tau(\|W^T\tau - w\|_2 \leqslant 2\beta\sqrt{\ell}) \leqslant 2e^{\pi\beta^2\ell/2}e^{-\nu m/2}\gamma_\ell(S_W(m))\,.$$

□

The proof of Lemma 2.5.2 proceeds in much the same way.

Proof of Lemma 2.5.2. Let us set $X = \|W^T \cdot \tau\|_2$ and write

$$\mathbb{E}_X e^{-\pi X^2/2} = \mathbb{E}_X \mathbb{1}(X \leqslant \beta\sqrt{\ell}) e^{-\pi X^2/2} + \mathbb{E}_X \mathbb{1}\big(X \geqslant \beta\sqrt{\ell}\big) e^{-\pi X^2/2} \leqslant \mathbb{P}_X(X \leqslant \beta\sqrt{\ell}) + e^{-\pi\beta^2\ell/2}$$

and therefore, using that $\exp(-\pi\beta^2\ell/2) \leqslant \exp(-\beta^2\ell)$,

$$\mathbb{E}_\tau \exp\left(\frac{-\pi\|W^T \cdot \tau\|_2^2}{2}\right) \leqslant \mathbb{P}_\tau(\|W^T \cdot \tau\|_2 \leqslant \beta\sqrt{\ell}) + e^{-\beta^2\ell}.$$

As before, we let φ be the characteristic function of $W^T\tau$, and let g be a standard ℓ-dimensional Gaussian random variable with standard deviation $(2\pi)^{-1/2}$. By Fact A.1.1 and (A.2) we obtain

$$\mathbb{E}_\tau \exp\left(-\frac{\pi\|W^T \cdot \tau\|_2^2}{2}\right) = \mathbb{E}_g[\varphi(g)] \geqslant \mathbb{E}_g[\exp(-32\mu\|Wg\|_{\mathbb{T}}^2)].$$

Similar to the proof of Lemma 2.3.2, we write

$$\mathbb{E}_g[\exp(-32\mu\|Wg\|_{\mathbb{T}}^2)] = 32\mu \int_0^\infty \gamma_\ell(S_W(u)) e^{-32\mu u} du \geqslant 32\mu\gamma_\ell(S_W(t)) \int_t^\infty e^{-32\mu u}\, du,$$

where we have used that $\gamma_\ell(S_W(b)) \geqslant \gamma_\ell(S_W(a))$ for all $b \geqslant a$. This completes the proof of Lemma 2.5.2. □

Appendix B

Relating A to the zeroed out matrix M.

This appendix presents joint work with Matthew Jenssen, Marcus Michelen and Julian Sahasrabudhe. In this appendix we prove Lemma 2.8.1 and Lemma 2.9.6. To prove these results, we compare Fourier transforms (that is the *characteristic functions*) of the random variables Mv and Av, for fixed v. We first record the characteristic functions of these random variables. For $\xi \in \mathbb{R}^n$ we have

$$\psi_v(\xi) := \mathbb{E}\, e^{2\pi i \langle Av, \xi\rangle} = \left(\prod_{k=1}^{n} \cos(2\pi v_k \xi_k)\right) \cdot \left(\prod_{j<k} \big(2\pi(\xi_j v_k + \xi_k v_j)\big)\right)$$

and

$$\chi_v(\xi) := \mathbb{E}\, e^{2\pi i \langle Mv, \xi\rangle} = \prod_{j=1}^{d} \prod_{k=d+1}^{n} \left(\frac{3}{4} + \frac{1}{4}\cos\big(2\pi(\xi_j v_k + \xi_k v_j)\big)\right).$$

Our comparison is based on two main points. First we have that $\chi_v(\xi) \geqslant 0$. Second, we have

$$\psi_v(\xi) \leqslant \chi_v(2\xi)\,, \tag{B.1}$$

which follows from $|\cos(t)| \leqslant \frac{3}{4} + \frac{1}{4}\cos(2t)$ and $|\cos(t)| \leqslant 1$.

Fact B.1.1. *For $v \in \mathbb{R}^n$, and $t \geqslant \mathcal{T}_L(v)$, we have*

$$\mathbb{E}\exp(-\pi\|Mv\|_2^2/t^2) \leqslant (9Lt)^n.$$

Proof. Now $\mathbb{E}\exp(-\pi\|Mv\|_2^2/t^2)$ is at most

$$\mathbb{P}(\|Mv\|_2 \leqslant t\sqrt{n}) + \sqrt{n}\int_t^\infty \exp\left(-\frac{s^2 n}{t^2}\right)\mathbb{P}(\|Mv\|_2 \leqslant s\sqrt{n})\, ds\,. \tag{B.2}$$

and since $t \geqslant \mathcal{T}_L(v)$, we have $\mathbb{P}(\|Mv\|_2 \leqslant s\sqrt{n}) \leqslant (8Ls)^n$ for all $s \geqslant t$, and so we may bound

$$\sqrt{n}\int_t^\infty \exp\left(-\frac{s^2 n}{t^2}\right)\mathbb{P}(\|Mv\|_2 \leqslant s\sqrt{n})\,ds \leqslant \sqrt{n}(8Lt)^n \int_t^\infty \exp\left(-\frac{s^2 n}{t^2}\right)(s/t)^n\,ds\,.$$

Changing variables $u = s/t$, the right hand side is equal to

$$t^{-1}\sqrt{n}(8Lt)^n \int_1^\infty \exp(-u^2 n)u^n\,du \leqslant t^{-1}\sqrt{n}(8Lt)^n \int_1^\infty \exp(-u^2/2)\,du \leqslant (9Lt)^n,$$

as desired. □

Proof of Lemma 2.8.1. Apply Markov's inequality to bound

$$\mathbb{P}(\|Av - w\|_2 \leqslant t\sqrt{n}) \leqslant \exp(\pi n/2)\mathbb{E}\exp\left(-\pi\|Av - w\|_2^2/2t^2\right)\,. \tag{B.3}$$

Using the Fourier inversion formula in Fact A.1.1 we write

$$\mathbb{E}_A \exp\left(-\pi\|Av - w\|_2^2/2t^2\right) = \int_{\mathbb{R}^n} e^{-\pi\|\xi\|_2^2}\cdot e^{-2\pi i t^{-1}\langle w,\xi\rangle}\psi_v(t^{-1}\xi)\,d\xi\,. \tag{B.4}$$

Rescaling, applying (B.1) and non-negativity of χ_v yields that the RHS of (B.4) is at most

$$\int_{\mathbb{R}^n} e^{-\pi\|\xi\|_2^2}\chi_v(2t^{-1}\xi)\,d\xi \leqslant \mathbb{E}_M \exp(-2\pi\|Mv\|_2^2/t^2).$$

Now use Fact B.1.1 along with the assumption $t \geqslant \mathcal{T}_L(v)$ to obtain

$$\mathbb{E}_M \exp(-2\pi\|Mv\|_2^2/t^2) \leqslant (9Lt)^n,$$

as desired. □

We prove Lemma 2.9.6 in a similar manner. Recall $\rho_\varepsilon(v) = \max_{b\in\mathbb{R}^n}\mathbb{P}\left(\sum_i v_i\varepsilon_i \in (b-\varepsilon, b+\varepsilon)\right)$.

Proof of Lemma 2.9.6. Set $\varepsilon = \mathcal{T}_L(v)$ and let B be a $n\times n$ matrix uniformly drawn from all matrices with entries in $\{\pm 1\}$ and apply Markov's inequality to bound

$$\rho_\varepsilon(v)^n \leqslant \max_{w\in\mathbb{R}^n}\mathbb{P}(\|Bv - w\|_2 \leqslant \varepsilon\sqrt{n}) \leqslant \max_{w\in\mathbb{R}^n}\exp(\pi n/2)\mathbb{E}\exp\left(-\pi\|Bv - w\|_2^2/2\varepsilon^2\right)\,. \tag{B.5}$$

Apply Fact A.1.1 to write

$$\mathbb{E}\exp\left(-\pi\|Bv - w\|_2^2/2\varepsilon^2\right) = \int_{\mathbb{R}^n} e^{-\pi\|\xi\|_2^2}\cdot e^{-2\pi i\varepsilon^{-1}\langle w,\xi\rangle}\prod_{1\leqslant j,k\leqslant n}\cos(2\pi\varepsilon^{-1}v_j\xi_k)\,d\xi \tag{B.6}$$

and use Hölder's inequality to bound the RHS of (B.6)

$$\leqslant \left(\int_{\mathbb{R}^n} e^{-2\pi\|\xi\|_2^2/3}\, d\xi\right)^{3/4} \left(\int_{\mathbb{R}^n} e^{-2\pi\|\xi\|_2^2} \prod_{1\leqslant j,k\leqslant n} \cos(2\pi\varepsilon^{-1} v_j \xi_k)^4\, d\xi\right)^{1/4} . \tag{B.7}$$

Now use $\int_{\mathbb{R}^n} e^{-2\pi\|\xi\|_2^2/3}\, d\xi = \left(\frac{3}{2}\right)^{n/2}$ and $(\cos(a)\cos(b))^4 \leqslant \frac{3}{4} + \frac{1}{4}\cos(2(a+b))$, to see (B.7) is

$$\leqslant \left(\frac{3}{2}\right)^{3n/8} \left(2^{-n/2}\int_{\mathbb{R}^n} e^{-\pi\|\xi\|_2^2} \chi_v(\sqrt{2}\varepsilon^{-1}\xi)\, d\xi\right)^{1/4} \leqslant \left(\frac{27}{128}\right)^{n/8} \left(\mathbb{E}\exp\left(-\pi\|Mv\|_2^2/\varepsilon^2\right)\right)^{1/4} . \tag{B.8}$$

Taken together, lines (B.5), (B.6), (B.7), (B.8) tell us that

$$\rho_\varepsilon(v)^n \leqslant (3/2)^{3n/8}(\exp(\pi/2)/\sqrt{2})^n \left(\mathbb{E}\exp\left(-\pi\|Mv\|_2^2/\varepsilon^2\right)\right)^{1/4} . \tag{B.9}$$

Now apply Fact B.1.1 to bound $\mathbb{E}\exp\left(-\pi\|Mv\|_2^2/\varepsilon^2\right) \leqslant (9L\varepsilon)^n$ and so $\rho_\varepsilon(v)^n \leqslant (2^{12}L\varepsilon)^{n/4}$, as desired. □

Bibliography

[1] Amol Aggarwal. Bulk universality for generalized Wigner matrices with few moments. *Probab. Theory Related Fields*, 173(1-2):375–432, 2019.

[2] Noga Alon, József Balogh, Robert Morris, and Wojciech Samotij. A refinement of the Cameron–Erdős conjecture. *Proceedings of the London Mathematical Society*, 108(1):44–72, 2013.

[3] Noga Alon, József Balogh, Robert Morris, and Wojciech Samotij. Counting sum-free sets in abelian groups. *Israel Journal of Mathematics*, 199(1):309–344, 2014.

[4] Noga Alon, Andrew Granville, and Adrián Ubis. The number of sumsets in a finite field. *Bull. London Math. Soc*, 42(5):784–794, 2010.

[5] Ryan Alweiss, Shachar Lovett, Kewen Wu, and Jiapeng Zhang. Improved bounds for the sunflower lemma. *Annals of Mathematics*, 194(3):795–815, 2021.

[6] Greg W Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*. Number 118. Cambridge University Press, 2010.

[7] Milla Anttila, Keith Ball, and Irini Perissinaki. The central limit problem for convex bodies. *Transactions of the American Mathematical Society*, 355(12):4723–4735, 2003.

[8] Gérard Ben Arous and Paul Bourgade. Extreme gaps between eigenvalues of random matrices. *The Annals of Probability*, 41(4):2648–2681, 2013.

[9] David Arthur, Bodo Manthey, and Heiko Röglin. Smoothed analysis of the k-means method. *Journal of the ACM (JACM)*, 58(5):1–31, 2011.

[10] Z. D. Bai and Y. Q. Yin. Convergence to the semicircle law. *Ann. Probab.*, 16(2):863–875, 1988.

[11] Zhi-Dong Bai and Yong-Qua Yin. Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. *The Annals of Probability*, pages 1729–1741, 1988.

[12] Zhidong Bai and Jack W Silverstein. *Spectral analysis of large dimensional random matrices*, volume 20. Springer, 2010.

[13] József Balogh, Hong Liu, and Maryam Sharifzadeh. The number of subsets of integers with no k-term arithmetic progression. *International Mathematics Research Notices*, 2017(20):6168–6186, 2017.

[14] József Balogh, Hong Liu, Maryam Sharifzadeh, and Andrew Treglown. Sharp bound on the number of maximal sum-free subsets of integers. *Journal of the European Mathematical Society*, 20(8):1885–1911, 2018.

[15] József Balogh, Robert Morris, and Wojciech Samotij. Independent sets in hypergraphs. *J. Amer. Math. Soc.*, 28(3):669–709, 2015.

[16] József Balogh, Robert Morris, and Wojciech Samotij. The method of hypergraph containers. In *Proceedings of the International Congress of Mathematicians*, 2018.

[17] József Balogh, Robert Morris, Wojciech Samotij, and Lutz Warnke. The typical structure of. *Transactions of the American Mathematical Society*, 368(9):6439–6485, 2016.

[18] Jozsef Balogh and Wojciech Samotij. An efficient container lemma. *Discrete Analysis*, 2020.

[19] Ross Berkowitz. A local limit theorem for cliques in $G(n, p)$. *arXiv preprint arXiv:1811.03527*, 2018.

[20] Aditya Bhaskara, Moses Charikar, Ankur Moitra, and Aravindan Vijayaraghavan. Smoothed analysis of tensor decompositions. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 594–603, 2014.

[21] Sergey G Bobkov and Alexander Koldobsky. On the central limit property of convex bodies. In *Geometric aspects of functional analysis*, pages 44–52. Springer, 2003.

[22] Béla Bollobás. *Random graphs*. Number 73. Cambridge University Press, 2001.

[23] Christer Borell. Inequalities of the Brunn–Minkowski type for Gaussian measures. *Probability Theory and Related Fields*, 140(1-2):195–205, 2008.

[24] Paul Bourgade, Laszlo Erdős, Horng-Tzer Yau, and Jun Yin. Fixed energy universality for generalized Wigner matrices. *Comm. Pure Appl. Math.*, 69(10):1815–1881, 2016.

[25] Jean Bourgain. Geometry of Banach spaces and harmonic analysis. In *Proceedings of the International Congress of Mathematicians*, 1986.

[26] Jean Bourgain. On high dimensional maximal functions associated to convex bodies. *American Journal of Mathematics*, 108(6):1467–1476, 1986.

[27] Jean Bourgain. On the distribution of polynomials on high dimensional convex sets. In *Geometric aspects of functional analysis*, pages 127–137. Springer, 1991.

[28] Jean Bourgain. On the isotropy-constant problem for "psi-2"-bodies. In *Geometric aspects of functional analysis*, pages 114–121. Springer, 2003.

[29] Jean Bourgain. On a problem of Farrell and Vershynin in random matrix theory. In *Geometric aspects of functional analysis*, volume 2169 of *Lecture Notes in Math.*, pages 65–69. Springer, Cham, 2017.

[30] Jean Bourgain, Van H. Vu, and Philip Matchett Wood. On the singularity probability of discrete random matrices. *J. Funct. Anal.*, 258(2):559–603, 2010.

[31] Marcelo Campos. On the number of sets with a given doubling constant. *Israel Journal of Mathematics*, 236(2):711–726, 2020.

[32] Marcelo Campos, Maurício Collares, Robert Morris, Natasha Morrison, and Victor Souza. The typical structure of sets with small sumset. *International Mathematics Research Notices*, 2022(14):11011–11055, 2022.

[33] Marcelo Campos, Matthew Jenssen, Marcus Michelen, and Julian Sahasrabudhe. The singularity probability of a random symmetric matrix is exponetially small. *arXiv preprint arXiv:2105.11384*, 2021.

[34] Marcelo Campos, Matthew Jenssen, Marcus Michelen, and Julian Sahasrabudhe. The least singular value of a random symmetric matrix. *arXiv preprint arXiv:2203.06141*, 2022.

[35] Marcelo Campos, Matthew Jenssen, Marcus Michelen, and Julian Sahasrabudhe. Singularity of random symmetric matrices revisited. *Proceedings of the American Mathematical Society*, 150(07):3147–3159, 2022.

[36] Marcelo Campos, Matthew Jenssen, Marcus Michelen, and Julian Sahasrabudhe. Supplementary paper to the least singular value of a random symmetric matrix. *arXiv preprint arXiv:2203.06141*, 2022.

[37] Marcelo Campos, Letícia Mattos, Robert Morris, and Natasha Morrison. On the singularity of random symmetric matrices. *Duke Mathematical Journal*, 170(5):881–907, 2021.

[38] Marcelo Campos, Peter van Hintum, Robert Morris, and Marius Tiba. Towards Hadwiger's conjecture via bourgain slicing. *arXiv preprint arXiv:2206.11227*, 2022.

[39] Yuansi Chen. An almost constant lower bound of the isoperimetric coefficient in the KLS conjecture. *Geometric and Functional Analysis*, 31(1):34–61, 2021.

[40] Kevin P. Costello. Bilinear and quadratic variants on the Littlewood-Offord problem. *Israel Journal of Mathematics*, 194(1):359–394, 2013.

[41] Kevin P. Costello, Terence Tao, and Van Vu. Random symmetric matrices are almost surely nonsingular. *Duke Math. J.*, 135(2):395–413, 2006.

[42] Domingos Dellamonica Jr, Yoshiharu Kohayakawa, Sang June Lee, Vojtěch Rödl, and Wojciech Samotij. The number of B3-sets of a given cardinality. *Journal of Combinatorial Theory, Series A*, 142:44–76, 2016.

[43] Alan Edelman. Eigenvalues and condition numbers of random matrices. *SIAM journal on matrix analysis and applications*, 9(4):543–560, 1988.

[44] Eugene Ehrhart. Une généralisation probable du théorème fondamental de Minkowski. *Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences*, 258(20):4885, 1964.

[45] Ronen Eldan. Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geometric and Functional Analysis*, 23(2):532–569, 2013.

[46] Ronen Eldan and Bo'az Klartag. Approximately gaussian marginals and the hyperplane conjecture. *Concentration, functional inequalities and isoperimetry*, 545:55–68, 2011.

[47] László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Spectral statistics of Erdos-Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues. *Comm. Math. Phys.*, 314(3):587–640, 2012.

[48] László Erdős, Sandrine Péché, José A. Ramírez, Benjamin Schlein, and Horng-Tzer Yau. Bulk universality for Wigner matrices. *Comm. Pure Appl. Math.*, 63(7):895–925, 2010.

[49] László Erdős, José Ramírez, Benjamin Schlein, Terence Tao, Van Vu, and Horng-Tzer Yau. Bulk universality for Wigner Hermitian matrices with subexponential decay. *Math. Res. Lett.*, 17(4):667–674, 2010.

[50] László Erdős, Benjamin Schlein, and Horng-Tzer Yau. Local semicircle law and complete delocalization for Wigner random matrices. *Comm. Math. Phys.*, 287(2):641–655, 2009.

[51] László Erdős, Benjamin Schlein, and Horng-Tzer Yau. Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab.*, 37(3):815–852, 2009.

[52] László Erdős, Benjamin Schlein, and Horng-Tzer Yau. Universality of random matrices and local relaxation flow. *Invent. Math.*, 185(1):75–119, 2011.

[53] László Erdős, Benjamin Schlein, Horng-Tzer Yau, and Jun Yin. The local relaxation flow approach to universality of the local statistics for random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.*, 48(1):1–46, 2012.

[54] László Erdős, Horng-Tzer Yau, and Jun Yin. Bulk universality for generalized Wigner matrices. *Probab. Theory Related Fields*, 154(1-2):341–407, 2012.

[55] Paul Erdős. On a lemma of Littlewood and Offord. *Bull. Amer. Math. Soc.*, 51:898–902, 1945.

[56] László Erdős. Universality of Wigner random matrices: a survey of recent results. *Russian Mathematical Surveys*, 66(3):507, 2011.

[57] László Erdős, Benjamin Schlein, and Horng-Tzer Yau. Wegner estimate and level repulsion for Wigner random matrices. *International Mathematics Research Notices*, 2010(3):436–479, 2010.

[58] László Erdős and Horng-Tzer Yau. Gap universality of generalized Wigner and beta-ensembles. *Journal of the European Mathematical Society*, 17(8):1927–2036, 2015.

[59] Carl-Gustav Esseen. On the Kolmogorov-Rogozin inequality for the concentration function. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 5(3):210–216, 1966.

[60] Brendan Farrell and Roman Vershynin. Smoothed analysis of symmetric random matrices with continuous distributions. *Proceedings of the American Mathematical Society*, 144(5):2257–2261, 2016.

[61] Ohad N. Feldheim and Sasha Sodin. A universality result for the smallest eigenvalues of certain sample covariance matrices. *Geom. Funct. Anal.*, 20(1):88–123, 2010.

[62] Asaf Ferber and Vishesh Jain. Singularity of random symmetric matrices—a combinatorial approach to improved bounds. *Forum Math. Sigma*, 7:Paper No. e22, 29, 2019.

[63] Asaf Ferber, Vishesh Jain, Kyle Luh, and Wojciech Samotij. On the counting problem in inverse Littlewood–Offord theory. *Journal of the London Mathematical Society*, 2021.

[64] Asaf Ferber, Vishesh Jain, and Yufei Zhao. On the number of Hadamard matrices via anti-concentration. *Combinatorics, Probability and Computing*, 31(3):455–477, 2022.

[65] Matthieu Fradelizi. Sections of convex bodies through their centroid. *Archiv der Mathematik*, 69(6):515–522, 1997.

[66] Péter Frankl and Zoltán Füredi. Solution of the Littlewood-Offord problem in high dimensions. *Annals of Mathematics*, pages 259–270, 1988.

[67] Keith Frankston, Jeff Kahn, Bhargav Narayanan, and Jinyoung Park. Thresholds versus fractional expectation-thresholds. *Annals of Mathematics*, 194(2):475–495, 2021.

[68] Gregory A. Freiman. The addition of finite sets I. *Izvestiya Vysshikh Uchebnykh Zavedenii. Matematika*, (6):202–213, 1959.

[69] F. Götze, A. Naumov, and A. Tikhomirov. Local semicircle law under fourth moment condition. *J. Theoret. Probab.*, 33(3):1327–1362, 2020.

[70] Friedrich Götze, Alexey Naumov, Alexander Tikhomirov, and Dmitry Timushev. On the local semicircular law for Wigner ensembles. *Bernoulli*, 24(3):2358–2400, 2018.

[71] Ben Green. The Cameron–Erdős conjecture. *Bulletin of the London Mathematical Society*, 36(6):769–778, 2004.

[72] Ben Green. Counting sets with small sumset, and the clique number of random Cayley graphs. *Combinatorica*, 25(3):307–326, 2005.

[73] Ben Green and Robert Morris. Counting sets with small sumset and applications. *Combinatorica*, 36(2):129–159, 2016.

[74] Ben Green and Imre Z. Ruzsa. Counting sumsets and sum-free sets modulo a prime. *Studia Scientiarum Mathematicarum Hungarica*, 41(3):285–293, 2004.

[75] Jerrold R. Griggs, Jeffrey C. Lagarias, Andrew M. Odlyzko, and James B. Shearer. On the tightest packing of sums of vectors. *European Journal of Combinatorics*, 4(3):231–236, 1983.

[76] Olivier Guédon and Emanuel Milman. Interpolating thin-shell and sharp large-deviation estimates for isotropic log-concave measures. *Geometric and Functional Analysis*, 21(5):1043–1068, 2011.

[77] Alice Guionnet. Bernoulli random matrices. *arXiv preprint arXiv:2112.05506.*, 2021.

[78] Hugo Hadwiger. Ungeloste problems nr. 20. *Elem. Math.*, 12(121), 1957.

[79] Gábor Halász. On the distribution of additive arithmetic functions. *Acta Arithmetica*, 1(27):143–152, 1975.

[80] Yahya O. Hamidoune and Oriol Serra. A note on Pollard's theorem. *arXiv preprint arXiv:0804.2593*, 2008.

[81] David L. Hanson and Farroll T. Wright. A bound on tail probabilities for quadratic forms in independent random variables. *The Annals of Mathematical Statistics*, 42(3):1079–1083, 1971.

[82] Martin Henk, Matthias Henze, and María A. Hernández Cifre. Variations of Minkowski's theorem on successive minima. In *Forum Mathematicum*, volume 28, pages 311–325. De Gruyter, 2016.

[83] Han Huang, Boaz A. Slomka, Tomasz Tkocz, and Beatrice-Helen Vritsiou. Improved bounds for Hadwiger's covering problem via thin-shell estimates. *Journal of the European Mathematical Society*, 24(4):1431–1448, 2021.

[84] Jiaoyang Huang, Benjamin Landon, and Horng-Tzer Yau. Bulk universality of sparse random matrices. *J. Math. Phys.*, 56(12):123301, 19, 2015.

[85] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney. On the smoothed analysis of the smallest singular value with discrete noise. *arXiv preprint arXiv:2009.01699*, 2020.

[86] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney. Singularity of discrete random matrices. *Geometric and Functional Analysis*, 31(5):1160–1218, 2021.

[87] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney. On the smallest singular value of symmetric random matrices. *Combinatorics, Probability and Computing*, 31(4):662–683, 2022.

[88] Jeff Kahn, János Komlós, and Endre Szemerédi. On the probability that a random ±1-matrix is singular. *J. Amer. Math. Soc.*, 8(1):223–240, 1995.

[89] Ravi Kannan, László Lovász, and Miklós Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete & Computational Geometry*, 13(3):541–559, 1995.

[90] Gy Katona. On a conjecture of Erdős and a stronger form of Sperner's theorem. *Studia Sci. Math. Hungar*, 1:59–63, 1966.

[91] Bo'az Klartag. On convex perturbations with a bounded isotropic constant. *Geometric & Functional Analysis GAFA*, 16(6):1274–1290, 2006.

[92] Bo'az Klartag and Joseph Lehec. Bourgain's slicing problem and KLS isoperimetry up to polylog. *arXiv preprint arXiv:2203.15551*, 2022.

[93] Bo'az Klartag and Vitali Milman. The slicing problem by bourgain. *To Appear) Analysis at Large, A Collection of Articles in Memory of Jean Bourgain*, 2021.

[94] Daniel J. Kleitman. On a lemma of Littlewood and Offord on the distributions of linear combinations of vectors. *Advances in Mathematics*, 5(1):155–157, 1970.

[95] János Komlós. On the determinant of (0, 1) matrices. *Studia Sci. Math. Hungar.*, 2:7–21, 1967.

[96] János Komlós. On the determinant of random matrices. *Studia Sci. Math. Hungar.*, 1968.

[97] Matthew Kwan and Lisa Sauermann. An algebraic inverse theorem for the quadratic Littlewood-Offord problem, and an application to Ramsey graphs. *Discrete Analysis*, 2020.

[98] Yin Tat Lee and Santosh S. Vempala. Eldan's stochastic localization and the KLS conjecture: Isoperimetry, concentration and mixing. *arXiv preprint arXiv:1612.01507*, 2016.

[99] Friedrich Wilhelm Levi. Überdeckung eines eibereiches durch parallelverschiebung seines offenen kerns. *Archiv der Mathematik*, 6(5):369–370, 1955.

[100] John E. Littlewood and Albert C. Offord. On the Number of Real Roots of a Random Algebraic Equation. *J. London Math. Soc.*, 13(4):288–295, 1938.

[101] John E. Littlewood and Albert C. Offord. On the number of real roots of a random algebraic equation. III. *Rec. Math. [Mat. Sbornik] N.S.*, 12(54):277–286, 1943.

[102] Alexander E Litvak and Konstantin E Tikhomirov. Singularity of sparse Bernoulli matrices. *Duke Mathematical Journal*, 171(5):1135–1233, 2022.

[103] Galyna V. Livshyts. The smallest singular value of heavy-tailed not necessarily iid random matrices via random rounding. *Journal d'Analyse Mathématique*, pages 1–50, 2021.

[104] Galyna V. Livshyts, Konstantin Tikhomirov, and Roman Vershynin. The smallest singular value of inhomogeneous square random matrices. *The Annals of Probability*, 49(3):1286 – 1309, 2021.

[105] Przemyslaw Mazur. A structure theorem for sets of small popular doubling. *Acta Arithmetica*, 171:221–239, 2015.

[106] Mark W. Meckes. Concentration of norms and eigenvalues of random matrices. *Journal of Functional Analysis*, 211(2):508–524, 2004.

[107] Madan Lal Mehta. *Random matrices*. Elsevier, 2004.

[108] Raghu Meka, Oanh Nguyen, and Van Vu. Anti-concentration for polynomials of independent random variables. *Theory of Computing*, 12(11):1–17, 2016.

[109] Vitali Milman and Alain Pajor. Isotropic position and inertia ellipsoids and zonoids of the unit ball of a normed n-dimensional space. *Geometric aspects of functional analysis*, pages 64–104, 1989.

[110] Vitali Milman and Alain Pajor. Entropy and asymptotic geometry of non-symmetric convex bodies. *Advances in Mathematics*, 152(2):314–335, 2000.

[111] Robert Morris, Wojciech Samotij, and David Saxton. An asymmetric container lemma and the structure of graphs with no induced 4-cycle. *submitted, arXiv:1806.03706*, 2018.

[112] Márton Naszódi. Flavors of translative coverings. In *New Trends in Intuitive Geometry*, pages 335–358. Springer, 2018.

[113] Hoi Nguyen. Inverse Littlewood-Offord problems and the singularity of random symmetric matrices. *Duke Math. J.*, 161(4):545–586, 2012.

[114] Hoi Nguyen. On the least singular value of random symmetric matrices. *Electron. J. Probab.*, 17:no. 53, 19, 2012.

[115] Hoi Nguyen. Random matrices: Overcrowding estimates for the spectrum. *Journal of Functional Analysis*, 275(8):2197–2224, 2018.

[116] Hoi Nguyen, Terence Tao, and Van Vu. Random matrices: tail bounds for gaps between eigenvalues. *Probability Theory and Related Fields*, 167(3):777–816, 2017.

[117] Hoi Nguyen and Van Vu. Optimal inverse Littlewood-Offord theorems. *Adv. Math.*, 226(6):5298–5319, 2011.

[118] Hoi Nguyen and Van Vu. Small probability, inverse theorems, and applications. In *Erdös centennial*, volume 25 of *Bolyai Soc. Math. Stud.*, pages 409–463. János Bolyai Math. Soc., Budapest, 2013.

[119] Jinyoung Park and Huy Tuan Pham. A proof of the Kahn-Kalai conjecture. *arXiv preprint arXiv:2203.17207*, 2022.

[120] L. A. Pastur. The spectrum of random matrices. *Teoret. Mat. Fiz.*, 10(1):102–112, 1972.

[121] Helmut Plünnecke. *Eigenschaften und Abschätzungen von Wirkungsfunktionen.* Number 22. Gesellschaft für Mathematik u. Datenverarbeitung, 1961.

[122] Helmut Plünnecke. Eine zahlentheoretische anwendung der graphentheorie. 1970.

[123] John M. Pollard. A generalisation of the theorem of Cauchy and Davenport. *Journal of the London Mathematical Society*, 2(3):460–462, 1974.

[124] Elizaveta Rebrova and Konstantin Tikhomirov. Coverings of random ellipsoids, and invertibility of matrices with iid heavy-tailed entries. *Israel Journal of Mathematics*, 227(2):507–544, 2018.

[125] Claude A. Rogers. A note on coverings. *Mathematika*, 4(1):1–6, 1957.

[126] Claude A. Rogers and Geoffrey C. Shephard. The difference body of a convex body. *Archiv der Mathematik*, 8(3):220–233, 1957.

[127] Claude A. Rogers and Chuanming Zong. Covering convex bodies by translates of convex bodies. *Mathematika*, 44(1):215–218, 1997.

[128] Mark Rudelson. Invertibility of random matrices: norm of the inverse. *Ann. of Math.*, 168(2):575–600, 2008.

[129] Mark Rudelson and Roman Vershynin. The Littlewood-Offord problem and invertibility of random matrices. *Adv. Math.*, 218(2):600–633, 2008.

[130] Mark Rudelson and Roman Vershynin. Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.*, 62(12):1707–1739, 2009.

[131] Mark Rudelson and Roman Vershynin. Non-asymptotic theory of random matrices: extreme singular values. In *Proceedings of the International Congress of Mathematicians. Volume III*, pages 1576–1602. Hindustan Book Agency, New Delhi, 2010.

[132] Mark Rudelson and Roman Vershynin. Hanson-Wright inequality and sub-Gaussian concentration. *Electronic Communications in Probability*, 18, 2013.

[133] Mark Rudelson and Roman Vershynin. Small ball probabilities for linear images of high-dimensional distributions. *International Mathematics Research Notices*, 2015(19):9594–9617, 2015.

[134] Mark Rudelson and Roman Vershynin. No-gaps delocalization for general random matrices. *Geometric and Functional Analysis*, 26(6):1716–1776, 2016.

[135] Imre Z. Ruzsa. On the cardinality of $a + a$ and $a - a$. In *Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976)*, volume 2, pages 933–938, 1978.

[136] Imre Z. Ruzsa. An application of graph theory to additive number theory. *Scientia, Ser. A*, 3(97-109):9, 1989.

[137] Imre Z. Ruzsa. An analog of Freıman's theorem in groups. *Astérisque*, 258:323–326, 1999.

[138] Attila Sali. Stronger form of an Ms-part Sperner theorem. *European Journal of Combinatorics*, 4(2):179–183, 1983.

[139] Arvind Sankar, Daniel A Spielman, and Shang-Hua Teng. Smoothed analysis of the condition numbers and growth factors of matrices. *SIAM Journal on Matrix Analysis and Applications*, 28(2):446–476, 2006.

[140] Alexander A. Sapozhenko. The Cameron-Erdős conjecture. *Dokl. Akad. Nauk*, 393(6):749–752, 2003.

[141] András Sárközy and Endre Szemeredi. Über ein problem von Erdös und Moser. *Acta Arithmetica*, 11(2):205–208, 1965.

[142] David Saxton and Andrew Thomason. Hypergraph containers. *Invent. Math.*, 201(3):925–992, 2015.

[143] Xuancheng Shao. On an almost all version of the Balog-Szemerédi-Gowers theorem. *Discrete Analysis*, 2019.

[144] Steve Smale. On the efficiency of algorithms of analysis. *Bulletin (New Series) of The American Mathematical Society*, 13(2):87–121, 1985.

[145] Daniel Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: why the simplex algorithm usually takes polynomial time. In *Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing*, pages 296–305. ACM, New York, 2001.

[146] Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis of algorithms. In *Proceedings of the International Congress of Mathematicians, Vol. I (Beijing, 2002)*, pages 597–606. Higher Ed. Press, Beijing, 2002.

[147] Richard P Stanley. Weyl groups, the hard Lefschetz theorem, and the Sperner property. *SIAM Journal on Algebraic Discrete Methods*, 1(2):168–184, 1980.

[148] Stanislaw J. Szarek. Condition numbers of random matrices. *Journal of Complexity*, 7(2):131–149, 1991.

[149] Terence Tao. *Topics in random matrix theory*, volume 132. American Mathematical Soc., 2012.

[150] Terence Tao. The asymptotic distribution of a single eigenvalue gap of a Wigner matrix. *Probability Theory and Related Fields*, 157(1-2):81–106, 2013.

[151] Terence Tao and Van Vu. *Additive Combinatorics*, volume 105. Cambridge University Press, 2006.

[152] Terence Tao and Van Vu. On random ± 1 matrices: singularity and determinant. *Random Structures Algorithms*, 28(1):1–23, 2006.

[153] Terence Tao and Van Vu. On the singularity probability of random Bernoulli matrices. *J. Amer. Math. Soc.*, 20(3):603–628, 2007.

[154] Terence Tao and Van Vu. Random matrices: A general approach for the least singular value problem. *preprint*, 2008.

[155] Terence Tao and Van Vu. Inverse Littlewood-Offord theorems and the condition number of random discrete matrices. *Ann. of Math.*, 169(2):595–632, 2009.

[156] Terence Tao and Van Vu. Random matrices: The distribution of the smallest singular values. *Geometric and Functional Analysis*, 20(1):260–297, 2010.

[157] Terence Tao and Van Vu. A sharp inverse Littlewood-Offord theorem. *Random Structures & Algorithms*, 37(4):525–539, 2010.

[158] Terence Tao and Van Vu. Smooth analysis of the condition number and the least singular value. *Mathematics of computation*, 79(272):2333–2352, 2010.

[159] Terence Tao and Van Vu. Random matrices: universality of local eigenvalue statistics. *Acta Math.*, 206(1):127–204, 2011.

[160] Terence Tao and Van Vu. The Wigner-Dyson-Mehta bulk universality conjecture for Wigner matrices. *Electron. J. Probab.*, 16:no. 77, 2104–2121, 2011.

[161] Terence Tao and Van Vu. The Littlewood-Offord problem in high dimensions and a conjecture of Frankl and Füredi. *Combinatorica*, 32(3):363–372, 2012.

[162] Terence Tao and Van Vu. Random covariance matrices: universality of local statistics of eigenvalues. *Ann. Probab.*, 40(3):1285–1315, 2012.

[163] Terence Tao and Van Vu. Random matrices have simple spectrum. *Combinatorica*, 37(3):539–553, 2017.

[164] Konstantin Tikhomirov. Invertibility via distance for noncentered random matrices with continuous distributions. *Random Structures Algorithms*, 57(2):526–562, 2020.

[165] Konstantin Tikhomirov. Singularity of random Bernoulli matrices. *Ann. of Math.*, 191(2):593–634, 2020.

[166] Roman Vershynin. Invertibility of symmetric random matrices. *Random Structures Algorithms*, 44(2):135–182, 2014.

[167] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge University Press, 2018.

[168] John Von Neumann. *Design of computers, theory of automata and numerical analysis*, volume 5. Pergamon Press, 1963.

[169] Van Vu. Random discrete matrices. In *Horizons of combinatorics*, volume 17 of *Bolyai Soc. Math. Stud.*, pages 257–280. Springer, Berlin, 2008.

[170] Van Vu. Recent progress in combinatorial random matrix theory. *Probability Surveys*, 18:179–200, 2021.

[171] Van Vu and Terence Tao. The condition number of a randomly perturbed matrix. In *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*, STOC '07, page 248–255, New York, NY, USA, 2007. Association for Computing Machinery.

[172] Eugene P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math.*, 62:548–564, 1955.

[173] Eugene P. Wigner. On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics*, 67(2):325–327, 1958.

[174] Eugene P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math.*, 67:325–327, 1958.

[175] Farroll T. Wright. A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric. *The Annals of Probability*, 1(6):1068–1070, 1973.