Random and exact structures in combinatorics

by

Ashwin Sah

B.S., Massachusetts Institute of Technology (2020)

Submitted to the Department of Mathematics
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN MATHEMATICS

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2024

© 2024 Ashwin Sah. This work is licensed under a CC BY-NC-ND 4.0 license.

The author hereby grants to MIT a nonexclusive, worldwide, irrevocable, royalty-free license to exercise any and all rights under copyright, including to reproduce, preserve, distribute and publicly display copies of the thesis, or release the thesis under an open-access license.

Authored by:	Ashwin Sah Department of Mathematics May 15, 2024
Certified by:	Yufei Zhao Associate Professor of Mathematics, Thesis Supervisor
Accepted by:	Jonathan Kelner Professor of Applied Mathematics Chairman, Department Committee on Graduate Theses

Random and exact structures in combinatorics

by

Ashwin Sah

Submitted to the Department of Mathematics
on May 15, 2024 in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN MATHEMATICS

ABSTRACT

In this thesis I aim to show several developments related to notions of randomness and structure in combinatorics and probability. One central notion, the pseudorandomness-structure dichotomy, has played a key role in additive combinatorics and extremal graph theory. More generally, however, such notions come into play in the study of combinatorial probability and the use of random processes in extremal combinatorics. In a broader view, randomness (and the pseudorandomness notions which resemble it along various axes) can be viewed as a type of structure in and of itself which has certain typical and global properties that may be exploited to exhibit or constrain combinatorial and probabilistic behavior.

These broader ideas often come in concert to allow the construction or extraction of exact behavior. I have chosen three directions along which to study this theme: the singularity of discrete random matrices, thresholds for Steiner triple systems, and improved bounds for Szemerédi's theorem. Each concerns breakthroughs in central questions of the fundamental areas of random matrices, combinatorial designs, and additive combinatorics.

Thesis supervisor: Yufei Zhao
Title: Associate Professor of Mathematics

Acknowledgments

I thank Yufei Zhao, my advisor and the key mentor in my journey into research mathematics and extremal combinatorics. Without his immense positive influence, I certainly would not be where I am today. His approach and philosophy to not only mathematics but also the process of research and learning as well as more intangible skills such as communication, writing, and clarity of thought have greatly impacted the way I approach math and life today. I cannot have asked for a better role model to guide me.

Equally important has been the influence of my longtime collaborator, Mehtaab Sawhney, with whom (among others) the three papers composing this thesis–and a majority of my work in graduate school–are joint. In this period I have benefited greatly from bouncing off almost every idea imaginable with Mehtaab, learning new fields and dealing with the struggles of our concerted efforts breaking against an open problem. It seems like there has not been a week gone by without a message about some crazy idea on Slack, or a discussion of how to improve a recent paper, or a session devoted to understanding the point of an unknown area. Truly, the best times in my research journey have been forged through this collaboration.

Throughout my time in graduate school I have also been distinctly influenced by some key collaborators who I have seen in part as mentors: Vishesh Jain, who introduced me to random matrix theory and has influenced my ability to choose research directions broadly; Matthew Kwan, from whom I learned much about matchings and designs, and who has constantly pushed my combinatorial intuition; and Asaf Ferber, whose undaunted optimism has inspired me to attack problems and learn fields that once seemed unimaginably difficult. I also wish to thank Joe Gallian, whose influential mentorship through the 2018 Duluth Research Experience for Undergraduates (REU), and Ken Ono and Jesse Thorner, whose leadership at the 2019 Emory REU contributed to my desire to pursue a graduate degree in mathematics. I additionally thank Dr. John Gorman for helping me in my first steps into higher mathematics.

I would also like to thank all the other collaborators I have met and done math with along this journey, not all of it published: Noga Alon, Aaron Berger, Ross Berkowitz, Matija Bucić, Colin Defant, Jacob Fox, Nate Gillman, Margalit Glasgow, Gopal Goel, Peter Keevash, Tom Kelly, Noah Kravitz, Janardhan Kulkarni, Michael Kural, Mitchell Lee, James Leng, Yang Liu, Vaughan McDonald, Dor Minzer, Assaf Naor, Bhargav Narayanan, Alexandru Pascadi, Sarah Peluse, Junyao Peng, Will Perkins, Huy Tuan Pham, Natesh Pillai, Cosmin Pohoata, Julian Sahasrabudhe, Wojciech Samotij, Lisa Sauermann, Michael Simkin, Aaron Smith, Cynthia Stoner, Jakub Tarnawski, Jonathan Tidor, Daniel Zhu, and Yizhe Zhu.

Important to my time in graduate school have also been academic siblings, officemates, and fellow students: Niven Achenjang, Anna Brandenberger, Evan Chen, Byron Chin, Anlong Chua, Alex Cohen, Travis Dillon, Dingding Dong, Marisa Gaetz, Benjamin Gunby, Arun Kannan, Nitya Mani, Timothy Ngotiaoco, Yannick Yao, Dmitrii Zakharov, and Kai Zheng. I also thank several friends in other graduate schools or programs including Ryan Alweiss, Sanath Devalapurkar, Brice Huang, Cathy Ji, Sujay Kazi, Alice Lin, Benjamin Quiring, Dhruv Rohatgi, Chris Xu, and Calvin Yost-Wolff.

I also thank various MIT faculty I have met in my time both as an undergraduate and graduate student; beyond those I have thanked prior, this includes Henry Cohn, David Gamarnik, Ju-Lee Kim, Ankur Moitra, Elchanan Mossel, Bjorn Poonen, Philippe Rigollet, Nike Sun, and John Urschel.

Finally, I am deeply indebted to my family for making this journey possible in the first place. To my brother, Varun, who has always given me the right advice and been a constant life mentor; to my dad, Anurag, who has been a steady rock throughout my life, and whose innocuous attempt to teach me algebra set me on the path towards research mathematics; and to my mom, Suneeta, who has supported me from the beginning, who first taught me arithmetic and the importance of learning in my childhood, and who is a guiding light in my life. Thank you for being there every step of the way, and helping me to become the person I am today.

Contents

Chapter 1

Introduction

1.1 Outline

The *pseudorandomness-structure dichotomy* has played a key role in the modern development of additive combinatorics and extremal graph theory, highlighted by authors such as Tao [100]. More broadly, as discussed in the abstract, we may view randomness and pseudorandomness as a certain form of structure itself; we develop this narrative further by demonstrating how *random structures* can aid the understanding of *exact structures*, and vice versa. I have chosen three representative works to highlight this narrative. An outline and summary of the key insights of each work in context of this theme constitutes the remainder of this first introductory chapter.

The second chapter is focused on work [41] on "Singularity of discrete random matrices", joint with Vishesh Jain and Mehtaab Sawhney, which in most cases provides an exact characterization of the probabilistic event that certain random matrices are singular. This work has appeared in Geometric and Functional Analysis.

The third chapter involves work [87] on the "Threshold for Steiner triple systems", joint with Mehtaab Sawhney and Michael Simkin, which proves the tight exponent for the *probabilistic threshold* for Steiner triple systems, a form of combinatorial design, to exist with high probability while using only a random set of allowed triples. This work has appeared in Geometric and Functional Analysis.

The fourth and final chapter contains recent work [68] on "Improved bounds for Szemerédi's theorem", joint with James Leng and Mehtaab Sawhney, which gives the first improvement to bounds for Szemerédi's theorem on sizes of sets avoiding arithmetic progressions of length $k \geq 5$ since the breakthrough work of Gowers [25, 26] at the turn of the century. This new work in turn relies on recent improved quasipolynomial bounds by the same authors [69] for the *inverse theorem for the Gowers $U^{s+1}[N]$-norm*, which I have chosen to treat as a black box in this thesis.

1.2 Singularity of discrete random matrices

1.2.1 Summary

The study of *nonasymptotic random matrix theory* has been central in establishing limiting results and behavior of the spectrum of random matrices, such as the circular law. A more detailed history of the combinatorial study of this area, focused on the singularity question, is given in Section 2.1, and we focus on the broad strokes here.

Question 1.2.1. Let ξ be a distribution with finite support on $\mathbb{R}$. What is the probability that an $n \times n$ matrix M with independent entries distributed as ξ is singular, i.e., $\det M = 0$?

The natural conjecture is that the "easiest" way to enforce singularity explains the entire behavior, namely, the event that $\det M = 0$ occurs with probability $(1 + o(1))q$, where q is the probability that M has a left- or right-kernel vector of the form e_i or $e_i \pm e_j$ where e_i, e_j are elementary basis vectors. One can conjecture an even stronger asymptotic expansion graded by the size of the support and height of potential kernel vectors. In the case where $\xi = \mathrm{Ber}(1/2)$ is uniform on $\{0, 1\}$, this implies a prediction of $(1/2+o(1))^n$ for the singularity probability, which was established only recently by Tikhomirov [103].

The content of this work is to extend this to all discrete distributions ξ, and furthermore to prove the stronger $(1 + o(1))q$ version for all ξ which are not uniform on its support (e.g., $\mathrm{Ber}(p)$ for $p \neq 1/2$). In fact, our result is stronger: for certain ξ we can pick up the second-order term in the conjectural asymptotic expansion (in particular, this applies to $\xi = \mathrm{Ber}(p)$ for $p \in (0, 1/2)$).

1.2.2 Overview of proof techniques

In the most naive approach, one wishes to look at each $v \in \mathbb{S}^{n-1}$ and consider the probability that $Mv = 0$, and then add up over all vectors. For a fixed v, the probability of $Mv = 0$ is the same as $\mathbb{P}[v \cdot x = 0]^n$, where x is a vector of n independent copies of ξ. When v has large support, this probability is superexponentially small and if v has constant size support, one can explicitly analyze the potential outcomes. However, the number of $v \in \mathbb{S}^{n-1}$ is infinite (and the number of potential v, after ruling out those which cannot be the kernel of such a matrix, is still quite large). So, these observations do not suffice on their own.

For future reference, we remark that each e_i can be a kernel vector precisely when ξ is supported at 0, and contributes $\mathbb{P}[\xi = 0]^n \leq \|\xi\|_\infty^n$ each, where $\|\xi\|_\infty = \max_{a \in \mathrm{supp}(\xi)} \mathbb{P}[\xi = a]$. Each $e_i - e_j$ can always be a kernel vector and contributes $\|\xi\|_2^{2n}$ each, where $\|\xi\|_2^2 = \sum_{a \in \mathrm{supp}(\xi)} \mathbb{P}[\xi = a]^2$. Each $e_i + e_j$ can be a kernel vector if ξ has two elements summing to 0 in its support. It always contributes at most $\|\xi\|_2^{2n}$, with equality when ξ is a symmetric distribution. Additionally, there are events corresponding to left-kernel vectors instead of the traditional right-kernel.

At a high level, the proof must handle two types of events: M has a "somewhat short" kernel vector (i.e., the approximate support has size at most cn for small c) on either side,

called a *compressible vector*, or the kernel of M on both sides has only *incompressible vectors*.

In the case of compressible vectors, one uses an ε-net and rounding argument to achieve the dream of a union bound over the continuous sphere $\mathbb{S}^{n-1}$; combined with a sharp analysis along the lines of the "naive approach" we can obtain the correct exponential prediction for the compressible sphere. For a sharper estimate, this works for every vector v not in the neighborhood of the special vectors e_i and $e_i \pm e_j$ which may be contributing to the tight probability. For these special vectors, we may use that we are in a close neighborhood to meaningfully improve the analysis. (We remark that technically this sharp analysis requires some level of understanding of incompressible vectors on a submatrix.)

In the more difficult case of incompressible vectors, we adapt methods of Tikhomirov [103] to work for more general distributions. To be more specific, consider the following simplified strategy. We focus on the usual right-kernel (the fact that we have no compressible left-kernel vectors is simply used for an intermediate *distance-to-hyperplane* reduction and related steps). We bucket incompressible vectors $v \in \mathbb{S}^{n-1}$ based on their *threshold*, which is the largest dyadic scale T such that the distribution of $v \cdot x$ at resolution $1/T$ no longer resembles $\mathcal{N}(\mu, \sigma^2)$ where μ, σ^2 are the mean and standard deviation of $v \cdot x$. Specifically, we care that $v \cdot x$ never concentrates in an interval of length $1/T$ by a much larger factor than what $\mathcal{N}(\mu, \sigma^2)$ can do, encoded via the *Lévy concentration function*. For vectors of threshold T, roughly by rounding them to an n-dimensional lattice $(T\sqrt{n})^{-1}\mathbb{Z}^n$ and attempting to run the same argument as above, it turns out that a union bound approach can still work given a key *inversion of randomness* estimate (Theorem 2.2.15) which essentially bounds the number of vectors in this lattice that actually have approximately the required threshold T and shows these occur at a superexponential level of sparsity.

The higher thresholds T we can handle, the smaller the error term in our estimate of $\mathbb{P}[\det M = 0]$ will be (it is roughly $1/T$ up to subexponential terms). Unfortunately, this generalization of Tikhomirov's inversion of randomness only goes up to thresholds $T \approx (\|\xi\|_\infty^{-1} - o(1))^n$, and this is best-possible for thresholds against rows x that have n independent entries. So, with this we can only handle a restricted class of distributions (roughly, those where $\mathbb{P}[\xi = 0] = \|\xi\|_\infty$); furthermore, we do not obtain the sharp version of the result for these distributions.

In order to go further, we need to be more discerning and utilize the fact that it is very rare for x to be very "atypical" for more than a few rows x of M. For example, with superexponentially good probability, all but $O(1)$ rows x of M are such that the multiset of elements of x are roughly in proportion to the distribution of ξ. Thus, in reality we use a *multi-slice* generalization of inversion of randomness (Theorem 2.2.1) which along with proper conditioning arguments allows us to push the argument through. This works since the best-possible threshold for such a statement is $T \approx \exp((H(\xi) - o(1))n)$ where $H(\xi)$ is the entropy of ξ. For ξ which are non-uniform, $\exp(-H(\xi)) < \|\xi\|_2^2 < \|\xi\|_\infty$ and this allows us to obtain an error term which is less than the contribution from the main term coming from the compressible vector cases (and in some cases, the second term of the asymptotic expansion). Here, a slice distribution is something like taking a uniform distribution on $\{0,1\}^n$ and conditioning to have sum in the range $(p \pm \gamma)n$, instead of using the independent

vector $\mathrm{Ber}(p)^{\otimes n}$. A multi-slice distribution is similar but with potentially larger support than $\{0,1\}$. This requires *rerandomization* arguments implemented delicately to avoid loss in the relevant estimates. (Earlier methods of Litvak and Tikhomirov [71] in some sense perform a cruder *switching* argument on top of inversion of randomness to handle $\mathrm{Ber}(p)$ for p of small constant order.)

The proof of the key inversion of randomness result is quite technical, but there are three key steps at a high level. First, we replace the notion of Lévy concentration with a log-Lipschitz smoothed version. Second, we prove an initial base estimate which counts the possible choices for most of the coordinates of vectors of threshold T allowing a weaker threshold. Third, we process the rest of the vector in small roughly linear-sized traunches, using the chunks to iteratively smooth the notion of threshold to obtain the correct condition on the threshold.

Thus, for the singularity problem we derive strong structure in a random distribution by isolating the key contributing factors (the compressible event), and demonstrating uniform global control in the "generic" portion of the distribution (the incompressible event). The latter step is accomplished by reformulating a question about vectors satisfying certain threshold structure using inversion of randomness.

1.3 Threshold for Steiner triple systems

1.3.1 Summary

The study of *designs* is one of the oldest areas of combinatorics. Indeed, one of the first theorems in combinatorics is that a *Steiner triple system* of order n exists precisely when $n \equiv 1, 3 \pmod 6$, due to Kirkman [58]. That is, for such n there is a collection of size 3 subsets of $[n] = \{1, \dots, n\}$ such that every pair is included in exactly one triple. The construction given by Kirkman is explicit, but much attention has been given to understanding less-structured combinatorial designs since much of the history of constructing designs has utilized algebraic or explicit methods. For instance, Steiner triple systems are closely related to Latin squares: we can view a Steiner triple system as a decomposition of the edges of the complete graph K_n into triangles, and a Latin square as a decomposition of the edges of the complete tripartite graph $K_{n,n,n}$ into triangles. In fact, until a recent breakthrough result of Keevash [54] (and a second proof by Glock, Kühn, Lo, and Osthus [24]), Steiner systems with more general parameters were not known to exist in most cases. We defer more detail to Section 3.1.

One question of particular interest concerns how random such objects can be, in various senses. One natural notion of randomness has to do with the *probabilistic threshold* for existence: if we sample each size 3 subset of $[n]$ with probability p, will there exist a Steiner triple system only using these subsets with high probability? This is closely related to problems on matchings and hypergraph matchings such as Shamir's problem, for which the answer is that this holds for p roughly larger than $(\log n)/n$.

Determining the threshold for existence of Steiner triple systems, and relatedly Latin squares, is a folklore question that was in particular asked by Johansson and highlighted in

Keevash's 2018 ICM talk. Similar considerations to matchings suggest that the probabilistic threshold must be at least $(\log n)/n$, which was conjectured to be correct.

The content of this work is to show that for an appropriate choice of $p = n^{-1+o(1)}$, with high probability a Steiner triple system does exist among p-sampled size 3 subsets of $[n]$ (and similar for Latin squares), determining the correct asymptotic constant. We remark that in later work, the upper bound was improved to $(\log n)^2/n$ [52] and then to the optimal order of magnitude $(\log n)/n$ [39, 55].

1.3.2 Overview of proof techniques

The requirement to construct fairly random-looking Steiner systems means that it is not enough to follow the classical explicit constructions. Additionally, the approach of Keevash [54] for constructing designs utilizes sampling from algebraically structured *templates* which simply do not exist at the optimal level of sparsity $p = n^{-1+o(1)}$. Thus, we start with the *iterative absorption* approach introduced by Glock, Kühn, Lo, and Osthus [24].

To construct a Steiner triple system, without concern for the p-sampling, we perform the following procedure. First, we fix a *vortex* $[n] = V_0 \supseteq \cdots \supseteq V_\ell$ of sets, each decreasing by roughly a large constant factor each step until the final set V_ℓ of very small size (say less than $\sqrt{\log n}$). Then we plant an *absorber* A in the edges of $K_{V_0} \setminus K_{V_1}$ with the property that for any possible *remainder* graph X on V_ℓ (specifically, a *triangle-divisible* graph with even degrees and number of edges a multiple of 3), the union $A \cup X$ can be decomposed into triangles. This can be done by disjointly embedding single-graph absorbers B_X with the property that $B_X \cup X$ and B_X both have triangle-decompositions, which are not hard to explicitly construct.

Then, our goal is to iteratively *cover down* by using all edges in $K_{V_0} \setminus (K_{V_1} \cup A)$, then all edges in $K_{V_1} \setminus K_{V_2}$, and so on until we reach a subset of K_{V_ℓ} which will be our X. Each step of the cover down is similar, and merely requires that the remaining graph G_i satisfies some pseudorandomness properties with respect to the current set V_i and the next set V_{i+1} (and technically, the future sets as well). It is executed in four steps.

First, we set aside a random *reserve graph* R_i which is bipartite between $V_i \setminus V_{i+1}$ and V_{i+1} and ensure that the random outcome satisfies various pseudorandomness conditions. Second, we use a random triangle removal process on $G_i \setminus R_i$ (with a slight bias to account for the fact that we are avoiding V_{i+1} and the reserve graph) to remove all but a small fraction of edges, smaller than the reserve graph. Furthermore, the outcome of this random process is pseudorandom according to various heuristics, so that e.g. the degree of the remainder at each vertex is small and roughly what it should be. Third, we use a random greedy process to cover every remaining edge in $V_i \setminus V_{i+1}$ via an extension to a vertex in V_{i+1} within the reserve graph (and again, ensure the outcome is appropriately pseudorandom).

Fourth and finally, every vertex in $V_i \setminus V_{i+1}$ has a neighborhood within V_{i+1} and we must cover the resulting edges. For each vertex, covering those edges with triangles corresponds to finding a matching in the neighborhood within the induced graph $G_i[V_{i+1}]$. This can be done since we were sure to include the reserve graph R_i which provides enough options

and pseudorandomness for each vertex. In fact, we use a greedy random process to take matchings for each vertex in some order, and use the pseudorandomness properties to show that this is likely to be successful in producing matchings which use disjoint edges across the different vertices (so that the resulting triples never reuse a pair).

Again, studying typical outcomes of this sequence of random processes allows us to conclude that a suitable version of this strategy will succeed with high probability and ensure that the next stage has appropriate pseudorandomness conditions on the resulting G_{i+1} to continue this process downwards.

Now, we turn to a discussion of introducing the concept of threshold. If we p-sample and try to find a Steiner system among limited choices, then optimizing all possible constructions and various other parts of the proof still leaves us stuck at roughly $p = n^{-1/2+o(1)}$ due to the requirement of having absorbers such as B_X above. (Furthermore, to even get to this, due to other issues with the construction one needs some of the ideas detailed below anyway.)

Instead, we utilize a recent breakthrough relating thresholds to *spread distributions*, or equivalently the *fractional expectation-threshold*, due to Frankston, Kahn, Narayanan, and Park [19]. (One can also use the more recent resolution of the Kahn–Kalai conjecture due to Park and Pham [77], but this is not needed.) A q-spread distribution μ on a family $\mathcal{H} \subseteq 2^X$ is a distribution such that $\mathbb{P}_{A\sim\mu}[S \subseteq A] \leq q^{|S|}$ for all $S \subseteq X$. Such a distribution is "dominated" by q-sampling in an appropriate sense. On the other hand, the *threshold* of a nontrivial family is the value p such that a p-sample of X contains an element of $\mathcal{H}$ with probability $1/2$. The result of [19] is that the threshold is $O(q \log \ell)$ if both $|S| \leq \ell$ for all $S \in \mathcal{H}$ and there is a q-spread distribution supported on $\mathcal{H}$.

Thus, instead of exhibiting a Steiner system in a random sample (threshold), we may instead focus on producing a spread distribution of many Steiner systems. In fact, the iterative absorption procedure described above already has many elements amenable to spread, such as the fact that most choices (except for the absorber) are made using random processes that satisfy various pseudorandomness estimates. Careful analysis can make this notion more precise. As it turns out, the key issue becomes finding out how big we can afford to let V_ℓ be: the spread that we can afford is morally closer to $n^{-1+o(1)}$ the closer $|V_\ell|$ is to $n^{1-o(1)}$.

Unfortunately, the strategy described above requires roughly $|V_\ell| = O(\sqrt{\log n})$. A more efficient naive absorption strategy can achieve $|V_\ell| = O(n^c)$ for some absolute $c > 0$, but this is nowhere near good enough.

The second key insight is that instead of using an absorber A, the final remainder graph X is itself a near-complete triangle-divisible graph. Thus, the iterative absorption argument for existence of Steiner triple systems (which is somewhat robust) can be applied to X as a black box to finish the construction. Of course, as stated this is merely a sleight of hand: if we unwrap the black box then there will be another vortex inside X until some point where there is an even smaller "final set" and there is still an absorber.

However, if we have some bound on the threshold for Steiner triple systems (or some appropriate generalization), then we know that we can find a decomposition of X that only uses a sample of the triangles. This is a bit more constraining than before, but again if we unwrap this argument then we can never improve the spread of our constructions by using

a black box construction for X.

Therefore, we introduce the concept of *spread boosters*. Essentially, we do plant something that serves as A with respect to X, but it is generated by randomly sampling triangles and including their edges (so that it is "inherently spread" to a near optimal degree). The purpose of A is not to complete the triangle-decomposition *per se*, but rather to replace triangles within X with triangles that are more "spread". This, converted to a boosting procedure, allows us to iteratively improve the spread of the construction to $n^{-1+o(1)}$.

Thus, for thresholds for Steiner triple systems we can create exact structures such as combinatorial designs using random processes (and some absorption to "fix" the structure). The success of such a procedure in turn relies on various global heuristics and pseudorandomness properties of the outcomes of these processes. Finally, converting this to a procedure to construct spread probability measures and performing spread boosting enables us to bound the probabilistic threshold which is measured with respect to ordinary Bernoulli measures.

1.4 Improved bounds for Szemerédi's theorem

1.4.1 Summary

Szemerédi's theorem states that in a subset of the integers with positive upper density, there exist arbitrarily long arithmetic progressions. This answers a classical question of Erdős and Turán. We defer more detail to Section 4.1, but this is a foundational question and quantitative bounds for this question remain central to the field of additive combinatorics. Study of this and related questions involving the *inverse theorem for the Gowers $U^{s+1}[N]$-norm* led to, among other things, the proof of the celebrated Green–Tao theorem on primes in arithmetic progressions.

To be more precise regarding quantitative bounds, we let $r_k(N)$ be the size of the largest subset of $[N] = \{1, \dots, N\}$ with no nontrivial k-term arithmetic progression. Prior to this work, the best known bounds were $r_3(N) \ll N \exp(-c(\log N)^{1/9})$ due to recent work of Bloom and Sisask [9] improving on a breakthrough of Kelley and Meka [57], $r_4(N) \ll N(\log N)^{-c}$ due to Green and Tao [28, 32], and $r_k(N) < N(\log\log N)^{-2^{-2^{k+9}}}$ due to seminal work of Gowers [25, 26].

We establish the bound

$$r_k(N) \ll N \exp(-(\log\log N)^{c_k})$$

for all $k \geq 5$, providing an improvement to the bounds for Szemerédi for general progressions for the first time since the work of Gowers. The key input to this is a new quasipolynomial bound on the inverse theorem for the Gowers U^{s+1}-norm [69] in companion work. In application, one takes $s = k-2$ (for instance, the proof of Roth's theorem on three-term arithmetic progressions can be seen as utilizing the proof of the U^2-inverse theorem, which follows from basic Fourier analysis).

1.4.2 Overview of proof techniques

Define

$$\Lambda_k(f_1, \dots, f_k) = \mathbb{E}_{x,y \in \{0,\dots,N\}} \prod_{j=1}^{k} f_j(x + (j-1)y)$$

for functions $f_j \colon [N] \to \mathbb{C}$, where we extend functions by 0 to $\mathbb{Z}$. We will generally be interested in 1-bounded functions such as indicators of sets or shifts of indicators.

The proof of Roth's theorem for three-term arithmetic progressions follows a *density increment* strategy. Given $A \subseteq [N]$ of size αN, or *density* α, we write

$$\Lambda_3(\mathbb{1}_A, \mathbb{1}_A, \mathbb{1}_A) = \Lambda_3(\alpha \mathbb{1}_{[N]}, \alpha \mathbb{1}_{[N]}, \alpha \mathbb{1}_{[N]}) + \cdots,$$

where there are 7 terms in the remainder each of the form $\Lambda_3(f_1, f_2, f_3)$ where one of the f_j is of the form $\mathbb{1}_A - \alpha \mathbb{1}_{[N]}$ and the rest are 1-bounded.

The first term is roughly α^3 in size up to a constant factor, so it is easy to see that we have a three-term arithmetic progression unless one of the remainder terms is large in magnitude. By an application of the *Gowers–Cauchy–Schwarz inequality* (e.g. Lemma 4.3.3) we see that the *Gowers U^2-norm* of $\mathbb{1}_A - \alpha \mathbb{1}_{[N]}$ is large. Or, we can merely use Fourier inversion here.

Then, by the inverse theorem for the Gowers U^2-norm, this function must correlate with a Fourier phase $n \mapsto \exp(2\pi i \theta n)$. Finally, by a *rational approximation* argument using Dirichlet's theorem, we may find a long subprogression $P \subseteq [N]$ on which A has a density increment, namely something like $|A \cap P|/|P| \geq \alpha + c\alpha^2$. Then we are done, since we can iterate this until the density hits 1 and we are guaranteed to find a progression by that point.

The original proof of Roth gives $r_3(N) \ll N/\log\log N$, but this was improved by Szemerédi [98] and Heath-Brown [35] by a strategy which extracts a set of multiple Fourier phases instead of just one, giving a multiplicative density increment. A robust version of this which can be extended to higher progressions was given in the work of Green and Tao [28] on progressions of length four.

We use this approach, and replace Fourier analysis (or the U^2-inverse theorem) with the U^{k-1}-inverse theorem. In order to obtain the desired bounds, we require that the correlation with a single degree $k-2$ *nilsequence* (which replaces Fourier phases) is quasipolynomial, and require appropriate complexity bounds on said nilsequence.

A higher degree nilsequence is essentially derived from polynomial phases such as the function $\exp(2\pi i \theta n^2)$, except we also allow *bracket polynomials* such as $\exp(2\pi i \alpha n \lfloor \beta n \rfloor)$ or $\exp(2\pi i \alpha n \lfloor \beta n \rfloor \lfloor \gamma n \lfloor \delta n \rfloor \rfloor)$, and combinations of such. As it turns out, the proof of the inverse theorems is only amenable when we represent such functions via polynomial sequences valued on nilpotent Lie groups of appropriate *step*, leading to the theory of nilsequences. For simplicity we do not elaborate further on this point and refer the interested reader to [33,69].

Beyond this broad approach, in order to be able to efficiently pass to subprogressions we must prove a robust rational approximation result for multiple nilsequences, which involves proving a *Schmidt-type decomposition* result. This is accomplished via an *iterative Schmidt refinement*. At a high level, we first use a more classical rational approximation argument

to decompose $[N]$ into subprogressions within which multiple pure floor functions such as $\lfloor \beta n \rfloor$ and $\lfloor \delta n \rfloor$ and $\lfloor \theta n^2 \rfloor$ behave purely as polynomials without rounding (perhaps with very different coefficients). Then, on each subprogression we plug in these functions into the nested bracket polynomials, simplify, and apply a similar argument to the current innermost polynomials. Upon iteration, this provides a decomposition into subprogressions on which our nilsequences are almost constant, in a way that yields good enough bounds for our argument.

Chapter 2

Singularity of discrete random matrices

2.1 Introduction

Let $M_n(\xi)$ be an $n \times n$ random matrix, each of whose entries is an independent copy of a random variable ξ. We will restrict attention to when ξ is a real-valued random variable whose support is finite and contains at least two points (which we call *discrete*). What is the probability that $M_n(\xi)$ is singular? This question, which has been studied since the 1960s, has attracted considerable attention over the years. A well-known folklore conjecture is that the dominant contribution to the probability of singularity is from the events that a row or column is zero, or that two rows or two columns are equal (possibly up to a sign). In order to facilitate discussion, let us introduce some notation. For a vector $v \in \mathbb{R}^n$, we define the event

$$\mathcal{E}_v := \{M_n(\xi)v = 0\}.$$

We will also denote the canonical basis vectors of $\mathbb{R}^n$ by $e_1, \dots, e_n$. Then, the aforementioned conjecture may be stated as follows.

Conjecture 2.1.1. *Let ξ be a discrete random variable, and let $M_n(\xi)$ be an $n \times n$ random matrix whose entries are independent copies of ξ. Then*

$$\mathbb{P}[M_n(\xi) \text{ is singular}] = (1 + o(1))\Big(2n\mathbb{P}[\mathcal{E}_{e_1}] + n(n-1)\mathbb{P}[\mathcal{E}_{e_1-e_2}] + n(n-1)\mathbb{P}[\mathcal{E}_{e_1+e_2}]\Big).$$

In this paper, as our first main result, we confirm a stronger version of Conjecture 2.1.1 for *all* discrete distributions which are not uniform on their support. Let $s_n(M_n)$ denote the least singular value of an $n \times n$ matrix M_n; recall that $s_n(M_n) = \inf_{x \in \mathbb{S}^{n-1}} \|M_n x\|_2$, where $\mathbb{S}^{n-1}$ denotes the unit sphere in $\mathbb{R}^n$ and $\|\cdot\|_2$ denotes the standard Euclidean norm on $\mathbb{R}^n$.

Theorem 2.1.2. *Let ξ be a discrete random variable which is not uniform on its support. There exist $c_\xi, C_\xi > 0$ so that for all sufficiently large n, and for all $t \geq 0$,*

$$\mathbb{P}[s_n(M_n(\xi)) \leq t/\sqrt{n}] \leq C_\xi t + 2n\mathbb{P}[\mathcal{E}_{e_1}] + (1 + O(\exp(-c_\xi n)))\Big(n(n-1)\mathbb{P}[\mathcal{E}_{e_1+e_2}] + n(n-1)\mathbb{P}[\mathcal{E}_{e_1-e_2}]\Big).$$

By applying Theorem 2.1.2 with $t = 0$ for the upper bound, and considering the probability that a row or column is zero, or that two rows or two columns are the same (up to a sign) for the lower bound (cf. the corresponding calculation in [71, Section 3.2]), we thus establish the following strengthening of Conjecture 2.1.1 for discrete distributions which are not uniform on their support.

Theorem 2.1.3. *Let ξ be a discrete random variable which is not uniform on its support. There exists $c_\xi > 0$ such that*

$$\mathbb{P}[M_n(\xi) \text{ is singular}] = 2n\mathbb{P}[\mathcal{E}_{e_1}] + (1 + O(\exp(-c_\xi n)))\Big(n(n-1)\mathbb{P}[\mathcal{E}_{e_1+e_2}] + n(n-1)\mathbb{P}[\mathcal{E}_{e_1-e_2}]\Big).$$

Let us record the consequence of this theorem for the special case of $\xi = \mathrm{Ber}(p)$, which has attracted considerable attention.

Theorem 2.1.4. *Fix $p \in (0, 1/2)$. There exists $c_p > 0$ such that*

$$\mathbb{P}[M_n(\mathrm{Ber}(p)) \text{ is singular}] = 2n(1-p)^n + (1 + O(\exp(-c_p n)))n(n-1)(p^2 + (1-p)^2)^n.$$

Remark 2.1.5. As discussed in Section 2.1.1, the above theorem resolves a conjecture of Litvak and Tikhomirov [71, Problem 8.2], thereby completing the program of determining the (dominant) mechanism leading to the singularity of sparse Bernoulli matrices. In fact, the above theorem provides the first *two* terms in the asymptotic expansion of the singularity probability of $M_n(\mathrm{Ber}(p))$; a result of this precision was not available before in any context.

Theorem 2.1.6. *Fix $p \in (1/2, 1)$. There exists $c_p > 0$ such that*

$$\mathbb{P}[M_n(\mathrm{Ber}(p)) \text{ is singular}] = (1 + O(\exp(-c_p n)))n(n-1)(p^2 + (1-p)^2)^n.$$

Remark 2.1.7. The above theorem provides the leading term in the asymptotic expansion of the singularity probability of $M_n(\mathrm{Ber}(p))$. Prior to this work, even the correct value of

$$\lim_{n\to\infty} n^{-1} \log \mathbb{P}[M_n(\mathrm{Ber}(p)) \text{ is singular})]$$

had not been determined; compared to the true value of $(p^2 + (1-p)^2)$ for this quantity, the previous best-known result of Bourgain, Vu, and Wood [13] provides a weaker upper bound of $\sqrt{p}$. The reason that the case $p \in (1/2, 1)$ is more challenging than $p \in (0, 1/2)$ (treated in [6, 36, 71], see the discussion below) is that in the former case, the dominant contribution to the probability of singularity comes from the event of *two* rows or columns being equal to each other, whereas in the latter case, the dominant contribution comes from the much simpler event of a single row or column being zero.

For general discrete distributions, we determine the value of $\lim_{n\to\infty} n^{-1} \log \mathbb{P}[M_n(\xi) \text{ is singular}]$. The only case not covered by Theorem 2.1.2 is that of uniform distributions, which we handle with a non-exact main term.

Theorem 2.1.8. *Let ξ be a discrete random variable. There exists $C_\xi > 0$ such that for any fixed $\epsilon > 0$ and for all sufficiently large n and all $t \geq 0$,*

$$\mathbb{P}[s_n(M_n) \leq t/\sqrt{n}] \leq C_\xi t + 2n\mathbb{P}[\mathcal{E}_{e_1}] + (1+\epsilon)^n \mathbb{P}[\mathcal{E}_{e_1-e_2}].$$

Remark 2.1.9. For non-uniform discrete distributions, Theorem 2.1.2 is strictly stronger.

2.1.1 Previous work

Let us put Theorems 2.1.2 and 2.1.8 in the context of known results. For convenience, we will use $q_n(\xi)$ to denote $\mathbb{P}[M_n(\xi) \text{ is singular}]$. The work of Komlós [60] was the first to show that $q_n(\mathrm{Ber}(1/2)) = o(1)$. Much later, an exponential bound on $q_n(\mathrm{Ber}(1/2))$ was obtained by Kahn, Komlós, and Szemerédi [50]. Subsequently, the base of the exponent was improved to 0.939 and $3/4+o(1)$ in a series of works by Tao and Vu [101,102], and later to $1/\sqrt{2}+o(1)$ by Bourgain, Vu, and Wood [13]. Finally, a truly breakthrough result of Tikhomirov [103] in 2018 established that $q_n(\mathrm{Ber}(p)) = (1-p+o(1))^n$ for fixed $p \in (0, 1/2]$. As mentioned earlier, for fixed $p \in (1/2, 1)$, the analogous result was not known prior to this work.

Conjecture 2.1.1 has been most accessible for sparse Bernoulli distributions, in which case, the right hand side simplifies considerably to $(1+o(1)) \cdot 2n\mathbb{P}[\mathcal{E}_{e_1}]$. Here, by the Bernoulli distribution with parameter p, which we will henceforth denote by $\mathrm{Ber}(p)$, we mean the two point distribution which attains the value 1 with probability p and the value 0 with probability $1-p$. Basak and Rudelson [6] confirmed the conjecture for $\xi = \mathrm{Ber}(p_n)$ for p_n in a certain range of sparsity limited to $n^{-1}\ln n - n^{-1}g(n) \le p_n \le n^{-1}\ln n + o(n^{-1}\ln\ln n)$, where $g(n)$ is some function which grows slowly with n. Subsequently, Litvak and Tikhomirov showed that the conjecture also holds for $\xi = \mathrm{Ber}(p_n)$ for $Cn^{-1}\ln n \le p_n \le c$, where $c > 0$ is a small absolute constant and $C > 0$ is a large absolute constant. Recent work of Huang [36] was able to bridge the gap between the regimes covered in [6] and [71], leaving open the regime $p \in (c, 1/2)$. Establishing Conjecture 2.1.1 (as opposed to the stronger Theorem 2.1.4) in this case does not require the full strength of the ideas in this paper; in particular the treatment of the 'compressible' part of the unit sphere is substantially simpler. Since this is a case of particular interest (see [71, Problem 8.2]), we have isolated the proof (given Theorem 2.2.1) of Conjecture 2.1.1 for sparse Bernoulli random variables in Section 2.3, which also serves as a 'warm-up' to subsequent sections.

For general discrete distributions ξ, the only previous systematic study in the literature is the work of Bourgain, Vu, and Wood [13]. They show [13, Corollary 1.2] that if ξ is a discrete distribution with $\sup_{r\in\mathbb{R}} \mathbb{P}[\xi = r] =: p$, then $q_n(\xi) \le (\sqrt{p}+o(1))^n$, which is far from optimal (the true bound is never more than $(p+o(1))^n$, although it may be much smaller). On the other hand, up to a possible $o(1)$ term, Theorem 2.1.8 in this work always obtains the correct base of the exponent.

For certain specific distributions, Bourgain, Vu, and Wood obtain the correct base of the exponent (again, up to a $o(1)$ term). Specifically, they show [13, Corollaries 3.1, 3.2] that if $\xi_{1,\mu}$ is a random variable taking on the value 0 with probability $1-\mu$ and ± 1 with probability $\mu/2$ each, and if $\xi_{2,\mu}$ is a random variable taking on the value 0 with probability $1-\mu$ and $\pm 1, \pm 2$ with probability $\mu/4$ each, then $q_n(\xi_{1,\mu}) = (1-\mu+o(1))^n$ for all $\mu \in (0, 1/2)$ and $q_n(\xi_{2,\mu}) = (1-\mu+o(1))^n$ for all $\mu \in (0, 16/25)$. For these random variables, Theorem 2.1.8 determines the correct base of the singularity probability for all fixed $\mu \in (0,1)$, and Theorem 2.1.2 determines the leading order in the asymptotic expansion for $\mu \in (0,1), \mu \ne 2/3$ in the first case, and $\mu \in (0,1), \mu \ne 4/5$ in the second case. In fact, for $\mu \in (0, 2/3)$ in the first case, and $\mu \in (0, 4/5)$ in the second case, Theorem 2.1.2 determines

the first two terms in the asymptotic expansion.

We remark that the results of [13] such as [13, Corollary 1.2] are also applicable to discrete random variables valued in the complex numbers, and settings where the entries of $M_n(\xi)$ are not identically distributed, and a small number of rows of $M_n(\xi)$ are possibly deterministic; we have not pursued these extensions.

Finally, we remark that there was a recent paper of Irmatov [37] which claimed to resolve Conjecture 2.1.1 for Rademacher random matrices. Experts have informed us that there are some unresolved issues in that work that its author is aware of, including [37, Theorem 3]. Furthermore, upon slight modification, the proof in [37] would appear to give impossibly good error terms.

2.1.2 Additional results

The next result addresses the main question left open by our work, namely, the resolution of Conjecture 2.1.1 for discrete distributions ξ which are uniform on their support. In this direction, we provide a sharp analysis of the contribution of a certain low-entropy part of the unit sphere; in fact, it is this contribution which forms the leading term of the conjectured asymptotic expansion of the singularity probability. This theorem is also central to the proofs of Theorems 2.1.2 and 2.1.8.

Theorem 2.1.10. *Fix a discrete distribution ξ. There exist $\delta, \rho, \eta > 0$ depending on ξ such that for all sufficiently large n and $t \leq 1$,*

$$\mathbb{P}\left[\inf_{x \in \mathrm{Cons}(\delta,\rho)} \|M_n(\xi)x\|_2 \leq t\right] \leq n\mathbb{P}[\mathcal{E}_{e_1}] + \binom{n}{2}(\mathbb{P}[\mathcal{E}_{e_1-e_2}] + \mathbb{P}[\mathcal{E}_{e_1+e_2}]) + (t + \mathbb{P}[\mathcal{E}_{e_1-e_2}])e^{-\eta n}.$$

The set $\mathrm{Cons}(\delta, \rho)$ appearing above is the set of unit vectors which have at least $(1-\delta)n$ coordinates within distance $\rho/\sqrt{n}$ of each other (see Definition 2.3.2), although a trivial modification shows this result holds for any sufficiently low-entropy subset of the unit sphere.

Our techniques also lend themselves naturally to studying a certain model of random matrices with combinatorially dependent entries. Let Q_n denote a random matrix with independent rows, each of which is chosen uniformly from among those vectors in $\{0,1\}^n$ which have sum exactly $\lfloor n/2 \rfloor$. In [76], Nguyen showed that $\mathbb{P}[Q_n \text{ is singular}] = O_C(n^{-C})$ for any $C > 0$, and conjectured [76, Conjecture 1.4] that $\mathbb{P}[Q_n \text{ is singular}] = (1/2+o(1))^n$. After intermediate work [18, 38], an exponential upper bound on the singularity probability was only very recently obtained in work of Tran [104]. Our next result settles [76, Conjecture 1.4].

Theorem 2.1.11. *For every $\epsilon > 0$, there exists C_ϵ depending on ϵ such that for all sufficiently large n, and for all $t \geq 0$,*

$$\mathbb{P}[s_n(Q_n) \leq t/\sqrt{n}] \leq C_\epsilon t + (1/2 + \epsilon)^n.$$

2.1.3 Overview of the techniques

As in many works in this area, we use the high-level strategy (going back to Kašin [53] and subsequently used in [70, 85, 86, 90]) of dividing the unit sphere into 'structured' and 'unstructured' components, and estimating the contribution of each part separately. However, compared to previous works, the treatment of both components require overcoming significant obstacles which unavoidably arise in the sharp analysis of the invertibility of random matrices in any amount of generality.

For instance, in the analysis of structured vectors, we need to additionally capture the event that two rows/columns of the matrix are equal (up to a sign) whereas previous considerations of sharp invertibility only addressed scenarios where the dominant contribution to the probability of singularity is due to a single row or column being zero. As discussed in the remark after Theorem 2.1.3, this is a fundamental issue. Moreover, in the analysis of unstructured vectors, we need precise metric entropy estimates for the anti-concentration problem with respect to random vectors on general multi-slices. Obtaining partial estimates of this nature (which are not sufficient to prove Conjecture 2.1.1) even for the special case of the Boolean slice is already a highly non-trivial endeavor which is at the heart of the recent work of Litvak and Tikhomirov [71], where it is accomplished using the substantially more involved notion of the 'UDLCD'.

Structured vectors: The structured vectors in our work are 'almost-constant vectors' i.e. those vectors on $\mathbb{S}^{n-1}$ which have $(1-\delta)n$ coordinates within distance $\rho/\sqrt{n}$ of each other, where $\delta, \rho > 0$ are sufficiently small constants. This class of structured vectors arises naturally in the consideration of the anti-concentration property of a sequence of numbers with respect to a random vector constrained to lie in a 'slice'. Moreover, since vectors which are close to the standard basis vectors e_i or to $e_i \pm e_j$ clearly play a special role in the problem under consideration, it is natural to separately handle 'elementary' and 'non-elementary' structured vectors.

Our treatment of structured vectors, culminating in Theorem 2.1.10, requires significant innovations compared to previous works on the sharp invertibility of sparse random Bernoulli matrices [6,36,71] – in the sparse Bernoulli case, the corresponding class of elementary vectors only needs to consist of those vectors which are close to some e_i, and the largest atom of the the random variable $\mathrm{Ber}(p)$ is conveniently at 0.

In the present work, in order to handle non-elementary vectors, we need to develop novel sharp anticoncentration estimates Propositions 2.6.2 and 2.6.3. (In contrast, the essentially standard estimate Lemma 2.3.7 is sufficient for the case of sparse Bernoulli random variables at the corresponding step). Even more involved is the analysis of elementary vectors, for which we develop a new technique. Let us begin by discussing this technique for $\xi = \mathrm{Ber}(p)$ for fixed $p \in (0, 1/2)$, in which case, the elementary vectors are those which are close to some standard basis vector. For concreteness, consider vectors which are sufficiently close to e_1. We show that, if any such vector has exponentially small image, then either the first column of the matrix is the zero vector, or it must belong to a universal subset of nonzero vectors of $\{0,1\}^n$ of measure at most $(1-p+\epsilon)^n$. The first case corresponds to the term

$\mathbb{P}[\mathcal{E}_{e_1}]$ in Conjecture 2.1.1; for the second case, we leverage the seminal work of Rudelson and Vershynin to show that, on our event, the probability that any vector in this universal subset appears as the first column of the matrix is at most $\exp(-4\epsilon n)$, at which point we can conclude using the union bound.

Of course, for general discrete random variables ξ, one must enlarge the class of elementary vectors to include unit vectors which are close to $(e_i \pm e_j)/\sqrt{2}$ and unit vectors which are close to e_i. In the first case (Propositions 2.6.5 and 2.6.7), we use a rotation trick to reduce to a situation where we can use an analysis similar to (but more complicated than) the one outlined in the previous paragraph. The second case requires a very careful treatment since we are aiming for a leading term of the form $(\mathbb{P}[\xi = 0])^n$ (as opposed to $(\sup_{r\in\mathbb{R}} \mathbb{P}[\xi = r])^n$), and moreover, the desired error is $(\mathbb{P}[\xi = \xi'] - \eta)^n$ which may be very small. To accomplish this, we first prove a version of Theorem 2.1.8 with an estimate on the singularity probability of the form $(\sup_{r\in\mathbb{R}} \mathbb{P}[\xi = r] + o(1))^n$ (Proposition 2.5.4 and Theorem 2.5.5), and then leverage these preliminary estimates to obtain the desired bound.

We emphasize that our treatment of structured vectors, as captured by Theorem 2.1.10, is not sensitive to the non-uniformity of the distribution ξ. In particular, given Theorems 2.1.2 and 2.1.10, the only missing case in the complete resolution of Conjecture 2.1.1 (in fact, in a stronger form) is a sharp analysis of unstructured vectors in the case when ξ is uniform on its support.

Unstructured vectors: The unstructured vectors are the complement of the structured vectors i.e. those which do not have a $(1-\delta)$-fraction of their coordinates within $\rho/\sqrt{n}$ of each other. Our treatment of these vectors relies on the non-uniformity of ξ by exploiting the gap between $\mathbb{P}[\xi = \xi']$ and the entropy of ξ; the idea to exploit such a gap to prove sharp invertibility results (in the case of Bernoulli random variables) is due to Litvak and Tikhomirov [71].

The main ingredient in our work for handling such vectors is Theorem 2.2.1, which is an extension of [103, Theorem B] to a (real) multislice, i.e., the set of vectors in $\{a_1, \dots, a_k\}^n$ which have a prescribed number of coordinates taking on each of the values $a_1, \dots, a_k$. Such a result was previously not known even for the Boolean slice; indeed, the work [71] uses a rather involved notion of arithmetic structure to study anti-concentration on Boolean slices, which is not powerful enough to handle slices that are *not* very far from the central slice. We remark that in general, even establishing much less precise versions of [103, Theorem B] on the Boolean slice has been very challenging, despite much work due to the natural connection to certain combinatorial models of random matrices (cf. [43] and the references therein).

Compared to [103, Theorem B], we need to overcome two challenges. The first, as mentioned above, is the lack of independence between the coordinates of a vector uniformly distributed on the multi-slice. The second challenge is that $a_1, \dots, a_k$ are now arbitrary real numbers (corresponding to the support of ξ), and hence, certain arguments tailored for integers no longer apply. Overcoming these challenges requires additional ideas, which we discuss in Section 2.2.

2.1.4 Notation

For a positive integer N, $\mathbb{S}^{N-1}$ denotes the set of unit vectors in $\mathbb{R}^N$, and if $x \in \mathbb{R}^N$ and $r \geq 0$ then $\mathbb{B}_2^N(x, r)$ denotes the radius r Euclidean ball in $\mathbb{R}^N$ centered at x. $\|\cdot\|_2$ denotes the standard Euclidean norm of a vector, and for a matrix $A = (a_{ij})$, $\|A\|$ is its spectral norm (i.e., $\ell^2 \to \ell^2$ operator norm).

We will let $[N]$ denote the interval $\{1, \dots, N\}$. For nonnegative integers $m \leq n$, we let $\{0,1\}_m^n$ be the set of vectors in $\{0,1\}^n$ with sum m.

Since it is essential throughout the paper, we formally record the definition of a discrete random variable and the corresponding random matrix.

Definition 2.1.12. We say that a random variable ξ is a discrete random variable (equivalently, has a discrete distribution) if it is real-valued, its support is finite, and the support contains at least two distinct points. $M_n(\xi)$ denotes the $n \times n$ random matrix, with independent entries that are copies of ξ.

For ξ a discrete random variable with $k = |\operatorname{supp}(\xi)|$ (so that $k \geq 2$), we will denote its support by $\vec{a} = (a_1, \dots, a_k)$, and the (nonzero) probabilities of attaining $a_1, \dots, a_k$ by $\vec{p} = (p_1, \dots, p_k)$. Note that $\|\vec{p}\|_1 = 1$, and $\|\vec{p}\|_2^2 \leq \|\vec{p}\|_\infty$ with equality if and only if ξ is uniform on its support. We will use $H(\xi)$ to denote the natural-logarithmic entropy of ξ, i.e., $H(\xi) = H(\vec{p}) = \sum_{i=1}^k -p_i \log(p_i)$. We will (somewhat abusively) use p_0 to denote $\mathbb{P}[\xi = 0]$.

For a random variable ξ and a real number $r \geq 0$, we let $\mathcal{L}(\xi, r) := \sup_{z \in \mathbb{R}} \mathbb{P}[|\xi - z| \leq r]$. We will use $\ell^1(\mathbb{Z})$ to denote the set of functions $f \colon \mathbb{Z} \to \mathbb{R}$ for which $\sum_{z \in \mathbb{Z}} |f(z)| < \infty$.

We will also make use of asymptotic notation. For functions f, g, $f = O_\alpha(g)$ (or $f \lesssim_\alpha g$) means that $f \leq C_\alpha g$, where C_α is some constant depending on α; $f = \Omega_\alpha(g)$ (or $f \gtrsim_\alpha g$) means that $f \geq c_\alpha g$, where $c_\alpha > 0$ is some constant depending on α, and $f = \Theta_\alpha(g)$ means that both $f = O_\alpha(g)$ and $f = \Omega_\alpha(g)$ hold. For parameters ϵ, δ, we write $\epsilon \ll \delta$ to mean that $\epsilon \leq c(\delta)$ for a sufficient function c.

Finally, we will omit floors and ceilings where they make no essential difference.

2.1.5 Organization

The remainder of this paper is organized as follows. In Section 2.2, we prove our key inversion of randomness estimate for conditional thresholds on the multislice (Theorem 2.2.1). In Section 2.3, we combine this with a much simpler analysis of the structured vectors (compared to the proof of Theorem 2.1.10) in order to complete the proof of Conjecture 2.1.1 for the special case of $\xi = \mathrm{Ber}(p)$ for a fixed $p \in (0, 1/2)$. This section also serves as a 'warm-up' to the subsequent sections. In Section 2.4, we use the results of Section 2.2 to prove the necessary invertibility estimate for unstructured vectors (Theorem 2.4.1). In Section 2.5, we prove a weaker version of Theorem 2.1.8; this is used in our significantly more involved treatment of structured vectors in general case (i.e., the proof of Theorem 2.1.10), which is the content of Section 2.6. In Section 2.7, we quickly combine Theorems 2.1.10 and 2.4.1

to prove Theorems 2.1.2 and 2.1.8. Finally, in Section 2.8, we prove Theorem 2.1.11 for the combinatorial model of random matrices discussed earlier.

2.1.6 Acknowledgements

We thank Mark Rudelson, Konstantin Tikhomirov, and Yufei Zhao for comments on the manuscript. We are grateful to an anonymous referee for their careful reading of our paper and for numerous comments which have improved the presentation. A.S. and M.S. were supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1745302. This work was done when V.J. was participating in a program at the Simons Institute for the Theory of Computing.

2.2 Inversion of randomness on the multislice

In this section, we prove our key inversion of randomness result, Theorem 2.2.1. We will focus on the non-independent "multislice" version as its deduction is strictly harder than the independent version, Theorem 2.2.15 (which we will only use to establish the preliminary estimate Theorem 2.5.5).

The proof of Theorem 2.2.1 follows a direction introduced by Tikhomirov [103]. In this approach, the relevant Lévy concentration function of a random vector is replaced with certain random averages of functions. One then shows that the random vectors with large values of the Lévy concentration function are super-exponentially rare, by first demonstrating a weaker notion of anticoncentration after revealing $(1-\epsilon)n$ coordinates of the random vector, and then iterating a smoothing procedure on linear-sized pieces of the vector which allows one to bootstrap the strength of anticoncentration considered.

Our major challenges lie in (i) the non-independence of the coordinates of a vector on the multislice, as the arguments in [103] rely strongly on the independence structure of the considered model, and (ii) the freedom to allow the support of ξ to consist of arbitrary real numbers, as certain arguments in [103] rely on the integrality of the support. For a more gentle introduction to the techniques in this section, we refer the reader to the expository paper [40, Theorem 3.1] where we record the proof for the Boolean slice, a setting which encounters the first challenge but not the second. We note that the presentation here is entirely self-contained and familiarity with [40] is not assumed.

2.2.1 Statement and preliminaries

Let $N, n \geq 1$ be integers and let $0 < \delta < 1/4$, $K_3 > K_2 > K_1 > 1$ be real parameters. We say that $\mathcal{A} \subseteq \mathbb{Z}^n$ is $(N, n, K_1, K_2, K_3, \delta)$-admissible if

- $\mathcal{A} = A_1 \times \cdots \times A_n$, where each A_i is a subset of $\mathbb{Z}$,
- $|A_1| \cdots |A_n| \leq (K_3 N)^n$,

- $\max_i \max\{|a|\colon a \in A_i\} \le nN$,
- A_i is an integer interval of size at least $2N+1$ for $i > 2\delta n$, and either (P1) and (P2) hold, or (Q1) and (Q2) hold:

(P1) A_{2i} is an integer interval of size at least $2N+1$ contained in $[-K_1N, K_1N]$ for $i \le \delta n$,

(P2) A_{2i-1} is symmetric about 0, is a union of two integer intervals of total size at least $2N$, and satisfies $A_{2i-1} \cap [-K_2N, K_2N] = \emptyset$ for $i \le \delta n$.

(Q1) A_{2i} is an integer interval of size at least $2N+1$ contained in $[K_1N, K_2N]$ for $i \le \delta n$,

(Q2) A_{2i-1} is an integer interval of size at least $2N+1$ contained in $[-K_2N, -K_1N]$ for $i \le \delta n$.

Recall at this point that ξ, which has (nonzero) probabilities $\vec{p} = (p_1, \dots, p_k)$ on atoms $\vec{a} = (a_1, \dots, a_k)$, is fixed. Let $\mathcal{A} = A_1 \times \cdots \times A_n$ be an $(N, n, K_1, K_2, K_3, \delta)$-admissible set, and let $(X_1, \dots, X_n)$ be the random vector uniformly distributed on $\mathcal{A}$. For any $f : \mathbb{R} \to \mathbb{R}$, any $0 \le \ell \le n$, and any $\vec{s} \in \mathbb{Z}^k_{\ge 0}$ with $\|\vec{s}\|_1 = \ell$, define the random function (depending on the randomness of $X_1, \dots, X_n$):

$$f_{\mathcal{A},\vec{s},\ell}(t) := \mathbb{E}_b\left[f\left(t + \sum_{i=1}^{\ell} b_i X_i \right) \middle| \#\{b_i = a_j\} = s_j \ \forall j \in [k] \right],$$

where $\mathbb{E}_b$ denotes the expectation over a random vector $b = (b_1, \dots, b_\ell) \in \mathbb{R}^\ell$ with coordinates independently distributed as ξ. The conditioning encodes that for all $j \in [k]$, there are exactly s_j coordinates (out of ℓ) where b hits the atom a_j.

Theorem 2.2.1. *Fix a discrete distribution ξ. For $0 < \delta < 1/4$, $K_3 > K_2 > K_1 > 1$, $\epsilon \ll \min(\vec{p})$, and a given parameter $M \ge 1$, there are $L_{2.2.1} = L_{2.2.1}(\xi, \epsilon, \delta, K_1, K_2, K_3) > 0$, and $\gamma_{2.2.1} = \gamma_{2.2.1}(\xi, \epsilon, \delta, K_1, K_2, K_3) \in (0, \epsilon)$ independent of M and $n_{2.2.1} = n_{2.2.1}(\xi, \epsilon, \delta, K_1, K_2, K_3, M) \ge 1$ and $\eta_{2.2.1} = \eta_{2.2.1}(\xi, \epsilon, \delta, K_1, K_2, K_3, M)$ such that the following holds.*

Let $n \ge n_{2.2.1}$, $1 \le N \le \exp((H(\vec{p}) - \epsilon)n)$, $f \in L^1(\mathbb{R})$ be a nonnegative function such that $\|f\|_1 = 1$ and $\log_2 f$ is $\eta_{2.2.1}$-Lipschitz, and $\mathcal{A}$ be $(N, n, K_1, K_2, K_3, \delta)$-admissible. Suppose also that $\|\vec{\gamma}\|_\infty \le \gamma_{2.2.1}$. Then, for any $\vec{m} \in \mathbb{Z}^k_{\ge 0}$ such that $\|\vec{m}\|_1 = n$ and $\|\vec{m} - \vec{p}n\|_\infty \le \gamma_{2.2.1} n$,

$$\mathbb{P}[\|f_{\mathcal{A},\vec{m},n}\|_\infty \ge L_{2.2.1}(N\sqrt{n})^{-1}] \le \exp(-Mn).$$

Given this we can deduce the following corollary which is crucial in our application.

Definition 2.2.2. Fix a discrete distribution ξ. Let $\vec{\gamma}$ be a nonnegative vector with $\|\vec{\gamma}\|_\infty \in (0, \min(\vec{p}))$ and let $r \ge 0$. For a vector $(x_1, \dots, x_n) \in \mathbb{R}^n$, we define

$$\mathcal{L}_{\xi,\vec{\gamma}}\left(\sum_{i=1}^n b_i x_i, r \right) := \sup_{z \in \mathbb{R}} \mathbb{P}\left[\left| \sum_{i=1}^n b_i x_i - z \right| \le r \middle| \#\{b_i = a_j\} \in [p_j n - \gamma_j n, p_j n + \gamma_j n] \ \forall j \in [k] \right],$$

where $b_1, \dots, b_n$ are independent ξ random variables. We also define

$$\mathcal{L}_\xi\bigg(\sum_{i=1}^n b_i x_i, r\bigg) = \sup_{z\in\mathbb{R}} \mathbb{P}\bigg[\bigg|\sum_{i=1}^n b_i x_i - z\bigg| \le r\bigg].$$

Corollary 2.2.3. *Fix a discrete distribution ξ. For $0 < \delta < 1/4$, $K_3 > K_2 > K_1 > 1$, $\epsilon \ll \min(\vec{p})$, and a given parameter $M \ge 1$, there are $L_{2.2.3} = L_{2.2.3}(\xi, \epsilon, \delta, K_1, K_2, K_3) > 0$ and $\gamma_{2.2.3} = \gamma_{2.2.3}(\xi, \epsilon, \delta, K_1, K_2, K_3) \in (0, \epsilon)$ independent of M and $n_{2.2.3} = n_{2.2.3}(\xi, \epsilon, \delta, K_1, K_2, K_3, M) \ge 1$ such that the following holds.*

Let $n \ge n_{2.2.3}$, $1 \le N \le \exp((H(\vec{p}) - \epsilon)n)$ and $\mathcal{A}$ be $(N, n, K_1, K_2, K_3, \delta)$-admissible. Suppose also that $\|\vec{\gamma}\|_\infty \le \gamma_{2.2.3}$. Then

$$\bigg|\bigg\{x \in \mathcal{A} \colon \mathcal{L}_{\xi,\vec{\gamma}}\bigg(\sum_{i=1}^n b_i x_i, \sqrt{n}\bigg) \ge L_{2.2.3} N^{-1}\bigg\}\bigg| \le e^{-Mn}|\mathcal{A}|.$$

Proof sketch. This is essentially the same as the deduction in [103, Corollary 4.3]. We apply Theorem 2.2.1 to $f(t) := 2^{-|t|/\sqrt{n}}/\iota$, where $t \in \mathbb{R}$ and ι is an appropriate normalization, separately for all $\vec{m} \in \mathbb{Z}^k_{\ge 0}$ such that $\|\vec{m} - \vec{p}n\|_\infty \le \gamma_{2.2.1} n$, and then conclude using a union bound. □

The proof of Theorem 2.2.1 makes use of an anticoncentration estimate on the multislice, which we record below (Lemmas 2.2.5 and 2.2.6), and is ultimately a consequence of the following standard anticoncentration inequality due to Kolmogorov–Lévy–Rogozin.

Lemma 2.2.4 ([83])**.** *Let $\xi_1, \dots, \xi_n$ be independent random variables. Then, for any real numbers $r_1, \dots, r_n > 0$ and any real $r \ge \max_{i\in[n]} r_i$, we have*

$$\mathcal{L}\bigg(\sum_{i=1}^n \xi_i, r\bigg) \le \frac{C_{2.2.4} r}{\sqrt{\sum_{i=1}^n (1 - \mathcal{L}(\xi_i, r_i)) r_i^2}},$$

where $C_{2.2.4} > 0$ is an absolute constant.

Lemma 2.2.5. *Fix $(a_1, \dots, a_k) \in \mathbb{R}^k$ with distinct coordinates. Let $\sigma, \lambda \in (0, 1/3)$ and $r > 0$. Let $Z = \{z_1, \dots, z_n\}$ be a set of real numbers for which there exist disjoint subsets $Z_1, Z_2 \subseteq Z$ such that $|Z_1|, |Z_2| \ge \sigma n$ and such that $|z_i - z_j| \ge r$ for all $z_i \in Z_1, z_j \in Z_2$. Then, there exists $C_{2.2.5} = C_{2.2.5}(\lambda, \sigma, k)$ such that for any $\vec{s} \in \mathbb{Z}^k_{\ge 0}$ with $\|\vec{s}\|_1 = n$ and with $s_\ell \in [\lambda n, (1-\lambda)n]$ for some $\ell \in [k]$, we have*

$$\mathcal{L}\bigg(\sum_{i=1}^n z_i b_i, r \cdot \min_{i<j} |a_i - a_j|\bigg) \le \frac{C_{2.2.5}}{\sqrt{n}},$$

where $(b_1, \dots, b_n)$ is a random vector uniformly chosen from among those with s_j coordinates equal to a_j for all $j \in [k]$.

Proof. By reindexing the coordinates of Z, we may assume that for $i \in [\sigma n]$, $z_{2i-1} \in Z_1$ and $z_{2i} \in Z_2$. In particular, for $i \in [\sigma n]$, we have $|z_{2i} - z_{2i-1}| \geq r$. Furthermore, by the pigeonhole principle, there exists some $\ell' \neq \ell$ such that $s_{\ell'} \geq \lambda n/k$. We will now use the randomness within the atoms a_ℓ and $a_{\ell'}$ in order to derive the anticoncentration result. Note that $\sum_{i=1}^n b_i z_i$ has the same distribution as

$$\sum_{i>2\sigma n} z_i b_i + \sum_{j \leq \sigma n} \Big(z_{2j-1} b_{2j-1} + z_{2j} b_{2j} + b'_j (b_{2j} - b_{2j-1})(z_{2j-1} - z_{2j}) \Big),$$

where $b'_1, \dots, b'_{\sigma n}$ are i.i.d. Ber(1/2) random variables. Next, note that by a standard large deviation estimate, we have

$$\mathbb{P}[|\{j \in [\sigma n] \colon \{b_{2j-1}, b_{2j}\} = \{a_\ell, a_{\ell'}\}| \leq c(\sigma, \lambda, k) n] \leq \exp(-c(\sigma, \lambda, k) n), \tag{2.2.1}$$

where $c(\sigma, \lambda, k) > 0$ is a constant depending only on σ, λ, and k. On the other hand, on the complement of this event, we may conclude by applying Lemma 2.2.4 to (2.2.1), using only the randomness in $b'_1, \dots, b'_{\sigma n}$. □

Lemma 2.2.6. *Fix a discrete distribution ξ, $\lambda \in (0, 1/3)$, $\delta_0 \in (0, 1/4)$. Let $\mathcal{A}$ be $(N, n, K_1, K_2, K_3, \delta)$-admissible for some integer parameters N, n and real parameters $\delta \in [\delta_0, 1/4)$, $K_3 > K_2 > K_1 > 1$. Suppose that $n > n_{2.2.6}(\lambda, \delta_0, K_1, K_2, K_3)$, $\ell \geq \delta_0 n$, and $\vec{s} \in \mathbb{Z}^k_{\geq 0}$ with $\|s\|_1 = \ell$ and $s_{j_0} \in [\lambda \ell, (1-\lambda)\ell]$ for some $j_0 \in [k]$. Then, for any interval J,*

$$\int_{t \in J} f_{\mathcal{A}, \vec{s}, \ell}(t) dt \leq \frac{C_{2.2.6}(\lambda, \xi, \delta_0, K_1, K_2) \max(|J|, N)}{N\sqrt{n}}.$$

Proof. The proof is nearly identical to that in [103, Lemma 4.4] though we provide details as we are in the slightly different setting of $L^1(\mathbb{R})$. Fix $X_1, \dots, X_\ell$. Then

$$\begin{aligned}
\int_{t \in J} f_{\mathcal{A}, \vec{s}, \ell}(t) dt &= \int_{t \in J} \mathbb{E}_b \Big[f\Big(t + \sum_{i=1}^\ell b_i X_i \Big) \Big| \#\{b_i = a_j\} = s_j \ \forall j \in [k] \Big] dt \\
&= \mathbb{E}_b \Big[\int_{t \in J} f\Big(t + \sum_{i=1}^\ell b_i X_i \Big) dt \Big| \#\{b_i = a_j\} = s_j \ \forall j \in [k] \Big] \\
&= \mathbb{E}_b \Big[\int_{t \in \mathbb{R}} f(t) \mathbb{1}_{J + \sum_{i=1}^\ell b_i X_i}(t) dt \Big| \#\{b_i = a_j\} = s_j \ \forall j \in [k] \Big] \\
&= \int_{t \in \mathbb{R}} f(t) \mathbb{E}_b \Big[\mathbb{1}_{J + \sum_{i=1}^\ell b_i X_i}(t) \Big| \#\{b_i = a_j\} = s_j \ \forall j \in [k] \Big] dt \\
&= \int_{t \in \mathbb{R}} f(t) \mathbb{P}_b \Big[\sum_{i=1}^\ell b_i X_i \in J - t \Big| \#\{b_i = a_j\} = s_j \ \forall j \in [k] \Big] dt \\
&\leq \mathcal{L}\Big(\sum_{i=1}^\ell b_i X_i, |J| \Big) \int_{t \in \mathbb{R}} |f(t)| dt \leq \mathcal{L}\Big(\sum_{i=1}^\ell b_i X_i, |J| \Big),
\end{aligned}$$

where $(b_1, \dots, b_\ell)$ is uniformly chosen from vectors which have s_j coordinates equal to a_j for all $j \in [k]$, and we have used that $\|f\|_1 = 1$. The required estimate now follows immediately from Lemma 2.2.5 applied with $r = (K_2 - K_1)N$, which is possible due to the admissibility of $\mathcal{A}$. □

2.2.2 Preprocessing on real-valued multislices

As in [103], we first prove a version of Theorem 2.2.1 in which L is allowed to depend on M.

Proposition 2.2.7. *Fix a discrete distribution ξ. For $0 < \delta < 1/4$, $K_3 > K_2 > K_1 > 1$, $\epsilon \ll \min(\vec{p})$, and a given parameter $M \geq 1$, there is $\gamma_{2.2.7} = \gamma_{2.2.7}(\xi, \epsilon, \delta, K_1, K_2, K_3) \in (0, \epsilon)$ independent of M and there are $L_{2.2.7} = L_{2.2.7}(\xi, \epsilon, \delta, K_1, K_2, K_3, M) > 0$ and $n_{2.2.7} = n_{2.2.7}(\xi, \epsilon, \delta, K_1, K_2, K_3, M) \geq 1$ such that the following holds.*

Let $n \geq n_{2.2.7}$, $1 \leq N \leq \exp((H(\vec{p}) - \epsilon)n)$, and $\mathcal{A}$ be $(N, n, K_1, K_2, K_3, \delta)$-admissible. Let f be a nonnegative function in $L^1(\mathbb{R})$ with $\|f\|_1 = 1$ such that $\log_2 f$ is 1-Lipschitz. Then, for all $\ell \in [(1 - \gamma_{2.2.7})n, n]$ and $\vec{s} \in \mathbb{Z}^k_{\geq 0}$ with $\|\vec{s}\|_1 = \ell$ and $\|\vec{s} - \vec{p}\ell\|_\infty \leq \gamma_{2.2.7}\ell$, we have

$$\mathbb{P}\left[\|f_{\mathcal{A},\vec{s},\ell}\|_\infty \geq L_{2.2.7}(N\sqrt{n})^{-1}\right] \leq \exp(-Mn).$$

Proposition 2.2.7 should be seen as an analogue of [103, Proposition 4.5] for the multislice. As mentioned earlier, compared to [103], our situation is much more delicate since we are working with a vector with non-independent coordinates and need to extract a term corresponding to the entropy of the multislice. (Such complications are already encountered when working with a Boolean slice.) Working on real multislices presents additional difficulties (along with significant notational complications), owing to the fact that we are working on $L^1(\mathbb{R})$; this extension is handled by using the log-Lipschitz condition on f. We note that the corresponding statements in [103] do not need to use any log-Lipschitz assumption at this stage of the argument since they are proved for $\ell^1(\mathbb{Z})$. We also note that, while the constant 1 in 1-log-Lipschitz is arbitrary, some condition of this nature is necessary to rule out f being very close to a Dirac mass ([103]).

We first note the trivial recursive relation

$$f_{\mathcal{A},\vec{s},\ell}(t) = \sum_{i=1}^{k} \frac{s_i}{\ell} f_{\mathcal{A},\vec{s}-e_i,\ell-1}(t + a_i X_\ell)$$

for all $1 \leq \ell \leq n$ and $\vec{s} \in \mathbb{Z}^k_{\geq 0}$ with $\|\vec{s}\|_1 = \ell$. If any coordinate of $\vec{s}$ is zero, note that the corresponding term (which would be undefined) has a coefficient of 0, and drops out. Note also that, by definition, $f_{\mathcal{A},\vec{0},0} = f$.

Definition 2.2.8 (Step record and averaging sequence)**.** Fix $f, \mathcal{A}, \vec{s}, \ell$, a point $t \in \mathbb{R}$, and a choice of $X = (X_1, \dots, X_n)$. For such a choice, we define the *averaging sequence* $(t_i)_{i=0}^{\ell}$ and *step record* $(w_i)_{i=1}^{\ell}$ as follows:

- $t_\ell := t$,
- Since
$$h_\ell := f_{\mathcal{A},\vec{s},\ell}(t_\ell) = \sum_{j=1}^{k} \frac{s_j}{\ell} f_{\mathcal{A},\vec{s}-e_j,\ell-1}(t_\ell + a_j X_\ell),$$
at least one of the k terms $f_{\mathcal{A},\vec{s}-e_j,\ell-1}(t_\ell + a_j X_\ell)$ has a positive coefficient and is at least h_ℓ. If it is index j, set $w_\ell = j$.
- Set $t_{\ell-1} := t_\ell + a_{w_\ell} X_\ell$, $h_{\ell-1} := f_{\mathcal{A},\vec{s}-e_{w_\ell},\ell-1}(t_{\ell-1})$, and repeat with $t_{\ell-1}, \vec{s} - e_{w_\ell}, \ell - 1$.

It will be convenient to write

- $W_i(j) := \#\{u \in [i] \colon w_u = j\}$ and $\overline{W}_i(j) := W_i(j)/i$ for all $i \in [\ell]$ and $j \in [k]$. We will view $W_i = (W_i(1), \ldots, W_i(k))$ as a vector in $\mathbb{Z}^k$.

We note some straightforward consequences of these definitions.

- $W_\ell = \vec{s}$.
- $W_{i-1} = W_i - e_{w_i}$ for $1 \le i \le \ell$, where we assume $W_0 = \vec{0}$.
- $\|W_i\|_1 = i$.
- $t_{i-1} = t_i + a_{w_i} X_i$ for all $i \in [\ell]$.
- $f_{\mathcal{A},W_i,i}(t_i) = \sum_{j=1}^{k} \overline{W}_i(j) f_{\mathcal{A},W_i-e_j,i-1}(t_i + a_j X_i)$.
- $h_i = f_{\mathcal{A},W_i,i}(t_i)$.
- $f(t_0) = h_0 \ge h_1 \ge \cdots \ge h_\ell = f_{\mathcal{A},\vec{s},\ell}(t)$.

Definition 2.2.9 (Drops and robust steps)**.** With notation as above, given $i \in [\ell]$:

- For $\lambda \in (0,1)$, we say that step i is λ-*robust* if
$$\overline{W}_i(w_i) \in (\lambda, 1-\lambda)$$
- For $R > 0$, we say that there is an R-*drop* at step i if
$$f_{\mathcal{A},W_i-e_j,i-1}(t_{i-1} + zX_i) \le \frac{R}{N\sqrt{n}}$$
for all $j \in [k]$ such that $W_i(j) > 0$ and for all $z \in \mathrm{supp}(\xi - \xi') \setminus \{0\}$.

Next we show that if $\|f_{\mathcal{A},\vec{s},\ell}\|_\infty$ is large in an appropriate sense, then there is a step record and averaging sequence with linearly many robust steps which do not participate in an R-drop.

Lemma 2.2.10. *Let $\xi, \mathcal{A}, f, N, \epsilon$ be as in Proposition 2.2.7, and let $L \geq 1$. Then, there exist $\lambda_{2.2.10} = \lambda_{2.2.10}(\xi, \epsilon) \in (0, 1/3)$, $\gamma_{2.2.10} = \gamma_{2.2.10}(\xi, \epsilon) \in (0, 1)$, and $n_{2.2.10} = n_{2.2.10}(\xi, \epsilon)$ for which the following holds.*

Let $n \geq n_{2.2.10}$, $R = \gamma_{2.2.10} L$, let $\ell \in [(1 - \gamma_{2.2.10})n, n]$ and $\vec{s} \in \mathbb{Z}^k_{\geq 0}$ satisfy $\|\vec{s}\|_1 = \ell$ and $\|\vec{s} - \vec{p}\ell\|_\infty \leq \gamma_{2.2.10}\ell$. Then, for $(X_1, \dots, X_n) \in \mathcal{A}$,

$$\|f_{\mathcal{A},\vec{s},\ell}\|_\infty \geq L(N\sqrt{n})^{-1}$$

implies that there exists some $t \in \mathbb{R}$ with $f_{\mathcal{A},\vec{s},\ell}(t) \geq L(N\sqrt{n})^{-1}$ so that its averaging sequence $(t_i)_{i=0}^{\ell}$ and step record $(w_i)_{i=1}^{\ell}$ satisfy

$$\#\{i \in [\ell] \colon \text{step } i \text{ is } \lambda_{2.2.10}\text{-robust and is not an } R\text{-drop}\} \geq \gamma_{2.2.10} n.$$

Proof. Consider $(X_1, \dots, X_n) \in \mathcal{A}$ satisfying $\|f_{\mathcal{A},\vec{s},\ell}\|_\infty \geq L(N\sqrt{n})^{-1}$. Then, there is some $t \in \mathbb{R}$ such that $f_{\mathcal{A},\vec{s},\ell}(t) \geq L(N\sqrt{n})^{-1}$. We will show that the conclusion of the lemma is satisfied for this t, for suitable choice of $\gamma_{2.2.10}, \lambda_{2.2.10}$. Below, we will make extensive use of the notation and relations in Definitions 2.2.8 and 2.2.9. Let $(t_i)_{i=0}^{\ell}$ and $(w_i)_{i=1}^{\ell}$ denote, respectively, the averaging sequence and step record of t. Note that

$$L(N\sqrt{n})^{-1} \leq f_{\mathcal{A},\vec{s},\ell}(t) = h_0 \prod_{i=1}^{\ell} \frac{h_i}{h_{i-1}} \leq h_{\ell-1} \leq \cdots \leq h_0.$$

We begin by controlling the ratios h_i/h_{i-1} at steps i which are R-drops. Hence, suppose that step i is an R-drop. If $w_i = u$, then $W_i = W_{i-1} + e_u$ and $t_i = t_{i-1} - a_u X_i$. Hence

$$\begin{aligned}
\frac{h_i}{h_{i-1}} &= \sum_{j=1}^{k} \overline{W}_i(j) \frac{f_{\mathcal{A},W_i - e_j, i-1}(t_i + a_j X_i)}{f_{\mathcal{A},W_{i-1},i-1}(t_{i-1})} \\
&= \overline{W}_i(u) + \sum_{j \neq u} \overline{W}_i(j) \frac{f_{\mathcal{A},W_i - e_j, i-1}(t_{i-1} + (a_j - a_u)X_i)}{h_{i-1}} \\
&\leq \overline{W}_i(u) + \sum_{j \neq u} \overline{W}_i(j) \frac{R(N\sqrt{n})^{-1}}{L(N\sqrt{n})^{-1}} \\
&= \overline{W}_i(u) + (1 - \overline{W}_i(u))\gamma_{2.2.10}.
\end{aligned}$$

The inequality uses the definition of R-drops (this is applicable since $a_j - a_u \in \operatorname{supp}(\xi - \xi') \setminus \{0\}$) along with $h_i \geq L(N\sqrt{n})^{-1}$. Note that if the condition $W_i(j) > 0$ in the definition of R-drops is not satisfied, then the jth term already drops out in the first line. Thus, we see that if step i is an R-drop, then

$$\frac{h_i}{h_{i-1}} \leq \overline{W}_i(w_i) + (1 - \overline{W}_i(w_i))\gamma_{2.2.10}. \tag{2.2.2}$$

Note that if step i is $\lambda_{2.2.10}$-robust, the right-hand side is at least $\lambda_{2.2.10}$. Therefore, for any step i which is $\lambda_{2.2.10}$-robust, we have

$$\lambda_{2.2.10} \le \overline{W}_i(w_i) + (1 - \overline{W}_i(w_i))\gamma_{2.2.10} \le \overline{W}_i(w_i)\left(1 + \frac{\gamma_{2.2.10}}{\lambda_{2.2.10}}\right), \tag{2.2.3}$$

where the final inequality uses $(1 - \overline{W}_i(w_i))/\overline{W}_i(w_i) \le 1/\lambda_{2.2.10}$ at any $\lambda_{2.2.10}$-robust step i.

Now, let $I \subseteq [\ell]$ denote the steps i which are $\lambda_{2.2.10}$-robust, and let $J \subseteq I$ denote the steps i which are *not* R-drops (so that $I \setminus J$ is the set of $\lambda_{2.2.10}$-robust R-drops). Our goal is to provide a lower bound on $|J|$.

Since $h_0 \le \|f\|_\infty \le \|f\|_1 = 1$ (this uses the 1-Lipschitz condition on $\log_2 f$), we have

$$\begin{aligned}
L(N\sqrt{n})^{-1} &\le \prod_{i \in I \setminus J} \frac{h_i}{h_{i-1}} \le \prod_{i \in I \setminus J} (\overline{W}_i(w_i) + (1 - \overline{W}_i(w_i))\gamma_{2.2.10}) \\
&= \frac{\prod_{i \in I}(\overline{W}_i(w_i) + (1 - \overline{W}_i(w_i))\gamma_{2.2.10})}{\prod_{i \in J}(\overline{W}_i(w_i) + (1 - \overline{W}_i(w_i))\gamma_{2.2.10})} \\
&\le \frac{(1 + \gamma_{2.2.10}/\lambda_{2.2.10})^{|I|} \prod_{i \in I} \overline{W}_i(w_i)}{\lambda_{2.2.10}^{|J|}} \\
&= (1 + \gamma_{2.2.10}/\lambda_{2.2.10})^{|I|} \lambda_{2.2.10}^{-|J|} \prod_{i \in [\ell]} \overline{W}_i(w_i) \prod_{i \in [\ell] \setminus I} \overline{W}_i(w_i)^{-1} \\
&= (1 + \gamma_{2.2.10}/\lambda_{2.2.10})^{|I|} \cdot \lambda_{2.2.10}^{-|J|} \cdot \binom{\ell}{\vec{s}}^{-1} \cdot \prod_{i \in [\ell] \setminus I} \overline{W}_i(w_i)^{-1};
\end{aligned} \tag{2.2.4}$$

here, the first line uses $h_i/h_{i-1} \le 1$ and (2.2.2), the third line uses (2.2.3), and the last line uses the identity

$$\prod_{i \in [\ell]} \overline{W}_i(w_i) = \binom{\ell}{\vec{s}}^{-1} := \binom{\ell}{s_1, \ldots, s_k}^{-1}.$$

This follows since both sides are equal to the probability that a uniformly random sample from $[k]^\ell$, conditioned on having s_j copies of j for each $j \in [k]$, returns $(w_1, \ldots, w_\ell)$.

Note that the first and the third terms in the final product in (2.2.4) are easy to suitably control (by taking $\gamma_{2.2.10}$ and $\lambda_{2.2.10}$ to be sufficiently small). As we will see next, these parameters also allow us to make the last term at most $\exp(c\epsilon n)$ for any constant $c > 0$.

Let $K \subseteq [\ell] \setminus I$ denote those indices i such that $\overline{W}_i(w_i) \ge 1 - \lambda_{2.2.10}$. Then,

$$\prod_{i \in K} \overline{W}_i(w_i)^{-1} \le (1 - \lambda_{2.2.10})^{-|K|}. \tag{2.2.5}$$

It remains to bound

$$\prod_{i \in [\ell] \setminus (I \cup K)} \overline{W}_i(w_i)^{-1}.$$

Note that for every $i \in [\ell] \setminus (I \cup K)$, we have $\overline{W}_i(w_i) \le \lambda_{2.2.10}$. Let J_j for $j \in [k]$ be the set of $i \in [\ell] \setminus (I \cup K)$ with $w_i = j$.

The following is the key point: let $i_1, \dots, i_{u_j} \in J_j$ be all elements of J_j in order. Then, for all $y \in [u_j]$, we have

$$y \le W_{i_y}(j) \le \lambda_{2.2.10}\ell.$$

Hence,

$$u_j \le \lambda_{2.2.10}\ell \quad \text{and} \quad \overline{W}_{i_y}(w_{i_y})^{-1} \le i_y/y \le \ell/y.$$

We derive

$$\prod_{i \in [\ell]\setminus(I\cup K)} \overline{W}_i(w_i)^{-1} = \prod_{j=1}^{k} \prod_{i \in J_j} \overline{W}_i(w_i)^{-1} \le \left(\prod_{u=1}^{\lceil \lambda_{2.2.10}\ell \rceil} \frac{\ell}{u} \right)^k \le \left(\frac{e}{\lambda_{2.2.10}} \right)^{2k\lambda_{2.2.10}\ell}. \tag{2.2.6}$$

Substituting (2.2.5) and (2.2.6) in (2.2.4), we have

$$Ln^{-1/2} \exp((\epsilon - H(\vec{p}))n) \le \lambda_{2.2.10}^{-|J|} \cdot \left(1 + \frac{\gamma_{2.2.10}}{\lambda_{2.2.10}}\right)^{\ell} \cdot \binom{\ell}{\vec{s}}^{-1} \cdot (1 - \lambda_{2.2.10})^{-\ell} \cdot \left(\frac{e}{\lambda_{2.2.10}} \right)^{2k\lambda_{2.2.10}\ell}. \tag{2.2.7}$$

We will first choose $\lambda_{2.2.10}$, and then choose some $\gamma_{2.2.10} < \lambda_{2.2.10}^2$. Note that, by enforcing the constraint $\gamma_{2.2.10} < \lambda_{2.2.10}^2$, we can choose $\lambda_{2.2.10}$ sufficiently small depending on ϵ and ξ so that the second term, the fourth term, and the fifth term in the product in (2.2.7) are each bounded above by $\exp(\epsilon n/10)$ and so that (using Stirling's approximation) the third term is bounded above by $\exp(\epsilon n/10 - H(\vec{p})n)$. Hence, we can choose $\lambda_{2.2.10}$ depending on ϵ and ξ such that

$$n^{-1/2} \exp(\epsilon n/2) \le \lambda_{2.2.10}^{-|J|}.$$

Now, for all n sufficiently large depending on ϵ, we can find $\gamma_{2.2.10}$ sufficiently small depending on $\epsilon, \lambda_{2.2.10}$ such that $|J| \ge \gamma_{2.2.10} n$. This completes the proof. □

We are now ready to prove Proposition 2.2.7.

Proof of Proposition 2.2.7. We use Lemma 2.2.10 along with a union bound. For controlling individual events in the union, we will use the following. Consider a step record $(w_i)_{i=1}^{\ell}$. We write $A_i = A_{i,0} \cup A_{i,1}$, where each of these is an integer interval of size at least N (this is possible by the admissibility of $\mathcal{A}$). Now suppose step i is $\lambda_{2.2.10}$-robust with respect to $(w_i)_{i=1}^{\ell}$. If $i > \delta_0 n$, then for any $t \in \mathbb{R}$, $j \in [k]$ and $z \in \operatorname{supp}(\xi - \xi') \setminus \{0\}$, by Lemma 2.2.6, we have

$$\mathbb{E}[f_{\mathcal{A},W_i-e_j,i-1}(t + zX_i) | X_1, \dots, X_{i-1}] = \frac{1}{|A_i|} \sum_{\tau \in t + zA_i} f_{\mathcal{A},W_i-e_j,i-1}(\tau)$$

$$\begin{aligned}
&\leq \max_{y\in\{0,1\}} \frac{1}{|A_{i,y}|} \sum_{\tau\in t+zA_{i,y}} f_{\mathcal{A},W_i-e_j,i-1}(\tau)\\
&\leq \max_{y\in\{0,1\}} \frac{2^{|z|}}{|A_{i,y}|}\left|\int_{t+z\min A_{i,y}}^{t+z\max A_{i,y}} f_{\mathcal{A},W_i-e_j,i-1}(\tau)d\tau\right|\\
&\leq \frac{2^{|z|+1}C_{2.2.6}(\lambda_{2.2.10}/2,\xi,\delta_0,K_1,K_2)\max(|z||A_{i,y}|,N)}{|A_{i,y}|N\sqrt{n}}\\
&\leq \frac{4^{|z|+1}C_{2.2.6}(\lambda_{2.2.10}/2,\xi,\delta_0,K_1,K_2)}{N\sqrt{n}}.
\end{aligned}$$

Here, we have used that $i-1\geq\delta_0 n$, that W_i-e_j has at least one coordinate in $[\lambda_{2.2.10}(i-1)/2,(1-\lambda_{2.2.10}/2)(i-1)]$ (since W_i satisfies a similar property with coordinate w_i), and that each $A_{i,y}$ is length at least N. We also used that $\log_2 f$ is 1-Lipschitz in the second inequality (where the absolute values are put just in case $z<0$ and the limits of integration are in the wrong direction).

Now, consider $t\in\mathbb{R}$ with averaging sequence $(t_i)_{i=0}^{\ell}$ and step record $(w_i)_{i=1}^{\ell}$. Note that, given the 'starting point' t_0 of the averaging sequence, the points $t_1,\dots,t_{i-1}$ are determined by $X_1,\dots,X_{i-1}$. In particular, the event that step i is not an R-drop is determined by $t_0,X_1,\dots,X_i,w_1,\dots,w_i$. Therefore, by Markov's inequality, we see that for any $\lambda_{2.2.10}$-robust step i with $i>\delta_0 n$, given the step record $(w_i)_{i=1}^{\ell}$ and the starting point t_0 of the averaging sequence $(t_i)_{i=0}^{\ell}$,

$$\mathbb{P}[\text{step } i \text{ is not an } R\text{-drop}|X_1,\dots,X_{i-1}]\leq\frac{k^3 4^{2\|\vec{a}\|_\infty+1}C_{2.2.6}(\lambda_{2.2.10}/2,\xi,\delta_0,K_1,K_2)}{R}. \quad (2.2.8)$$

This follows from a union bound over the at most k^3 possible conditions for an R-drop and the fact that all $z\in\operatorname{supp}(\xi-\xi')\setminus\{0\}$ have magnitude at most $2\|\vec{a}\|_\infty$.

From here on, the proof closely follows the proof of [103, Proposition 4.5]. Fix parameters as given in the proposition statement. Let $\lambda_{2.2.10}=\lambda_{2.2.10}(\xi,\epsilon)$. We choose $\gamma_{2.2.7}=\gamma_{2.2.10}(\xi,\epsilon)$. Further, we set $R'=\gamma_{2.2.10}L/2$, where $L\geq 1$ will be chosen later.

Let $\mathcal{E}_L$ denote the event that $\|f_{\mathcal{A},\vec{s},\ell}\|_\infty\geq L(N\sqrt{n})^{-1}$. For $(X_1,\dots,X_n)\in\mathcal{E}_L$, by Lemma 2.2.10, there exists $t\in\mathbb{R}$ with $f_{\mathcal{A},\vec{s},\ell}(t)\geq L(N\sqrt{n})^{-1}$ with averaging sequence $(t_i)_{i=0}^{\ell}$ and step record $(w_i)_{i=1}^{\ell}$ such that

$$\#\{i\in[\ell]\colon \text{step } i \text{ is } \lambda_{2.2.10}\text{-robust and is not a } 2R'\text{-drop in } (t_i)_{i=0}^{\ell}\}\geq\gamma_{2.2.10}n.$$

We then shift t_0 to the nearest integer $\widetilde{t_0}$. We also shift $(t_i)_{i=1}^{\ell}$ by the same amount to obtain points $(\widetilde{t_i})_{i=1}^{\ell}$ (note that these points are not necessarily integers). We call the sequence $(\widetilde{t_i})_{i=0}^{\ell}$, which technically may no longer be an averaging sequence, a *witnessing sequence*. We see that every index which is not a $2R'$-drop in $(t_i)_{i=0}^{\ell}$ will not be an R'-drop in $(\widetilde{t_i})_{i=0}^{\ell}$ as $\log_2 f$ is 1-Lipschitz.

Taking a union bound over the choice of the step record is not costly, and note that given $(X_1,\dots,X_n)$ and the step record, the witnessing sequence is completely determined by its

starting point $\widetilde{t_0}$. Furthermore, the definition of the witnessing sequence and the definition of $f_{\mathcal{A},\vec{s},\ell}$ easily show that

$$\widetilde{t_0} \in \{\tau \in \mathbb{Z} \colon f(\tau) > (2N\sqrt{n})^{-1}\} =: \mathcal{D}.$$

Note that $\mathcal{D}$ is a deterministic set depending only on f. Further, since $\|f\|_1 = 1$ and $\log_2 f$ is 1-Lipschitz, we see that

$$|\mathcal{D}| \leq 4N\sqrt{n}.$$

To summarize, we have shown that if $(X_1, \dots, X_n) \in \mathcal{E}_L$, then there exists a witnessing sequence $(\widetilde{t_i})_{i=0}^{\ell}$ with step record $(w_i)_{i=1}^{\ell}$ such that $\widetilde{t_0} \in \mathcal{D}$, and such that

$$\#\{i \in [\ell] : \text{step } i \text{ is } \lambda_{2.2.10}\text{-robust and is not an } R'\text{-drop in } (\widetilde{t_i})_{i=0}^{\ell}\} \geq \gamma_{2.2.10} n.$$

Therefore, by the union bound and since $N \leq k^n$ (as $H(\vec{p}) \leq \log k$), it follows that

$$\mathbb{P}[\mathcal{E}_L] \leq (2k^2)^n \sup_{\substack{I \subseteq [\ell], |I| = \lceil \gamma_{2.2.10} n \rceil \\ \widetilde{t_0} \in \mathcal{D}, (w_i)_{i=1}^{\ell} \in [k]^{\ell}}} \mathbb{P}[\text{The witnessing sequence starts at } \widetilde{t_0}, \text{ has step record } (w_i)_{i=1}^{\ell}, \text{ and every } i \in I \text{ is } \lambda_{2.2.10}\text{-robust and is not an } R'\text{-drop}],$$

where the supremum is only over those $(w_i)_{i=1}^{\ell}$ which have s_j coordinates equal to j for all $j \in [k]$.

From (2.2.8), taking $\delta_0 = \gamma_{2.2.10}/2$, it follows that the probability appearing on the right hand side above is bounded by

$$\left(\frac{2k^3 4^{2\|\vec{a}\|_\infty + 1} C_{2.2.6}(\lambda_{2.2.10}/2, \xi, \gamma_{2.2.10}/2, K_1, K_2)}{\gamma_{2.2.10} L} \right)^{\gamma_{2.2.10} n/2},$$

since there are at least $\gamma_{2.2.10} n/2$ values of $i \in I$ with $i > \delta_0 n$ and since $R' = \gamma_{2.2.10} L/2$ by definition. Therefore, taking L and n sufficiently large depending on M and the parameters appearing above gives the desired conclusion. □

2.2.3 Refining the initial estimate

We now need to remove the dependence of L on M. This is accomplished by the main result of this subsection, Proposition 2.2.11, which is a multislice and $L^1(\mathbb{R})$ analogue of [103, Proposition 4.10]. Even though we are working in the much more complicated setting of real multislices, remarkably, our proof of Proposition 2.2.11 is able to use [103, Proposition 4.10] as a black box: roughly, we first use a re-randomization procedure to reduce smoothing on the multislice for $L^1(\mathbb{R})$ to smoothing on the hypercube, also for $L^1(\mathbb{R})$. At this juncture, the necessary smoothing estimate on the hypercube for $L^1(\mathbb{R})$ can in fact be lifted from the smoothing estimate for the hypercube for $\ell^1(\mathbb{Z})$, proved in [103]. In particular, we reduce the smoothing estimate for general log-Lipschitz functions in $L^1(\mathbb{R})$ to that of a simpler class of "step" functions, which in turn is equivalent to $\ell^1(\mathbb{Z})$.

Proposition 2.2.11. *Fix a discrete distribution ξ. There exists $h = h(\xi) \geq 1$ so that the following holds. For any $\epsilon \in (0,1)$, $\widetilde{R} \geq 1$, $L_0 \geq h\widetilde{R}$, and $M \geq 1$, there is $\gamma_{2.2.11} = \gamma_{2.2.11}(\xi)$ and there are $n_{2.2.11} = n_{2.2.11}(\xi, \epsilon, L_0, \widetilde{R}, M) > 0$ and $\eta_{2.2.11} = \eta_{2.2.11}(\xi, \epsilon, L_0, \widetilde{R}, M) \in (0,1)$ with the following property. Let $L_0 \geq L \geq h\widetilde{R}$, let $n \geq n_{2.2.11}$, $N \in \mathbb{N}$, and let $g \in L^1(\mathbb{R})$ be a nonnegative function satisfying*

(A) $\|g\|_1 = 1$,

(B) $\log_2 g$ *is* $\eta_{2.2.11}$*-Lipschitz,*

(C) $\int_{t \in I} g(t) \leq \widetilde{R}/\sqrt{n}$ *for any interval* I *of size* N*, and*

(D) $\|g\|_\infty \leq L/(N\sqrt{n})$.

For each $i \leq 2\lfloor \epsilon n \rfloor$, let Y_i be a random variable uniform on some disjoint union of integer intervals of cardinality at least N each, and assume that $Y_1, \dots, Y_{2\lfloor \epsilon n \rfloor}$ are mutually independent. Define a random function $\widetilde{g} \in L^1(\mathbb{R})$ by

$$\widetilde{g}(t) = \mathbb{E}_b \left[g\left(t + \sum_{i=1}^{2\lfloor \epsilon n \rfloor} b_i Y_i \right) \middle| \#\{b_i = a_j\} = s_j \ \forall j \in [k] \right]$$

where $b = (b_1, \dots, b_{2\lfloor \epsilon n \rfloor})$ is a vector of independent ξ components and $\vec{s} \in \mathbb{Z}^k_{\geq 0}$ satisfies $\|\vec{s}\|_1 = 2\lfloor \epsilon n \rfloor$ and

$$\left\| \frac{\vec{s}}{2\lfloor \epsilon n \rfloor} - \vec{p} \right\|_\infty \leq \gamma_{2.2.11}.$$

Then

$$\mathbb{P}\left[\|\widetilde{g}\|_\infty > \frac{19L/20}{N\sqrt{n}} \right] \leq \exp(-Mn).$$

We now state an analogue of Proposition 2.2.11 for independent scaled Bernoulli random variables, which in fact is strong enough to imply Proposition 2.2.11.

Proposition 2.2.12. *Fix $h \geq 1$, and let $z \in [h^{-1}, h]$. For any $\epsilon \in (0,1)$, $\widetilde{R} \geq 1$, $L_0 \geq 64h^2\widetilde{R}$, and $M \geq 1$, there are $n_{2.2.12} = n_{2.2.12}(h, \epsilon, L_0, \widetilde{R}, M) > 0$ and $\eta_{2.2.12} = \eta_{2.2.12}(h, \epsilon, L_0, \widetilde{R}, M) \in (0,1)$ with the following property. Let $L_0 \geq L \geq 64h^2\widetilde{R}$, let $n \geq n_{2.2.12}$, $N \in \mathbb{N}$, and let $g \in L^1(\mathbb{R})$ be a nonnegative function satisfying*

(A) $\|g\|_1 = 1$,

(B) $\log_2 g$ *is* $\eta_{2.2.12}$*-Lipschitz,*

(C) $\int_{t \in I} g(t) \leq \widetilde{R}/\sqrt{n}$ *for any interval* I *of size* N*, and*

(D) $\|g\|_\infty \leq L/(N\sqrt{n})$.

For each $i \leq \lfloor \epsilon n \rfloor$, let Y_i be a random variable uniform on some disjoint union of integer intervals of cardinality at least N each, and assume that $Y_1, \dots, Y_{\lfloor \epsilon n \rfloor}$ are mutually independent. Define a random function $\widetilde{g} \in L^1(\mathbb{R})$ by

$$\widetilde{g}(t) = \mathbb{E}_b g\left(t + z\sum_{i=1}^{\lfloor \epsilon n \rfloor} b_i Y_i\right)$$

where b is a vector of independent $\mathrm{Ber}(1/2)$ *components. Then*

$$\mathbb{P}\left[\|\widetilde{g}\|_\infty \geq \frac{9L/10}{N\sqrt{n}}\right] \leq \exp(-Mn).$$

This follows almost immediately from an $\ell^\infty(\mathbb{Z})$ decrement result established by Tikhomirov [103].

Proposition 2.2.13 ([103, Proposition 4.10])**.** *For any $p \in (0, 1/2]$, $\epsilon \in (0,1)$, $\widetilde{R} \geq 1$, $L_0 \geq 16\widetilde{R}$, and $M \geq 1$ there are $n_{2.2.13} = n_{2.2.13}(p, \epsilon, L_0, \widetilde{R}, M) > 0$ and $\eta_{2.2.13} = \eta_{2.2.13}(p, \epsilon, L_0, \widetilde{R}, M) \in (0,1)$ with the following property. Let $L_0 \geq L \geq 16\widetilde{R}$, let $n \geq n_{2.2.13}$, $N \in \mathbb{N}$, and let $g \in \ell^1(\mathbb{Z})$ be a nonnegative function satisfying*

(A) $\|g\|_1 = 1$,

(B) $\log_2 g$ *is $\eta_{2.2.13}$-Lipschitz,*

(C) $\sum_{t \in I} g(t) \leq \widetilde{R}/\sqrt{n}$ *for any integer interval I of size N, and*

(D) $\|g\|_\infty \leq L/(N\sqrt{n})$.

For each $i \leq \lfloor \epsilon n \rfloor$, let Y_i be a random variable uniform on some disjoint union of integer intervals of cardinality at least N each, and assume that $Y_1, \dots, Y_{\lfloor \epsilon n \rfloor}$ are mutually independent. Define a random function $\widetilde{g} \in \ell^1(\mathbb{Z})$ by

$$\widetilde{g}(t) = \mathbb{E}_b g\left(t + \sum_{i=1}^{\lfloor \epsilon n \rfloor} b_i Y_i\right)$$

where b is a vector of independent $\mathrm{Ber}(p)$ *components. Then*

$$\mathbb{P}\left[\|\widetilde{g}\|_\infty > \frac{(1 - p(1 - 1/\sqrt{2}))L}{N\sqrt{n}}\right] \leq \exp(-Mn).$$

Remark 2.2.14. In [103, Proposition 4.10], there is a condition $N \leq 2^n$ which is not necessary (indeed, it is not used anywhere in the proof) and so has been dropped. In fact, we will only need values $N \leq k^n$, in which case one can actually replace n by kn and ϵ by ϵ/k (and adjust other parameters appropriately) in order to deduce what we need directly from the statement as written in [103]. We will only need this statement for $p = 1/2$.

We first prove Proposition 2.2.12.

Proof of Proposition 2.2.12. Consider the operator $\mathcal{O} : L^1(\mathbb{R}) \to \ell^1(\mathbb{Z})$ given by

$$(\mathcal{O}\omega)(t) = \int_{-z/2}^{z/2} \omega(zt + u)\, du.$$

We note that $\|\omega\|_1 = \|\mathcal{O}\omega\|_1$ and if ω is nonnegative and $\log_2 \omega$ is η-Lipschitz, then

$$z2^{-\eta h/2}\|\omega\|_\infty \le \|\mathcal{O}\omega\|_\infty \le z\|\omega\|_\infty.$$

Given $g \in L^1(\mathbb{R})$ satisfying the given conditions, we consider $g' \in \ell^1(\mathbb{Z})$ defined via

$$g' = \mathcal{O}g.$$

We see that g' satisfies properties (A), (B), (C), (D) of Proposition 2.2.13 with log-Lipschitz constant slightly changed (depending on z, hence h), L changed to zL, and $\widetilde{R}$ increased to $4h\widetilde{R}$. These last changes are responsible for the condition $L_0 \ge 64h^2\widetilde{R}$.

Since $zL \ge h^{-1}L \ge 16(4h\widetilde{R})$, we may apply Proposition 2.2.13 to g' to deduce that $\|\widetilde{g'}\|_\infty$ is small, except with superexponentially small probability. Here $\widetilde{g'}$ is averaged in the sense of Proposition 2.2.13 with respect to the same $Y_1, \dots, Y_{\lfloor \epsilon n \rfloor}$.

Now, by Fubini's theorem, note that

$$\widetilde{g'} = \mathcal{O}\widetilde{g},$$

where $\widetilde{g}$ is averaged in the sense of Proposition 2.2.12. Therefore,

$$\mathbb{P}\left[\|\mathcal{O}\widetilde{g}\|_\infty > \frac{(2+\sqrt{2})zL/4}{N\sqrt{n}}\right] \le \exp(-Mn),$$

so that

$$\mathbb{P}\left[\|\widetilde{g}\|_\infty > \frac{(2+\sqrt{2})2^{\eta_{2.2.12}h/2}L/4}{N\sqrt{n}}\right] \le \exp(-Mn).$$

Finally, if $\eta_{2.2.12}$ is appropriately small, we deduce the desired as

$$\frac{2+\sqrt{2}}{4} < \frac{9}{10}.$$

□

Finally, we are able to deduce Proposition 2.2.11.

Proof of Proposition 2.2.11. Similar to the proof of Lemma 2.2.5, we can use an equivalent method of sampling from the $\vec{s}$-multislice to rewrite $\widetilde{g}(t)$ as

$$\widetilde{g}(t) = \mathbb{E}_b\left[g\left(t + \sum_{i=1}^{\lfloor \epsilon n \rfloor}(b_{2i-1}Y_{2i-1} + b_{2i}Y_{2i})\right)\middle| \#\{b_i = a_j\} = s_j \ \forall j \in [k]\right]$$

$$= \mathbb{E}_{b,b'}\left[g\left(t + \sum_{i=1}^{\lfloor \epsilon n \rfloor} b_{2i-1}Y_{2i-1} + b_{2i}Y_{2i} + b'_i(b_{2i} - b_{2i-1})(Y_{2i-1} - Y_{2i})\right)\middle| \#\{b_i = a_j\} = s_j \ \forall j \in [k]\right],$$

where b' is an $\lfloor \epsilon n \rfloor$-dimensional vector with *independent* Ber(1/2) components. Below, we will fix b and use only the randomness in b'. In order to do this, let

$$B_0 := \left\{b_1, \dots, b_{2\lfloor \epsilon n \rfloor} \colon \#\{i \colon b_{2i-1} = a_1, b_{2i} = a_2\} \geq \min(\vec{p})^2 \epsilon n/8\right\}.$$

Then, provided that $\gamma_{2.2.11}$ is chosen sufficiently small depending on ξ, and n is sufficiently large depending on ξ and ϵ, we have

$$\mathbb{E}_b[1_{B_0} | \#\{b_i = a_j\} = s_j \ \forall j \in [k]] > \frac{1}{2}.$$

Let $\mathcal{E}_L$ denote the event (depending on $Y_1, \dots, Y_{2\lfloor \epsilon n \rfloor}$) that $\|\widetilde{g}\|_\infty > 19L/(20N\sqrt{n})$. Now, suppose $Y_1, \dots, Y_{2\lfloor \epsilon n \rfloor} \in \mathcal{E}_L$, and suppose further that $\|\widetilde{g}\|_\infty$ is attained at $t \in \mathbb{R}$. Let

$$B_1 := \left\{b_1, \dots, b_{2\lfloor \epsilon n \rfloor} \colon \mathbb{E}_{b'}\left[g\left(t + \sum_{i=1}^{\lfloor \epsilon n \rfloor} (b_{2i-1}Y_{2i-1} + b_{2i}Y_{2i} + b'_i(b_{2i} - b_{2i-1})(Y_{2i-1} - Y_{2i})\right)\middle| b\right] \geq \frac{9L/10}{N\sqrt{n}}\right\}.$$

Since $\|g\|_\infty \leq L/(N\sqrt{n})$, it follows from the reverse Markov inequality that

$$\mathbb{E}_b[1_{B_1} | \#\{b_i = a_j\} = s_j \ \forall j \in [k]] > \frac{1}{2}.$$

Thus, we see that for every $(Y_1, \dots, Y_{2\lfloor \epsilon n \rfloor}) \in \mathcal{E}_L$, there exists some $b \in B_0 \cap B_1$. Hence, taking a union bound, we see that

$$\begin{aligned}
\mathbb{P}\left[\|\widetilde{g}\|_\infty > \frac{19L/20}{N\sqrt{n}}\right] &\leq \mathbb{P}[\exists b \in B_0 : b \in B_1] \\
&\leq |B_0| \sup_{b \in B_0} \mathbb{P}\left[\exists t \colon \mathbb{E}_{b'}\left[g\left(t + \sum_{i=1}^{\lfloor \epsilon n \rfloor} (b_{2i-1}Y_{2i-1} + b_{2i}Y_{2i} + b'_i(b_{2i} - b_{2i-1})(Y_{2i-1} - Y_{2i}))\right)\right] \geq \frac{9L/10}{N\sqrt{n}}\right] \\
&\leq |B_0| \sup_{b \in B_0} \mathbb{P}\left[\exists t : \mathbb{E}_{b'}\left[g\left(t + \sum_{i=1}^{\lfloor \epsilon n \rfloor} b'_i(b_{2i} - b_{2i-1})(Y_{2i-1} - Y_{2i})\right)\right] \geq \frac{9L/10}{N\sqrt{n}}\right]. \qquad (2.2.9)
\end{aligned}$$

We now bound the probability appearing on the right hand side of the above equation uniformly for $b \in B_0$. We fix $b \in B_0$, and note that, by definition, there is a set $I = \{i_1, \dots, i_m\} \subseteq \lfloor \epsilon n \rfloor$ such that $|I| = m \geq \min(\vec{p})^2 \epsilon n/8$ and such that for all $j \in [m]$,

$$b'_{i_j}(b_{2i_j} - b_{2i_j-1})(Y_{2i_j-1} - Y_{2i_j}) = b'_{i_j}(a_2 - a_1)(Y_{2i_j-1} - Y_{2i_j}).$$

For $j \in [k]$, let $Y_j^b := Y_{2i_j} - Y_{2i_j-1}$. Let $Y_{-2\cdot I}$ denote all components of $Y_1, \dots, Y_{2\lfloor \epsilon n \rfloor}$, except those corresponding to indices in $2 \cdot I$, and let $Y_{2\cdot I}$ denote the remaining components. Then, for $b \in B_0$ and a choice of $Y_{-2\cdot I}$, we define the random function (depending on $Y_{2\cdot I}$),

$$\widetilde{g}_{b,Y_{-2\cdot I}}(t) := \mathbb{E}_{b'} g\left(t + (a_1 - a_2) \sum_{j=1}^{\lfloor \min(\vec{p})^2 \epsilon n/8 \rfloor} b'_j Y_j^b\right).$$

Thus, we see that for any $b \in B_0$ and $Y_{-2\cdot I}$, the probability appearing on the right hand side of (2.2.9) is bounded by

$$\mathbb{P}\left[\|\widetilde{g}_{b,Y_{-2\cdot I}}\|_\infty \geq \frac{9L/10}{N\sqrt{n}}\right],$$

where the probability is over the choice of $Y_{2\cdot I}$.

At this point, we can apply Proposition 2.2.12 to $\widetilde{g}_{b,Y_{-2\cdot I}}$. Let us quickly check that the hypotheses of Proposition 2.2.12 are satisfied. The assumptions on g needed in Proposition 2.2.12 are satisfied because the same properties are assumed in Proposition 2.2.11 (see below for the log-Lipschitz condition). Moreover, $b'_1, \dots, b'_{\lfloor \min(\vec{p})^2 \epsilon n/8 \rfloor}$ are independent $\mathrm{Ber}(1/2)$ random variables. Finally, notice that, given $Y_{-2\cdot I}$, each Y_j^b is a random variable uniform on some disjoint intervals of cardinality at least N each (since Y_j^b is a translation of Y_{2i_j} which is assumed to satisfy this property). Also, $a_1 - a_2$ is bounded away from 0 (in terms of ξ).

Thus, Proposition 2.2.12 shows that the expression on the right hand side of (2.2.9) is bounded above by

$$|B_0| \sup_{b \in B_0, Y_{-2\cdot I}} \mathbb{P}\left[\|\widetilde{g}_{b,Y_{-2\cdot I}}\|_\infty > \frac{9L/10}{N\sqrt{n}}\right] \leq k^n \exp(-M \min(\vec{p})^2 n/8),$$

provided that we choose $\eta_{2.2.11}$ sufficiently small compared to $\eta_{2.2.12}(d, \min(\vec{p})^2\epsilon/8, L_0, \tilde{R}, M)$, where $d = \max(|a_2 - a_1|, |a_2 - a_1|^{-1})$. The desired result now follows after rescaling M by a constant factor (depending on ξ). □

2.2.4 Deriving the final result

We now prove the main result of this section, Theorem 2.2.1. The proof of Theorem 2.2.1 given Propositions 2.2.7 and 2.2.11 is similar to the derivation in [103, Theorem 4.2] however we record the argument in full detail below.

Proof of Theorem 2.2.1. Fix ξ and any admissible parameters $\delta, K_1, K_2, K_3, \epsilon, N$, and the given parameter $M \geq 1$. We need to choose $L_{2.2.1}, \gamma_{2.2.1}, \eta_{2.2.1}, n_{2.2.1}$, where the first two quantities are allowed to depend on all the parameters *except* M, and the last two quantities are allowed to depend on all the parameters.

We let

$$L' := L_{2.2.7}(\xi, \epsilon/2, \delta, K_1, K_2, K_3, 2M); \quad \gamma_{2.2.1} := \gamma = \min\{\gamma_{2.2.7}(\xi, \epsilon/2, \delta, K_1, K_2, K_3), \gamma_{2.2.11}(\xi)\}/4;$$

note that $\gamma_{2.2.1} = \gamma$ does not depend on M. We choose

$$\widetilde{R} := C_{2.2.6}(1/4, \delta/2, K_1, K_2); \quad L_{2.2.1} := 16\widetilde{R};$$

note that $L_{2.2.1}$ does not depend on M, as desired. We choose q to be the smallest positive integer for which

$$0.95^q L' \leq 16\widetilde{R}.$$

Now, let

$$\eta_{2.2.1} = \eta_{2.2.11}(p, \gamma n/2q, \max\{L', 16\widetilde{R}\}, \widetilde{R}, 2M),$$

and suppose that $f \in \ell^1(\mathbb{Z})$ with $\|f\|_1 = 1$, and that $\log_2 f$ is $\eta_{2.2.1}$-Lipschitz.

Step 1: Let $\ell := \lceil (1-\gamma)n \rceil$. Since $N \leq \exp((H(\vec{p})-\epsilon)n)$, it follows from Proposition 2.2.7 and the choice of parameters that, as long as $\|\vec{s} - \vec{p}\ell\|_\infty \leq 4\gamma\ell$, then for all sufficiently large n,

$$\mathbb{P}[\|f_{\mathcal{A},\vec{s},\ell}\|_\infty \geq L'(N\sqrt{n})^{-1}] \leq \exp(-2Mn).$$

Let $\mathcal{E}_0$ be the event that $\|f_{\mathcal{A},\vec{s},\ell}\|_\infty < L'(N\sqrt{n})^{-1}$ simultaneously for all $\vec{s}$ satisfying $\|\vec{s} - \vec{p}\ell\|_\infty \leq 4\gamma\ell$. Then, by the union bound, we see that

$$\mathbb{P}[\mathcal{E}_0^c] \leq n^k \exp(-2Mn).$$

Step 2: We split the interval $[\ell+1, n]$ into q subintervals of size $\gamma n/q$ each, which we denote by $I_1, \ldots, I_q$. Note that

$$f_{\mathcal{A},\vec{s},n}(t) = \mathbb{E}_b f_{\mathcal{A},\vec{s'},\ell}\bigg(t + \sum_{i=1}^{q}\sum_{j\in I_i} b_j X_j\bigg);$$

here, we have sampled a uniform point in the multislice $\#\{b_i = a_j\} = s_j \ \forall j \in [k]$ by first sampling from the distribution of its last γn coordinates, which we denote by $b = (b_{\ell+1}, \ldots, b_n)$, and then sampling the remaining coordinates, subject to the constraint that the amounts of each value a_j is in total equal to s_j. For each $j \in [k]$ let s'_j be the number of values in b equal to a_j (which is therefore a random variable).

Note that if $s_j \in [p_j n - \gamma n, p_j n + \gamma n]$, then we always have $s'_j \in [p_j\ell - 2\gamma\ell, p_j\ell + 2\gamma\ell]$. In particular, on the event $\mathcal{E}_0$, we have for all $\vec{s}$ satisfying $\|\vec{s} - \vec{p}n\|_\infty \leq \gamma n$ and for all possible realizations of $\vec{s'}$ that

$$\|f_{\mathcal{A},\vec{s'},\ell}\|_\infty < L'(N\sqrt{n})^{-1}.$$

Step 3: For $i \in [q]$, let $\vec{s}^{(i)}$ be the vector of values $s_j^{(i)} = \#\{u \in I_i : b_u = a_j\}$ for $j \in [k]$. Let $\mathcal{G}$ be the event that $\|\vec{s}^{(i)} - \vec{p}|I_i|\|_\infty \leq \gamma|I_i|$ for all $i \in [q]$. Then, by a standard large deviation estimate, it follows that for all n sufficiently large, $\mathbb{P}[\mathcal{G}^c] \leq n^{-1/2}$ (say). Hence, using the conclusion of the previous step, conditioning on the values $\vec{s}^{(1)}, \ldots, \vec{s}^{(q)}$, and using the law of total probability, we see that on the event $\mathcal{E}_0$,

$$f_{\mathcal{A},\vec{s},n}(t) \leq L' n^{-1/2} \cdot (N\sqrt{n})^{-1} + \sup_{\substack{\|\vec{s}^{(i)}/|I_i| - \vec{p}\|_\infty \leq \gamma \\ \text{for all } i\in[q]}} \mathbb{E}_b f_{\mathcal{A},\vec{s'},\ell}\bigg(t + \sum_{i=1}^{q}\sum_{u\in I_i} b_u X_j\bigg), \tag{2.2.10}$$

where each vector $(b_u)_{u\in I_i}$ is independently sampled uniformly from the multislice corresponding to $\vec{s}^{(i)}$. Fix vectors $\vec{s}^{(1)},\dots,\vec{s}^{(q)}$ such that $\|\vec{s}^{(i)}/|I_i| - \vec{p}\|_\infty \le \gamma$ and $\|\vec{s}^{(i)}\|_1 = |I_i|$. In particular, this fixes $\vec{s'} = \vec{s} - \vec{s}^{(1)} - \cdots - \vec{s}^{(q)}$. We define the sequence of functions $(g_r)_{r=0}^q$ by

$$g_0(t) = f_{\mathcal{A},\vec{s'},\ell}(t)$$
$$g_r(t) = \mathbb{E}_b\left[f_{\mathcal{A},\vec{s'},\ell}\left(t + \sum_{i=1}^{r}\sum_{j\in I_i} b_j X_j\right)\middle| \#\{u\in I_i : b_u = a_j\} = \vec{s}_j^{(i)}\ \forall j\in[k], i\in[r]\right] \text{ for } r\ge 1.$$

Note that for all $r\in[q]$,

$$g_r(t) = \mathbb{E}_b\left[g_{r-1}\left(t + \sum_{j\in I_k} b_j X_j\right)\middle| \#\{u\in I_r : b_u = a_j\} = \vec{s}_j^{(r)}\ \forall j\in[k]\right].$$

Step 4: We now wish to apply Proposition 2.2.11 to $g_0,\dots,g_{q-1}$ successively with parameters $L_0,\dots,L_{q-1}$ given by

$$L_r := L'\cdot(19/20)^r.$$

More precisely, for $r\ge 1$, let $\mathcal{E}_r$ denote the event that $\|g_r\|_\infty \le L_r(N\sqrt{n})^{-1}$. We claim that

$$\mathbb{P}[\mathcal{E}_r|\mathcal{E}_{r-1}] \ge 1 - \exp(-2Mn) \text{ for all } r\in[q]. \tag{2.2.11}$$

Let us quickly check that on the event $\mathcal{E}_{r-1}$, the hypotheses of Proposition 2.2.11 are satisfied for g_{r-1}. We have $\|g_{r-1}\|_1 = 1$ and $\log_2 g_{r-1}$ is $\eta_{2.2.1}$-Lipschitz since it is a convex combination of functions satisfying the same properties. The condition $\|g_{r-1}\|_\infty \le L_{r-1}(N\sqrt{n})^{-1}$ holds on $\mathcal{E}_{r-1}$ by definition. Moreover, the condition that for any interval I of size N,

$$\int_{t\in I} g_{r-1}(t) \le \widetilde{R}/\sqrt{n}$$

follows since, by Lemma 2.2.6 and our choice of $\widetilde{R}$, the analogous property holds for $f_{\mathcal{A},\vec{s'},\ell}$, and hence for g_{r-1}, which is a convex combination of translates of $f_{\mathcal{A},\vec{s'},\ell}$. Proposition 2.2.11 now justifies (2.2.11). In particular, by the union bound, we have

$$\mathbb{P}[\mathcal{E}_q \mid \mathcal{E}_0] \ge 1 - q\exp(-2Mn).$$

Combine this with the estimate on $\mathcal{P}[\mathcal{E}_0^c]$ (with an at most n^{kq} sized union bound to account for the choice of $\vec{s}^{(1)},\dots,\vec{s}^{(q)}$), then use (2.2.10), and finally take n sufficiently large (so that all the quoted results hold). This gives the desired conclusion. □

2.2.5 Independent model

We conclude this section with an analogue of Corollary 2.2.3 in the independent case.

Theorem 2.2.15. *Fix a discrete distribution ξ. For $0 < \delta < 1/4$, $K_3 > K_2 > K_1 > 1$, $\epsilon \ll \|\vec{p}\|_\infty$, and a given parameter $M \geq 1$, there is $L_{2.2.15} = L_{2.2.15}(\xi, \epsilon, \delta, K_1, K_2, K_3) > 0$ independent of M and $n_{2.2.15} = n_{2.2.15}(\xi, \epsilon, \delta, K_1, K_2, K_3, M) \geq 1$ such that the following holds.*

Let $n \geq n_{2.2.15}$, $1 \leq N \leq \|\vec{p}\|_\infty^{-n} \exp(-\epsilon n)$ and $\mathcal{A}$ be $(N, n, K_1, K_2, K_3, \delta)$-admissible. Then

$$\left| \left\{ x \in \mathcal{A} \colon \mathcal{L}_\xi \left(\sum_{i=1}^n b_i x_i, \sqrt{n} \right) \geq L_{2.2.15} N^{-1} \right\} \right| \leq e^{-Mn} |\mathcal{A}|.$$

The proof of Theorem 2.2.15 is analogous to that of Theorem 2.2.1 followed by Corollary 2.2.3, except that the random variables b_i are now independent copies of ξ. This independence simplifies matters dramatically, as one can derive an analogue of Proposition 2.2.7 by simply considering drops (as in [103, Proposition 4.5]) instead of "well-conditioned" drops, and then using subsampling arguments similar to those appearing above to prove analogues of Proposition 2.2.11 and Theorem 2.2.1. We leave the details to the interested reader.

2.3 Sharp invertibility of sparse Bernoulli matrices

In this section, we use Theorem 2.2.1 to confirm Conjecture 2.1.1 for the case $\xi = \mathrm{Ber}(p)$ for fixed $p \in (0, 1/2)$, thereby resolving [71, Problem 8.2]. More precisely, we will show the following.

Theorem 2.3.1. *Fix $p \in (0, 1/2)$. There exist constants $C_p, \epsilon_p, n_p > 0$ such that for all $n \geq n_p$ and $t \geq 0$,*

$$\mathbb{P}[s_n(B_n(p)) \leq t/\sqrt{n}] \leq C_p t + (2 + (1 - \epsilon_p)^n) n (1 - p)^n.$$

Compared to the proof of Theorem 2.1.2, the main difference in this section is the substantially simpler treatment of structured vectors, due to the reasons mentioned in the introduction. At the same time, the arguments used in this section form the basis of developments in subsequent sections, and we hope that encountering them in this simpler setting will clarify their role later in the paper.

2.3.1 Almost-constant and almost-elementary vectors

We recall the usual notion of almost-constant vectors, a modification of compressible vectors (see, e.g., [104]).

Definition 2.3.2 (Almost-constant vectors)**.** For $\delta, \rho \in (0, 1)$, we define $\mathrm{Cons}(\delta, \rho)$ to be the set of $x \in \mathbb{S}^{n-1}$ for which there exists some $\lambda \in \mathbb{R}$ such that $|x_i - \lambda| \leq \rho/\sqrt{n}$ for at least $(1 - \delta)n$ values $i \in [n]$.

The main result of this subsection is the following.

Proposition 2.3.3. *For any $c > 0$, there exist $\delta, \rho, \epsilon, n_0 > 0$ depending only on c, such that for all $n \geq n_0$,*

$$\mathbb{P}[\exists x \in \mathrm{Cons}(\delta, \rho) \colon \|B_n(p)x\|_2 \leq 2^{-cn}] \leq n(1-p)^n + (1-p-\epsilon)^n.$$

For later use, we record the following simple property of non-almost-constant vectors.

Lemma 2.3.4. *For $\delta, \rho \in (0, 1/4)$, there exist $\nu, \nu' > 0$ depending only on δ, ρ, and a finite set $\mathcal{K}$ of* positive *real numbers, also depending only on δ, ρ, such that if $x \in \mathbb{S}^{n-1} \setminus \mathrm{Cons}(\delta, \rho)$, then at least one of the following two conclusions is satisfied.*

1. *There exist $\kappa, \kappa' \in \mathcal{K}$ such that*

$$|x_i| \leq \frac{\kappa}{\sqrt{n}} \text{ for at least } \nu n \text{ indices } i \in [n], \text{ and}$$

$$\frac{\kappa + \nu'}{\sqrt{n}} < |x_i| \leq \frac{\kappa'}{\sqrt{n}} \text{ for at least } \nu n \text{ indices } i \in [n].$$

2. *There exist $\kappa, \kappa' \in \mathcal{K}$ such that*

$$\frac{\kappa}{\sqrt{n}} < x_i < \frac{\kappa'}{\sqrt{n}} \text{ for at least } \nu n \text{ indices } i \in [n], \text{ and}$$

$$-\frac{\kappa'}{\sqrt{n}} < x_i < -\frac{\kappa}{\sqrt{n}} \text{ for at least } \nu n \text{ indices } i \in [n].$$

Proof. Let $I_0 := \{i \in [n] : |x_i| \leq 4/\sqrt{\delta n}\}$. Since $\|x\|_2 = 1$, it follows that $|I_0| \geq (1 - \delta/16)n$. We consider the following cases.

Case I: $|\{i \in I_0 : |x_i| < \rho/(10\sqrt{n})\}| \geq \delta n/16$. Since $x \notin \mathrm{Cons}(\delta, \rho)$, there are at least δn indices $j \in [n]$ such that $|x_j| \geq \rho/\sqrt{n}$. Moreover, at least $\delta n/8$ of these indices satisfy $|x_j| \leq 4/\sqrt{\delta n}$. Hence, in this case, the first conclusion is satisfied for suitable choice of parameters.

Case II: $|\{i \in I_0 : |x_i| < \rho/(10\sqrt{n})\}| < \delta n/16$ and $|\{i \in I_0 : x_i \leq -\rho/(10\sqrt{n})\}| \geq \delta n/16$ and $|\{i \in I_0 : x_i \geq \rho/(10\sqrt{n})\}| \geq \delta n/16$. In this case, the second conclusion is clearly satisfied for suitable choice of parameters.

Case III: Either $|\{i \in I_0 : x_i \geq \rho/(10\sqrt{n})\}| \geq (1 - \delta/4)n$ or $|\{i \in I_0 : x_i \leq -\rho/(10\sqrt{n})\}| \geq (1 - \delta/4)n$. We assume that we are in the first sub-case; the argument for the second sub-case is similar. We decompose

$$[0, 4/\sqrt{\delta n}] = \cup_{\ell=1}^{L} J_\ell,$$

where $J_\ell := [(\ell - 1) \cdot \rho/(10\sqrt{n}), \ell \cdot \rho/(10\sqrt{n}))$, and $L = O_{\delta,\rho}(1)$. Let

$$\ell_0 := \min\{\ell \in L \colon |\{i \in [n] \colon x_i \in J_1 \cup \cdots \cup J_\ell\}| \geq \delta n/16\}.$$

Since $x \notin \mathrm{Cons}(\delta, \rho)$, there are at least δn indices $j \in [n]$ such that $|x_j - \ell_0 \cdot \rho/(10\sqrt{n})| > \rho/\sqrt{n}$. On the other hand, by the assumption of this case and the definition of ℓ_0, there exist at least $\delta n - (\delta n/4) - (\delta n/16) - (\delta n/16) > \delta n/2$ indices $j \in I_0$ for which $x_j > (\ell_0 + 2) \cdot \rho/(10\sqrt{n})$. Thus, the first conclusion is satisfied for suitable choice of parameters. □

We also isolate the following set of almost-elementary vectors.

Definition 2.3.5 (Almost-elementary vectors)**.** For $\delta > 0$ and $i \in [n]$, let

$$\mathrm{Elem}_i(\delta) := \mathbb{S}^{n-1} \cap \mathbb{B}_2^n(e_i, \delta) = \{x \in \mathbb{S}^{n-1} \colon \|x - e_i\|_2 \leq \delta\}.$$

We define the set of δ-almost-elementary vectors by

$$\mathrm{Coord}(\delta) := \bigcup_{i=1}^{n} \mathrm{Elem}_i(\delta).$$

We will need a standard concentration estimate for the operator norm of a random matrix with independent centered sub-Gaussian entries.

Lemma 2.3.6 ([105, Lemma 4.4.5])**.** *There exists an absolute constant $C > 0$ such that the following holds. Let A be an $m \times n$ i.i.d. matrix with mean 0, sub-Gaussian entries with sub-Gaussian norm at most K. Then for any $t \geq 0$ we have*

$$\mathbb{P}[\|A\| \leq C(\sqrt{m} + \sqrt{n} + t)] \leq 2\exp(-t^2/K^2).$$

To prove Proposition 2.3.3, we will handle almost-elementary vectors and non-almost-elementary vectors separately. We begin with the easier case of non-almost-elementary vectors.

2.3.1.1 Invertibility on non-almost-elementary vectors

Using Lemma 2.2.4, we can prove the following elementary fact about sums of independent $\mathrm{Ber}(p)$ random variables.

Lemma 2.3.7. *Fix $p, \delta \in (0, 1/2)$. There exists $\theta = \theta(\delta, p) > 0$ such that for all $x \in \mathbb{S}^{n-1} \setminus \mathrm{Coord}(\delta)$,*

$$\mathcal{L}(b_1 x_1 + \cdots + b_n x_n, \theta) \leq 1 - p - \theta,$$

where $b = (b_1, \ldots, b_n)$ is a random vector whose coordinates are independent $\mathrm{Ber}(p)$ random variables.

Proof. Since $\mathrm{Coord}(\delta)$ is increasing with δ, it suffices to prove the statement for all sufficiently small δ (depending on p). We may assume that $|x_1| \geq |x_2| \geq \cdots \geq |x_n|$. The desired conclusion follows by combining the following two cases.

Case 1: Suppose $|x_2| > \delta^4$. We claim that there is some $\theta = \theta(\delta, p)$ for which

$$\mathcal{L}(b_1 x_1 + \cdots + b_n x_n, \theta) \leq \mathcal{L}(b_1 x_1 + b_2 x_2, \theta) \leq 1 - p - \theta.$$

We borrow elements from [71, Proposition 3.11]. The first inequality is trivial. For the second inequality, we note that the random variable $b_1 x_1 + b_2 x_2$ is supported on the four points $\{0, x_1, x_2, x_1 + x_2\}$. Moreover, the sets $\{0, x_1 + x_2\}$ and $\{x_1, x_2\}$ are δ^4-separated, and each of

these two sets is attained with probability at most $\max\{p^2+(1-p)^2, 2p(1-p)\} < 1-p-\theta$, where the final inequality uses $p < 1/2$.

Case 2: $|x_2| \le \delta^4$. Note that we must have $\|(x_2,\dots,x_n)\|_2 \ge \delta/2$, since otherwise, we would have $\|x - e_1\|_2 < \delta$, contradicting $x \notin \mathrm{Coord}(\delta)$. We claim that there is some $\theta = \theta(\delta, p) > 0$ such that

$$\mathcal{L}(b_1x_1 + \cdots + b_nx_n, \theta) \le \mathcal{L}(b_2x_2 + \cdots + b_nx_n, \theta) \le 1 - p - \theta.$$

Once again, the first inequality is trivial. The second inequality follows from Lemma 2.2.4 applied with $\xi_i = b_ix_i$ for $i = 2,\dots,n$, $r_i = |x_i|/4$, and $r = \delta^4$, and from our assumption that δ was small enough in terms of p. □

By combining the preceding statement with a standard net argument exploiting the low metric entropy of almost-constant vectors and the well-controlled operator norm of random matrices with independent centered subgaussian entries (Lemma 2.3.6), we obtain invertibility on non-almost-elementary vectors.

Proposition 2.3.8. *Fix $p \in (0, 1/2)$. Then, for any $\delta' > 0$, there exist $\delta, \rho, \epsilon', n_0 > 0$ (depending on p, δ') such that for all $n \ge n_0$,*

$$\mathbb{P}[\exists x \in \mathrm{Cons}(\delta,\rho) \setminus \mathrm{Coord}(\delta') \colon \|B_n(p)x\|_2 \le \epsilon'\sqrt{n}] \le (1-p-\epsilon')^n.$$

Proof. The argument is closely related to the proof of [103, Proposition 3.6], and we omit the standard details. The point is that we can first choose ϵ' depending on δ', p, then choose ρ sufficiently small depending on δ', p, ϵ', and finally, choose δ sufficiently small depending on all prior choices. □

2.3.1.2 Invertibility on almost-elementary vectors

We now prove the much more delicate claim that $\mathrm{Coord}(\delta')$ contributes the appropriate size to the singularity probability.

Proposition 2.3.9. *Fix $p \in (0, 1/2)$. Given $\theta' > 0$, there exist $\delta', \theta, n_0 > 0$ depending on p and θ' such that for all $n \ge n_0$,*

$$\mathbb{P}[\exists x \in \mathrm{Coord}(\delta') \colon \|B_n(p)x\|_2 \le \exp(-\theta' n)] \le n \cdot \Big((1-p)^n + (1-p-\theta)^n \Big).$$

Before proceeding to the proof, we need the following preliminary fact, which essentially follows from the seminal work of Rudelson and Vershynin [86] (although we were not able to locate the precise statement needed here in the literature).

Lemma 2.3.10. *Fix $p \in (0,1)$. For any $c > 0$, there exist $c', n_0 > 0$ depending on c and p for which the following holds. For all $n \ge n_0$ and for any $v \in \mathbb{R}^n$ with $\|v\|_2 \ge 1$, we have*

$$\mathbb{P}[\exists x \in \mathbb{R}^{n-1} \colon \|Ax - v\|_2 \le 2^{-cn}] \le 2^{-c'n},$$

where A is a random $n \times (n-1)$ matrix with independent $\mathrm{Ber}(p)$ entries.

Proof. By reindexing the coordinates, we may write

$$A = \begin{bmatrix} R \\ B \end{bmatrix}, \quad v = \begin{bmatrix} v_1 \\ v' \end{bmatrix},$$

where B is an $(n-1) \times (n-1)$ matrix, $v' \in \mathbb{R}^{n-1}$ and $\|v'\|_2 \geq 1/2$. Let $\mathcal{E} = \{s_{n-1}(B) \leq 2^{-cn/2}\}$. Then, by an extension of the main result of Rudelson and Vershynin [86] (see, e.g., [42, Theorem 1.3] for a concrete reference), $\mathbb{P}[\mathcal{E}] \leq 2^{-c_1 n}$ for some $c_1 > 0$ depending on c and p.

Moreover, on the event $\mathcal{E}^c$, if there exists some $x \in \mathbb{R}^{n-1}$ such that $\|Ax - v\|_2 \leq 2^{-cn}$, then

$$\|Bx - v'\|_2 \leq 2^{-cn} \implies \|x - B^{-1}v'\|_2 \leq 2^{-cn/2}, \quad \text{and}$$
$$|Rx - v_1| \leq 2^{-cn} \implies |R(B^{-1}v') - v_1| \leq 2^{-cn} + n2^{-cn/2}.$$

Let $x_0 := B^{-1}v'/\|B^{-1}v'\|_2$; this is a random vector depending on B. It follows from a straightforward modification of the argument of Rudelson and Vershynin that there exists a constant $c_2 > 0$, depending on c and p, such that with probability at least $1 - 2^{-c_2 n}$, $\mathcal{L}(\sum_{i=1}^{n-1} b_i \cdot (x_0)_i, 2^{-cn/4}) \leq c_2^{-1} 2^{-c_2 n}$, where $b_1, \ldots, b_n$ are independent $\mathrm{Ber}(p)$ random variables (again, one has to take into account that $\mathrm{Ber}(p)$ does not have mean 0, but this is not an issue, see, e.g., [42]). Let $\mathcal{G}$ denote the event (depending on B) that this occurs.

Finally, note that since $\|B\| \leq n$ deterministically, we have $\|B^{-1}v'\|_2 \geq 1/(2n)$. Hence, we see that for all n sufficiently large depending on c and p, we have

$$\begin{aligned}
\mathbb{P}[\mathcal{E}^c \wedge \|Ax - v\|_2 \leq 2^{-cn}] &\leq \mathbb{P}[\mathcal{E}^c \wedge |R(B^{-1}v') - v_1| \leq 2n \cdot 2^{-cn/2}] \\
&\leq \mathbb{P}[\mathcal{E}^c \wedge \mathcal{G} \wedge |R(B^{-1}v') - v_1| \leq 2n \cdot 2^{-cn/2}] + 2^{-c_2 n} \\
&\leq \mathcal{L}\left(\sum_{i=1}^{n-1} b_i \cdot (x_0)_i, 4n^2 \cdot 2^{-cn/2} \right) + 2^{-c_2 n} \leq 2^{-c_2 n/2},
\end{aligned}$$

which completes the proof. □

Now we are ready to prove Proposition 2.3.9.

Proof of Proposition 2.3.9. By taking a union bound, and exploiting the permutation invariance of the distribution of the matrix, it suffices to prove the statement for $\mathrm{Elem}_1(\delta')$ and without the additional factor of n on the right hand side. Let $\mathcal{G}_K$ denote the event that $\|B_n(p) - pJ_{n\times n}\| \leq K\sqrt{n}$, where $J_{n\times n}$ denotes the $n \times n$ all ones matrix. We fix a choice of K such that $\mathbb{P}[\mathcal{G}_K^c] \leq \exp(-2n)$, which is possible by Lemma 2.3.6.

For $\delta' \in (0, 1/4)$, let $\mathcal{E}_{\delta'}$ denote the event that there exists some $x \in \mathrm{Elem}_1(\delta')$ such that $\|B_n(p)x\|_2 \leq \exp(-\theta' n)$. By rescaling, we see that on the event $\mathcal{E}_{\delta'}$, there exists $y = e_1 + u \in \mathbb{R}^n$ with $u_1 = 0$ and $\|u\|_2 \leq 4\delta'$ for which $\|B_n(p)y\|_2 \leq 2\exp(-\theta' n)$. For convenience, let $u' := (u_2, \ldots, u_n) \in \mathbb{R}^{n-1}$. Writing

$$B_n(p) = \begin{bmatrix} b_{11} & R \\ S & B_{n-1} \end{bmatrix},$$

we see that

$$|Ru' + b_{11}| \le 2\exp(-\theta' n), \qquad \|B_{n-1}u' + S\|_2 \le 2\exp(-\theta' n).$$

Let $B^{(1)}$ denote the first column of $B_n(p)$ and let $B^{(-1)}$ denote the $n \times (n-1)$ matrix formed by excluding the first column of $B_n(p)$. Then, on the event $\mathcal{E}_{\delta'} \wedge \mathcal{G}_K$, we have

$$\begin{aligned}\|B^{(1)} + pJ_{n\times n-1}u'\|_2 &= \|B^{(1)} + B^{(-1)}u' - B^{(-1)}u' + pJ_{n\times n-1}u'\|_2 \\ &\le \|B^{(1)} + B^{(-1)}u'\|_2 + \|(B^{(-1)} - pJ_{n\times n-1})u'\|_2 \\ &\le 4\exp(-\theta' n) + K\sqrt{n}\cdot 4\delta' \\ &\le 8K\delta'\sqrt{n}.\end{aligned}$$

The key point is the following. Let $\mathcal{C} := \{x \in \{0,1\}^n : \exists \lambda \in \mathbb{R} \text{ such that } \|x - \lambda 1_n\|_2 \le 8K\delta'\sqrt{n}\}$, where 1_n denotes the n-dimensional all ones vector. Then, it is readily seen that for any $\epsilon > 0$, there exists $\delta' > 0$ sufficiently small so that

$$\mathbb{P}[B^{(1)} \in \mathcal{C}] \le (1 - p + \epsilon)^n.$$

To summarize, we have shown that for any $\theta', \epsilon > 0$, there exists $\delta' \in (0, 1/4)$ such that

$$\begin{aligned}\mathbb{P}[\mathcal{E}_{\delta'}] &\le \mathbb{P}[\mathcal{E}_{\delta'} \wedge \mathcal{G}_K] + \exp(-2n) \\ &\le \sum_{a\in\mathcal{C}} \mathbb{P}[B^{(1)} = a] \cdot \mathbb{P}[\exists u' \in \mathbb{R}^{n-1}, \|u'\|_2 \le 4\delta' : \|B^{(-1)}u' + a\| \le 2\exp(-\theta' n)]. \quad (2.3.1)\end{aligned}$$

If we only wanted a bound of the form $(1 - p + \epsilon)^n$ on the right hand side, then we would be done. However, since we want a more precise bound, we need to perform a more refined analysis based on whether or not $a = 0$.

Case I: $a = 0$. The contribution of this term to the sum in (2.3.1) is exactly $(1-p)^n$.

Case II: $a \neq 0$. Since $a \in \{0,1\}^n$, we have $\|a\|_2 \ge 1$. In this case, we can apply Lemma 2.3.10 with $c = \theta'$ to see that, for all n sufficiently large,

$$\mathbb{P}[\exists u' \in \mathbb{R}^{n-1}, \|u'\|_2 \le 4\delta' : \|B^{(-1)}u' + a\| \le 2\exp(-\theta' n)] \le 2^{-c'n},$$

where $c' > 0$ depends only on θ' and p. Thus, we see that the contribution of $a \in \mathcal{C}$, $a \neq 0$ to the sum in (2.3.1) is at most

$$(1 - p + \epsilon)^n \times 2^{-c'n}.$$

Since c' does not depend on δ', we can fix c' (depending on θ' and p), and then choose δ' sufficiently small so that $\epsilon > 0$ is small enough to make the above product at most $(1-p-\theta)^n$, for some $\theta > 0$ depending on the previous parameters. □

We now put everything together to prove Proposition 2.3.3.

Proof of Proposition 2.3.3. Let $c > 0$ be as in the statement of the proposition, and choose $\theta' > 0$ such that $\exp(-\theta' n) = 2^{-cn}$. Then, applying Proposition 2.3.9 with this choice of θ', we find δ', θ, n_0 such that for all $n \ge n_0$, we have the desired estimate for $x \in \mathrm{Coord}(\delta')$ (provided that ϵ is chosen small enough). Then, we apply Proposition 2.3.8 with this choice of δ' to find $\delta, \rho, n_0 > 0$ such that for all $n \ge n_0$, we have the desired estimate for $x \in \mathrm{Cons}(\delta, \rho) \setminus \mathrm{Coord}(\delta')$, provided again that we choose $\epsilon > 0$ sufficiently small. □

2.3.2 The structure theorem for Boolean slices

The following is a natural extension of the threshold function appearing in [103] to the Boolean slice.

Definition 2.3.11. Let $p \in (0, 1/2]$, $\gamma \in (0, p)$, and $L \geq 1$. Then, for any integer $n \geq 1$ and $x \in \mathbb{S}^{n-1}$, we define

$$\mathcal{T}_{p,\gamma}(x, L) := \sup \left\{ t \in (0,1) \colon \mathcal{L}_{p,\gamma}\left(\sum_{i=1}^{n} b_i x_i, t \right) > Lt \right\}.$$

For $p \in (0, 1/2)$, let $H = H(p)$ denote an $(n-1) \times n$ random matrix, each of whose entries is an independent copy of a $\mathrm{Ber}(p)$ random variable. We fix a function $v(H)$ which takes as input an $(n-1) \times n$ matrix and outputs a unit vector in its right kernel. The goal of this subsection is to prove the following result about the threshold function of $v(H)$.

Proposition 2.3.12. *Let $\delta, \rho, \epsilon \in (0,1)$. Let $p \in (0, 1/2)$ and let $H = H(p)$ denote an $(n-1) \times n$ random matrix as above. There exist $L_{2.3.12} = L_{2.3.12}(\delta, \rho, p, \epsilon)$, $\gamma_{2.3.12} = \gamma_{2.3.12}(\delta, \rho, p, \epsilon)$ and $n_{2.3.12} = n_{2.3.12}(\delta, \rho, p, \epsilon)$ such that for all $n \geq n_{2.3.12}$, with probability at least $1 - 4^{-n}$, exactly one of the following holds.*

- *$v(H) \in \mathrm{Cons}(\delta, \rho)$, or*
- *$\mathcal{T}_{p,\gamma_{2.3.12}}(v(H), L_{2.3.12}) \leq \binom{n}{pn}^{-1} \exp(\epsilon n)$.*

Remark 2.3.13. We note that $p \in (0, 1/2)$ is not actually necessary for this statement or its proof, and this is crucial in the more general setting of all discrete random variables.

We will need the following lemma, proved using randomized rounding (cf. [72]), which is a straightforward generalization of [103, Lemma 5.3]. We omit details since the proof is identical.

Lemma 2.3.14. *Let $y = (y_1, \dots, y_n) \in \mathbb{R}^n$ be a vector, and let $\mu > 0$, $\lambda \in \mathbb{R}$ be fixed. Let Δ denote a probability distribution which is supported in $[-s, s]^n$. There exist constants $c_{2.3.14}$ and $C_{2.3.14}$, depending only on s, for which the following holds. Suppose that for all $t \geq \sqrt{n}$,*

$$\mathbb{P}\left[\left| \sum_{i=1}^{n} b_i y_i - \lambda \right| \leq t \right] \leq \mu t,$$

where $(b_1, \dots, b_n)$ is distributed according to Δ. Then, there exists a vector $y' = (y'_1, \dots, y'_n) \in \mathbb{Z}^n$ satisfying

(R1) $\|y - y'\|_\infty \leq 1$,

(R2) $\mathbb{P}[|\sum_{i=1}^n b_i y'_i - \lambda| \leq t] \leq C_{2.3.14} \mu t$ *for all* $t \geq \sqrt{n}$,

(R3) $\mathcal{L}(\sum_{i=1}^n b_i y'_i, \sqrt{n}) \geq c_{2.3.14} \mathcal{L}(\sum_{i=1}^n b_i y_i, \sqrt{n})$,

(R4) $|\sum_{i=1}^n y_i - \sum_{i=1}^n y_i'| \le C_{2.3.14}\sqrt{n}$.

We also record two useful tensorization statements.

Lemma 2.3.15 ([103, Lemma 3.2])**.** *Let $\chi_1, \dots, \chi_m$ be independent random variables.*

- *Assume that for all $\epsilon \ge \epsilon_0$,*
$$\mathbb{P}[|\chi_i| \le \epsilon] \le K\epsilon.$$
Then for $\epsilon \ge \epsilon_0$
$$\mathbb{P}[\|(\chi_1, \dots, \chi_m)\|_2 \le \epsilon\sqrt{m}] \le (CK\epsilon)^m,$$
where C is an absolute constant.

- *Assume that for some $\eta, \tau > 0$,*
$$\mathbb{P}[|\chi_i| \le \eta] \le \tau.$$
Then for $\epsilon \in (0, 1]$,
$$\mathbb{P}[\|(\chi_1, \dots, \chi_m)\|_2 \le \eta\sqrt{\epsilon m}] \le (e/\epsilon)^{\epsilon m}\tau^{m-\epsilon m}.$$

Proof of Proposition 2.3.12. For lightness of notation, we will often denote $v(H)$ simply by v. If $v \notin \mathrm{Cons}(\delta, \rho)$, it follows from Lemma 2.2.5 that for all $\gamma < p/4$, there exists some $L_0 = L_0(\delta, \rho, p)$ and $C_0 = C_0(\delta, \rho, p)$ such that $\mathcal{T}_{p,\gamma}(v, L) \le C_0 \cdot n^{-1/2}$ for all $L \ge L_0$. Fix K_0 such that the event $\mathcal{E}_{K_0} := \{\|H - pJ_{n-1\times n}\| \le K_0\sqrt{n}\}$ holds with probability at least $1 - 2^{-1729n}$.

Let $L > L_0$ be a parameter to be chosen later, depending on $\delta, \rho, p, \epsilon$. Let $\gamma < 1/4$ be a parameter to be chosen later depending on $\delta, \rho, p, \epsilon$. Fix some $N \in [C_0^{-1} \cdot \sqrt{n}, \binom{n}{pn}\exp(-\epsilon n)]$, and let $\mathcal{U}_N$ denote the event that $\mathcal{T}_{p,\gamma}(v, L) \in [1/N, 2/N]$. We proceed to bound $\mathbb{P}[\mathcal{U}_N \wedge \mathcal{E}_{K_0}]$.

Let $D := C_1\sqrt{n}N$, where $C_1 = C_1(\delta, \rho) \ge 1$ will be an integer chosen later. Let $y := Dv$. Since for all $t \ge \sqrt{n}$,

$$\begin{aligned}\mathbb{P}\left[\left|\sum_{i=1}^n b_i y_i\right| \le t\right] &= \mathbb{P}\left[\left|\sum_{i=1}^n b_i v_i\right| \le \frac{t}{C_1\sqrt{n}N}\right] \\ &\le \mathbb{P}\left[\left|\sum_{i=1}^n b_i v_i\right| \le \frac{t}{\sqrt{n}N}\right] \le \frac{L}{N} \cdot \frac{2t}{\sqrt{n}},\end{aligned}$$

it follows that by applying Lemma 2.3.14 to y, with $\mu := 2L/(N\sqrt{n})$, $\lambda = 0$, and the distribution on $\mathbb{R}^n$ coinciding with that of n independent $\mathrm{Ber}(p)$ random variables conditioned to have sum in $[pn - \gamma n, pn + \gamma n]$, we see that for all sufficiently large n, there exists some $y' \in \mathbb{Z}^n$ satisfying the conclusions of Lemma 2.3.14 (note that $C_{2.3.14}, c_{2.3.14}$ in this case are absolute constants). By (R3), we have

$$\mathcal{L}\left(\sum_{i=1}^n b_i y_i', \sqrt{n}\right) \ge c_{2.3.14}\mathcal{L}\left(\sum_{i=1}^n b_i v_i, 1/(C_1 N)\right)$$

$$\geq (2C_1)^{-1} \cdot c_{2.3.14}\mathcal{L}\left(\sum_{i=1}^n b_i v_i, 2/N\right)$$
$$\geq (2C_1)^{-1} \cdot c_{2.3.14} \cdot 2LN^{-1}. \tag{2.3.2}$$

Moreover, by (R1) and (R4), we have on the event $\mathcal{E}_{K_0}$ that

$$\begin{aligned}\|Hy'\|_2 &= \|H(y'-y)\|_2 \\ &\leq \|(H - pJ_{n-1\times n})(y'-y)\|_2 + \|pJ_{n-1\times n}(y'-y)\|_2 \\ &\leq K_0' n, \end{aligned} \tag{2.3.3}$$

where K_0' is a constant depending only on K_0.

We claim that there is an absolute constant $C_2 > 0$ and a constant $C_3 = C_3(\delta,\rho)$, a collection of real numbers (depending on δ,ρ) $(K_3)_j > (K_2)_j > (K_1)_j > 1$ for $j \in [C_3 \cdot C_2^n]$, a positive real number $\delta' > 0$ (depending on δ,ρ), and a collection of $(N, n, (K_1)_j, (K_2)_j, (K_3)_j, \delta')$-admissible sets $\mathcal{A}_j$ (depending on δ,ρ) for $j \in [C_3 \cdot C_2^n]$ such that $y' \in \mathcal{A}_j$ for some $j \in [C_3 \cdot C_2^n]$. Let $\widetilde{v} := y'/D$. By (R1), it follows that $\|\widetilde{v} - v\|_\infty \leq D^{-1}$. Moreover, by Lemma 2.3.4, there exist ν, ν' depending on δ,ρ, and a finite set $\mathcal{K}$ of positive real numbers, also depending on δ,ρ, such that either the first conclusion or the second conclusion of Lemma 2.3.4 is satisfied for v. Since $D^{-1} \leq C_1^{-1}C_0/n$, we see that there exists n_0 depending on δ,ρ,p such that for all $n \geq n_0$, $\widetilde{v}$ satisfies either the first conclusion or the second conclusion of Lemma 2.3.4, with $\nu/2, \nu'/2$ and $2^{-1}\cdot\mathcal{K} \cup 2\cdot\mathcal{K}$. After paying an overall factor of at most 2^n, we may assume that the νn coordinates of $\widetilde{v}$ satisfying this conclusion are the first νn coordinates. The remaining $(1-\nu)n$ coordinates of $\widetilde{v}$ lie in the $(1-\nu)n$-dimensional ball of radius 1. By a volumetric argument, we see that this ball can be covered by at most 100^n translates of $[0, n^{-1/2}]^{(1-\nu)n}$. By paying an overall factor of 100^n, we may fix the translate of $[0, n^{-1/2}]^{(1-\nu)n}$ that the remaining $(1-\nu)n$ coordinates lie in. Note that each such translate contains at most $(2D/\sqrt{n})^{(1-\nu)n}$ points in $(1/D)\mathbb{Z}^n$. Finally, taking $C_1(\delta,\rho)$ sufficiently large so that $C_1(\delta,\rho)\cdot\min(2^{-1}\cdot\mathcal{K}) > 1$ and rescaling by D proves the claim.

To summarize, we have so far shown the following. For parameters L and γ depending on δ,ρ,p,ϵ (to be chosen momentarily), on the event $\mathcal{U}_N \wedge \mathcal{E}_{K_0}$, the event $\mathcal{B}_j$ holds for some $j \in [C_2 \cdot C_3^n]$, where $\mathcal{B}_j$ is the event that there exists some $y' \in \mathcal{A}_j$ satisfying (2.3.2), (2.3.3), and (by (R2)),

$$\mathbb{P}\left[\left|\sum_{i=1}^n b_i y_i'\right| \leq t\right] \leq C_{2.3.14}\mu t \text{ for all } t \geq \sqrt{n}, \tag{2.3.4}$$

where recall that $\mu = 2L/(N\sqrt{n})$.

We are now ready to specify the parameters L and γ. First, let

$$L' := \max_j L_{2.2.3}(\mathrm{Ber}(p), \epsilon, \delta', (K_1)_j, (K_2)_j, (K_3)_j); \gamma' := \min_j \gamma_{2.2.3}(\mathrm{Ber}(p), \epsilon, \delta', (K_1)_j, (K_2)_j, (K_3)_j)/4.$$

Then, let

$$L := (2C_1)\cdot c_{2.3.14}^{-1}\cdot L' + L_0; \gamma := \gamma'.$$

Our goal is to bound $\mathbb{P}[\cup_j \mathcal{B}_j]$. Let $H_1, \dots, H_{n-1}$ denote the rows of H. By a standard large deviation estimate, we can find an absolute constant $Q \geq 1$ such that the event

$$\mathcal{W}_Q := \{|\{i \in [n-1] \colon \sum_{j=1}^{n} H_{i,j} \notin [pn - \gamma n, pn + \gamma n]\}| \leq Q\}$$

holds with probability at least $1 - 2^{-1729n}$. Then, it suffices to bound $\mathbb{P}[\cup_j (\mathcal{B}_j \wedge \mathcal{W}_Q)]$. We will provide a uniform (in j) upper bound on $\mathbb{P}[\mathcal{B}_j \wedge \mathcal{W}_Q]$, and then conclude using the union bound. Note that on the event $\mathcal{B}_j$, y' belongs to the set $\mathcal{D}_j$ defined by

$$\mathcal{D}_j := \left\{ x \in \mathcal{A}_j \colon \mathcal{L}_{p,\gamma}\left(\sum_{i=1}^{n} b_i x_i, \sqrt{n} \right) \geq LN^{-1} \right\}.$$

By the choice of L, it follows from Corollary 2.2.3 that for any $M \geq 1$, for all sufficiently large n,

$$|\mathcal{D}_j| \leq e^{-Mn} |\mathcal{A}_j| \leq e^{-Mn} (K_3 N)^n, \tag{2.3.5}$$

where $K_3 := \max_j (K_3)_j$. Moreover, it follows from (2.3.4) and Lemma 2.3.15 that

$$\mathbb{P}[\{\|Hy'\|_2 \leq K_0' n\} \wedge \mathcal{W}_Q] \leq \left(\frac{C_4 L K_0'}{N} \right)^{n-Q}, \tag{2.3.6}$$

where $C_4 \geq 1$ is an absolute constant. Here, we have used that on the event $\mathcal{W}_Q$, the entries of at least $n - Q$ rows have sum in $[pn - \gamma n, pn + \gamma n]$.

Finally, from (2.3.5) and (2.3.6), we see that first taking M to be sufficiently large (compared to various constants depending on $\delta, \rho, p, \epsilon$), and then taking n sufficiently large, gives the desired conclusion. □

2.3.3 Proof of Theorem 2.3.1

We now have all the ingredients needed to prove Theorem 2.3.1. The proof uses the insight from [71] of exploiting the exponential gap between $\binom{n}{pn}$ and $(1-p)^n$ for $p < 1/2$ by using a 'row boosting' argument to reduce to an anticoncentration problem on a well-conditioned slice.

Proof of Theorem 2.3.1. Throughout, we fix functions $x(A), y(A)$ which take as input a matrix A and output (fixed, but otherwise arbitrary) right and left least singular unit vectors, respectively. Let $B = B_n(p)$ for simplicity. Fix $\epsilon > 0$ such that $\binom{n}{pn} \exp(\epsilon n) \leq (1 - p - \epsilon)^n$.

Step 1: By the work of Rudelson and Vershynin [86], there is some $c_p > 0$ so that for all $t > 2^{-2c_p n}$,

$$\mathbb{P}[s_n(B) \leq t/\sqrt{n}] \leq C_p t;$$

note that there is a slight complication since $\mathrm{Ber}(p)$ is not centered, but this can be handled using standard techniques (see, e.g., [42, Theorem 1.3]) Therefore, it suffices to consider the case $t \le 2^{-2c_p n}$.

Step 2: For $\delta, \rho \in (0,1)$, we define

$$\mathcal{E}_L(\delta,\rho) = \{\exists y \in \mathrm{Cons}(\delta,\rho)\colon \|y(B)^T B\|_2 \le 2^{-c_p n}\},$$
$$\mathcal{E}_R(\delta,\rho) = \{\exists x \in \mathrm{Cons}(\delta,\rho)\colon \|Bx(B)\|_2 \le 2^{-c_p n}\}.$$

Applying Proposition 2.3.3 with $c_p > 0$, we find that there exist $\delta, \rho, \epsilon' > 0$ such that for all sufficiently large n,

$$\mathbb{P}[s_n(B) \le t/\sqrt{n}] \le 2n(1-p)^n + 2(1-p-\epsilon')^n + \mathbb{P}[s_n(B) \le t/\sqrt{n} \wedge \mathcal{E}_L(\delta,\rho)^c \wedge \mathcal{E}_R(\delta,\rho)^c].$$

Here, we have used that the distribution of B is invariant under transposition.

Step 3: Let $\gamma = \gamma_{2.3.12}(\delta,\rho,p,\epsilon)$. Let $W_\gamma \subseteq \{0,1\}^n$ denote the set of vectors $x \in \{0,1\}^n$ such that $\sum_{i=1}^n x_i \in [pn - \gamma n, pn + \gamma n]$. As in the proof of Proposition 2.3.12, let $Q \ge 1$ be a constant such that the event

$$\mathcal{W}_Q := \{|\{i \in [n]\colon B_i \notin W_\gamma\}| \le Q\}$$

holds with probability at least $1 - 2^{-1729n}$. Then, it suffices to bound $\mathbb{P}[s_n(B) \le t/\sqrt{n} \wedge \mathcal{E}_L^c \wedge \mathcal{E}_R^c \wedge \mathcal{W}_Q]$, where for simplicity, we have omitted the parameters δ, ρ fixed in the previous step.

Let $B_1, \dots, B_n$ denote the rows of B, and for simplicity, let $y = y(B)$. On the event that $s_n(B) \le t/\sqrt{n}$, we have

$$\|y_1 B_1 + \cdots + y_n B_n\|_2 \le t/\sqrt{n}.$$

Moreover, on the event $\mathcal{E}_L^c$, using Lemma 2.3.4, there is a set $I \subseteq [n]$ such that $|I| \ge \nu n$ and such that for all $i \in I$, $|y_i| \ge \kappa/\sqrt{n}$, for some $\kappa := \kappa(\delta,\rho) > 0$. In particular, since for any $i \in [n]$, $\|y_1 B_1 + \cdots + y_n B_n\|_2 \ge |y_i| \operatorname{dist}(B_i, H_i)$, where H_i denotes the span of rows $B_1, \dots, B_{i-1}, B_{i+1}, \dots B_n$, it follows that

$$\operatorname{dist}(B_i, H_i) \le \frac{t}{\kappa} \text{ for all } i \in I.$$

Also, on the event $\mathcal{W}_Q$, there are at least $\nu n/2$ indices $i \in I$ such that $B_i \in W_\gamma$. Thus, we see that

$$\mathbb{P}[s_n(B) \le t/\sqrt{n} \wedge \mathcal{E}_L^c \wedge \mathcal{E}_R^c \wedge \mathcal{W}_Q] \le \frac{2}{\nu n} \sum_{i=1}^n \mathbb{P}[\operatorname{dist}(B_i, H_i) \le t/\kappa \wedge \mathcal{E}_L^c \wedge \mathcal{E}_R^c \wedge B_i \in W_\gamma].$$

Step 4: By symmetry, it suffices to bound $\mathbb{P}[\mathcal{B}_1]$, where

$$\mathcal{B}_1 := \operatorname{dist}(B_1, H_1) \le t/\kappa \wedge \mathcal{E}_R^c \wedge B_1 \in W_\gamma.$$

Let $v(H_1)$ be a unit vector normal to H_1. Then, by Proposition 2.3.12, except with probability 4^{-n} (over the randomness of H_1), exactly one of the following holds.

- $v(H_1) \in \mathrm{Cons}(\delta, \rho)$, or
- $\mathcal{T}_{p,\gamma}(v(H_1), L) \le \binom{n}{pn}^{-1} \exp(\epsilon n)$,

where $L := L_{2.3.12}(\delta, \rho, p, \epsilon)$. If the first possibility occurs, then $\mathcal{B}_1$ cannot hold, since then, $v(H_1) \in \mathrm{Cons}(\delta, \rho)$ satisfies

$$\|Bv(H_1)\|_2 = |\langle B_1, v(H_1)\rangle| \le \mathrm{dist}(B_1, H_1) \le t/\kappa \le 2^{-2c_p n}/\kappa,$$

which contradicts $\mathcal{E}_R^c$ for all n sufficiently large. Hence, the second possibility must hold. But then, using $\mathrm{dist}(B_1, H_1) \ge |\langle B_1, v(H_1)\rangle|$, we have that (over the randomness of B_1),

$$\begin{aligned}\mathbb{P}[\mathrm{dist}(B_1, H_1) \le t/\kappa \wedge B_1 \in W_\gamma] &\le \mathbb{P}[|\langle B_1, v(H_1)\rangle| \le t/\kappa \mid B_1 \in W_\gamma] \\ &\le \frac{Lt}{\kappa} + \binom{n}{pn}^{-1} \exp(\epsilon n) \\ &\le \frac{Lt}{\kappa} + (1 - p - \epsilon)^n.\end{aligned}$$

This completes the proof. □

2.4 Non-almost-constant vectors

In this short section, we prove the following result, which controls the invertibility of random matrices with i.i.d. discrete entries on the bulk of the unit sphere. We note that this is a generalization of the discussion in Sections 2.3.2 and 2.3.3.

Theorem 2.4.1. *Fix a discrete distribution ξ. For any $\delta, \rho, \epsilon > 0$, there exists $C_{2.4.1} = C_{2.4.1}(\xi, \delta, \rho, \epsilon) > 0$ and $n_{2.4.1}(\xi, \delta, \rho, \epsilon) \ge 1$ such that for all $n \ge n_{2.4.1}$ and $t \ge 0$,*

$$\mathbb{P}\left[\inf_{x \in \mathbb{S}^{n-1} \setminus \mathrm{Cons}(\delta,\rho)} \|M_n(\xi)x\|_2 \le t/\sqrt{n} \wedge \inf_{y \in \mathrm{Cons}(\delta,\rho)} \|yM_n(\xi)\|_2 > C_{2.4.1}t\right] \le C_{2.4.1}t + \exp((\epsilon - H(\vec{p}))n).$$

Let ξ be a discrete distribution, and let $A = A(\xi)$ denote an $(n-1) \times n$ random matrix, each of whose entries is an independent copy of a ξ random variable. We fix a function $v(A)$ which takes as input an $(n-1) \times n$ matrix and outputs a unit vector in its right kernel. As in Section 2.3, a key ingredient in the proof of Theorem 2.4.1 is a structure theorem for kernel vectors of A, which encodes the fact that (with very high probability) non-almost-constant kernel vectors of A must be maximally unstructured in the relevant sense.

Definition 2.4.2. Fix a discrete distribution ξ. Let $\vec{\gamma} \in \mathbb{R}^k_{\ge 0}$ with $\|\vec{\gamma}\|_\infty < \min(\vec{p})$, and let $L \ge 1$. Then, for any integer $n \ge 1$ and $x \in \mathbb{S}^{n-1}$, we define

$$\mathcal{T}_{\xi,\vec{\gamma}}(x, L) := \sup\left\{t \in (0,1) \colon \mathcal{L}_{\xi,\vec{\gamma}}\left(\sum_{i=1}^n b_i x_i, t\right) > Lt\right\}.$$

We also define

$$\mathcal{T}_\xi(x, L) := \sup\left\{t \in (0,1)\colon \mathcal{L}_\xi\left(\sum_{i=1}^n b_i x_i, t\right) > Lt\right\}.$$

Proposition 2.4.3. *Let $\delta, \rho, \epsilon \in (0,1)$ and $k = |\operatorname{supp}(\xi)|$. There exist $L_{2.4.3} = L_{2.4.3}(\delta, \rho, \xi, \epsilon)$, $\gamma_{2.4.3} = \gamma_{2.4.3}(\delta, \rho, \xi, \epsilon)$ and $n_{2.4.3} = n_{2.4.3}(\delta, \rho, \xi, \epsilon)$ such that for all $n \geq n_{2.4.3}$, with probability at least $1 - k^{-2n}$, exactly one of the following holds.*

- *$v(A) \in \operatorname{Cons}(\delta, \rho)$, or*
- *$\mathcal{T}_{\xi, \gamma_{2.4.3} 1_k}(v(A), L_{2.4.3}) \leq \exp((\epsilon - H(\vec{p}))n)$.*

Proposition 2.4.3 follows from Corollary 2.2.3 and Lemma 2.3.14 in an identical fashion to the proof given in Section 2.3.2, so we do not repeat it here. The only minor difference is that in the last part of the proof, we now restrict ourselves to the event that all but $O_{\xi, \gamma_{2.2.3}}(1)$ rows belong to a well-conditioned multislice corresponding to ξ (instead of simply restricting to well-conditioned slices).

Given this, we can complete the proof of Theorem 2.4.1.

Proof of Theorem 2.4.1. Let $M := M_n(\xi)$ for simplicity, and let $\delta, \rho, \epsilon > 0$ be as in the statement of the theorem. Let $k = |\operatorname{supp}(\xi)|$ and denote the points in the support of ξ by $a_1, \ldots, a_k$. We will denote the columns of M by $M^{(1)}, \ldots, M^{(n)}$. Also, for each $i \in [n]$, $M^{(-i)}$ denotes the subspace spanned by all columns of M except for $M^{(i)}$.

Step 1: Let $\gamma = \gamma_{2.4.3}(\delta, \rho, \xi, \epsilon)$. Let $W_\gamma \subseteq \operatorname{supp}(\xi)^n$ denote the set of vectors $x \in \operatorname{supp}(\xi)^n$ such that $\#\{x_i = a_j\} \in [p_j n - \gamma n, p_j n + \gamma n]$ for all $j \in [k]$. Let $Q \geq 1$ be a constant such that the event

$$\mathcal{W}_Q := \{|\{i \in [n]\colon M^{(i)} \notin W_\gamma| \leq Q\}$$

holds with probability at least $1 - k^{-1729n}$. Then, it suffices to bound

$$\mathbb{P}\left[\inf_{x \in \mathbb{S}^{n-1} \setminus \operatorname{Cons}(\delta,\rho)} \|Mx\|_2 \leq t/\sqrt{n} \wedge \inf_{y \in \operatorname{Cons}(\delta,\rho)} \|yM\|_2 > Ct \wedge \mathcal{W}_Q\right]. \tag{2.4.1}$$

Let us denote the first of the three events in the equation above by $\mathcal{E}_R$, and the second event by $\mathcal{E}_L$.

Let $x = x(M)$ denote a vector in $\mathbb{S}^{n-1} \setminus \operatorname{Cons}(\delta, \rho)$ certifying the event $\mathcal{E}_R$, so that

$$\|x_1 M^{(1)} + \cdots + x_n M^{(n)}\|_2 \leq t/\sqrt{n}.$$

Using Lemma 2.3.4, there is a set $I \subseteq [n]$ such that $|I| \geq \nu n$ and such that for all $i \in I$, $|x_i| \geq \kappa/\sqrt{n}$, for some $\kappa := \kappa(\delta, \rho) > 0$. In particular, since for any $i \in [n]$, $\|x_1 M^{(1)} + \cdots + x_n M^{(n)}\|_2 \geq |x_i| \operatorname{dist}(M^{(i)}, M^{(-i)})$, it follows that

$$\operatorname{dist}(M^{(i)}, M^{(-i)}) \leq \frac{t}{\kappa} \text{ for all } i \in I.$$

Also, on the event $\mathcal{W}_Q$, there are at least $\nu n/2$ indices $i \in I$ such that $M^{(i)} \in W_\gamma$. Thus, we see that

$$(2.4.1) = \mathbb{P}[\mathcal{E}_R \wedge \mathcal{E}_L \wedge \mathcal{W}_Q] \le \frac{2}{\nu n} \sum_{i=1}^{n} \mathbb{P}[\operatorname{dist}(M^{(i)}, M^{(-i)}) \le t/\kappa \wedge \mathcal{E}_L \wedge M^{(i)} \in W_\gamma].$$

Step 2: By symmetry, it suffices to bound $\mathbb{P}[\mathcal{M}_1]$, where

$$\mathcal{M}_1 := \operatorname{dist}(M^{(1)}, M^{(-1)}) \le t/\kappa \wedge \mathcal{E}_L \wedge M^{(1)} \in W_\gamma.$$

Let $v(M^{(-1)})$ be a unit vector normal to $M^{(-1)}$. Then, by Proposition 2.4.3, except with probability k^{-2n} (over the randomness of $M^{(-1)}$), exactly one of the following holds.

- $v(M^{(-1)}) \in \operatorname{Cons}(\delta, \rho)$, or
- $\mathcal{T}_{\xi,\gamma 1_k}(v(M^{(-1)}), L) \le \exp((\epsilon - H(\vec{p}))n)$,

where $L := L_{2.4.3}(\delta, \rho, \xi, \epsilon)$. If the first possibility occurs, then $\mathcal{M}_1$ cannot hold as $v(M^{(-1)}) \in \operatorname{Cons}(\delta, \rho)$ satisfies

$$\|v(M^{(-1)})M\|_2 = |\langle M^{(1)}, v(M^{(-1)})\rangle| \le \operatorname{dist}(M^{(1)}, M^{(-1)}) \le t/\kappa \le Ct,$$

(choosing C appropriately), which contradicts $\mathcal{E}_L$. Hence, the second possibility must hold. But then, using $\operatorname{dist}(M^{(1)}, M^{(-1)}) \ge |\langle M^{(1)}, v(M^{(-1)})\rangle|$, we have that (over the randomness of $M^{(1)}$),

$$\begin{aligned}\mathbb{P}[\operatorname{dist}(M^{(1)}, M^{(-1)}) \le t/\kappa \wedge M^{(1)} \in W_\gamma] &\le \mathbb{P}[|\langle M^{(1)}, v(M^{(-1)})\rangle| \le t/\kappa \mid M^{(1)} \in W_\gamma] \\ &\le \frac{Lt}{\kappa} + \exp((\epsilon - H(\vec{p}))n)).\end{aligned}$$

□

2.5 Preliminary invertibility estimates

In this section, we will prove a version of Theorem 2.1.8 with the weaker singularity estimate $(\|\vec{p}\|_\infty + o(1))^n$. This estimate, which generalizes [103, Theorem A], will be used crucially in our refined treatment of invertibility for almost-constant vectors in the next section. The techniques in this section also serve as a gentle warm-up to the next section, where much more involved versions of the arguments are presented.

We begin with the following elementary fact regarding sums of ξ random variables.

Lemma 2.5.1. *Fix a discrete distribution ξ. There is $\theta = \theta(\xi) > 0$ such that for all $x \in \mathbb{S}^{n-1}$,*

$$\mathcal{L}_\xi(b_1 x_1 + \cdots + b_n x_n, \theta) \le \|\vec{p}\|_\infty.$$

Proof. This is essentially identical to the proof given in [103, Lemma 3.5]. Briefly, if $\|x\|_\infty \ge \delta$, then we can choose θ small enough (depending on δ and ξ) so the claim is immediate. Otherwise $\|x\|_\infty < \delta$ and $\|x\|_2 = 1$, in which case the claim follows from Lemma 2.2.4 as long as δ is sufficiently small depending on ξ. □

Combining the above estimate with the second part of Lemma 2.3.15, we have the following.

Corollary 2.5.2. *Fix a discrete distribution ξ. For every $\epsilon > 0$, there exists $c > 0$ depending on ϵ and ξ such that for any $x \in \mathbb{S}^{n-1}$ and $y \in \mathbb{R}^n$, we have*

$$\mathbb{P}[\|M_n(\xi)x - y\|_2 \leq c\sqrt{n}] \leq (\|\vec{p}\|_\infty + \epsilon)^n.$$

Moreover, combining this corollary with the low metric entropy of $\mathrm{Cons}(\delta, \rho)$ and Lemma 2.3.6, we obtain the following (weak) estimate for invertibility on almost-constant vectors.

Corollary 2.5.3. *Fix a discrete distribution ξ. For every $\epsilon > 0$, there exist $\delta, \rho, c > 0$ depending on ϵ and ξ such that for any $y \in \mathbb{R}^n$,*

$$\mathbb{P}\left[\inf_{x \in \mathrm{Cons}(\delta,\rho)} \|M_n(\xi)x - y\|_2 \leq c\sqrt{n}\right] \leq (\|\vec{p}\|_\infty + \epsilon)^n.$$

Next, we show that with very high probability, the inverse of any fixed vector is unstructured.

Proposition 2.5.4. *Fix a discrete distribution ξ. For every $\epsilon, \eta > 0$, there exist $\delta, \rho, L > 0$ depending on ϵ, η, ξ such that for any $y \in \mathbb{R}^n$,*

$$\mathbb{P}\left[\exists x \in \mathbb{S}^{n-1} : M_n(\xi)x \parallel y \wedge x \in \mathrm{Cons}(\delta, \rho) \vee \mathcal{T}_\xi(x, L) \geq (\|\vec{p}\|_\infty + \eta)^n\right] \leq (\|\vec{p}\|_\infty + \epsilon)^n.$$

Proof. This follows essentially from combining Corollary 2.5.3 with a cruder analogue of Proposition 2.4.3, the only difference being that we are considering $M_n(\xi)x \parallel y$ for arbitrary $y \in \mathbb{R}^n$ as opposed to only $y = 0$.

To handle this last point, we begin by choosing (using Lemma 2.3.6) a sufficiently large constant K so that $\mathcal{E}_K = \{\|M_n(\xi) - \mathbb{E}[\xi]J_{n\times n}\| \leq K\sqrt{n}\}$ satisfies $\mathbb{P}[\mathcal{E}_K^c] \leq \|\vec{p}\|_\infty^{2n}$. Then, it suffices to restrict to $\mathcal{E}_K$. Moreover, by the triangle inequality, we see that on the event $\mathcal{E}_K$, $\|M_n(\xi)\| \leq K\sqrt{n} + \mathbb{E}[\xi]n$, so that in particular, on the event in the proposition (intersected with $\mathcal{E}_K$), we have that $M_n(\xi)x = ty_0$ with $y_0 \in \mathbb{S}^{n-1}$ fixed and for some $t \in \mathbb{R}$ with $|t| \leq K\sqrt{n} + \mathbb{E}[\xi]n$.

Now, for the treatment of vectors in $\mathrm{Cons}(\delta, \rho)$, we can divide the range of t into n^3 uniformly spaced intervals, apply Corollary 2.5.3 with y equal to the mid-point of an interval times y_0, and use the union bound. For the treatment of vectors x satisfying $(x \in \mathbb{S}^{n-1} \setminus \mathrm{Cons}(\delta, \rho)) \wedge \mathcal{T}_\xi(x, L) \leq (\|\vec{p}\|_\infty + \eta)^n$, we divide the range of t into $\|\vec{p}\|_\infty^{-2n}$ equally spaced intervals, use a slight generalization of the argument in the proof of Proposition 2.3.12 with M sufficiently large (depending on ξ) for each y equal to the mid-point of an interval times y_0, and finally use the union bound. We leave the details to the interested reader. □

Using Corollary 2.5.3 and Proposition 2.5.4, we can prove the following weaker version of Theorem 2.1.8.

Theorem 2.5.5. *Let ξ be a discrete random variable. For any $\epsilon > 0$, there exist $C, n_0 > 0$ depending on ξ, ϵ such that for all $n \geq n_0$ and $t \geq 0$,*

$$\mathbb{P}[s_n(M_n) \leq t/\sqrt{n}] \leq Ct + (\|\vec{p}\|_\infty + \epsilon)^n.$$

Proof. The deduction of this theorem follows from the argument in [103, Section 5] with the application of Corollary 2.5.3 and Proposition 2.5.4 at the appropriate steps. A similar deduction appears in Section 2.3.3 and a more complicated version of this deduction also appears in Section 2.7, so we omit the details. □

2.6 Almost-constant vectors

The goal of this section is to prove Theorem 2.1.10. The proof is presented at the end of the section and needs a few intermediate steps.

For the proof, we will need to isolate the following natural class of almost-elementary vectors.

Definition 2.6.1. (Almost-elementary vectors) For $\delta > 0$ and $i, j \in [n]$, $i \neq j$, let

$$\begin{aligned}
\mathrm{Elem}_i(\delta) &:= \{x \in \mathbb{S}^{n-1} \colon \|x - e_i\|_2 \leq \delta\}, \\
\mathrm{Elem}_{i,j}(\delta) &:= \{x \in \mathbb{S}^{n-1} \colon \|x - (e_i - e_j)/\sqrt{2}\|_2 \leq \delta\}, \\
\mathrm{Elem}'_{i,j}(\delta) &:= \{x \in \mathbb{S}^{n-1} \colon \|x - (e_i + e_j)/\sqrt{2}\|_2 \leq \delta\}.
\end{aligned}$$

Also, let

$$\begin{aligned}
\mathrm{Elem}(\delta) &:= \bigcup_{i \in [n]} \mathrm{Elem}_i(\delta) \cup \bigcup_{i,j \in [n], i \neq j} \mathrm{Elem}_{i,j}(\delta), \\
\mathrm{Elem}'(\delta) &:= \mathrm{Elem}(\delta) \cup \bigcup_{i,j \in [n], i \neq j} \mathrm{Elem}'_{i,j}(\delta).
\end{aligned}$$

Note that $\bigcup_{i \in [n]} \mathrm{Elem}_i(\delta)$ is exactly the set $\mathrm{Coord}(\delta)$ defined in Section 2.3.1.

For excluding almost-constant vectors which are not almost-elementary, we will need to develop sharp results regarding the Lévy concentration function of discrete random variables.

Proposition 2.6.2. *Fix a discrete distribution ξ and $\delta \in (0, 1/2)$. There exists $\theta = \theta(\delta, \xi) > 0$ such that for all $x \in \mathbb{S}^{n-1} \setminus \mathrm{Elem}'(\delta)$,*

$$\mathcal{L}_\xi(b_1 x_1 + \cdots + b_n x_n, \theta) \leq \|\vec{p}\|_2^2 - \theta.$$

Proof. Since $\mathrm{Elem}'(\delta)$ is increasing with δ, it suffices to prove the statement for sufficiently small δ (depending on ξ), which will be chosen during the course of the proof. Moreover, we may assume that $|x_1| \geq |x_2| \geq \cdots \geq |x_n|$.

Since $x \notin \mathrm{Elem}_1(\delta)$, we must have $\|(x_2, \dots, x_n)\|_2 \geq \delta/2$. In case $|x_2| \leq \delta^4$, then we are done using Lemma 2.2.4 (cf. the proof of Lemma 2.3.7) for all sufficiently small δ. Similarly, if $\|(x_3, \dots, x_n)\|_2 \geq \delta/4$ and $|x_3| \leq \delta^4$, we are done. We now analyze the remaining situations via case analysis.

Case I: $\delta^4 \leq |x_2| < (1-\delta^5)|x_1|$. Since $\mathcal{L}_\xi(b_1x_1 + \cdots + b_nx_n, \theta) \leq \mathcal{L}_\xi(b_1x_1 + b_2x_2, \theta)$, it suffices to bound the latter. Let ξ' be an independent copy of ξ. For any $s \in \mathbb{R}$, we have

$$\begin{aligned}\mathbb{P}[x_1\xi + x_2\xi' \in [s-c, s+c]]^2 &= \left(\sum_a \mathbb{P}[\xi' = a]\mathbb{P}[|\xi - x_1^{-1}(s - x_2a)| \leq c|x_1|^{-1}]\right)^2 \\ &\leq \left(\sum_a \mathbb{P}[\xi' = a]^2\right)\left(\sum_a \mathbb{P}[|\xi - x_1^{-1}(s-x_2a)| \leq c|x_1|^{-1}]^2\right) \leq \|\vec{p}\|_2^4,\end{aligned}$$

where the sum is over $a \in \mathrm{supp}(\xi)$. Here, the equality is by definition, the first inequality is Cauchy–Schwarz, and the last inequality holds as long as $c > 0$ is chosen small enough in terms of δ, ξ. Let us elaborate on this final point. We choose $c > 0$ small enough so that $c|x_1|^{-1} \leq c\delta^{-4}$ is smaller than $|x_2/x_1|$ times half the minimum gap in $\mathrm{supp}(\xi)$, which is possible since $|x_2/x_1| \geq \delta^4$. Now, such a choice of c clearly implies that each summand in $\sum_a \mathbb{P}[|\xi - x_1^{-1}(s-x_2a)| \leq c|x_1|^{-1}]^2$ covers at most a single atom in $\mathrm{supp}(\xi)$, and that different choices of $a, a' \in \mathrm{supp}(\xi)$ cover distinct atoms in $\mathrm{supp}(\xi)$.

Moreover, for such a choice of c, equality in the final inequality holds if and only if there is a permutation σ on $\mathrm{supp}(\xi)$ such that for all $a \in \mathrm{supp}(\xi)$,

$$\mathbb{P}[\xi' = \sigma(a)] = \mathbb{P}[|\xi - x_1^{-1}(s - x_2a)| \leq c|x_1|^{-1}].$$

Summing over all the atoms in $\mathrm{supp}(\xi)$, we see that if equality holds in the final inequality, then

$$\mathrm{supp}(\xi) \subseteq \bigcup_{j=1}^{k} [x_1^{-1}(s-x_2a) - c|x_1|^{-1}, x_1^{-1}(s-x_2a) + c|x_1|^{-1}],$$

so that in particular, $\mathrm{supp}(\xi)$ is contained in an interval of length at most $|x_2/x_1|m_\xi + 2c|x_1|^{-1}$, where $m_\xi = \max \mathrm{supp}(\xi) - \min \mathrm{supp}(\xi)$. But since $|x_2/x_1| \leq 1 - \delta^5$ and $c|x_1|^{-1} \leq c\delta^{-4}$, we see (by taking $c > 0$ sufficiently small) that $\mathrm{supp}(\xi)$ is contained in an interval of length at most $(1 - \delta^5/2)m_\xi$, which contradicts the definition of m_ξ. Hence, we see that equality cannot hold in the final inequality.

Since equality does not hold, it follows from the above discussion that (for $c > 0$ sufficiently small), we have the stronger inequality

$$\mathbb{P}[x_1\xi + x_2\xi' \in [s-c, s+c]]^2 \leq \|\vec{p}\|_2^2(\|\vec{p}\|_2^2 - (\min \vec{p})^2),$$

which completes the analysis in this case, noting that the choice of c depends only on ξ, δ.

Case II: $|x_2| \geq (1-\delta^5)|x_1|, \|(x_3, \dots, x_n)\|_2 \leq \delta/4$. This implies that $x \in \mathrm{Elem}'_{1,2}(\delta) \cup \mathrm{Elem}_{1,2}(\delta)$, thereby violating our assumption.

Case III: $\delta^4 \leq |x_3| \leq (1-\delta^5)|x_1|$. This can be treated in exactly the same way as Case I.

Case IV: $(1-\delta^5)|x_1| \le |x_3| \le |x_2|$ and $|x_2| \ge \delta^4$. It suffices to bound $\mathcal{L}_\xi(b_1x_1 + b_2x_2 + b_3x_3)$. Let $u_i \in \{\pm 1\}$ be defined via $u_i = \operatorname{sgn}(x_i) = x_i/|x_i|$. Let $m'_\xi > 0$ be the smallest positive real such that $\operatorname{supp}(\xi) \subseteq [-m'_\xi, m'_\xi]$.

We begin by noting that for any $s \in \mathbb{R}$,

$$\begin{aligned}\mathbb{P}[x_1\xi_1 + x_2\xi_2 + x_3\xi_3 \in [s-c, s+c]] \\ &= \mathbb{P}[|x_1|(u_1\xi_1 + |x_1|^{-1}|x_2|u_2\xi_2 + |x_1|^{-1}|x_3|u_3\xi_3) \in [s-c, s+c]] \\ &\le \mathbb{P}[|x_1|(u_1\xi_1 + u_2\xi_2 + u_3\xi_3) \in [s-c-3\delta^5 m'_\xi, s+c+3\delta^5 m'_\xi]],\end{aligned}$$

where the inequality uses $(1-\delta^5) \le |x_1|^{-1}|x_3| \le |x_1|^{-1}|x_2| \le 1$, $|x_1| \le 1$, and the definition of m'_ξ.

Since $|x_1| \ge |x_2| \ge \delta^4$, this localizes the value of $u_1\xi_1 + u_2\xi_2 + u_3\xi_3$ to an interval of length at most $2(c\delta^{-4} + 3\delta m'_\xi)$. As discussed at the beginning, we can assume that δ is sufficiently small based on ξ. By first choosing $\delta > 0$ sufficiently small depending on ξ, and then choosing $c > 0$ sufficiently small depending on δ and ξ, we may assume that $2(c\delta^{-4} + 3\delta m'_\xi)$ is smaller than the minimum distance between two distinct atoms in both $\operatorname{supp}(\xi + \xi' + \xi'')$ and $\operatorname{supp}(\xi + \xi' - \xi'')$, where ξ, ξ', ξ'' are independent copies of ξ. Note that, after possibly multiplying by an overall negative sign, $u_1\xi_1 + u_2\xi_2 + u_3\xi_3$ is distributed as either $\xi + \xi' + \xi''$ or $\xi + \xi' - \xi''$.

Therefore, by our choice of δ and c, we see that it suffices to show that for all $s \in \mathbb{R}$,

$$\mathbb{P}[\xi_1 + \xi_2 + \xi_3 = s] \le \|\vec{p}\|_2^2 - c_\xi, \quad \mathbb{P}[\xi_1 + \xi_2 - \xi_3 = s] \le \|\vec{p}\|_2^2 - c_\xi,$$

for some $c_\xi > 0$ depending only on ξ. Now for $u_3 \in \{\pm 1\}$, we have

$$\begin{aligned}\mathbb{P}[\xi_1 + \xi_2 + u_3\xi_3 = s]^2 &= \Big(\sum_a \mathbb{P}[\xi_3 = a]\mathbb{P}[\xi_1 + \xi_2 = s - u_3 a]\Big)^2 \\ &\le \Big(\sum_a \mathbb{P}[\xi_3 = a]^2\Big)\Big(\sum_a \mathbb{P}[\xi_1 + \xi_2 = s - u_3 a]^2\Big) \\ &\le \Big(\sum_a \mathbb{P}[\xi_3 = a]^2\Big)\Big(\sum_{a' \in \operatorname{supp}(\xi_1+\xi_2)} \mathbb{P}[\xi_1 + \xi_2 = a']^2\Big) \\ &\le \|\vec{p}\|_2^4,\end{aligned}$$

where the first line is by definition, the second line is Cauchy–Schwarz, and the last line follows by Young's convolution inequality. To obtain the inequality with a positive constant $c_\xi > 0$, we note that equality cannot hold in the third line since $\operatorname{supp}(\xi_1 + \xi_2)$ has strictly more positive atoms than $\operatorname{supp}(\xi)$ (since ξ is supported on at least 2 points), and this leads to the desired improvement since ξ has finite support. □

When ξ is not a translate of an origin-symmetric distribution, the above result can be strengthened.

Proposition 2.6.3. *Fix a discrete distribution ξ and $\delta \in (0, 1/2)$. Suppose that ξ is not a translate of any origin-symmetric distribution. Then, there exists $\theta = \theta(\delta, \xi) > 0$ such that for all $x \in \mathbb{S}^{n-1} \setminus \mathrm{Elem}(\delta)$,*

$$\mathcal{L}_\xi(b_1 x_1 + \cdots + b_n x_n, \theta) \le \|\vec{p}\|_2^2 - \theta.$$

Proof. As before, since $\mathrm{Elem}(\delta)$ is increasing with δ, it suffices to prove the statement for sufficiently small δ depending on ξ. By Proposition 2.6.2, we can choose $\theta = \theta(\delta, \xi) > 0$ such that for all $x \in \mathbb{S}^{n-1} \setminus \mathrm{Elem}'(\delta)$,

$$\mathcal{L}_\xi(b_1 x_1 + \cdots + b_n x_n, \theta) \le \|\vec{p}\|_2^2 - \theta.$$

Hence, it remains to prove the result for $x \in \mathrm{Elem}'(\delta) \setminus \mathrm{Elem}(\delta)$. By symmetry, it suffices to consider $x \in \mathrm{Elem}'_{1,2}(\delta)$. We will bound $\mathcal{L}_\xi(b_1 x_1 + b_2 x_2, \theta)$.

We use an argument similar to **Case IV** of the proof of Proposition 2.6.2. Let $m'_\xi > 0$ be the smallest positive real for which $\mathrm{supp}(\xi) \subseteq [-m'_\xi, m'_\xi]$. We have

$$\mathbb{P}[x_1\xi_1 + x_2\xi_2 \in [s-c, s+c]] \le \mathbb{P}\left[\frac{1}{\sqrt{2}}(\xi_1 + \xi_2) \in [s - c - 2m'_\xi\delta, s + c + 2m'_\xi\delta]\right].$$

Once again, by choosing δ and c sufficiently small (depending on ξ), we may assume that $2(c+2m'_\xi\delta)$ is smaller than the minimum distance between two distinct atoms in $\mathrm{supp}(\xi+\xi')$, where ξ, ξ' are independent copies of ξ. With this choice of δ and c, the problem reduces to showing that there exists some $c_\xi > 0$ depending only on ξ such that for all $s \in \mathbb{R}$,

$$\mathbb{P}[\xi_1 + \xi_2 = s] \le \|\vec{p}\|_2^2 - c_\xi.$$

We have

$$\mathbb{P}[\xi_1+\xi_2 = s] = \sum_a \mathbb{P}[\xi_1 = a]\mathbb{P}[\xi_2 = s-a] \le \left(\sum_a \mathbb{P}[\xi_1 = a]^2\right)^{1/2}\left(\sum_a \mathbb{P}[\xi_2 = s-a]^2\right)^{1/2} \le \|\vec{p}\|_2^2,$$

where the first inequality is Cauchy–Schwarz. To obtain the improved inequality with $c_\xi > 0$, we note that equality can hold in both inequalities if and only if $\mathbb{P}[\xi_1 = a] = \mathbb{P}[\xi_2 = s - a]$, which implies that ξ is a shift (by $s/2$) of an origin-symmetric random variable. Since we have assumed that ξ is not a shift of an origin-symmetric random variable, we see that equality cannot hold, and using that the support of ξ is finite, we can conclude. □

Using the preceding lemmas, and exploiting the low metric entropy of $\mathrm{Cons}(\delta, \rho)$ along with Lemma 2.3.6, we obtain the following corollary. Note that since ξ may not have mean 0, one must perform the standard trick of densifying the net of these vectors along the direction 1_n (see [103, Proposition 3.6]). We omit the (standard) proof; as in Proposition 2.3.8, it is closely related to the proof of [103, Proposition 3.6].

Corollary 2.6.4. *Fix a discrete distribution ξ. For all $\delta' > 0$, there exist $\delta, \rho, \epsilon', n_0 > 0$, depending on ξ and δ', such that for all $n \geq n_0$,*

$$\mathbb{P}[\exists x \in \mathrm{Cons}(\delta, \rho) \setminus \mathrm{Elem}'(\delta') \colon \|M_n(\xi)x\|_2 \leq \epsilon'\sqrt{n}] \leq (\|\vec{p}\|_2^2 - \epsilon')^n.$$

Further, if ξ is not a shift of any origin-symmetric random variable, then the same conclusion holds with $\mathrm{Elem}(\delta')$ instead of $\mathrm{Elem}'(\delta')$.

Given the previous corollary, it remains to analyze vectors in $\mathrm{Elem}'(\delta')$ (or only in $\mathrm{Elem}(\delta')$ if ξ is not a shift of any origin-symmetric random variable), which is the content of the remainder of this section.

2.6.1 Two columns

We first handle vectors in $\mathrm{Elem}_{i,j}(\delta')$. By the invariance of the distribution of $M_n(\xi)$ under permuting columns, it suffices to analyze vectors in $\mathrm{Elem}_{1,2}(\delta')$. We show the following.

Proposition 2.6.5. *Fix a discrete distribution ξ. There exist $\delta', \eta, n_0 > 0$ depending on ξ such that for all $n \geq n_0$ and $t \leq 1$,*

$$\mathbb{P}[\exists x \in \mathrm{Elem}_{1,2}(\delta') \colon \|M_n(\xi)x\|_2 \leq t] \leq \|\vec{p}\|_2^{2n} + (\|\vec{p}\|_2^2 - \eta)^n + t\exp(-\eta n).$$

We will need the following preliminary lemma, which essentially follows from the seminal work of Rudelson and Vershynin [86]. Since we were not able to locate the statement we need in the literature, we provide details below.

Lemma 2.6.6. *Fix $S, s > 0$. There exist $C', c', n_0 > 0$ depending on s, S such that the following holds. For all $n \geq n_0$, any $v \in \mathbb{R}^n$ with $\|v\|_2 \geq 1$, any $\kappa \in (0,1)$, and all $t \leq 1$, we have*

$$\mathbb{P}[\exists x \in \mathbb{R}^{n-1} \colon \|Ax - v\|_2 \leq t] \leq C'n^3\sqrt{t}\exp(\kappa n) + \exp(-c'n)\exp(\kappa n),$$

where A is an $n \times (n-1)$ random matrix, each of whose entries is an independent random variable with sub-Gaussian norm at most S, and such that all but a collection of κn specified entries have variance at least s.

Proof. By the law of total probability, it suffices to assume that the κn specified entries are deterministic, and take the values $a_1, \dots, a_{\kappa n}$. Consider the $n \times (n-1)$ random matrix A', which has the same distribution as A, except for the κn specified entries, which are now replaced by $a_1 + b_1, \dots, a_{\kappa n} + b_{\kappa n}$, where $b_1, \dots, b_{\kappa n}$ are independent $\mathrm{Ber}(1/2)$ random variables.

From a slight generalization of Lemma 2.3.10 (specifically, one should replace the application of [86] with an inhomogeneous version due to [73] and replace 2^{-cn} by t, see the proof of Lemma 2.6.9), we get that there exist C', c', n_0 depending on s, S such that for all $n \geq n_0$, for any $v \in \mathbb{R}^n$ with $\|v\|_2 \geq 1$, and for all $t \geq 1$, we have

$$\mathbb{P}[\exists x \in \mathbb{R}^{n-1} \colon \|A'x - v\|_2 \leq t] \leq C'n^3\sqrt{t} + \exp(-c'n).$$

The conclusion now follows since, with probability $2^{-\kappa n}$, $b_1 = \cdots = b_{\kappa n} = 0$. □

We now prove Proposition 2.6.5.

Proof of Proposition 2.6.5. By Lemma 2.3.6, we can choose $K > 0$ depending on ξ such that $\mathbb{P}[\mathcal{E}_K] \le \|\vec{p}\|_2^{3n}$, where

$$\mathcal{E}_K := \{\|M_n(\xi) - \mathbb{E}[\xi] J_{n\times n}\| \le K\sqrt{n}\}.$$

For $\delta' \in (0, 1/4)$, which will be chosen later in terms of ξ, let

$$\mathcal{E} := \{\exists x \in \mathbb{B}_2^n(e_1, \delta') \cap \mathbb{S}^{n-1} \colon \|M_n(\xi) Q x\|_2 \le t\},$$

where Q is the rotation matrix whose bottom-right $(n-2) \times (n-2)$ minor is the identity matrix, and the top-left 2×2 minor is the rotation matrix given by

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Up to scaling δ' by a constant factor, this is clearly equivalent to the event that we wish to bound.

Note that on the event $\mathcal{E}$, there exists some vector $y = e_1 + u \in \mathbb{R}^n$ with $u_1 = 0$ and $\|u\|_2 \le 4\delta'$ such that

$$\|M_n(\xi) Q y\|_2 \le 2t.$$

Let $u' = (u_2, \dots, u_n) \in \mathbb{R}^{n-1}$, let $\widetilde{M}^{(1)}$ be the first column of $M_n(\xi)Q$, and let $\widetilde{M}^{(-1)}$ denote the $n \times (n-1)$ matrix obtained by removing this column. Then, on the event $\mathcal{E} \wedge \mathcal{E}_K$, we have

$$\begin{aligned} \|\widetilde{M}^{(1)} - \mathbb{E}[\xi] J_{n\times n-1} u'\|_2 &\le \|\widetilde{M}^{(1)} + \widetilde{M}^{(-1)} u'\|_2 + \|(\widetilde{M}^{(-1)} - \mathbb{E}[\xi] J_{n\times n-1}) u'\|_2 \\ &\le 2t + K\sqrt{n} \cdot 4\delta' \\ &\le 8K\delta'\sqrt{n} \end{aligned}$$

for all sufficiently large n, since $t \le 1$.

The key point is the following. Let $\Xi := \operatorname{supp}(\xi - \xi')/\sqrt{2} \subseteq \mathbb{R}$. Let

$$\mathcal{C} := \{a \in \Xi^n \colon \exists \lambda \in \mathbb{R} \text{ with } \|a - \lambda 1_n\|_2 \le 8K\delta'\sqrt{n}\},$$

and for $\kappa = \kappa(\delta', \xi) > 0$, to be chosen later depending on δ', ξ, and for $z \in \Xi$, let

$$\mathcal{C}_z := \mathcal{C} \cap \{a \in \mathbb{R}^n \colon |\operatorname{supp}(a - z1_n)| \le \kappa n\}.$$

It is easy to see that

$$\mathcal{C} \subseteq \bigcup_{z \in \Xi} \mathcal{C}_z$$

for an appropriate choice of κ which goes to 0 as δ' goes to 0. Furthermore,

$$\mathbb{P}[\widetilde{M}^{(1)} \in \mathcal{C}_z] \le \mathbb{P}[(\xi - \xi')/\sqrt{2} = z]^n \exp(c_{\kappa,\delta',\xi} n),$$

where $c_{\kappa,\delta',\xi} > 0$ goes to 0 as κ, δ' go to 0. Therefore, we have

$$\mathbb{P}[\widetilde{M}^{(1)} \in \mathcal{C}] \leq \|\vec{p}\|_2^{2n} \exp(2c_{\kappa,\delta',\xi} n), \text{ and}$$

$$\mathbb{P}[\widetilde{M}^{(1)} \in \mathcal{C} \setminus \mathcal{C}_0] \leq (\|\vec{p}\|_2^2 - c_\xi)^n$$

for some $c_\xi > 0$ depending only on ξ, provided that δ' (hence κ) is chosen sufficiently small. Here, for the second inequality, we have used that by Cauchy–Schwarz (as in the proof of Proposition 2.6.2), the unique most probable atom of $(\xi - \xi')/\sqrt{2}$ is at 0, and is $\|\vec{p}\|_2^2$, so that any other atom in Ξ has probability at most $\|\vec{p}\|_2^2 - 2c_\xi$ for some $c_\xi > 0$.

So far, we have shown that for all κ and δ' sufficiently small (depending on ξ), we have

$$\begin{aligned}\mathbb{P}[\mathcal{E}] &\leq \|\vec{p}\|_2^{3n} + \sum_{a \in \mathcal{C}} \mathbb{P}[\widetilde{M}^{(1)} = a] \mathbb{P}[\exists u' \in \mathbb{R}^{n-1} \colon \|\widetilde{M}^{(-1)} u' + a\|_2 \leq 2t |\widetilde{M}^{(1)}] \\ &\leq \|\vec{p}\|_2^{3n} + (\|\vec{p}\|_2^2 - c_\xi)^n + \sum_{a \in \mathcal{C}_0} \mathbb{P}[\widetilde{M}^{(1)} = a] \mathbb{P}[\exists u' \in \mathbb{R}^{n-1} \colon \|\widetilde{M}^{(-1)} u' + a\|_2 \leq 2t |\widetilde{M}^{(1)}].\end{aligned}$$

We proceed to bound the third term in the above sum.

Case I: If $a = 0$, we have $\mathbb{P}[\widetilde{M}^{(1)} = 0] = \|p\|_2^{2n}$.

Case II: If $a \neq 0$, we have in particular that $\|a\|_2 \geq h_\xi > 0$. The crucial observation is the following. Given $\widetilde{M}^{(1)} = a$, the entries of the first column of $\widetilde{M}^{(-1)}$ are independent random variables, each of which is distributed as the sum of two i.i.d. copies of $\xi/\sqrt{2}$, conditioned on knowing their difference. In particular, for the coordinates $i \in [n]$ for which $a_i = 0$, the corresponding coordinate of the first column of $\widetilde{M}^{(-1)}$ is distributed as $\sqrt{2} \cdot \xi^*$, where ξ^* has the same support as ξ but takes on atom a_i with probability proportional to p_i^2. Thus, we see that conditioned on $\widetilde{M}^{(1)} = a \in \mathcal{C}_0$, all entries of $\widetilde{M}^{(-1)}$ are independent with sub-Gaussian norm at most S_ξ, and all but at most κn entries have variance at least $s_\xi > 0$. Hence, by Lemma 2.6.6, and by using the lower bound $\|a\|_2 \geq h_\xi$, we find that there exist C', c', n_1 depending on ξ such that for all $n \geq n_1$,

$$\mathbb{P}[\exists u' \in \mathbb{R}^{n-1} \colon \|\widetilde{M}^{(-1)} u' + a\|_2 \leq 2t |\widetilde{M}^{(1)}] \leq C' n^3 \sqrt{t} \exp(\kappa n) + \exp(-c' n) \exp(\kappa n).$$

Thus, the contribution of this case is at most

$$\|\vec{p}\|_2^{2n} \exp(2c_{\kappa,\delta',\xi} n) \exp(\kappa n) \Big(C' n^3 \sqrt{t} + 2 \exp(-c' n) \Big).$$

By the AM-GM inequality, we have $\|\vec{p}\|_2^{2n} \sqrt{t} \leq t \|\vec{p}\|_2^n + \|\vec{p}\|_2^{3n}$. The desired conclusion now follows by taking $\eta > 0$ sufficiently small so that $\|\vec{p}\|_2^n \leq \exp(-2\eta n)$, and then taking δ' (hence κ) sufficiently small so that $2c_{\kappa,\delta',\xi} + \kappa < \min(c'/2, \eta/2)$. □

The preceding proposition handles vectors in $\mathrm{Elem}_{i,j}(\delta)$. If the distribution ξ is a translate of an origin-symmetric distribution, we also need to handle vectors in $\mathrm{Elem}'_{i,j}(\delta)$. In case the distribution ξ is itself an origin-symmetric distribution, the desired bound follows

immediately from the previous proposition, using that the distribution of any column of $M_n(\xi)$ is invariant under negation in this case. Therefore, it remains to handle vectors in $\mathrm{Elem}'_{i,j}(\delta)$ when ξ is a nonzero translate of an origin-symmetric distribution, which is done by the next proposition.

Proposition 2.6.7. *Fix a discrete distribution ξ that is a* nonzero *translate of an origin-symmetric distribution. There exist $\delta', \eta, n_0 > 0$ depending on ξ such that for all $n \geq n_0$ and $t \leq 1$,*

$$\mathbb{P}[\exists x \in \mathrm{Elem}'_{1,2}(\delta') \colon \|M_n(\xi)x\|_2 \leq t] \leq (\|\vec{p}\|_2^2 - \eta)^n + t\exp(-\eta n).$$

Proof. The proof is essentially the same as that of Proposition 2.6.5. The lack of the "main term" $\|\vec{p}\|_2^2$ comes from the fact that $e_1 + e_2$ is unlikely to be a kernel vector since ξ is not origin-symmetric.

We quickly discuss the main modifications to the proof of Proposition 2.6.5. Throughout, $s \neq 0$ denotes a real number such that ξ and $s - \xi$ have the same distribution (such an s exists by our assumption about ξ). First, the top-left 2×2 minor of Q is now

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Next, we let $\Xi = \mathrm{supp}(\xi + \xi')/\sqrt{2}$ and as before, let

$$\mathcal{C} := \{a \in \Xi^n \colon \exists \lambda \in \mathbb{R} \text{ with } \|a - \lambda 1_n\|_2 \leq 8K\delta'\sqrt{n}\}$$

and for $z \in \Xi$,

$$\mathcal{C}_z := \mathcal{C} \cap \{a \in \mathbb{R}^n \colon |\mathrm{supp}(a - z1_n)| \leq \kappa n\},$$

where $\kappa = \kappa(\delta', \xi) > 0$ is chosen as in the previous argument. For such a choice of κ, we have

$$\mathbb{P}[\widetilde{M}^{(1)} \in \mathcal{C}] \leq \|\vec{p}\|_2^{2n}\exp(2c_{\kappa,\delta',\xi}n), \text{ and}$$
$$\mathbb{P}[\widetilde{M}^{(1)} \in \mathcal{C} \setminus \mathcal{C}_{s/\sqrt{2}}] \leq (\|\vec{p}\|_2^2 - c_\xi)^n.$$

This time the inequalities are derived as follows. We note that, by Cauchy–Schwarz, for any $z \in \Xi$, $\mathbb{P}[\xi + \xi' = z\sqrt{2}] \leq \|\vec{p}\|_2^2$, with equality holding if and only if $\mathbb{P}[\xi = a] = \mathbb{P}[\xi' = z\sqrt{2} - a]$ for all $a \in \mathrm{supp}(\xi)$, which happens if and only if $z = s/\sqrt{2}$.

Using this, we have as before that

$$\mathbb{P}[\mathcal{E}] \leq \|\vec{p}\|_2^{3n} + (\|\vec{p}\|_2^2 - c_\xi)^n + \sum_{a \in \mathcal{C}_{s/\sqrt{2}}} \mathbb{P}[\widetilde{M}^{(1)} = a]\mathbb{P}[\exists u' \in \mathbb{R}^{n-1} \colon \|\widetilde{M}^{(-1)}u' + a\|_2 \leq 2t|\widetilde{M}^{(1)}].$$

The most important detail is that for $\kappa \leq 1/2$ (say), every $a \in \mathcal{C}_{s/\sqrt{2}}$ is nonzero, since it has at least $(1 - \kappa)n$ coordinates equal to $s/\sqrt{2}$. Since s is a nonzero constant depending only on ξ, we can now use the analysis in **Case II** of the proof of Proposition 2.6.5. The final thing to note is that the distribution of the random variable $(\xi - \xi')/\sqrt{2}$, conditioned on $(\xi + \xi')/\sqrt{2} = s/\sqrt{2}$ coincides with the distribution of $(2\xi^* - s)/\sqrt{2}$, where ξ^* has the same support as ξ, but takes on atom a_i with probability proportional to p_i^2. The remaining details of the proof are essentially the same. □

2.6.2 One column

We now handle vectors in $\mathrm{Elem}_i(\delta')$. Once again, by permutation invariance, it suffices to handle $\mathrm{Elem}_1(\delta')$. We will prove the following.

Proposition 2.6.8. *Fix a discrete distribution ξ. There exist $C', \delta', \eta, n_0 > 0$ depending on ξ such that for all $n \geq n_0$ and $t \leq 1$,*

$$\mathbb{P}[\exists x \in \mathrm{Elem}_1(\delta') \colon \|M_n(\xi)x\|_2 \leq t] \leq p_0^n + C' t \exp(-\eta n) + (\|\vec{p}\|_2^2 - \eta)^n.$$

The analysis is more delicate than the two column case, since (i) we may have $p_0 < \|\vec{p}\|_\infty$, but we still want to isolate p_0 as the major contribution coming from these events, and (ii) we are aiming for an error term of $(\|\vec{p}\|_2^2 - \eta)^n$, which may be smaller than $(p_0 - \eta)^n$. However, given the preparation above, the rest of the proof is similar to the proof in the sparse Bernoulli case, isolated in Proposition 2.3.9, except that we need to replace the application of the results of Rudelson and Vershynin [86] with the much sharper Proposition 2.5.4 and Theorem 2.5.5.

We begin with the following proposition.

Lemma 2.6.9. *Fix a discrete distribution ξ. For any $\eta \in (0,1)$, there exist $C, n_0 > 0$ depending on ξ, η for which the following holds. For any $v \in \mathbb{R}^n$ with $\|v\|_2 \geq 1$, $n \geq n_0$, and $t \geq 0$, we have*

$$\mathbb{P}[\exists x \in \mathbb{R}^{n-1} \colon \|Ax - v\|_2 \leq t] \leq C \cdot n^3 t^{1/2} + (\|\vec{p}\|_\infty + \eta)^n,$$

where A is a random $n \times (n-1)$ matrix with independent ξ entries.

Proof. Fix $\eta > 0$, and let $\mathcal{E}$ be the event whose probability we are trying to control. After potentially reindexing the coordinates, we may write

$$A = \begin{bmatrix} R \\ A_{n-1} \end{bmatrix}, \quad v = \begin{bmatrix} v_1 \\ v' \end{bmatrix}$$

where A_{n-1} is an $(n-1) \times (n-1)$ matrix and $v' \in \mathbb{R}^{n-1}$ satisfies $\|v'\|_2 \geq 1/2$. Let $\mathcal{E}_S = \{s_{n-1}(A_{n-1}) \leq \sqrt{t}\}$. By Theorem 2.5.5, we have that for all sufficiently large n, there exists a constant C' depending on ξ and η such that

$$\mathbb{P}[\mathcal{E}_S] \leq C'\sqrt{nt} + (\|\vec{p}\|_\infty + \eta/2)^n.$$

It therefore suffices to bound the probability of $\mathcal{E} \wedge \mathcal{E}_S^c$. In such a situation, we see that $y := (A_{n-1})^{-1} v'$ is unique. Let $y_0 := y/\|y\|_2$, and for δ, ρ, L to be chosen momentarily, let

$$\mathcal{E}_U = \{y_0 \in \mathrm{Cons}(\delta, \rho) \vee \mathcal{T}_\xi(y_0, L) \geq (\|\vec{p}\|_\infty + \eta/2)^n\}.$$

By Proposition 2.5.4, we can choose $\delta, \rho, L > 0$ depending on ξ and η so that

$$\mathbb{P}[\mathcal{E}_U] \leq (\|\vec{p}\|_\infty + \eta/2)^n.$$

Hence, it suffices to bound the probability of $\mathcal{E} \wedge \mathcal{E}_S^c \wedge \mathcal{E}_U^c$. Let $x \in \mathbb{R}^{n-1}$ be a vector certifying this event. Then, we have for all sufficiently large n that

$$\|A_{n-1}x - v'\|_2 \le t \implies \|x - y\|_2 \le t^{1/2}, \text{ and}$$

$$|Rx - v_1| \le t \implies |Ry - v_1| \le t + nt^{1/2}.$$

Furthermore, since $\|v\|_2 \ge 1$, we have $\|y\|_2 \ge 1/C''n^2$, for some constant C'' depending on ξ.

We now fix a realization of A_{n-1} satisfying $\mathcal{E}_S^c \wedge \mathcal{E}_U^c$. In particular, this fixes y, y_0 satisfying the conditions in $\mathcal{E}_U^c$ and with $\|y\|_2 \ge 1/C''n^2$. Now, we use the independence of R and A_{n-1} and the fact that $\mathcal{E}$ implies

$$|Ry - v_1| \le t + nt^{1/2} \le 2nt^{1/2}.$$

Since

$$\mathcal{T}_\xi(y_0, L) < (\|\vec{p}\|_\infty + \eta/2)^n$$

and $\|y\|_2 \ge 1/C''n^2$, we find that the desired probability is bounded by

$$2LC''n^3t^{1/2} + L(\|\vec{p}\|_\infty + \eta/2)^n. \qquad \square$$

Now we are ready to conclude Proposition 2.6.8.

Proof of Proposition 2.6.8. A completely identical argument to the proof of Proposition 2.6.5 shows that for a sufficiently large constant K depending on ξ, and for $\Xi := \operatorname{supp}(\xi) \subseteq \mathbb{R}$,

$$\mathbb{P}[\mathcal{E}] \le p_0^{3n} + \sum_{a \in \mathcal{C}} \mathbb{P}[M^{(1)} = a]\mathbb{P}[\exists u' \in \mathbb{R}^{n-1} \colon \|M^{(-1)}u' + a\|_2 \le 2t],$$

where $M^{(1)}$ denotes the first column of M_n, $M^{(-1)}$ denotes the $n \times (n-1)$ matrix formed by excluding this column, and

$$\mathcal{C} = \{a \in \Xi^n : \exists \lambda \in \mathbb{R} \text{ with } \|a - \lambda 1_n\|_2 \le 8K\delta'\sqrt{n}\}.$$

We want to bound the contribution of the sum on the right hand side.

Case I: If $a = 0$, $\mathbb{P}[M^{(1)} = a] = p_0^n$.

Case II: If $a \neq 0$, then $\|a\|_2 \ge h_\xi > 0$. Hence, by Lemma 2.6.9, there is a constant $C > 0$ depending on ξ, η such that

$$\mathbb{P}[\exists u' \in \mathbb{R}^{n-1} \colon \|M^{(-1)}u' + a\|_2 \le 2t] \le Cn^3\sqrt{t} + (\|\vec{p}\|_\infty + \eta/2)^n.$$

Moreover, a similar (but easier) argument as in the proof of Proposition 2.6.5 shows that

$$\mathbb{P}[M^{(1)} \in \mathcal{C}] \le \|\vec{p}\|_\infty^n \exp(c_{\xi,\delta'}n),$$

where $c_{\xi,\delta'}$ goes to 0 as δ' goes to 0.

Hence, we see that the contribution to the sum from this case is bounded by

$$\|\vec{p}\|_\infty^n \exp(c_{\xi,\delta'} n) \cdot \left(Cn^3\sqrt{t} + (\|\vec{p}\|_\infty + \eta/2)^n \right).$$

By choosing δ' sufficiently small depending on ξ and η, and using $\|\vec{p}\|_\infty^2 \le \|\vec{p}\|_2^2 - c_\xi$ for some $c_\xi > 0$, we see as before (using the AM-GM inequality) that the above quantity is at most

$$t\exp(-\eta' n) + (\|\vec{p}\|_2^2 - \eta')^n$$

for a sufficiently small η' depending on ξ and η. This completes the proof. □

The proof of Theorem 2.1.10 is now immediate.

Proof of Theorem 2.1.10. First, assume that ξ is not a shift of an origin-symmetric random variable. We choose δ' small enough so that the conclusions of Propositions 2.6.5 and 2.6.8 are satisfied. By the union bound, this shows that the contribution of $\mathrm{Elem}(\delta')$ to the probability is at most

$$n\mathbb{P}[\mathcal{E}_{e_1}] + \binom{n}{2}\mathbb{P}[\mathcal{E}_{e_1-e_2}] + (t + \|\vec{p}\|_2^{2n})e^{-\eta n},$$

for a sufficiently small $\eta > 0$ depending on ξ, and for all sufficiently large n depending on ξ. Now, we can conclude using Corollary 2.6.4.

Next, if ξ is a nonzero shift of an origin-symmetric random variable, we do the same, except we require Propositions 2.6.5, 2.6.7, and 2.6.8 and then conclude with Corollary 2.6.4.

Finally, we consider the case when ξ is an origin-symmetric random variable. As before, we begin by using Propositions 2.6.5 and 2.6.8. The only thing to note is that, by the symmetry of ξ about the origin, for all $i \neq j$, $\mathbb{P}[\mathcal{E}_{e_i-e_j}] = \mathbb{P}[\mathcal{E}_{e_i+e_j}]$. Hence, by the union bound, the contribution of $\mathrm{Elem}'(\delta')$ to the probability is at most

$$n\mathbb{P}[\mathcal{E}_{e_1}] + \binom{n}{2}(\mathbb{P}[\mathcal{E}_{e_1-e_2}] + \mathbb{P}[\mathcal{E}_{e_1+e_2}]) + (t + \|\vec{p}\|_2^{2n})e^{-\eta n}.$$

Now, we can conclude using Corollary 2.6.4. □

2.7 Deduction of Theorems 2.1.2 and 2.1.8

Given the results in Sections 2.4 and 2.6, the deduction of Theorems 2.1.2 and 2.1.8 is immediate. Fix a discrete distribution ξ, and let $\delta, \rho, \eta, n_0 > 0$ be parameters depending on ξ coming from Theorem 2.1.10. Then, for the proof of Theorem 2.1.8, let $\epsilon > 0$ be as in the statement of the theorem (it suffices to assume that $\epsilon < 1$), and for the proof of Theorem 2.1.2, let $\epsilon > 0$ be such that

$$\exp(2\epsilon - H(\vec{p})) < \|\vec{p}\|_2^2,$$

which is possible since, by the weighted AM-GM inequality, we have

$$\exp(-H(\vec{p})) = \prod_i p_i^{p_i} \le \sum_i p_i^2 = \|\vec{p}\|_2^2,$$

and equality holds if and only if ξ is uniform on its support.

Let $C = C_{2.4.1}(\xi, \delta, \rho, \epsilon/2)$. By taking $C_{\xi,\epsilon}$ in Theorem 2.1.8 and C_ξ in Theorem 2.1.2 to be at least C, we may restrict our attention to $0 \le t \le 1/C$ (since for $t \ge 1/C$, the right-hand sides of Theorems 2.1.2 and 2.1.8 are at least 1). By Theorem 2.1.10 and Theorem 2.4.1, for all $0 \le t \le 1/C$, we have

$$\begin{aligned}\mathbb{P}[s_n(M_n(\xi)) \le t/\sqrt{n}] &\le \mathbb{P}\left[\inf_{x \in \mathrm{Cons}(\delta,\rho)} \|M_n(\xi)x\|_2 \le t/\sqrt{n}\right] + \mathbb{P}\left[\inf_{y \in \mathrm{Cons}(\delta,\rho)} \|yM_n(\xi)\|_2 \le Ct\right] \\ &\quad + \mathbb{P}\left[\inf_{x \in \mathbb{S}^{n-1} \setminus \mathrm{Cons}(\delta,\rho)} \|M_n(\xi)x\|_2 \le t/\sqrt{n} \wedge \inf_{y \in \mathrm{Cons}(\delta,\rho)} \|yM_n(\xi)\|_2 > Ct\right] \\ &\le 2n\mathbb{P}[\mathcal{E}_{e_1}] + (n^2 - n)(\mathbb{P}[\mathcal{E}_{e_1 - e_2}] + \mathbb{P}[\mathcal{E}_{e_1 + e_2}]) + 2(Ct + \|\vec{p}\|_2^{2n})e^{-\eta n} \\ &\quad + Ct + \exp((\epsilon/2 - H(\vec{p}))n)\end{aligned}$$

for all sufficiently large n. Here, we have used that $M_n(\xi)$ and $M_n(\xi)^\intercal$ have the same distribution.

For Theorem 2.1.2, we are done by our choice of ϵ.

For Theorem 2.1.8, we note that by Cauchy–Schwarz (as in Proposition 2.6.2), $\mathbb{P}[\mathcal{E}_{e_1+e_2}] \le \mathbb{P}[\mathcal{E}_{e_1-e_2}]$ and recall from above that $\exp(-H(\vec{p})) \le \|\vec{p}\|_2^2$. Using this, we can bound the right hand side of the above computation by

$$2Ct + 2n\mathbb{P}[\mathcal{E}_{e_1}] + 2n^2 \exp(\epsilon n/2)\mathbb{P}[\mathcal{E}_{e_1-e_2}].$$

The desired conclusion follows since $2n^2 \exp(\epsilon n/2) \le (1+\epsilon)^n$ for all $\epsilon < 1$ and n sufficiently large.

2.8 Singularity of random combinatorial matrices

In this section, we discuss the proof of Theorem 2.1.11. Given Theorem 2.4.1, the only ingredient we need is the following estimate for invertibility on almost constant vectors.

Proposition 2.8.1. *For any $\epsilon > 0$, there exist δ, ρ, c, n_0 depending on ϵ such that for all $n \ge n_0$,*

$$\mathbb{P}\left[\inf_{x \in \mathrm{Cons}(\delta,\rho)} \|Q_n x\|_2 \le c\sqrt{n} \vee \inf_{y \in \mathrm{Cons}(\delta,\rho)} \|yQ_n\|_2 \le c\sqrt{n}\right] \le \left(\frac{1}{2} + \epsilon\right)^n.$$

We begin with the easier case of $\|yQ_n\|_2$.

Lemma 2.8.2. *For any $\epsilon > 0$, there exists c, n_0 depending on ϵ such that for all $n \geq n_0$ and for any $y \in \mathbb{S}^{n-1}$,*

$$\mathbb{P}[\|yQ_n\|_2 \leq c\sqrt{n}] \leq (1/2 + \epsilon)^n.$$

Proof. Without loss of generality, we may assume that $|y_1| \geq \cdots \geq |y_n|$. We divide the proof into two cases depending on $|y_1|$. Let $\delta > 0$ be a constant to be chosen at the end of the proof.

Case I: $|y_1| < \delta$. Note that any entry in the first $n/4$ columns of Q_n, conditioned on all the remaining entries in the first $n/4$ columns of Q_n, is distributed as $\mathrm{Ber}(p)$ for some $p \in [1/3, 2/3]$. Moreover, by Lemma 2.2.4, it follows that for independent random variables $\xi_1, \ldots, \xi_n$, where $\xi_i \sim \mathrm{Ber}(p_i)$ for some $p_i \in [1/3, 2/3]$,

$$\mathcal{L}(y_1\xi_1 + \cdots + y_n\xi_n, \delta) \leq 3C_{2.2.4}\delta.$$

Therefore, a slight conditional generalization of the second part of Lemma 2.3.15 (which has the same proof) shows that

$$\mathbb{P}[\|yQ_n\|_2 \leq \delta\sqrt{n/8}] \leq (20C_{2.2.4}\delta)^{n/8} \leq (1/4)^n,$$

provided that δ is chosen sufficiently small depending on $C_{2.2.4}$.

Case II: $|y_1| \geq \delta$. Let $R_1, \ldots, R_n$ denote the rows of Q_n. Then,

$$\begin{aligned}\mathbb{P}[\|yQ_n\|_2 \leq \delta c\sqrt{n}] &\leq \sup_{R_2,\ldots,R_n} \mathbb{P}[\|y_1R_1 + y_2R_2 + \cdots + y_nR_n\|_2 \leq \delta c\sqrt{n} | R_2, \ldots, R_n] \\ &\leq \sup_{v \in \mathbb{R}^n} \mathbb{P}[\|R_1 - v\|_2 \leq c\sqrt{n}] \leq \binom{n}{n/2}^{-1}\binom{n}{2c^2n} \leq \left(\frac{1}{2} + \epsilon\right)^n,\end{aligned}$$

provided that $c > 0$ is chosen to be sufficiently small depending on $\epsilon > 0$. This completes the proof. □

Next, we deal with the harder case of $\|Q_nx\|_2$. We will need the following analogue of Lemma 2.5.1.

Lemma 2.8.3. *For any $\epsilon \in (0, 1/8)$, there exist $\theta = \theta(\epsilon) > 0$ and n_0 depending on ϵ for which the following holds. For all $n \geq n_0$ and for all $x \in \mathbb{S}^{n-1}$ such that $|\langle x, 1_n/\sqrt{n}\rangle| \leq 1/2$, we have*

$$\mathcal{L}(q \cdot x, \theta) \leq 1/2 + \epsilon,$$

where q is distributed uniformly on $\{0,1\}^n_{n/2}$.

Proof. Without loss of generality, we may assume that $|x_1| \geq \cdots \geq |x_n|$. Again, we divide the proof into two cases depending on $|x_1|$. Let $\delta > 0$ be a constant to be chosen at the end of the proof.

Case I: $|x_1| < \delta$. Let $\mu := \mathbb{E}[q \cdot x]$ and $\sigma^2 := \mathrm{Var}(q \cdot x)$. Since $\langle x, 1_n\rangle \leq \sqrt{n}/2$, a direct computation shows that $\sigma^2 \geq 3/16$. Moreover, a quantitative combinatorial central limit theorem due to Bolthausen [10] shows that the L^∞ distance between the cumulative

distribution function of $(q \cdot x - \mu)/\sigma$ and that of the standard Gaussian is at most $C\delta$, where C is an absolute constant. Hence, for all δ sufficiently small, we have $\mathcal{L}(q \cdot x, \delta) \leq 1/4$ whenever $|x_1| < \delta$.

Case II: $|x_1| \geq \delta$. Let $\mathcal{G}$ denote the event (depending on q) that

$$(n - \epsilon^2 n - 1)^{-1} \sum_{i=2}^{n-\epsilon^2 n} q_i \in [1/2 - \epsilon^4/2, 1/2 + \epsilon^4/2].$$

Then for all sufficiently large n, we have

$$\begin{aligned}\sup_{r \in \mathbb{R}} \mathbb{P}[|q \cdot x - r| \leq \theta] &\leq \sup_{r \in \mathbb{R}} \mathbb{P}[|q \cdot x - r| \leq \theta \wedge \mathcal{G}] + \mathbb{P}[\mathcal{G}^c] \\ &\leq \sup_{r \in \mathbb{R}} \mathbb{P}[|q \cdot x - r| \leq \theta \wedge \mathcal{G}] + 2\exp(-\epsilon^8 n/128),\end{aligned}$$

where the final inequality is by a standard large deviation estimate. It remains to control $\mathbb{P}[|q \cdot x - r| \leq \theta \wedge \mathcal{G}]$. For this, fix any realization $q' := (q_2, \ldots, q_{n-\epsilon^2 n})$ satisfying $\mathcal{G}$. Note that

$$1/2 - 2\epsilon^2 \leq \inf_{q' \in \mathcal{G}} \mathbb{P}[q_1 = 0 \mid q'] \leq \sup_{q' \in \mathcal{G}} \mathbb{P}[q_1 = 0 \mid q'] \leq 1/2 + 2\epsilon^2.$$

Note also that, since $\sum_{i \geq n-\epsilon^2 n} x_i^2 \leq \epsilon^2$ (this uses $\|x\|_2 = 1$ and $|x_1| \geq |x_2| \geq \cdots \geq |x_n|$), it follows that

$$\sup_{q' \in \mathcal{G}, q_1} \mathrm{Var}\left[\sum_{i \geq n-\epsilon^2 n} q_i x_i \middle| q_1, q'\right] \leq \epsilon^2,$$

so that by Markov's inequality,

$$\sup_{q' \in \mathcal{G}, q_1} \mathbb{P}\left[\left|\sum_{i \geq n-\epsilon^2 n} q_i x_i - f(q', q_1)\right| \geq \frac{\delta}{8} \middle| q', q_1\right] \leq \frac{32\epsilon^2}{\delta^2},$$

where $f(q', q_1)$ denotes the mean of $\sum_{i \geq n-\epsilon^2 n} q_i x_i$ conditioned on q', q_1. Finally, since $|x_1| \geq \delta$, and since

$$\sup_{q' \in \mathcal{G}} |f(q', 0) - f(q', 1)| \leq |x_{n-\epsilon^2 n}| \leq 2/\sqrt{n},$$

it follows by putting everything together that

$$\sup_{r \in \mathbb{R}} \mathbb{P}[|q \cdot x - r| \leq \theta \wedge \mathcal{G}] \leq \sup_{r \in \mathbb{R}} \sup_{q' \in \mathcal{G}} \mathbb{P}[|q \cdot x - r| \leq \theta \mid q'] \leq 1/2 + 2\epsilon^2 + 64\epsilon^2/\delta^2,$$

provided that θ is chosen sufficiently small compared to δ, and n is sufficiently large. Indeed, the two values of $q_1 x_1$ (for $q_1 = 1$ and $q_1 = 0$) differ by $|x_1|$, which is at least δ by assumption, and the above discussion shows that given q' and q_1, $\sum_{i \geq n-\epsilon^2 n} q_i x_i$ is localized in an interval of length $\delta/2 + 2/\sqrt{n}$ except with probability at most $32\epsilon^2/\delta^2$. Since δ is an absolute constant coming from **Case I**, this gives the desired conclusion for all sufficiently small ϵ, which completes the proof. □

Given the previous two lemmas, the proof of Proposition 2.8.1 is by now standard.

Proof of Proposition 2.8.1. The estimate for $\inf_{y \in \mathrm{Cons}(\delta,\rho)} \|yQ_n\|_2 \leq c\sqrt{n}$ (for a suitable choice of δ, ρ, c) follows immediately by combining Lemma 2.8.2 with the low metric entropy of $\mathrm{Cons}(\delta, \rho)$. To exploit the latter, one could either use a randomized rounding based net construction due to Livshyts [72, Theorem 4], which uses that $\|Q_n\|_{\mathrm{HS}}^2 \leq n^2$, or one could use the fact that there exists a constant K such that with probability at least $1-4^{-n}$, all singular values of Q_n except for the top singular value are at most $K\sqrt{n}$ (see [104, Proposition 2.8]).

For the estimate on $\inf_{x \in \mathrm{Cons}(\delta,\rho)} \|Q_n x\|_2 \leq c\sqrt{n}$ (for suitable δ, ρ, c), we begin by using the fact [104, Proposition 2.8] noted above that there exists a constant $K > 0$ such that with probability at least $1-4^{-n}$, the operator norm of Q_n restricted to the subspace perpendicular to 1_n is at most $K\sqrt{n}$. Let us denote this event by $\mathcal{E}_K$. Then, on $\mathcal{E}_K$, for any $x \in \mathbb{S}^{n-1}$ such that $\langle x, 1_n/\sqrt{n} \rangle \geq 1/2$, we have $\|Q_n x\|_2 \geq n/4 - K\sqrt{n}$. Hence, on the event $\mathcal{E}_K$, and for all n sufficiently large, it suffices to consider the infimum over those vectors $x \in \mathrm{Cons}(\delta, \rho)$ which also satisfy $\langle x, 1_n/\sqrt{n} \rangle < 1/2$. For this, we can use Lemma 2.2.6 followed by the tensorization lemma Lemma 2.3.15, and then exploit the low metric entropy of $\mathrm{Cons}(\delta, \rho)$ as above. We leave the details to the interested reader. □

Finally, given Proposition 2.8.1, the proof of Theorem 2.1.11 follows exactly as in Theorem 2.1.8.

Chapter 3

Threshold for Steiner triple systems

3.1 Introduction

A foundational result in the theory of random graphs, due to Erdős and Rényi [17], is that the threshold for the appearance of perfect matchings in $\mathbb{G}(n,p)$ is $\log n/n$. It is natural to seek higher-dimensional analogues of this result. As the simplest case, consider perfect matchings in 3-uniform hypergraphs. These are collections of 3-edges, or *triangles*, such that each vertex is contained in exactly one triangle. Let $\mathbb{G}^{(3)}(n,p)$ denote the binomial random n-vertex hypergraph in which each triangle is present with probability p. Determining the threshold for the appearance of perfect matchings in this model is the well-known "Shamir's problem" [91], which was resolved (up to a constant factor) in a seminal paper by Johansson, Kahn, and Vu [45] with sharp threshold and hitting time results obtained in later work of Kahn [47, 48].

Of course, there is more than one high-dimensional analogue to (graphical) perfect matchings. It is just as natural to consider (spanning) *Steiner triple systems* (i.e., triangle sets in which each *pair* of vertices is contained in exactly one triangle) and their appearance in $\mathbb{G}^{(3)}(n,p)$. Here, much less is known and until recently such results seemed out of reach. Indeed, even the log-asymptotics of the number of Steiner triple systems was a mystery until the work of Keevash [56] (which built on his earlier breakthrough establishing the existence of designs [54]). Furthermore, both of these problems can be viewed under the common umbrella of determining the threshold for the existence of designs relative to a random set, with Shamir's problem corresponding to $(r,1)$-designs and Steiner triple systems to $(3,2)$-designs.

Our main result is the determination of the threshold for Steiner triple systems up to a sub-polynomial factor, which is the first higher-dimensional generalization of Erdős–Rényi and Johansson–Kahn–Vu incorporating nontrivial designs.

Theorem 3.1.1. *Let $n \in \mathbb{N}$ satisfy $n \equiv 1,3 \pmod 6$. Let $\mathcal{H} \sim \mathbb{G}^{(3)}\big(n, \exp(C(\log n)^{3/4})/n\big)$, with $C > 0$ a sufficiently large constant. With high probability[1], $\mathcal{H}$ contains an order-n*

[1] We say that a sequence of events, parameterized by n, holds *with high probability* (*w.h.p.*) if the probabilities of their occurrence tend to 1.

Steiner triple system.

Remark 3.1.2. An order-n Steiner triple system is equivalent to a triangle-decomposition of the complete graph K_n. A graph has a triangle-decomposition only if it is *triangle-divisible*, i.e., its every degree is even and the number of edges is a multiple of 3. For K_n this is equivalent to the arithmetic condition $n \equiv 1, 3 \pmod 6$, demonstrating the necessity of this assumption. That Steiner triple systems indeed exist whenever n satisfies this condition is a famous classical theorem of Kirkman [58].

Note that $\exp((\log n)^{3/4}) = n^{o(1)}$. Thus, Theorem 3.1.1 implies that the threshold for the appearance of Steiner triple systems is bounded above by $n^{-1+o(1)}$. A corresponding lower bound is obtained by observing that if $\mathcal{H}$ contains a Steiner triple system then every edge of K_n is contained in at least one triangle of $\mathcal{H}$. A straightforward calculation (analogous to that for isolated vertices in $\mathbb{G}(n,p)$) reveals that the (sharp) threshold for this modified property is $2\log n/n = n^{-1+o(1)}$. Hence, our result establishes the threshold up to a subpolynomial factor.

A recent breakthrough relating thresholds and fractional expectation-thresholds by Frankston, Kahn, Narayanan, and Park [19] (and also a very recent breakthrough of Park and Pham [77] resolving the Kahn–Kalai conjecture), as a corollary, gave an alternate and substantially simpler proof of the result of Johansson, Kahn, and Vu. This proof hinges on the ability to determine the *fractional expectation-threshold*, which can be viewed as a linear program whose variables are all sets of hyperedges.

The sheer size of such programs suggests that determining the expectation-threshold or fractional expectation threshold is a difficult task in general. Instead, in previous applications, the uniform distribution on some desired class of objects is used to witness a lower bound, via a parameter known as "spread". In this light, the application of [19] to Shamir's problem relies crucially on the fact that the enumeration of $(r,1)$-designs (as well as extensions of partial $(r,1)$-designs) is straightforward. However, determining the fractional expectation-threshold of Steiner triple systems is nontrivial; as noted in [19, Section 8.D] the error terms in the enumerative results of Keevash [56] are too large to prove that the uniform distribution on Steiner triple systems has sufficiently small spread. Therefore the authors of [19] raise applying these methods to combinatorial designs as an interesting open problem. In this paper, we circumvent this difficulty by *constructing* a (non-uniform) distribution on Steiner triple systems with small spread. We expect this approach to have further applications.

Steiner triple systems are closely related to *Latin squares* (i.e., $n \times n$ matrices in which every row and column is a permutation of $\{1, 2, \dots, n\}$). The latter are naturally equivalent to (labeled) triangle-decompositions of $K_{n,n,n}$, with the three vertex parts corresponding to rows, columns, and symbols. With a few adjustments the proof of Theorem 3.1.1 yields a similar threshold result for Latin squares.

We use the following terminology: Let $S\colon [n]^2 \to 2^{[n]}$ be a function that assigns, to each cell in an $n \times n$ grid, a set of symbols from $\{1, 2, \dots, n\}$. Say that S *supports* an order-n Latin square L if for every $i, j \in [n]$ there holds $L(i,j) \in S(i,j)$. For $p \in [0,1]$ let $\mathbb{M}(n,p)$

be the distribution on functions $S\colon [n]^2 \to 2^{[n]}$ where for every $i,j,k \in [n]$, the symbol k is included in $S(i,j)$ with probability p independent of all other choices.

Theorem 3.1.3. *Let $n \in \mathbb{N}$ and let $\mathcal{S} \sim \mathbb{M}(n, \exp(C(\log n)^{3/4})/n)$, with $C > 0$ a sufficiently large constant. W.h.p. $\mathcal{S}$ supports a Latin square.*

Prior to this work the only known upper bounds on the thresholds in Theorems 3.1.1 and 3.1.3 were quite far from the $n^{o(1)-1}$ proved here. For Latin squares, Andrén, Casselgren, and Öhman [3] proved that there exists a constant $p < 1$ such that w.h.p. $\mathbb{M}(n,p)$ supports a Latin square. For Steiner triple systems, Simkin [94] observed that Keevash's method of randomized algebraic construction [54] can be used to show that for a sufficiently small $\varepsilon > 0$, w.h.p. $\mathbb{G}^{(3)}(n, n^{-\varepsilon})$ contains a Steiner triple system.

Before moving on to proofs we mention an interesting consequence of Theorem 3.1.1: the threshold for the appearance of Steiner triple systems is sharp, in the following sense.

Corollary 3.1.4. *There is a function $p_{\mathrm{STS}}(n)$ such that for all $\varepsilon > 0$, when $n \equiv 1,3 \pmod 6$ w.h.p. $\mathbb{G}^{(3)}(n, (1+\varepsilon)p_{\mathrm{STS}}(n))$ contains a Steiner triple system but w.h.p. $\mathbb{G}^{(3)}(n, (1-\varepsilon)p_{\mathrm{STS}}(n))$ does not.*

This is surprising, since we have not determined what the threshold actually is. Nevertheless, Corollary 3.1.4 follows from Friedgut's characterization of sharp thresholds [21, 22]. Indeed, for $n \equiv 1,3 \pmod 6$, let $p_{\mathrm{STS}}(n)$ be the threshold for containing Steiner triple systems (i.e., $\mathbb{G}^{(3)}(n, p_{\mathrm{STS}}(n))$ contains a Steiner triple system with probability $1/2$). Theorem 3.1.1 tells us that $p_{\mathrm{STS}}(n) \leq n^{o(1)-1}$. On the other hand, by considering the disappearance of vertex pairs not contained in a triangle, we concluded that $p_{\mathrm{STS}}(n) = \Omega(n^{-1}\log n)$. Hence, $p_{\mathrm{STS}}(n) \neq \Theta(n^{\alpha})$ for any $\alpha \in \mathbb{R}$. However, a consequence of [22, Theorem 2.1] and the remarks immediately after is that in our setting (sampling hypergraphs), coarse thresholds are limited to the form $\Theta(n^{\alpha})$, implying that our threshold is sharp.

Regarding Latin squares, the threshold is sharp for essentially the same reasons. Although triangle-decompositions of $K_{n,n,n}$ are not invariant under vertex permutations so that [22] does not directly apply, similar methods can be used to deduce a sharp threshold [20].

3.1.1 Techniques for bounding thresholds

Finding thresholds for spanning structures in random graphs and hypergraphs has played a major role in the field since its inception. Prominent examples include thresholds for containing a spanning tree (which is equivalent to connectivity) [15], a perfect matching [16, 17], a Hamilton cycle [79], a triangle-factor [45], and a given bounded-degree spanning tree [75].

Lower bounds on the thresholds for each of these properties (and many others) can be obtained in more or less the same way: fixing a vertex v, it is contained in a spanning tree or perfect matching only if its degree is at least 1. Similarly, it is contained in a Hamilton cycle only if its degree is at least 2. Finally, it is contained in a triangle-factor only if it is contained in at least one triangle. By computing the expectation of each of these random variables and

applying Markov's inequality we obtain a lower bound of $\Omega(n^{-1})$ for connectivity, perfect matchings, and Hamiltonicity, and a lower bound of $\Omega(n^{-2/3})$ for existence of a triangle-factor. Avoiding formal definitions (which can be found in [49]), the maximal lower bound obtained by similar arguments is known as the *expectation-threshold* for the property.

Surprisingly, these easily-obtained lower bounds turn out to be within a logarithmic factor of the true thresholds. However, in sharp contrast to the lower bounds, the original proofs of the corresponding upper bounds are problem-specific. This disparity (and the associated difficulty of obtaining thresholds for some properties, as in Shamir's problem) motivated a family of beautiful conjectures of Kahn and Kalai [49]. The main conjecture is that the threshold for a monotone property is always within a logarithmic factor of its expectation-threshold.

In a very recent breakthrough, Park and Pham [77] gave an ingenious proof of the Kahn–Kalai conjecture. However, for our application (and many others), a fractional version of the conjecture, due to Talagrand [99], suffices. The so-called *fractional expectation-threshold vs. threshold conjecture* was proved by Frankston, Kahn, Narayanan, and Park [19] in an earlier breakthrough. These works are related to yet another, yet earlier, breakthrough: the advance on the *sunflower conjecture* due to Alweiss, Lovett, Wu, and Zhang [2]. For our purposes it suffices to consider a corollary of these results, for which we need the next definition.

Definition 3.1.5. Consider a finite ground set Z and fix a nonempty collection of subsets $\mathcal{H} \subseteq 2^Z$. Let μ be a probability measure on $\mathcal{H}$. For $q > 0$ we say that μ is *q-spread* if for every set $S \subseteq Z$:

$$\mu(\{A \in \mathcal{H} \colon S \subseteq A\}) \leq q^{|S|}.$$

The next theorem, relating spread measures and thresholds, is due to Frankston, Kahn, Narayanan, and Park [19]. We have slightly tailored it to our setting.

Theorem 3.1.6 (From [19, Theorem 1.6])**.** *There exists a constant $C = C_{3.1.6} > 0$ such that the following holds. Consider a non-empty ground set Z and fix a nonempty collection of subsets $\mathcal{H} \subseteq 2^Z$. Suppose that there exists a q-spread probability measure on $\mathcal{H}$. Then a random binomial subset of Z where each element is sampled with probability $\min(Cq \log |Z|, 1)$ contains an element of $\mathcal{H}$ as a subset with probability at least $3/4$.*

For many graph families, including those mentioned above, the spread of the uniform distribution is easily seen to match the lower bounds on the threshold. Thus, Theorem 3.1.6 immediately determines these thresholds up to a logarithmic factor. However, Theorem 3.1.6 does not immediately imply any bound at all on the threshold for $\mathbb{G}^{(3)}(n,p)$ to contain a Steiner triple system. The issue is that currently, our understanding of the uniform distribution on Steiner triple systems is rather poor. Indeed, as remarked in [19, Section 8.D], even the uncertainty in the *number* of order-n Steiner triple systems is large enough that it precludes any useful bounds on the spread.

Although the uniform distribution is the most natural one with which to apply Theorem 3.1.6, this is certainly not required. We prove Theorem 3.1.1 by designing a distribution

on Steiner triple systems that is $n^{o(1)-1}$-spread *by construction*, and then applying Theorem 3.1.6.

3.1.2 Spread distributions and iterative absorption

The $n^{o(1)-1}$-spread distribution used to prove Theorem 3.1.1 is defined implicitly by a randomized algorithm to construct Steiner triple systems. In order to outline this algorithm we briefly recount some recent breakthroughs in design theory.

We begin with the *triangle removal process*, which is closely related to the influential *Rödl nibble* [81]. This is the following random greedy algorithm to construct a partial Steiner triple system: Beginning with $G = K_n$ repeatedly and for as long as possible delete, uniformly at random, a triangle from G and add it to a growing collection of triples. Spencer [95] and Rödl and Thoma [82] independently proved that w.h.p. this process terminates when G has only $o(n^2)$ edges. Equivalently, this method produces an approximate Steiner triple system.

It is straightforward to adapt the triangle removal process so that the triangle set it produces has spread $O(n^{o(1)-1})$. Perhaps the simplest way is to first restrict the available triangles to a prescribed binomial random subset of density (say) $(\log n)^2/n$, and then show that the process is still likely to construct an approximate Steiner triple system.

Given the success of the triangle removal process, a natural way to construct an exact Steiner triple system is to find a triangle-decomposition of the edges remaining at the end of the process[2]. This is essentially what Keevash does with his breakthrough method of *randomized algebraic constructions* [54, 56]. Moreover, Keevash's method is incredibly powerful in that it proves the existence of designs with arbitrary parameters, which was a central question in combinatorics since the nineteenth century. Unfortunately, the algebraic component of Keevash's construction has rather poor spread, and so is unsuitable for our application.

An alternative to Keevash's method is *iterative absorption*, developed by Kühn, Osthus, and collaborators [5, 59, 63]. This method gave an alternate proof of the existence of designs [24] using purely probabilistic and combinatorial methods. In this paper we mostly follow its specialization by Barber, Glock, Kühn, Lo, Montgomery, and Osthus [4] to triangle-decompositions.

A key insight is that by using a modified version of the triangle removal process, the uncovered edges at the end of the process can be "localized" to a small vertex set $U_1 \subseteq V(K_n)$. That is, after fixing $U_1 \subseteq V(K_n)$ (satisfying, say, $|U_1| \approx \varepsilon n$ for a small $\varepsilon > 0$), a multi-stage randomized "cover-down" procedure can produce a partial Steiner triple system that covers all edges in K_n not spanned by U_1. Furthermore, the graph of uncovered edges in U_1 is nearly-complete. By repeating this process the uncovered edges can be iteratively localized to sets $U_1 \supseteq U_2 \supseteq \cdots \supseteq U_\ell = X$, where X may be quite small. Since the goal is to construct an exact Steiner triple system, the iterative process is preceded by setting aside an "absorber" for X. This is a graph $H \subseteq K_n$ with the property that for any possible remainder graph L on X, the graph $H \cup L$ admits a triangle-decomposition.

[2]Strictly speaking, one must stop the triangle removal process before its natural termination, at which point there are no triangles in G by definition.

Iterative absorption, as outlined in [4], does not itself produce a distribution with sufficient spread. The issue is that for each U_i, the triangle set constructed on U_i forms a nearly complete graph and contains $\Omega(|U_i|^2)$ triangles. Thus, the best spread one can hope for in this method is $\Omega(|X|^{-1})$, which is far larger than $1/n$ since X must be small in order to construct the absorber H.

As a remedy, our algorithm combines iterative absorption with a bootstrapping scheme that iteratively constructs distributions with better and better spread. Concretely, let $P(\eta)$ be the proposition that for every sufficiently large, near-complete, and triangle-divisible graph G there exists an n^{η}/n-spread distribution over triangle-decompositions of G. Note that $P(0)$ would imply Theorem 3.1.1, since we may take $G = K_n$. We remark that here, and in the remainder of this outline, our goal is to provide a clear and concise summary of our argument. Thus, we take some leeway and are not as precise with some statements as we will be in the proof. For example, proposition $P(\eta)$ is slightly different from its analogue, Theorem 3.3.1(η).

We define the sequence $\eta_k = 2/(2+k)$, and we will inductively show that $P(\eta_k)$ holds. Since $\eta_k \to 0$, this implies that there exists an $n^{o(1)-1}$-spread distribution of Steiner triple systems. The fact that $P(\eta_0 = 1)$ holds is itself a non-trivial fact; this follows implicitly from [56] and explicitly from [4].

Now suppose that $P(\eta_{k-1})$ holds. We wish to show that $P(\eta_k)$ holds as well. Let G be a large, near-complete, triangle-divisible graph. We wish to construct an $n^{\eta_k - 1}$-spread distribution on the triangle-decompositions of G. We proceed as follows: We first set aside a vertex set $X \subseteq V(G)$ of size approximately $n^{1-\eta_k}$. Next, we set aside a small, random, absorbing triangle set $\mathcal{H}$ in G. The triangle set $\mathcal{H}$ resembles a binomial random triangle set of density $1/n$, ensuring that it does not negatively impact the spread. As we will explain momentarily, this absorber serves a different purpose than the absorber in [4].

After setting aside the absorber, we use iterative absorption to find a triangle set $\mathcal{S}$ that covers all edges in $G \setminus (E(\mathcal{H}) \cup G[X])$. Let $L \subseteq G[X]$ denote the graph of uncovered edges. We remark that since $\mathcal{S}$ is constructed by processes resembling the triangle removal process, it is straightforward to ensure that the spread of $\mathcal{S}$ is approximately $|X|^{-1} \approx n^{\eta_k - 1}$.

Now, by assumption, there exists an $|X|^{\eta_{k-1}-1}$-spread distribution over the triangle-decompositions of L. Let $\mathcal{L}$ be a triangle-decomposition sampled according to this distribution. Observe that $\mathcal{S} \cup \mathcal{L}$ is a triangle-decomposition of G. However, due to the presence of $\mathcal{L}$, its spread is $|X|^{\eta_{k-1}-1} \gg n^{\eta_k - 1}$. This is where the absorber comes in: its purpose is to "spread" the probability mass of the triangles induced by X over the rest of the graph. This reduces the overall spread of the Steiner triple system. Specifically, $\mathcal{H}$ has the property that for every triangle T in X, there exist many configurations $\mathcal{H}' \subseteq \mathcal{H}$ such that $\mathcal{H}' \cup \{T\}$ can be replaced by a triangle set that does not use T. Furthermore, given the triangle-decomposition $\mathcal{L}$ of L, it is possible to choose a set of such configurations $\{\mathcal{H}'_T\}_{T \in \mathcal{L}}$ that are mutually disjoint. Thus, all the triangles used in $\mathcal{L}$ can be replaced by a set of triangles that are not spanned by X. Finally, if these configurations are randomly chosen in an appropriate way, then the spread of the resulting Steiner triple system is less than $n^{\eta_k - 1}$. Since there exists an $n^{\eta_k - 1}$-spread distribution of triangle-decompositions of G, the proposition $P(\eta_k)$

holds.

Finally, the modification for Latin squares is straightforward; we detail the minor changes in Section 3.5.

3.1.3 Absorbers as spread boosters

We stress that unlike traditional uses of absorbers, we cannot intentionally plant specific absorbers for each triangle that might appear in $\mathcal{L}$ (though it may seem like there is available space within K_n for such constructions). The reason is that any specific finite absorber w.h.p. will not appear at all in $\mathbb{G}(n, n^{\theta-1})$ for $\theta > 0$ sufficiently small. In fact, as far as algorithmically finding absorbers in $\mathcal{H}$ goes, there seems to be a "barrier" around density $1/\sqrt{n}$ which is polynomially far from the conjectured threshold of $O(\log n/n)$.

This is different to other threshold problems. For example, absorber-based algorithms were used to bound the threshold for the appearance of the square of a Hamilton cycle in $\mathbb{G}(n,p)$ up to a subpolynomial factor [62]. (Eventually, the true threshold was recovered by Kahn, Narayanan, and Park [51] using ideas related to [19] and without use of absorbers.) In contrast, we use the richness of possible configurations within $\mathcal{H}$, which is essentially spread by definition, to show that there is some way to absorb $\mathcal{L}$ in a spread manner. However, the absorbers are not necessarily themselves contained in $\mathcal{H}$. Thus one can think of our absorber as a sparsification template that facilitates boosting the spread, after which the non-algorithmic [19] is applied. To our knowledge this is the first such use of absorbers.

3.1.4 Further directions

We briefly remark on a few natural questions arising from this work. First, the next two conjectures convey our intuition that the disappearance of uncovered edges probabilistically tells the whole story regarding the appearance of Steiner triple systems. The first conjecture locates a sharp threshold at $2\log n/n$, while the second is the corresponding hitting time statement.

Conjecture 3.1.7. *Let $n \in \mathbb{N}$ satisfy $n \equiv 1,3 \pmod 6$ and fix $\epsilon > 0$. If $\mathcal{H} \sim \mathcal{G}^{(3)}(n, (2+\epsilon)\log n/n)$, then with high probability $\mathcal{H}$ contains an order-n Steiner triple system.*

Conjecture 3.1.8. *Let $n \in \mathbb{N}$ satisfy $n \equiv 1,3 \pmod 6$ and let $T_1, T_2, \dots$ be a uniformly random ordering of $\binom{[n]}{3}$. W.h.p. the first prefix $T_1, \dots, T_k$ that covers each 2-edge at least once contains a Steiner triple system.*

We note that a threshold of $O(\log n/n)$ would follow from the existence of an $O(n^{-1})$-spread distribution, while Conjecture 3.1.7 seems to require ideas beyond those in [19][3].

[3]Since this paper was released the bounds on the threshold have been improved, first to $O((\log n)^2/n)$ [52] and then to $O(\log n/n)$ [39, 55]. The latter is within a constant factor of the lower bound. As in this paper, the proofs rely in part on a cover-down procedure and construction of non-uniform spread measures; however, the implementations are substantially different.

It is also natural to ask for the threshold of more general Steiner systems in random hypergraphs. Since iterative absorption can famously construct designs with arbitrary parameters, we expect that some of the ideas in this paper might extend to this setting. However, to highlight just one potential difficulty, the current argument hinges on the ability to find absorbers that are sufficiently sparse so as not to contribute adversely to the spread. It is unclear whether this is possible for designs with other parameters.

Finally, we wonder whether there exists an efficient algorithm to *find* Steiner triple systems in $\mathbb{G}^{(3)}(n,p)$, with p sufficiently large for them to exist w.h.p. This question is relevant to Shamir's problem too: while the sharp threshold for the existence of perfect matchings in $\mathbb{G}^{(3)}(n,p)$ is known, it is not known whether a perfect matching can be found algorithmically. We note that a slight modification of the proof given for Theorem 3.3.1 (though it is not stated this way) allows one to construct the measure on Steiner triple systems which is well-spread in an algorithmic fashion. The difficulty lies in the black-box application of [19] to show that $\mathbb{G}^{(3)}(n,n^{-1+o(1)})$ therefore contains Steiner triple systems w.h.p.

3.1.5 Organization

The paper is organized as follows. At the end of this section we introduce some notation. Section 3.2 introduces basic concepts connected to triangle-decompositions and also collects useful probabilistic tools. In Section 3.3 we lay out the bootstrapping technique that is the heart of our argument. In Section 3.4 we adapt the framework of iterative absorption to the sparse random setting. Finally, as mentioned, in Section 3.5 we describe the modifications to our proof required to obtain Theorem 3.1.3.

3.1.6 Notation

For a graph G we write $V(G)$ for its vertex set, and G^c for its complement within that set. For $v \in V(G)$ we write $N_G(v)$ for its set of neighbors in G. If $X \subseteq V(G)$ then $G[X]$ is the induced subgraph on vertex set X.

If $\mathcal{H}$ is a 3-uniform hypergraph, we may refer to its 3-edges as triangles. We denote by $e(\mathcal{H})$ the number of triangles in $\mathcal{H}$, and we denote by $E(\mathcal{H})$ the graph of edges that are contained in a triangle of $\mathcal{H}$.

3.2 Preliminaries

We remind the reader of the following definitions.

Definition 3.2.1. A graph G is *triangle-divisible* if every vertex degree is even and the number of edges is a multiple of 3. A *triangle-decomposition* of a graph G is a collection of triangles in G such that every edge of G is contained in exactly one triangle.

Remark 3.2.2. It is easy to see that triangle-divisibility is a necessary but insufficient condition for G to admit a triangle decomposition. The main result of [56] (and also [4]) is that

if G is sufficiently large, dense, and typical (pseudorandom in an appropriate sense) then triangle-divisibility is sufficient for G to admit a triangle-decomposition.

We will repeatedly use the Chernoff bound for binomial and hypergeometric distributions (see for example [44, Theorems 2.1 and 2.10]) without further comment.

Lemma 3.2.3 (Chernoff bound)**.** *Let X be either:*

- *a sum of independent random variables, each of which take values in $\{0,1\}$, or*
- *hypergeometrically distributed (with any parameters).*

Then for any $\delta > 0$ we have

$$\mathbb{P}[X \leq (1-\delta)\mathbb{E}X] \leq \exp(-\delta^2\mathbb{E}X/2), \qquad \mathbb{P}[X \geq (1+\delta)\mathbb{E}X] \leq \exp(-\delta^2\mathbb{E}X/(2+\delta)).$$

Next we will require that if a sequence of random variables is stochastically dominated by a sequence of Bernoulli random variables it satisfies an identical set of tail bounds.

Lemma 3.2.4 ([80, Lemma 8])**.** *Let $X_1, \dots, X_n$ be $\{0,1\}$-valued random variables such that for all $i \in [n]$, we have that $\mathbb{P}[X_i = 1 | X_1, \dots, X_{i-1}] \leq p$ then $\mathbb{P}[\sum_{i\in[n]} X_i \geq t] \leq \mathbb{P}[\mathrm{Bin}(n,p) \geq t]$ for all $t \geq 0$.*

Finally we will need the symmetric form of the Lovász Local Lemma.

Lemma 3.2.5 ([1, Corollary 5.1.2])**.** *Let $A_1, A_2, \dots, A_n$ be events in a probability space such that each A_i is mutually independent of all but d other events. If $\mathbb{P}[A_i] \leq p$ for every $i \in [n]$ and $ep(d+1) \leq 1$ then $\mathbb{P}[\bigwedge_{i\in[n]} A_i^c] > 0$.*

3.3 Bootstrapping with Spread Families

In order to prove Theorem 3.1.1, we will iteratively prove the following result for all $\eta \in (0,1]$. We fix a constant $C_{3.3.1} > 0$, large enough for various inequalities we encounter later to hold.

Theorem 3.3.1 (Theorem(η))**.** *Fix a triangle-divisible graph G on n vertices with $\Delta(G^c) \leq n/\log n$ and $n \geq \exp(C_{3.3.1}/\eta^4)$. Let $\mathcal{H} \sim \mathbb{G}^{(3)}(n, n^\eta/n)$. With probability at least $1/2$ the collection $\mathcal{H}$ contains a triangle-decomposition of G.*

Let $\eta' = c(\log n)^{-1/4}$, with $c > 0$ a large constant. Note that Theorem 3.3.1(η') implies Theorem 3.1.1. Indeed, assuming Theorem 3.3.1(η'), if $\mathcal{H}_1, \mathcal{H}_2, \dots, \mathcal{H}_m$ are (say) $m = \log n$ independent samples of $\mathbb{G}^{(3)}(n, n^{\eta'}/n)$ then w.h.p. at least one of them contains a Steiner triple system. On the other hand $\bigcup_{i=1}^m \mathcal{H}_i$ is distributed as $\mathbb{G}^{(3)}(n,p)$ with $p = \exp(O((\log n)^{3/4}))/n$.

We note that Theorem 3.3.1($\eta = 1$) was proved by Gustavsson [34] and also follows immediately from the results of Keevash [56, Theorem 2.1] or Barber, Kühn, Lo, and Osthus

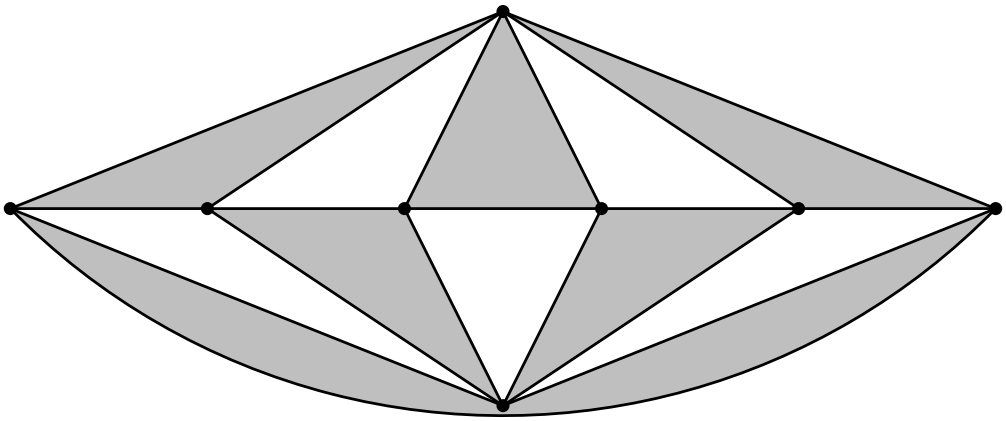

Figure 3.1: The absorber $\mathcal{F}_6$ consists of the shaded triangles, while the unshaded triangles (including the external triangle) comprise its absorber flip $\mathcal{F}_6^*$.

[5, Theorem 1.2]. This will serve as the base case for our results, and we will inductively show that the result is true for smaller and smaller η.

We will need the following result, which can be thought of as stating that a triangle-decomposition iterative absorption scheme that attempts to cover $G \subseteq K_n$ and ultimately has a leftover contained in a smaller set $X \subseteq V(K_n)$ can be performed using only a $|X|^{-1+o(1)}$-fraction of triangles of G. (This is essentially the limit for a pure iterative absorption framework as in [4, 24].) We defer its proof to Section 3.4. The remaining argument is independent of its justification.

Proposition 3.3.2. *There exists a constant $C = C_{3.3.2} > 0$ such that the following holds. Let $n \in \mathbb{N}$. Fix a subset X of $V(K_n)$ such that $|X| \in [C, n/(\log n)^3]$. Furthermore fix a triangle-divisible graph $G \subseteq K_n$ such that $\Delta(G^c) \leq n/\log n + n/(\log n)^2$ and $|X \setminus N_G(v)| \leq |X|(1/\log|X| - 1/(\log|X|)^2)$ for all $v \in V(G)$. Given a sample $\mathcal{H}' \sim \mathbb{G}^{(3)}(n, (\log|X|)^C/|X|))$, with probability at least $3/4$ there exists an edge-disjoint triangle set $\mathcal{H} \subseteq \mathcal{H}'$ such that $G^* := E(\mathcal{H})$ satisfies:*

1. *$G \setminus G[X] \subseteq G^*$ (i.e., $\mathcal{H}$ covers all edges of G outside of X),*
2. *$G^* \subseteq G$ (i.e., $\mathcal{H}$ consists of triangles in G), and*
3. *$\Delta((G^c \cup G^*)[X]) \leq |X|/\log|X|$ (the graph of uncovered edges in $G[X]$ is nearly complete).*

As outlined in the introduction, we describe a randomized construction of Steiner triple systems. It uses "probabilistic absorbers", whose role is to "spread out" the distribution of triangles that are too highly concentrated on a small vertex set. The absorbers are copies of the 3-uniform hypergraph $\mathcal{F}_{2m}$, defined as follows: Consider a cycle C_{2m} and add two vertices that are each connected with edges to the $2m$ vertices in the cycle. The resulting

graph naturally has a 2-colorable triangulation. Let $\mathcal{F}_{2m}$ be one of the color classes. Define its "absorber-flip" $\mathcal{F}_{2m}^*$ as the hypergraph on the same vertex set but with the opposite color class of triangles. For an illustration, see Figure 3.1.

The next lemma allows us to find such absorbers in random linear hypergraphs. Crucially, it requires only that $\mathcal{F}_{2m}$ be present. It uses the following construction, akin to the Rödl nibble: Let G be a graph with a distinguished vertex set $X \subseteq V(G)$ and let $p \in [0,1]$. Let $\mathbb{G}_*^{(3)}(G, X, p)$ be the distribution on pairs of triangle sets $(\mathcal{H}, \mathcal{H}')$ defined as follows: First, include each triangle of $G \setminus G[X]$ in $\mathcal{H}'$ with probability p, independently. Then, let $\mathcal{H} \subseteq \mathcal{H}'$ be the set of triangles that are edge-disjoint from all other triangles in $\mathcal{H}'$.

Lemma 3.3.3. *The following holds for a sufficiently large $C_{3.3.3} > 0$. Fix a graph G on $n \geq C_{3.3.3}$ vertices with $\Delta(G^c) \leq n/(\log n)$, an integer $2 \leq m \leq (\log n)^{3/4}$, a set $X \subseteq V(G)$ of size at most $n/2$, and a triangle T in $G[X]$. Let $(\mathcal{H}, \mathcal{H}') \sim \mathbb{G}_*^{(3)}(G, X, 1/n)$. Then with probability at least $1/(4e)^{12m-6}$, there exists a collection $\mathcal{C}$ of $2m-1$ triangles in $\mathcal{H}$ such that $\mathcal{F} \coloneqq \mathcal{C} \cup \{T\}$ is a copy of $\mathcal{F}_{2m}$ with $V(\mathcal{F}) \cap X = V(T)$.*

Proof. Let $\mathfrak{F}$ be the collection of copies $\mathcal{F}$ of $\mathcal{F}_{2m}$ using triangles from G such that $V(\mathcal{F}) \cap X = V(T)$. Let Z be the number of copies $\mathcal{F} \in \mathfrak{F}$ such that $\mathcal{F} \setminus \{T\} \subseteq \mathcal{H}$. We observe that

$$|\mathfrak{F}| \geq 3(2m-1)! \binom{n/2}{2m-1} \left(1 - O\left(\frac{m}{\log n}\right)\right). \tag{3.3.1}$$

This accounts for the three non-isomorphic ways to embed T into $\mathcal{F}_{2m}$ and the number of ways to choose the remaining vertices. The factor of $1 - O(m/\log n)$ accounts for the fact that not all edges of K_n are present in G.

Let $\mathcal{F} \in \mathfrak{F}$ and let $\mathcal{C} \coloneqq \mathcal{F} \setminus \{T\}$. We claim that

$$\mathbb{P}[\mathcal{C} \subseteq \mathcal{H}] \geq \left(\frac{1}{2e^3 n}\right)^{2m-1}. \tag{3.3.2}$$

Indeed, by definition, $\mathbb{P}[\mathcal{C} \subseteq \mathcal{H}'] = n^{-(2m-1)}$. Next, we bound $\mathbb{P}[\mathcal{C} \subseteq \mathcal{H} | \mathcal{C} \subseteq \mathcal{H}']$. Conditioning on $\mathcal{C} \subseteq \mathcal{H}'$, this is equal to the probability that the 2-skeleton of $\mathcal{C}$ does not participate in any triangles of $\mathcal{H}'$ besides $\mathcal{C}$. There are at most $|E(\mathcal{C})|n = (6m-3)n$ such triangles, and so the probability that none of them are in $\mathcal{H}'$ is at least $(1-1/n)^{(6m-3)n} > (1.1e)^{-(6m-3)}$. This proves (3.3.2).

By linearity of expectation, (3.3.1), and (3.3.2) we have that

$$\mathbb{E}Z \geq \left(\frac{1}{2e}\right)^{6m-3}. \tag{3.3.3}$$

Now note that

$$\mathbb{E}\left[Z^2\right] = \mathbb{E}Z + \mathbb{E}[Z(Z-1)] \leq \mathbb{E}Z + \left(3(2m-1)!\binom{n}{2m-1}\right)^2 n^{-2(2m-1)} \leq \mathbb{E}Z + 9.$$

In calculating $\mathbb{E}[Z(Z-1)]$, we have used the fact that any pair of distinct configurations that appear in $\mathcal{H}$ do not share a triangle. Applying the second moment method we obtain

$$\mathbb{P}[Z>0] \geq \frac{(\mathbb{E}Z)^2}{\mathbb{E}[Z^2]} \geq \frac{(\mathbb{E}Z)^2}{\mathbb{E}Z+9} \overset{(3.3.3)}{>} \left(\frac{1}{4e}\right)^{12m-6},$$

completing the proof. □

The next lemma follows via a direct union-bound computation.

Lemma 3.3.4. *Fix a graph G on n vertices and a pair of distinct triangles T_1, T_2 in G, and sample each triangle in G with probability $1/n$. Call this random hypergraph $\mathcal{H}$. Let $2 \leq m \leq (\log n)^{3/4}$. Let q be the probability that $\mathcal{H} \cup \{T_1, T_2\}$ contains two copies, $\mathcal{F}^1$ and $\mathcal{F}^2$, of $\mathcal{F}_{2m}$, with $T_i \in \mathcal{F}^i$ and $T_{3-i} \notin \mathcal{F}^i$ for $i \in \{1,2\}$, that share a triangle. There exists a constant $C_{3.3.4} > 0$ such that $q \leq C_{3.3.4} 2^{3m}/n^2$ if T_1 and T_2 are vertex-disjoint and $q \leq C_{3.3.4} 2^{3m}/n$ if T_1 and T_2 share exactly one vertex.*

Proof. Consider two copies of $\mathcal{F}_{2m}$ and all possible ways to identify vertices of one copy to vertices of the other (if there is a repeated triangle as a result of this gluing, we keep only one copy). There are at most $(2^m)^2$ resulting hypergraphs. Suppose that $\mathcal{F}'$ is obtained in this way and has a repeated triangle. Consider the probability that a copy of $\mathcal{F}'$ that extends T_1, T_2 simultaneously can be found in $\mathcal{H}$. There are $O(m^2)$ ways to choose which triangles correspond to T_1, T_2. Fixing such a choice the probability is bounded by $n^{v(\mathcal{F}')-6}(1/n)^{e(\mathcal{F}')-2}$ in the vertex-disjoint case and $n^{v(\mathcal{F}')-5}(1/n)^{e(\mathcal{F}')-2}$ when T_1 and T_2 share a vertex. Therefore, it suffices to show that in both cases $e(\mathcal{F}') \geq v(\mathcal{F}') - 2$.

We will use the following property of $\mathcal{F}_{2m}$, which is related to it being an *Erdős configuration* in the sense of [23, 64]. Let $\mathcal{S} \subseteq \mathcal{F}_{2m}$ be a set of $s > 0$ triangles. Then $\mathcal{S}$ is incident to at least $s+2$ vertices. Indeed, if $s = 2m$ then $\mathcal{S} = \mathcal{F}_{2m}$ and so $\mathcal{S}$ is incident to all $2m+2 = s+2$ vertices. Otherwise, the triangles in $\mathcal{S}$ cover $s < 2m$ edges in the cycle C_{2m}. The covered edges form a collection of paths and so have at least $s+1$ vertices. Furthermore, $\mathcal{S}$ is incident to at least one of the two additional vertices in $\mathcal{F}_{2m}$, and so $\mathcal{S}$ is incident to at least $s+2$ vertices, as desired.

Returning to the main argument, let v be the number of pairs of glued vertices and t be the number of triangles that occur as repeated triangles. We have $v(\mathcal{F}') = 2(2m+2) - v$ and $e(\mathcal{F}') = 2(2m) - t$, so we equivalently must show $v \geq t+2$. Suppose that the t repeated triangles, within a single copy of $\mathcal{F}_{2m}$, span w vertices. By the argument above $w \geq t+2$. Moreover, it is evident that $v \geq w$, completing the proof. □

3.3.1 Bootstrapping

We are now in a position to lay out the bootstrapping argument that establishes iteratively improved versions of Theorem 3.3.1 and, eventually, Theorem 3.1.1.

Proof of Theorem 3.3.1. For $k \geq 0$ let $\eta_k = 2/(k+2)$. As discussed, Theorem 3.3.1($\eta = \eta_0$) follows from [54, Theorem 1.4] or [24, Theorem 1.1]. We will show Theorem 3.3.1($\eta = \eta_k$) via induction on k. Then, to see the result for arbitrary $\eta > 0$ we may round down to the nearest η_k and appropriately adjust the constant $C_{3.3.1}$.

Let $k \geq 1$ and assume that Theorem 3.3.1($\eta = \eta_{k-1}$) holds. Set $\gamma_k = 4/(2k+5)$ and consider a graph G on $n \geq \exp(C_{3.3.1}/\eta_k^4)$ vertices satisfying the conditions of Theorem 3.3.1($\eta = \eta_k$). Our goal is to construct a measure μ on triangle-decompositions of G which is $O(n^{\eta_k - 1}/(\log n)^2)$-spread, since then Theorem 3.1.6 will imply the result.

Our first task is to construct the absorber. As it concerns the measure μ, the absorber is deterministic. However, we will need it to satisfy certain structural properties. For this reason we will use the probabilistic method. We define the following quantities:

$$M = n^{1-\gamma_k}\exp\big((\log n)^{1/3}\big),\ m = \sqrt{\log n},$$
$$\ell_1 = \sqrt{\frac{n}{M^{1+\eta_{k-1}}}},\ \ell_2 = \frac{\sqrt{M^{1+\eta_{k-1}}n}}{e^m},\ \ell_1' = \frac{\ell_1}{2(\log M)^6}.$$

Claim 3.3.5. *There exist sets $X, X_1, \dots, X_{\ell_1}, Y_1, \dots, Y_{\ell_1} \subseteq V(G)$ satisfying the following conditions.*

Ab1 *$|X| = M$ and $Y_1, \dots, Y_{\ell_1}$ are disjoint sets of vertices within $V(G) \setminus X$, each of size ℓ_2, and $X_1, \dots, X_{\ell_1} \subseteq X$ are sets of vertices of size $|X|/(\log|X|)^2$. We set $G_i = G[X_i \cup Y_i]$.*

Ab2 *Every triangle in $G[X]$ is contained in at least ℓ_1' graphs G_i.*

Ab3 *Every $v \in V(G)$ satisfies $|(V(G) \setminus N_G(v)) \cap X| \leq |X|(1/\log|X| - 2/(\log|X|)^2)$.*

Ab4 *Every $i \in [\ell_1]$ satisfies $\Delta(G_i^c) \leq |V(G_i)|/\log|V(G_i)|$.*

Proof. First, let $X \subseteq V(G)$ be a uniformly random set of size M. Then, sample ℓ_1 disjoint sets of vertices $Y_1, \dots, Y_{\ell_1}$ of size ℓ_2 uniformly at random from $V(G) \setminus X$. For each $i \in [\ell_1]$ choose a set $X_i \subseteq X$ of size $|X|/(\log|X|)^2$ uniformly at random. Then **Ab1** is satisfied by definition.

Observe that the number of graphs G_i containing a fixed triangle in $G[X]$ is distributed binomially with parameters $(\ell_1, (\log M)^{-2})$. Therefore, by Chernoff's inequality and a union bound, with probability $1 - O(|X|^{-1})$ every triangle in $G[X]$ is contained in at least ℓ_1' graphs G_i. This establishes **Ab2**.

Recall that $\Delta(G^c) \leq n/\log n$. Thus, since X was chosen randomly, by a similar application of Chernoff's inequality we conclude that with probability $1 - O(|X|^{-1})$ each $v \in V(G)$ satisfies $|(V(G) \setminus N(v)) \cap X| \leq |X|(1/\log|X| - 2/(\log|X|)^2)$. This proves **Ab3**. **Ab4** follows from a similar argument. □

For the remainder of the proof we fix (deterministic) sets $X, X_1, \dots, X_{\ell_1}, Y_1, \dots, Y_{\ell_1}$ and graphs $G_1, \dots, G_{\ell_1}$ satisfying **Ab1** to **Ab4**.

We now define μ, which corresponds to the output of a randomized algorithm to find a triangle-decomposition of G. At a high level, we (i) sample a rich random *template* of

triangles within $X \cup Y_1 \cup \cdots \cup Y_{\ell_1}$ to use as an absorber, (ii) use Proposition 3.3.2 to run iterative absorption to cover the remainder apart from X in a spread fashion, (iii) use Theorem 3.3.1($\eta = \eta_{k-1}$) to decompose this remainder in a $|X|^{-1+\eta_{k-1}}$-spread fashion, and (iv) simultaneously flip all of the resulting triangles within X using the template to improve the spread. The algorithm is as follows:

Alg1 For each $i \in [\ell_1]$ let $G_i' := G_i[X_i \cup Y_i] \setminus G_i[X_i]$ and sample $(\mathcal{H}_i, \mathcal{H}_i') \sim \mathbb{G}_*^{(3)}(G_i', X_i, |V(G_i')|^{-1})$ (with independent samples for each $i \in [\ell_1]$).

Alg2 Let $\mathcal{H}_{\mathrm{IA}} \sim \mathbb{G}^{(3)}(n, (\log |X|)^{C_{3.3.2}}/|X|)$.

Alg3 Sample every triangle in $G[X]$ with probability $|X|^{-1+\eta_{k-1}}$ and call this family $\mathcal{H}_{\mathrm{Rand}}$.

Alg4 Condition on the event $\mathcal{E}_{\mathrm{IA}}$ that there is a triangle set $\mathcal{H}_{\mathrm{Dec}} \subseteq \mathcal{H}_{\mathrm{IA}}$ satisfying the conditions Proposition 3.3.2(1-3) for $G' = G \setminus (\bigcup_{i=1}^{\ell_1} E(\mathcal{H}_i))$.

Alg5 Condition on the event $\mathcal{E}_{\mathrm{Ind}}$ that $(G' \setminus E(\mathcal{H}_{\mathrm{Dec}}))[X]$ has a triangle-decomposition $\mathcal{H}_{\mathrm{Ind}} \subseteq \mathcal{H}_{\mathrm{Rand}}$.

Alg6 Condition on the event $\mathcal{E}_{\mathrm{Abs}}$ that we can find, for each $T \in \mathcal{H}_{\mathrm{Ind}}$, an index $i_T \in [\ell_1]$ and a 3-uniform hypergraph $F_T \simeq \mathcal{F}_{2m}$ contained in $\mathcal{H}_{i_T} \cup \{T\}$ such that (a) $V(F_T) \cap X = V(T)$ and (b) the triangle sets $\{F_T\}_{T \in \mathcal{H}_{\mathrm{Ind}}}$ are edge-disjoint.

Alg7 Let $\mathcal{H}_{\mathrm{Abs}} = \bigcup_{T \in \mathcal{H}_{\mathrm{Ind}}} (F_T \setminus \{T\})$ and let $\mathcal{H}_{\mathrm{Flip}} = \bigcup_{T \in \mathcal{H}_{\mathrm{Ind}}} F_T^*$, where for each $T \in \mathcal{H}_{\mathrm{Ind}}$ the 3-graph F_T^* consists of all triangles of the corresponding "absorber-flip" $\mathcal{F}_{2m}^*$ associated to F_T.

Alg8 Finally, output the triple system $\mathcal{H} = \mathcal{H}_{\mathrm{Dec}} \cup \mathcal{H}_{\mathrm{Flip}} \cup (\bigcup_{i=1}^{\ell_1} \mathcal{H}_i \setminus \mathcal{H}_{\mathrm{Abs}})$.

Remark 3.3.6. Although we have described an algorithm to *sample* from μ, this should not be misconstrued as an algorithm to *find* the triangle-decomposition guaranteed by Theorem 3.3.1. The main reason is that we will ultimately use the non-algorithmic Theorem 3.1.6 applied to μ.

Nevertheless, with a slight modification to the algorithm one can efficiently sample from μ. To do so, in **Alg5**, one should inductively invoke the ability to efficiently sample from an $|X|^{-1+\eta_{k-1}}$-spread distribution on triangle-decompositions. However, we would still rely on Theorem 3.1.6 to convert the spread distribution into a threshold result. Additionally, the analysis is slightly more involved. As thresholds, rather than spread distributions, are our main concern, we have not made this change.

To verify that μ satisfies the desired properties, we must check that μ is supported on triangle-decompositions and also that the above process succeeds with nonzero probability (otherwise μ is not well-defined). We do so in Claims 3.3.7 and 3.3.8. Finally, in Claim 3.3.9 we show that μ has the appropriate spread.

Claim 3.3.7. *$\mathcal{H}$ is a triangle-decomposition of G.*

Proof. First note that $\bigcup_{i=1}^{\ell_1} \mathcal{H}_i$ is a set of edge-disjoint triangles contained in G. Indeed, every $\mathcal{H}_i$ is such a set by definition. Additionally, the sets $Y_1, \dots, Y_{\ell_1}$ are disjoint, and by definition every edge in G_i' contains at least one vertex from Y_i. Hence, the triangle sets $\mathcal{H}_1, \dots, \mathcal{H}_{\ell_1}$ are mutually edge-disjoint.

Next, by **Alg4** we have that $\mathcal{H}_{\text{Dec}}$ is a set of edge-disjoint triangles covering all edges of G' except for some in X. By **Alg5** these remaining edges are covered by $\mathcal{H}_{\text{Ind}}$. That is, $\mathcal{H}_{\text{Dec}} \cup \mathcal{H}_{\text{Ind}} \cup \bigcup_{i=1}^{\ell_1} \mathcal{H}_i$ is a triangle-decomposition of G. Finally, we have $E(F_T) = E(F_T^*)$ and the F_T are edge-disjoint by **Alg6**. Hence $\mathcal{H}_{\text{Ind}} \cup \mathcal{H}_{\text{Abs}}$ and $\mathcal{H}_{\text{Flip}}$ are each edge-disjoint triangle-decompositions, and they decompose the same underlying graph. Since $\mathcal{H}_{\text{Abs}} \subseteq \bigcup_{i=1}^{\ell_1} \mathcal{H}_i$ by **Alg6**, the result follows. □

Next we show that μ is well-defined. In fact, we show that the above process defining μ succeeds with probability at least $(3/4)(1/2)(2/3) = 1/4$.

Claim 3.3.8. *We have* $\mathbb{P}[\mathcal{E}_{\text{IA}}] \geq 3/4$, $\mathbb{P}[\mathcal{E}_{\text{Ind}} | \mathcal{E}_{\text{IA}}] \geq 1/2$, *and* $\mathbb{P}[\mathcal{E}_{\text{Abs}} | \mathcal{E}_{\text{IA}} \wedge \mathcal{E}_{\text{Ind}}] \geq 2/3$.

Proof. We first claim that G' and X satisfy the assumptions of Proposition 3.3.2 regardless of the outcome of $\bigcup_{i=1}^{\ell_1} \mathcal{H}_i$. Note that by **Alg1** and **Alg4**, G' differs from G only by edges in $\bigcup_{i=1}^{\ell_1} E(G_i')$. Additionally, since $Y_1, \dots, Y_{\ell_1}$ are disjoint, every $v \in V(G) \setminus X$ is contained in at most one graph G_i'. By construction, every $v \in Y_i$ has at most $|X_i| = |X|/(\log |X|)^2$ neighbors in X within G_i'. Therefore $|X \setminus N_{G'}(v)| \leq |X|(1/\log |X| - 2/(\log |X|)^2) + |X|/(\log |X|)^2$, using **Ab3**. For $v \in X$, the number of neighbors within X is unchanged upon removing $\bigcup_{i=1}^{\ell_1} E(G_i')$. Additionally, $C_{3.3.2} \leq |X| \leq n/(\log n)^3$ is immediate from **Ab1** and the assumption $n \geq \exp(C_{3.3.1}/\eta_k^4)$. Finally,

$$\Delta((G')^c) \leq \Delta(G^c) + |X| + \sum_{i=1}^{\ell_1} |Y_i| \overset{\textbf{Ab1}}{\leq} \Delta(G^c) + n/(\log n)^2 \leq n/\log n + n/(\log n)^2.$$

Thus all conditions are satisfied, and by Proposition 3.3.2 the necessary $\mathcal{H}_{\text{Dec}} \subseteq \mathcal{H}_{\text{IA}}$ exists with probability at least $3/4$.

We condition on $\mathcal{E}_{\text{IA}}$ occurring. This means that $(G' \setminus E(\mathcal{H}_{\text{Dec}}))[X]$ has maximum degree at most $|X|/\log |X|$ due to Proposition 3.3.2(3). Therefore the inductive hypothesis Theorem 3.3.1($\eta = \eta_{k-1}$) applies and shows that the necessary $\mathcal{H}_{\text{Ind}} \subseteq \mathcal{H}_{\text{Rand}}$ exists with probability at least $1/2$, since

$$|X| \geq n^{1-\gamma_k} \geq (\exp(C_{3.3.1}/\eta_k^4))^{1-\gamma_k} \geq \exp(C_{3.3.1}/\eta_{k-1}^4).$$

It remains to prove that $\mathbb{P}[\mathcal{E}_{\text{Abs}} | \mathcal{E}_{\text{IA}} \wedge \mathcal{E}_{\text{Ind}}] \geq 2/3$. The proof is more involved, and we break it into three steps.

Step 1: Potential absorbers with few overlaps. We first establish that certain conditions hold with high probability in the unconditional model. For any triangle T of $G[X]$ and any $i \in [\ell_1]$, let $\mathfrak{A}_{T,i}$ be the set of copies $\mathcal{F}$ of $\mathcal{F}_{2m}$ that contain T, use only triangles in $\mathcal{H}_i \cup \{T\}$, and $V(\mathcal{F}) \cap X = V(T)$. Note that this is a random collection depending only on $\mathcal{H}_i$. Given T and i, we consider (a) the number of choices $N_{T,i} = |\mathfrak{A}_{T,i}|$ and (b) the number

$M_{T,i}$ of triangles T' of $G[X]$ such that $E(T') \cap E(T) = \emptyset$ and $\mathfrak{A}_{T,i}, \mathfrak{A}_{T',i}$ contain a pair of copies of $\mathcal{F}_{2m}$ that are not edge-disjoint (equivalently, these copies share a triangle). Let $\mathcal{E}_{\text{Tem}}$ be the event that the following conditions for the random template hold:

Tem1 For each triangle T of $G[X]$, $|\{i \in [\ell_1] \colon N_{T,i} > 0\}| \geq \ell_1'/(8e)^{12m-6}$.

Tem2 For each T, at most $\ell_1'/(32e)^{12m-6}$ indices $i \in [\ell_1]$ satisfy $M_{T,i} \geq 2^{90m}|X|^2/\ell_2$.

We claim that $\mathbb{P}[\mathcal{E}_{\text{Tem}}] \geq 1 - 1/n$. For **Tem1**, fix T and note that $N_{T,i}$ for $i \in [\ell_1]$ are independent, as each depends only on $\mathcal{H}_i$. Furthermore, Lemma 3.3.3 implies that $\mathbb{P}[N_{T,i} > 0] \geq 1/(4e)^{12m-6}$ whenever $V(T) \subseteq X_i$. By **Ab2** we have at least ℓ_1' such indices, so Chernoff's inequality implies that $|\{i \in [\ell_1] : N_{T,i} > 0\}| \geq \ell_1'/(8e)^{12m-6}$ with probability at least $1 - n^{-5}$. Applying a union bound, **Tem1** holds with probability at least $1 - n^{-2}$.

For **Tem2**, fix T and note that by linearity of expectation and Lemma 3.3.4 we have

$$\mathbb{E}M_{T,i} \leq |X|^2 \cdot O(2^{3m}/\ell_2) + |X|^3 \cdot O(2^{3m}/\ell_2^2) \leq 2^{4m}|X|^2/\ell_2,$$

since every triangle in $G[X]$ shares a vertex with at most $|X|^2$ other triangles in $G[X]$. By Markov's inequality, we have

$$\mathbb{P}[M_{T,i} \geq 2^{90m}|X|^2/\ell_2] \leq 2^{-86m}.$$

Now let Z_T be the number of indices $i \in [\ell_1]$ satisfying $M_{T,i} \geq 2^{90m}|X|^2/\ell_2$. Applying Chernoff's inequality:

$$\mathbb{P}[Z_T \geq \ell_1'/(32e)^{12m-6}] \leq \frac{1}{n^5}.$$

By a union bound, **Tem2** holds with probability at least $1 - n^{-2}$. Thus $\mathcal{E}_{\text{Tem}}$ holds with probability at least $1 - 1/n$, as desired.

Step 2: Few overlaps in absorbers for $\mathcal{H}_{\text{Rand}}$. Next we show that the number of potential conflicts counted by $M_{T,i}$ significantly diminishes (w.h.p.) when we consider only the sparse random set $\mathcal{H}_{\text{Rand}}$. First, for each triangle T in $G[X]$ we define the index set I_T of indices i satisfying $N_{T,i} > 0$ and $M_{T,i} < 2^{90m}|X|^2/\ell_2$. If $\mathcal{E}_{\text{Tem}}$ holds then $|I_T| \geq \ell_1'/(32e)^{12m-6}$ for all T.

Next, given T, let $\mathcal{M}_T'$ be the set of triangles $T' \in \mathcal{H}_{\text{Rand}}$ with $E(T') \cap E(T) = \emptyset$ such that for some $i \in I_T$, the collections $\mathfrak{A}_{T,i}, \mathfrak{A}_{T',i}$ contain a pair of copies of $\mathcal{F}_{2m}$ that are not edge-disjoint. Let $\mathcal{E}_{\text{Pack}}$ be the event that the following holds:

Pack1 For all $T \in \mathcal{H}_{\text{Rand}}$, we have $|\mathcal{M}_T'| \leq 2^{95m}|X|^{1+\eta_{k-1}}\ell_1/\ell_2$.

We claim that $\mathbb{P}[\mathcal{E}_{\text{Pack}}|\mathcal{E}_{\text{Tem}}] \geq 1 - 1/n$. Indeed, reveal all of $\bigcup_{i=1}^{\ell_1} \mathcal{H}_i$ and fix any triangle T of $G[X]$. Condition on $T \in \mathcal{H}_{\text{Rand}}$ and on $\mathcal{E}_{\text{Tem}}$. Then, expose the remainder of $\mathcal{H}_{\text{Rand}}$. It follows that $|\mathcal{M}_T'|$ is distributed as $\text{Bin}(M^*, |X|^{-1+\eta_{k-1}})$ for some

$$M^* \leq \sum_{i \in I_T} M_{T,i} \leq 2^{90m}|X|^2\ell_1/\ell_2.$$

Chernoff's inequality and a union bound now imply that $\mathbb{P}[\mathcal{E}_{\text{Pack}}|\mathcal{E}_{\text{Tem}}] \geq 1 - 1/n$, as desired.

Step 3: Finding simultaneous edge-disjoint absorbers. Now we condition on $\mathcal{E}_{\text{IA}}$ and $\mathcal{E}_{\text{Ind}}$ occurring. We have

$$\mathbb{P}[(\mathcal{E}_{\text{Tem}} \wedge \mathcal{E}_{\text{Pack}})^c|\mathcal{E}_{\text{IA}} \wedge \mathcal{E}_{\text{Ind}}] \leq \frac{O(1/n)}{1/2 \cdot 3/4} = O(1/n),$$

so $\mathbb{P}[\mathcal{E}_{\text{Tem}} \wedge \mathcal{E}_{\text{Pack}}|\mathcal{E}_{\text{IA}} \wedge \mathcal{E}_{\text{Ind}}] \geq 1 - O(1/n)$.

We now argue that if $\mathcal{E}_{\text{Tem}} \wedge \mathcal{E}_{\text{Pack}} \wedge \mathcal{E}_{\text{IA}} \wedge \mathcal{E}_{\text{Ind}}$ holds then $\mathcal{E}_{\text{Abs}}$ holds, which will finish the proof. Assuming $\mathcal{E}_{\text{Ind}}$, **Alg5** succeeds so there exists some edge-disjoint collection $\mathcal{H}_{\text{Ind}} \subseteq \mathcal{H}_{\text{Rand}}$. We will use the Lovász Local Lemma (Lemma 3.2.5) to prove the existence of the absorbers necessary for **Alg6**. As we are assuming **Tem1** and **Tem2**, for each $T \in \mathcal{H}_{\text{Ind}}$ we have a nonempty set of indices I_T such that for all $i \in I_T$, $N_{T,i} > 0$. For every $T \in \mathcal{H}_{\text{Ind}}$ we choose, uniformly at random, an index $i_T \in I_T$ and then uniformly at random choose one of the extensions of T (isomorphic to $\mathcal{F}_{2m}$) counted by N_{T,i_T}. We make these choices independently for each triangle in $\mathcal{H}_{\text{Ind}}$. We claim that with nonzero probability, all of these extensions are edge-disjoint.

We define a "disjointness graph" H with vertex set $\mathcal{H}_{\text{Ind}}$. For each $T, T' \in \mathcal{H}_{\text{Ind}}$, we put an edge between them if there is some $i \in I_T \cap I_{T'}$ such that $\mathfrak{A}_{T,i}, \mathfrak{A}_{T',i}$ contain a pair of copies of $\mathcal{F}_{2m}$ that are not edge-disjoint. We see that $N_H(T) \subseteq \mathcal{H}_{\text{Ind}} \cap \mathcal{M}'_T$. Hence, by **Pack1**,

$$\Delta(H) \leq 2^{95m}|X|^{1+\eta_{k-1}}\ell_1/\ell_2 = 2^{90m}e^m = e^{O(\sqrt{\log n})}.$$

For each edge $f \in E(H)$, let $\mathcal{B}_f$ be the "bad" event that the random extensions chosen for T and T' share an edge. We wish to show that with nonzero probability, we can simultaneously avoid all the bad events. This will prove the result, since by definition the only pairs T, T' that can have a conflict with this process are those corresponding to some $f \in E(H)$. To apply the Lovász Local Lemma we observe that each $\mathcal{B}_f$ is mutually independent from all other events except for $\mathcal{B}_{f'}$ where f, f' share a vertex. There are at most $2\Delta(H)$ such events. Additionally, for each bad event $\mathcal{B}_f$ where $f = \{T, T'\}$ we have

$$\mathbb{P}[\mathcal{B}_f] \leq \mathbb{P}[i_T = i_{T'}] \leq \frac{1}{|I_T|} \leq \frac{(32e)^{12m-6}}{\ell'_1}.$$

Since $\ell_1 \geq n^{1/(20k^2)}$, we see that $\ell'_1 = n^{\Omega(1)}$ and therefore

$$e \max_{f \in E(H)} \mathbb{P}[\mathcal{B}_f] \cdot (2\Delta(H) + 1) = \frac{e^{O(\sqrt{n})}}{n^{\Omega(1)}} < 1.$$

The result follows. □

Finally, we verify the spread condition.

Claim 3.3.9. *μ is $O(n^{\eta_k - 1}/(\log n)^2)$-spread.*

Proof. **Step 1: Spread of $\mathcal{H} \setminus \mathcal{H}_{\text{Flip}}$.** Let $\mathcal{T} = \{T_1, \dots, T_t\}$ be a set of triangles in G, and consider the probability that they simultaneously appear in a sample $\mathcal{S} \sim \mu$. Since $\mathcal{S}$ consists of edge-disjoint triangles we may assume that the triangles in $\mathcal{T}$ are edge-disjoint as well. Additionally, recall that the underlying probability space associated to μ is defined by independent samples $\mathcal{H}_{\text{IA}}$, $\mathcal{H}_{\text{Rand}}$, and $\mathcal{H}'_1, \dots, \mathcal{H}'_{\ell_1}$ conditional on $\mathcal{E}_{\text{IA}} \wedge \mathcal{E}_{\text{Ind}} \wedge \mathcal{E}_{\text{Abs}}$. Finally, we have $\mathcal{H} = \mathcal{H}_{\text{Dec}} \cup \mathcal{H}_{\text{Flip}} \cup (\bigcup_{i=1}^{\ell_1} \mathcal{H}_i \setminus \mathcal{H}_{\text{Abs}})$.

Suppose that $\mathcal{T} \subseteq \mathcal{S}$. Then each T_j is either in $\mathcal{H}_{\text{Dec}} \subseteq \mathcal{H}_{\text{IA}}$, in $\bigcup_{i=1}^{\ell_1} \mathcal{H}_i \setminus \mathcal{H}_{\text{Abs}} \subseteq \bigcup_{i=1}^{\ell_1} \mathcal{H}'_i$, or in $\mathcal{H}_{\text{Flip}}$. Let $J_1, J_2, J_3 \subseteq [t]$ be the index sets of the triangles contained in each of these respective sets. Observe that for triangles T_j such that $j \in J_3$, this means that T_j is in some "absorber-flip" $\mathcal{F}^*_{2m}$ where the corresponding $\mathcal{F}_{2m}$ consists of a triangle $S_j \in \mathcal{H}_{\text{Ind}} \subseteq \mathcal{H}_{\text{Rand}}$ and $2m-1$ triangles of $\bigcup_{i=1}^{\ell_1} \mathcal{H}'_i$ by **Alg6** and **Alg7**. Furthermore, those triangles cannot share an edge with any $\{T_j : j \in J_2\}$.

Now, for a partition $J_1 \sqcup J_2 \sqcup J_3 = [t]$, let $\mathcal{E}_{J_1,J_2,J_3}$ be the event that:

E1 $T_j \in \mathcal{H}_{\text{IA}}$ for all $j \in J_1$;

E2 $T_j \in \bigcup_{i=1}^{\ell_1} \mathcal{H}'_i$ for all $j \in J_2$;

E3 For all $j \in J_3$, there is a copy of $F_j \simeq \mathcal{F}_{2m}$, consisting of one triangle $S_j \in \mathcal{H}_{\text{Rand}}$ and $2m-1$ triangles of $\bigcup_{i=1}^{\ell_1} \mathcal{H}'_i$ such that T_j is in the associated "absorber-flip" $\mathcal{F}^*_{2m}$;

E4 F_j is edge-disjoint from $\{T_{j'} : j' \in J_2\}$;

E5 Every pair $F_j, F_{j'}$ for distinct $j, j' \in J_3$ is edge-disjoint or identical.

The above analysis shows that the union of the 3^t events $\mathcal{E}_{J_1,J_2,J_3}$ covers all possible situations where $\{T_1, \dots, T_t\} \subseteq \mathcal{S}$. Thus, we have

$$\mathbb{P}[\{T_1, \dots, T_t\} \subseteq \mathcal{S}] \leq \sum_{J_1,J_2,J_3} \mathbb{P}[\mathcal{E}_{J_1,J_2,J_3} | \mathcal{E}_{\text{IA}} \wedge \mathcal{E}_{\text{Ind}} \wedge \mathcal{E}_{\text{Abs}}] \leq 4 \cdot 3^t \max_{J_1,J_2,J_3} \mathbb{P}[\mathcal{E}_{J_1,J_2,J_3}]$$

by Claim 3.3.8 and Bayes' theorem. Furthermore, the event $\mathcal{E}_{J_1,J_2,J_3}$ does not depend directly on $\mathcal{S}$, but rather on an independent model of triangles. Thus, we can essentially disregard the complicated process **Alg1** to **Alg8** in favor of this substantially simpler situation.

We will reduce the situation further to studying the triangles in J_3. Given $J_3 \subseteq [t]$, let $\mathcal{E}_{J_3}$ be the event that **E3** and **E5** hold. We see that

$$\mathbb{P}[\mathcal{E}_{J_1,J_2,J_3}] \leq \left(\prod_{j \in J_1} \mathbb{P}[T_j \in \mathcal{H}_{\text{IA}}] \prod_{j \in J_2} \mathbb{P}\left[T_j \in \bigcup_{i=1}^{\ell_1} \mathcal{H}'_i \right] \right) \mathbb{P}[\mathcal{E}_{J_3}].$$

Indeed, the first term can be extracted due to independence of $\mathcal{H}_{\text{IA}}$ and conditions **E2** to **E5**. Additionally, careful scrutiny shows that in fact **E2** is independent from the event that events **E3** to **E5** hold by construction.

By definition, we have $\mathbb{P}[T_j \in \mathcal{H}_{\mathrm{IA}}] = (\log |X|)^{C_{3.3.2}}/|X|$ and $\mathbb{P}[T_j \in \bigcup_{i=1}^{\ell_1} \mathcal{H}_i'] \le 1/|V(G_i')| \le 1/|X|$. Since $|X| = n^{1-\gamma_k}\exp((\log n)^{1/3})$, these terms are bounded by $n^{\eta_k - 1}/(\log n)^2$. Putting everything together, we find

$$\mathbb{P}[\{T_1, \dots, T_t\} \subseteq \mathcal{S}] \le 4 \cdot 3^t \max_{J \subseteq [t]} (n^{\eta_k - 1}/(\log n)^2)^{t-|J|} \mathbb{P}[\mathcal{E}_J]. \tag{3.3.4}$$

Step 2: Preliminaries for understanding $\mathcal{H}_{\mathrm{Flip}}$. Now we analyze the absorber-flips. Call a copy of $\mathcal{F}_{2m}$ with a distinguished triangle a *rooted absorber*, and call any nonempty subset of triangles $\mathcal{P} \subseteq \mathcal{F}_{2m}^*$ within the flip of a rooted absorber a *polymer*. If we additionally distinguish a triangle of a polymer we call it a *rooted polymer*. We see that there are $2^{2m} - 1$ (labeled) polymers. At a high level, we wish to count the number of ways to break $\{T_j : j \in J\}$ into polymers, then count extensions to a full $\mathcal{F}_{2m}^*$, and then consider the probability that the corresponding $\mathcal{F}_{2m}$ is in $\mathcal{H}_{\mathrm{Rand}} \cup \bigcup_{i=1}^{\ell_1} \mathcal{H}_i'$.

Let

$$p(s) = \max_{T_1, \dots, T_t} \max_{J \in \binom{[t]}{s}} \mathbb{P}[\mathcal{E}_J],$$

where the maximum is over edge-disjoint collections of triangles.

Given a triangle T of G_i' for some i and an edge-disjoint triangle set $\mathcal{T}$, both within K_n, and given a rooted polymer $\mathcal{P}$, let $f(\mathcal{P}, T, \mathcal{T})$ be the number of ways to extend T to a copy of $\mathcal{P}$ within G_i' where T is the root of the polymer, and the other triangles are all in $\mathcal{T}$. We claim that

$$f(\mathcal{P}, T, \mathcal{T}) \le \begin{cases} (2\ell_2)^{v(\mathcal{P}) - e(\mathcal{P}) - 2} & e(\mathcal{P}) \le 2m - 1 \\ (2\ell_2) & e(\mathcal{P}) = 2m. \end{cases} \tag{3.3.5}$$

(Here, $v(\mathcal{P})$ is the number of vertices incident to triangles in $\mathcal{P}$.)

We first reduce to the case where $e(\mathcal{P}) \le 2m-1$. Indeed, if $e(\mathcal{P}) = 2m$ let $\mathcal{P}'$ be a polymer obtained by removing a non-root triangle from $\mathcal{P}$. It then holds that $f(\mathcal{P}, T, \mathcal{T}) \le f(\mathcal{P}', T, \mathcal{T})$ and $v(\mathcal{P}') - e(\mathcal{P}') - 2 = 2m + 2 - (2m-1) - 2 = 1$. Thus, (3.3.5) with $e(\mathcal{P}) = 2m$ follows from (3.3.5) with $e(\mathcal{P}) = 2m - 1$.

We prove (3.3.5) when $e(\mathcal{P}) \le 2m - 1$ by induction on $e(\mathcal{P})$. When $e(\mathcal{P}) = 1$, the corresponding polymer consists only of the root triangle and hence must be $\{T\}$; the result follows. Now assume that (3.3.5) holds for all $e(\mathcal{P}) < k$ where $2 \le k \le 2m - 1$. Let $\mathcal{P}$ be a polymer with $e(\mathcal{P}) = k$. Now fix a rooted polymer $\mathcal{P}' \subseteq \mathcal{P}$ such that $e(\mathcal{P} \setminus \mathcal{P}') = 1$ and $v(\mathcal{P}) - v(\mathcal{P}') \ge 1$ (this is possible since $2 \le e(\mathcal{P}) \le 2m - 1$: any such polymer covers a proper subgraph of the cycle C_{2m}, and we can remove a non-root triangle containing a degree 1 vertex from $\mathcal{P}$). By the inductive hypothesis

$$f(\mathcal{P}', T, \mathcal{T}) \le (2\ell_2)^{v(\mathcal{P}') - e(\mathcal{P}') - 2}.$$

We now consider the possible ways to extend a copy of $\mathcal{P}'$ to a copy of $\mathcal{P}$, using only triangles from $\mathcal{T}$. There are three cases. If $v(\mathcal{P}) - v(\mathcal{P}') = 3$, note that since the triangles in $\mathcal{T}$ are edge-disjoint there are at most $|\mathcal{T}| \le |V(G_i')|^2 \le (2\ell_2)^2$ possible extensions. If $v(\mathcal{P}) - v(\mathcal{P}') = 2$, there are at most $(2\ell_2)$ possible triangles in $\mathcal{T}$ within G_i' which could be

in $\mathcal{P} \setminus \mathcal{P}'$ as this triangle must contain a fixed vertex (given $\mathcal{P}'$). Finally, if $v(\mathcal{P}) - v(\mathcal{P}') = 1$, note that there is at most 1 triangle in $\mathcal{T}$ within G_i' which could be in $\mathcal{P} \setminus \mathcal{P}'$ as this triangle must contain a fixed edge (given $\mathcal{P}'$) and the triangles in $\mathcal{T}$ are edge disjoint. This completes the inductive proof.

Next, given a polymer $\mathcal{P}$ within some G_i', let $g(\mathcal{P})$ be the number of ways to extend it to a full copy of $\mathcal{F}_{2m}^*$ within G_i' with its flip having the property that at least one triangle is fully within X. We claim that

$$g(\mathcal{P}) \le (2\ell_2)^{2m - v(\mathcal{P}) + 2}$$

and that if $e(\mathcal{P}) = 1$ then $g(\mathcal{P}) \le |X|(2\ell_2)^{2m-2}$. To see this note that given the polymer there are at most $2m - v(\mathcal{P}) + 2$ labelled vertices to be specified and at most $(2\ell_2)$ choices for each vertex. Additionally, when $e(\mathcal{P}) = 1$, at least one vertex remaining to be chosen must be in X (since the flip of $\mathcal{F}_{2m}^*$ has at least one triangle fully in X, while $\mathcal{F}_{wm}^*$ does not), improving the bound to $|X| \cdot (2\ell_2)^{2m+2-3-1} = |X|(2\ell_2)^{2m-2}$ as desired.

Step 3: Spread of $\mathcal{H}_{\text{Flip}}$. Finally, we bound $p(s)$. Suppose that $J \in \binom{[t]}{s}$. For each $j \in J$ in increasing order, there are at most $2^{2m} - 1$ ways to choose which polymer type $\mathcal{P}$ the set $\{T_{j'} \colon j' \in J,\ F_j = F_{j'}\}$ creates. Given $\mathcal{P}$, there are at most $f(\mathcal{P}, T_j, \{T_{j'} \colon j' \in J\})$ ways to choose how T_j actually extends to that polymer within $\{T_{j'} \colon j' \in J\}$. There are then at most $g(\mathcal{P})$ ways to count the number of extensions to a full $\mathcal{F}_{2m}^*$. Given these choices, the probability that the flip of this $\mathcal{F}_{2m}^*$ is contained in $\mathcal{H}_{\text{Rand}} \cup \bigcup_{i=1}^{\ell_1} \mathcal{H}_i'$ is at most $|X|^{\eta_{k-1}-1}(1/\ell_2)^{2m-1}$.

It follows that

$$\begin{aligned} p(s) &\le \sum_{s'=1}^{2m} \binom{2m}{s'} \Big(\max_{\substack{e(\mathcal{P})=s' \\ T,\mathcal{T}}} f(\mathcal{P}, T, \mathcal{T}) g(\mathcal{P}) \cdot |X|^{\eta_{k-1}-1}(1/\ell_2)^{2m-1} \Big) p(s - s'), \\ &\le \max_{\substack{s' \in [2m] \\ e(\mathcal{P})=s' \\ T,\mathcal{T}}} (2m)^{s'} f(\mathcal{P}, T, \mathcal{T}) g(\mathcal{P}) \cdot |X|^{\eta_{k-1}-1}(1/\ell_2)^{2m-1} p(s - s') \end{aligned}$$

where we let $p(s) = 0$ for $s < 0$ and $p(0) = 1$.

When $e(\mathcal{P}) = s' \in \{2, \dots, 2m-1\}$ we have

$$f(\mathcal{P}, T, \mathcal{T}) g(\mathcal{P}) \cdot |X|^{\eta_{k-1}-1}(1/\ell_2)^{2m-1} \le (2\ell_2)^{1-s'} \cdot |X|^{\eta_{k-1}-1} \le (n^{\eta_k - 1}/(\log n)^3)^{s'}.$$

The last inequality is true since it holds for $s' = 2$ and since $2\ell_2 \ge n^{1-\eta_k}(\log n)^3$.

When $e(\mathcal{P}) = 2m$ we have

$$f(\mathcal{P}, T, \mathcal{T}) g(\mathcal{P}) \cdot |X|^{\eta_{k-1}-1}(1/\ell_2)^{2m-1} \le 2\ell_2^{2-2m} \cdot |X|^{\eta_{k-1}-1} \le (n^{\eta_k - 1}/(\log n)^3)^{2m}.$$

The inequality holds as $|X|^{\eta_{k-1}-1} \le 1$ and $2\ell_2^{(2-2m)/(2m)} \le \exp(\sqrt{\log n})/(2\ell_2) \le n^{\eta_k - 1}/(\log n)^3$ as $n^{\Omega(1/k^2)} \ge \exp(O(m))$.

When $e(\mathcal{P}) = 1$ we have

$$f(\mathcal{P}, T, \mathcal{T}) g(\mathcal{P}) \cdot |X|^{\eta_{k-1}-1}(1/\ell_2)^{2m-1} \le |X|^{\eta_{k-1}} 2^{2m-2}/\ell_2$$

$$\leq (4e)^m/(n^{1/2}|X|^{(1-\eta_{k-1})/2}) \leq n^{\eta_k-1}/(\log n)^3$$

where the final inequality follows as $n^{\eta_k-1/2}|X|^{(1-\eta_{k-1})/2} = n^{\Omega(1/k^2)} \geq \exp(O(m))$.

Putting this together, we obtain

$$p(s) \leq \max_{s' \in [2m]} (4mn^{\eta_k-1}/(\log n)^3)^{s'} p(s-s').$$

Along with the initial conditions, this immediately yields $p(s) \leq (4mn^{\eta_k-1}/(\log n)^3)^s$. Finally, combining with (3.3.4) yields

$$\mathbb{P}[\{T_1, \dots, T_t\} \subseteq \mathcal{S}] \leq (O(n^{\eta_k-1}/(\log n)^2))^t,$$

as desired. □

Claims 3.3.7 and 3.3.9 imply that μ is an $O(n^{\eta_k-1/(\log n)^2})$-spread distribution on triangle-decompositions of G. Applying Theorem 3.1.6 to μ yields Theorem 3.3.1($\eta = \eta_k$), completing the induction. □

3.4 Iterative Absorption in Random Hypergraphs

In this section we use the machinery of iterative absorption to prove Proposition 3.3.2. Informally, it states that given a nearly complete graph G and a specified set $X \subseteq V(G)$, one can use edge-disjoint triangles to cover all edges in $G \setminus G[X]$ while only covering a small fraction of edges in $G[X]$. Moreover, the triangles can be restricted to a sparse random set. It is proved via iterating the following lemma.

Lemma 3.4.1. *There exists a constant $C_{3.4.1} > 0$ such that the following holds. Let $n \in \mathbb{N}$. Fix a subset $V_1 \subseteq V(K_n)$ such that $|V_1| \in (n/(\log n)^4, n/(\log n)^2)$. Furthermore fix $G \subseteq K_n$ such that $\Delta(G^c) \leq 2n/\log n$, and $|N(v)^c \cap V_1| \leq 2|V_1|/\log|V_1|$ for every $v \in V(G)$, and for every $v \notin V_1$ we have $\deg_G(v) \equiv 0 \pmod 2$.*

Let $\mathcal{H}' \sim \mathbb{G}^{(3)}(n, (\log n)^{C_{3.4.1}}/n)$. Then there exists an edge-disjoint triangle set $\mathcal{H} \subseteq \mathcal{H}'$, with $G^* := E(\mathcal{H})$, such that:

1. *$G^*[V_1]$ is stochastically dominated by sampling every edge independently with probability $(\log|V_1|)^{-20}$,*
2. *$G \setminus G[V_1] \subseteq G^*$ (i.e., $\mathcal{H}$ covers all edges of G outside of V_1) with probability $1 - n^{-\omega(1)}$,*
3. *$G^* \subseteq G$ (i.e., $\mathcal{H}$ consists of triangles in G).*

Before proving Lemma 3.4.1 we show how it implies Proposition 3.3.2.

Proof of Proposition 3.3.2 given Lemma 3.4.1. Let $n = t_0 > t_1 > \cdots > t_\ell = |X|$ be a sequence of integers such that $t_{i+1} \in (t_i/(\log t_i)^4, t_i/(\log t_i)^2)$ for every $0 \le i < \ell$. Observe that $\ell = O(\log n)$. Sample a uniformly random descending sequence of sets $V(K_n) = V_0 \supseteq V_1 \supseteq V_2 \supseteq \cdots \supseteq V_\ell = X$ such that $|V_i| = t_i$ for every i. We call this sequence of sets the *vortex*.

We now consider the respective degrees from each vertex set into the next. By applying the Chernoff bound (Lemma 3.2.3), a union bound, and the assumed upper bound on $|X \setminus N_G(v)|$, with probability at least 0.99 (over the random choice of the vortex), we have that every vertex $v \in V_i$ has $|N(v)^c \cap V_{i+1}| \le 3|V_{i+1}|/(2\log|V_{i+1}|)$ and $\Delta(K_n[V_i] \setminus G[V_i]) \le 3|V_i|/(2\log|V_i|)$ for $0 \le i < \ell$. We assume these conditions hold.

We now apply Lemma 3.4.1 inductively. Suppose that for some $0 \le i < \ell$ we have already applied Lemma 3.4.1 i times, leaving the graph $L_i \subseteq G[V_i]$ of uncovered edges. In order to simplify the analysis we will not apply Lemma 3.4.1 to L_i directly. Rather, we will apply it to the graph G_i^{curr}, defined as follows: If $i = \ell - 1$ then $G_i^{\text{curr}} = L_i$. Otherwise let $G_i^{\text{curr}} = L_i \setminus G[V_{i+2}]$. Applying Lemma 3.4.1 in this way implies that an edge in $G[V_i]$ can only be covered in the i-th or $(i-1)$-th stage of the algorithm (but not before).

We next note that as $|V_{i+2}| \le |V_i|/(\log|V_i|)^{5/4}$ we see that $G_i' := G[V_i] \setminus G[V_{i+2}]$ has the property that for all $v \in V_i$, we have $|N_{G_i'}(v)^c \cap V_i| \le 7|V_i|/(4\log|V_i|)$ and $|N_{G_i'}(v)^c \cap V_{i+1}| \le 7|V_{i+1}|/(4\log|V_{i+1}|)$. Hence G_i^{curr} satisfies the necessary conditions for Lemma 3.4.1 with high probability. Indeed, by the inductive assumption, after i steps of the process L_i is stochastically dominated by sampling every edge independently with probability $(\log|V_i|)^{-20}$. Thus, by Chernoff's inequality and a union bound, the two minimum degree assumptions hold w.h.p. Moreover, L_i is obtained from the triangle-divisible graph G by removing a set of edge-disjoint triangles. Therefore all degrees in L_i are even. Since G_i^{curr} is obtained from L_i by removing only edges spanned by V_{i+2}, the degrees of all vertices in $V_i \setminus V_{i+1}$ in G_i^{curr} are even as well. Therefore we can apply Lemma 3.4.1 to G_i^{curr} and continue the process. Assuming $C_{3.3.2}$ is sufficiently large, the failure probability in stage i is less than $|V_i|^{-2}$. Thus, the total failure probability is less than

$$\sum_{i \ge 0} |V_i|^{-2} \le \sum_{k=C_{3.3.2}}^{\infty} k^{-2} \le 1/8.$$

Finally, note that in this procedure no edges in $G[X]$ are covered until the final step. Therefore the degree bound on $G[X]$ follows by noting that $G^*[X]$ is stochastically dominated by sampling every edge with probability $(\log|X|)^{-20}$. □

3.4.1 Fractional matching

In order to find the existence of fractional matchings within a sparse set of triangles we will use the following result of Barber, Glock, Kühn, Lo, Montgomery, and Osthus [4].

Lemma 3.4.2. *There exists an $\varepsilon = \varepsilon_0 > 0$ such that the following holds. Given a graph G on n vertices with minimum degree at least $(1-\varepsilon)n$, let $\mathcal{T}$ denote the set of triangles in*

G, and, for $e \in E(G)$, let $\mathcal{T}(e)$ be the set of edges containing e. There exists a function $\gamma : \mathcal{T} \to [0,1]$ such that $\sum_{T \in \mathcal{T}(e)} \gamma(T) = n/8$ for every e.

Remark 3.4.3. This follows from [4, Lemma 4.2], noting that any sufficiently dense graph is regular in the appropriate sense and letting $\gamma(\cdot) = \psi(\cdot) \cdot 1/(2p^2)$ where $\psi(\cdot)$ is defined as in [4, Lemma 4.2] and p is the density of G.

The crucial tool for our setting is that given such a fractional matching, one can subsample every triangle with weight proportional to γ and obtain a nearly perfect fractional matching inside the sampled hypergraph.

Lemma 3.4.4. *There exists an $\varepsilon = \varepsilon_0 > 0$ such that the following holds. Given a graph G on n vertices with minimum degree at least $(1-\varepsilon)n$ let $\mathcal{T}$ denote the set of triangles in G. Sample every triangle in G with probability p with $p \geq (\log n)^2/n$ and call this collection $\mathcal{H}$. Then with probability $1 - n^{-\omega(1)}$, there exists a triangle set $\mathcal{H}_1 \subseteq \mathcal{H}$ such that every edge is contained in $pn/8 \pm \sqrt{pn}\log n$ triangles of $\mathcal{H}_1$.*

Proof. Let $\gamma(\cdot)$ be as in Lemma 3.4.2. Let $\mathcal{H}_1$ be the random model where every triangle is sampled with probability $\gamma(T) \cdot p$ and note we can couple $\mathcal{H}_1 \subseteq \mathcal{H}$. The result then follows immediately from the Chernoff bound, noting that the expected number of triangles containing a given edge e is $\sum_{T \in \mathcal{T}(e)} p\gamma(T) = pn/8$. □

3.4.2 Covering process within regular triangle subset

We now show that we can cover most of the edges of an almost-complete graph using a sparse random triangle set. We will first require a set of notions with regards to hypergraph matchings. For a hypergraph $\mathcal{H}$, define

$$\Delta(\mathcal{H}) := \max_{v \in V(\mathcal{H})} \deg_{\mathcal{H}}(v), \ \Delta^{\mathrm{co}}(\mathcal{H}) := \max_{v_1, v_2 \in V(\mathcal{H})} \mathrm{codeg}_{\mathcal{H}}(v_1, v_2).$$

Call a function $\omega \colon E(\mathcal{H}) \to \mathbb{R}_{\geq 0}$ a *weight function*, and for $X \subseteq E(\mathcal{H})$ let $\omega(X) = \sum_{x \in X} \omega(x)$. We will require the following result of Ehard, Glock, and Joos [14] which guarantees the existence of hypergraph matchings which are pseudorandom with respect to a collection of weight functions.

Theorem 3.4.5 ([14, Theorem 1.2])**.** *Suppose $\delta \in (0,1)$ and $r \in \mathbb{N}$ with $r \geq 2$, and let $\varepsilon := \delta/(50r^2)$. Then there exists Δ_0 such that for all $\Delta \geq \Delta_0$ the following holds: Let $\mathcal{H}$ be an r-uniform hypergraph with $\Delta(\mathcal{H}) \leq \Delta$ and $\Delta^{\mathrm{co}}(\mathcal{H}) \leq \Delta^{1-\delta}$ as well as $e(\mathcal{H}) \leq \exp(\Delta^{\varepsilon^2})$. Suppose that $\mathcal{W}$ is a set of at most $\exp(\Delta^{\varepsilon^2})$ weight functions on $E(\mathcal{H})$. Then, there exists a matching $\mathcal{M}$ in $\mathcal{H}$ such that $\omega(\mathcal{M}) = (1 \pm \Delta^{-\varepsilon})\omega(E(\mathcal{H}))/\Delta$ for all $\omega \in \mathcal{W}$ with $\omega(E(\mathcal{H})) \geq \max_{e \in E(\mathcal{H})} \omega(e)\Delta^{1+\delta}$.*

This immediately implies the following lemma. We include the proof for completeness.

Lemma 3.4.6. *There exists $\varepsilon = \varepsilon_0 > 0$, $C = C_{3.4.6} > 0$ such that the following holds for sufficiently large n. Fix $p \in ((\log n)^C/n, 1)$, a graph G on n vertices with minimum degree at least $(1-\varepsilon_0)n$, and let $\mathcal{H}_1$ be a collection of triangles in G such that each edge is in $pn/8 \pm \sqrt{pn}\log n$ triangles. Then there exists a set of edge disjoint triangles $\mathcal{H}_2 \subseteq \mathcal{H}_1$ such that $\Delta(G \setminus E(\mathcal{H}_2)) \leq n/(\log n)^{1000}$.*

Proof. Define the auxiliary 3-uniform hypergraph $\mathcal{H}$ with vertices corresponding to the set of edges in G and 3-edges corresponding to the triangles in $\mathcal{H}_1$. Notice that $\Delta(\mathcal{H}) = pn/8 + O(\sqrt{pn}\log n)$. Furthermore as any pair of edges are contained in at most 1 triangle, it follows that $\Delta^{\text{co}}(\mathcal{H}) \leq 1$. Thus Theorem 3.4.5 applies with $\delta = 1/2$ and therefore $\varepsilon = 1/(900)$, and $\Delta = \Delta(\mathcal{H})$.

We now define the weight functions. For a vertex v, let w_v be 1 on all 3-edges of $\mathcal{H}$ corresponding to triangles containing v, and 0 elsewhere. Note that $w_v(E(\mathcal{H})) \geq pn^2/32 \geq \Delta(\mathcal{H})^{1+\delta}$. Furthermore if C is sufficiently large we have that $\exp(\Delta^{\varepsilon^2}) \geq n$ and thus there is a matching $\mathcal{M}$ in $\mathcal{H}$ such that

$$w_v(\mathcal{M}) \geq \frac{(1 \pm \Delta^{-\varepsilon})}{\Delta} \cdot \frac{(\Delta - O(\sqrt{pn}\log n))\deg_G(v)}{2} \geq \frac{(1-2\Delta^{-\varepsilon})\deg_G(v)}{2}$$

for all $v \in V(G)$. This implies that the matching, which corresponds to triangles of G, covers all but $2\Delta^{-\varepsilon}n$ edges incident to v. Taking C sufficiently large the result follows immediately. □

In the next three sections we prove Lemma 3.4.1. Our proof closely follows the proof of [4, Lemma 3.8], with the necessary adaptations to account for the random triangle set $\mathcal{H}'$ and the fact that $|V_1|/|V(G)|$ is relatively smaller in our setting than in [4].

3.4.3 Setup for iterative absorption

We are now in position to apply the results of Lemma 3.4.4 and Lemma 3.4.6. However, we cannot simply invoke these results on the whole graph G; a more delicate approach is required.

Recall that we have a graph $G \subseteq K_n$ with a distinguished vertex subset V_1. Let $q = (\log|V_1|)^{-30}$. Let R be a set of edges in $G[V(K_n) \setminus V_1, V_1]$ with the following properties:

(**A1**) For all $v \in V(G) \setminus V_1$, $\deg_R v = q|V_1| + O(q|V_1|(\log|V_1|)^{-1})$.

(**A2**) For all $v \in V_1$, $\deg_R v = qn + O(qn(\log n)^{-1})$.

(**A3**) For all $v \in V(G)\setminus V_1$, $v' \in V_1$, we have that $|N_R(v)\cap N_G(v')| = q|V_1| + O(q|V_1|(\log|V_1|)^{-1})$.

(**A4**) For all $v, v' \in V(G) \setminus V_1$, we have that $|N_R(v) \cap N_R(v')| \geq q^2|V_1|/2$.

(**A5**) For all $v, v' \in V_1$ we have that $|N_R(v) \cap N_R(v')| \leq 2q^2 n$.

That such a graph R exists is established by noting that if each edge in $G[V(K_n)\setminus V_1, V_1]$ is independently sampled with probability q then, by Chernoff's inequality and a union bound, these properties hold with positive probability.

Let $G_1 = G \setminus (R \cup G[V_1])$. It is easy to see that $\delta(G_1) \geq |V(G)| - O(|V(G)|/(\log |V(G)|)$. Let $\mathcal{T}$ denote the set of triangles in G_1 and sample each triangle in $\mathcal{T}$ with probability $p = (\log n)^{2C_{3.4.6}}/n$ to form the random set $\mathcal{H}''$. By Lemma 3.4.4, with probability $1 - n^{-\omega(1)}$ there exists a subset of triangles $\mathcal{H}_1 \subseteq \mathcal{H}''$ such that every edge of G_1 is in $pn/8 \pm \sqrt{pn}\log n$ triangles. Applying Lemma 3.4.6, we find that there exists a set of edge-disjoint triangles in $\mathcal{H}_1$ that covers all edges of G_1 except at most $n/(\log n)^{100}$ incident to each vertex. Let $L \subseteq G_1$ be the graph of uncovered edges. Let $L_1 \subseteq L$ be the "internal" edges with no vertex in V_1 and let $L_2 \coloneqq L \setminus L_1$ be the uncovered "crossing" edges with an endpoint in V_1. We remark that since G_1 contains no edges with both vertices in V_1, neither does L.

It remains to cover $G_2 \coloneqq L_1 \sqcup L_2 \sqcup R$ with triangles while not covering too many edges in $G[V_1]$. Let $R_2 \coloneqq L_2 \cup R$. Observe that R_2 satisfies:

(**B1**) For all $v \in V(G) \setminus V_1$, we have that $\deg_{R_2}(v) = q|V_1| + O(q|V_1|(\log |V_1|)^{-1})$.

(**B2**) For all $v \in V_1$, we have that $\deg_{R_2}(v) = qn + O(qn/\log n)$.

(**B3**) For all $v \in V(G)\setminus V_1$, $v' \in V_1$, we have that $|N_{R_2}(v) \cap N_G(v')| = q|V_1| + O(q|V_1|(\log |V_1|)^{-1})$.

(**B4**) For all $v, v' \in V(G) \setminus V_1$, we have that $|N_{R_2}(v) \cap N_{R_2}(v')| \geq q^2|V_1|/2$.

(**B5**) For all $v, v' \in V_1$ we have that $|N_{R_2}(v) \cap N_{R_2}(v')| \leq 3q^2 n$.

We complete the construction by first covering the internal edges that comprise L_1 and then covering the remaining crossing edges.

3.4.4 Cover-down stage 1: internal edges

Lemma 3.4.7. *With the above setup, let $\mathcal{T}_2$ denote the set of triangles in G_2 and let $\mathcal{H}_3 \subseteq \mathcal{T}_2$ be a random set of triangles with each triangle included with probability $(\log n)^{100}/n$. Then with probability $1 - n^{-\omega(1)}$, one can choose edge disjoint triangles $\mathcal{H}_4 \subseteq \mathcal{H}_3$ such that $L_1 \subseteq E(\mathcal{H}_4)$.*

Proof. We construct $\mathcal{H}_4$ with a random greedy algorithm. Order the edges in L_1 arbitrarily. When processing an edge e, expose the triangles of $\mathcal{H}_3$ containing e and then choose one such triangle, not overlapping with previous choices, uniformly at random and add it to $\mathcal{H}_4$. This procedure only fails if for some e, all triangles containing it in $\mathcal{H}_3$ overlap previous choices. However note that initially each edge in L_1 is contained in at least $q^2|V_1|/2$ triangles in $\mathcal{T}_2$, and that triangles added previously to $\mathcal{H}_4$ eliminate at most $2n/(\log n)^{100} \leq q^2|V_1|/4$ of these. Thus the expected number of extensions at each stage is at least $q^2|V_1|/4 \cdot (\log n)^{100}/n \gtrsim (\log n)^{50}$. Applying Chernoff's inequality and a union bound, with probability $1 - n^{-\omega(1)}$ there is at least one choice for each stage. □

Given Lemma 3.4.7, the remaining graph to cover is $R_3 := R_2 \setminus E(\mathcal{H}_4)$. We note that since every triangle in $\mathcal{H}_4$ involves an edge of L_1, we have:

(**C1**) For all $v \in V(G) \setminus V_1$, we have that $\deg_{R_3}(v) = q|V_1| + O(q|V_1|/\log|V_1|)$.

(**C2**) For all $v \in V_1$, we have that $\deg_{R_3}(v) = qn + O(qn/\log n)$.

(**C3**) For all $v \in V(G)\setminus V_1$, $v' \in V_1$, we have that $|N_{R_3}(v) \cap N_G(v')| = q|V_1| + O(q|V_1|/\log|V_1|)$.

(**C4**) For all $v, v' \in V_1$, we have that $|N_{R_3}(v) \cap N_{R_3}(v')| \le 3q^2 n$.

3.4.5 Cover-down stage 2: crossing edges

Our goal is to cover R_3 using only a small number of edges from $G[V_1]$. This will be accomplished by reducing the problem to a simultaneous matching problem on link graphs of vertices in $V(G) \setminus V_1$.

We first require the following lemma. It is an immediate consequence of the (substantially stronger) main results in [46, 61]. We include an elementary proof for completeness.

Lemma 3.4.8. *Let G' be a graph on N vertices, with N even, and with minimum degree at least $3N/4$. Let H be a random subgraph of G' where each edge is sampled independently with probability $(\log N)^2/N$. Then H has a perfect matching with probability $1 - N^{-\omega(1)}$.*

The proof of Lemma 3.4.8 uses the following convenient Hall-type criterion for a bipartite graph to have a perfect matching. It is an immediate consequence of the main theorem in [93].

Lemma 3.4.9. *Let $G' = (X \cup Y, E)$ be a bipartite graph with $|X| = |Y| = N$. Suppose that for every $S \subseteq X$, $S' \subseteq Y$ with $|S'| < |S| \le \lceil N/2 \rceil$ we have $e(S, Y \setminus S') \neq 0$, and that for every $T' \subseteq X$, $T \subseteq Y$ with $|T'| < |T| \le \lceil N/2 \rceil$ we have $e(T, X \setminus T') \neq 0$. Then G has a perfect matching.*

Proof of Lemma 3.4.8. Consider a uniformly random equipartition $X \cup Y$ of $V(G')$ and let $G^\dagger := G'[X, Y]$. By the Chernoff bound for hypergeometric random variables and a union bound with probability $1 - N^{-\omega(1)}$ we have $\deg_{G^\dagger}(v) \ge N/3$ for each vertex v. Now consider some $S \subseteq X, S' \subseteq Y$ satisfying $\lceil N/4 \rceil \ge |S| > |S'|$. It holds that

$$e_{G^\dagger}(S, Y \setminus S') = \sum_{v \in S} \deg_{G^\dagger}(v) - e_{G'}(S, S') \ge (N/3)|S| - |S|^2 \ge N|S|/12.$$

Similarly, if $T' \subseteq X, T \subseteq Y$ with $|T'| < |T| \le \lceil N/4 \rceil$ then $e_{G^\dagger}(T, X \setminus T') \ge N|S|/12$.

We observe that with probability at least $1 - \exp(-\Omega((\log N)^2))$, we have $e_H(S, Y \setminus S') > 0$ for every pair of sets $S \subseteq X, S' \subseteq Y$ with $\lceil N/4 \rceil \ge |S| > |S'|$. Indeed, by a union bound over S, S' and the Chernoff bound, the probability that this fails to hold is at most

$$\sum_{k=1}^{\lceil N/4 \rceil} \binom{N/2}{k} \sum_{\ell=0}^{k-1} \binom{N/2}{\ell} \exp(-\Omega(k(\log N)^2)) \le \sum_{k=1}^{\lceil N/4 \rceil} N^{2k} \exp(-\Omega(k(\log N)^2)))$$

$$\leq \exp(-\Omega((\log N)^2)).$$

By symmetry, the same is true (with probability at least $1 - \exp(-\Omega((\log N)^2))$ when switching the roles of X and Y. The desired result then follows from Lemma 3.4.9. □

Lemma 3.4.10. *Fix R_3 as in Section 3.4.4 and suppose that it satisfies (C1)-(C4). Let $\mathcal{T}_3$ be the set of triangles in $R_3 \cup G[V_1]$ and let $\mathcal{H}_5 \subseteq \mathcal{T}_3$ be a random set of triangles with each triangle included with probability $(\log |V_1|)^2/(q|V_1|)$. We then have:*

- *For each edge in $G[V_1]$, the probability that it is contained in a triangle in $\mathcal{H}_5$ is at most $(\log n)^{-25}$. Moreover, these events are mutually independent.*
- *With probability $1 - n^{-\omega(1)}$, there exists a set of edge disjoint triangles $\mathcal{H}_6 \subseteq \mathcal{H}_5$ such that $R_3 \subseteq E(\mathcal{H}_6)$.*

Proof. For the first point, recall that by **(C4)** every pair of distinct $u, v \in V_1$ has at most $3q^2n$ common neighbors in R_3. Thus, the probability for an edge to be contained in a triangle in $\mathcal{H}_5$ is at most $3q^2n \cdot (\log |V_1|)^2/(q|V_1|) \leq (\log n)^{-25}$. Moreover, for distinct edges in $G[V_1]$, their extensions into triangles in $\mathcal{H}_5$ are disjoint, implying mutual independence of these events.

For the second point, order the vertices in $V(G) \setminus V_1$ and consider them sequentially. Suppose we are processing v and note first that by **(C1)** the graph spanned by the triangles containing v in $R_3 \cup G[V_1]$ has $q|V_1| + O(q|V_1|(\log |V_1|)^{-1})$ vertices. Furthermore by **(C3)** its link graph (i.e., the subgraph spanned by V_1)has minimum degree $q|V_1| + O(q|V_1|(\log |V_1|)^{-1})$. Our goal is to find a perfect matching in this link graph that is edge-disjoint from previously found perfect matchings. These perfect matchings correspond to the desired $\mathcal{H}_6$.

At every step the set of edges removed is stochastically dominated by our random sample of edges at rate $(\log n)^{-25}$, so we find that the minimum degree in this link is essentially unchanged by the previous triangles removed in this process. Therefore if one samples each triangle in the link with probability $(\log |V_1|)^2/(q|V_1|)$, or equivalently each edge in the link with probability $(\log |V_1|)^2/(q|V_1|)$, by Lemma 3.4.8 we can construct a matching for v with probability $1 - n^{-\omega(1)}$. This immediately gives the desired result. □

We are now in position to prove Lemma 3.4.1.

Proof of Lemma 3.4.1. Using the construction above, let $\mathcal{H}^* := \mathcal{H}_2 \cup \mathcal{H}_4 \cup \mathcal{H}_6$. Observe that $\mathcal{H}^*$ is stochastically dominated by a random hypergraph of the specified density. Furthermore, if R_3 is suitable in the sense of Section 3.4.4 then $E(\mathcal{H}^*)[V_1]$ is stochastically dominated by a random hypergraph of the appropriate density as well. Hence, we may take $\mathcal{H} = \mathcal{H}^*$ if R_3 is suitable. Otherwise we take $\mathcal{H} = \emptyset$. Since R_3 is suitable with probability $1 - n^{-\omega(1)}$ and $\mathcal{H}^*$ covers all edges in $G \setminus G[V_1]$ with probability $1 - n^{-\omega(1)}$, the result follows. □

3.5 Modifications for Latin squares

In this section we briefly discuss the necessary changes to prove Theorem 3.1.3, as opposed to Theorem 3.1.1. Since these changes are largely superficial, we do not repeat the arguments in detail.

A Latin square can be thought of as a triangle-decomposition of $K_{n,n,n}$, with vertex parts V^1, V^2, V^3, each of size n. We say that a tripartite subgraph of $K_{n,n,n}$ is *triangle-divisible* if for every $j \in [3]$ and vertex $v \in V^j$, its degrees into V^{j-1} and V^{j+1} are the same (taking indices mod3). The analogue of Theorem 3.3.1 is the following:

Theorem 3.5.1. *Fix a triangle-divisible tripartite graph $G \subseteq K_{n,n,n}$ with $\Delta(K_{n,n,n} \setminus G) \leq n/\log n$ and $n \geq \exp(C_{3.5.1}/\eta^4)$. Let $\mathcal{H}$ be the result of randomly sampling each triangle of $K_{n,n,n}$ with probability n^η/n. With probability at least $1/2$ the collection $\mathcal{H}$ contains a triangle-decomposition of G.*

The proof strategy is similar, and we detail the necessary changes.

- To prove an analogue of Proposition 3.3.2, the vortex $V(K_{N,N,N}) = V_0 \supseteq \cdots \supseteq V_\ell = X$ should be chosen so that each V_k has the same number of vertices in each V^j.
- During the iteration, replace the various degree typicality assumptions (e.g. **(C1)** to **(C4)**) with the obvious tripartite analogues.
- In the final step of the cover-down procedure (Section 3.4.5), in the original setup we reduced to a bipartite matching problem by taking a random bipartition of U_{i+1}. In the Latin square setting this is not necessary since the bipartite structure is already induced by $K_{n,n,n}$.
- The existence of the regular triangle subset (Lemma 3.4.2) relies on weight-shifting gadgets which are not tripartite. It is possible to adapt work of Montgomery [74] to obtain a suitable approximate tripartite fractional matching result; see e.g. [65, Lemma 8.11].
- The absorbing structures used in **Alg6** and **Alg7** within the proof of Theorem 3.3.1 must be tripartite. However, this is not an obstruction since the vertices of $\mathcal{F}_{2m}$ can be split into three classes so that all hyperedges are tripartite.

Chapter 4

Improved bounds for Szemerédi's theorem

4.1 Introduction

Let $[N] = \{1, \dots, N\}$ and $r_k(N)$ denote the size of the largest $S \subseteq [N]$ such that S has no k-term arithmetic progressions. The first nontrivial upper bound on $r_3(N)$ came from work of Roth [84] which proved

$$r_3(N) \ll N(\log\log N)^{-1}.$$

A long series of works improved this bound, including works of Heath-Brown [35], Szemerédi [98], Bourgain [11,12], Sanders [88,89], Bloom [7], and Bloom and Sisask [8]. In breakthrough work, Kelley and Meka [57] very recently proved

$$r_3(N) \ll N \exp(-c(\log N)^{1/12});$$

the constant $1/12$ was refined to $1/9$ in work of Bloom and Sisask [9].

For higher k, a long-standing conjecture of Erdős and Turán stated that $r_k(N) = o(N)$. In seminal works, Szemerédi [96, 97] first established the estimate $r_4(N) = o(N)$ and then established his eponymous theorem that

$$r_k(N) = o(N).$$

Due to uses of van der Waerden theorem and the regularity lemma (which was introduced in this work), Szemerédi's density saving was exceedingly small. In breakthrough work, Gowers [25, 26] introduced higher order Fourier analysis and proved the first "reasonable" bounds for Szemerédi's theorem:

$$r_k(N) < N(\log\log N)^{-2^{-2^{k+9}}}.$$

The only improvement to this result for $k \geq 4$ was work of Green and Tao [28, 32] which ultimately established that

$$r_4(N) \ll N(\log N)^{-c},$$

and recent work of the authors [67] which proved

$$r_5(N) \ll N \exp(-(\log\log N)^c).$$

Our main result is an extension of this bound for all $k \geq 5$.

Theorem 4.1.1. *Fix $k \geq 5$. There is $c_k \in (0,1)$ such that*

$$r_k(N) \ll N \exp(-(\log\log N)^{c_k}).$$

4.1.1 Proof outline and techniques

4.1.1.1 Local and global inverse theorems

The primary input to our result will be the main result of recent work of the authors [69], i.e., quasipolynomial bounds on the inverse theorem for the Gowers U^{k+1}-norm. Given an inverse theorem, the deduction of Szemerédi's theorem via a standard density increment strategy is essentially folklore and was recorded in work of Green and Tao [30] (although, prior to [69] the resulting bounds would be far from matching those of Gowers [26]). However, if one naively follows this script using [69], one obtains a bound of $N \exp(-(\log\log\log N)^{-\Omega_k(1)})$ which is weaker than the work of Gowers [26]. Furthermore, Gowers's argument makes use of a "local" inverse theorem that in fact gives a slightly stronger correlation compared to the bound given for the "global" inverse theorem in [69] (namely, polynomial versus quasipolynomial). Thus, this global nature of [69] must be exploited. Additionally, use of global inverse theorems necessitates understanding of nilsequences and polynomial sequences on nilpotent Lie groups, as opposed to merely polynomials as in the work of Gowers [26].

4.1.1.2 Schmidt-type decomposition problems

This is done via the improved density increment strategy of Heath-Brown [35] and Szemerédi [98] which involves extracting a set of functions to correlate with instead of simply one and using this to give a multiplicative density increment. Such a strategy was given a robust formulation in work of Green and Tao [28] on four-term progressions; in particular, their reformulation avoided the explicit Fourier-analytic formulas used in [35, 98] and thus is applicable to the higher order setting. The strategy here runs smoothly given the inverse theorem, modulo resolving a certain Schmidt-type problem for nilsequences. In particular, given a polynomial sequence $g(n)$ with $g(0) = \mathrm{id}_G$ on a nilmanifold G/Γ of degree k with complexity M and dimension d, one needs to prove that

$$\min_{1 \leq n \leq N} d_{G/\Gamma}(\mathrm{id}_G, g(n)\Gamma) \ll M^{O_k(d^{O_k(1)})} N^{-1/d^{O_k(1)}}.$$

In particular, the polynomial dependence on dimension within the exponent is key.

We in fact require a certain slightly stronger result (decomposing $[N]$ into long arithmetic progressions P such that the diameters of the sets $\{g(n)\Gamma : n \in P\}$ are small), which is the

heart of the matter for this work. When the underlying nilpotent group G is abelian, this is easily deduced from a result of Schmidt [92] (see Lemma 4.2.3, or [28, Section 6] in the quadratic case).

For general degree 2 nilmanifolds such a problem was implicitly solved in work of Green and Tao [28] and for degree 3 nilmanifolds it was essentially solved in recent work of the authors [67]. More precisely, [67] essentially proves that given a list of bracket expressions $(a_i n\lfloor b_i n\rfloor\lfloor c_i n\rfloor)_{1\leq i\leq d}$ that

$$\min_{1\leq n\leq N}\|a_i n\lfloor b_i n\rfloor\lfloor c_i n\rfloor\|_{\mathbb{R}/\mathbb{Z}}\leq N^{-1/d^{O(1)}}$$

and via an explicit computation with fundamental domains on degree 3 nilmanifolds one may reduce to such a situation. The proof given in [67] relies on the fact that 3 is sufficiently small and in particular that it is possible to reduce to a situation in which there are no "nested integer part operations" as one attempts to solve the "bracket Schmidt" problem in one go.

4.1.1.3 Iterative Schmidt refinement

The key observation required for our work, at least at a heuristic level, is a procedure for solving such "bracket Schmidt" problems even when there are nested brackets. As a simple example, consider bracket expressions $(a_i n\lfloor b_i n\lfloor c_i n\rfloor\rfloor)_{1\leq i\leq d}$. We will solve the Schmidt problem via iteratively "reducing" the number of brackets from the inside-out (at the cost of passing to subprogressions). In particular, using Dirichlet's theorem, one can break $[N]$ into arithmetic progressions P each of length $N^{1/d^{O(1)}}$ such that when restricted to each arithmetic progression, every function $\lfloor c_i n\rfloor$ is a linear function (i.e., it is a *locally linear function* on each P). Since the only locally linear functions on a progression agree with genuinely linear functions, we can replace $\lfloor c_i n\rfloor$ by $d_{i,P}n+e_{i,P}$ and reduce to considering the bracket expression $a_i n\lfloor b_i n(d_{i,P}n+e_{i,P})\rfloor$ when restricted to P. One can then iterate this argument on the "inner quadratics" $b_i n(d_{i,P}n+e_{i,P})$ (essentially using abelian Schmidt as discussed above for degree 2 in this case). We may find a decomposition into long arithmetic progressions Q such that $\lfloor b_i n(d_{i,P}n+e_{i,P})\rfloor$ is locally quadratic (and hence agrees with a global quadratic) on each Q. Thus, restricted to any such Q, our original functions $a_i n\lfloor b_i n\lfloor c_i n\rfloor\rfloor$ agrees with a genuine cubic. Finally, we can decompose these progressions Q into ones where the cubics are approximately constant mod 1 (using abelian Schmidt for degree 3). While in theory this approach can be made to work for all such bracket Schmidt problems, this however necessitates working with bracket functions and rather quickly becomes messy to handle.

This procedure can be adapted to work with polynomial sequences on nilmanifolds directly due to an unpublished observation of Green and Tao. This is the approach we take in the present work. The crucial point is that given a polynomial sequence $g(n)$ with respect to a group G given a filtration $G_0=G_1\geqslant G_2\geqslant\cdots\geqslant G_k\geqslant \mathrm{Id}_G$, the polynomial sequence $g(n)$ mod G_2 is a standard polynomial. Thus one can apply Schmidt to a standard polynomial and therefore (after passing to long subprogressions) one may factor $g(n)$ mod $G_2=\varepsilon(n)\cdot\gamma(n)$ where ε is smooth and γ lies in the lattice Γ mod G_2. One may then lift ε,γ from G mod G_2

to $\widetilde{\varepsilon}, \widetilde{\gamma}$ on G and analyze the polynomial sequence $\widetilde{\varepsilon}^{-1} g \widetilde{\gamma}^{-1}$ which now lives in the group G_2. One can iterate this procedure and inductively reduce G_2 to G_3 and so on, which allows us to solve the Schmidt problem for our nilmanifold. We remark that this procedure is an induction on the length of the filtration whereas the (closely related) approach taken in [30] is phrased as an induction on dimension. This difference is crucial for getting bounds in which the exponent depends polynomially on dimension.

4.1.2 Organization and notation

All definitions regarding nilsequences and associated complexity will be exactly as in [69, Sections 3–4]. We refer the reader to that paper for all such definitions; we will only require degree filtrations in this paper.

We use standard asymptotic notation. Given functions $f = f(n)$ and $g = g(n)$, we write $f = O(g)$, $f \ll g$, $g = \Omega(f)$, or $g \gg f$ to mean that there is a constant C such that $|f(n)| \leq Cg(n)$ for sufficiently large n. We write $f \asymp g$ or $f = \Theta(g)$ to mean that $f \ll g$ and $g \ll f$, and write $f = o(g)$ or $g = \omega(f)$ to mean $f(n)/g(n) \to 0$ as $n \to \infty$. Subscripts on asymptotic notation indicate dependence of the bounds on those parameters. We will use the notation $[x] = \{1, 2 \ldots, \lfloor x \rfloor\}$. In this paper $x = \lfloor x \rfloor + \{x\}$ where $\{x\} \in [0, 1)$ and $\lfloor x \rfloor \in \mathbb{Z}$; we remark this is different than in [69]. We write $\|x\|_{\mathbb{R}/\mathbb{Z}} = \operatorname{dist}(x, \mathbb{Z})$ for $x \in \mathbb{R}$. Furthermore throughout this paper we abusively write log for $\max(\log(\cdot), e^e)$; this is to avoid trivial issues with small numbers.

Finally, in terms of organization, in Section 4.2 we solve the Schmidt problem for nilsequences and in Section 4.3 we prove Theorem 4.1.1.

Acknowledgments

The third author thanks Mark Sellke and Dmitrii Zakharov for helpful and motivating conversations. We thank Ben Green for helpful comments on the manuscript. We thank Zach Hunter for various minor corrections.

4.2 Schmidt's problem for nilsequences

In this section, we prove that given a list of nilsequences on $[N]$, one can decompose $[N]$ into a controlled set of arithmetic progressions such that the nilsequences are almost constant on these sequences.

Lemma 4.2.1. *Consider nilmanifolds G_i/Γ_i for $1 \leq i \leq T$, each given a degree k filtration, having complexity bounded by M, dimension bounded by d, and for each $1 \leq i \leq T$ let $g_i(n)$ be a polynomial sequence with respect to the specified degree k-filtration on G_i.*

We may decompose $[N]$ into disjoint arithmetic progressions $\mathcal{P}_1, \ldots, \mathcal{P}_L$ such that following conditions hold:

- $N/L \geq N^{\Omega_k(1/(Td)^{O_k(1)})}/2$;

- *We have*

$$\max_{\substack{1\le i\le T\\ 1\le j\le L}} \max_{n,n'\in\mathcal{P}_j} d_{G_i/\Gamma_i}(g_i(n)\Gamma_i, g_i(n')\Gamma_i) \le M^{O_k(d^{O_k(1)})}\cdot N^{-\Omega_k(1/(Td)^{O_k(1)})}.$$

The key ingredient in this proof is a result of Schmidt [92] regarding finding small fractional parts of polynomials. We will need a version of this result with explicit quantification; this is explicitly stated in work of the authors [67, Proposition 3.7] although the argument is essentially verbatim from a paper of Green and Tao [28, Appendix A] generalized from quadratics to all degrees.

Proposition 4.2.2. *Fix an integer $k \ge 1$. There exist $c_k > 0$ such that the following holds. Let $\alpha_1, \dots, \alpha_d$ be real numbers. Then*

$$\min_{1\le n\le N} \max_{1\le i\le d} \|\alpha_i n^k\|_{\mathbb{R}/\mathbb{Z}} \ll_k dN^{-c_k/d^2}.$$

As stated this result is for pure monomial phases and only provides a single point with small fractional part. This statement however can be "upgraded" via a straightforward iterative argument which is implicit in say [28, Proposition 6.4] (where the quadratic case is handled).

Lemma 4.2.3. *Fix an integer $k \ge 0$. Consider polynomials $Q_1, \dots, Q_d$ of degree k. Then there exist disjoint arithmetic progressions $\mathcal{P}_1, \dots, \mathcal{P}_L$ such that following conditions hold:*

- $N/L \ge N^{\Omega_k(1/d^{O_k(1)})}/2$
- *We have*

$$\max_{\substack{1\le i\le d\\ 1\le j\le L}} \max_{n,n'\in\mathcal{P}_j} \|Q_i(n) - Q_i(n')\|_{\mathbb{R}/\mathbb{Z}} \le 2\cdot N^{-\Omega_k(1/d^{O_k(1)})}.$$

Proof. We proceed by induction on k. The case $k = 0$ is trivial as $Q_j(\cdot)$ are constant. Furthermore we may assume that $N \ge \exp(d^{\Omega_k(1)})$ else we may break $[N]$ into singleton arithmetic progressions.

Let $Q_j(n) = \sum_{\ell=0}^k \alpha_{j,\ell} n^\ell$. Applying Proposition 4.2.2, there exists $D \le N^{1/2}$ such that

$$\max_{1\le j\le d} \|\alpha_{j,k} D^k\|_{\mathbb{R}/\mathbb{Z}} \ll_k dN^{-c_k/(2d^2)} =: \tau.$$

We break $[N]$ into arithmetic progressions of common difference D and with lengths between $2^{-1}\tau^{-1/(2k)}$ and $\tau^{-1/(2k)}$. Label these progressions $\mathcal{R}_1, \dots, \mathcal{R}_{L'}$ with starting points s_i for $1 \le i \le L'$. We have

$$Q_j(Dn + s_i) = \alpha_{j,k} D^k n^k + Q_{j,i}(n)$$

for appropriately defined polynomials $Q_{j,i}(n)$ of degree at most $k-1$. Note that for $n, n' \in [\tau^{-1/(2k)}]$, we have

$$\|Q_j(Dn + s_i) - Q_j(Dn' + s_i)\|_{\mathbb{R}/\mathbb{Z}} = \|\alpha_{j,k} D^k (n^k - (n')^k) + Q_{j,i}(n) - Q_{j,i}(n')\|_{\mathbb{R}/\mathbb{Z}}$$

$$\leq 2\tau^{-1/2} \cdot \|\alpha_{j,k} D^k\|_{\mathbb{R}/\mathbb{Z}} + \|Q_{j,i}(n) - Q_{j,i}(n')\|_{\mathbb{R}/\mathbb{Z}}$$
$$\leq 2\tau^{1/2} + \|Q_{j,i}(n) - Q_{j,i}(n')\|_{\mathbb{R}/\mathbb{Z}}.$$

The result now follows by induction applied to each $Q_{j,i}(n)$ for $1 \leq i \leq L'$ on the interval $[\tau^{-1/(2k)}]$ and using these decompositions to split the $\mathcal{R}_i$ into our final decomposition. Letting $N' = \tau^{-1/(2k)}$, the number of arithmetic progressions resulting is bounded by

$$(2N/N') \cdot 2(N')^{1-c_1/d^{c_2}} \leq 2N^{1-\Omega_k(1/d^{O_k(1)})},$$

where c_1, c_2 are the implicit constants for the inductive hypothesis $k-1$. The result follows. □

We next require the following lemma controlling coefficients of polynomials which live in a restricted range mod 1. It will be convenient to recall the smoothness norm of a polynomial $P(n) = \sum_{i=0}^{k} \alpha_i \binom{n}{i}$ which is defined as

$$\|P\|_{C^\infty[N]} := \max_{1 \leq i \leq k} N^i \|\alpha_i\|_{\mathbb{R}/\mathbb{Z}}.$$

Lemma 4.2.4. *Fix an integer $k \geq 1$. There exists $c_k > 0$ such that if $\varepsilon \in (0, c_k)$ and $N \geq c_k^{-1}$ then the following holds. Consider a polynomial $P(n) = \sum_{i=0}^{k} \alpha_i \binom{n}{i}$. Suppose that for $n, n' \in [N]$, we have $\|P(n) - P(n')\|_{\mathbb{R}/\mathbb{Z}} \leq \varepsilon$. Then*

$$\|P\|_{C^\infty[N]} \ll_k \varepsilon.$$

Proof. Note that

$$\left| \sum_{n=1}^{N} e(P(n)) \right| \geq N/2.$$

By a quantitative version of Weyl's inequality, which may be found in Green and Tao [31, Proposition 4.3], there exists $q \in \mathbb{N}$ with $q \ll_k 1$ such that

$$\|qP\|_{C^\infty[N]} \ll_k 1.$$

Let $1 \leq t \leq \lfloor N/(2k) \rfloor$ be an integer and note that

$$\alpha_k \cdot t^k = \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} \cdot P(t \cdot i + 1).$$

Via the triangle inequality, we therefore have

$$\|\alpha_k \cdot t^k\|_{\mathbb{R}/\mathbb{Z}} \leq 2^{k-1}\varepsilon.$$

Take t to be a prime between $\lfloor N/(2C) \rfloor$ and $\lfloor N/C \rfloor$ where C is a sufficiently large absolute constant in terms of k. Combining this with the estimate $\|q\alpha_k\|_{\mathbb{R}/\mathbb{Z}} \ll_k N^{-k}$ implies that $\|\alpha_k\|_{\mathbb{R}/\mathbb{Z}} \ll_k \varepsilon \cdot N^{-k}$. The result then follows by induction on k and applying the result for the degree $(k-1)$ polynomial $P'(n) = \sum_{i=0}^{k-1} \alpha_i \binom{n}{i}$. □

With this we are in position to deduce the result for nilsequences along the lines sketched in Section 4.1.1.3.

Proof of Lemma 4.2.1. Consider the degree k filtration of the group G_i, $G_{i,0} = G_{i,1} \geqslant G_{i,2} \geqslant \cdots \geqslant G_{i,k} \geqslant \mathrm{Id}_{G_i}$. We say the group G_i has a degree k filtration of type t if $G_{i,t} = G_i$ (i.e., the first $(t+1)$ groups in the filtration match). We prove the result by backwards induction on t assuming that all groups G_i have degree k filtrations of type t; note that the result is trivial when $t = k+1$ and we aim to prove the claim when $t = 1$. So, consider the case where the filtration has type t for some $1 \leq t \leq k$ and suppose that we already know cases of larger type.

Let $\mathcal{X}_i = \{X_{i,1}, \ldots, X_{i,\dim(G_i)}\}$ denote the Mal'cev basis for G_i. By the classification of polynomial sequences (see [29, Lemma 6.7]), we have

$$g_i(n) = \exp\left(\sum_{j=1}^{\dim(G_i)} P_{i,j}(n) \cdot X_{i,j}\right)$$

where if $X_{i,j} \in (\mathcal{X}_i \cap \log(G_{i,\ell})) \setminus (\mathcal{X}_i \cap \log(G_{i,\ell+1}))$ then the degree of polynomial $P_{i,j}(n)$ is bounded by ℓ.

We consider the polynomials $P_{i,j}(n)$ for $1 \leq i \leq T$ and $1 \leq j \leq \dim(G_i) - \dim(G_{i,t+1})$. The degrees of $P_{i,j}(n)$ are all at most $t \leq k$ and the total number of polynomials number consideration is bounded by $T \cdot d$. By applying Lemma 4.2.3, there exists a decomposition of $[N]$ into arithmetic progressions $\mathcal{P}_1, \ldots, \mathcal{P}_L$ such that:

- $N/L \geq N^{\Omega_k(1/(dT)^{O_k(1)})}/2$
- We have

$$\max_{\substack{1\leq i\leq T\\ 1\leq j\leq \dim(G_i)-\dim(G_{i,t+1})}} \max_{1\leq s\leq L} \max_{n,n'\in\mathcal{P}_s} \|P_{i,j}(n) - P_{i,j}(n')\|_{\mathbb{R}/\mathbb{Z}} \leq 2 \cdot N^{-\Omega_k(1/(dT)^{O_k(1)})}.$$

We break the progressions $\mathcal{P}_s$ into two classes: the first class ($s \in \mathcal{S}$) if the progression has length bounded by $\sqrt{N/L}$ and the second class ($s \in \mathcal{L}$) otherwise. For progressions which are short, we break each such progression into singletons; after this there are at most $L + \sqrt{N/L} \cdot L \leq 2\sqrt{NL}$ progressions which is qualitatively identical to before. For each $s \in \mathcal{L}$, we write $\mathcal{P}_s = \{a_s n + b_s\}_{n \in [|\mathcal{P}_s|]}$ where $|\mathcal{P}_s|$ denotes the length of the progression.

Using the second condition above and applying Lemma 4.2.4, we see that for each long progression $\mathcal{P}_s$, we have for all i, j that

$$P_{i,j}(a_s n + b_s) = P_{i,j,s,\mathrm{small}}(n) + P_{i,j,s,\mathrm{int}}(n)$$

where:

- $\deg(P_{i,j,s,\mathrm{int}}), \deg(P_{i,j,s,\mathrm{small}}) \leq \deg(P_{i,j})$
- $P_{i,j,s,\mathrm{int}}$ maps $\mathbb{Z} \to \mathbb{Z}$

- If $P_{i,j,s,\mathrm{small}}(n) = \sum_{r=0}^{t} \alpha_{i,j,s,\mathrm{small},r} \binom{n}{r}$ then

$$|\alpha_{i,j,s,\mathrm{small},r}| \leq 2N^{-r} \cdot N^{-\Omega_k(1/(dT)^{O_k(1)})}$$

  for $1 \leq r \leq t$ and $|\alpha_{i,j,s,\mathrm{small},0}| \leq 1$.

We have implicitly used $|\mathcal{P}_s| \geq \sqrt{N/L} \geq N^{\Omega_k(1/(dT)^{O_k(1)})}$ for $s \in \mathcal{L}$ here.

The key trick is to now "reduce" the polynomial sequence g_i to one which lives in $G_{i,t+1}$. Define

- $\varepsilon_{i,s}(n) = \exp\left(\sum_{j=1}^{\dim(G_i)-\dim(G_{i,t+1})} P_{i,j,s,\mathrm{small}}(n) \cdot X_{i,j}\right)$
- $\gamma_{i,s}(n) = \prod_{j=1}^{\dim(G_i)-\dim(G_{i,t+1})} \exp(X_{i,j})^{P_{i,j,s,\mathrm{int}}(n)}$
- $g'_{i,s}(n) = \varepsilon_{i,s}(n)^{-1} \cdot g(a_s n + b_s) \cdot \gamma_{i,s}(n)^{-1}$

Note that $\varepsilon_{i,s}$, $\gamma_{i,s}$ are polynomial sequences with respect to the filtration given on G_i by the classification of polynomial sequences (see [29, Lemma 6.7]) and the fact that the set of polynomial sequences form a group. Therefore $g'_{i,s}$ is also seen to be a polynomial sequence. The crucial point, however, is that by the Baker–Campbell–Hausdorff formula, we have that $g'_{i,s}$ only takes on values in $G_{i,t+1}$. (We are using the assumption on type that $G_i = G_{0,i} = G_{t,i}$, so any commutator is in $G_{2t,i} \leqslant G_{t+1,i}$ since $t \geq 1$.)

Therefore we may inductively apply the claim for each long progression $\mathcal{P}_s$, to the polynomials $g'_{i,s}$ on $G_{i,t+1}$ where we take the filtration on G_i intersected with $G_{i,t+1}$ (note that the filtration is still degree k). The corresponding Mal'cev basis is given by taking the last $\dim(G_{i,t+1})$ elements of $\mathcal{X}_i$. By induction therefore we may break each long $\mathcal{P}_s$ into L_s such progressions $\mathcal{P}_{s,r}$ where $L_s \leq |\mathcal{P}_s|^{1-\Omega_k(1/(Td)^{O_k(1)})}$ and such that

$$\max_{\substack{s\in\mathcal{L}\\1\leq r\leq L_s}} \max_{n,n'\in\mathcal{P}_{s,r}} d_{G_i/\Gamma_i}(g'_{i,s}(n)\Gamma_i, g'_{i,s}(n')\Gamma_i) \leq M^{O_k(d^{O_k(1)})} \cdot N^{-\Omega_k(1/d^{O_k(1)})}.$$

Here we are using [66, Lemma B.9] to compare distances between G_i and $G_{i,t+1}$.

Furthermore note that $\gamma_{i,s}$ takes values only in Γ by the definition of a Mal'cev basis and that for $n, n' \in [|\mathcal{P}_s|]$ we have

$$d_{G_i}(\varepsilon_{i,s}(n), \mathrm{id}_{G_i}) \leq M^{O_k(d^{O_k(1)})} \text{ and } d_{G_i}(\varepsilon_{i,s}(n), \varepsilon_{i,s}(n')) \leq M^{O_k(d^{O_k(1)})} \cdot N^{-\Omega_k(1/(dT)^{O_k(1)})}.$$

This is due to our bounds on the smoothness norm of $P_{i,j,s,\mathrm{small}}$ and [66, Lemma B.3].

It therefore follows by [66, Lemma B.4] that for any s, r we have

$$\begin{aligned}
&\max_{n,n'\in\mathcal{P}_{s,r}} d_{G_i/\Gamma_i}(g_i(a_s n + b_s)\Gamma_i, g_i(a_s n' + b_s)\Gamma_i) \\
&\quad = \max_{n,n'\in\mathcal{P}_{s,r}} d_{G_i/\Gamma_i}(\varepsilon_{i,s}(n)g'_{i,s}(n)\Gamma_i, \varepsilon_{i,s}(n')g'_{i,s}(n')\Gamma_i) \\
&\quad \leq \max_{n,n'\in\mathcal{P}_{s,r}} d_{G_i/\Gamma_i}(\varepsilon_{i,s}(n)g'_{i,s}(n)\Gamma_i, \varepsilon_{i,s}(n)g'_{i,s}(n')\Gamma_i)
\end{aligned}$$

$$+ \max_{n,n' \in \mathcal{P}_{s,r}} d_{G_i/\Gamma_i}(\varepsilon_{i,s}(n)g'_{i,s}(n')\Gamma_i, \varepsilon_{i,s}(n')g'_{i,s}(n')\Gamma_i)$$
$$\leq M^{O_k(d^{O_k(1)})}\big(\max_{n,n' \in \mathcal{P}_{s,r}} d_{G_i/\Gamma_i}(g'_{i,s}(n)\Gamma_i, g'_{i,s}(n')\Gamma_i)\big) + \max_{n,n' \in \mathcal{P}_{s,r}} d_{G_i}(\varepsilon_{i,s}(n), \varepsilon_{i,s}(n'))$$
$$\leq M^{O_k(d^{O_k(1)})} \cdot N^{-\Omega_k(1/(dT)^{O_k(1)})}$$

which completes the inductive step (our final decomposition is composed of all elements of the short $\mathcal{P}_s$ indexed by $s \in \mathcal{S}$ and all $\mathcal{P}_{s,r}$ arising from the long progressions indexed by $s \in \mathcal{L}$). We are done, noting that the number of inductive steps (hence the decay in parameters) is bounded in terms of k. □

4.3 Completing the proof

We are now run the Heath-Brown [35] and Szemerédi [98] density increment strategy as reformulated by Green and Tao [28]. In the first subsection we recall a number of preliminaries for the proof and in the second subsection we prove Theorem 4.1.1. Our treatment at this point is quite close to that of [28] and we borrow certain elements from the density increment portion of [78] as well.

4.3.1 Preliminaries for density increment

We first recall the definition of the Gowers U^s-norm over the integers.

Definition 4.3.1. Given $f\colon \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ and $s \geq 1$, we define

$$\|f\|^{2^s}_{U^s(\mathbb{Z}/N\mathbb{Z})} = \mathbb{E}_{x,h_1,\dots,h_s \in \mathbb{Z}/N\mathbb{Z}} \Delta_{h_1,\dots,h_s} f(x)$$

where $\Delta_h f(x) = f(x)\overline{f(x+h)}$ is the multiplicative discrete derivative (extended to vectors h in the natural way).

Given a natural number N and a function $f\colon [N] \to \mathbb{C}$, we choose a number $\widetilde{N} \geq 2^s N$ and define $\widetilde{f}\colon \mathbb{Z}/\widetilde{N}\mathbb{Z} \to \mathbb{C}$ via $\widetilde{f}(x) = f(x)$ for $x \in [N]$ and 0 otherwise. Then

$$\|f\|_{U^s[N]} := \|\widetilde{f}\|_{U^s(\mathbb{Z}/\widetilde{N}\mathbb{Z})}/\|\mathbb{1}_{[N]}\|_{U^s(\mathbb{Z}/\widetilde{N}\mathbb{Z})}.$$

One can check that this definition does not depend on the choice of $\widetilde{N}$. This is well known to be a seminorm for $s \geq 1$ and a norm for $s \geq 2$.

As mentioned, the main input for our result will be the following improved bound for the U^s-norm inverse theorem given as [69, Theorem 1.2].

Theorem 4.3.2. *Fix $\delta \in (0, 1/2)$. Suppose that $f\colon [N] \to \mathbb{C}$ is 1-bounded and*

$$\|f\|_{U^{s+1}[N]} \geq \delta.$$

Then there exists a nilmanifold G/Γ of degree s, complexity at most M, and dimension at most d as well as a function F on G/Γ which is at most K-Lipschitz such that

$$|\mathbb{E}_{n\in[N]}[f(n)\overline{F(g(n)\Gamma)}]| \geq \varepsilon,$$

where we may take

$$d \leq \log(1/\delta)^{O_s(1)} \text{ and } \varepsilon^{-1}, K, M \leq \exp(\log(1/\delta)^{O_s(1)}).$$

We now define the k-fold linear operator corresponding to counting k-term arithmetic progressions. Given functions $f_i\colon [N] \to \mathbb{C}$, define

$$\Lambda_k(f_1, \dots, f_k) = \mathbb{E}_{x,y\in\{0,\dots,N\}} \prod_{j=1}^{k} f_j(x + (j-1)y)$$

where f_i are extended by 0 outside of $[N]$. We also write

$$\Lambda_k(f) := \Lambda_k(f, \dots, f).$$

We have the following basic inequalities regarding the operator Λ_k. The proof is by now standard and hence is omitted (see [28, Lemma 3.2] and [27, Theorem 3.2]).

Lemma 4.3.3. *Consider functions $f_i\colon [N] \to \mathbb{C}$ for $1 \leq i \leq k$. Then we have*

$$\Lambda_k(f_1, \dots, f_k) \leq \min_{1\leq i\leq k} \|f_i\|_{L^1[N]} \cdot \prod_{j\neq i} \|f_j\|_{L^\infty[N]},$$
$$\Lambda_k(f_1, \dots, f_k) \ll_k \min_{1\leq i\leq k} \|f_i\|_{U^{k-1}[N]} \cdot \prod_{j\neq i} \|f_j\|_{L^\infty[N]}.$$

We next define factors and the factor induced by function g with a resolution K.

Definition 4.3.4. We define a *factor* $\mathcal{B}$ of $[N]$ to be a partition $[N] = \bigsqcup_{B\in\mathcal{B}} B$. We define $\mathcal{B}(x)$ for $x \in [N]$ to be the part of $\mathcal{B}$ that contains x. We say $\mathcal{B}'$ refines $\mathcal{B}$ if every part of $\mathcal{B}$ can be written as a disjoint union of parts of $\mathcal{B}'$. We define a *join* of a sequence of factors to be the partition (discarding empty parts)

$$\mathcal{B}_1 \vee \cdots \vee \mathcal{B}_d := \{B_1 \cap \cdots \cap B_d\colon B_i \in \mathcal{B}_i\}.$$

Next given a function $g\colon [N] \to \mathbb{R}$ and a resolution K, we define the factor induced by g of resolution K to be

$$\mathcal{B}_{g,K} = \bigsqcup_{j\in\mathbb{Z}} \{x \in [N]\colon g(x) \in [j/K, (j+1)/K).$$

Finally, given a factor $\mathcal{B}$, we define $\Pi_{\mathcal{B}} f$ by

$$\Pi_{\mathcal{B}} f(x) = \mathbb{E}_{y\in\mathcal{B}(x)} f(y).$$

A technical annoyance is that one may potentially have a large set of points near the cutoffs when defining $\mathcal{B}_{g,K}$. We define a notion of regularity capturing when a function g avoids such issues, which is related to an idea introduced by Bourgain [11] with regards to Bohr sets.

Definition 4.3.5. The factor $\mathcal{B}_{g,K}$ is C-regular if

$$\sup_{r>0}\left(\frac{1}{2r}\frac{1}{N}\big|\{x\in[N]\colon \|K\cdot g(x)\|_{\mathbb{R}/\mathbb{Z}}\le r\}\big|\right)\le C.$$

It turns out to be easy to obtain "regular" factors; a useful trick (motivated by the proof of [29, Corollary 2.3]) is to consider a random shift of g and then apply the Hardy—Littlewood maximal inequality. Given a function g and resolution K, we define the maximal function

$$M_{g,K}(t) := \sup_{r>0}\frac{1}{2r}\frac{1}{N}\big|\{x\in[N]\colon \|K\cdot g(x)-t\|_{\mathbb{R}/\mathbb{Z}}\le r\}\big|.$$

The Hardy–Littlewood maximal inequality (on the torus $\mathbb{R}/\mathbb{Z}$) implies that

$$\mathbb{E}_{t\in[0,1]}[M_{g,K}(t)] = O(1).$$

Therefore we have the following elementary fact which will prove useful.

Fact 4.3.6. There exists a constant $C = C_{4.3.6} > 0$ such that the following holds. Given a function $g\colon [N]\to\mathbb{R}$ and a resolution K, there exists a shift $t\in[0,1/K)$ such that $\mathcal{B}_{g-t,K}$ is C-regular.

4.3.2 Constructing factor approximation and density increment

The key claim which we need to prove Theorem 4.1.1 is the following density increment lemma, phrased as a trichotomy.

Lemma 4.3.7. *Fix an integer $k\ge 5$ and a constant $c>0$. Consider a function $f\colon[N]\to[0,1]$ such that $\mathbb{E}_{n\in[N]}f(n)=\delta$. There exist $c'=c'(c,k)$ and $C=C(c,k)$ such that one of the following always holds:*

- $N\le \exp(\exp(\log(1/\delta)^C))$;
- $|\Lambda_k(f)-\Lambda_k(\delta\cdot\mathbb{1}_{[N]})|\le c\delta^k$;
- *There exists an arithmetic progression $\mathcal{P}\subseteq[N]$ of length at least $N^{1/\exp(\log(1/\delta)^C)}$ such that*
$$\mathbb{E}_{n\in\mathcal{P}}f(n)\ge(1+c')\delta.$$

We prove Theorem 4.1.1 given Lemma 4.3.7; this is the standard density increment strategy.

Proof of Theorem 4.1.1 given Lemma 4.3.7. Suppose $A \subseteq [N]$ has no k-term arithmetic progressions. We iteratively increase the density of A; set $A = A_1$, $N = N_1$ and $\delta = \delta_1$ and we iteratively define $A_i \subseteq [N_i]$, and $\delta_i = |A_i|/N_i$.

If $N_i \leq \exp(\exp(\log(1/\delta_i)^C))$, we immediately terminate. Otherwise, note that as A_i is free of k-term arithmetic progressions, we have that

$$|\Lambda_k(\mathbb{1}_{A_i}) - \Lambda_k(\delta_i \cdot \mathbb{1}_{[N_i]})| \geq \delta_i^k \cdot |\Lambda_k(\mathbb{1}_{[N_i]})| - |A_i| \cdot N_i^{-2} \gg_k \delta_i^k$$

where we have used that $N_i \geq \exp(\exp(\log(1/\delta_i)^C)) \gg \delta_i^{-k}$. Therefore, the third case in Lemma 4.3.7 occurs and there exists $\mathcal{P}_{i+1}$ such that

$$|A_i \cap \mathcal{P}_{i+1}|/|\mathcal{P}_{i+1}| \geq (1 + c')\delta_i$$

and $|\mathcal{P}_{i+1}| \geq N_i^{1/\exp(\log(1/\delta_i)^C)}$. We now rescale the arithmetic progression $\mathcal{P}_{i+1}$ to $[|\mathcal{P}_{i+1}|] =: [N_{i+1}]$, which sends $A_i \cap \mathcal{P}_{i+1}$ to a new set A_{i+1}, and then we continue the iteration.

Note that at every iteration the density δ_i increases by a multiplicative factor of at least $(1+c')$, so we must terminate in at most $O_k(\log(1/\delta))$ iterations. Thus there exists an index $j \leq O_k(\log(1/\delta))$ such that

$$N^{1/\exp(O_k(\log(1/\delta)^{C+1}))} \leq N_j \leq \exp(\exp(\log(1/\delta_j)^C)) \leq \exp(\exp(\log(1/\delta)^C)).$$

This implies that

$$\log N \leq \exp(O_k(\log(1/\delta)^{O_k(1)}))$$

and thus

$$\delta \leq \exp(-(\log\log N)^{\Omega_k(1)}).$$ □

In order to prove Lemma 4.3.7, we first iterate Theorem 4.3.2 to obtain the following result.

Lemma 4.3.8. *Fix a parameter $\eta \in (0, 1/2)$ and $k \geq 5$. There exists a constant $C = C_k > 0$ such that the following statement holds. If $N \geq \exp(\log(1/\eta)^C)$ and $f\colon [N] \to \mathbb{R}$ is 1-bounded then there exist functions $h_1, \dots, h_T\colon [N] \to \mathbb{R}$ and $d, M, K \geq 1$ such that:*

- *$\mathcal{B} = \bigvee_{1 \leq i \leq T} \mathcal{B}_{h_i,K}$ satisfies $\|f - \Pi_{\mathcal{B}} f\|_{U^{k-1}[N]} \leq \eta$;*
- *$T, M, K \leq \exp(\log(1/\eta)^C)$ and $d \leq \log(1/\eta)^C$;*
- *$h_i = F_i(g_i(n)\Gamma_i)$ is a nilsequence where $g_i(n)$ takes values in a group G_i which is given a degree $(k-2)$ filtration, G_i/Γ_i has complexity bounded by M and dimension bounded by d, and $F_i\colon G_i/\Gamma_i \to \mathbb{R}$ is M-Lipschitz;*
- *$\mathcal{B}_{h_i,K}$ is C-regular for $1 \leq i \leq T$.*

Proof. The proof follows via applying Theorem 4.3.2 repeatedly. We begin the iteration by setting $\mathcal{B}_0 = [N]$ (i.e., the trivial partition). At each stage we will construct h_{i+1} and then set $\mathcal{B}_{i+1} = \mathcal{B}_i \vee \mathcal{B}_{h_{i+1},K}$ with $K = \lceil \exp(\log(1/\eta)^{O_k(1)}) \rceil$, where the implicit constant is chosen sufficiently large.

Step 1: If $\|f - \Pi_{\mathcal{B}_i} f\|_{U^{k-1}[N]} \leq \eta$, we terminate.

Step 2: If $\|f - \Pi_{\mathcal{B}_i} f\|_{U^{k-1}[N]} > \eta$, by Theorem 4.3.2, there exists a nilsequence $F_{i+1}(g_{i+1}(n)\Gamma_{i+1})$ such that

$$\left|\mathbb{E}_{n\in[N]}[(f - \Pi_{\mathcal{B}_i} f)(n)\overline{F_{i+1}(g_{i+1}(n)\Gamma_{i+1})}]\right| \geq \exp(-\log(1/\eta)^{O_k(1)})$$

and where G_{i+1}/Γ_{i+1} has complexity bounded by $\exp(\log(1/\eta)^{O_k(1)})$, dimension bounded by $\log(1/\eta)^{O_k(1)}$, $F_{i+1}\colon G_{i+1}/\Gamma_{i+1} \to \mathbb{C}$ is $\exp(\log(1/\eta)^{O_k(1)})$-Lipschitz, G_{i+1} has been given a degree $(k-2)$ filtration, and where $g_{i+1}(n)$ is a polynomial sequence with respect to this filtration. Taking either the real or imaginary part of F_{i+1}, we may assume that $F_{i+1}\colon G_{i+1}/\Gamma_{i+1} \to \mathbb{R}$ and thus that

$$\left|\mathbb{E}_{n\in[N]}[(f - \Pi_{\mathcal{B}_i} f)(n)F_{i+1}(g_{i+1}(n)\Gamma_{i+1})]\right| \geq \exp(-\log(1/\eta)^{O_k(1)}).$$

Note that for any $t \in [0, 1/K)$, this implies that

$$\begin{aligned}\left|\mathbb{E}_{n\in[N]}\left[(f - \Pi_{\mathcal{B}_i} f)(n)\frac{\lfloor K(F_{i+1}(g_{i+1}(n)\Gamma_{i+1}) + t)\rfloor}{K}\right]\right| &\\ \geq \left|\mathbb{E}_{n\in[N]}[(f - \Pi_{\mathcal{B}_i} f)(n)F_{i+1}(g_{i+1}(n)\Gamma_{i+1})]\right| - 2/K &\\ \geq \exp(-\log(1/\eta)^{O_k(1)}) &\end{aligned}$$

given that the implicit constant defining K is chosen sufficiently large. Recall here $\lfloor x \rfloor$ is defined in the standard manner that $x = \lfloor x \rfloor + \{x\}$ where $\lfloor x \rfloor \in \mathbb{Z}$ and $\{x\} \in [0, 1)$. We then take $t \in [0, 1/K)$, such that $\mathcal{B}_{F_{i+1}(g_{i+1}(n)\Gamma_{i+1})+t,K}$ is C-regular; this exists for C larger than an absolute constant by Fact 4.3.6.

Set $h_{i+1}(n) := F_{i+1}(g_{i+1}(n)\Gamma_{i+1}) + t$. Note that

$$\frac{\lfloor K(F_{i+1}(g_{i+1}(n)\Gamma_{i+1}) + t)\rfloor}{K}$$

is measurable with respect to $\mathcal{B}_{h_{i+1},K}$ by construction and it is bounded by $\exp(\log(1/\eta)^{O_k(1)})$. Therefore since $\Pi_{\mathcal{B}_{h_{i+1},K}}$ is self-adjoint we have

$$\begin{aligned}&\mathbb{E}_{n\in[N]}[|\Pi_{\mathcal{B}_{h_{i+1},K}}(f - \Pi_{\mathcal{B}_i} f)(n)|] \\ &\quad\geq (1 + \|F_{i+1}\|_{L^\infty(G_{i+1}/\Gamma_{i+1})})^{-1} \cdot \left|\mathbb{E}_{n\in[N]}\left[(f - \Pi_{\mathcal{B}_i} f)(n)\frac{\lfloor K(F_{i+1}(g_{i+1}(n)\Gamma_{i+1}) + t)\rfloor}{K}\right]\right| \\ &\quad\geq \exp(-\log(1/\eta)^{O_k(1)}).\end{aligned}$$

Step 3: We now return back to Step 1 and keep on iterating this procedure until it terminates. This completes the proof modulo showing that the iteration terminates in a small number of steps. To show this, note that

$$\|\Pi_{\mathcal{B}_{h_{i+1},K}}(f - \Pi_{\mathcal{B}_i} f)\|_{L^1[N]} \leq \|\Pi_{\mathcal{B}_{h_{i+1},K}}(f - \Pi_{\mathcal{B}_i} f)\|_{L^2[N]} = \|\Pi_{\mathcal{B}_{h_{i+1},K}}\Pi_{\mathcal{B}_{i+1}}(f - \Pi_{\mathcal{B}_i} f)\|_{L^2[N]}$$

$$\leq \|\Pi_{\mathcal{B}_{i+1}}(f - \Pi_{\mathcal{B}_i} f)\|_{L^2[N]} = \|\Pi_{\mathcal{B}_{i+1}} f - \Pi_{\mathcal{B}_i} f\|_{L^2[N]}$$
$$= (\|\Pi_{\mathcal{B}_{i+1}} f\|_{L^2[N]}^2 - \|\Pi_{\mathcal{B}_i} f\|_{L^2[N]}^2)^{1/2}.$$

The final equality is the Pythagorean theorem with respect to projections (this follows from e.g. [78, Lemma 4.3(iv)]). We deduce

$$\|\Pi_{\mathcal{B}_{i+1}} f\|_{L^2[N]}^2 - \|\Pi_{\mathcal{B}_i} f\|_{L^2[N]}^2 \geq \exp(-\log(1/\eta)^{O_k(1)}).$$

Since for all i we have $\|\Pi_{\mathcal{B}_i} f\|_{L^2[N]} \leq \|f\|_{L^2[N]} \leq 1$, there are at most $\exp(\log(1/\eta)^{O_k(1)})$ iterations as desired. □

We now complete the proof of Lemma 4.3.7 and therefore the proof of Theorem 4.1.1. The first part of the proof is finding a density increment on a factor derived from nilsequences, which is essentially identical to that of [28, Lemma 5.8]. In the second part, we apply our nilsequence Schmidt-type result Lemma 4.2.1 to find a long arithmetic progression with density increment.

Proof of Lemma 4.3.7. Without loss of generality, we may assume that c is smaller than an absolute constant. Furthermore we may assume that $N \geq \exp(\exp(\log(1/\delta)^{\Omega(1)}))$ (where the implicit constant may depend on c, k) and $|\Lambda_k(f) - \Lambda_k(\delta \cdot \mathbb{1}_{[N]})| \geq c\delta^k$.

Step 1: Increment on a factor. By applying Lemma 4.3.8, there exists a factor $\mathcal{B}$ (derived from nilsequences of appropriate complexity, with parameters below) such that

$$\|\Pi_{\mathcal{B}} f - f\|_{U^{k-1}[N]} \leq c^* \delta^k$$

where we choose c^* sufficiently small in terms of c. Via telescoping and the second inequality in Lemma 4.3.3, we have

$$|\Lambda_k(f) - \Lambda_k(\Pi_{\mathcal{B}} f)| \leq c\delta^k/2$$

as long as c^* was chosen appropriately, and therefore

$$|\Lambda_k(\Pi_{\mathcal{B}} f) - \Lambda_k(\delta \cdot \mathbb{1}_{[N]})| \geq c\delta^k/2.$$

Take $c' = \min(c, 1)/(10k)^5$. Let $g = \min(\Pi_{\mathcal{B}} f, (1 + c')\delta)$. The crucial claim is that if $\Omega' = \{n \in [N] \colon g(n) \neq \Pi_{\mathcal{B}} f(n)\} = \{n \in [N] \colon \Pi_{\mathcal{B}} f(n) > (1 + c')\delta\}$ then Ω' must have sufficiently large measure. To see this note that:

$$|\Lambda_k(\Pi_{\mathcal{B}} f) - \Lambda_k(g)| \leq k\|\Pi_{\mathcal{B}} f - g\|_{L^1[N]} \leq k\mathbb{P}_{n\in[N]}[n \in \Omega'],$$
$$|\Lambda_k(\delta\mathbb{1}_{[N]}) - \Lambda_k(g)| \leq k(1 + c')^{k-1}\delta^{k-1}\|\delta\mathbb{1}_{[N]} - g\|_{L^1[N]},$$
$$\|g - \delta\mathbb{1}_{[N]}\|_{L^1[N]} \leq \mathbb{P}_{n\in[N]}[n \in \Omega'] + \|\delta\mathbb{1}_{[N]} - \Pi_{\mathcal{B}} f\|_{L^1[N]}.$$

The first and second inequality follow from the first part of Lemma 4.3.3 and telescoping while the final inequality follows from the triangle inequality. We simplify the inequalities slightly; as $\mathbb{E}_{n\in[N]}[\delta\mathbb{1}_{[N]}] = \mathbb{E}_{n\in[N]}[f] = \mathbb{E}_{n\in[N]}[\Pi_{\mathcal{B}} f]$, we have

$$\|\delta\mathbb{1}_{[N]} - \Pi_{\mathcal{B}} f\|_{L^1[N]} = 2\|\max(\Pi_{\mathcal{B}} f - \delta\mathbb{1}_{[N]}, 0)\|_{L^1[N]} \leq 2c'\delta + 2\mathbb{P}_{n\in[N]}[n \in \Omega'].$$

Given this and using the upper bound on c', we deduce

$$\begin{aligned}|\Lambda_k(\Pi_{\mathcal{B}} f) - \Lambda_k(g)| &\leq k\|\Pi_{\mathcal{B}} f - g\|_{L^1[N]} \leq k\mathbb{P}_{n\in[N]}[n \in \Omega'],\\ |\Lambda_k(\delta \mathbb{1}_{[N]}) - \Lambda_k(g)| &\leq 2k\delta^{k-1}\|\delta \mathbb{1}_{[N]} - g\|_{L^1[N]},\\ \|g - \delta \mathbb{1}_{[N]}\|_{L^1[N]} &\leq 3\mathbb{P}_{n\in[N]}[n \in \Omega'] + 2c'\delta.\end{aligned}$$

Therefore

$$\begin{aligned}c\delta^k/2 \leq |\Lambda_k(\delta \cdot \mathbb{1}_{[N]}) - \Lambda_k(\Pi_{\mathcal{B}} f)| &\leq |\Lambda_k(\delta \cdot \mathbb{1}_{[N]}) - \Lambda_k(g)| + |\Lambda_k(\Pi_{\mathcal{B}} f) - \Lambda_k(g)|\\ &\leq k\mathbb{P}_{n\in[N]}[n \in \Omega'] + 2k\delta^{k-1}\|\delta \mathbb{1}_{[N]} - g\|_{L^1[N]} \leq 7k\mathbb{P}_{n\in[N]}[n \in \Omega'] + 4kc'\delta^k;\end{aligned}$$

thus we have $\mathbb{P}_{n\in[N]}[n \in \Omega'] \geq c\delta^k/(20k)$.

Step 2: Increment on a progression. We are now in position to apply the nilsequence Schmidt-type result Lemma 4.2.1. Recall that we applied Lemma 4.3.8 to find $\mathcal{B}$, and hence we may write $\mathcal{B} = \bigvee_{1\leq i\leq T} \mathcal{B}_{h_i,K}$ where:

- $T, M, K \leq \exp(\log(1/\delta)^C)$ and $d \leq \log(1/\delta)^C$;
- $h_i = F_i(g_i(n)\Gamma_i)$ is a nilsequence where $g_i(n)$ takes values in a group G_i which is given a degree $(k-2)$ filtration, G_i/Γ_i has complexity bounded by M and dimension bounded by d, and $F_i\colon G_i/\Gamma_i \to \mathbb{R}$ is M-Lipschitz;
- $\mathcal{B}_{h_i,K}$ is C-regular for $1 \leq i \leq T$.

Here C is a slightly larger value than the constant C_k in Lemma 4.3.8, depending only on k.

We now apply Lemma 4.2.1 to $g_i(n)$ for $1 \leq i \leq T$. We obtain a decomposition of $[N]$ into arithmetic progressions $\mathcal{P}_1, \dots, \mathcal{P}_L$ such that

- $N/L \geq N^{-1/\exp(\log(1/\delta)^{O_k(1)})}$;
- We have

$$\max_{\substack{1\leq i\leq T\\ 1\leq j\leq L}} \max_{n,n'\in\mathcal{P}_j} d_{G_i/\Gamma_i}(g_i(n)\Gamma_i, g_i(n')\Gamma_i) \leq \exp(\log(1/\delta)^{O_k(1)}) \cdot N^{-1/\exp(\log(1/\delta)^{O_k(1)})}.$$

We now consider $\mathcal{P}_j$ which intersect Ω'. Call a progression in the decomposition *crossing* if it intersects Ω' and $[N] \setminus \Omega'$ and a progression *contained* if it is fully within in Ω'. Since Ω' is measurable in terms of $\mathcal{B}$, for a progression to be crossing it must "cross a boundary" defining $\mathcal{B}_{h_i,K}$ for at least one $1 \leq i \leq T$. If a progression $\mathcal{P}_j$ crosses one of these boundaries defined by h_i then all points in $\mathcal{P}_j$ map close to this boundary, since the function F_i is M-Lipschitz. In particular, by regularity of each $\mathcal{B}_{h_i,K}$, the measure (with respect to the uniform distribution on $[N]$) of improper progressions is bounded by

$$\ll_k T \cdot \exp(\log(1/\delta)^{O_k(1)}) \cdot N^{-1/\exp(\log(1/\delta)^{O_k(1)})} = \exp(\log(1/\delta)^{O_k(1)}) \cdot N^{-1/\exp(\log(1/\delta)^{O_k(1)})}.$$

Let Ω^* denote the union of all the contained progressions which have length at least $N' = cc'\delta^{k+1}/(400k) \cdot N/L$ (hence certainly $\Omega^* \subseteq \Omega'$). Let $\mathcal{I}$ be the set of all $1 \leq i \leq L$ so that $\mathcal{P}_i$ either has length at most N' or is crossing. We easily see that

$$0 \leq \mathbb{P}_{n\in[N]}[n \in \Omega'] - \mathbb{P}_{n\in[N]}[n \in \Omega^*] \leq \sum_{i\in\mathcal{I}} \mathbb{P}_{n\in[N]}[n \in \mathcal{P}_i] \leq cc'\delta^{k+1}/(200k);$$

in the final inequality we have used that $N \geq \exp(\log(1/\delta)^{\Omega(1)})$ for a sufficiently large implicit constant.

Finally, this implies that

$$\begin{aligned}
\mathbb{E}_{n\in\Omega^*}[f] = \frac{\mathbb{E}_{n\in[N]}[f \cdot \mathbb{1}_{n\in\Omega^*}]}{\mathbb{P}_{n\in[N]}[n \in \Omega^*]} &\geq \frac{\mathbb{E}_{n\in[N]}[f \cdot \mathbb{1}_{n\in\Omega'}] - cc'\delta^{k+1}/(200k)}{\mathbb{P}_{n\in[N]}[n \in \Omega']} \\
&\geq \frac{\mathbb{E}_{n\in[N]}[f \cdot \mathbb{1}_{n\in\Omega'}]}{\mathbb{P}_{n\in[N]}[n \in \Omega']} - \frac{cc'\delta^{k+1}/(200k)}{c\delta^k/(20k)} \\
&= \frac{\mathbb{E}_{n\in[N]}[\Pi_{\mathcal{B}} f \cdot \mathbb{1}_{n\in\Omega'}]}{\mathbb{P}_{n\in[N]}[n \in \Omega']} - c'\delta/10 \\
&\geq (1+c')\delta - c'\delta/10 \geq (1+c'/2)\delta.
\end{aligned}$$

By pigeonhole, this implies that there exists a contained arithmetic progression $\mathcal{P}_i$ having length at least $cc'\delta^{k+1}/(400k) \cdot N/L \geq N^{-1/\exp(\log(1/\delta)^{O_k(1)})}$ on which the density of f is at least $(1+c'/2)\delta$. Adjusting the value of c', this completes the proof. □

Bibliography

[1] Noga Alon and Joel H. Spencer, *The probabilistic method*, fourth ed., Wiley Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., Hoboken, NJ, 2016.

[2] Ryan Alweiss, Shachar Lovett, Kewen Wu, and Jiapeng Zhang, *Improved bounds for the sunflower lemma*, Ann. of Math. (2) **194** (2021), 795–815.

[3] Lina J. Andrén, Carl Johan Casselgren, and Lars-Daniel Öhman, *Avoiding arrays of odd order by Latin squares*, Combin. Probab. Comput. **22** (2013), 184–212.

[4] Ben Barber, Stefan Glock, Daniela Kühn, Allan Lo, Richard Montgomery, and Deryk Osthus, *Minimalist designs*, Random Structures Algorithms **57** (2020), 47–63.

[5] Ben Barber, Daniela Kühn, Allan Lo, and Deryk Osthus, *Edge-decompositions of graphs with high minimum degree*, Adv. Math. **288** (2016), 337–385.

[6] Anirban Basak and Mark Rudelson, *Sharp transition of the invertibility of the adjacency matrices of sparse random graphs*, Probab. Theory Related Fields **180** (2021), 233–308.

[7] Thomas F. Bloom, *A quantitative improvement for Roth's theorem on arithmetic progressions*, J. Lond. Math. Soc. (2) **93** (2016), 643–663.

[8] Thomas F. Bloom and Olof Sisask, *Breaking the logarithmic barrier in Roth's theorem on arithmetic progressions*, arXiv:2007.03528.

[9] Thomas F. Bloom and Olof Sisask, *An improvement to the Kelley-Meka bounds on three-term arithmetic progressions*, arXiv:2309.02353.

[10] Erwin Bolthausen, *An estimate of the remainder in a combinatorial central limit theorem*, Z. Wahrsch. Verw. Gebiete **66** (1984), 379–386.

[11] Jean Bourgain, *On triples in arithmetic progression*, Geom. Funct. Anal. **9** (1999), 968–984.

[12] Jean Bourgain, *Roth's theorem on progressions revisited*, J. Anal. Math. **104** (2008), 155–192.

[13] Jean Bourgain, Van H. Vu, and Philip Matchett Wood, *On the singularity probability of discrete random matrices*, J. Funct. Anal. **258** (2010), 559–603.

[14] Stefan Ehard, Stefan Glock, and Felix Joos, *Pseudorandom hypergraph matchings*, Combin. Probab. Comput. **29** (2020), 868–885.

[15] Paul Erdős and Alfréd Rényi, *On random graphs. I*, Publ. Math. Debrecen **6** (1959), 290–297.

[16] Paul Erdős and Alfréd Rényi, *On random matrices*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **8** (1964), 455–461 (1964).

[17] Paul Erdős and Alfréd Rényi, *On the existence of a factor of degree one of a connected random graph*, Acta Math. Acad. Sci. Hungar. **17** (1966), 359–368.

[18] Asaf Ferber, Vishesh Jain, Kyle Luh, and Wojciech Samotij, *On the counting problem in inverse Littlewood–Offord theory*, J. Lond. Math. Soc. (2) **103** (2021), 1333–1362.

[19] Keith Frankston, Jeff Kahn, Bhargav Narayanan, and Jinyoung Park, *Thresholds versus fractional expectation-thresholds*, Ann. of Math. (2) **194** (2021), 475–495.

[20] Ehud Friedgut, *personal communication.*

[21] Ehud Friedgut, *Sharp thresholds of graph properties, and the k-SAT problem*, J. Amer. Math. Soc. **12** (1999), 1017–1054, With an appendix by Jean Bourgain.

[22] Ehud Friedgut, *Hunting for sharp thresholds*, Random Structures Algorithms **26** (2005), 37–51.

[23] Stefan Glock, Daniela Kühn, Allan Lo, and Deryk Osthus, *On a conjecture of Erdős on locally sparse Steiner triple systems*, Combinatorica **40** (2020), 363–403.

[24] Stefan Glock, Daniela Kühn, Allan Lo, and Deryk Osthus, *The existence of designs via iterative absorption: hypergraph F-designs for arbitrary F*, Mem. Amer. Math. Soc. **284** (2023), 1–131.

[25] W. T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, Geom. Funct. Anal. **8** (1998), 529–551.

[26] W. T. Gowers, *Arithmetic progressions in sparse sets*, Current developments in mathematics, 2000, Int. Press, Somerville, MA, 2001, pp. 149–196.

[27] W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), 465–588.

[28] Ben Green and Terence Tao, *New bounds for Szemerédi's theorem. II. A new bound for $r_4(N)$*, Analytic number theory, Cambridge Univ. Press, Cambridge, 2009, pp. 180–204.

[29] Ben Green and Terence Tao, *An arithmetic regularity lemma, an associated counting lemma, and applications*, An irregular mind, Bolyai Soc. Math. Stud., vol. 21, János Bolyai Math. Soc., Budapest, 2010, pp. 261–334.

[30] Ben Green and Terence Tao, *Yet another proof of Szemerédi's theorem*, An irregular mind, Bolyai Soc. Math. Stud., vol. 21, János Bolyai Math. Soc., Budapest, 2010, pp. 335–342.

[31] Ben Green and Terence Tao, *The quantitative behaviour of polynomial orbits on nilmanifolds*, Ann. of Math. (2) **175** (2012), 465–540.

[32] Ben Green and Terence Tao, *New bounds for Szemerédi's theorem, III: a polylogarithmic bound for* $r_4(N)$, Mathematika **63** (2017), 944–1040.

[33] Ben Green, Terence Tao, and Tamar Ziegler, *An inverse theorem for the Gowers* $U^{s+1}[N]$*-norm*, Ann. of Math. (2) **176** (2012), 1231–1372.

[34] Torbjörn Gustavsson, *Decompositions of large graphs and digraphs with high minimum degree*, Ph.D. thesis, Univ. of Stockholm, 1991.

[35] D. R. Heath-Brown, *Integer sets containing no arithmetic progressions*, J. London Math. Soc. (2) **35** (1987), 385–394.

[36] Han Huang, *Rank of sparse Bernoulli matrices*, arXiv:2009.13726.

[37] Anwar Irmatov, *Asymptotics of the number of threshold functions and the singularity probability of random* $\{\pm 1\}$*-matrices*, Doklady Mathematics **101** (2020), 247–249.

[38] Vishesh Jain, *Approximate Spielman-Teng theorems for the least singular value of random combinatorial matrices*, Israel J. Math. **242** (2021), 461–500.

[39] Vishesh Jain and Huy Tuan Pham, *Optimal thresholds for Latin squares, Steiner Triple Systems, and edge colorings*, Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 2024, pp. 1425–1436.

[40] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney, *Sharp invertibility of random Bernoulli matrices*, arXiv:2010.06553.

[41] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney, *Singularity of discrete random matrices*, Geom. Func. Anal. **31** (2021), 1160–1218.

[42] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney, *On the smoothed analysis of the smallest singular value with discrete noise*, Bull. Lond. Math. Soc. **54** (2022), 369–388.

[43] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney, *The Smallest Singular Value of Dense Random Regular Digraphs*, Int. Math. Res. Not. IMRN (2022), 19300–19334.

[44] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.

[45] Anders Johansson, Jeff Kahn, and Van Vu, *Factors in random graphs*, Random Structures Algorithms **33** (2008), 1–28.

[46] Tony Johansson, *On Hamilton cycles in Erdős-Rényi subgraphs of large graphs*, Random Structures Algorithms **57** (2020), 132–149.

[47] Jeff Kahn, *Hitting times for Shamir's problem*, Trans. Amer. Math. Soc. **375** (2022), 627–668.

[48] Jeff Kahn, *Asymptotics for Shamir's problem*, Adv. Math. **422** (2023), Paper No. 109019, 39.

[49] Jeff Kahn and Gil Kalai, *Thresholds and expectation thresholds*, Combin. Probab. Comput. **16** (2007), 495–502.

[50] Jeff Kahn, János Komlós, and Endre Szemerédi, *On the probability that a random* ± 1*-matrix is singular*, J. Amer. Math. Soc. **8** (1995), 223–240.

[51] Jeff Kahn, Bhargav Narayanan, and Jinyoung Park, *The threshold for the square of a Hamilton cycle*, Proc. Amer. Math. Soc. **149** (2021), 3201–3208.

[52] Dong Yeap Kang, Tom Kelly, Daniela Kühn, Abhishek Methuku, and Deryk Osthus, *Thresholds for Latin squares and Steiner triple systems: Bounds within a logarithmic factor*, Trans. Amer. Math. Soc. **376** (2023), 6623–6662.

[53] Boris S. Kašin, *Diameters of some finite-dimensional sets and classes of smooth functions*, Izvestiya: Mathematics **11** (1977), 317–333.

[54] Peter Keevash, *The existence of designs*, arXiv:1401.3665.

[55] Peter Keevash, *The optimal edge-colouring threshold*, arXiv:2212.04397.

[56] Peter Keevash, *Counting designs*, J. Eur. Math. Soc. (JEMS) **20** (2018), 903–927.

[57] Zander Kelley and Raghu Meka, *Strong Bounds for 3-Progressions*, 2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS), 2023, pp. 933–973.

[58] Thomas P. Kirkman, *On a problem in combinations*, Cambridge and Dublin Mathematical Journal **2** (1847), 191–204.

[59] Fiachra Knox, Daniela Kühn, and Deryk Osthus, *Edge-disjoint Hamilton cycles in random graphs*, Random Structures Algorithms **46** (2015), 397–445.

[60] János Komlós, *On the determinant of* (0, 1) *matrices*, Studia Sci. Math. Hungar. **2** (1967), 7–21.

[61] Michael Krivelevich, Choongbum Lee, and Benny Sudakov, *Robust Hamiltonicity of Dirac graphs*, Trans. Amer. Math. Soc. **366** (2014), 3095–3130.

[62] Daniela Kühn and Deryk Osthus, *On Pósa's conjecture for random graphs*, SIAM J. Discrete Math. **26** (2012), 1440–1457.

[63] Daniela Kühn and Deryk Osthus, *Hamilton decompositions of regular expanders: a proof of Kelly's conjecture for large tournaments*, Adv. Math. **237** (2013), 62–146.

[64] Matthew Kwan, Ashwin Sah, Mehtaab Sawhney, and Michael Simkin, *High-girth Steiner triple systems*, arXiv:2201.04554.

[65] Matthew Kwan, Ashwin Sah, Mehtaab Sawhney, and Michael Simkin, *Substructures in Latin squares*, Israel J. Math. **256** (2023), 363–416.

[66] James Leng, *Efficient Equidistribution of Nilsequences*, arXiv:2312.10772.

[67] James Leng, Ashwin Sah, and Mehtaab Sawhney, *Improved bounds for five-term arithmetic progressions*, arXiv:2312.10776.

[68] James Leng, Ashwin Sah, and Mehtaab Sawhney, *Improved bounds for Szemerédi's theorem*, arXiv:2402.17995.

[69] James Leng, Ashwin Sah, and Mehtaab Sawhney, *Quasipolynomial bounds for the inverse theorem for the Gowers $U^{s+1}[N]$-norm*, arXiv:2402.17994.

[70] Alexander E. Litvak, Alain Pajor, Mark Rudelson, and Nicole Tomczak-Jaegermann, *Smallest singular value of random matrices and geometry of random polytopes*, Adv. Math. **195** (2005), 491–523.

[71] Alexander E. Litvak and Konstantin Tikhomirov, *Singularity of sparse Bernoulli matrices*, Duke Math. J. **171** (2022), 1135–1233.

[72] Galyna V. Livshyts, *The smallest singular value of heavy-tailed not necessarily i.i.d. random matrices via random rounding*, J. Anal. Math. **145** (2021), 257–306.

[73] Galyna V. Livshyts, Konstantin Tikhomirov, and Roman Vershynin, *The smallest singular value of inhomogeneous square random matrices*, Ann. Probab. **49** (2021), 1286–1309.

[74] Richard Montgomery, *Fractional clique decompositions of dense partite graphs*, Combin. Probab. Comput. **26** (2017), 911–943.

[75] Richard Montgomery, *Spanning trees in random graphs*, Adv. Math. **356** (2019), 106793, 92.

[76] Hoi H. Nguyen, *On the singularity of random combinatorial matrices*, SIAM J. Discrete Math. **27** (2013), 447–458.

[77] Jinyoung Park and Huy Pham, *A proof of the Kahn–Kalai conjecture*, J. Amer. Math. Soc. **37** (2024), 235–243.

[78] Sarah Peluse and Sean Prendiville, *A polylogarithmic bound in the nonlinear Roth theorem*, Int. Math. Res. Not. IMRN (2022), 5658–5684.

[79] Lajos Pósa, *Hamiltonian circuits in random graphs*, Discrete Math. **14** (1976), 359–364.

[80] Rajeev Raman, *The power of collision: randomized parallel algorithms for chaining and integer sorting*, Foundations of software technology and theoretical computer science (Bangalore, 1990), Lecture Notes in Comput. Sci., vol. 472, Springer, Berlin, 1990, pp. 161–175.

[81] Vojtěch Rödl, *On a packing and covering problem*, European J. Combin. **6** (1985), 69–78.

[82] Vojtěch Rödl and Luboš Thoma, *Asymptotic packing and the random greedy algorithm*, Random Structures Algorithms **8** (1996), 161–177.

[83] Boris A. Rogozin, *On the increase of dispersion of sums of independent random variables*, Teor. Verojatnost. i Primenen **6** (1961), 106–108.

[84] Klaus F. Roth, *On certain sets of integers. II*, J. London Math. Soc. **29** (1954), 20–26.

[85] Mark Rudelson, *Invertibility of random matrices: norm of the inverse*, Ann. of Math. (2) **168** (2008), 575–600.

[86] Mark Rudelson and Roman Vershynin, *The Littlewood-Offord problem and invertibility of random matrices*, Adv. Math. **218** (2008), 600–633.

[87] Ashwin Sah, Mehtaab Sawhney, and Michael Simkin, *Threshold for Steiner triple systems*, Geom. Func. Anal. **33** (2023), 1141–1172.

[88] Tom Sanders, *On Roth's theorem on progressions*, Ann. of Math. (2) **174** (2011), 619–636.

[89] Tom Sanders, *On certain other sets of integers*, J. Anal. Math. **116** (2012), 53–82.

[90] Gideon Schechtman, *Special orthogonal splittings of L_1^{2k}*, Israel J. Math. **139** (2004), 337–347.

[91] Jeanette Schmidt and Eli Shamir, *A threshold for perfect matchings in random d-pure hypergraphs*, Discrete Math. **45** (1983), 287–295.

[92] Wolfgang M. Schmidt, *Small fractional parts of polynomials*, Regional Conference Series in Mathematics, No. 32, American Mathematical Society, Providence, RI, 1977.

[93] Eli Shamir and Benny Sudakov, *Two-sided, unbiased version of Hall's marriage theorem*, Amer. Math. Monthly **124** (2017), 79–80.

[94] Michael Simkin, $(n, k, k-1)$-*Steiner systems in random hypergraphs*, arXiv:1711.01975.

[95] Joel Spencer, *Asymptotic packing via a branching process*, Random Structures Algorithms **7** (1995), 167–172.

[96] Endre Szemerédi, *On sets of integers containing no four elements in arithmetic progression*, Number Theory (Colloq., János Bolyai Math. Soc., Debrecen, 1968), Colloq. Math. Soc. János Bolyai, vol. 2, North-Holland, Amsterdam-London, 1970, pp. 197–204.

[97] Endre Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245.

[98] Endre Szemerédi, *Integer sets containing no arithmetic progressions*, Acta Math. Hungar. **56** (1990), 155–158.

[99] Michel Talagrand, *Are many small sets explicitly small?*, STOC'10—Proceedings of the 2010 ACM International Symposium on Theory of Computing, ACM, New York, 2010, pp. 13–35.

[100] Terence Tao, *Structure and randomness in combinatorics*, 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07), 2007, pp. 3–15.

[101] Terence Tao and Van H. Vu, *On random* ± 1 *matrices: singularity and determinant*, Random Structures Algorithms **28** (2006), 1–23.

[102] Terence Tao and Van H. Vu, *On the singularity probability of random Bernoulli matrices*, J. Amer. Math. Soc. **20** (2007), 603–628.

[103] Konstantin Tikhomirov, *Singularity of random Bernoulli matrices*, Ann. of Math. (2) **191** (2020), 593–634.

[104] Tuan Tran, *The smallest singular value of random combinatorial matrices*, arXiv:2007.06318.

[105] Roman Vershynin, *High-dimensional probability*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 47, Cambridge University Press, Cambridge, 2018, An introduction with applications in data science, With a foreword by Sara van de Geer.