Title
Spectral properties of non-Hermitian random matrices

Author
Cook, Nicholas Andrew

Publication Date
2016

Peer reviewed|Thesis/dissertation

University of California
Los Angeles

Spectral properties of non-Hermitian random matrices

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Mathematics

by

Nicholas Andrew Cook

2016

Abstract of the Dissertation

Spectral properties of non-Hermitian random matrices

by

Nicholas Andrew Cook

Doctor of Philosophy in Mathematics

University of California, Los Angeles, 2016

Professor Terence Chi-Shen Tao, Chair

This thesis presents new results concerning the spectral properties of certain families of large random matrices. The overarching goal is to extend some well-known results for matrices with independent and identically distributed (iid) entries to random matrices whose entries are either dependent or non-identically distributed. Particular attention is given to the adjacency matrix of a random regular digraph.

Making use of the method of exchangeable pairs, we establish concentration bounds for codegrees and edge densities of induced subgraphs of a random regular digraph. We apply these bounds along with other coupling techniques to prove that the associated adjacency matrix is invertible with high probability, assuming a mild growth condition on the degree of the graph.

We next prove lower tail estimates for the smallest singular value of matrices with independent but non-identically distributed entries, including matrices with most entries set to zero deterministically. In particular, we show that for small diagonal perturbations of centered random matrices with independent entries of bounded variance, the smallest singular value is of inverse-polynomial order with high probability.

Finally, we prove that the circular law holds for a random matrix obtained from the adjacency matrix of a dense random regular digraph by multiplying each entry by a random sign.

The dissertation of Nicholas Andrew Cook is approved.

Igor Pak

Ali H. Sayed

Terence Chi-Shen Tao, Committee Chair

University of California, Los Angeles

2016

For my grandfather

Table of Contents

List of Figures

Acknowledgments

I am grateful to have been supported at various times during my PhD by an NSF RTG Grant and a Girsky Fellowship Award.

With regard to this thesis, I am most indebted to my adviser, Terry Tao, for all of his guidance and support, and for making himself available for weekly meetings and over email despite many other demands on his time. From Terry I have learned a great deal about how to approach math problems; perhaps more importantly, his consistently positive outlook and the great care he puts into everything he does have left a deep impression on me.

I also thank the other members of my committee – Marek Biskup, Igor Pak, and Ali Sayed – for their interest in this work and for helpful comments.

To a beginning graduate student, the UCLA Mathematics Department can seem overwhelmingly large. I am grateful to Terry and the other analysis faculty – in particular, Mario Bonk, John Garnett, Michael Hitrik, Rowan Killip, Jim Ralston, Christoph Thiele (now at Bonn), and Monica Visan – for making the analysis group feel more like a small department within the Department. I have come to greatly value their faithful attendance at the weekly participating seminar, and for always keeping me on my toes with their laser-sharp focus. I enjoyed our meetings outside for cookies between talks (weather always permitting), and even the long carpool rides through L.A. traffic to attend the joint seminar at Caltech.

The probability group at UCLA is somewhat smaller, but this was more than compensated by the hard work of Marek Biskup, who ran numerous seminars and brought in several visitors and postdocs. I learned a lot from his excellent courses on random Schrödinger operators and Gaussian processes, and I am thankful for all of the advice he has given me over the years.

I also want to thank Olga Radko, who provided invaluable mentorship and support throughout my time as a teaching assistant, instructor, and my involvement with the L.A. Math Circle.

I thank my collaborators – Walid Hachem, Jamal Najim, David Renfrew, Larry Goldstein and Toby Johnson – from whom I have learned a great deal of mathematics. I am particularly grateful to Walid and Jamal for hosting me for a week in Paris in November 2014.

Among the many friends I made at UCLA I particularly want to thank my fellow students Will Feldman, Alan Mackey, and Brent Nelson for the countless hours we shared discussing homework problems and preparing for qualifying exams (with occasional breaks for ping pong). While I am sure none of us would willingly go through all of that again, their companionship through our first year at UCLA and each year since made the experience worthwhile.

As regards my life's trajectory leading up to graduate school, I am especially thankful to my parents for giving me what I increasingly realize was a good upbringing. I could not have asked for better role models, and I am grateful to have had their constant support for my interests and ambitions.

I would not be pursuing this doctorate were it not for my grandfather, Roy Jackson, who himself became a professor of engineering at Princeton despite never getting a PhD (which was slightly less uncommon in those days). To him I owe my entire interest in science and mathematics. From an early age he showed me the joy of unraveling how the world works, and that learning is something one can do outside of school. If not for him, I wouldn't have known until much later how challenging math can be. To him I dedicate this thesis.

Finally, I want to thank my family for their support and for giving my life balance. I thank my children for their love and understanding, and for providing a constant stream of problems of a very different sort from the ones tackled in this thesis. Above all, I thank my wife Shemra, who is my partner in everything, and whom I admire more than anyone.

Los Angeles, August 2016

Vita

2009	B.S. (Mathematics), University of North Carolina, Chapel Hill, NC.
2008	Alfred Brauer Prize, UNC Chapel Hill Mathematics Dept.
2015–2016	Girsky Fellowship Award, UCLA Mathematics Dept.
2016	Heaviside Wealth Management Award, UCLA Mathematics Dept.
2016	Pacific Journal of Mathematics Dissertation Prize, UCLA Mathematics.

Publications

N. A. Cook. On the singularity of adjacency matrices for random regular digraphs. *Probab. Theory Relat. Fields*, to appear.

N. A. Cook. Discrepancy properties for random regular digraphs. *Random Struct. Algor.*, to appear.

N. A. Cook, L. Goldstein, and T. Johnson. Size biased couplings and the spectral gap for random regular graphs. Submitted.

CHAPTER 1

Introduction

In the past decade or so, much progress has been made in the understanding of spectral properties for *iid matrices* – that is, matrices of the form $X = (\xi_{ij})_{i,j=1}^n$ where the entries $\xi_{ij} \in \mathbb{C}$ are independent and identically distributed random variables, and n is either large or going to infinity. Note that we impose no symmetry constraint on X, so that in general it will be a non-Hermitian matrix. Many of these results fall into the following broad categories:

- Bounds on the probability that certain discrete iid matrices are singular;
- Quantitative lower tail estimates for the smallest singular value;
- Strong versions of the universal *circular law* describing the limiting global distribution of the spectrum in the complex plane;
- Various *local universality* results describing the asymptotic fine-scale distribution of eigenvalues.

The above items cover various results by many authors, which we discuss in more detail in the chapters below. "Universality" refers to the phenomenon that in the large n limit, many spectral statistics of random matrices depend on only a few details of the distribution of the entries, such as their first two moments.

An overarching goal of this thesis is to extend some of these results to cover various *non-iid* matrix models – that is, random matrices with entries that are either dependent or not identically distributed. We present results falling into the first three categories above. Particular attention is given to adjacency matrices of large random regular digraphs ("rrd matrices"), which are 0/1 matrices drawn uniformly at random under the constraint that all row and column sums are equal to a parameter d (the degree), which may grow with the matrix dimension n. Here the entries are identically distributed Bernoulli variables which are jointly dependent. We also consider "structured random matrices", which have independent entries with a specified pattern of means and variances, as well as a hybrid model where one attaches iid weights to the nonzero entries of an rrd matrix.

The thesis is based on the following four papers written by the author in the last two years:

1. "Discrepancy properties for random regular digraphs",
2. "On the singularity probability for random regular digraphs",
3. "The smallest singular value for structured random matrices",
4. "The circular law for signed random regular digraphs".

The first two papers have been accepted for publication in *Random Struct. Algor.* and *Probab. Theory Relat. Fields*, respectively. These papers are roughly reproduced in the four subsequent chapters of this thesis. Some some minor edits and rearrangements have been made from the original papers in order to avoid repetition and to give a more unified presentation.

A central theme that emerges from these works is a close connection between the invertibility of a random matrix and graph expansion properties enjoyed by its support. Here "support" can either mean the set of nonzero entries or, in the case of structured random matrices, the set of entries with variance exceeding some threshold; the support is associated to a directed graph in the natural way. A prior instance of this connection is a result of Rudelson and Zeitouni [RZ16] on the smallest singular value for a class of Gaussian matrices, which we extend to matrices with more general entry distributions – see Chapter 4.

1.1 Overview of chapters

Discrepancy properties for random regular digraphs

The material in Chapter 2 is based on the author's work in [Coo16]. This chapter is concerned with establishing various regularity properties holding with high probability for the distribution of edges in a random regular digraph. In particular, in Theorem 2.1.5 we give sharp tail estimates for the codegree of a fixed pair of vertices, and for the number of edges passing from one fixed set of vertices to another. Applying the union bound, we conclude that off a very small event the densities of edges between all sufficiently large sets of vertices are uniformly close to their expectation. In terms of the adjacency matrix this means that with high probability all sufficiently large submatrices have density very close to the density of the whole matrix. Such estimates are called *discrepancy properties* in the random graphs literature.

The tail estimates in Theorem 2.1.5 are established using a theorem of Chatterjee for concentration of measure by the *method of exchangeable pairs* [Cha07], a technique originally developed by Stein for proving normal approximation of random variables [Ste72]. A natural way to create exchangeable pairs of random regular digraphs is to apply the well-known *switching* operation, which we use to obtain our concentration estimates for edge densities. However, it turns out that to obtain tight control on the influence of a random switching on edge densities we first need to prove that codegrees are uniformly close to their expectations. The exchangeable pairs developed to prove concentration of codegrees are formed using a novel "reflection" involution on regular digraphs.

This is the only chapter where we are not directly concerned with *spectral* properties of random matrices. However, we note that the discrepancy properties we obtain are closely related to the property of having a spectral gap, though the latter is not implied by our bounds. In a separate work [CGJ] with Goldstein and Johnson we apply the method of *size biased couplings*, another offshoot of Stein's method (see [CGS11, GR96, GG11]), to prove a bound of optimal order on the spectral gap for random undirected regular graphs of growing degree, extending a classic result of Kahn and Szemerédi for graphs of bounded degree [FKS89]. The work [CGJ] is not included in this thesis.

The main result of this chapter, Theorem 2.1.5, is applied extensively in Chapter 3 (indeed, this result was obtained as an offshoot of our work on the singularity probability problem), and is also used in Chapter 5.

The singularity probability for random regular digraphs

The material in Chapter 3 is based on the author's work in [Coo15]. The main result of this chapter is Theorem 3.1.2, which says that an rrd matrix is invertible with probability tending to one provided the degree of the graph and of its complement grows at least poly-logarithmically. We also establish Theorem 3.1.13, which proves invertibility with high probability for a general class of matrices obtained by taking the Hadamard (entrywise) product between a matrix of iid Bernoulli signs and a fixed 0/1 matrix whose associated digraph satisfies certain graph expansion properties. These expansion properties can easily be shown to hold for random regular digraphs using the results from Chapter 2, which allows us to deduce that a signed rrd matrix – the Hadamard product between an rrd matrix and a matrix of iid signs – is invertible with high probability.

Our approach for all of these results is broadly influenced by the proof by Komlós [Kom67] that an iid Bernoulli matrix is invertible with high probability. What makes rrd matrices more difficult is the dependencies among the entries. To get around this, we use a key technique of "injecting independence" by applying several independent switching operations within a pair of rows of the matrix (giving a "shuffling coupling" – see Definition 3.3.1), allowing us to mimic parts of Komlós's argument. The discrepancy properties from Chapter 2 are used repeatedly to avoid various bad events that can arise due to the dependencies among entries.

The smallest singular value for structured random matrices

In Chapter 4 we turn from rrd matrices to a different class of matrix models which we call *structured random matrices*. These take the form $M = A \circ X + B = (a_{ij}\xi_{ij} + b_{ij})$, where A

and B are fixed matrices (which we call the "standard deviation profile" and "mean profile" for M) with $a_{ij} \geq 0$, and X is an iid matrix (here $\circ$ denotes the Hadamard product). Our main question is: under what conditions on A and B, as well as the distribution of ξ_{ij}, can we obtain polynomial lower bounds on the smallest singular value of M holding with high probability? As in the previous chapter, the main theme is that we can establish such bounds if A satisfies certain graph discrepancy or expansion properties.

We have two main results. The first, Theorem 4.1.14, is an extension of a result of Rudelson and Zeitouni [RZ16] for the case that X is an iid Gaussian matrix and A satisfies a certain set of expansion assumptions called "broad connectivity". We extend their result to a general class of entry distributions ξ_{ij}.

The main result of the chapter, Theorem 4.1.17, shows that if one takes B to be a diagonal matrix with entries bounded away from zero, then we obtain the desired lower tail estimates for the smallest singular value for essentially arbitrary A (assuming only that the entries a_{ij} are uniformly bounded). The proof makes use of Szemerédi's regularity lemma (more precisely, a version for digraphs due to Alon and Shapira – see Lemma 4.6.2) to partition the matrix A into submatrices that have good discrepancy properties. The invertibility of submatrices is then lifted back to the whole matrix using the *Schur complement formula* and some graph theoretic arguments.

The work in this chapter was motivated by the problem of proving convergence of the empirical spectral distribution for centered random matrices with a "variance profile", i.e. $n \times n$ matrices of the form

$$Y = \frac{1}{\sqrt{n}} A \circ X = \Big(\frac{1}{\sqrt{n}} a_{ij} \xi_{ij} \Big) \tag{1.1}$$

(normalized so that the eigenvalues of Y are at scale ~ 1). This problem aims for a generalization of the circular law, which covers the case that A is the all-ones matrix. A crucial step is to obtain lower tail estimates for the smallest singular value of scalar shifts $Y - z$ for arbitrary fixed $z \in \mathbb{C}$ (here we identify z with a scalar multiple of the identity matrix). As a consequence of Theorem 4.1.17 we obtain such bounds for any fixed z nonzero (see

Corollary 4.1.21), which is sufficient for the purposes of proving convergence of spectral distributions. This result is used in a separate work [CHNR] with Hachem, Najim and Renfrew to establish the limiting spectral distribution for a wide class of matrices Y of the form (1.1). The work in [CHNR] is not included in this thesis.

The circular law for signed random regular digraphs

In the final chapter we combine some of our results from Chapters 2 and 4 to prove that the circular law holds for the limiting spectral distribution of *signed rrd matrices*, which are matrices of the form $A \circ X = (a_{ij}\xi_{ij})$, where A is the adjacency matrix of a uniform random regular digraph and X is a matrix of iid Bernoulli signs. Here we cover the case that the degree of the digraph is linear in n. The introduction to Chapter 5 also contains an overview of existing results and conjectures on the extent of the circular law universality class.

1.2 Notation

The following notation is used repeatedly throughout the thesis. Each chapter also contains a section that lists notational conventions specific to that chapter.

$[n]$	The set of natural numbers $\{1, \dots, n\}$.
$\mathcal{M}_{n,m}(S)$	The set of $n \times m$ matrices with entries lying in the set S (which will generally be $\mathbb{R}, \mathbb{C}, \{+1, -1\}$ or the interval $[0, 1]$).
$\mathcal{M}_n(S)$	Abbreviation for $\mathcal{M}_{n,n}(S)$ as above.
$\{s_k(M)\}_{k=1}^n$	The singular values of a matrix $M \in \mathcal{M}_n(\mathbb{C})$, which are counted with multiplicity and arranged in non-increasing order $s_1(M) \geq \cdots \geq s_n(M) \geq 0$.
$\{\lambda_k(M)\}_{k=1}^n$	The eigenvalues of a matrix $M \in \mathcal{M}_n(\mathbb{C})$, which are counted with multiplicity and labeled in some arbitrary fashion.

$\|\cdot\|$	Denotes the Euclidean ℓ_2^n norm when applied to vectors in $\mathbb{R}^n$ or $\mathbb{C}^n$, and the $\ell_2^m \to \ell_2^n$ operator norm when applied to $n \times m$ matrices. Other norms are indicated with subscripts.
S^{n-1}	The unit sphere in $\mathbb{C}^n$ with respect to the Euclidean norm.
C, C', c, c_0, etc.	Unspecified constants whose value may change from line to line. Understood to be absolute unless indicated otherwise.
$f \ll g$, $g \gg f$	(Synonymous) There is a constant C such that $\|f\| \le Cg$. Unless otherwise remarked the implied constant C will be independent of all parameters. We indicate dependence of the implied constant on parameters with a subscript, e.g. $f \ll_\alpha g$.
$f = O(g)$, $g = \Omega(f)$	Both are synonymous to the above.
$f \asymp g$	$f \ll g$ and $g \ll f$.
$f = \Theta(g)$	Synonymous to the above.
$f = o(g)$	$f/g \to 0$ as $n \to \infty$ (for quantities f, g depending on a natural number n, which will always denote the dimension of the random matrix under consideration). We will occasionally use this notation when the limit is taken with respect to some other parameter, in which case we indicate the limit with a subscript, e.g. $f = o_{r\to 0}(g)$ means $f(r)/g(r) \to 0$ as $r \to 0$. We indicate that the rate of convergence depends on the value of some parameter with a subscript, e.g. $f = o_\alpha(g)$.
$g = \omega(f)$	Synonymous to the above.
$\mathbb{E}\, X$	Mathematical expectation of a random variable X defined on an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ (which will remain hidden throughout the thesis).
$X \overset{d}{=} Y$	(For random variables X, Y) X is identically distributed to Y.
$\mathbb{1}_{\mathcal{E}}$	The indicator random variable corresponding to an event $\mathcal{E}$. For a statement P, $\mathbb{1}(P) := \mathbb{1}_{\{P \text{ holds}\}}$.

CHAPTER 2

Discrepancy properties for random regular digraphs

2.1 Introduction

For $n \geq 1$ and $d \in [n] = \{1, \ldots, n\}$, let $\mathcal{D}_{n,d}$ denote the set of d-regular directed graphs on n labeled vertices – that is, with each vertex having d in-neighbors and d out-neighbors (allowing self-loops). Let $\Gamma = (V, E)$ be a uniform random element of $\mathcal{D}_{n,d}$. One may identify Γ with a uniform random d-regular bipartite graph on $n + n$ vertices in the obvious way. We will stick with the digraph interpretation, though we note that all of our results can be extended to cover (d, d')-regular bipartite graphs on $m + n$ vertices; see Section 2.1.4.

Our aim in this chapter is to show that two types of statistics of Γ are sharply concentrated when n is large and d is sufficiently large depending on n. We identify V with $[n]$ throughout, and view E as a subset of $[n]^2$. We sometimes write $i \to j$ to mean $(i, j) \in E$.

1. **Codegrees:** Denote the number of common out-neighbors of a fixed pair of vertices $i_1, i_2 \in [n]$ by
$$\overrightarrow{\mathrm{co}}_\Gamma(i_1, i_2) := \left| \left\{ j \in [n] : i_1 \to j \text{ and } i_2 \to j \right\} \right|$$
and the number of common in-neighbors by
$$\overleftarrow{\mathrm{co}}_\Gamma(i_1, i_2) := \left| \left\{ j \in [n] : i_1 \leftarrow j \text{ and } i_2 \leftarrow j \right\} \right|.$$
We expect these statistics to be of size roughly $d^2/n = p^2 n$, where we denote by $p := d/n$ the average edge density for Γ.

2. **Edge counts:** For fixed subsets of vertices $A, B \subset [n]$, denote the number of edges passing from A to B by

$$e_\Gamma(A, B) := |E \cap (A \times B)|.$$

We expect this statistic to be of size roughly $p|A||B| =: \mu(A, B)$. We refer to the deviation

$$\mathrm{disc}_\Gamma(A, B) := \big|e_\Gamma(A, B) - \mu(A, B)\big| \tag{2.1}$$

as the *(edge) discrepancy of* Γ *at* (A, B).

We will loosely use the term *edge discrepancy property* to refer to a bound on edge discrepancies holding uniformly for all pairs (A, B), or at least for all pairs of "sufficiently large" sets A, B.

The discrepancy properties and control on codegrees proved in this chapter are an important component for the proof in Chapter 3 that random 0/1 matrices with constant row and column sum d are invertible with high probability, assuming $\min(d, n - d) \geq C_0 \log^2 n$ for sufficiently large absolute constant C_0. We expect that the results of the present chapter will also be useful for questions of a more graph-theoretic nature.

2.1.1 Background on random regular graphs

Random graphs have been studied intensively since their popularization by Erdős as a tool for proving of the existence of graphs with certain properties, often when no constructive approach was known (such as graphs with arbitrarily large chromatic number and girth; see [AS08]). They have since found myriad applications in computer science, physics, biology, and other fields. The most commonly used model is the *binomial* or *Erdős–Rényi* random graph $G(n, p)$, in which each of the $\binom{n}{2}$ possible edges is present independently of all others with probability p. We may similarly define the Erdős–Rényi digraph $D(n, p)$, which has n^2 possible directed edges.

Random regular graphs emerged as a popular model much later, and their origin can also be traced back to a question in extremal combinatorics: are there expander graphs of bounded degree? (Strictly speaking the term "expander" only makes sense for a sequence of graphs; the reader may consult the survey [HLW06] for a precise statement of this question.) This was answered in the affirmative by Pinsker in 1973 [Pin73] (and independently by Barzdin and Kolmogorov in the bipartite case [Bar93]) who showed that certain random regular graphs of constant degree are expanders with positive probability.

Since then, much of the interest in random regular graphs has been due to their robust connectivity properties as compared to Erdős–Rényi graphs. Indeed, while Erdős–Rényi graphs are asymptotically almost surely *disconnected* when the average degree is smaller than $\log n$, random regular graphs are not only connected with high probability for degree as small as 3, they are *nearly Ramanujan* (meaning they are near-optimal expanders in a certain sense; see [Fri08]).

Random regular graphs are often harder to analyze than their binomial counterparts since the d-regularity constraint destroys the independence of the edges. Nevertheless, asymptotic enumeration results were obtained in [BC78], [Bol80] and [Wor80]. The introduction by Bollobás in [Bol80] of the *configuration model* for the uniform random regular graph allowed for many later developments (ideas similar to the configuration model were also present in [BC78] and [Wor80]). Here one generates a uniform random regular graph by the following procedure:

1. Associate to each vertex $v \in V$ a "fiber" F_v of d "points", so that there are

$$\left| \bigcup_{v \in V} F_v \right| = nd$$

points in total.

2. Select a pairing $\mathcal{P}$ of the nd points uniformly at random.

3. Now collapse each fiber F_v to the associated vertex v: we say that v is connected to w if there are points $v' \in F_v$, $w' \in F_w$ such that $v'w' \in \mathcal{P}$. In general the resulting graph $G = G(\mathcal{P})$ is a d-regular *multi-graph*; however, conditional on the event $\mathcal{E}_{\text{simple}}$ that that $\mathcal{P}$ collapses to a simple graph, it is easy to check that $G(\mathcal{P})$ is a uniform random

d-regular graph. Hence we may

4. repeat this process if necessary until we obtain a simple graph.

The procedure can be modified to generate uniform random d-regular directed or bipartite graphs in the obvious manner. When using the configuration model to bound the probability of an event $\mathcal{B}$ holding for a uniform random regular graph, one "lifts" $\mathcal{B}$ to the corresponding event $\mathcal{B}'$ for the pairing $\mathcal{P}$, which is often easier to analyze. Then one can bound

$$\mathbb{P}(\mathcal{B}) = \mathbb{P}(\mathcal{B}'|\mathcal{E}_{\text{simple}}) \leq \frac{\mathbb{P}(\mathcal{B}')}{\mathbb{P}(\mathcal{E}_{\text{simple}})}. \tag{2.2}$$

A particularly nice feature is that working with the random pairing $\mathcal{P}$ rather than with the graph G gives access to concentration of measure inequalities for martingale sequences (e.g. the Azuma–Hoeffding inequality). (However, we will see below that the method of exchangeable pairs can be applied directly to the uniform measure on random regular graphs.) A drawback is that $\mathbb{P}(\mathcal{E}_{\text{simple}})$ becomes quite small when the degree d is large. Indeed, the enumeration result of [BC78] implies the estimate

$$\mathbb{P}(\mathcal{E}_{\text{simple}}) = \exp\left(-\Theta(d^2)\right) \tag{2.3}$$

for d fixed (see Section 2.1.5 for definitions of asymptotic notation used in this chapter). This asymptotic was later shown by McKay and Wormald in [MW91] to hold when $d = o(\sqrt{n})$.

An advantage of the concentration results given in Theorem 2.1.5 below is that they are proved for the uniform random regular digraph directly, rather than through the configuration model, and hence do not have to compete with with the small probability in (2.3). In particular, our results are not limited to $d = o(\sqrt{n})$ (in fact the bounds are strongest for dense graphs).

Our approach is to exploit symmetries of the set $\mathcal{M}_n(d)$ of a "local nature", i.e. ones that change only a small number of edges. This leads naturally to the method of *switchings*, developed by McKay and Wormald in several works (see the survey [Wor99]). For regular digraphs, perhaps the most obvious symmetry is to change between the following two configurations of edges at fixed vertices i_1, i_2, j_1, j_2:

where we use a solid arrow to depict an edge and a dashed arrow to indicate the absence of an edge. We refer to this modification as a *simple switching*. Roughly speaking, the probability that an event $\mathcal{B}$ holds for random regular graphs can be estimated by performing a switching in a random fashion (such as by sampling the indices i_1, i_2, j_1, j_2 at random) and estimating the probability that the graph enters or leaves the event $\mathcal{B}$ under the application of the switching.

McKay introduced the method of switchings in [McK81b], and in [McK81a] used it to prove bounds on the probability of occurrence of cycles of various length in a random regular graph of bounded degree. Through the trace method this allowed him to deduce that the limiting spectral distribution of the adjacency matrix is that of the infinite d-regular tree, now known as the *Kesten–McKay distribution*. (For d tending to infinity with n, the spectral distribution is instead governed by the *semi-circle law*, as was proved by Dumitriu and Pal [DP12] in the sparse regime $d = n^{o(1)}$, and by Tran, Vu and Wang in the general case by a different strategy [TVW13], which we describe in Section 3.1.1 – see the discussion around (3.7).) Since then, the method has been extended and applied to several problems on random regular graphs, such as to extend the asymptotic enumeration results of [BC78], [Bol80], [Wor80] to $d = o(\sqrt{n})$ in [MW91]. See the survey [Wor99] for more background on switchings. See also [McK] for a simple illustration of the method for the problem of estimating the probability that a random permutation has a fixed point.

2.1.2 Codegrees, edge discrepancy, and pseudo-randomness

Parallel to the study of random graphs, there has been a rich literature on *pseudo-random graphs*, which is an imprecise term for deterministic graphs that exhibit properties held by (Erdős–Rényi) random graphs with high probability. Systematic research into pseudo-random graphs was initiated by Thomason in [Tho85], [Tho87], where he introduced the notion of *jumbled graphs* (see Definition 2.1.11 below). In [CGW89], Chung, Graham and Wilson defined *quasi-random graphs* and proved that several "pseudo-randomness" properties are in fact equivalent. See also the survey [KS06] and Chapter 9 of [AS08].

In particular, the works [Tho85] and [CGW89] highlighted a close connection between codegrees and edge discrepancy, the quantities of interest for the present work. We have the following result from [KSVW01] deducing a discrepancy property from uniform control on codegrees, proved earlier for the Erdős–Rényi case in [AKS99], and essentially going back to [Tho85]. (While the result in [KSVW01] was stated for undirected graphs, the following version can be obtained by following similar lines to the proof given there.)

Lemma 2.1.1 (Pseudorandomness [KSVW01])**.** *Let Γ be a fixed element of $\mathcal{D}_{n,d}$ with the property that for some $\varepsilon > 0$ and for every $i_1, i_2 \in [n]$ distinct,*

$$\max\Big[\, \overrightarrow{\mathrm{co}}_\Gamma(i_1,i_2)\ ,\ \overleftarrow{\mathrm{co}}_\Gamma(i_1,i_2)\,\Big] \le (1+\varepsilon)p^2 n. \tag{2.4}$$

Then for any pair of sets $A, B \subset [n]$ such that $|A|, |B| \ge \frac{1}{\varepsilon}\frac{n}{d} = (\varepsilon p)^{-1}$, we have

$$\left|\frac{e_\Gamma(A,B)}{p|A||B|} - 1\right| \le \left[\frac{2\varepsilon n}{\max(|A|,|B|)}\right]^{1/2}. \tag{2.5}$$

Remark 2.1.2. Note that in order to have concentration of $e_\Gamma(A,B)$ at the scale of the mean $p|A||B|$, the lemma requires that one of the sets be of size linear in n. Theorem 2.1.5 below will allow us to extend this to much smaller sets.

Lemma 2.1.1 can be used to deduce control on edge discrepancy for random regular digraphs holding asymptotically almost surely (a.a.s.), as soon as one can show that (2.4) holds a.a.s. This was the route taken in [KSVW01] for the undirected case by Krievelevich, Sudakov, Vu and Wormald, who obtained the following concentration result for codegrees in sufficiently dense d-regular graphs.

Theorem 2.1.3 (From Theorem 2.1 in [KSVW01])**.** *Let G be a uniform random d-regular* *undirected graph on n vertices. Suppose that*

$$\omega(\sqrt{n \log n}) \le d < n - cn/\log n$$

for some constant $c > 2/3$. Then asymptotically almost surely we have

$$\max_{i_1, i_2 \in V} \left| \mathrm{co}_G(i_1, i_2) - \frac{d^2}{n} \right| < C\frac{d^3}{n^2} + 6d\sqrt{\frac{\log n}{n}} \tag{2.6}$$

for some $C > 0$ absolute. If $d \ge cn/\log n$ we may take C to be zero.

Remark 2.1.4. Theorem 2.1 in [KSVW01] also states some weaker upper bounds on codegrees valid for smaller d, which we have omitted.

The proof of Theorem 2.1.3 divides into two (overlapping) cases. For $\min(d, n - d) \ge cn/\log n$ the proof uses an asymptotic enumeration formula for dense graphs with given degree sequence, proved in [MW90]. The method of switchings is used for the case $d = o(n)$. The proof shows that the estimate $o(1)$ for the probability that (2.6) fails is in fact $O(n^{-c})$ for some $c > 0$ absolute.

2.1.3 Results

We combine variants of the switching method of McKay and Wormald with the method of exchangeable pairs for concentration of measure, as developed by Chatterjee in [Cha07], to prove exponential tail bounds on codegrees and edge discrepancies. For edge discrepancies we use the simple switching coupling, reviewed in Section 2.3.1, while for concentration of codegrees we employ a novel (to our knowledge) "reflection" coupling, described in Section 2.3.2.

For both $e_\Gamma(A, B)$ and $\vec{\mathrm{co}}_\Gamma(i_1, i_2)$ we are able to prove tail bounds that match (up to constant factors in the exponential) what can be obtained in the Erdős–Rényi case using Chernoff bounds (specifically, Bernstein's inequality). As a consequence, we can combine our concentration estimates with union bounds to prove discrepancy properties essentially matching those available for Erdős–Rényi digraphs. We review the (brief and completely standard) proofs of analogous results for the Erdős–Rényi case in Section 2.2.1 for comparison.

It is possible that our approach can be extended to prove similar results for undirected (non-bipartite) random regular graphs, but we do not pursue this matter here. It is also likely that our methods can be applied to the study of directed multi-graphs with given (non-constant) degree sequence.

Before stating our main theorem we set up some notation. Due to the constraint of d-regularity, a deviation of $e_\Gamma(A,B)$ from its mean coincides with an equal deviation of $e_\Gamma(A^c,B^c)$, where we denote $A^c := [n]\setminus A$. Indeed, if $e_\Gamma(A,B)=k$, we have from d-regularity that

$$\begin{aligned} e_\Gamma(A^c,B) &= d|B|-k,\\ e_\Gamma(A,B^c) &= d|A|-k,\\ e_\Gamma(A^c,B^c) &= d(n-|A|-|B|)+k. \end{aligned}$$

It follows from the last line that for any $t\in\mathbb{R}$, the following identity of events holds:

$$\big\{e_\Gamma(A,B)-\mu(A,B)\ge t\big\} = \big\{e_\Gamma(A^c,B^c)-\mu(A^c,B^c)\ge t\big\}. \tag{2.7}$$

It is hence natural to consider deviations of $e_\Gamma(A,B)$ at the scale

$$\begin{aligned} \hat\mu(A,B) &:= \min\big(\mu(A,B),\mu(A^c,B^c)\big) && (2.8)\\ &= p\min\big(|A||B|,(n-|A|)(n-|B|)\big). && (2.9) \end{aligned}$$

We will often suppress the dependence of μ and $\hat\mu$ on A,B. We also denote

$$\hat d := \min(d,n-d) \tag{2.10}$$

and $\hat p := \hat d/n$, the minimum of the edge density of $\Gamma=([n],E)$ and its complement $\Gamma' = ([n],[n]^2\setminus E)$.

The main result of this chapter can be summarized as follows:

1. With high probability, codegrees are uniformly close to p^2n.

2. Restricted to the (likely) event that all codegrees are roughly p^2n, we have concentration with exponential tails for the edge discrepancy at fixed pairs of sets A, B.

Theorem 2.1.5 (Main result)**.** *For $\eta \geq 0$ define the event*

$$\mathcal{G}^{\mathrm{co}}(\eta) = \left\{\forall \{i_1, i_2\} \subset [n], \ \left|\vec{\mathrm{co}}_\Gamma(i_1, i_2) - p^2n\right| \leq \eta p(1-p)n\right\}. \tag{2.11}$$

We have

1. *(Uniform control on codegrees) For any $\eta \geq 0$, $\mathcal{G}^{\mathrm{co}}(\eta)$ holds except with probability*

$$O\left(n^2\hat{d}^2 \exp\left(-c\eta \min\left\{\hat{d}, \eta n\right\}\right)\right). \tag{2.12}$$

   In particular, for any $K_1 > 0$ there exists $K_2 > 0$ such that $\mathcal{G}^{\mathrm{co}}(\eta)$ holds with probability $1 - O\big(n^{-K_1}\big)$ if

$$\eta \geq K_2 \max\left\{\frac{\log n}{\min(d, n-d)}, \sqrt{\frac{\log n}{n}}\right\}. \tag{2.13}$$

2. *(Concentration of edge counts) For any $A, B \subset [n]$ and any $\tau \geq 0$,*

$$\mathbb{P}\Big(\Big\{e_\Gamma(A,B) - \mu \geq \tau\hat{\mu}\Big\} \wedge \mathcal{G}^{\mathrm{co}}\big(\eta\big)\Big) \leq \exp\left(-\frac{\tau^2\hat{\mu}}{C_1 + C_2\tau}\right) \tag{2.14}$$

   provided $\eta \leq \min\left(\frac{1}{4}, \frac{\tau}{8}\right)$, and

$$\mathbb{P}\Big(\Big\{e_\Gamma(A,B) - \mu \leq -\tau\hat{\mu}\Big\} \wedge \mathcal{G}^{\mathrm{co}}\big(\eta\big)\Big) \leq \exp\left(-\frac{\tau^2\hat{\mu}}{C_1}\right) \tag{2.15}$$

   provided $\eta \leq \frac{\tau}{4}$, where $C_1, C_2 > 0$ are absolute constants. In particular, if $\eta \leq$

$\min\left(\frac{1}{4}, \frac{\tau}{8}\right)$, *we have*

$$\mathbb{P}\Big(\Big\{ \operatorname{disc}_\Gamma(A,B) \ge \tau\hat{\mu}(A,B)\Big\} \wedge \mathcal{G}^{\mathrm{co}}(\eta)\Big) \le 2\exp\left(-\frac{\tau^2}{C_1 + C_2\tau}\hat{\mu}(A,B)\right). \tag{2.16}$$

Remark 2.1.6. The proof shows that one may take $C_1 = 64$, $C_2 = 8$, though we make little effort to optimize these values.

Remark 2.1.7. In order to deduce that $e_\Gamma(A,B)$ is within an arbitrarily small fixed multiplicative error of its mean $\mu = p|A||B|$ using Theorem 2.1.5, one must assume $\min(d, n-d) = \omega(\log n)$. Indeed, we want to take τ as small as we like in (2.16), which requires taking $\eta \le \tau/8$. Now to deduce that $\mathcal{G}^{\mathrm{co}}(\tau/8)$ holds a.a.s. from part (1), we must take $\hat{d} = \min(d, n-d) \ge C\tau^{-1}\log n$ for a sufficiently large constant C. See Theorem 2.1.13 below for a result which is valid for $d = O(\log n)$, but for a slightly different model of random regular digraph (the permutation model).

Remark 2.1.8 (Comparison to the Erdős–Rényi case). For A, B such that $|A| + |B| \le n$ (i.e. such that $\mu(A,B) = \hat{\mu}(A,B)$), the bound (2.16) is the same as what one obtains in the Erdős–Rényi case from Bernstein's inequality, up to modification of the constants C_1, C_2 – see Section 2.2.1 and the bound (2.46). For the case $|A| + |B| > n$ the bound (2.16) becomes superior to (2.46). This is due to the identity (2.7) (which comes from d-regularity): if A, B are of size close to n, a large deviation of $e_\Gamma(A,B)$ coincides with a *very* large deviation of $e_\Gamma(A^c, B^c)$. (Of course, the most concentrated statistic of all is $e_\Gamma(V,V) = dn$, which is deterministic, while this random variable has variance $p(1-p)n^2 \asymp n\min(d, n-d)$ in the Erdős–Rényi model.)

Our proof of both parts of Theorem 2.1.5 is by the method of exchangeable pairs. Roughly speaking, to prove concentration of a statistic $f(\Gamma)$ of the random digraph Γ, the method is to analyze the change in f under a small random change to Γ. To prove the concentration of edge counts in part (2) we will use the simple switching operation on digraphs, reviewed in Section 2.3.1. For the concentration of codegrees in part (1) we use an operation on digraphs which we call "reflection". Reflections are less local in nature than simple switchings; the construction is given in Section 2.3.2.

Theorem 2.1.5 can be viewed as an improvement on the deterministic Lemma 2.1.1 for the setting of random graphs. Like Lemma 2.1.1 it deduces some control on edge discrepancy after restricting to a "good" event on which there is some uniform control on the codegrees. The key differences are the following:

1. Rather than deduce a deterministic bound on edge discrepancy from the control on codegrees as in Lemma 2.1.1 (which obtained an essentially optimal bound), Theorem 2.1.5 gives much tighter bounds holding with high probability.

2. The control on codegrees summarized in the event $\mathcal{G}^{\mathrm{co}}(\eta)$ differs in two respects: on the one hand we allow fluctuations at scale $p(1-p)n$ rather than p^2n, which is less stringent for sparse graphs, while on the other hand we need both lower and upper bounds.

One can deduce various discrepancy properties for Γ holding with high probability using Theorem 2.1.5, and essentially matching standard discrepancy properties for Erdős–Rényi digraphs (since our tail bounds match the bounds (2.45) and (2.46) for Erdős–Rényi digraphs up to constants in the exponential). There is flexibility with the range of sets to consider and the tolerance level for edge discrepancy; the choice will be dictated by the application at hand. We now give one example.

Suppose one desires to have $e_\Gamma(A,B)$ within a small factor of its expectation $p|A||B|$ for all pairs of sufficiently large sets A,B. The following corollary shows that this is satisfied with high probability.

Corollary 2.1.9. *Let $C_0>0$ be a sufficiently large absolute constant. For $\varepsilon\in(0,1)$, let $\mathcal{G}(\varepsilon)$ denote the event that for all $A,B\subset[n]$ such that*

$$|A|,|B|\ge\frac{C_0\log n}{\varepsilon^2 p} \tag{2.17}$$

we have $\mathrm{disc}_\Gamma(A,B)\le\varepsilon\hat{\mu}(A,B)$. If $\hat{d}=\min(d,n-d)\ge C_0\varepsilon^{-1}\log n$, then $\mathcal{G}(\varepsilon)$ holds except with probability

$$O\left(\exp\left(-c\min\left\{\varepsilon\hat{d},\varepsilon^2 n,\frac{n}{\varepsilon^2 d}\log^2 n\right\}\right)\right). \tag{2.18}$$

Proof. By the lower bound on $\hat{d}$ and part (1) of Theorem 2.1.5 we have

$$\mathbb{P}(\mathcal{G}^{\mathrm{co}}(\varepsilon)) \geq 1 - \exp\left(-c \min\left\{\varepsilon \hat{d}, \varepsilon^2 n\right\}\right)$$

(taking C_0 sufficiently large to beat the polynomial factors). By abuse of notation we restrict the sample space to $\mathcal{G}^{\mathrm{co}}(\varepsilon)$. It now suffices to show

$$\mathbb{P}(\mathcal{G}(\varepsilon)) \geq 1 - C \exp\left(-c \frac{n \log^2 n}{\varepsilon^2 d}\right). \tag{2.19}$$

Since

$$\left\{ \operatorname{disc}_M(A, B) \geq \varepsilon \hat{\mu} \right\} = \left\{ \operatorname{disc}_M(A^c, B^c) \geq \varepsilon \hat{\mu} \right\}$$

it suffices to consider pairs (A, B) with $|A| + |B| \leq n$. By giving up a factor of 2 we may also assume $|A| \leq |B|$.

Set

$$a_0 = \frac{C_0 \log n}{\varepsilon^2 p}. \tag{2.20}$$

For $a_0 \leq a \leq b \leq n$ let

$$\mathcal{B}_{ab}(\varepsilon) = \Big\{ \exists A, B \subset [n] : \ |A| = a, |B| = b, \ \operatorname{disc}_M(A, B) \geq \varepsilon \hat{\mu}(A, B) \Big\}.$$

Applying part (2) of Theorem 2.1.5 and a union bound (and by our restriction to $\mathcal{G}^{\mathrm{co}}(\varepsilon)$),

$$\mathbb{P}(\mathcal{B}_{ab}(\varepsilon)) \ll \binom{n}{a}\binom{n}{b} \exp\left(-c\varepsilon^2 pab\right) \tag{2.21}$$

$$\ll \exp\left(Cb \log n - c\varepsilon^2 pab\right) \tag{2.22}$$

$$\ll \exp\left(-c\varepsilon^2 pab\right) \tag{2.23}$$

where in the last line we used that $a \geq a_0$ and took C_0 sufficiently large (adjusting the constant c). By another union bound,

$$\begin{aligned}\mathbb{P}(\mathcal{G}^e(\varepsilon)^c) &\ll \sum_{b=a_0}^{n}\sum_{a=a_0}^{b}\mathbb{P}(\mathcal{B}_{ab}(\varepsilon))\\ &\ll \sum_{b=a_0}^{n}\sum_{a=a_0}^{b}\exp\left(-c\varepsilon^2 pab\right)\\ &\ll \exp\left(-c\varepsilon^2 pa_0^2\right)\end{aligned}$$

where in the last line we performed the geometric sums. Substituting the expression (2.20) completes the proof. ☐

Remark 2.1.10. Note that in going from (2.21) to (2.23) we actually only needed

$$a \geq \frac{C_0}{\varepsilon^2}\log\frac{en}{b}.$$

Hence, we could have taken the wider class

$$\mathcal{F}(\varepsilon) = \left\{(A,B):\ A,B\subset[n],\ \min\left(|A|,\,|B|\right) \geq \frac{C_0}{\varepsilon^2}\frac{n}{d}\log\frac{en}{\max(|A|,|B|)}\right\} \tag{2.24}$$

which includes some pairs (A,B) where, say, $|A| \asymp 1/p$ and $|B| \asymp n$.

Next we state a conjecture concerning the singular value distribution for the adjacency matrix of Γ, which we denote by $M = M_\Gamma$. Denote the singular values of M by

$$d = s_1(\Gamma) \geq s_2(\Gamma) \geq \cdots \geq s_n(\Gamma) \geq 0$$

(where $s_1(\Gamma) = d$ follows from d-regularity and the Cauchy-Schwarz inequality). It is well known that control on edge discrepancy follows from a spectral gap. We recall the notion of a *jumbled* graph, introduced by Thomason [Tho85] and adapted here to the setting of digraphs.

Definition 2.1.11. Say that a digraph $D = (V, E)$ is *α-jumbled* if for all $A, B \subset [n]$ we have

$$\mathrm{disc}_D(A, B) := \big|e_D(A, B) - \mu(A, B)\big| \leq \alpha\sqrt{|A||B|}.$$

It is a straightforward exercise to show that a d-regular digraph on n vertices whose adjacency matrix has second singular value s_2 is s_2-jumbled (see for instance Theorem 2.11 in [KS06] for the undirected case; the directed case follows similar lines).

Conjecture 2.1.12. *Assume $1 \leq d \leq n$. Then asymptotically almost surely, $s_2(\Gamma) = O\big(\sqrt{d}\big)$. In particular, Γ is $O(\sqrt{d})$-jumbled.*

The singular vector corresponding to $s_1(\Gamma) = d$ is the constant vector $\frac{1}{\sqrt{n}}\mathbf{1} := \frac{1}{\sqrt{n}}(1, \ldots, 1)$. By the Courant-Fischer minimax theorem, letting S_0^{n-1} denote the set of unit vectors in $\mathbb{R}^n$ orthogonal to $\mathbf{1}$, we have

$$\begin{aligned} s_2(\Gamma) &= \sup_{u \in S_0^{n-1}} \|M_\Gamma u\| \\ &= \sup_{u \in S_0^{n-1}} \|(\mathbf{1}\,\mathbf{1}^\mathsf{T} - M_\Gamma)u\| \\ &= \sup_{u \in S_0^{n-1}} \|M_{\Gamma'} u\| \\ &= s_2(\Gamma') \end{aligned}$$

where Γ' is the complementary $(n-d)$-regular digraph. Hence, it suffice to consider $1 \leq d \leq n/2$. Using Theorem 2.1.5 and union bounds, one can show that Γ is $O(\sqrt{d})$-jumbled for the dense case $n \ll d \leq n/2$, following similar lines to the proof of Corollary 2.1.9. For the sparse case, this approach can only show that Γ is $O(\sqrt{d \log n})$-jumbled.

Conjecture 2.1.12 parallels a conjecture of Vu for the undirected case [Vu08]. For an undirected graph G with adjacency matrix M_G having real eigenvalues $\lambda_1(G) \geq \cdots \geq \lambda_n(G)$ we simply have

$$s_2(G) = \lambda(G) := \max\big\{|\lambda_2(G)|,\, |\lambda_n(G)|\big\}.$$

In [KS06], Kahn and Szemerédi proved a bound of $O(\sqrt{d})$ for $\lambda(\Pi)$, with d fixed independent of n, and with the graph Π drawn from a different distribution on random regular graphs which we call the *permutation model*. Let $P_1, \dots, P_d$ be iid uniform $n \times n$ permutation matrices, and put

$$M_\Lambda = P_1 + \cdots + P_d. \tag{2.25}$$

We may interpret M_Λ as the adjacency matrix for a random d-regular directed multi-graph Λ, and we may also associate $M_\Lambda + M_\Lambda^\mathsf{T}$ to a $2d$-regular undirected multi-graph Π. Kahn and Szemerédi proved that if d is fixed independent of n, we have

$$s_2(\Lambda) = O(\sqrt{d}) \tag{2.26}$$

asymptotically almost surely. By the triangle inequality this implies $\lambda(\Pi) = O(\sqrt{d})$ a.a.s. Their argument was later extended to allow $d = o(\sqrt{n})$ in [DJPP13], and was also adapted to the configuration model with $d = o(\sqrt{n})$ in [BFSU99]. Furthermore, the optimal bound $\lambda(\Pi) \le 2\sqrt{2d-1} + o(1)$ was obtained for fixed d by Friedman in [Fri08] by a completely different argument.

For small degree, the permutation model Λ is "close" to the uniform model Γ in the following precise sense. It was proved in [Jan95] and [MRRW97] that if d is fixed, the models

1. Γ (a uniform random element of $\mathcal{D}_{n,d}$), and

2. Λ conditioned to be simple

are *contiguous*, meaning that a sequence of events holding a.a.s. for one model will hold a.a.s. for the other. In particular, for the case that d is fixed Conjecture 2.1.12 follows from contiguity and the bound (2.26). It was also shown that the model Π is contiguous to a uniform random regular graph of fixed even degree. We believe that these models continue to be contiguous if $d = O(\log n)$, though we are not aware of any such results in the literature.

We record an analogue of our main theorem for the permutation model Λ. The following result has no restrictions on d and hence can serve as a substitute for (2.16) for sparser regular digraphs (recall that Theorem 2.1.5 is most useful when $\min(d, n-d) = \omega(\log n)$ – see Remark 2.1.7).

Theorem 2.1.13 (Concentration of edge counts, permutation model)**.** *Let $n, d \geq 1$, and $A, B \subset [n]$. For any $\tau \geq 0$ we have*

$$\mathbb{P}\left(\left|e_\Lambda(A,B) - \mu\right| \geq \tau\mu\right) \leq 2\exp\left(-\frac{\tau^2\mu}{2+\tau}\right) \tag{2.27}$$

where $\mu = p|A||B|$ as before, and $e_\Lambda(A,B)$ is the number of directed edges from A to B, counting multiplicity.

The above theorem is considerably easier to establish than part (2) of Theorem 2.1.5 – it turns out that the independence between the d factors allows one to proceed with the method of switchings without needing a priori bounds on codegrees. We will hence prove Theorem 2.1.13 as a warmup in Section 2.2.

Sharper bounds for larger deviations (i.e. when τ is large) can be proved by directly estimating $\mathbb{P}(e(A,B) = t)$ for all $t \in \mathbb{N}$, leading to an estimate on the moment generating function $m(\theta) = \mathbb{E}\exp(\theta\, e(A,B))$. This was the route taken in [DJPP13] to prove a certain discrepancy property for the permutation model.

2.1.4 Extension to general bipartite regular graphs

Theorem 2.1.5 above easily extends to the following more general setting. For $m, n \geq 1$ and $d \in [n]$, $d' \in [m]$, draw $\Gamma = (U, V, E)$ uniformly from the set of bipartite graphs on parts U, V with $|U| = m$, $|V| = n$ and edge set $E \subset U \times V$, with the constraint that each $i \in U$ has degree d and each $j \in V$ has degree d'. Since the total number of edges is

$$md = nd'$$

we denote

$$\theta = \frac{m}{n} = \frac{d'}{d}. \tag{2.28}$$

The random regular digraph considered above corresponds to the case $\theta = 1$. As before, we identify U with $[m]$ and V with $[n]$, denote by $p = d/n = d'/m$ the edge density of Γ, $\hat{d} := \min(d, n-d)$, and $\hat{\mu}(A,B) := p\min\{|A||B|, (m-|A|)(n-|B|)\}$.

The following result is proved by the same lines as Theorem 2.1.5, only with slightly more burdensome notation.

Theorem 2.1.14 (Extension to bipartite graphs)**.** *For $\eta \geq 0$ define the event*

$$\mathcal{G}^{\mathrm{co}}(\eta) = \left\{\forall\,\{i_1, i_2\} \subset [m],\ \left|\vec{\mathrm{co}}_\Gamma(i_1,i_2) - p^2 n\right| \leq \eta p(1-p)n\right\}. \tag{2.29}$$

We have

1. *(Uniform control on codegrees) For any $\eta \geq 0$, $\mathcal{G}^{\mathrm{co}}(\eta)$ holds except with probability*

$$O\left(m^2\hat{d}^2 \exp\left(-\frac{c\eta n^2}{m}\right)\right) + O\left(m^2 \exp\left(-c\eta \min\left\{\hat{d}, \eta n\right\}\right)\right). \tag{2.30}$$

   In particular, $\mathcal{G}^{\mathrm{co}}(\eta)$ holds a.a.s. in the limit $m, n \to \infty$ as long as

$$\eta \geq C \max\left\{\frac{m\log(m\hat{d})}{n^2}, \frac{\log m}{\hat{d}}, \sqrt{\frac{\log m}{n}}\right\}$$

   for some $C > 0$ sufficiently large.

2. *(Concentration of edge counts) For any $A \subset [m]$, $B \subset [n]$ and any $\tau \geq 0$, if $\eta \leq \min\left(\frac{1}{4}, \frac{\tau}{8}\right)$, we have*

$$\mathbb{P}\Big(\Big\{\operatorname{disc}_\Gamma(A,B) \geq \tau\hat{\mu}(A,B)\Big\} \wedge \mathcal{G}^{\mathrm{co}}(\eta)\Big) \leq 2\exp\left(-\frac{\tau^2}{C_1 + C_2\tau}\hat{\mu}(A,B)\right). \tag{2.31}$$

The rest of the chapter is organized as follows. In Section 2.2 we introduce and motivate Chatterjee's method of exchangeable pairs in the context of two random digraph models that are simpler to analyze than the uniform random regular digraph, namely the Erdős–Rényi model and the permutation model (as defined in (2.25)). The proof of Theorem 2.1.13 is given in Section 2.2.3. In Section 2.3 we construct the switching and reflection couplings, which will be used to create exchangeable pairs of random regular digraphs. In Section 2.4 we use the reflection coupling to prove an upper tail bound for the codegree of a fixed pair of vertices. For technical reasons the proof of the lower tail bounds requires more care, in particular using the control on the upper tail as input – this is carried out in Section 2.5, completing the proof of part (1) of Theorem 2.1.5. The tail bounds for edge discrepancy in part (2) of Theorem 2.1.5 are proved using the simple switching coupling in Section 2.6.

2.1.5 Notation and terminology

In addition to the notation given in Section 1.2, the following conventions will be used in this chapter.

Events will be denoted by the letters $\mathcal{E}, \mathcal{B}$, and $\mathcal{G}$, where the latter two denote "bad" and "good" events, respectively. Their meaning may vary from proof to proof, but will remain fixed for the duration of each proof. $\mathbb{E}_X$ and $\mathbb{P}_X$ denote expectation and probability, respectively, conditional on all random variables but X. We say that an event $\mathcal{E}$ depending on n holds *asymptotically almost surely* if $\mathbb{P}(\mathcal{E}^c) = o(1)$.

It will be convenient to express codegrees and edge counts in terms of the adjacency matrix associated to Γ. The following terminology will be used throughout the thesis:

Definition 2.1.15 (rrd matrix)**.** For $1 \le d \le n-1$, denote by $\mathcal{M}_n(d)$ the set adjacency matrices associated to the elements of $\mathcal{D}_{n,d}$, i.e. the set of $n \times n$ matrices $A = (a_{ij})$ with $a_{ij} \in \{0,1\}$ for all $i, j \in [n]$, and satisfying the constraint that for all $k \in [n]$,

$$d = \sum_{i=1}^{n} a_{ik} = \sum_{j=1}^{n} a_{kj}. \tag{2.32}$$

We refer to a uniform random element $M \in \mathcal{M}_n(d)$ as a *random regular digraph matrix*, or *rrd matrix.*

We identify V with $[n]$ and index the rows and columns of M by i and j, respectively. By abuse of notation we refer to i, j as "vertices". Given ordered tuples of row and column indices $(i_1, \dots, i_a)$ and $(j_1, \dots, j_b)$, we denote by $M_{(i_1,\dots,i_a)\times(j_1,\dots,j_b)}$ the $a \times b$ matrix with (k, l)

entry equal to the (i_k, j_l) entry of M. (Note for instance that the sequence $(i_1, \dots, i_a)$ need not be increasing.)

For $i \in [n]$, let

$$\mathcal{N}_M(i) = \{j \in [n] : M(i,j) = 1\} \tag{2.33}$$

so that $\mathcal{N}_M(i)$ and $\mathcal{N}_{M^\intercal}(i)$ are the out- and in-neighborhoods of the vertex i, respectively. For the neighborhood of a pair of distinct vertices $i_1, i_2 \in [n]$, denote the set of common out-neighbors by

$$\begin{aligned} \mathrm{Co}_M(i_1, i_2) &= \mathcal{N}_M(i_1) \cap \mathcal{N}_M(i_2) && (2.34)\\ &= \left\{ j \in [n] : M_{(i_1,i_2)\times j} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} && (2.35)\end{aligned}$$

and denote also

$$\begin{aligned} \mathrm{Ex}_M(i_1, i_2) &= \mathcal{N}_M(i_1) \setminus \mathcal{N}_M(i_2) && (2.36)\\ &= \left\{ j \in [n] : M_{(i_1,i_2)\times j} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\} && (2.37)\end{aligned}$$

so that

$$\mathrm{Ex}_M(i_2, i_1) = \left\{ j \in [n] : M_{(i_1,i_2)\times j} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}. \tag{2.38}$$

We write $\mathrm{co}_M(i_1, i_2)$ and $\mathrm{ex}_M(i_1, i_2)$ for the cardinality of these sets, so that in our previous notation

$$\begin{aligned} \overrightarrow{\mathrm{co}}_\Gamma(i_1, i_2) &= \mathrm{co}_M(i_1, i_2)\\ \overleftarrow{\mathrm{co}}_\Gamma(j_1, j_2) &= \mathrm{co}_{M^\intercal}(j_1, j_2) = \left| \left\{ i \in [n] : M_{i\times(j_1,j_2)} = \begin{pmatrix} 1 & 1 \end{pmatrix} \right\} \right| . \end{aligned}$$

We note the following identities. From the constraints $\sum_{j=1}^n M(i_1, j) = d$ and $\sum_{j=1}^n M(i_2, j) = d$ we have

$$\mathrm{ex}_M(i_1, i_2) = d - \mathrm{co}_M(i_1, i_2) = \mathrm{ex}_M(i_2, i_1). \tag{2.39}$$

Finally we have that

$$\left|\left\{j \in [n] : M_{(i_1,i_2)\times j} = \begin{pmatrix}0\\0\end{pmatrix}\right\}\right| = n - 2d + \mathrm{co}_M(i_1, i_2). \tag{2.40}$$

We will also write $e_M(A, B)$ instead of $e_\Gamma(A, B)$.

2.2 Concentration of measure and exchangeable pairs

In this section we prove analogues of the bounds in Theorem 2.1.5 for two digraph models possessing more independence than the uniform d-regular digraph: the Erdős–Rényi model, in which all edges are independent, and the permutation model, as defined in (2.25). The proofs for the former model illustrate the application of concentration of measure tools, and are completely standard. Their use of Chernoff-type bounds (namely Bernstein's inequality), which are unavailable for random regular graphs, motivate the method of exchangeable pairs (Chatterjee's Theorem 2.2.2) as a substitute. We prove Theorem 2.1.13 for the permutation model in Section 2.2.3 as a simple illustration of the method. The reader who is primarily interested in getting a feel for applying the method to combinatorial problems may prefer to read the proof of Theorem 2.1.13 to the more technical proof of part (2) of Theorem 2.1.5 in Section 2.6.

2.2.1 The Erdős–Rényi model

Let $D = (V, E)$ be drawn from the distribution $D(n, p)$ over digraphs on n vertices, where each directed edge is included independently with probability p.

Proposition 2.2.1 (Uniform control of codegrees and edge counts, Erdős–Rényi case)**.**

(i) For any $\varepsilon > 0$, except with probability $O\Big(n^2 \exp\left(-\frac{c\varepsilon^2}{1+\varepsilon}p^2 n\right)\Big)$ we have that for all $i_1, i_2 \in [n]$ distinct,

$$\Big|\mathrm{co}_D(i_1, i_2) - p^2 n\Big| \le \varepsilon p^2 n.$$

(ii) For $\varepsilon \in (0,1)$, let

$$\mathcal{F}(\varepsilon) = \left\{ (A,B) : \ A, B \subset [n], \ \min\big(|A|, |B|\big) \geq \frac{C_0 \log n}{\varepsilon^2 p} \right\} \tag{2.41}$$

where $C_0 > 0$ is a sufficiently large absolute constant. For any $\varepsilon \in (0,1)$, with probability $1 - O\Big(\exp\Big(-c \frac{\log^2 n}{\varepsilon^2 p} \Big) \Big)$ we have that for all $(A,B) \in \mathcal{F}(\varepsilon)$,

$$\Big| e_D(A,B) - p|A||B| \Big| \leq \varepsilon p |A||B|. \tag{2.42}$$

Proof. For fixed vertices $i_1, i_2 \in V$ and subsets $A, B \subset V$, the statistics $\mathrm{co}_D(i_1,i_2)$ and $e_D(A,B)$ can be expressed as sums of iid indicator variables:

$$\mathrm{co}_D(i_1,i_2) = \sum_{j=1}^{n} \mathbb{1}((i_1,j) \in E)\, \mathbb{1}((i_2,j) \in E), \tag{2.43}$$

$$e_D(A,B) = \sum_{i \in A, j \in B} \mathbb{1}((i,j) \in E). \tag{2.44}$$

It follows that $\mathbb{E}\, \mathrm{co}_D(i_1,i_2) = p^2 n$ and $\mathbb{E}\, e_D(A,B) = p|A||B|$. Furthermore, by Bernstein's inequality we have that for any $\varepsilon \geq 0$,

$$\mathbb{P}\Big[\big| \mathrm{co}_D(i_1,i_2) - p^2 n \big| \geq \varepsilon p^2 n \Big] \leq 2 \exp\left(-\frac{c\varepsilon^2}{1+\varepsilon} p^2 n \right) \tag{2.45}$$

and

$$\mathbb{P}\Big[\big| e_D(A,B) - p|A||B| \big| \geq \varepsilon p |A||B| \Big] \leq 2 \exp\left(-\frac{c\varepsilon^2}{1+\varepsilon} p |A||B| \right) \tag{2.46}$$

for some absolute constant $c > 0$. From (2.45) and a union bound we obtain uniform control of codegrees off a small event:

$$\mathbb{P}\Big(\exists \text{ distinct } i_1, i_2 \in [n] : \Big| \mathrm{co}_D(i_1,i_2) - p^2 n \Big| \geq \varepsilon p^2 n \Big) \ll n^2 \exp\left(-\frac{c\varepsilon^2}{1+\varepsilon} p^2 n \right) \tag{2.47}$$

which establishes (i).

The proof of (ii) follows the same lines as in the proof of Corollary 2.1.9 (in particular the part establishing (2.19)), using the bounds (2.46) in place of (2.16). □

2.2.2 The method of exchangeable pairs

The main challenge for proving analogous results for d-regular digraphs is that the entries of M are all dependent on one another, and so we cannot apply off-the-shelf concentration of measure tools like Bernstein's inequality. The method of exchangeable pairs provides a convenient framework for analyzing dependent structures possessing measure preserving actions of a "local" nature. It was developed by Stein for normal-approximation [Ste72] and by Chatterjee for concentration of measure in his PhD thesis [Cha05, Cha07]. We will use the method to obtain bounds of the form (2.45) and (2.46) for the random regular digraph Γ.

Recall that a pair of $\mathcal{M}$-valued random variables (M_1, M_2) is exchangeable if

$$(M_1, M_2) \stackrel{d}{=} (M_2, M_1).$$

In particular we have $M_1 \stackrel{d}{=} M_2$. We will consider exchangeable pairs $(M, \Phi(M))$ formed by the application of a transformation $\Phi : \mathcal{M} \to \mathcal{M}$ with certain properties. Roughly speaking, the method derives properties of a statistic $f(M)$, such as concentration or approximate normality, by analyzing the change in $f(M)$ under the application of Φ.

An example of a "local" measure-preserving operation for a sequence of independent variables is to resample one of the variables independently of all others. For d-regular graphs, there are switching operations (described in Section 2.3.1).

The following is a version of Theorem 1.5 from [Cha07] suitable for our purposes:

Theorem 2.2.2 (Chatterjee [Cha07]). *Let $\mathcal{M}$ be a separable metric space, and suppose $(M, \widetilde{M})$ is an exchangeable pair of $\mathcal{M}$-valued random variables, i.e.*

$$(M, \widetilde{M}) \stackrel{d}{=} (\widetilde{M}, M).$$

Suppose $f : \mathcal{M} \to \mathbf{R}$ and $F : \mathcal{M} \times \mathcal{M} \to \mathbf{R}$ are square-integrable functions such that

$F(M,\widetilde{M}) = -F(\widetilde{M},M)$ a.s. and $\mathbb{E}(F(M,\widetilde{M})|M) = f(M)$ a.s.. Assume

$$\mathbb{E}\left[e^{\theta f(M)}\big|F(M,\widetilde{M})\big|\right] < \infty \tag{2.48}$$

for all $\theta \in \mathbb{R}$. Let

$$v_f(M) := \frac{1}{2}\,\mathbb{E}\left[\big|\big(f(M) - f(\widetilde{M})\big)F(M,\widetilde{M})\big|\;\Big|\;M\right].$$

If there are non-negative constants K_1, K_2 such that $v_f(M) \le K_1 + K_2 f(M)$ a.s., then for any $t \ge 0$,

$$\mathbb{P}(f(M) \ge t) \le \exp\left(-\frac{t^2}{2(K_1+K_2t)}\right), \quad \mathbb{P}(f(M) \le -t) \le \exp\left(-\frac{t^2}{2K_1}\right). \tag{2.49}$$

Remark 2.2.3. The qualitative integrability conditions on f and F will be satisfied automatically in our applications as we will only consider bounded (depending on n) functions on a finite set.

The quantity $v_f(M)$ is referred to by Chatterjee as a "stochastic measure of the variance of $f(M)$", and one can view a bound of the form

$$v_f(M) \le K_1 + K_2 f(M)$$

as a generalization of the "Lipschitz" conditions assumed in other commonly used concentration bounds such as McDiarmid's inequality [McD89]. We point the reader to [Cha07] for further discussion of Theorem 2.2.2 and its relation to other concentration inequalities.

2.2.3 The permutation model: Proof of Theorem 2.1.13

In this section we illustrate how one applies Theorem 2.2.2 by proving the edge discrepancy bounds of Theorem 2.1.13 for the permutation model Λ. The proof is a cartoon of the proof of the analogous bound from Theorem 2.1.5 for the uniform model, given in Section 2.6. Various technical issues that must be addressed for the case of the uniform model are absent here; in particular, the independence between the permutation matrices allows us to proceed without any a priori control on codegrees.

We recall from Section 2.1.3 that the permutation model d-regular directed multigraph Λ has adjacency matrix given by

$$M_\Lambda = P_1 + \cdots + P_d$$

where $P_1, \ldots, P_d$ are iid uniform $n \times n$ permutation matrices. We may hence view the statistics $e_\Lambda(A, B)$ as functions of a uniform random element $\pi = (\pi_1, \ldots, \pi_d)$ of $\mathrm{Sym}(n)^d$, where $\mathrm{Sym}(n)$ denotes the symmetric group over $[n]$. For $\sigma \in \mathrm{Sym}(n)$ and $A, B \subset [n]$, denote

$$e_\sigma(A, B) = \big|\big\{i \in A : \sigma(i) \in B\big\}\big| \tag{2.50}$$

and for $\pi = (\pi_1, \ldots, \pi_d) \in \mathrm{Sym}(n)^d$ we set

$$e_\pi(A, B) = \sum_{k=1}^{d} e_{\pi_k}(A, B). \tag{2.51}$$

If π is a uniform random element of $\mathrm{Sym}(n)^d$ we hence have

$$e_\Lambda(A, B) \stackrel{d}{=} e_\pi(A, B).$$

Theorem 2.1.13 is then a consequence of the following

Proposition 2.2.4. *If $\pi = (\pi_1, \ldots, \pi_d)$ is a uniform random element of $\mathrm{Sym}(n)^d$ and A, B are fixed subsets of $[n]$, we have that for any $\tau \geq 0$,*

$$\mathbb{P}\left\{e_\pi(A, B) \geq (1+\tau)\frac{d}{n}|A||B|\right\} \leq \exp\left(-\frac{\tau^2}{2+\tau}\frac{d}{n}|A||B|\right) \tag{2.52}$$

and

$$\mathbb{P}\left\{e_\pi(A, B) \leq (1-\tau)\frac{d}{n}|A||B|\right\} \leq \exp\left(-\frac{\tau^2}{2}\frac{d}{n}|A||B|\right). \tag{2.53}$$

The proof is similar to the proof of Proposition 1.1 in [Cha07], which was concerned with a more general statistic but for the case of $d = 1$. Here and in the remainder of the chapter we will make use of the following

Observation 2.2.5 (Exchangeable pair from an involution)**.** *Let $\mathcal{M}$ be a finite set, and suppose $\Phi : \mathcal{M} \to \mathcal{M}$ is an involution. Let M be a uniform random element of $\mathcal{M}$, and set $\widetilde{M} = \Phi(M)$. Then $(M, \widetilde{M})$ is an exchangeable pair of uniformly distributed elements of $\mathcal{M}$.*

Proof. Since M is uniform and Φ is a permutation we have $\Phi(M) \stackrel{d}{=} M$, and so

$$(M, \widetilde{M}) = (M, \Phi(M)) \stackrel{d}{=} (\Phi(M), \Phi^2(M)) = (\widetilde{M}, M).$$

□

Proof of Proposition 2.2.4. Define the anti-symmetric function $F : \mathrm{Sym}(n)^d \times \mathrm{Sym}(n)^d \to \mathbb{R}$ by

$$F(\pi, \pi') = K\big[e_\pi(A, B) - e_{\pi'}(A, B)\big]$$

where K is a normalizing constant. With foresight we take

$$K = \frac{d}{n} a(n-a) \tag{2.54}$$

where we denote $|A| = a$, $|B| = b$.

We construct an exchangeable pair $(\pi, \tilde{\pi})$ of uniform random elements of $\mathrm{Sym}(n)^d$ as follows. We draw the following random variables, uniformly at random from their respective ranges:

- $\pi = (\pi_1, \ldots, \pi_d) \in \mathrm{Sym}(n)^d$,
- $J \in [d]$,
- $I_1 \in A$
- $I_2 \in [n] \setminus A$

with π, J, I_1, I_2 jointly independent. We form $\tilde{\pi}$ by replacing π_J with $\tau_{\{I_1,I_2\}} \circ \pi_J$, where $\tau_{\{i_1,i_2\}}$ denotes the transposition of $i_1, i_2 \in [n]$; $(\pi_k)_{k\neq J}$ are left unchanged. $(\pi, \tilde{\pi})$ is an exchangeable pair by Observation 2.2.5. We have

$$\begin{aligned} e_\pi(A,B) - e_{\tilde{\pi}}(A,B) &= e_{\pi_J}(A,B) - e_{\tau_{\{I_1,I_2\}}\circ\pi_J}(A,B) && (2.55)\\ &= \mathbb{1}(\pi_J(I_1)\in B)\,\mathbb{1}(\pi_J(I_2)\notin B) - \mathbb{1}(\pi_J(I_1)\notin B)\,\mathbb{1}(\pi_J(I_2)\in B) && (2.56) \end{aligned}$$

and so

$$\begin{aligned} f(\pi) &:= \mathbb{E}\left[F(\pi,\tilde{\pi})\middle|\pi\right]\\ &= K\Big[\mathbb{P}\Big\{\pi_J(I_1)\in B,\ \pi_J(I_2)\notin B\Big|\pi\Big\} - \mathbb{P}\Big\{\pi_J(I_1)\notin B,\ \pi_J(I_2)\in B\Big|\pi\Big\}\Big] && (2.57)\\ &= K\,\mathbb{E}_J\left[\frac{e_{\pi_J}(A,B)}{a}\frac{e_{\pi_J}(A^c,B^c)}{n-a} - \frac{e_{\pi_J}(A,B^c)}{a}\frac{e_{\pi_J}(A^c,B)}{n-a}\right]\\ &= \frac{K}{a(n-a)}\,\mathbb{E}_J\Big[e_{\pi_J}(A,B)\big(n-a-b+e_{\pi_J}(A,B)\big) - \big(a-e_{\pi_J}(A,B)\big)\big(b-e_{\pi_J}(A,B)\big)\Big]\\ &= \frac{Kn}{a(n-a)}\,\mathbb{E}_J\,e_{\pi_J}(A,B) - \frac{dab}{n} && (2.58)\\ &= e_\pi(A,B) - \frac{dab}{n} && (2.59) \end{aligned}$$

where in the last line we applied (2.54).

It remains to bound the quantity $v_f(\pi)$ from Theorem 2.2.2. We have

$$\begin{aligned} (f(\pi)-f(\tilde{\pi}))^2 &= (e_\pi(A,B) - e_{\tilde{\pi}}(A,B))^2\\ &= \mathbb{1}(\pi_J(I_1)\in B)\,\mathbb{1}(\pi_J(I_2)\notin B) + \mathbb{1}(\pi_J(I_1)\notin B)\,\mathbb{1}(\pi_J(I_2)\in B) \end{aligned}$$

so

$$
\begin{aligned}
v_f(\pi) &:= \frac{1}{2}\,\mathbb{E}\left[|f(\pi)-f(\tilde\pi)||F(\pi,\tilde\pi)|\middle|\pi\right] \\
&= \frac{K}{2}\,\mathbb{E}\left[\big(f(\pi)-f(\tilde\pi)\big)^2\middle|\pi\right] \\
&= \frac{K}{2}\Big[\mathbb{P}\Big\{\pi_J(I_1)\in B,\ \pi_J(I_2)\notin B\Big|\pi\Big\} + \mathbb{P}\Big\{\pi_J(I_1)\notin B,\ \pi_J(I_2)\in B\Big|\pi\Big\}\Big] \\
&= \frac{1}{2}f(\pi) + K\,\mathbb{P}\Big\{\pi_J(I_1)\notin B,\ \pi_J(I_2)\in B\Big|\pi\Big\} \\
&= \frac{1}{2}f(\pi) + K\,\mathbb{E}_J\,\frac{\big(a-e_{\pi_J}(A,B)\big)\big(b-e_{\pi_J}(A,B)\big)}{a(n-a)} \\
&\le \frac{1}{2}f(\pi) + \frac{d}{n}ab
\end{aligned}
$$

where in the fourth line we applied (2.57).

By Theorem 2.2.2 we conclude that for any $t \ge 0$,

$$\mathbb{P}\left\{e_\pi(A,B) - \frac{d}{n}ab \ge t\right\} \le \exp\left(-\frac{t^2}{2\frac{d}{n}ab + t}\right) \tag{2.60}$$

$$\mathbb{P}\left\{e_\pi(A,B) - \frac{d}{n}ab \le -t\right\} \le \exp\left(-\frac{t^2}{2\frac{d}{n}ab}\right). \tag{2.61}$$

Setting $t = \tau\frac{d}{n}ab$ completes the proof. □

2.3 Exchangeable pairs constructions

In this section we define two involutions on $\mathcal{M}_n(d)$ – *simple switchings* and *reflections* – which we use to create exchangeable pairs $(M, \widetilde{M})$ of rrd matrices via Observation 2.2.5.

2.3.1 Simple switching

Below we set up our notation for switchings on a digraph in terms of the adjacency matrix M.

Definition 2.3.1 (Simple switching)**.** For $M \in \mathcal{M}_n(d)$ and $i_1, i_2, j_1, j_2 \in [n]$, we say that the 2×2 minor $M_{(i_1,i_2)\times(j_1,j_2)}$ is *switchable* if it is equal to either

$$\mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{or} \quad \mathbf{J}_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{2.62}$$

By *perform a switching at* $(i_1, i_2) \times (j_1, j_2)$ *on* M we mean to replace the minor $M_{(i_1,i_2)\times(j_1,j_2)}$ with $\mathbf{J}_2$ if it is $\mathbf{I}_2$ and $\mathbf{I}_2$ if it is $\mathbf{J}_2$, and to leave M unchanged if this minor is not switchable.

In the associated digraph Γ, the switching operation changes between the following edge configurations at vertices i_1, i_2, j_1, j_2:

where we use solid arrows to depict directed edges, and dashed arrows to indicate places where there is no edge (i.e. "non-edges").

Lemma 2.3.2 (Switching coupling)**.** *For* $i_1, i_2, j_1, j_2 \in [n]$*, let* $\Phi_{(i_1,i_2)\times(j_1,j_2)} : \mathcal{M}_n(d) \to \mathcal{M}_n(d)$ *denote the map which performs a simple switching at the minor* $(i_1, i_2) \times (j_1, j_2)$*. If* M *is an rrd matrix (i.e. a uniform random element of* $\mathcal{M}_n(d)$*) and* $I_1, I_2, J_1, J_2 \in [n]$ *are random (or deterministic) indices independent of* M*, then setting*

$$\widetilde{M} := \Phi_{(I_1,I_2)\times(J_1,J_2)}(M) \tag{2.63}$$

we have that $(M, \widetilde{M})$ *is an exchangeable pair of rrd matrices.*

Proof. We may condition on I_1, I_2, J_1, J_2. Note that the map $\Phi_{(I_1,I_2)\times(J_1,J_2)}$ is an involution on $\mathcal{M}_n(d)$. The result now follows from Observation 2.2.5. $\square$

2.3.2 Reflection

In order to prove that the random variables $\mathrm{co}_M(i_1, i_2)$ are concentrated we will need a different operation on random regular digraphs of switching-type which we call "reflection".

We pause to give some motivation and intuition for the rigorous definition below.

Suppose first that we only want to prove an upper tail bound on $\mathrm{co}_M(1,2)$. Hence, we want to show it is unlikely that for most $j \in [n]$ we have

$$M_{(1,2)\times j} = \begin{pmatrix}1\\1\end{pmatrix} \quad \text{or} \quad \begin{pmatrix}0\\0\end{pmatrix}$$

i.e., that the first two rows of M are nearly parallel. The idea is to show that for a pair of column indices $j_1, j_2 \in [n]$, the event that

$$M_{(1,2)\times(j_1,j_2)} = \begin{pmatrix}1 & 0\\1 & 0\end{pmatrix} \tag{2.64}$$

is roughly just as likely as the event that

$$M_{(1,2)\times(j_1,j_2)} = \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}. \tag{2.65}$$

We will do this by defining a "reflection" operation which switches the $(1,2)\times(j_1,j_2)$ minor between these two outcomes. If we can perform reflections independently at random at several disjoint pairs of column indices, we can then deduce from Hoeffding's inequality that with high probability there are many columns j for which

$$M_{(1,2)\times j} = \begin{pmatrix}1\\0\end{pmatrix} \quad \text{or} \quad \begin{pmatrix}0\\1\end{pmatrix}$$

as desired. While this approach can be made precise, we can do much better by instead using Theorem 2.2.2, which gives upper and lower tail estimates for $\mathrm{co}_M(1,2)$ around its mean.

While it is possible to alternate between the minors (2.64) and (2.65) using simple switchings involving entries from a third row, it turns out that when one tries to apply Theorem 2.2.2 with this coupling some control on the quantities $\mathrm{co}_{M^\intercal}(j_1,j_2)$ is needed, so that such an approach is circular.

The reflection involution is most natural to state in terms of a *walk* $w_{(j_1,j_2)} : [n] \to \mathbb{Z}$

associated to an ordered pair of columns X_{j_1}, X_{j_2} of M. For $(j_1, j_2) \in [n]^2$ we define

$$w_{(j_1,j_2)}(i) = \sum_{k=1}^{i} \mathbb{1}\left(M_{k\times(j_1,j_2)} = (1\ 0)\right) - \mathbb{1}\left(M_{k\times(j_1,j_2)} = (0\ 1)\right). \tag{2.66}$$

If we think of $w_{(j_1,j_2)}$ as giving the position of a walker on $\mathbb{Z}$, the walker starts at 0 and, reading down the pair of columns (X_{j_1}, X_{j_2}) of M, takes a step in the + direction each time it sees a row equal to (1 0), a step in the − direction each time it reads (0 1), and does not move otherwise. By d-regularity, the walker takes an even number of steps, half to the left and half to the right, ending its walk at 0. The number of steps is between 0 and $2d$; in the former case X_{j_1} and X_{j_2} are parallel, and in the latter case they are orthogonal.

Definition 2.3.3 (Reflecting pair)**.** With $w_{(j_1,j_2)}$ as in (2.66), we say that an ordered pair of column indices $(j_1, j_2) \in [n]^2$ is *reflecting* for $M \in \mathcal{M}_n(d)$ if

1. $w_{(j_1,j_2)}(1) = +1$,
2. $w_{(j_1,j_2)}(2) \neq +1$, and
3. there exists $i \in [3, n]$ such that $w_{(j_1,j_2)}(i) = +1$

that is, if the walker moves to +1 on the first step, leaves +1 on the second step, and returns again to +1 at some later time.

Conditions (1) and (2) above assert that the minor $M_{(1,2)\times(j_1,j_2)}$ is either

$$\mathbf{K}_2 := \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \quad \text{or} \quad \mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

We pause to note that condition (3) usually holds if (1) and (2) hold. Indeed, note that if $w_{(j_1,j_2)}(2) = +2$ then condition (3) follows automatically from (1) and (2) since the walk must pass through +1 on its way back to 0. Hence, any pair (j_1, j_2) of column indices such that $M_{(1,2)\times(j_1,j_2)} = \mathbf{K}_2$ is reflecting.

On the other hand, note that a non-reflecting pair (j_1, j_2) for which $M_{(1,2)\times(j_1,j_2)} = \mathbf{I}_2$ corresponds to a walk $w_{(j_1,j_2)}$ that reaches +1 on the first step, then turns back and never returns to $+1$. Non-reflecting pairs satisfying (1) and (2) but not (3) hence correspond to walks that do not cross the line $w = 0$ after the second step, so we can bound the probability of this happening by a standard enumerative argument involving Catalan numbers. This is carried out in the proof of Lemma 2.5.3. Consequently, one may think of reflecting pairs as essentially being those $(j_1, j_2) \in [n]^2$ such that $M_{(1,2)\times(j_1,j_2)} = \mathbf{K}_2$ or $\mathbf{I}_2$.

If (j_1, j_2) is reflecting for M, denote by

$$i^*(j_1, j_2) := \min\left\{i \in [3, n] : w_{(j_1,j_2)}(i) = +1\right\}$$

the first return time to $+1$.

Lemma 2.3.4 (Reflection coupling)**.** *For $(j_1, j_2) \in [n]^2$, let $\Psi_{(j_1,j_2)} : \mathcal{M}_n(d) \to \mathcal{M}_n(d)$ denote the map which replaces the minor $M_{[2,i^*]\times(j_1,j_2)}$ with the "reflected" minor $M_{[2,i^*]\times(j_2,j_1)}$ if (j_1, j_2) is reflecting, and leaves M unchanged otherwise. If M is an rrd matrix and $J_1, J_2 \in [n]$ are random column indices independent of M, then setting*

$$\widetilde{M} := \Psi_{(J_1,J_2)}(M) \tag{2.67}$$

we have that $(M, \widetilde{M})$ is an exchangeable pair of rrd matrices.

Proof. By conditioning on J_1, J_2, from Observation 2.2.5 it suffices to show that $\Psi_{(j_1,j_2)}$ is an involution on $\mathcal{M}_n(d)$ for $j_1, j_2 \in [n]$ fixed.

$\Psi_{(j_1,j_2)}$ acts trivially if $j_1 = j_2$, so we may fix $j_1, j_2 \in [n]$ distinct. We can now divide $\mathcal{M}_n(d)$ into three classes:

1. $\mathcal{M}^0_{(j_1,j_2)}$, the set of M such that (j_1, j_2) is not reflecting for M.
2. $\mathcal{M}^+_{(j_1,j_2)}$, the set of M such that (j_1, j_2) is reflecting for M and $w_{(j_1,j_2)}(2) = +2$.
3. $\mathcal{M}^-_{(j_1,j_2)}$, the set of M such that (j_1, j_2) is reflecting for M and $w_{(j_1,j_2)}(2) = 0$.

We dispense with the subscripts (j_1, j_2) for the remainder of the proof.

Ψ acts trivially on $\mathcal{M}^0$. We will show that Ψ is a bijection between $\mathcal{M}^+$ with $\mathcal{M}^-$ with $\Psi^2 = \mathrm{Id}$.

We define a pairing $\mathcal{P}$ of the elements of $\mathcal{M}^+$ with those of $\mathcal{M}^-$ (in particular, these sets have the same cardinality). For $(M^+, M^-) \in \mathcal{M}^+ \times \mathcal{M}^-$, let w^+ and w^- denote the associated walks for the columns (j_1, j_2). We say that (M^+, M^-) is in $\mathcal{P}$ if the first return time i^* of the walks w^+, w^- to $+1$ is the same, and if the walk w^+ is obtained from w^- by reflecting the portion of the trajectory of $w^-(i)$ with $i \in [2, i^*]$ across the line $w = +1$. We conclude the proof by noting that Ψ sends each $M \in \mathcal{M}^+ \cup \mathcal{M}^-$ to its mate in $\mathcal{P}$. $\square$

Remark 2.3.5. The bijection $\Psi_{(j_1,j_2)}$ above is an application of the well-known *reflection principle* from the theory of random walks – see for instance [Fel68, Chapter III].

2.4 The upper tail for codegrees

Our aim in this section is to prove the following

Proposition 2.4.1 (Upper tail for codegree)**.** *For any $\varepsilon \geq 0$ and any distinct $i_1, i_2 \in [n]$,*

$$\mathbb{P}\left\{ \mathrm{co}_M(i_1, i_2) - p^2 n \geq \varepsilon \hat{p}^2 n \right\} \leq \exp\left(-\frac{\varepsilon^2}{4 + 2\varepsilon} \hat{p}^2 n \right). \tag{2.68}$$

Remark 2.4.2 (Comparison to the Erdős–Rényi case)*.* Up to constants in the exponential, this matches the upper tail for the Erdős–Rényi digraph given in (2.45).

Remark 2.4.3. As a corollary one may obtain some control on edge discrepancy by applying the above proposition (with a union bound over pairs of vertices) with Lemma 2.1.1. This will only be effective when $d = \omega(\sqrt{n})$ and for pairs of sets A, B with $\max(|A|, |B|) \gg n$, and is hence inferior to Corollary 2.1.9.

Proof. We will apply Theorem 2.2.2 and the reflection coupling of Lemma 2.3.4.

We first note the trivial deterministic bounds

$$d \geq \mathrm{co}_M(i_1, i_2) \geq \max(0, 2d - n). \tag{2.69}$$

The lower bound is equivalent to

$$\mathrm{ex}_M(i_1, i_2) \leq \min(d, n - d) = \hat{d} \tag{2.70}$$

which can be seen from the obvious bound

$$\mathrm{ex}_M(i_1, i_2) = \big|\mathcal{N}_M(i_1) \setminus \mathcal{N}_M(i_2)\big| \leq \big|\mathcal{N}_M(i_1)\big| = d$$

and the fact that $\mathrm{ex}_M(i_1, i_2) = \mathrm{ex}_{M'}(i_1, i_2) \leq n - d$ (where M' is the adjacency matrix of the complementary digraph Γ').

Since the rows of M are exchangeable we may take $(i_1, i_2) = (1, 2)$. Let us abbreviate

$$\mathrm{co}(M) := \mathrm{co}_M(1, 2).$$

We construct a coupled pair $(M, \widetilde{M})$ of rrd matrices as follows: letting M be an rrd matrix and J_1, J_2 be iid uniform random elements of $[n]$, independent of M, we set

$$\widetilde{M} = \Psi_{(J_1, J_2)}(M). \tag{2.71}$$

Then $(M, \widetilde{M})$ is an exchangeable pair of rrd matrices by Lemma 2.3.4. We denote the sampled 2×2 minor of the first two rows by

$$\hat{M} := M_{(1,2)\times(J_1,J_2)}.$$

Define the antisymmetric function $F(M_1, M_2) = n\big(\mathrm{co}(M_1) - \mathrm{co}(M_2)\big)$ on $\mathcal{M}_n(d) \times \mathcal{M}_n(d)$.

Recall the notation

$$\mathbf{I}_2 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{J}_2 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \mathbf{K}_2 := \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}.$$

Defining

$$\mathcal{E} = \{(J_1, J_2) \text{ is reflecting}\}$$

we have

$$\left\{\hat{M} = \mathbf{K}_2\right\} \subset \mathcal{E} \subset \left\{\hat{M} = \mathbf{K}_2\right\} \vee \left\{\hat{M} = \mathbf{I}_2\right\}$$

(see the discussion under Definition 2.3.3), and

$$F(M, \widetilde{M}) = n\big(\, \mathbb{1}(\hat{M} = \mathbf{K}_2) - \mathbb{1}(\hat{M} = \mathbf{I}_2)\, \mathbb{1}_{\mathcal{E}}\,\big). \tag{2.72}$$

Hence

$$\begin{aligned} f(M) &:= \mathbb{E}\left(F(M, \widetilde{M})\middle|M\right) \\ &= n\Big[\mathbb{P}\Big(\hat{M} = \mathbf{K}_2\big|M\Big) - \mathbb{P}\Big(\Big\{\hat{M} = \mathbf{I}_2\Big\} \wedge \mathcal{E}\big|M\Big)\Big] \\ &= g(M) + \frac{1}{n} b(M) \end{aligned}$$

where we define the "main term"

$$g(M) := n\big[\, \mathbb{P}(\hat{M} = \mathbf{K}_2|M) - \mathbb{P}(\hat{M} = \mathbf{I}_2|M)\big] \tag{2.73}$$

and the "error term"

$$\begin{aligned} \frac{1}{n} b(M) &:= n\, \mathbb{P}\Big(\Big\{\hat{M} = \mathbf{I}_2\Big\} \wedge \mathcal{E}^c \,\big|\, M\Big) \\ &= \frac{1}{n} \left|\left\{(j_1, j_2) \in \mathrm{Ex}_M(1,2) \times \mathrm{Ex}_M(2,1) : (j_1, j_2) \text{ not reflecting}\right\}\right|. \end{aligned} \tag{2.74}$$

Let us call a pair $(j_1, j_2) \in [n]^2$ "bad" if $(j_1, j_2) \in \mathrm{Ex}_M(1,2) \times \mathrm{Ex}_M(2,1)$ and (j_1, j_2) is not reflecting. In other words, (j_1, j_2) is bad if it satisfies conditions (1) and (2) from

Definition 2.3.3 but not (3). We have

$$b(M) = \left|\left\{\text{bad } (j_1, j_2) \in [n]^2\right\}\right|. \tag{2.75}$$

Using the identities (2.39), (2.40), we see that the main term $g(M)$ is simply a shift of $\operatorname{co}(M)$:

$$\begin{aligned} g(M) &= n\left(\frac{\operatorname{co}(M)}{n}\frac{n-2d+\operatorname{co}(M)}{n} - \frac{(d-\operatorname{co}(M))^2}{n^2}\right) \\ &= \operatorname{co}(M) - p^2 n. \end{aligned}$$

Hence, if we can show that the number of bad pairs $b(M)$ is small, then we can deduce tail bounds for $\operatorname{co}(M)$ around the value p^2n from tail bounds for $f(M)$.

To deduce a tail bound for $f(M)$ from Theorem 2.2.2, we must bound the quantity

$$\begin{aligned} v_f(M) &:= \frac{1}{2}\,\mathbb{E}\left[|f(M) - f(\widetilde{M})||F(M,\widetilde{M})|\middle| M\right] \\ &\le \frac{n}{2}\,\mathbb{E}\left[|\operatorname{co}(M) - \operatorname{co}(\widetilde{M})|^2\middle| M\right] + \frac{1}{2}\,\mathbb{E}\left[|\operatorname{co}(M) - \operatorname{co}(\widetilde{M})||b(M) - b(\widetilde{M})|\middle| M\right] \end{aligned}$$

and so we need to control the expressions $|\operatorname{co}(M) - \operatorname{co}(\widetilde{M})|$ and $|b(M) - b(\widetilde{M})|$.

Now $\operatorname{co}(M) - \operatorname{co}(\widetilde{M}) = \mathbb{1}(\hat{M} = \mathbf{K}_2) - \mathbb{1}(\hat{M} = \mathbf{I}_2)\,\mathbb{1}_{\mathcal{E}}$, and since these events are disjoint,

$$\begin{aligned} |\operatorname{co}(M) - \operatorname{co}(\widetilde{M})| &= \mathbb{1}(\hat{M} = \mathbf{K}_2) + \mathbb{1}(\hat{M} = \mathbf{I}_2)\,\mathbb{1}_{\mathcal{E}} \\ &\le \mathbb{1}(\hat{M} = \mathbf{K}_2) + \mathbb{1}(\hat{M} = \mathbf{I}_2). \end{aligned}$$

Since the map $\Psi_{(J_1,J_2)}$ only alters the columns indexed by J_1, J_2, it follows that at most $2\operatorname{ex}_M(1,2)$ pairs $(j_1, j_2) \in \operatorname{Ex}_M(1,2) \times \operatorname{Ex}_M(2,1)$ either become or cease to be reflecting under the application of $\Psi_{(J_1,J_2)}$, whence

$$\begin{aligned} |b(M) - b(\widetilde{M})| &\le 2\operatorname{ex}_M(1,2)\,\mathbb{1}_{\mathcal{E}} && (2.76) \\ &\le 2\hat{d}\,\mathbb{1}_{\mathcal{E}} && (2.77) \end{aligned}$$

where we used (2.70) in the second line. Combining these bounds with the identities (2.39)-(2.40),

$$\begin{aligned}
v_f(M) &\le \frac{n}{2}\,\mathbb{E}\left[|\operatorname{co}(M)-\operatorname{co}(\widetilde{M})|^2\middle|M\right] + \hat{d}\,\mathbb{E}\left[|\operatorname{co}(M)-\operatorname{co}(\widetilde{M})|\middle|M\right]\\
&\le \left(\frac{n}{2}+\hat{d}\right)\left[\mathbb{P}(\hat{M}=\mathbf{K}_2|M)+\mathbb{P}(\hat{M}=\mathbf{I}_2|M)\right]\\
&\le n\left[\frac{\operatorname{co}(M)(n-2d+\operatorname{co}(M))}{n^2}+\frac{(d-\operatorname{co}(M))^2}{n^2}\right]\\
&= \operatorname{co}(M)-\frac{d^2}{n}+\frac{2}{n}(d-\operatorname{co}(M))^2\\
&\le f(M)+\frac{2}{n}\hat{d}^2
\end{aligned} \tag{2.78}$$

where in the last line we used (2.70) and

$$\begin{aligned}
\operatorname{co}(M)-\frac{d^2}{n} &= g(M)\\
&= f(M)-\frac{1}{n}b(M)\\
&\le f(M)
\end{aligned}$$

since $b(M)\ge 0$. Applying Theorem 2.2.2 with constants

$$K_1=\frac{2}{n}\hat{d}^2,\quad K_2=1,$$

we conclude that for any $t\ge 0$,

$$\begin{aligned}
\mathbb{P}\big(\operatorname{co}(M)-p^2n\ge t\big) &\le \mathbb{P}\left(\operatorname{co}(M)-p^2n+\frac{1}{n}b(M)\ge t\right)\\
&= \mathbb{P}(f(M)\ge t)\\
&\le \exp\left(-\frac{t^2/2}{\frac{2}{n}\hat{d}^2+t}\right)
\end{aligned}$$

where we again used that $b(M)\ge 0$. The terms in the denominator are balanced by scaling $t=\varepsilon\hat{p}^2n$, where we recall

$$\hat{p}:=\hat{d}/n=\frac{1}{n}\big(\min(d,n-d)\big) \tag{2.79}$$

giving the desired bound

$$\mathbb{P}\big(\mathrm{co}(M) - p^2 n \geq \varepsilon \hat{p}^2 n\big) \leq \exp\left(-\frac{\varepsilon^2}{4+2\varepsilon}\hat{p}^2 n\right).$$

$\square$

2.5 Uniform control on codegrees

In this section we complete the proof of part (1) of Theorem 2.1.5.

In the previous section, we could pass from control on the upper tail of $f(M)$ to control on the upper tail of

$$\mathrm{co}(M) = p^2 n + f(M) - \frac{1}{n} b(M)$$

using the fact that the number of bad pairs $b(M)$ (defined in (2.75)) is non-negative. In order to control the lower tail of $\mathrm{co}(M)$, we will need to improve on the trivial upper bound

$$b(M) \leq |\,\mathrm{ex}_M(1,2)|^2 \leq \hat{d}^2$$

(from monotonicity and (2.70)). In this section we show that $b(M) \leq \varepsilon \hat{d}^2$ with high probability for $\varepsilon > 0$ small. A key ingredient will be the control on the upper tail of the codegrees obtained in the previous section.

Part (1) of Theorem 2.1.5 follows from substituting $\varepsilon = \eta \frac{\max(p,1-p)}{\min(p,1-p)}$ in the following proposition.

Proposition 2.5.1 (Uniform bounds on codegrees)**.** *For any $\varepsilon \geq 0$,*

$$\mathbb{P}\left(\exists\, \{i_1, i_2\} \subset [n] : \left|\vec{\mathrm{co}}_\Gamma(i_1,i_2) - \frac{d^2}{n}\right| \geq \varepsilon \frac{\hat{d}^2}{n}\right) \leq C_1 n^2 \hat{d}^2 \exp\left(-c\varepsilon \hat{d}\right) + C_2 n^2 \exp\left(-\frac{c\varepsilon^2}{1+\varepsilon}\frac{\hat{d}^2}{n}\right) \tag{2.80}$$

where $\hat{d} = \min(d, n-d)$, and $C_1, C_2, c > 0$ are absolute constants. If $\varepsilon \geq 1$ we may take $C_1 = 0$.

Remark 2.5.2. For $\varepsilon \in (0,1)$ fixed independent of n and $\hat{d} = \omega(\log n)$ the first term on the right hand side of (2.80) is of lower order. The second term matches the bound (2.47) for Erdős–Rényi digraphs, up to the constants in the exponential.

Proof. For $i_1, i_2 \in [n]$ distinct, define

$$b_{(i_1,i_2)}(M) = \left|\left\{(j_1, j_2) \in \mathrm{Ex}_M(i_1, i_2) \times \mathrm{Ex}_M(i_2, i_1) : (j_1, j_2) \text{ not reflecting}\right\}\right| \tag{2.81}$$

so that in the notation of (2.74) we have $b(M) = b_{(1,2)}(M)$. By row-exchangeability it suffices to get control on $b_{(1,2)}(M)$ and apply a union bound over all $(i_1, i_2) \in [n]^2$.

Lemma 2.5.3. *For any $\lambda > 0$,*

$$\mathbb{P}\left\{ b_{(1,2)}(M) \geq \lambda \hat{d} \right\} \ll \hat{d}^2 \exp\left(-c\hat{d}\right) + \hat{d}e^{-c\lambda}.$$

Proof. Defining the subsets of $[n]^2$

$$\begin{aligned} Q(M) &:= \mathrm{Ex}_M(1, 2) \times \mathrm{Ex}_M(2, 1), \\ B(M) &:= \{(j_1, j_2) \text{ not reflecting}\} \end{aligned}$$

we have

$$b(M) = |Q(M) \cap B(M)|.$$

Denote

$$x = |\operatorname{Ex}_M(1, 2)| = |\operatorname{Ex}_M(2, 1)|.$$

Now we decompose $b(M)$ as a sum of x terms, each of which can be expressed as a sum of independent indicators. We enumerate the elements of $\mathrm{Ex}_M(1, 2), \mathrm{Ex}_M(2, 1)$ in increasing order as $j_1^+ < \cdots < j_x^+$ and $j_1^- < \cdots < j_x^-$, respectively. For each $s \in [0, x-1]$, define

$$Q_s(M) = \left\{(j_m^+, j_{m+s}^-) : m \in [x]\right\}$$

with the sum $m+s$ understood to be mod x, and put

$$b_s(M) = |Q_s(M) \cap B(M)|$$

so that

$$b(M) = \sum_{s=0}^{x-1} b_s(M). \tag{2.82}$$

Fix $0 \le s \le x-1$. We now construct an exchangeable pair $(M, \widetilde{M})$ by resampling a certain subset of the entries of M. For each element $(j^+, j^-) \in Q_s(M)$ write

$$\begin{aligned} \mathcal{I}(j^+, j^-) &= [3, n] \cap \left(\mathrm{Ex}_{M^\intercal}(j^+, j^-) \cup \mathrm{Ex}_{M^\intercal}(j^-, j^+) \right) \\ &= \left\{ i \in [3, n] : M(i, j^+) + M(i, j^-) = 1 \right\}. \end{aligned}$$

We form the pair $(M, \widetilde{M})$ by first drawing $M \in \mathcal{M}_n(d)$ uniformly, then forming $\widetilde{M}$ by independently and uniformly resampling the k sub-matrices

$$\left\{ \widetilde{M}_{\mathcal{I}(j^+, j^-) \times (j^+, j^-)} \right\}_{(j^+, j^-) \in Q_s(M)} \tag{2.83}$$

conditional on all other entries of M. We can do this resampling independently since our conditioning has already fixed all of the row and column sums of each of these sub-matrices. For exchangeability it is important to note that $Q_s(M) = Q_s(\widetilde{M})$, as this set is determined by the first two rows of M, which are not resampled.

We will restrict to an event on which we have an upper bound on codegrees. Let

$$\mathcal{G}_s = \bigwedge_{(j^+, j^-) \in Q_s(M)} \left\{ \mathrm{co}_{M^\intercal}(j^+, j^-) \le \left(\frac{1+p}{2} \right) d \right\} \tag{2.84}$$

enforcing a slight improvement on the deterministic upper bound $\mathrm{co}_{M^\intercal}(j^+, j^-) \le d$. By a union bound and Proposition 2.4.1 (taking ε to be a small multiple of $n/\hat{d}$) we have

$$\mathbb{P}(\mathcal{G}_s) \ge 1 - xe^{-c\hat{d}} \ge 1 - \hat{d}e^{-c\hat{d}}. \tag{2.85}$$

Note that $\mathcal{G}_s$ holds for M if and only if it holds for $\widetilde{M}$, since the resampling does not change the value of $\mathrm{co}_{M^\mathsf{T}}(j^+, j^-)$ for any $(j^+, j^-) \in Q_s(M)$.

Conditional on M, from the joint independence of the sub-matrices (2.83) we see that $b_s(\widetilde{M})$ is a sum of independent indicators. Hence, we can control the upper tail of $b_s(\widetilde{M})$ using Bernstein's inequality, once we have estimates on $\mathbb{E}\left[b_s(\widetilde{M})\middle|M\right]$. We will then deduce the desired bound on $b(M)$ through the decomposition (2.82) and a union bound.

To estimate $\mathbb{E}\left[b_s(\widetilde{M})\middle|M\right]$ we have the following

Claim 2.5.4. For each $s \in [0, x-1]$ and $(j^+, j^-) \in Q_s(M)$,

$$\mathbb{P}\left[(j^+, j^-) \in B(\widetilde{M})\middle|M\right] \mathbb{1}_{\mathcal{G}_s} \ll 1/\hat{d}. \tag{2.86}$$

Let us assume this claim for now. Restricting to $\mathcal{G}_s$, from (2.86) we have

$$\mathbb{E}\left[b_s(\widetilde{M})\middle|M\right] \mathbb{1}_{\mathcal{G}_s} \ll x/\hat{d} \le 1$$

for each $s \in [0, x-1]$, where we used (2.70). Moreover, since $b_s(\widetilde{M})|M$ is a sum of independent indicator variables, from Bernstein's inequality we have that for any $\lambda > 0$,

$$\mathbb{P}\Big(b_s(\widetilde{M}) \ge \lambda \,\Big|\, M\Big) \mathbb{1}_{\mathcal{G}_s} \ll \exp(-c\lambda)$$

and so

$$\begin{aligned}\mathbb{P}(b_s(M) \ge \lambda) &= \mathbb{P}(b_s(\widetilde{M}) \ge \lambda) \\ &\le \mathbb{P}(\mathcal{G}_s^c) + \mathbb{E}\,\mathbb{P}\big(b_s(\widetilde{M}) \ge \lambda\big|M\big) \mathbb{1}_{\mathcal{G}_s} \\ &\ll \hat{d} \exp\left(-c\hat{d}\right) + e^{-c\lambda}.\end{aligned}$$

By pigeonholing and a union bound it follows that

$$\mathbb{P}\Big(b(M) \geq \lambda \hat{d}\Big) \leq \hat{d}\, \mathbb{P}(b_0(M) \geq \lambda)$$
$$\ll \hat{d}^2 \exp\Big(-c\hat{d}\Big) + \hat{d}e^{-c\lambda}.$$

It remains to establish Claim 2.5.4. Fix s and (j^+, j^-) as in the claim. Consider the walk $w = w_{(j^+,j^-)} : [n] \to \mathbb{Z}$ associated to the pair of columns (X_{j^+}, X_{j^-}) as defined in (2.66). Since $(j^+, j^-) \in \mathrm{Ex}_M(1,2) \times \mathrm{Ex}_M(2,1)$ by assumption, we have $M_{(1,2)\times(j^+,j^-)} = \mathbf{I}_2$, and so $w(1) = 1$ and $w(2) = 0$. The event that $(j^+, j^-) \in B(M)$ is the event that there is no $i \in [3,n]$ such that $w(i) = 1$, i.e. that w is "non-crossing" in this range. Let us condition on the number r of steps taken to the right by w; by our restriction to $\mathcal{G}_s$ we have

$$r = d - \mathrm{co}_{M^\intercal}(j^+, j^-) \geq \frac{1}{2}(1-p)d.$$

Conditional on r, in the randomness of the resampling of (X_{j^+}, X_{j^-}) we have that every ordering of the $r-1$ left steps of w and $r-1$ right steps in the range $[3,n]$ is equally likely. There are $\binom{2(r-1)}{r-1}$ such orderings, while the number of these giving non-crossing walks is the Catalan number

$$\frac{1}{r}\binom{2(r-1)}{r-1}.$$

It follows that under the resampling, the probability that $(j^+, j^-) \in B(M)$ is

$$\frac{1}{r} \ll \frac{1}{(1-p)d} \ll \frac{1}{\hat{d}}.$$

Undoing the conditioning on r, the claim follows. □

Now we can get a good lower tail estimate on $\mathrm{co}(M)$ and complete the proof of Proposition 2.5.1.

Fix $\varepsilon \geq 0$. If $\varepsilon \geq 1$ then the result already follows from Proposition 2.4.1 and a union bound as the lower tail event is empty in this case. Hence we may assume $\varepsilon < 1$. We may further assume that ε is sufficiently small by adjusting the constant c in the statement of the theorem.

For $\lambda \geq 0$ and $i_1, i_2 \in [n]$ distinct, let

$$\mathcal{B}_{(i_1,i_2)}(\lambda) = \Big\{ b_{(i_1,i_2)}(M) \geq \lambda \hat{d} \Big\}$$

and

$$\mathcal{G}(\lambda) := \bigwedge_{i_1 \neq i_2 \in [n]} \mathcal{B}_{(i_1,i_2)}(\lambda)^c$$

From Lemma 2.5.3 and a union bound, we have

$$\mathbb{P}(\mathcal{G}(\lambda)^c) \ll n^2 \Big(\hat{d}^2 \exp\left(-c\hat{d} \right) + \hat{d} e^{-c\lambda} \Big). \tag{2.87}$$

Restricting to the good event, we can bound

$$\begin{aligned}
\mathbb{P}\Big(\mathcal{G}(\lambda) \wedge \big\{ \mathrm{co}(M) \leq (1-\varepsilon)\hat{p}^2 n \big\} \Big) &= \mathbb{P}\left(\mathcal{G}(\lambda) \wedge \left\{ f(M) \leq -\varepsilon \hat{p}^2 n + \frac{1}{n} b(M) \right\} \right) \\
&\leq \mathbb{P}\Big(f(M) \leq -\hat{p}(\varepsilon \hat{d} - \lambda) \Big).
\end{aligned}$$

Taking $\lambda = \varepsilon \hat{d}/2$ and applying Theorem 2.2.2 (with the bound (2.78) on $v_f(M)$) the last quantity is bounded by $\exp(-c\varepsilon^2 \hat{p}^2 n)$. Putting it all together, denoting

$$\mathcal{B} = \bigvee_{\{i_1,i_2\} \subset [n]} \left\{ \left| \mathrm{co}_M(i_1,i_2) - p^2 n \right| \geq \varepsilon \hat{p}^2 n \right\}$$

we have

$$\begin{aligned}
\mathbb{P}(\mathcal{B}) &\leq \mathbb{P}(\mathcal{G}(\lambda)^c) + \sum_{\{i_1,i_2\} \subset [n]} \mathbb{P}\Big(\mathcal{G}(\lambda) \wedge \Big\{ \left| \mathrm{co}_M(i_1,i_2) - p^2 n \right| \geq \varepsilon \hat{p}^2 n \Big\} \Big) \\
&\ll n^2 \hat{d}^2 \exp\left(-c\varepsilon \hat{d} \right) + n^2 \exp\left(-\frac{c\varepsilon^2}{1+\varepsilon} \frac{\hat{d}^2}{n} \right).
\end{aligned}$$

$\square$

2.6 Concentration of edge counts

In this section we prove part (2) of Theorem 2.1.5, using Theorem 2.2.2 with the switching coupling of Lemma 2.3.2. A crucial ingredient will be the control on codegrees enforced by restriction to the event $\mathcal{G}^{\mathrm{co}}(\eta)$. The reader may wish to read the simpler proof of Theorem 2.1.13 in Section 2.2.3 first, as it uses a similar switching on permutation matrices, but does not require restriction to the event $\mathcal{G}^{\mathrm{co}}(\eta)$.

Fix $A, B \subset [n]$, and let us denote $|A| = a$, $|B| = b$. Without loss of generality we may assume

$$a + b \leq n. \tag{2.88}$$

Indeed, as noted in (2.7), for any $t \in \mathbb{R}$,

$$\{e_M(A,B) - \mu(A,B) \geq t\} = \{e_M(A^c,B^c) - \mu(A^c,B^c) \geq t\}\,.$$

Hence, if we establish the claim assuming $a + b \leq n$, then for the case that $a + b > n$ we can apply the claim to (A^c, B^c) rather than (A, B). Under assumption (2.88) we have

$$\hat{\mu}(A,B) = p\min\left[ab, (n-a)(n-b)\right] = pab = \mu(A,B). \tag{2.89}$$

We define an exchangeable pair of rrd matrices $(M, \widetilde{M})$ as follows. Draw M, and sample

$$I_1 \in A,\ I_2 \in [n]\setminus A,\ J_1 \in B,\ J_2 \in [n]\setminus B$$

uniformly from their respective ranges, independently of each other and of M. Conditional on M, I_1, I_2, J_1, J_2, form $\widetilde{M}$ by performing a switching at the minor $(I_1, I_2)\times(J_1, J_2)$. $(M, \widetilde{M})$ is an exchangeable pair by Lemma 2.3.2.

Let us denote

$$K_{ab} := a(n-a)b(n-b).$$

Define the antisymmetric function $F : \mathcal{M}_n(d) \times \mathcal{M}_n(d) \to \mathbb{R}$ by

$$F(M_1, M_2) = K_{ab}\Big[e_{M_1}(A, B) - e_{M_2}(A, B)\Big].$$

Denote the sampled 2×2 minor $M_{(I_1,I_2)\times(J_1,J_2)}$ by $\hat{M}$. We have

$$\mathcal{E} := \Big\{\hat{M} \text{ is switchable}\Big\} = \Big\{\hat{M} = \mathbf{I}_2\Big\} \vee \Big\{\hat{M} = \mathbf{J}_2\Big\}$$

and

$$F(M, \widetilde{M}) = K_{ab}\big[\,\mathbb{1}(\hat{M} = \mathbf{I}_2) - \mathbb{1}(\hat{M} = \mathbf{J}_2)\big].$$

We have

$$\begin{aligned} f(M) &:= \mathbb{E}\left[F(M, \widetilde{M})\big|M\right] \\ &= K_{ab}\Big[\mathbb{P}\Big(\hat{M} = \mathbf{I}_2 \,\Big|\, M\Big) - \mathbb{P}\Big(\hat{M} = \mathbf{J}_2 \,\Big|\, M\Big)\Big] \qquad (2.90) \\ &= \sum_{(i_1,i_2)\in A\times A^c} \Big(|\operatorname{Ex}_M(i_1, i_2) \cap B||\operatorname{Ex}_M(i_2, i_1) \cap B^c| \\ &\qquad\qquad - |\operatorname{Ex}_M(i_2, i_1) \cap B||\operatorname{Ex}_M(i_1, i_2) \cap B^c|\Big) \qquad (2.91) \end{aligned}$$

(with $A^c = [n] \setminus A$, $B^c = [n] \setminus B$).

Before proceeding to control the expression $v_f(M)$ from Theorem 2.2.2, let us show how $f(M)$ is related to $e_M(A, B)$. Recalling the notation

$$\begin{aligned} \operatorname{ex}_M(i_1, i_2) &= d - \operatorname{co}_M(i_1, i_2) \\ &= |\operatorname{Ex}_M(i_1, i_2)| = |\operatorname{Ex}_M(i_2, i_1)| \end{aligned}$$

we re-express the summand in (2.91) as

$$\begin{aligned}
&|\operatorname{Ex}_M(i_1,i_2)\cap B|\Big(\operatorname{ex}_M(i_1,i_2)-|\operatorname{Ex}_M(i_2,i_1)\cap B|\Big)\\
&\qquad\qquad -|\operatorname{Ex}_M(i_2,i_1)\cap B|\Big(\operatorname{ex}_M(i_1,i_2)-|\operatorname{Ex}_M(i_1,i_2)\cap B|\Big)\\
&\quad=\operatorname{ex}_M(i_1,i_2)\big(|\operatorname{Ex}_M(i_1,i_2)\cap B|-|\operatorname{Ex}_M(i_2,i_1)\cap B|\big)\\
&\quad=\operatorname{ex}_M(i_1,i_2)\big(|\mathcal{N}_M(i_1)\cap B|-|\mathcal{N}_M(i_2)\cap B|\big).
\end{aligned}$$

Putting this in (2.91) we have

$$f(M)=\sum_{(i_1,i_2)\in A\times A^c}\operatorname{ex}_M(i_1,i_2)\Big(|\mathcal{N}_M(i_1)\cap B|-|\mathcal{N}_M(i_2)\cap B|\Big). \tag{2.92}$$

On $\mathcal{G}^{\mathrm{co}}(\eta)$ the quantities $\operatorname{ex}_M(i_1,i_2)$ all lie in $p(1-p)n[1-\eta,1+\eta]$. Writing

$$\operatorname{ex}_M(i_1,i_2)=p(1-p)n+\big(\operatorname{ex}_M(i_1,i_2)-p(1-p)n\big)$$

we can express

$$f(M)=f_1(M)+f_2(M) \tag{2.93}$$

where we define the “main term”

$$\begin{aligned}
f_1(M)&:=p(1-p)n\sum_{(i_1,i_2)\in A\times A^c}|\mathcal{N}_M(i_1)\cap B|-|\mathcal{N}_M(i_2)\cap B|\\
&=p(1-p)n\big[(n-a)e_M(A,B)-ae_M(A^c,B)\big]\\
&=p(1-p)n^2\big[e_M(A,B)-pab\big].
\end{aligned}$$

and the “error term”

$$f_2(M):=\sum_{(i_1,i_2)\in A\times A^c}\big(\operatorname{ex}_M(i_1,i_2)-p(1-p)n\big)\Big(|\mathcal{N}_M(i_1)\cap B|-|\mathcal{N}_M(i_2)\cap B|\Big).$$

We now show that $f_2(M)$ is small on $\mathcal{G}^{\mathrm{co}}(\eta)$ if η is sufficiently small, so that on this event

$f(M) \approx f_1(M)$, a scaling and centering of $e_M(A,B)$. Indeed, letting $\eta > 0$ to be chosen later,

$$\begin{aligned}
|f_2(M)|\,\mathbb{1}_{\mathcal{G}^{\mathrm{co}}(\eta)} &\le \eta p(1-p)n \sum_{(i_1,i_2)\in A\times A^c} |\mathcal{N}_M(i_1)\cap B| + |\mathcal{N}_M(i_2)\cap B| \\
&= \eta p(1-p)n\big[\,(n-a)e_M(A,B) + a e_M(A^c,B)\,\big] \\
&\le \eta p(1-p)n^2\,[\,e_M(A,B) - \mu(A,B) + 2\mu(A,B)\,] \\
&= \eta\big[\,f_1(M) + 2p(1-p)n^2\mu\,\big] \qquad (2.94)
\end{aligned}$$

where in the third line we added and subtracted $\mu(A,B) = dab/n$ and used $e_M(A^c,B) \le db$.

Now we will bound the quantity

$$v_f(M) := \frac{1}{2}\,\mathbb{E}\left[\big|\big(f(M) - f(\widetilde{M})\big)F(M,\widetilde{M})\big|\,\Big|\,M\right]$$

from Theorem 2.2.2. First we bound $|f(M) - f(\widetilde{M})|$ by considering the expression (2.91). Since M and $\widetilde{M}$ only differ on the $(I_1,I_2)\times(J_1,J_2)$ minor, the only summands in (2.91) that do not cancel in $f(M) - f(\widetilde{M})$ have indices in the set

$$\mathcal{I} := \Big\{(i_1,i_2)\in A\times A^c:\ \text{either } i_1 = I_1 \text{ or } i_2 = I_2 \text{ (or both)}\Big\}.$$

Now note that for any $(i_1,i_2)\in\mathcal{I}$,

$$\big|\,|\operatorname{Ex}_M(i_1,i_2)\cap B| - |\operatorname{Ex}_{\widetilde{M}}(i_1,i_2)\cap B|\,\big| \le \mathbb{1}_{\mathcal{E}}$$

since this set only changes (possibly) if $\hat{M}$ is switchable. We have the same bound for the pair neighborhoods $\operatorname{Ex}_M(i_2,i_1)$ and with B replaced by B^c. Using these bounds with (2.91) we have

$$
\begin{aligned}
|f(M) - f(\widetilde{M})| &\le \sum_{(i_1,i_2)\in\mathcal{I}} \mathbb{1}_{\mathcal{E}} \Big[|\operatorname{Ex}_M(i_1,i_2)\cap B| + |\operatorname{Ex}_{\widetilde{M}}(i_2,i_1)\cap B^c| \\
&\qquad\qquad + |\operatorname{Ex}_{\widetilde{M}}(i_2,i_1)\cap B| + |\operatorname{Ex}_M(i_1,i_2)\cap B^c| \Big] \\
&= \sum_{(i_1,i_2)\in\mathcal{I}} \mathbb{1}_{\mathcal{E}} \Big[|\operatorname{Ex}_M(i_1,i_2)| + |\operatorname{Ex}_{\widetilde{M}}(i_2,i_1)| \Big] \\
&\le 2|\mathcal{I}|\hat{d}\,\mathbb{1}_{\mathcal{E}} \\
&= 2n\hat{d}\,\mathbb{1}_{\mathcal{E}}
\end{aligned}
$$

where in the third line we applied the upper bound (2.70) for M and $\widetilde{M}$.

Now since

$$\frac{1}{K_{ab}}|F(M,\widetilde{M})| = |\,\mathbb{1}(\hat{M} = \mathbf{I}_2) - \mathbb{1}(\hat{M} = \mathbf{J}_2)| \le \mathbb{1}_{\mathcal{E}}$$

we have

$$v_f(M) \le n\hat{d}K_{ab}\,\mathbb{P}(\mathcal{E}|M).$$

We want to show that $f(M)$ is "self bounding" in the sense that we can control $v_f(M)$ by an expression of the form $K_1 + K_2 f(M)$ for some constants $K_1, K_2 > 0$ (possibly depending on n, d, a, b). Since

$$\mathbb{P}(\mathcal{E}|M) = \mathbb{P}\Big(\hat{M} = \mathbf{I}_2 \,\Big|\, M\Big) + \mathbb{P}\Big(\hat{M} = \mathbf{J}_2 \,\Big|\, M\Big)$$

we have

$$
\begin{aligned}
v_f(M) &\le n\hat{d}K_{ab}\Big[\mathbb{P}\Big(\hat{M} = \mathbf{I}_2 \,\Big|\, M\Big) + \mathbb{P}\Big(\hat{M} = \mathbf{J}_2 \,\Big|\, M\Big)\Big] \\
&= n\hat{d}\Big[f(M) + 2K_{ab}\,\mathbb{P}\Big(\hat{M} = \mathbf{J}_2 \,\Big|\, M\Big)\Big] \qquad (2.95)
\end{aligned}
$$

where we have used (2.90) in the second line. Writing

$$\mathbb{P}\Big(\hat{M} = \mathbf{J}_2 \Big| M\Big) = \frac{1}{K_{ab}} \sum_{(i_1,i_2)\in A\times A^c} |\operatorname{Ex}_M(i_2,i_1)\cap B||\operatorname{Ex}_M(i_1,i_2)\cap B^c|$$

we can crudely bound $|\operatorname{Ex}_M(i_1,i_2)\cap B^c| \leq \hat{d}$ and $|\operatorname{Ex}_M(i_2,i_1)\cap B| \leq |\mathcal{N}_M(i_2)\cap B|$ (from monotonicity) to get

$$\begin{aligned}
K_{ab}\mathbb{P}\Big(\hat{M} = \mathbf{J}_2 \Big| M\Big) &\leq \hat{d}a \sum_{i_2\in A^c} |\mathcal{N}_M(i_2)\cap B| \\
&= \hat{d}ae_M(A^c, B) \\
&\leq \hat{d}a(db) \\
&= \hat{d}n\mu(A,B).
\end{aligned}$$

Combining the last line with (2.95) we conclude

$$v_f(M) \leq n\hat{d}\big[f(M) + 2n\hat{d}\mu\big]. \tag{2.96}$$

From (2.94), on $\mathcal{G}^{\mathrm{co}}(\eta)$ we have

$$\begin{aligned}
f(M) &= f_1(M) + f_2(M) \\
&\geq f_1(M) - |f_2(M)| \\
&\geq (1-\eta)f_1(M) - 2\eta p(1-p)n^2\mu.
\end{aligned}$$

It follows that for $\eta, t \geq 0$ fixed,

$$\begin{aligned}
\mathbb{P}\Big(\mathcal{G}^{\mathrm{co}}(\eta) \wedge \Big\{e_M(A,B) - \mu \geq t\Big\}\Big) &= \mathbb{P}\Big(\mathcal{G}^{\mathrm{co}}(\eta) \wedge \Big\{f_1(M) \geq p(1-p)n^2 t\Big\}\Big) \\
&\leq \mathbb{P}\left(f(M) \geq \big[(1-\eta)t - 2\eta\mu\big]p(1-p)n^2\right) \qquad (2.97)
\end{aligned}$$

and

$$\mathbb{P}\Big(\mathcal{G}^{\mathrm{co}}(\eta) \wedge \Big\{e_M(A,B) - \mu \le -t\Big\}\Big) = \mathbb{P}\Big(\mathcal{G}^{\mathrm{co}}(\eta) \wedge \Big\{f_1(M) \le -p(1-p)n^2 t\Big\}\Big)$$
$$\le \mathbb{P}\left(f(M) \le -\big[t - 2\eta\mu\big]p(1-p)n^2\right). \tag{2.98}$$

Let us scale $t = \tau\mu$. If we take $\eta \le \min(1/4, \tau/8)$, then from (2.97)

$$\mathbb{P}\Big(\mathcal{G}^{\mathrm{co}}(\eta) \wedge \Big\{e_M(A,B) - \mu \ge \tau\mu\Big\}\Big) \le \mathbb{P}\Big(f(M) \ge \frac{\tau}{2}p(1-p)n^2\mu\Big) \tag{2.99}$$

Now we may apply Theorem 2.2.2 to the right hand side of (2.99) with

$$K_1 = 2n^2\hat{d}^2\mu, \quad K_2 = n\hat{d} \tag{2.100}$$

from (2.96) to bound

$$\begin{aligned}\mathbb{P}\Big(\mathcal{G}^{\mathrm{co}}(\eta) \wedge \Big\{e_M(A,B) - \mu \ge \tau\mu\Big\}\Big) &\le \exp\left(-\frac{\big(\frac{1}{2}\tau p(1-p)n^2\mu\big)^2}{2n\hat{d}\big(2n\hat{d}\mu + \frac{1}{2}\tau p(1-p)n^2\mu\big)}\right) \\ &= \exp\left(-\frac{\tau^2\mu}{8\frac{\hat{d}}{p(1-p)n}\big(2\frac{\hat{d}}{p(1-p)n} + \frac{1}{2}\tau\big)}\right) \\ &\le \exp\left(-\frac{\tau^2\mu}{64 + 8\tau}\right)\end{aligned}$$

where in the last line we used that

$$\frac{\hat{d}}{p(1-p)n} = \frac{\min(d, n-d)}{\frac{1}{n}d(n-d)} \le 2.$$

The lower tail is obtained similarly from (2.98) and Theorem 2.2.2 (and only requiring that we take $\eta \le \tau/4$).

CHAPTER 3

The singularity probability for random regular digraphs

3.1 Introduction

We begin by recalling some notation and terminology from Chapter 2. We use $\mathcal{M}_n(d)$ to denote the set of $n \times n$ matrices with entries in $\{0, 1\}$ satisfying the constraint that all row and column sums are equal to d (see Definition 2.1.15). (Thus, for instance, $\mathcal{M}_n(1)$ is the set of $n \times n$ permutation matrices.) One may interpret the elements of $\mathcal{M}_n(d)$ as the adjacency matrices of d-regular digraphs – that is, directed graphs on n labeled vertices with each vertex having d in-neighbors and d out-neighbors (allowing self-loops). One can also identify $\mathcal{M}_n(d)$ with the set of d-regular bipartite graphs on $n + n$ vertices in the obvious way.

We denote by M a uniform random element of $\mathcal{M}_n(d)$, and refer to M as an "rrd matrix" (for "random regular digraph"). Our objective in this chapter is to determine whether M is invertible with high probability when n is large and for some range of the parameter d. Before stating our main result, we give an overview of related work on other random matrix models.

3.1.1 Background

Much work on the singularity of random matrices has focused on *iid sign matrices* Ξ, whose entries are iid uniform ± 1 Bernoulli random variables. It is already a non-trivial problem to prove that Ξ is invertible with probability tending to 1; this was first accomplished by Komlós in the 1960s [Kom67, Kom68] (see also [Kom]). His proof was later refined to give the following quantitative bound:

Theorem 3.1.1 (Komlós)**.** *Let Ξ be an $n \times n$ matrix of iid uniform signs. Then*

$$\mathbb{P}\left(\det(\Xi) = 0\right) = O\big(n^{-1/2}\big). \tag{3.1}$$

(See Section 1.2 for our conventions on asymptotic notation.)

A key ingredient in the proof of Theorem 3.1.1 was a bound of Littlewood-Offord type due to Erdős (Theorem 3.2.1 below) from the seemingly unrelated field of *additive combinatorics*. This inspired a sequence of works improving (3.1) to exponential bounds

$$\mathbb{P}\left(\det(\Xi) = 0\right) \ll c^n \tag{3.2}$$

for some constant $c < 1$ by making heavier use of additive combinatorics machinery. Specifically, the base $c = .999$ was obtained by Kahn, Komlós and Szemerédi in [KKS95], and was lowered to $c = 3/4 + o(1)$ by Tao and Vu [TV07], and to $c = 1/\sqrt{2} + o(1)$ by Bourgain, Vu and Wood [BVW10]. The latter two works relied on the *inverse Littlewood-Offord theory* developed in [TV07]. These bounds still fall short of the conjecture

$$\mathbb{P}\left(\det(\Xi) = 0\right) = n^2(1 + o(1))2^{1-n} \tag{3.3}$$

which was stated in [KKS95], though the authors noted that it is best regarded as folklore. The lower bound in (3.3) is easily proved by considering the event that Ξ has a pair of rows or columns that are parallel.

One source of motivation for controlling the singularity probability is its relation to the problem of proving limit laws for the distribution of eigenvalues. Define the (rescaled) *empirical spectral distribution* of Ξ to be the random probability measure

$$\mu_{\frac{1}{\sqrt{n}}\Xi} := \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(\frac{1}{\sqrt{n}}\Xi)},$$

distributed uniformly over the eigenvalues of the normalized matrix $\frac{1}{\sqrt{n}}\Xi$. In [TV10b] Tao and Vu proved the *circular law* for Ξ, which states that almost surely, as $n \to \infty$, $\mu_{\frac{1}{\sqrt{n}}\Xi}$ converges weakly to the uniform measure on the unit disc in $\mathbb{C}$. They actually proved a *universality principle*, which implies that the circular law holds for any matrix with iid entries having mean 0 and variance 1. See Chapter 5 for further background on the circular law.

The main technical hurdle in the proof of the circular law was to obtain good lower bounds on the smallest singular value $s_n(\Xi)$ holding with high probability. (Actually, it was necessary to do this for arbitrary scalar shifts $\frac{1}{\sqrt{n}}\Xi - zI$.) Proving lower bounds on $s_n(\Xi)$ is an extension of the singularity probability problem – indeed, the latter is to bound $\mathbb{P}(s_n(\Xi) = 0)$. Polynomial lower bounds on the smallest singular value of a general class of iid matrices were first obtained by Rudelson in [Rud06], and were subsequently improved by Tao and Vu [TV09, TV10c] and Rudelson and Vershynin [RV08]. See Chapter 4 for more background on lower bounds for the smallest singular value.

Apart from iid matrices, a lot of activity has concentrated on random matrix models with constraints on row and column sums. In [BCC12], Bordenave, Caputo and Chafaï proved the circular law for random Markov matrices, obtained by normalizing the rows of an iid matrix with continuous entry distributions. On the discrete side, in [Ngu13] Nguyen proved that a uniform random 0/1 matrix constrained to have all row-sums equal to $n/2$ (say n is even) is invertible with probability $1 - O_C(n^{-C})$. Nguyen and Vu subsequently proved the circular law for a more general class of random discrete matrices with constant row sums [NV13].

The approach in [Ngu13] and [NV13] was to use a conditioning trick, which we now briefly sketch. As in [Ngu13], assume n is even, and let Q be a uniform random $n \times n$ 0/1 matrix with all row sums equal to $n/2$. Suppose we want to control the probability that some property P holds for the first row R_1 of Q. We draw $Y_1 \in \{0,1\}^n$ uniformly at random, and let $\mathcal{E}$ be the event that the components of Y_1 sum to $n/2$. We have $Y_1 \,\big|\, \mathcal{E} \overset{d}{=} R_1$. Moreover, one can easily show that $\mathbb{P}(\mathcal{E}) \gg n^{-1/2}$. It follows that we can bound

$$\begin{aligned}\mathbb{P}(\, P \text{ holds for } R_1\,) &= \mathbb{P}\big(P \text{ holds for } Y_1 \,\big|\, \mathcal{E}\big) \\ &\le \frac{\mathbb{P}(P \text{ holds for } Y_1)}{\mathbb{P}(\mathcal{E})} \\ &\ll n^{1/2}\mathbb{P}(P \text{ holds for } Y_1). \end{aligned} \tag{3.4}$$

This last term can be controlled using the existing theory for iid matrices (the loss of a factor $O(n^{1/2})$ turns out to be acceptable).

The results from [BCC12], [Ngu13] and [NV13] still relied on the independence between rows. For the rrd matrix M considered in the present work there is no independence among rows or columns. In particular, an approach by conditioning iid variables as in [Ngu13] can not treat each row separately, and instead must condition on the event that the entire iid matrix is in $\mathcal{M}_n(d)$. The probability of this event can be estimated using asymptotic enumeration results. Letting $p := d/n$, we draw a random 0/1 matrix M_p with iid Bernoulli(p) entries, and let

$$\mathcal{E}_{n,d} = \{M_p \in \mathcal{M}_n(d)\}\,. \tag{3.5}$$

Then $M_p|\mathcal{E}_{n,d} \overset{d}{=} M$. We have

$$\mathbb{P}(\mathcal{E}_{n,d}) \sim \sqrt{2\pi d(n-d)}\exp\left(-n\log\left(\frac{2\pi d(n-d)}{n}\right)\right) \tag{3.6}$$

which follows from an asymptotic formula for the cardinality of $\mathcal{M}_n(d)$, established for the sparse case $d = np = o(\sqrt{n})$ by McKay and Wang in [MW03] and for the dense range $\min(d, n-d) \gg n/\log n$ by Canfield and McKay in [CM05].

Although enumeration results for the range $\sqrt{n} \ll d \ll n/\log n$ are unavailable as of this writing (though it is natural to conjecture that the formula (3.6) extends to hold in this range), in [Tra] Tran used an argument from [SU84] of Shamir and Upfal to show that for $d = \Omega(\log n)$,

$$\mathbb{P}(\mathcal{E}_{n,d}) \ge \exp\left(-O\big(n\sqrt{d}\big)\right). \tag{3.7}$$

While weaker than (3.6), this lower bound was enough to prove the quarter-circular law for the singular value distribution of M using the conditioning trick (in fact Tran treated the more general case of rectangular 0/1 matrices with constant row and column sums, for which he proved the Marchenko–Pastur law). This followed earlier work in [TVW13] establishing the semi-circular law for undirected random regular graphs with $d \to \infty$ by a similar approach. It is worth noting that the Marchenko–Pastur and semi-circular laws were also obtained in [DP12] and [DJ16] for the sparse regime $\omega(1) \leq d \leq n^{o(1)}$, using the fact that d-regular graphs converge locally (in a quantitative Benjamini–Schramm sense) to d-regular trees.

Hence, we see that with (3.7) we are limited to importing properties of the iid matrix M_p that hold with probability $1 - O(\exp(-Cn\sqrt{np}))$ for some sufficiently large C, and this can be slightly relaxed by using the formula (3.6) for the appropriate range of d. We note in particular that the results of the present work cannot be obtained by the conditioning trick.

On the continuous side, the conditioning trick was used to study uniform random doubly stochastic matrices in [CDS] and [Ngu14]. In [CDS], Chatterjee, Diaconis and Sly noted that this distribution can be obtained by conditioning a matrix with iid exponentially distributed entries. They proved the quarter circular law by similar lines to [Tra], relying on an asymptotic formula of Canfield and McKay from [CM10] for the volume of the Birkhoff polytope. Nguyen built on this work in [Ngu14] to prove the circular law for this model.

3.1.2 Main results and conjectures

Our main result for this chapter is an analogue of Komlós's Theorem 3.1.1 for rrd matrices, assuming that the matrix is not too sparse or too dense. Specifically, we assume that $\min(d, n-d) \geq C_0 \log^2 n$ for a sufficiently large constant $C_0 > 0$. As in Chapter 2 we make use of couplings rather than the conditioning trick discussed above an in Section 2.1.1. We give more detail and motivation for the proof strategy in Section 3.1.3 below.

Theorem 3.1.2 (Main result)**.** *There are absolute constants $C_0, c_0 > 0$ such that the following holds. Assume $\min(d, n-d) \geq C_0 \log^2 n$, and let M be a uniform random element of*

$\mathcal{M}_n(d)$. *Then*

$$\mathbb{P}\left(\det(M) = 0\right) = O(d^{-c_0}). \tag{3.8}$$

Remark 3.1.3. The proof shows that we may take $c_0 = 1/18$, though we do not expect this bound to be optimal (see Conjectures 3.1.7 and 3.1.8 below).

Remark 3.1.4 (Lower bound on d). It is possible that our argument could be extended to only assume $\min(d, n-d) \geq C_0 \log n$, but a new approach will be certainly necessary beyond that – see Remark 3.2.6.

Remark 3.1.5. One can easily show that a matrix $M \in \mathcal{M}_n(d)$ is invertible if and only if the "complementary" matrix M' with entries $M'(i,j) = 1 - M(i,j)$ is invertible. Hence, in the proof of Theorem 3.1.2 we may assume that $p := \frac{d}{n} \leq \frac{1}{2}$.

Remark 3.1.6. Very recently (after the final version of these results were sent for publication) an extension of the bound (3.8) to lower values of d has been accomplished in [LLT+], along with an improvement in the exponent c_0. Specifically, they are able to show that for some absolute constants $C, c > 0$, $\mathbb{P}(\det(M) = 0) = O((\log^3 d)/\sqrt{d})$ if $C \leq d \leq cn/\log^2 n$. Together with Theorem 3.1.2 this implies that a uniform random element $M \in \mathcal{M}_n(d)$ is invertible with probability tending to 1 as $n \to \infty$ if $\min(d, n-d)$ grows to ∞ at any speed, rather than at speed at least $\log^2 n$. See Remark 3.2.7 for some additional comments on this result.

We believe that when d is of linear size, the singularity probability is exponentially small, similarly to the bound (3.2) for iid sign matrices.

Conjecture 3.1.7. *Fix* $p_0 \in \left(0, \frac{1}{2}\right)$ *and assume* $\min(d, n-d) \geq p_0 n$. *Then*

$$\mathbb{P}\left(\det(M) = 0\right) \leq Ce^{-cn}$$

for constants $C, c > 0$ *depending only on* p_0.

We also conjecture that rrd matrices are invertible with high probability for much smaller values of d:

Conjecture 3.1.8. *There are absolute constants $C, c > 0$ such that for any $3 \le d \le n-3$ we have*

$$\mathbb{P}\left(\det(M) = 0\right) \le Cn^{-c}.$$

This mirrors a similar conjecture of Vu in [Vu08] on the adjacency matrices of undirected d-regular graphs, which are the symmetric analogue of M. When d is bounded, considering the event that two columns of M are parallel shows that we cannot hope for better than a polynomial bound on the singularity probability. M is obviously invertible when $d = 1$ as it is a permutation matrix in this case. On the other hand, for the case $d = 2$ we have the following:

Observation 3.1.9. *Let M be a uniform random element of $\mathcal{M}_{n,2}$. Then M is singular asymptotically almost surely.*

Proof Sketch. One first observes that M is identically distributed to $P(I + P_0)$ where P and P_0 are independent permutation matrices, with P uniform random and P_0 uniform among permutation matrices with 0 diagonal (i.e. P_0 is associated to a uniform random *derangement*). Hence, the probability that M is invertible is equal to the probability that $I+P_0$ is invertible. Now we conjugate by a permutation matrix Q to put P_0 in block diagonal form according to its cycle structure. The resulting block matrix $I + Q^{\mathsf{T}}P_0Q$ has blocks of the form

$$\begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Such a matrix is invertible if and only if it is of odd dimension. Hence, the probability that M is invertible is equal to the probability that a uniform random derangement decomposes into only odd cycles. The reader may verify that for $\sigma \in \mathrm{Sym}(n)$ a uniform random permutation, we have

$$\mathbb{P}(\sigma \text{ contains only odd cycles}) = o(1) \tag{3.9}$$

(for the precise asymptotics see Exercise 5.10 in [Sta99]). The result then follows from the fact that a uniform random permutation is a derangement with probability $\Omega(1)$ – see for instance [McK]. $\square$

Remark 3.1.10. One may similarly show that the sum of two independent and uniformly distributed permutation matrices $P_1 + P_2$ is singular asymptotically almost surely.

Next we give a consequence of Theorem 3.1.2 for random sign matrices. Note that if we draw an iid matrix of signs Ξ as in Theorem 3.1.1 and condition on the event that all rows and columns sum to 0, the resulting matrix Ξ_0 will be singular with null vector $\mathbf{1} = (1, 1, \dots, 1) \in \mathbb{R}^n$. It is an easy consequence of Theorem 3.1.2 that this is usually the only obstruction for invertibility.

Corollary 3.1.11. *Assume n is even, and let Ξ_0 be an $n \times n$ matrix of iid uniform ± 1 signs conditioned to have each row and column sum to 0. Then with high probability,* $\ker(\Xi_0) = \langle \mathbf{1} \rangle$.

Proof. Let $M = \frac{1}{2}\big(\Xi_0 + \mathbf{1}\,\mathbf{1}^{\mathsf{T}}\big)$. Then M is an rrd matrix with $d = n/2$. For $x \in \mathbb{R}^n$, write

$$x = \overline{x}\,\mathbf{1} + x_0$$

with $\overline{x} = \frac{1}{n}\sum_i x(i)$ and x_0 the orthogonal projection of x to $\langle \mathbf{1} \rangle^{\perp}$, the space of mean-zero vectors. One then verifies that $x \in \ker(\Xi_0)$ if and only if $x_0 \in \ker(M)$, and the result follows from Theorem 3.1.2. $\square$

Our next result concerns *signed rrd matrices*. Let $\mathcal{M}_{n,d}^{\pm}$ denote the set of $n \times n$ matrices $M_{\pm}$ with entries in $\{+1, 0, -1\}$ satisfying the constraints

$$d = \sum_{i=1}^{n} \big|M_{\pm}(i,k)\big| = \sum_{j=1}^{n} \big|M_{\pm}(k,j)\big| \tag{3.10}$$

for all $k \in [n]$. We have the following analogue of Theorem 3.1.2:

Theorem 3.1.12 (Signed rrd matrices are invertible a.a.s.)**.** *Assume $C \log^2 n \le d \le n$ for a* *sufficiently large constant $C > 0$, and let $M_\pm$ be a uniform random element of $\mathcal{M}^\pm_{n,d}$. Then*

$$\mathbb{P}\left(\det(M_\pm) = 0\right) = O(d^{-1/4}). \tag{3.11}$$

Note that in contrast to Theorem 3.1.2, in the above result we don't need to assume an upper bound on d (apart from the trivial one). This is because of the additional randomness of the Bernoulli signs in Ξ: as d approaches n, M approaches the non-random (and singular) matrix of all 1s, while $M_\pm$ approaches an iid sign matrix. In particular, by considering $d = n$ we see that Theorem 3.1.12 is a generalization of Komlós' Theorem 3.1.1, up to a small loss in the exponent of n from the bound (3.1).

The signed rrd matrix $M_\pm$ is easier to work with than the unsigned rrd matrix M due to the following alternative description. Letting M be the rrd matrix with entries

$$M(i,j) := |M_\pm(i,j)| \tag{3.12}$$

we have $M_\pm \stackrel{d}{=} M \circ \Xi$, where Ξ is an iid sign matrix independent of M. (Here $\circ$ denotes the Hadamard (or Schur) product, so that $M_\pm(i,j) = M(i,j)\Xi(i,j)$ for each $i,j \in [n]$.) We refer to M as the "base" or "support" of the signed rrd matrix $M_\pm$. Roughly speaking, our approach to proving Theorem 3.1.12 will be to condition on a "good" realization of the base rrd matrix M and proceed using only the randomness of the iid signs. We will then have to show that such good configurations M occur with high probability.

The conditions of a good configuration are most naturally stated in terms of the d-regular digraph $\Gamma = (V, E)$ which has M as its adjacency matrix. We will use the notation $\mathcal{N}_M(i)$, $e_M(A,B)$, $\mathrm{Co}_M(i_1,i_2)$, etc. introduced in Section 2.1.5. Additionally, for $S \subset [n]$ we denote

$$\mathcal{N}_M(S) := \bigcup_{i \in S} \mathcal{N}_M(i). \tag{3.13}$$

Roughly speaking, the base matrix M is a good configuration if the associated digraph satisfies certain *expansion* properties. In Section 3.3.2 we prove that the rrd matrix M

satisfies all of the necessary properties with overwhelming probability if $d = \omega(\log n)$. The proofs rely on the sharp tail bounds for the edge counts $e_M(A,B)$ provided by Theorem 2.1.5. The proof of Theorem 3.1.2 for the unsigned rrd matrix M will rely more heavily on the expansion properties of Section 3.3.2.

It turns out that to prove Theorem 3.1.12 by this approach we do not need all of the expansion properties enjoyed by M. Hence, we will actually prove the following result, where M is replaced by a general 0/1 matrix Σ, and the event $\mathcal{G}(d)$ distills the required expansion properties. In particular, Theorem 3.1.13 is independent of the results in Chapter 2.

Theorem 3.1.13 (0/±1 matrices with expanding support are invertible a.a.s.)**.** *Let Σ be a random or deterministic 0/1 matrix. For $d \in [n]$, let $\mathcal{G}(d)$ be the event that Σ enjoys the following* expansion properties *with constants $c_1, c_2, C_2 > 0$ and a parameter $\kappa_3 \geq 1$:*

0. *(Minimum degree) Every row and column of Σ has at least d nonzero entries. That is, for all $i \in [n]$,*
$$|\mathcal{N}_\Sigma(i)|,\ |\mathcal{N}_{\Sigma^\mathsf{T}}(i)| \geq d. \tag{3.14}$$

1. *(Expansion of small sets) There is some constant $c_1 > 0$ such that for all $\gamma \in (0, c_1]$, for all $S \subset [n]$ such that $|S| \leq \frac{\log n}{2\gamma}\frac{n}{d}$, we have*
$$|\mathcal{N}_\Sigma(S)|,\ |\mathcal{N}_{\Sigma^\mathsf{T}}(S)| \geq \frac{\gamma}{\log n} d|S|.$$

2. *(No large sparse minors) There are constants $C_2, c_2 > 0$ such that for all $A, B \subset [n]$ satisfying $|A|, |B| \geq C_2 \frac{n}{d}\log n$ we have*
$$e_\Sigma(A,B) \geq c_2 \frac{d}{n}|A||B|.$$

3. *(No thin dense minors) There is a parameter $\kappa_3 \in [1, \infty)$, possibly depending on n, such that for any $S, B \subset [n]$,*
$$e_\Sigma(S,B),\ e_\Sigma(B,S) \leq \kappa_3 d|S|.$$

(In particular, taking S to be a singleton we have the degree bounds

$$|\mathcal{N}_\Sigma(i)|, |\mathcal{N}_{\Sigma^\mathsf{T}}(i)| \le \kappa_3 d$$

to complement (3.14) *above.)*

Let Ξ be an iid sign matrix independent of Σ, and put $H = \Sigma \circ \Xi$. There is a constant $C_0' > 0$ depending on c_1, c_2, C_2 such that if $d \ge C_0' \log^2 n$, then

$$\mathbb{P}\Big(\{\det(H) = 0\} \wedge \mathcal{G}(d) \Big) = O(\kappa_3 d^{-1/4}). \tag{3.15}$$

In particular, if properties (0)-(3) hold a.a.s. for Σ with $d \ge C_0' \log^2 n$ and $\kappa_3 = o(d^{1/4})$, then H is invertible a.a.s.

In Section 3.3.2 we will show that with $\Sigma = M$ the rrd matrix, the event $\mathcal{G}(d)$ holds with overwhelming probability for some constants $c_1, c_2, C_2 > 0$ (we can take $\kappa_3 = 1$ in this case) assuming $d = \omega(\log n)$, at which point Theorem 3.1.12 follows from Theorem 3.1.13.

The proof of Theorem 3.1.13 will follow the general outline of Theorem 3.1.2, but each stage will be easier due to the independence of the entries of Ξ. Hence, we believe it will benefit the reader to first see the arguments for H as warmups to the more complicated arguments involving couplings for the rrd matrix M, and have structured the chapter accordingly. However, nothing from the proof of Theorem 3.1.13 is needed for the proof of Theorem 3.1.2, so the reader who is only interested in the proof of the main theorem can skip the sections devoted to H (namely, Sections 3.4.1 and 3.5.1).

3.1.3 The general strategy

Now we give a high level discussion of our couplings approach to proving invertibility of an rrd matrix M. The strategy is similar in spirit to the one used by Rudelson and Vershynin in the recent work [RV14] on the smallest singular value of perturbations of deterministic matrices by Haar unitary or orthogonal matrices. As the rrd matrix M has discrete distribution, the couplings we define will be of a very different nature from the ones considered in that work. Nevertheless, our approach for dealing with dependent random variables owes much to the work [RV14] on a conceptual level.

In order to improve on the strategy of conditioning on an iid Bernoulli(p) matrix M_p as in (3.5), we would like to show that the events $\{\det(M_p) = 0\}$ and $\{M_p \in \mathcal{M}_n(d)\}$ are approximately independent in some sense. Indeed, proceeding as in (3.4) gives

$$\begin{aligned} \mathbb{P}(\,\det(M) = 0\,) &= \mathbb{P}\big(\,\det(M_p) = 0 \,\big|\, M_p \in \mathcal{M}_n(d)\,\big) \\ &\le \frac{\mathbb{P}(\det(M_p) = 0)}{\mathbb{P}(M_p \in \mathcal{M}_n(d))} \end{aligned} \tag{3.16}$$

which is only sharp for the worst case that we have the containment

$$\{\det(M_p) = 0\} \subset \{M_p \in \mathcal{M}_n(d)\}\,. \tag{3.17}$$

Of course, (3.17) is likely far from the truth. This motivates us to better understand the structure of the set $\mathcal{M}_n(d)$; specifically, we try to identify *symmetries* of this set. If we can identify a large class of operations $\Phi : \mathcal{M}_n(d) \to \mathcal{M}_n(d)$ which leave the distribution of a uniform random element $M \in \mathcal{M}_n(d)$ invariant, then we could select such an operation Φ *at random* from this class and form a new rrd matrix $\widetilde{M} = \Phi(M)$. Now to bound the event that some property P holds for M, we may replace M with $\widetilde{M}$:

$$\begin{aligned} \mathbb{P}(\,P \text{ holds for } M\,) &= \mathbb{P}(\,P \text{ holds for } \widetilde{M}\,) \\ &= \mathbb{E}\,\mathbb{P}\big(\,P \text{ holds for } \widetilde{M} \,\big|\, M\big) \end{aligned}$$

and proceed to bound the inner probability using only the randomness we have "injected" via the map Φ. This approach can be very powerful if we can design the map Φ to involve a large number of independent random variables.

This strategy was used in [RV14] to obtain bounds of the form

$$\mathbb{P}(s_n(D+U) \le t) \ll t^c n^C \tag{3.18}$$

for some absolute constants $C, c > 0$, where D is a deterministic matrix (satisfying some additional hypotheses) and U is a Haar-distributed unitary or orthogonal matrix. Since the random matrices in this case are drawn from a group, there is no shortage of symmetries to consider for injecting independence. Furthermore, the availability of *continuous* symmetries allowed for the injection of random variables possessing smooth bounded density (such as iid Gaussians). This gave quick access to anti-concentration or "small ball" estimates, which play a fundamental role in all currently known (to this author at least) proofs of invertibility for random matrices.

The bound (3.18) had implications for the Single Ring Theorem, proved by Guionnet, Krishnapur and Zeitouni in [GKZ11], for the limiting spectral distribution of certain random matrices with prescribed singular values; specifically, it was shown that a hypothesis in [GKZ11] could be disposed of. It was also used in the proof by Basak and Dembo in [BD13] of the limiting spectral distribution for the sum of a fixed number of independent Haar unitary or orthogonal matrices. It was conjectured in [BC12] that the same law should hold for the random matrix

$$M_{\mathrm{Perm}} := P_1 + \cdots + P_d$$

where the summands are iid uniform $n \times n$ permutation matrices – this can be viewed as a sparse version of the rrd matrix M. It is possible that the smallest singular value of M_{Perm} could be controlled by an extension of the ideas used in the present work, though the extreme sparsity of this matrix will likely call for new ideas.

The present setting of rrd matrices is a little more complicated than the case of Haar unitaries as the distribution is discrete, and $\mathcal{M}_n(d)$ is not a group (except when $d = 1$ of course, but then the problem is trivial). The basic building block for our coupled pairs $(M, \widetilde{M})$ is the "simple switching" operation introduced in Chapter 2, which we now recall. Letting

$$\mathbf{I}_2 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{J}_2 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{3.19}$$

we can replace a 2×2 minor of M by $\mathbf{I}_2$ if it is $\mathbf{J}_2$ and $\mathbf{J}_2$ if it is $\mathbf{I}_2$ – indeed, note that this preserves the row and column sums. If i_1, i_2 and j_1, j_2 are the row and column indices, respectively, of such a minor, then in the associated digraph $\Gamma = (V, E)$ we are alternating between the following edge configurations at vertices i_1, i_2, j_1, j_2:

where we use solid arrows to depict directed edges, and dashed arrows to indicate the absence of a directed edge (i.e. "non-edges").

We will want to form $\widetilde{M}$ by applying several switchings at non-overlapping 2×2 minors. Each minor is replaced with $\mathbf{I}_2$ or $\mathbf{J}_2$ uniformly at random, independently of all other switchings. We can encode the outcomes of the random switchings with iid uniform signs – this will give us access to anti-concentration estimates for random walks (specifically Erdős' Theorem 3.2.1 below). The formal construction, called the "shuffling coupling", is given in Section 3.3.1.

3.1.4 Organization of the chapter

The rest of the chapter is organized as follows. In Section 3.2 we describe the ideas of the proof in more detail, reviewing the approach introduced by Komlós to classify potential null vectors as structured and unstructured, and illustrating our use of couplings by solving a toy problem. Section 3.3 gives the formal statements and proofs for the tools that were motivated in Section 3.2, namely the "shuffling" coupling of rrd matrices, the discrepancy properties from Chapter 2, and a concentration inequality for the symmetric group due to Chatterjee. In Section 3.3.2 we use the discrepancy properties to deduce Theorem 3.1.12 from Theorem 3.1.13. In Sections 3.4 and 3.5 we bound the events that our random matrices have "unstructured" and "structured" null vectors, respectively – for the signed matrix H "structured" will mean "sparse", while for the rrd matrix M it will mean that the null vector has a large level set. (Note that in the recent literature on invertibility of iid matrices the term "structured" is used for vectors whose components lie in a set that is well-approximated in some sense by a *generalized arithmetic progression* – see [TV07, TV09, RV08].) In each section, we first treat the signed matrix H as a warmup to the more complicated arguments for M. However, the reader who is only interested in the proof of the main theorem concerning the rrd matrix M can skip Sections 3.4.1 and 3.5.1.

3.1.5 Notation and terminology

In addition to the notation listed in Section 1.2 we will use the following conventions in the present chapter.

Some distinguished constants from the statements of Theorems and Propositions (such as C_0, c_0 in Theorem 3.1.2) have numbered subscripts so that they can be more easily tracked through the arguments. We still allow hidden constants in asymptotic notation to depend on these numbered constants.

Most events will be denoted by the letters $\mathcal{E}, \mathcal{B}$, and $\mathcal{G}$, where the latter two denote "bad" and "good" events, respectively. Their meaning may vary from proof to proof, but will remain fixed for the duration of each proof. $\mathbb{E}_X$ and $\mathbb{P}_X$ denote expectation and probability, respectively, conditional on all random variables but X.

We make use of the following terminology for sequences of events.

Definition 3.1.14 (Frequent events)**.** An event $\mathcal{E}$ depending on n holds

- *asymptotically almost surely* (a.a.s.) if $\mathbb{P}(\mathcal{E}) = 1 - o(1)$,
- *with high probability* (w.h.p.) if $\mathbb{P}(\mathcal{E}) = 1 - O(n^{-c})$ for some absolute constant $c > 0$,
- *with overwhelming probability* (w.o.p.) if $\mathbb{P}(\mathcal{E}) = 1 - O_C(n^{-C})$ for any constant $C > 0$.

Given ordered tuples of row and column indices $(i_1, \dots, i_a)$ and $(j_1, \dots, j_b)$, we denote by $M_{(i_1,\dots,i_a)\times(j_1,\dots,j_b)}$ the $a \times b$ matrix with (k, l) entry equal to the (i_k, j_l) entry of M. (Note for instance that the sequence $(i_1, \dots, i_a)$ need not be increasing.) For $A, B \subset [n]$, with $M_{A\times B}$ the increasing ordering of the elements of A, B is implied. We also recall from (3.19) our notation

$$\mathbf{I}_2 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{J}_2 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{3.20}$$

We use $\circ$ to denote the matrix Hadamard product; that is, for $n \times m$ matrices M_1, M_2, $M_1 \circ M_2$ is the $n \times m$ matrix with entries $M_1(i,j)M_2(i,j)$. We will use this notation for row vectors as well (which is the case $n = 1$).

We use notation that views a vector $x \in \mathbb{R}^n$ as a function $x : [n] \to \mathbb{R}$. In particular, the ith component of x is denoted $x(i)$. We also define the support of x as

$$\operatorname{spt}(x) := \big\{ i \in [n] : x(i) \neq 0 \big\} = [n] \setminus x^{-1}(0).$$

We write $\mathbb{R}^T \subset \mathbb{R}^n$ for the subspace of vectors supported on $T \subset [n]$. We let $\mathbf{1}$ denote the all-ones vector $(1, \dots, 1) \in \mathbb{R}^n$. "Null vector" will mean "right null vector" unless otherwise stated. The span of a single vector x is denoted by $\langle x \rangle$.

We will also make use of the graph-theoretic notation introduced in Section 2.1.5.

3.2 Ideas of the proof

Our general approach to Theorem 3.1.2 is inspired by Komlós's proof of the analogous theorem for iid sign matrices (Theorem 3.1.1 above). After briefly reviewing Komlós's argument below, we will discuss the new ideas that are necessary to treat rrd matrices.

3.2.1 The strategy of Komlós

A key ingredient of Komlós's proof is the following "discrete small ball estimate" for random walks due to Erdős.

Theorem 3.2.1 (Anti-concentration for random walks [Erd45])**.** *Let $x \in \mathbb{R}^n$ be a fixed* *nonzero vector, and let* $\xi : [n] \to \{\pm 1\}$ *be a sequence of iid uniform signs. Then*

$$\sup_{a\in\mathbb{R}} \mathbb{P}\left\{ \sum_{i=1}^{n} x(i)\xi(i) = a \right\} \ll |\operatorname{spt}(x)|^{-1/2} \tag{3.21}$$

where we recall the notation $\operatorname{spt}(x) = \big\{ i \in [n] : x(i) \neq 0 \big\}$.

Proof. of Theorem 3.1.1. We want to bound the bad event

$$\mathcal{B} := \big\{ \det(\Xi) = 0 \big\} = \big\{ \exists \text{ nonzero } x \in \mathbb{R}^n : \Xi x = 0 \big\}. \tag{3.22}$$

The idea is to separately consider the possibility of "structured" and "unstructured" null vectors x. Here the right notion of structure is *sparsity*. Say that $x \in \mathbb{R}^n$ is k-sparse if $|\operatorname{spt}(x)| \le k$.

Proposition 3.2.2 (No structured null vectors for Ξ)**.** *For any fixed* $\eta \in (0,1)$*, with overwhelming probability* Ξ *has no nontrivial* $(1-\eta)n$*-sparse null vectors.*

We defer the proof of this proposition to the end. Fix $\eta \in (0,1)$. We say that $x \in \mathbb{R}^n$ is "structured" if x is $(1-\eta)n$-sparse, and "unstructured" otherwise. Since Ξ is identically distributed to its transpose, we may now restrict to the event $\mathcal{G}$ on which Ξ has no structured left or right null vectors.

For each $i \in [n]$, let R_i denote the ith row of Ξ, and denote $V_i = \operatorname{span}(R_{i'} : i' \neq i)$. Define the events

$$\mathcal{B}_i := \mathcal{G} \wedge \{R_i \in V_i\}.$$

On $\mathcal{B} \wedge \mathcal{G}$, Ξ must have an unstructured left null vector, which implies that $\mathcal{B}_i$ holds for at least $(1-\eta)n$ values of $i \in [n]$. By double counting we then have that

$$\sum_{i=1}^{n} \mathbb{P}(\mathcal{B}_i) \ge (1-\eta)n\, \mathbb{P}(\mathcal{B} \wedge \mathcal{G}). \tag{3.23}$$

Since the rows of Ξ are exchangeable, all of the summands on the left hand side are equal to $\mathbb{P}(\mathcal{B}_1)$, say, and so

$$\mathbb{P}(\mathcal{B} \wedge \mathcal{G}) \leq \frac{1}{1-\eta}\, \mathbb{P}(\mathcal{B}_1). \tag{3.24}$$

By our bound on $\mathbb{P}(\mathcal{G}^c)$ from Proposition 3.2.2, it only remains to show that $\mathbb{P}(\mathcal{B}_1) \ll n^{-1/2}$.

We condition on the rows $R_2, \ldots, R_n$ of Ξ, which fixes their span V_1. Condition also on a unit normal vector $u \in V_1^{\perp}$, drawn independently of R_1. We have

$$\{R_1 \in V_1\} \subset \{R_1 \cdot u = 0\}\,. \tag{3.25}$$

On $\mathcal{B}_1$ we have that u is perpendicular to every row of Ξ, and is hence a left null vector. By our restriction to $\mathcal{G}$ we may hence assume that u is unstructured. By Theorem 3.2.1 we may now use the randomness of R_1 to conclude the desired bound

$$\mathbb{P}(\mathcal{B}_1) \leq \mathbb{P}(\mathcal{G} \wedge \{R_1 \cdot u = 0\}\,) \ll_\eta n^{-1/2}.$$

We turn to the proof of Proposition 3.2.2. Define the events

$$\mathcal{E}_k = \{\exists\, x \in \mathbb{R}^n :\ 0 < |\,\mathrm{spt}(x)| \leq k,\ \Xi\, x = 0\}\,.$$

Our aim is to bound

$$\mathbb{P}\big(\mathcal{E}_{(1-\eta)n}\big) = \sum_{k=2}^{\lfloor (1-\eta)n \rfloor} \mathbb{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1}) \tag{3.26}$$

(noting that $\mathcal{E}_1$ is empty). It suffices to show that $\mathbb{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1})$ is exponentially small for arbitrary fixed $2 \leq k \leq (1-\eta)n$.

Fix k in this range. On $\mathcal{E}_k \setminus \mathcal{E}_{k-1}$ there is a right null vector x with exactly k nonzero components. We may spend a factor $\binom{n}{k}$ to assume that x is supported on $[k]$ (using column exchangeability). Now on the complement of $\mathcal{E}_{k-1}$, the first k columns of Ξ must span a space of dimension $k-1$. It follows that there are $k-1$ linearly independent rows of the left $n \times k$ minor of Ξ. By row exchangeability we may spend another factor $\binom{n}{k-1}$ to assume the first $k-1$ rows are linearly independent. To summarize,

$$\mathbb{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1}) \le \binom{n}{k}\binom{n}{k-1}\mathbb{P}(\mathcal{E}_k') \tag{3.27}$$

where

$$\mathcal{E}_k' := \Big\{\exists x \in \mathbb{R}^n : \Xi\, x = 0,\ \mathrm{spt}(x) = [k],\ R_1, \dots, R_{k-1} \text{ are linearly independent}\Big\}.$$

Now note that by linear independence, on $\mathcal{E}_k'$ we have that x is determined by the first $k-1$ rows. Conditioning on these rows fixes x. Then by the independence of the rows of Ξ we have

$$\begin{aligned}\mathbb{P}\big(\mathcal{E}_k' \,\big|\, R_1, \dots, R_{k-1}\big) &\le \mathbb{P}(R_i \cdot x = 0\ \forall i \in [k, n]) \\ &= \mathbb{P}(R_n \cdot x = 0)^{n-k+1}. \end{aligned} \tag{3.28}$$

Since $|\,\mathrm{spt}(x)| = k$, by Theorem 3.2.1 we can bound

$$\mathbb{P}(R_n \cdot x = 0) \le \min\big[1/2, O(k^{-1/2})\big]. \tag{3.29}$$

Combining this bound with (3.28), (3.27) and the inequality $\binom{n}{n-k} \le (en/(n-k))^{n-k}$ we conclude

$$\mathbb{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1}) \ll \exp\left\{(n-k)\left[C + 2\log\frac{n}{n-k} - \log\sqrt{k}\right]\right\} \tag{3.30}$$

which is more than sufficiently small if $n-k \ge \eta n$ for any fixed $\eta \in (0,1)$ (in fact we can allow $n-k$ as small as $C'n^{-1/4}$ for a sufficiently large absolute constant $C' > 0$). □

3.2.2 Structured and unstructured null vectors

It turns out that Proposition 3.2.2 is robust under some zeroing out of the entries of M. Specifically, we can show an analogous result for the matrix $H = \Sigma \circ \Xi$ from Theorem 3.1.13.

Proposition 3.2.3 (No structured null vectors for H)**.** *For $\eta \in (0,1]$, let $\mathcal{G}_{\pm}^{\mathrm{sp}}(\eta)$ be the event* *that H has no nontrivial $(1-\eta)n$-sparse left or right null vectors. With hypotheses as in Theorem 3.1.13, we have that on $\mathcal{G}(d)$ the event $\mathcal{G}^{\mathrm{sp}}_{\pm}(\eta)$ holds with probability $1-O(n^{-100})$ if $\eta \in [C_1' d^{-1/4}, 1]$ for a sufficiently large absolute constant $C_1' > 0$.*

As for the rrd matrix M and Theorem 3.1.2, we will also treat structured null vectors separately, but it turns out that sparsity is no longer the right notion of structure. Instead, we will need to show that null vectors of M have *small level sets*:

Proposition 3.2.4 (No structured null vectors for M)**.** *For $\eta \in (0,1]$, let $\mathcal{G}^{\mathrm{sls}}(\eta)$ be the event that for any nontrivial left or right null vector x of M and for any $\lambda \in \mathbb{R}$,*

$$|x^{-1}(\lambda)| \leq \eta n. \tag{3.31}$$

With hypotheses as in Theorem 3.1.2, we have that $\mathcal{G}^{\mathrm{sls}}(\eta)$ holds with probability $1-O(n^{-100})$ if $\eta \geq C_1 d^{-c_0}$ for some absolute constants $C_1, c_0 > 0$ sufficiently large and small, respectively (the constant c_0 is the same as in Theorem 3.1.2 and can be taken to be $1/18$).

We prove Propositions 3.2.3 and 3.2.4 in Section 3.5. The reason for ruling out null vectors with large level sets will be apparent in the next section, where we describe our couplings approach.

3.2.3 Injecting a random walk

The proof in Section 3.2.1 proceeded by reducing to the event that $R_1 \cdot u = 0$, where R_1 is the first row of Ξ and u is a unit vector in $V_1^{\perp} = \mathrm{span}(R_2, \dots, R_n)^{\perp}$. Then we used independence of the entries of Ξ in two ways:

1. Independence of the rows of Ξ allowed us to condition on $R_2, \dots, R_n$ to fix u, without affecting the distribution of R_1.
2. Independence of the components of R_1 allowed us to view the dot product $R_1 \cdot u$ as a random walk, to which we could apply the anti-concentration result Theorem 3.2.1.

The rrd matrix M enjoys neither of these properties. However, we will be able to accomplish something like (2) above by defining an appropriate coupling of rrd matrices using switchings. It will take some care to implement this without having the independence between rows (1).

To illustrate our couplings approach, let us consider a toy problem: to control the event that the first two rows lie in the span of the remaining rows, i.e. to show

$$\mathbb{P}\left\{R_1, R_2 \in V_{(1,2)}^{\perp}\right\} = o(1) \tag{3.32}$$

where $V_{(1,2)} := \mathrm{span}(R_3, \ldots, R_n)$. We will see later that this can be used to control the event that M has corank at least 2 (see Lemma 3.4.3). For now we will operate under the following

Assumption 3.2.5. $n \ll d \leq \frac{n}{2}$.

Thus, we are assuming M is a dense rrd matrix. In the next section we will discuss some of the new ideas necessary to treat sparse matrices.

Blindly following the proof from Section 3.2.1, we condition on the rows $R_3, \ldots, R_n$ to fix the space $V_{(1,2)}$, and pick a unit vector $u \in V_{(1,2)}^{\perp}$, say uniformly and independently of R_1, R_2 under the conditioning. Now it suffices to show

$$\mathbb{P}\left(R_1 \cdot u = 0 \,\middle|\, R_3, \ldots, R_n\right) = o(1). \tag{3.33}$$

We need to understand how R_1 and R_2 are distributed under the conditioning on $R_3, \ldots, R_n$. Recall from Section 2.1.5 the sets $\mathrm{Co}_M(1,2)$, $\mathrm{Ex}_M(1,2)$, $\mathrm{Ex}_M(2,1)$, which the partition the vertex-pair neighborhood $\mathcal{N}_M(\{1,2\})$. Now since the entries of each column sum to d, by fixing $R_3, \ldots, R_n$ we have fixed which columns of M need both, neither, or just one of their first two components equal to 1 in order to meet the constraint. This fixes the sets $\mathrm{Co}_M(1,2)$ and $\mathrm{Ex}_M(1,2) \cup \mathrm{Ex}_M(2,1)$. Furthermore, by the row sums constraint, we must have

$$|\,\mathrm{Ex}_M(1,2)| = d - |\,\mathrm{Co}_M(1,2)| = |\,\mathrm{Ex}_M(2,1)| \tag{3.34}$$

$$
\begin{array}{cl}
 & \qquad\qquad\qquad\qquad\qquad \overset{\pi}{\curvearrowright} \\
 & \underbrace{Co(1,2)} \qquad \underbrace{Ex(1,2)}_{j} \qquad\qquad \underbrace{Ex(2,1)}_{\pi(j)} \\
\begin{matrix}1\\2\\ \\ \end{matrix} & \begin{bmatrix} 1 & \cdots & 1 & 1 & \color{red}{1} & \cdots & 1 & 0 & \color{red}{0} & \cdots & 0 & 0 & \cdots \\ 1 & \cdots & 1 & 0 & \color{red}{0} & \cdots & 0 & 1 & \color{red}{1} & \cdots & 1 & 0 & \cdots \\ & & & \vdots & & & & & & \vdots & & & \end{bmatrix}
\end{array}
$$

Figure 3.1: The shuffling coupling: the minor $M_{(1,2)\times(j,\pi(j))}$ (in red) will be replaced with $\mathbf{I}_2$ or $\mathbf{J}_2$ according to a random sign $\xi(j)$. We do this independently for each $j \in \mathrm{Ex}(1,2)$.

It follows that with $R_3, \ldots, R_n$ fixed, the only remaining randomness is in the uniform random equipartition of the deterministic set $\mathrm{Ex}_M(1,2) \cup \mathrm{Ex}_M(2,1)$ into the sets $\mathrm{Ex}_M(1,2)$, $\mathrm{Ex}_M(2,1)$. See Figure 3.1.

We re-randomize the sets $\mathrm{Ex}_M(1,2)$, $\mathrm{Ex}_M(2,1)$ in the following way. Under this conditioning, pick a bijection $\pi : \mathrm{Ex}_M(1,2) \to \mathrm{Ex}_M(2,1)$ uniformly at random. Now for each $j \in \mathrm{Ex}_M(1,2)$ we have

$$M_{(1,2)\times(j,\pi(j))} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} =: \mathbf{I}_2.$$

Having obtained a sequence of "switchable" 2×2 minors, we can apply random switchings (with terminology as in Section 3.1.3). Let $\xi : [n] \to \{\pm 1\}$ be a sequence of iid uniform signs, independent of all other variables. For each $j \in \mathrm{Ex}_M(1,2)$, we replace the minor $M_{(1,2)\times(j,\pi(j))}$ with the random minor

$$\mathbf{I}_2 \,\mathbb{1}(\xi(j) = +1) + \mathbf{J}_2 \,\mathbb{1}(\xi(j) = -1)$$

(with notation as in (3.20)). Call the resulting matrix $\widetilde{M}$. It is not hard to show that $\widetilde{M}$ is also an rrd matrix after undoing all of the conditioning (see the proof of Lemma 3.3.2 below).

We have hence obtained a coupled pair $(M, \widetilde{M})$ of rrd matrices (more precisely, we have defined a coupling $(M, \widetilde{M}, \pi, \xi)$ on an enlarged probability space such that the marginals for the first two entries are uniform). Let $\widetilde{R}_i$ denote the ith row of $\widetilde{M}$. Replacing M with $\widetilde{M}$ in

(3.33), it suffices to show

$$\mathbb{P}\left(\widetilde{R}_1 \cdot u = 0 \,\middle|\, M\right) = o(1). \tag{3.35}$$

Now in the randomness of the iid signs $\xi(j)$, one sees that the dot product $\widetilde{R}_1 \cdot u$ is a random walk:

$$\widetilde{R}_1 \cdot u = W_0 + \sum_{j \in \mathrm{Ex}_M(1,2)} \xi(j) \frac{u(j) - u(\pi(j))}{2} \tag{3.36}$$

where W_0 is a term that does not depend on π or ξ. Applying Theorem 3.2.1 we have

$$\mathbb{P}_\xi \left\{\widetilde{R}_1 \cdot u = 0\right\} \ll \left|\left\{j \in \mathrm{Ex}_M(1,2) : \, u(j) \neq u(\pi(j))\right\}\right|^{-1/2}.$$

It remains to get a lower bound on the number of $j \in \mathrm{Ex}_M(1,2)$ for which $u(j) \neq u(\pi(j))$.

First we deal with the possibility that $\mathrm{Ex}_M(1,2)$ is a very small set. On average, we expect $\mathrm{Co}_M(1,2)$ to be of size roughly $p^2 n = d^2/n$. By (3.34) and our assumption $p \leq 1/2$ (from Remark 3.1.5) we have $\mathbb{E}\left|\mathrm{Ex}_M(1,2)\right| \gg d$. It was shown in Chapter 2 that codegrees in random regular digraphs are sharply concentrated (see Theorem 2.1.5) from which we can deduce that

$$\left|\mathrm{Ex}_M(1,2)\right| \gg d \tag{3.37}$$

off a negligibly small event.

Now we apply Proposition 3.2.4 and the randomness of π to argue that for most $j \in \mathrm{Ex}_M(1,2)$ we have $u(j) \neq u(\pi(j))$. Since $u \in V_{(1,2)}^\perp$ we have that u is a (right) null vector of the $(n-2) \times n$ matrix $M_{[3,n]\times[n]}$. By a small extension of Proposition 3.2.4 we may assume that u is unstructured, i.e. that all of its level sets are of size at most ηn, with η of size $\Theta(d^{-c_0})$ for some $c_0 > 0$ absolute. (In the actual proof we will argue that u is unstructured in a slightly different way, but in any case it comes down to an application of Proposition 3.2.4.) Now by (3.37) and Assumption 3.2.5, the sets $\mathrm{Ex}_M(1,2)$, $\mathrm{Ex}_M(2,1)$ are much larger than the level sets of u. Hence, in the randomness of π, it is very unlikely that we have $\pi(j) \in u^{-1}(u(j))$ for a large number of indices $j \in \mathrm{Ex}_M(1,2)$. Thus, off a negligibly small event we can deduce that most of the steps taken by the random walk (3.36) are nonzero, and hence

$$\mathbb{P}_{\xi,\pi}\left\{\widetilde{R}_1 \cdot u = 0\right\} \ll \left|\operatorname{Ex}_M(1,2)\right|^{-1/2} \ll d^{-1/2}. \tag{3.38}$$

Since we are assuming $d = \omega(1)$, we have completed the proof of (3.32).

To summarize, we bounded $\mathbb{P}(R_1, R_2 \in V_{(1,2)}^{\perp})$ by defining a coupling $(M, \widetilde{M}, \pi, \xi)$ on an enlarged probability space, with $\widetilde{M} \stackrel{d}{=} M$, and replacing M with $\widetilde{M}$ in (3.33). The variables M, π and ξ each played a special role:

1. In the randomness of M, we simply restricted to a couple of "good events": the event $\mathcal{G}^{\mathrm{sls}}(\eta)$ that null vectors are unstructured, and the event that codegrees are close to their expectations.

2. Conditional on M satisfying the good events, π was used to pair indices in $\operatorname{Ex}_M(1,2)$ with indices in $\operatorname{Ex}_M(2,1)$ to show that, off a small event, the random walk $\widetilde{R}_1 \cdot u$ takes many nonzero steps $\frac{1}{2}(u(j) - u(\pi(j)))$.

3. Conditional on good realizations of M and π, the randomness of ξ was used with Theorem 3.2.1 to finish the proof.

As remarked above, (3.32) can be used to deduce that M has corank at most 1 a.a.s. It then remains to deal with the event that $\operatorname{corank}(M) = 1$. This task is a little more complicated, and involves expressing a certain 2×2 determinant involving two randomly sampled rows of M as a random walk. See Section 3.4.5 for details.

3.2.4 Dealing with sparsity

In the previous section, we used Assumption 3.2.5 to guarantee that the level sets of the normal vector u were small in comparison to the neighborhood $\mathcal{N}_M(\{1,2\})$ (more precisely, the sets $\operatorname{Ex}_M(1,2)$ and $\operatorname{Ex}_M(2,1)$). Indeed, since the level sets are of size at most $\eta n = \Theta(nd^{-c_0})$ by Proposition 3.2.4, and since $|\operatorname{Ex}_M(1,2)| \gg d$ with high probability, we see upon rearranging that in the above argument we must assume $d \geq Cn^{1/(1+c_0)}$ for a sufficiently large constant $C > 0$. It turns out that the value $c_0 = 1/8$ is the limit of what can be obtained by our arguments in the proof of Proposition 3.2.4 for the case that $d = \omega(\sqrt{n})$. Hence, the argument of the previous section is limited to $d \geq Cn^{8/9}$.

In the present work we are able to take d as small as $C_0 \log^2 n$ using some new ideas. Rather than consider the event that $R_1, R_2 \in V_{(1,2)}^{\perp}$, we will draw row indices I_1, I_2 *at random* and seek to bound $\mathbb{P}\left\{ R_{I_1}, R_{I_2} \in V_{(I_1,I_2)}^{\perp} \right\}$. It can be shown that this leads to control on the event that $\operatorname{corank}(M) \geq 2$ (see Lemma 3.4.3). Conditional on I_1, I_2 and the remaining rows $(R_i)_{i \notin \{I_1, I_2\}}$, we will again select a unit normal vector u uniformly at random.

Whereas in Section 3.2.3 the distribution of u played no special role, here we will use it along with the randomness of I_1, I_2 to argue that it is very unlikely that a level set of u has large overlap with the neighborhood $\mathcal{N}_M(\{I_1, I_2\})$. Under conditioning on I_1, I_2, one can see that the "bad" realizations of u form an algebraic subset of the sphere. We will then use the simple fact that a proper algebraic subset of the sphere has surface measure zero. (This is perhaps the only part of the proof that is not strictly combinatorial.) The argument requires some care as the vector u and the set $\mathcal{N}_M(\{I_1, I_2\})$ are both dependent on I_1, I_2. See Section 3.4.4 for the detailed proof.

Remark 3.2.6 (Necessary lower bounds on d). While we need to assume $\min(d, n-d) \geq C_0 \log^2 n$ in Theorem 3.1.2, various parts of the argument work under a weaker lower bound assumption. Specifically, Theorem 3.1.2 follows from Proposition 3.4.2 and our ability to restrict to the following "good events":

1. $\mathcal{G}^{\mathrm{sls}}(\eta)$ (from Proposition 3.2.4), with η of order d^{-c_0};

2. $\mathcal{G}^{\mathrm{ex}}(\delta)$ (from Theorem 3.3.4), with δ a small fixed constant.

Proposition 3.2.4 shows 1. holds with high probability under the hypothesis $\min(d, n-d) \geq C_0 \log^2 n$ for $C_0 > 0$ sufficiently large. Theorem 3.3.4 establishes 2. holding with overwhelming probability if $\min(d, n-d) = \omega(\log n)$ (and in fact holds with high probability if $d \geq C_0 \log n$). Finally, Proposition 3.4.2 itself assumes no lower bound on d. Hence, the only real barrier to assuming a lower bound of order $\log n$ is Proposition 3.2.4. The lower bound $\min(d, n-d) \geq C_0 \log^2 n$ is only needed there for technical reasons, and we believe that an improvement to $C_0 \log n$ is possible. Beyond that, it is likely that an entirely different approach will be needed for the case $\min(d, n-d) = o(\log n)$, as the discrepancy properties in Section 3.3.2 would no longer hold with high probability, and these are essential to several parts of our argument.

Similar comments apply to Theorem 3.1.13, where the lower bound assumption $d \geq C_0' \log^2 n$ comes from Proposition 3.2.3.

Remark 3.2.7. As was mentioned in Remark 3.1.6, an extension of Theorem 3.1.2 to the range $C \leq d \leq cn/\log^2 n$ has recently been accomplished in [LLT$^+$], for some absolute constants $C, c > 0$. The argument in [LLT$^+$] builds on the approach of the present work, and is similar in its use of a shuffling coupling (much like Lemma 3.3.2) and graph discrepancy properties. To take d below the barrier $\log n$ discussed in Remark 3.2.6, they are able to make use of weaker discrepancy properties than the ones employed in the present work. Another notable difference from the present work is that they are able to effectively apply the shuffling coupling with much less control on "structured null vectors" than is provided by Proposition 3.2.4.

3.3 Preliminaries

3.3.1 The shuffling coupling

In this section we formally define the pair $(M, \widetilde{M})$ of rrd matrices described in the previous section, where $\widetilde{M}$ is obtained by re-randomizing the neighborhood $\mathcal{N}_M(\{i_1, i_2\})$ of a pair of distinct vertices $i_1, i_2 \in [n]$ in a certain way. Recall that from the row sums constraint we have

$$\mathrm{ex}_M(i_1, i_2) = d - \mathrm{co}_M(i_1, i_2) = \mathrm{ex}_M(i_2, i_1)$$

for any distinct $i_1, i_2 \in [n]$ (recall from Section 3.1.5 our notation $\mathrm{ex}_M(i_1, i_2) = |\,\mathrm{Ex}_M(i_1, i_2)|$, $\mathrm{co}_M(i_1, i_2) = |\,\mathrm{Co}_M(i_1, i_2)|$). On an intuitive level, the shuffling operation is somewhat similar to performing a "riffle shuffling" of the "deck" $\mathrm{Ex}_M(i_1, i_2) \cup \mathrm{Ex}_M(i_2, i_1)$, then cutting the deck into two equal parts to obtain $\mathrm{Ex}_{\widetilde{M}}(i_1, i_2)$, $\mathrm{Ex}_{\widetilde{M}}(i_2, i_1)$. The set of common neighbors

$\mathrm{Co}_M(i_1, i_2) = \mathcal{N}_M(i_1) \cap \mathcal{N}_M(i_2)$ is preserved by the shuffling.

Definition 3.3.1 (Shuffling)**.** Let $M \in \mathcal{M}_n(d)$ and $i_1, i_2 \in [n]$ distinct. For a bijection

$$\pi : \mathrm{Ex}_M(i_1, i_2) \to \mathrm{Ex}_M(i_2, i_1)$$

and a sequence of signs $\xi : [n] \to \{\pm 1\}$, by *perform a shuffling on M at rows (i_1, i_2) according to π, ξ,* we mean to replace the 2×2 minors $M_{(i_1,i_2)\times(j,\pi(j))}$ with

$$\mathbf{I}_2 \, \mathbb{1}(\xi(j) = +1) + \mathbf{J}_2 \, \mathbb{1}(\xi(j) = -1)$$

for each $j \in \mathrm{Ex}_M(i_1, i_2)$, and to leave all other entries of M unchanged.

The key to applying the shuffling operation in the proof of Theorem 3.1.2 will be to take π and ξ to be random.

Lemma 3.3.2 (Shuffling coupling)**.** *Let M be an rrd matrix, and fix $i_1, i_2 \in [n]$ distinct. Conditional on M, let $\pi : \mathrm{Ex}_M(i_1, i_2) \to \mathrm{Ex}_M(i_2, i_1)$ be a uniform random bijection. Draw a sequence $\xi : [n] \to \{\pm 1\}$ of iid uniform signs, independent of all other variables. Form $\widetilde{M}$ by performing a shuffling on M at rows (i_1, i_2) according to π and ξ. Then $\widetilde{M} \stackrel{d}{=} M$.*

At one part of the proof we will need the following slightly more general version (which implies the above lemma) in which there is a fixed set of "frozen" columns which we cannot modify. For a set A and an integer $0 \le k \le |A|$, we use the notation $\binom{A}{k}$ for the set of subsets of A of size k.

Lemma 3.3.3 (Restricted shuffling)**.** *Let M be an rrd matrix and fix $i_1, i_2 \in [n]$ distinct. Let* $\mathrm{Frozen} \subset [n]$ *be a set of column indices that is fixed by conditioning on the rows $(R_i)_{i \notin \{i_1, i_2\}}$. Set*

$$A_1 = \mathrm{Ex}_M(i_1, i_2) \setminus \mathrm{Frozen}, \quad A_2 = \mathrm{Ex}_M(i_2, i_1) \setminus \mathrm{Frozen} \tag{3.39}$$

and let $s \le \min(|A_1|, |A_2|)$ also be fixed by conditioning on the rows $(R_i)_{i \notin \{i_1, i_2\}}$ (i.e. chosen measurably with respect to the sigma algebra generated by these rows). Conditional on M let

$S_1 \in \binom{A_1}{s}$ and $S_2 \in \binom{A_2}{s}$ be chosen independently and uniformly. Conditional on M, S_1, S_2, let $\pi : S_1 \to S_2$ be a uniform random bijection. Finally, let $\xi : [n] \to \{\pm 1\}$ be a sequence of iid uniform signs, independent of all other variables.

Form $\widetilde{M}$ from M by replacing the 2×2 minors $M_{(i_1,i_2)\times(j,\pi(j))}$ with

$$\mathbf{I}_2 \,\mathbb{1}(\xi(j) = +1) + \mathbf{J}_2 \,\mathbb{1}(\xi(j) = -1)$$

for each $j \in S_1$, leaving all other entries of M unchanged. Then $\widetilde{M} \stackrel{d}{=} M$.

Lemma 3.3.2 follows from Lemma 3.3.3 by taking Frozen to be empty and $s = \mathrm{ex}_M(i_1, i_2)$.

Proof. Condition on the rows $(R_i)_{i \notin \{i_1,i_2\}}$. This fixes Frozen and the set $\mathrm{Ex}_M(i_1, i_2) \cup \mathrm{Ex}_M(i_2, i_1)$. Condition also on the columns of M with indices in Frozen – this fixes $a_1 := |A_1|$ and $a_2 := |A_2|$.

The only remaining randomness of M is in the uniform random partition of

$$A := \big(\mathrm{Ex}_M(i_1, i_2) \cup \mathrm{Ex}_M(i_2, i_1)\big) \setminus \mathrm{Frozen}$$

into the sets A_1, A_2 of prescribed sizes. It hence suffices to show that $\widetilde{A}_1 := \mathrm{Ex}_{\widetilde{M}}(i_1, i_2) \setminus \mathrm{Frozen}$ is also distributed uniformly over $\binom{A}{a_1}$. We may write $\widetilde{A}_1$ as the disjoint union

$$\widetilde{A}_1 = (A_1 \setminus S_1) \sqcup \big(S_1 \cap \xi^{-1}(+1)\big) \sqcup \pi\big(S_1 \cap \xi^{-1}(-1)\big). \tag{3.40}$$

To see that $\widetilde{A}_1 \stackrel{d}{=} A_1$ it is clearer to use the following alternative description of the coupling $(M, \widetilde{M}, S_1, S_2, \pi, \xi)$. Denote

$$E_1 := [a_1] \supset [s] =: F_1$$
$$E_2 := [a_1 + 1, a_1 + a_2] \supset [a_1 + 1, a_1 + s] =: F_2$$

and denote $E = E_1 \cup E_2$, $F = F_1 \cup F_2$. Under the above conditioning, draw bijections

$$\Phi : E \to A, \quad \check{\pi} : F_1 \to F_2 \tag{3.41}$$

independently and uniformly at random, and let $\check{\xi} : [n] \to \{\pm 1\}$ be a sequence of iid uniform signs independent of all other variables. Then

$$(A_1, A_2, S_1, S_2, \pi, \xi) \stackrel{d}{=} \big(\Phi(E_1), \Phi(E_2), \Phi(F_1), \Phi(F_2), \Phi \circ \check{\pi} \circ \Phi^{-1}, \check{\xi} \circ \Phi^{-1}\big).$$

We have shifted the randomness of the sets A_1, A_2, S_1, S_2 to the randomness of the map Φ. We want to show that $\widetilde{A}_1 \stackrel{d}{=} \Phi(E_1)$, where $E_1 = [a_1]$ is now a deterministic set. From (3.40),

$$\begin{aligned}\widetilde{A}_1 &\stackrel{d}{=} \Phi(E_1 \setminus F_1) \sqcup \Phi\big(F_1 \cap \check{\xi}^{-1}(+1)\big) \sqcup (\Phi \circ \check{\pi})\big(F_1 \cap \check{\xi}^{-1}(-1)\big) \\ &= \Phi\Big[(E_1 \setminus F_1) \sqcup \big(F_1 \cap \check{\xi}^{-1}(+1)\big) \sqcup \check{\pi}\big(F_1 \cap \check{\xi}^{-1}(-1)\big)\Big].\end{aligned}$$

Conditioning on $\check{\pi}$ and $\check{\xi}$ (which doesn't affect the distribution of Φ) we have that $\widetilde{A}_1$ is the image under Φ of a fixed set of size a_1, which completes the proof. □

3.3.2 Discrepancy properties

In this section we collect various "good events" concerning the distribution of edges in the random regular digraph Γ associated to M. In all cases, the good event is shown to hold with overwhelming probability, for a suitable range of parameters and assuming $\min(d, n-d) = \omega\big(\log n\big)$ (note this is a wider range of d than is assumed in Theorems 3.1.2 and 3.1.13). This will allow us to restrict to these events without further comment in subsequent stages of the proof. (Indeed, note that we are ultimately aiming for only a polynomially-small bound on the singularity probability, so the failure probabilities for the good events will be negligible.) At the end of the section we prove that Theorem 3.1.12 follows from Theorem 3.1.13 by showing that the event $\mathcal{G}(d)$ from the latter theorem holds with overwhelming probability for M. The results of this section are all corollaries of Theorem 2.1.5 from the previous chapter.

The shuffling coupling from Lemma 3.3.2 will only be useful if the sets $\mathrm{Ex}_M(i_1, i_2)$ are large (see Section 3.2.3). Hence, the following immediate consequence of Theorem 2.1.5 will be essential for our arguments. Recall that $p := d/n$ denotes the average edge density for the digraph.

Theorem 3.3.4 (Concentration of codegrees)**.** *For $\delta \in (0,1)$, let $\mathcal{G}^{\mathrm{ex}}(\delta)$ denote the event that for every pair of distinct $i_1, i_2 \in [n]$ we have*

$$\left|\frac{\mathrm{ex}_M(i_1,i_2)}{p(1-p)n} - 1\right| \le \delta \quad \textit{and} \quad \left|\frac{\mathrm{ex}_{M^\mathsf{T}}(i_1,i_2)}{p(1-p)n} - 1\right| \le \delta. \tag{3.42}$$

Then

$$\mathbb{P}(\mathcal{G}^{\mathrm{ex}}(\delta)) = 1 - n^{O(1)} \exp\left(-c\delta \min\left\{d, n-d, \delta n\right\}\right). \tag{3.43}$$

In particular, for any fixed $\delta \in (0,1)$ independent of n we have that $\mathcal{G}^{\mathrm{ex}}(\delta)$ holds with overwhelming probability if $\min(d, n-d) = \omega(\log n)$.

We will also need the tail estimates on edge counts provided by Theorem 2.1.5, which we restate in terms of the event $\mathcal{G}^{\mathrm{ex}}(\delta)$ as follows:

Theorem 3.3.5 (Concentration of edge counts)**.** *With $\mathcal{G}^{\mathrm{ex}}(\delta)$ as in Theorem 3.3.4, we have that for any $A, B \subset [n]$ and any $\tau \ge 0$,*

$$\mathbb{P}\Big(\Big\{\big|e_M(A,B) - \mu(A,B)\big| \ge \tau\hat{\mu}(A,B)\Big\} \wedge \mathcal{G}^{\mathrm{ex}}(\delta)\Big) \le 2\exp\left(-\frac{c\tau^2}{1+\tau}\hat{\mu}(A,B)\right) \tag{3.44}$$

provided $\delta \le \min\left(\frac{1}{4}, \frac{\tau}{8}\right)$.

The following is an immediate consequence of Corollary 2.1.9 from Chapter 2.

Corollary 3.3.6 (Discrepancy for large minors)**.** *Let $C > 0$ be a sufficiently large absolute constant. For $\varepsilon \in (0,1)$, define the family of pairs of sets*

$$\mathcal{F}(\varepsilon) = \left\{(A,B) : \ A, B \subset [n], \ \min(|A|, |B|) \ge \frac{C}{\varepsilon^2}\frac{\log n}{p}\right\} \tag{3.45}$$

and the event

$$\mathcal{G}^{\text{edge}}(\varepsilon) = \Big\{\forall (A,B) \in \mathcal{F}(\varepsilon),\ \big|e_M(A,B) - \mu(A,B)\big| \le \varepsilon\hat{\mu}(A,B)\Big\}.$$

If $\min(d, n-d) = \omega(\log n)$ *and* $\varepsilon \in (0,1)$ *is fixed independent of* n*, then* $\mathcal{G}^{\text{edge}}(\varepsilon)$ *holds with overwhelming probability.*

Note that for $S, B \subset [n]$ we have the deterministic bound

$$e_M(S,B) \le d|S| \tag{3.46}$$

which is effective when S is small (and we have equality when $B = [n]$). While this bound will be sufficient for many purposes, we will sometimes need a little more when $|B| = o(n)$. Theorem 3.3.5 allows us to improve on (3.46) off a small event:

Corollary 3.3.7 (Discrepancy for thin minors)**.** *For* $\gamma > 0$ *set*

$$s_0(\gamma) := \frac{\log n}{2\gamma}\frac{n}{d} \tag{3.47}$$

and for $\varepsilon_0 \in (0,1]$ *define the family of "thin minors"*

$$\mathcal{F}_{\text{thin}}(\varepsilon_0, \gamma) = \Big\{(S,B) : |S| \le s_0(\gamma),\ |B| \le \frac{\varepsilon_0\gamma}{\log n} d|S|\Big\}. \tag{3.48}$$

Let

$$\mathcal{B}(\varepsilon_0,\gamma) = \Big\{\exists\,(S,B) \in \mathcal{F}_{\text{thin}}(\varepsilon_0,\gamma) :\ \max\big(e_M(S,B), e_M(B,S)\big) \ge \varepsilon_0 d|S|\Big\}. \tag{3.49}$$

There are absolute constants $C, c > 0$ *such that if* $\min(d, n-d) \ge C\varepsilon_0^{-1}\log n$ *and* $\gamma \in (0,c]$*, then*

$$\mathbb{P}(\mathcal{B}(\varepsilon_0,\gamma)) \ll \exp\left(-c\varepsilon_0 d\right). \tag{3.50}$$

In particular, if $\min(d, n-d) = \omega(\log n)$ *and* ε_0 *is fixed independent of* n*, then* $\mathcal{B}(\varepsilon_0,\gamma)^c$ *holds with overwhelming probability.*

Proof. Let $\varepsilon_0 \in (0,1]$ and $\gamma > 0$. Denote

$$b_0(\varepsilon_0, \gamma, s) := \frac{\varepsilon_0 \gamma}{\log n} ds \tag{3.51}$$

and fix $s \leq s_0(\gamma)$, $b \leq b_0(\varepsilon_0, \gamma, s)$. For $S, B \subset [n]$ with $|S| = s$, $|B| = b$ we have

$$\mu(S, B) = \frac{dsb}{n} \leq \frac{b_0 ds}{n} \leq \frac{\varepsilon_0 ds}{2}. \tag{3.52}$$

Let $\delta > 0$ to be chosen. For fixed S, B as above, from Theorem 3.3.5 we have

$$\begin{aligned}\mathbb{P}\Big(\mathcal{G}^{\text{ex}}(\delta) \wedge \Big\{e_M(S,B) \geq \varepsilon_0 ds\Big\}\Big) &\leq \mathbb{P}\Big(\mathcal{G}^{\text{ex}}(\delta) \wedge \Big\{e_M(S,B) - \mu(S,B) \geq \frac{\varepsilon_0}{2} ds\Big\}\Big) \\ &\leq \exp\left(-c\varepsilon_0 ds\right) \end{aligned} \tag{3.53}$$

provided we take $\delta \leq \min\left(\frac{1}{4}, \frac{\varepsilon_0 ds}{16\mu}\right)$. Since $\frac{\varepsilon_0 ds}{16\mu} \geq \frac{1}{8}$ by (3.52), we can take $\delta = \frac{1}{8}$. With this choice of δ we have

$$\mathbb{P}(\mathcal{G}^{\text{ex}}(\delta)^c) \ll n^{O(1)} \exp\left(-cd\right) \tag{3.54}$$

from Theorem 3.3.4. Now by a union bound, (3.53) and the assumed lower bound on d,

$$\begin{aligned}\mathbb{P}(\mathcal{G}^{\text{ex}}(1/8) \wedge \mathcal{B}(\varepsilon_0, \gamma)) &\leq \sum_{s \leq s_0(\gamma)} \binom{n}{s} \sum_{b \leq b_0(\varepsilon_0, \gamma, s)} \binom{n}{b} \exp\left(-c\varepsilon_0 ds\right) \\ &\ll \sum_{s \leq s_0(\gamma)} n^{b_0} \exp\left(s(\log n - c\varepsilon_0 d)\right) \\ &\leq \sum_{s \leq s_0(\gamma)} \exp\left(b_0 \log n - c\varepsilon_0 ds\right) \\ &\leq \sum_{s \leq s_0(\gamma)} \exp\left(-c\varepsilon_0 ds\right) \\ &\ll \exp\left(-c\varepsilon_0 d\right)\end{aligned}$$

where in the fourth line we used the definition (3.51) of b_0 and took γ sufficiently small. Combining with the bound (3.54) and the lower bound on d completes the proof. □

We have the following quick consequence that with high probability, the size of the neighborhood $\mathcal{N}_M(S)$ of any small set S is within a logarithmic factor of the upper bound $d|S|$.

Corollary 3.3.8 (Expansion of small sets)**.** *For $\gamma > 0$, let $\mathcal{G}^{\exp}(\gamma)$ be the event that for every $S \subset [n]$ with $|S| \le s_0(\gamma) = \frac{\log n}{2\gamma}\frac{n}{d}$, we have*

$$\big|\mathcal{N}_M(S)\big| \ge \frac{\gamma}{\log n} d|S|.$$

Assume $d = \omega(\log n)$. Then there is a constant $c_1 > 0$ such that for all $\gamma \in (0, c_1]$,

$$\mathbb{P}(\mathcal{G}^{\exp}(\gamma)) = 1 - O\big(e^{-cd}\big). \tag{3.55}$$

Proof. Let $\gamma > 0$. From the crude lower bound $|\mathcal{N}_M(S)| \ge d$ we have that $\mathcal{G}^{\exp}(\gamma)$ trivially holds if $d \ge n/2$, so assume $\omega(\log n) \le d \le n/2$. On $\mathcal{G}^{\exp}(\gamma)^c$ there exist $S, B \subset [n]$ with $|S| \le s_0(\gamma)$ and

$$|B| < \frac{\gamma}{\log n} d|S| = b_0(1, \gamma, |S|)$$

in the notation of (3.47), (3.51), such that $e_M(S, B) = d|S|$ (simply from taking $B = \mathcal{N}_M(S)$). Hence, $\mathcal{G}^{\exp}(\gamma)^c$ is contained in the event $\mathcal{B}(1, \gamma)$ from Corollary 3.3.7, and the result follows by taking γ sufficiently small. $\square$

Now we can prove that Theorem 3.1.12 follows from Theorem 3.1.13.

Proof. of Theorem 3.1.12. Assume $\omega(\log n) \le d \le n$. Write $M_\pm = M \circ \Xi$ as in (3.12), where M is an rrd matrix and Ξ is matrix of iid uniform signs, independent of M. It suffices to show that the event $\mathcal{G}(d)$ in Theorem 3.1.13 holds with overwhelming probability for $\Sigma = M$.

Conditions (0) and (3) of $\mathcal{G}(d)$ are immediate for M (and hold with probability 1), taking $\kappa_3 = 1$. From Corollary 3.3.8 we have that condition (1) is satisfied with probability $1-O(e^{-cd})$. From Corollary 3.3.6, we have that if $\min(d, n-d) = \omega(\log n)$, then condition (2) holds with overwhelming probability with $c_2 \in (0, 1)$ fixed arbitrarily and $C_2 > 0$ sufficiently large depending only on c_2.

It only remains to show that condition (2) holds with overwhelming probability for the high density case $d = n - O(\log n)$. Let $A, B \subset [n]$ such that $|A|, |B| \geq C_2 \frac{n}{d} \log n$ for some $C_2 > 0$ to be chosen sufficiently large. Let M' denote the complementary rrd matrix with entries $M'(i,j) = 1 - M(i,j)$. Then by $(n-d)$-regularity, $e_{M'}(A,B) \leq (n-d)|A|$. It follows that

$$\begin{aligned} e_M(A,B) &= |A||B| - e_{M'}(A,B) \\ &\geq |A|(|B| - (n-d)) \\ &\gg |A||B| \\ &\geq \frac{d}{n}|A||B| \end{aligned}$$

where in the third line we used the upper bound $n - d = O(\log n)$ and the lower bound on $|B|$, taking C_2 sufficiently large. It follows that for the case $d = n - O(\log n)$, we may take $c_2 \in (0,1)$ sufficiently small such that condition (2) holds with probability 1 for all n sufficiently large. □

3.3.3 Concentration of measure

The following concentration inequality for certain functions on the symmetric group will be useful when working with the bijections π in the shuffling coupling of Lemma 3.3.2, and follows from the $d = 1$ case of Theorem 2.1.13, or alternatively from Proposition 1.1 in [Cha07].

Lemma 3.3.9 (Concentration for the symmetric group)**.** *For $m \geq 1$, $\pi \in \mathrm{Sym}(m)$ a permutation on $[m]$, and $A, B \subset [m]$, denote*

$$e_\pi(A,B) = \big|\{i \in A : \pi(i) \in B\}\big|.$$

If π is a uniform random element of $\mathrm{Sym}(m)$*, we have that for any $\tau \geq 0$,*

$$\mathbb{P}\left\{\left|e_\pi(A,B) - |A||B|/m\right| \geq \tau|A||B|/m\right\} \leq 2\exp\left\{-\frac{c\tau^2}{1+\tau}\frac{|A||B|}{m}\right\}. \tag{3.56}$$

Remark 3.3.10. Note that the above lemma is essentially the $d = 1$ case of Theorem 3.3.5, the only difference (apart from constants in the exponential) being that we do not need to restrict to any "good event" like $\mathcal{G}^{\mathrm{ex}}(\delta)$.

3.4 Unstructured null vectors

In this section we prove Theorems 3.1.2 and 3.1.13, taking as black boxes Propositions 3.2.4 and 3.2.3 ruling out structured null vectors. These propositions are proved in Section 3.5. We remark that the proof of Theorem 3.1.13 is not needed for the proof of Theorem 3.1.2, so the reader who is only interested in the proof of the main theorem can begin at Section 3.4.2.

3.4.1 Warmup: Proof of Theorem 3.1.13

We restrict the sample space to the event $\mathcal{G}(d)$ defined in Theorem 3.1.13. For convenience, we let $\mathcal{G}_i(d)$ denote the event that condition i of $\mathcal{G}(d)$ holds for Σ, so that $\mathcal{G}(d) = \bigwedge_{i=0}^3 \mathcal{G}_i(d)$. In this section we will only use the "minimum degree" and "no thin dense minors" properties enjoyed on $\mathcal{G}_0(d) \wedge \mathcal{G}_3(d)$. We denote the rows of Σ by r_i and the rows of Ξ by Y_i, so that the ith row of H is $R_i = r_i \circ Y_i$. Our aim is to control the event

$$\mathcal{R}_1 = \{\mathrm{corank}(H) \geq 1\}. \tag{3.57}$$

The following lemma reduces this task to bounding the event that a randomly sampled row lands in the span of the remaining rows. We will extend this to larger corank with Lemma 3.4.3. Recall that a vector $x \in \mathbb{R}^n$ is k-sparse if $|\operatorname{spt}(x)| \leq k$.

Lemma 3.4.1. *Let H be a random $n \times n$ matrix with rows R_i, $1 \leq i \leq n$. For $\eta \in (0,1)$,* *let $\mathcal{G}_L(\eta)$ be the event that H has no non-trivial $(1-\eta)n$-sparse left null vectors. For $i \in [n]$ denote $V_i := \mathrm{span}(R_j : j \neq i)$, and define the events*

$$\mathcal{S}_i := \big\{ R_i \in V_i \big\}. \tag{3.58}$$

Draw I uniformly from $[n]$, independently of H. Then with $\mathcal{R}_1$ as in (3.57) we have

$$\mathbb{P}\left(\mathcal{R}_1 \wedge \mathcal{G}_L(\eta) \right) \le \frac{1}{1-\eta} \mathbb{P}\left(\mathcal{S}_I \wedge \mathcal{G}_L(\eta) \right). \tag{3.59}$$

Proof. On $\mathcal{R}_1 \wedge \mathcal{G}_L(\eta)$, H has a left null vector with support of size at least $(1-\eta)n$. It follows that on this event, $\mathcal{S}_i$ holds for at least $(1-\eta)n$ values of $i \in [n]$. By double counting,

$$(1-\eta)n\, \mathbb{P}\left(\mathcal{R}_1 \wedge \mathcal{G}_L(\eta)\right) \le \sum_{i=1}^{n} \mathbb{P}\left(\mathcal{S}_i \wedge \mathcal{G}_L(\eta)\right) \tag{3.60}$$

and the result follows by rearranging. □

From Proposition 3.2.3 we have that $\mathcal{G}^{\mathrm{sp}}_{\pm}(\eta)$ holds with probability $1 - O(n^{-100})$ for $H = \Sigma \circ \Xi$ for any $\eta \in [C_1' d^{-1/4}, 1]$, where $C_1' > 0$ is a sufficiently large absolute constant. Since $\mathcal{G}^{\mathrm{sp}}_{\pm}(\eta)$ is simply the event that $\mathcal{G}_L(\eta)$ holds for H and H^{T}, by the above lemma it suffices to show

$$\mathbb{P}\left(\mathcal{S}_I \wedge \mathcal{G}^{\mathrm{sp}}_{\pm}(\eta)\right) \ll \kappa_3 \eta + d^{-1/2} \tag{3.61}$$

for arbitrary $\eta \in [C_1' d^{-1/4}, 0.1]$ (say), where I is drawn uniformly of $[n]$, independently of Σ and Ξ. From now on we restrict the sample space to $\mathcal{G}^{\mathrm{sp}}_{\pm}(\eta)$ for η in this range, in order to lighten the notation.

Draw u uniformly from the unit sphere in $V_I^{\perp}$, in a way such that u, r_I and Y_I are jointly independent conditional on I and the remaining rows of Σ and Ξ. Now it would be enough to show

$$\mathbb{P}\left\{ (r_I \circ Y_I) \cdot u = 0 \right\} \ll \kappa_3 \eta + d^{-1/2}. \tag{3.62}$$

From Theorem 3.2.1 we have

$$\mathbb{P}_{Y_I}\big((r_I \circ Y_I) \cdot u = 0 \,\big|\, \Sigma, I\big) \ll |\mathrm{spt}(u) \cap \mathrm{spt}(r_I)|^{-1/2} \tag{3.63}$$

so we need to argue that $\mathrm{spt}(u)$ and $\mathrm{spt}(r_I) = \mathcal{N}_\Sigma(I)$ have large overlap.

By our restriction to $\mathcal{G}_0(d)$ we have $|\mathcal{N}_\Sigma(i)| \geq d$ for all $i \in [n]$. We identify the set of undesirable realizations of u as

$$\mathcal{H}_{\Sigma,I} := \left\{ x \in \mathbb{R}^n : \,\big|\mathrm{spt}(x) \cap \mathcal{N}_\Sigma(I)\big| \leq \frac{d}{2} \right\} \tag{3.64}$$

and define the bad event

$$\mathcal{B} = \big\{\mathbb{P}_u(u \in \mathcal{H}_{\Sigma,I}) > 0\big\}. \tag{3.65}$$

We note that $\mathcal{B}$ is decided by the randomness of Σ, I, and the rows $(Y_i)_{i \neq I}$ of Ξ.

First we bound $\mathbb{P}(\mathcal{S}_I \wedge \mathcal{B})$ using the randomness of I and our restriction to $\mathcal{G}_3(d) \wedge \mathcal{G}_\pm^{\mathrm{sp}}(\eta)$. The crucial observation is that $\mathcal{H}_{\Sigma,I}$ is a finite union of subspaces, each of co-dimension at least $d/2$:

$$\mathcal{H}_{\Sigma,I} = \bigcup_{B \subset \mathcal{N}_\Sigma(I):\, |B| \leq \frac{d}{2}} \mathbb{R}^{([n] \setminus \mathcal{N}_\Sigma(I)) \cup B}.$$

Since we picked u according to the surface measure on the unit sphere of $V_I^\perp$, it follows that on $\mathcal{B}$ we actually have $V_I^\perp \subset \mathcal{H}_{\Sigma,I}$. On $\mathcal{R}_1$ we may pick a nontrivial vector $x \in \ker(H)$ (note that the kernel is nontrivial on this event). Crucially, we may do this with x independent of I. We have

$$x \in \ker(H) \subset V_I^\perp$$

and so on $\mathcal{B}$ we have

$$x \in \mathcal{H}_{\Sigma,I}. \tag{3.66}$$

Summarizing our progress so far,

$$\mathbb{P}(\mathcal{S}_I \wedge \mathcal{B}) \leq \mathbb{P}\big(\mathcal{R}_1 \wedge \{x \in \mathcal{H}_{\Sigma,I}\}\big) \tag{3.67}$$

where we used that $\mathcal{S}_I \subset \mathcal{R}_1$. Letting

$$S_\Sigma(x) = \left\{ i \in [n] : \big|\mathcal{N}_\Sigma(i) \cap \operatorname{spt}(x)\big| \le \frac{d}{2} \right\} \tag{3.68}$$

we have

$$\begin{aligned} |S_\Sigma(x)| \frac{d}{2} &< \sum_{i \in S_\Sigma(x)} \big|\mathcal{N}_\Sigma(i) \cap x^{-1}(0)\big| \\ &= e_\Sigma\big(S_\Sigma(x), x^{-1}(0)\big) \\ &\le \kappa_3 d |x^{-1}(0)| \end{aligned}$$

where in the last line we applied our restriction to $\mathcal{G}_3(d)$. By our restriction to $\mathcal{G}_\pm^{\mathrm{sp}}(\eta)$ we conclude

$$|S_\Sigma(x)| \le 2\kappa_3 |x^{-1}(0)| \le 2\kappa_3 \eta n. \tag{3.69}$$

It follows that conditional on Σ and Ξ such that $\mathcal{R}_1$ holds,

$$\mathbb{P}_I(x \in \mathcal{H}_{\Sigma,I}) = \mathbb{P}_I(I \in S_\Sigma(x)) \le 2\kappa_3 \eta$$

and so we conclude from (3.67) that

$$\mathbb{P}(\mathcal{S}_I \wedge \mathcal{B}) \le 2\kappa_3 \eta. \tag{3.70}$$

It remains to bound $\mathbb{P}(\mathcal{S}_I \wedge \mathcal{B}^c)$. Condition on I, Σ and $(Y_i)_{i \neq I}$ such that $\mathcal{B}$ does not hold. Off a null event we may assume that $u \notin \mathcal{H}_{\Sigma,I}$. Now since

$$\big|\operatorname{spt}(u) \cap \mathcal{N}_\Sigma(I)\big| > \frac{d}{2}$$

applying (3.63) we have

$$\begin{aligned}\mathbb{P}\left(\mathcal{S}_I \wedge \mathcal{B}^c\right) &\le \mathbb{P}\big(\big\{(r_I \circ Y_I)\cdot u = 0\big\} \wedge \big\{u \notin \mathcal{H}_{\Sigma,I}\big\}\big)\\ &= \mathbb{E}\,\mathbb{P}_{Y_I}\left((r_I \circ Y_I)\cdot u = 0 \,\middle|\, \Sigma, I, (Y_i)_{i\neq I}\right)\mathbb{1}(u \notin \mathcal{H}_{\Sigma,I})\\ &\ll d^{-1/2}\end{aligned}$$

which combines with (3.70) to give the claim. □

3.4.2 Preliminary reductions

Now we turn to the proof of Theorem 3.1.2. We may assume

$$C_0 \log^2 n \le d \le n/2 \tag{3.71}$$

(for the upper bound see Remark 3.1.5). This will allow us to restrict to the following "good events":

- By Theorem 3.3.4, the event $\mathcal{G}^{\mathrm{ex}}(\delta)$ holds with overwhelming probability for any fixed $\delta \in (0,1)$ independent of n (here we only need $d = \omega(\log n)$).
- From Proposition 3.2.4 (which we are assuming for now, deferring the prove to Section 3.5) (3.71) implies that $\mathcal{G}^{\mathrm{sls}}(\eta)$ holds with probability $1 - O(n^{-100})$ for any $\eta \in [C_1 d^{-c_0}, 1]$. Moreover, we note that $\mathcal{G}^{\mathrm{sls}}(\eta) \subset \mathcal{G}^{\mathrm{sp}}(\eta)$, the event that M has no left or right $(1-\eta)n$-sparse null vectors.

We leave the parameters $\delta, \eta \in (0,1)$ unspecified for now.

For $k \in [n]$ define the event

$$\mathcal{R}_k := \{\mathrm{corank}(M) \ge k\}\,. \tag{3.72}$$

Our aim is to bound $\mathbb{P}(\mathcal{R}_1)$. Unlike the proof for H in the previous section, we will need to separately handle $\mathcal{R}_2$ and $\mathcal{R}_1 \setminus \mathcal{R}_2$ by different arguments. The argument for $\mathcal{R}_2$ will follow a similar approach to the proof of Theorem 3.1.13, after invoking the shuffling coupling to inject iid signs. Controlling $\mathcal{R}_1 \setminus \mathcal{R}_2$ will require more care. Theorem 3.1.2 follows from the next proposition and Proposition 3.2.4.

Proposition 3.4.2. *For all $\eta \in (0, 1]$ we have*

$$\mathbb{P}\big(\mathcal{R}_2 \wedge \mathcal{G}^{\mathrm{sls}}(\eta)\big) \ll \eta + d^{-1/2} \tag{3.73}$$

and

$$\mathbb{P}\big(\mathcal{R}_1 \wedge \mathcal{R}_2^c \wedge \mathcal{G}^{\mathrm{sls}}(\eta)\big) \ll \eta + d^{-1/2}. \tag{3.74}$$

We will use the following extension of Lemma 3.4.1 for controlling the event $\mathcal{R}_k$ when there are no sparse null vectors. We only need this for $k = 2$, but the result for larger values of k comes with little additional effort.

Lemma 3.4.3 (Control by random sampling)**.** *Assume M is a random $n \times n$ matrix with rows R_i, $1 \le i \le n$. Let $\mathcal{R}_k$ be as in* (3.72) *and for $\eta \in (0, 1)$ let $\mathcal{G}_L(\eta)$ be the event that M has no non-trivial $(1-\eta)n$-sparse left null vectors. For an arbitrary k-tuple of row indices $(i_1, \dots, i_k) \in [n]^k$, denote the subspaces*

$$V_{(i_1,\dots,i_k)} = \operatorname{span}\left(R_i : i \notin \{i_l\}_{l=1}^k\right) \tag{3.75}$$

and the events

$$\mathcal{S}_{(i_1,\dots,i_k)} = \left\{R_{i_1}, \dots, R_{i_k} \in V_{(i_1,\dots,i_k)}\right\}.$$

For $k \in [n/2]$, let $\mathcal{I} = (I_1, \dots, I_k)$ be a vector of indices sampled uniformly without replacement from $[n]$, independently of M. Then if $\eta \in \left(0, \frac{1}{2}\right)$, we have

$$\mathbb{P}\left(\mathcal{R}_k \wedge \mathcal{G}_L(\eta)\right) \le (1-2\eta)^{-k}\, \mathbb{P}\left(\mathcal{S}_{\mathcal{I}} \wedge \mathcal{G}_L(\eta)\right).$$

Proof. Since

$$\mathbb{P}\left(\mathcal{S}_{\mathcal{I}} \wedge \mathcal{G}_L(\eta)\right) = \mathbb{P}\left(\mathcal{S}_{\mathcal{I}} \,\middle|\, \mathcal{R}_k \wedge \mathcal{G}_L(\eta)\right) \mathbb{P}\left(\mathcal{R}_k \wedge \mathcal{G}_L(\eta)\right)$$

it suffices to show that for fixed M such that $\mathcal{R}_k \wedge \mathcal{G}_L(\eta)$ holds, we have

$$\mathbb{P}_{\mathcal{I}}\left(\mathcal{S}_{\mathcal{I}}\right) \geq (1-2\eta)^{-k}. \tag{3.76}$$

Condition on such M. We may pick k linearly independent left null vectors $y_1, \ldots, y_k$, so that for each $j \in [k]$,

$$\sum_{i \in [n]} y_j(i) R_i = 0.$$

Next we apply row reduction to the $k \times n$ matrix with rows y_j. For $\mathcal{S}_{\mathcal{I}}$ to hold, it suffices that there exist $Z = (z_1, \ldots, z_k) \in \mathbb{R}^{n \times k}$ with $\mathrm{span}(z_1, \ldots, z_k) = \mathrm{span}(y_1, \ldots, y_k)$ such that the $k \times k$ matrix

$$Z_{\mathcal{I} \times [k]} = (z_j(I_l))_{1 \leq j, l, \leq k} \tag{3.77}$$

is upper triangular with nonzero diagonal entries. Indeed, this implies that $R_{I_1}, \ldots, R_{I_k}$ can each be expressed as linear combinations of the rows $\{R_i : i \notin \{I_1, \ldots, I_k\}\}$.

Set $z_1 = y_1$, and let $\mathcal{B}_1 = \{z_1(I_1) = 0\}$. By our restriction to $\mathcal{G}_L(\eta)$ we have

$$\mathbb{P}_{I_1}(\mathcal{B}_1) \leq \frac{|z_1^{-1}(0)|}{n} \leq \eta.$$

For $j \in [k-1]$, having defined linearly independent vectors $z_1, \ldots, z_j$ and events $\mathcal{B}_1, \ldots, \mathcal{B}_j$, on $\bigwedge_{1 \leq l \leq j} \mathcal{B}_l^c$ we can find $z_{j+1} \in \mathrm{span}(y_{j+1}, z_j, \ldots, z_1)$ such that $z_{j+1}(I_l) = 0$ for all $1 \leq l \leq j$ (by linear independence). Let $\mathcal{B}_{j+1} = \{z_{j+1}(I_{j+1}) = 0\}$. Since $z_{j+1} \in \ker(M^\mathsf{T})$, by our restriction to $\mathcal{G}_L(\eta)$ we have

$$\mathbb{P}_{I_{j+1}}\left(\mathcal{B}_{j+1} \Big| \bigwedge_{1 \leq l \leq j} \mathcal{B}_l^c\right) \leq \frac{|z_{j+1}^{-1}(0)|}{n-j} \leq 2\eta$$

(using the upper bound on k). Applying the above bound iteratively with Bayes' rule we conclude that $\bigwedge_{1 \leq l \leq k} \mathcal{B}_l^c$ holds with probability at least $(1-2\eta)^k$ in the randomness of $\mathcal{I}$, and on this event the matrix (3.77) has the desired properties. $\square$

3.4.3 Injecting a random walk

We now turn to Proposition 3.4.2. Without the randomness of the independent signs enjoyed by H, we must use the shuffling coupling of Lemma 3.3.2 to express $\mathcal{S}_{\mathcal{I}}$ as the event that a random walk lands at a particular point. We define a coupled pair of rrd matrices $(M, \widetilde{M})$ as in that lemma, but with the pair of rows selected randomly. That is, we draw:

1. an rrd matrix M,
2. $I_1, I_2 \in [n]$ sampled uniformly without replacement from $[n]$, independently of M,
3. a uniform random bijection

$$\pi : \mathrm{Ex}_M(I_1, I_2) \to \mathrm{Ex}_M(I_2, I_1),$$

4. a sequence $\xi : [n] \to \{\pm 1\}$ of iid uniform signs independent of all other variables.

We form $\widetilde{M}$ by performing a shuffling on M at the rows (I_1, I_2) with respect to π, ξ. By Lemma 3.3.2 and conditioning on I_1, I_2 we have that $M \overset{d}{=} \widetilde{M}$.

Now we wish to control the events $\mathcal{S}_{I_1} = \{R_{I_1} \in V_{I_1}\}$ and $\mathcal{S}_{(I_1,I_2)} = \left\{R_{I_1}, R_{I_2} \in V_{(I_1,I_2)}\right\}$. Note that on $\mathcal{S}_{(I_1,I_2)}$, R_{I_1} and R_{I_2} are orthogonal to any vector in the orthocomplement of $V_{(I_1,I_2)}$. We are hence interested in the dot products $R_{I_1} \cdot u$, $R_{I_2} \cdot u$ for u taken from the unit sphere (say) of $V_{(I_1,I_2)}^{\perp}$. Let us examine the joint distribution of these dot products when we replace M by $\widetilde{M}$. Letting $\widetilde{R}_i$ denote the ith row of $\widetilde{M}$, we can express

$$
\begin{aligned}
\begin{pmatrix} \widetilde{R}_{I_1} \cdot u \\ \widetilde{R}_{I_2} \cdot u \end{pmatrix} &= \begin{pmatrix} \sum_{j \in \mathcal{N}_{\widetilde{M}}(I_1)} u(j) \\ \sum_{j \in \mathcal{N}_{\widetilde{M}}(I_2)} u(j) \end{pmatrix} \\
&= \begin{pmatrix} \sum_{j \in \mathrm{Co}_M(I_1,I_2)} u(j) + \sum_{j \in \mathrm{Ex}_{\widetilde{M}}(I_1,I_2)} u(j) \\ \sum_{j \in \mathrm{Co}_M(I_1,I_2)} u(j) + \sum_{j \in \mathrm{Ex}_{\widetilde{M}}(I_2,I_1)} u(j) \end{pmatrix} \\
&= \sum_{j \in \mathrm{Co}_M(I_1,I_2)} u(j) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \sum_{j \in \mathrm{Ex}_M(I_1,I_2)} \begin{pmatrix} u\big(j\, \mathbb{1}(\xi(j) = +1) + \pi(j)\, \mathbb{1}(\xi(j) = -1)\big) \\ u\big(\pi(j)\, \mathbb{1}(\xi(j) = +1) + j\, \mathbb{1}(\xi(j) = -1)\big) \end{pmatrix} \\
&= \sum_{j \in \mathrm{Co}_M(I_1,I_2)} u(j) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \sum_{j \in \mathrm{Ex}_M(I_1,I_2)} \frac{u(j) + u(\pi(j))}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \xi(j) \frac{u(j) - u(\pi(j))}{2} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
&= \left(\frac{1}{2} (R_{I_1} + R_{I_2}) \cdot u \right) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \left(\sum_{j \in \mathrm{Ex}_M(I_1,I_2)} \xi(j) \partial_j^\pi(u) \right) \begin{pmatrix} 1 \\ -1 \end{pmatrix} && (3.78) \\
&=: A(u) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + W(u) \begin{pmatrix} 1 \\ -1 \end{pmatrix} && (3.79)
\end{aligned}
$$

where in the penultimate line we have defined

$$\partial_j^\pi(u) = \frac{u(j) - u(\pi(j))}{2}. \tag{3.80}$$

Note that the term $A(u)$ is fixed by conditioning on M, I_1, I_2. Furthermore, the sequence $(\partial_j^\pi(u))_{j \in \mathrm{Ex}_M(I_1,I_2)}$ is fixed by additionally conditioning on π. Hence, conditional on M, I_1, I_2, π, in the randomness of the $\xi(j)$ this pair of dot products is a random walk in the $(1, -1)$ direction with steps $\partial_j^\pi(u)$.

The following lemma isolates the role of the randomness of the signs $\xi(j)$ and reduces the problem to the study of structural properties of the normal vector u. While it is stated for an arbitrary fixed pair of row indices (i_1, i_2), it can be applied to the random pair (I_1, I_2) after conditioning.

Lemma 3.4.4 (The role of the signs $\xi(j)$)**.** *For $u \in \mathbb{R}^n$, $i_1, i_2 \in [n]$ distinct, and a bijection*

$\pi : \mathrm{Ex}_M(i_1, i_2) \to \mathrm{Ex}_M(i_2, i_1)$, define

$$\mathrm{Steps}(u) = \mathrm{Steps}_{M,\pi}^{(i_1,i_2)}(u) := \{j \in \mathrm{Ex}_M(i_1, i_2) : u(j) \neq u(\pi(j))\}\,. \tag{3.81}$$

Then with $(M, \widetilde{M})$ coupled as in Lemma 3.3.2 and u deterministic or random depending only on $(R_i : i \notin \{i_1, i_2\})$ we have

$$\mathbb{P}_\xi \left(\widetilde{R}_{i_1} \cdot u = 0\right) \ll |\,\mathrm{Steps}(u)|^{-1/2}. \tag{3.82}$$

Proof. From the representation (3.78) we have

$$\widetilde{R}_{i_1} \cdot u = A(u) + \sum_{j \in \mathrm{Ex}_M(i_1,i_2)} \xi(j)\partial_j^\pi(u)$$

and the claim follows by conditioning on M, π and applying Theorem 3.2.1. □

3.4.4 Ruling out corank ≥ 2

In this section we establish the bound (3.73) from Proposition 3.4.2. By increasing the hidden constant in (3.73) we may assume η is at most a sufficiently small absolute constant.

From Lemma 3.4.3, for η sufficiently small it suffices to bound $\mathbb{P}(\mathcal{S}_{(I_1,I_2)})$. Conditional on M, I_1, I_2, let u be drawn from the uniform surface measure of the unit sphere in $V_{(I_1,I_2)}^\perp$, independently of π, R_{I_1}, R_{I_2}. We have

$$\begin{aligned}\mathbb{P}(\mathcal{S}_{(I_1,I_2)}) &\leq \mathbb{P}\big(\widetilde{R}_{I_1} \in V_{(I_1,I_2)}\big)\\ &\leq \mathbb{P}(\widetilde{R}_{I_1} \cdot u = 0).\end{aligned}$$

We want to bound this using Lemma 3.4.4, so we will need to argue that the set $\mathrm{Steps}_{M,\pi}^{(I_1,I_2)}(u)$ defined there is large. For this task we use the randomness of u, I_1, I_2, π, and restrict M to the good events $\mathcal{G}^{\mathrm{ex}}(\delta)$ and $\mathcal{G}^{\mathrm{sls}}(\eta)$.

First we identify the set of undesirable realizations of u. Let

$$\mathcal{H}'_{M,I_1,I_2} = \left\{ x \in \mathbb{R}^n : \exists \lambda \in \mathbb{R} \text{ with } \min_{l=1,2} |\mathcal{N}_M(I_l) \cap x^{-1}(\lambda)| > d/100 \right\}. \tag{3.83}$$

That is, $\mathcal{H}'_{M,I_1,I_2}$ is the set of vectors with a level set intersecting at least 1% of the support of both R_{I_1} and R_{I_2}. Note that $\mathcal{H}'_{M,I_1,I_2}$ is a finite union of proper subspaces of $\mathbb{R}^n$. Indeed, we may express

$$\mathcal{H}'_{M,I_1,I_2} = \bigcup_{(T_1,T_2)} \mathcal{H}_{T_1 \cup T_2}$$

where the union ranges over pairs of subsets $T_1 \subset \mathcal{N}_M(I_1), T_2 \subset \mathcal{N}_M(I_2)$ of size at least $d/100$, and $\mathcal{H}_T$ denotes the subspace of vectors that are constant on T. Define the bad event

$$\mathcal{B}' = \left\{ \mathbb{P}_u(u \in \mathcal{H}'_{M,I_1,I_2}) > 0 \right\}. \tag{3.84}$$

Since $\mathcal{H}'_{M,I_1,I_2}$ is a finite union of proper subspaces, and u is drawn from the uniform surface measure of the subspaces $V^{\perp}_{(I_1,I_2)}$, it follows that if $\mathcal{B}'$ holds then we actually have the inclusion

$$V^{\perp}_{(I_1,I_2)} \subset \mathcal{H}'_{M,I_1,I_2}.$$

Note also that $\ker(M) \subset V^{\perp}_{(I_1,I_2)}$. On $\mathcal{R}_2$, we may fix an arbitrary nontrivial element $x \in \ker(M)$ (note that the kernel is nonempty on this event), independent of I_1, I_2. Now we have

$$\mathbb{P}(\mathcal{R}_2 \wedge \mathcal{B}') \leq \mathbb{P}(\mathcal{R}_2 \wedge \left\{ x \in \mathcal{H}'_{M,I_1,I_2} \right\}).$$

We will bound the latter quantity using the randomness of I_1, I_2. For $\lambda \in \mathbb{R}$, let

$$S_M(x,\lambda) = \left\{ i \in [n] : |\mathcal{N}_M(i) \cap x^{-1}(\lambda)| \geq d/100 \right\}.$$

We can control the size of these sets using only a crude bound on edge counts:

$$|S_M(x,\lambda)| \frac{d}{100} < e_M(S_M(x,\lambda), x^{-1}(\lambda)) \leq d|x^{-1}(\lambda)|$$

whence

$$|S_M(x,\lambda)| \le 100|x^{-1}(\lambda)|.$$

Now for M such that $\mathcal{G}^{\mathrm{sls}}(\eta)$ holds we have $|x^{-1}(\lambda)| \le \eta n$ for all $\lambda \in \mathbb{R}$. Conditional on M such that $\mathcal{R}_2$ and $\mathcal{G}^{\mathrm{sls}}(\eta)$ hold (which fixes $x \neq 0$), we can bound

$$\begin{aligned}
\mathbb{P}_{I_1,I_2}\left(x \in \mathcal{H}'_{M,I_1,I_2}\right) &= \mathbb{P}_{I_1,I_2}\left(\exists \lambda \in \mathbb{R} : I_1, I_2 \in S_M(x,\lambda)\right) \\
&\le \sum_{\lambda : x^{-1}(\lambda)\neq \phi} \mathbb{P}_{I_1,I_2}(I_1, I_2 \in S_M(x,\lambda)) \\
&\ll \sum_{\lambda : x^{-1}(\lambda)\neq \phi} \left(\frac{|x^{-1}(\lambda)|}{n}\right)^2 \\
&\le \eta \sum_{\lambda : x^{-1}(\lambda)\neq \phi} \frac{|x^{-1}(\lambda)|}{n} \\
&\le \eta.
\end{aligned}$$

Undoing the conditioning on M, we have shown that

$$\mathbb{P}(\mathcal{B}' \wedge \mathcal{R}_2 \wedge \mathcal{G}^{\mathrm{sls}}(\eta)) \ll \eta. \tag{3.85}$$

It remains to bound $\mathbb{P}\big(\{R_{I_1} \cdot u = 0\} \setminus \mathcal{B}'\big)$. Condition on M, I_1, I_2 such that $\mathcal{B}'$ does not hold. Off a null event we may assume that $u \notin \mathcal{H}'_{M,I_1,I_2}$. That is, for every $\lambda \in \mathbb{R}$, we may assume

$$|\mathcal{N}_M(I_1) \cap u^{-1}(\lambda)| \le d/100 \quad \text{or} \quad |\mathcal{N}_M(I_2) \cap u^{-1}(\lambda)| \le d/100. \tag{3.86}$$

We will now get a lower bound on $|\operatorname{Steps}(u)|$ (as defined in (3.81)). It will be more convenient to work with the complementary set

$$\begin{aligned}
\operatorname{Flats}(u) &:= \operatorname{Ex}_M(I_1,I_2) \setminus \operatorname{Steps}(u) \qquad (3.87)\\
&= \{j \in \operatorname{Ex}_M(I_1,I_2) : u(j) = u(\pi(j))\}.
\end{aligned}$$

We have

$$\begin{aligned}
\mathbb{E}_\pi |\operatorname{Flats}(u)| &= \sum_{j\in \operatorname{Ex}_M(I_1,I_2)} \mathbb{P}_\pi(\pi(j)\in u^{-1}(u(j))) \\
&= \sum_{j\in \operatorname{Ex}_M(I_1,I_2)} \frac{|u^{-1}(u(j))\cap \operatorname{Ex}_M(I_2,I_1)|}{|\operatorname{Ex}_M(I_2,I_1)|} \\
&= \frac{1}{|\operatorname{Ex}_M(I_1,I_2)|}\sum_{\lambda: u^{-1}(\lambda)\neq\phi} |u^{-1}(\lambda)\cap \operatorname{Ex}_M(I_1,I_2)||u^{-1}(\lambda)\cap \operatorname{Ex}_M(I_2,I_1)|
\end{aligned}$$

where the last line follows from double counting. Now we apply (3.86) to get

$$\begin{aligned}
\mathbb{E}_\pi |\operatorname{Flats}(u)| &\le \frac{d}{100}\frac{1}{|\operatorname{Ex}_M(I_1,I_2)|}\sum_{\lambda: u^{-1}(\lambda)\neq\phi} \max\left(|u^{-1}(\lambda)\cap \operatorname{Ex}_M(I_1,I_2)|, |u^{-1}(\lambda)\cap \operatorname{Ex}_M(I_2,I_1)|\right) \\
&\le \frac{d}{100}\frac{1}{|\operatorname{Ex}_M(I_1,I_2)|}\sum_{\lambda: u^{-1}(\lambda)\neq\phi} \left|u^{-1}(\lambda)\cap [\operatorname{Ex}_M(I_1,I_2)\cup \operatorname{Ex}_M(I_2,I_1)]\right| \\
&= \frac{d}{50}. \qquad (3.88)
\end{aligned}$$

We want to show that $|\operatorname{Flats}(u)|$ is concentrated around its expectation (we only need control on the upper tail). In the notation of Lemma 3.3.9 we have $|\operatorname{Flats}(u)| = e_\pi(A,B)$ with $A = \operatorname{Ex}_M(I_1,I_2)$ and $B = u^{-1}(u(j_1))$ (which are fixed by conditioning on M, I_1, I_2). Applying Lemma 3.3.9 and (3.88) we conclude that for any $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{P}_\pi\left(|\operatorname{Flats}(u)| \ge (1+\varepsilon)\frac{d}{50}\right) &\le \mathbb{P}_\pi\left(|\operatorname{Flats}(u)| \ge \mathbb{E}_\pi|\operatorname{Flats}(u)| + \varepsilon\frac{d}{50}\right) \\
&\ll \exp\left(-\frac{c\varepsilon^2 d^2/50^2}{\mathbb{E}_\pi|\operatorname{Flats}(u)| + \varepsilon\frac{d}{50}}\right) \\
&\le \exp\left(-\frac{c\varepsilon^2}{1+\varepsilon}d\right). \qquad (3.89)
\end{aligned}$$

On the other hand, on $\mathcal{G}^{\mathrm{ex}}(\delta)$ we have (applying our assumption $d \le n/2$)

$$|\operatorname{Ex}_M(I_1,I_2)| \ge (1-\delta)d\left(1-\frac{d}{n}\right) \ge \frac{1-\delta}{2}d$$

so that fixing $\delta \le 1/2$ and $\varepsilon \le 4$ (say), we conclude that on $\mathcal{B}'^c \wedge \mathcal{G}^{\mathrm{ex}}(\delta)$, except with probability at most $O(\exp(-cd))$ we have

$$|\,\mathrm{Steps}(u)| = |\,\mathrm{Ex}_M(I_1, I_2)| - |\,\mathrm{Flats}(u)| \geq \frac{d}{10}. \tag{3.90}$$

Summarizing our work so far,

$$\begin{aligned}
\mathbb{P}\left[\mathcal{S}_{(I_1,I_2)} \wedge \mathcal{G}^{\mathrm{sls}}(\eta) \wedge \mathcal{G}^{\mathrm{ex}}(\delta)\right] &\leq \mathbb{P}\left[\mathcal{R}_2 \wedge \mathcal{B}' \wedge \mathcal{G}^{\mathrm{sls}}(\eta)\right] + \mathbb{P}\left[\mathcal{S}_{(I_1,I_2)} \wedge \mathcal{B}'^c \wedge \mathcal{G}^{\mathrm{sls}}(\eta) \wedge \mathcal{G}^{\mathrm{ex}}(\delta)\right] \\
&\leq \mathbb{P}\left[\mathcal{R}_2 \wedge \mathcal{B}' \wedge \mathcal{G}^{\mathrm{sls}}(\eta)\right] + \mathbb{P}\left[\mathcal{B}'^c \wedge \left\{|\,\mathrm{Steps}(u)| < \frac{d}{10}\right\} \wedge \mathcal{G}^{\mathrm{ex}}(\delta)\right] \\
&\quad + \mathbb{P}\left[\left\{\widetilde{R}_{I_1} \cdot u = 0\right\} \wedge \left\{|\,\mathrm{Steps}(u)| \geq \frac{d}{10}\right\}\right] \\
&\ll \eta + e^{-cd} + d^{-1/2}
\end{aligned}$$

where in the last line we substituted our bounds (3.90) and (3.85) and applied Lemma 3.4.4. Combined with our estimates for $\mathcal{G}^{\mathrm{sls}}(\eta)$ and $\mathcal{G}^{\mathrm{ex}}(\delta)$, together with Lemma 3.4.3 we have

$$\mathbb{P}(\mathcal{R}_2) \ll \eta + d^{-1/2} \tag{3.91}$$

as desired.

3.4.5 Ruling out corank 1

In this section we establish the bound (3.74), which completes the proof of Proposition 3.4.2 and hence of Theorem 3.1.2. By increasing the hidden constant in (3.74) we may assume η is at most a sufficiently small absolute constant.

By Lemma 3.4.3 it suffices to bound

$$\mathbb{P}(\mathcal{S}_{I_1} \setminus \mathcal{R}_2) = \mathbb{P}\left(R_{I_1} \in V_{I_1},\ \mathrm{corank}(M) = 1\right)$$

(taking η smaller if necessary). We cannot simply condition on all rows but R_{I_1} and pick a normal vector $u \in V_{I_1}^{\perp}$, since this conditioning fixes R_{I_1} as well by d-regularity. Instead, we will leave R_{I_2} random and express the event $\mathcal{S}_{I_1}$ in terms of a certain 2×2 determinant involving the rows R_{I_1}, R_{I_2}. We now have the advantage that on the bad event, we can condition on a unique (up to dilation) null vector x of M, which is independent of I_1, I_2.

We turn to the details. Conditional on M, I_1, I_2, we pick a pair of orthonormal vectors $u_1 \perp u_2 \in V_{(I_1,I_2)}^{\perp}$ uniformly at random, and independently of (R_{I_1}, R_{I_2}). (On $\mathcal{S}_{I_1} \wedge \mathcal{R}_2^c$ we have $\dim(V_{(I_1,I_2)}) = n-2$, so u_1, u_2 are an orthonormal basis for $V_{(I_1,I_2)}^{\perp}$ on this event.) In terms of u_1, u_2 we may construct a vector which is also orthogonal to R_{I_2} as follows:

$$z_1 := (R_{I_2} \cdot u_2)u_1 - (R_{I_2} \cdot u_1)u_2 \ \in V_{I_1}^{\perp}.$$

Since z_1 lies in the orthocomplement of V_{I_1}, on $\mathcal{S}_{I_1}$ we have

$$\begin{aligned} 0 &= R_{I_1} \cdot z_1 \\ &= (R_{I_1} \cdot u_1)(R_{I_2} \cdot u_2) - (R_{I_2} \cdot u_1)(R_{I_1} \cdot u_2) \\ &=: D_M(I_1, I_2). \end{aligned} \tag{3.92}$$

Hence,

$$\mathbb{P}\big(\mathcal{S}_{I_1} \setminus \mathcal{R}_2\big) \le \mathbb{P}\big(D_M(I_1,I_2) = 0,\ \operatorname{corank}(M) = 1\big). \tag{3.93}$$

Substituting $\widetilde{M}$ for M, may express the 2×2 determinant using (3.79):

$$\begin{aligned} D_{\widetilde{M}}(I_1,I_2) &= \big[A(u_1) + W(u_1)\big]\big[A(u_2) - W(u_2)\big] - \big[A(u_2) + W(u_2)\big]\big[A(u_1) - W(u_1)\big] \\ &= 2A(u_2)W(u_1) - 2A(u_1)W(u_2) \\ &= \sum_{j \in \mathrm{Ex}_M(I_1,I_2)} \xi(j)\big[2A(u_2)\partial_j^{\pi}(u_1) - 2A(u_2)\partial_j^{\pi}(u_2)\big] \\ &= \sum_{j \in \mathrm{Ex}_M(I_1,I_2)} \xi(j)\partial_j^{\pi}(v) && (3.94) \\ &= W(v) && (3.95) \end{aligned}$$

where we have defined

$$\begin{aligned} v &:= 2A(u_2)u_1 - 2A(u_2)u_2 \\ &= \big[(R_{I_1} + R_{I_2}) \cdot u_2\big]u_1 - \big[(R_{I_1} + R_{I_2}) \cdot u_1\big]u_2 \\ &\in V^{\perp}_{(I_1,I_2)}. \end{aligned} \tag{3.96}$$

We would like to replace M with $\widetilde{M}$ and bound (3.93) using the random walk representation (3.95) with Theorem 3.2.1. First we must reduce to an event on which many of the steps $\partial^{\pi}_j(v)$ are nonzero. We will do this in two stages. First we must remove a bad event on which $v = 0$; in light of (3.96) this is the event

$$\mathcal{B}_0 := \left\{ R_{I_1} + R_{I_2} \in V_{(I_1,I_2)} \right\}.$$

Once we have done this, we will be able to argue that v is unstructured in a manner similar to the way we controlled the event $\mathcal{B}'$ in Section 3.4.4.

We begin with $\mathcal{B}_0$. Since we are free to restrict to $\mathcal{R}_2^c \wedge \mathcal{G}^{\text{sls}}_{\eta}$, let us condition on M such that these events hold. On $\mathcal{B}_0 \wedge \mathcal{R}_2^c$, M has exactly one nontrivial left null vector (up to dilation) which we denote by y; furthermore, on $\mathcal{G}^{\text{sls}}(\eta)$ the level sets of y are of size at most ηn. Now $\mathcal{B}_0$ is the event that M has a left null vector y' with $y'(I_1) = y'(I_2) \neq 0$, so we have

$$\mathcal{B}_0 \wedge \mathcal{R}_2^c \subset \{y(I_1) = y(I_2)\}.$$

It follows that

$$\begin{aligned} \mathbb{P}_{I_1,I_2}\big(\mathcal{B}_0 \wedge \mathcal{R}_2^c \wedge \mathcal{G}^{\text{sls}}(\eta)\big) &\le \mathbb{P}_{I_1,I_2}\left(y(I_1) = y(I_2)\right) \\ &\le \eta(1 + o(1)) \end{aligned} \tag{3.97}$$

which is small enough.

Similarly to what we did in Section 3.4.4, for $\varepsilon_1 \in (0,1)$ we define

$$\mathcal{H}'_{M,I_1,I_2}(\varepsilon_1) = \left\{ x \in \mathbb{R}^n : \exists \lambda \in \mathbb{R} \text{ with } \min_{l=1,2} \left| x^{-1}(\lambda) \cap \mathcal{N}_M(I_l) \right| > \varepsilon_1 d \right\} \tag{3.98}$$

but we also set for $\varepsilon_2 \in (0,1)$

$$\mathcal{H}''_{M,I_1,I_2}(\varepsilon_2) = \left\{ x \in \mathbb{R}^n : \exists \lambda \in \mathbb{R} \text{ with } \left| x^{-1}(\lambda) \cap \left[\mathrm{Ex}_M(I_1,I_2) \cup \mathrm{Ex}_M(I_2,I_1) \right] \right| > \varepsilon_2 p(1-p)n \right\}. \tag{3.99}$$

For $\varepsilon \in (0,1)$ we have the inclusion

$$\begin{aligned} \mathcal{H}''_{M,I_1,I_2}(1+2\varepsilon) &\subset \mathcal{H}'_{M,I_1,I_2}(2\varepsilon(1-p)) \\ &\subset \mathcal{H}'_{M,I_1,I_2}(\varepsilon) \end{aligned} \tag{3.100}$$

(by our assumption $p \le 1/2$). Since $\mathcal{H}''_{M,I_1,I_2}$ is determined by $\mathrm{Ex}_M(I_1,I_2) \cup \mathrm{Ex}_M(I_2,I_1) = \mathrm{Ex}_{\widetilde{M}}(I_1,I_2) \cup \mathrm{Ex}_{\widetilde{M}}(I_2,I_1)$, we have

$$\mathcal{H}''_{M,I_1,I_2}(\varepsilon) = \mathcal{H}''_{\widetilde{M},I_1,I_2}(\varepsilon) \tag{3.101}$$

for any $\varepsilon \in (0,1)$, whereas this invariance does not hold for $\mathcal{H}'_{M,I_1,I_2}(\varepsilon)$.

Let $\varepsilon > 0$ to be chosen later. On $\{\mathrm{corank}(M) = 1\}$, let x denote a fixed nontrivial null vector of M, so that $\ker(M) = \langle x \rangle$. From (3.92), on $\{D_M(I_1,I_2) = 0\}$ we have $z_1, z_2 \in \ker(M)$, where

$$z_2 := (R_{I_1} \cdot u_1)u_2 - (R_{I_1} \cdot u_2)u_1 \ \in V_{I_2}^{\perp}$$

and (as before)

$$z_1 := (R_{I_2} \cdot u_2)u_1 - (R_{I_2} \cdot u_1)u_2 \ \in V_{I_1}^{\perp}.$$

It follows that on $\{D_M(I_1,I_2) = 0\}$ we have $v = z_1 - z_2 \in \ker(M)$. On the intersection of this event with $\{\mathrm{corank}(M) = 1\}$ and the event $\mathcal{B}_0^c$ on which v is non-zero, we have $0 \ne v \in \langle x \rangle$.

Hence,

$$\mathcal{B}_0^c \wedge \left\{D_M(I_1, I_2) = 0,\ \operatorname{corank}(M) = 1, v \in \mathcal{H}'_{M,I_1,I_2}(\varepsilon)\right\}$$
$$\subset \left\{\operatorname{corank}(M) = 1,\ x \in \mathcal{H}'_{M,I_1,I_2}(\varepsilon)\right\}. \tag{3.102}$$

We may now argue exactly as in Section 3.4.4 to conclude

$$\mathbb{P}_{I_1,I_2}\big(\mathcal{G}_\eta^{\mathrm{sls}} \wedge \left\{\operatorname{corank}(M) = 1,\ x \in \mathcal{H}'_{M,I_1,I_2}(\varepsilon)\right\}\big) \ll_\varepsilon \eta. \tag{3.103}$$

It only remains to bound

$$\mathbb{P}\left(D_M(I_1, I_2) = 0,\ v \notin \mathcal{H}'_{M,I_1,I_2}(\varepsilon)\right). \tag{3.104}$$

From (3.100) this is bounded by

$$\mathbb{P}\left(D_M(I_1, I_2) = 0,\ v \notin \mathcal{H}''_{M,I_1,I_2}(1+2\varepsilon)\right).$$

Now we replace M with $\widetilde{M}$. We make the crucial observation that the second event is unchanged by this substitution. Indeed, $\mathcal{H}''_{M,I_1,I_2}(1+2\varepsilon)$ is unchanged as noted in (3.101). Similarly, v is the same or M and $\widetilde{M}$ since

$$v = \big[(R_{I_1} + R_{I_2}) \cdot u_2\big]u_1 - \big[(R_{I_1} + R_{I_2}) \cdot u_1\big]u_2$$

and $R_{I_1} + R_{I_2} = \widetilde{R}_{I_1} + \widetilde{R}_{I_2}$ as the shuffling preserves the sets $\mathcal{N}_M(I_1) \cap \mathcal{N}_M(I_2)$ and $\mathcal{N}_M(I_1) \cup \mathcal{N}_M(I_2)$. Hence, (3.104) is bounded by

$$\mathbb{P}\left(D_{\widetilde{M}}(I_1, I_2) = 0,\ v \notin \mathcal{H}''_{M,I_1,I_2}(1+2\varepsilon)\right) = \mathbb{P}\left(W(v) = 0,\ v \notin \mathcal{H}''_{M,I_1,I_2}(1+2\varepsilon)\right). \tag{3.105}$$

In the final step of the argument, we must show that the set $\operatorname{Steps}(v)$ (as defined in (3.81)) is usually large off the event

$$\mathcal{B}'' := \left\{ v \in \mathcal{H}''_{M,I_1,I_2}(1+2\varepsilon) \right\}$$

with high probability in the randomness of π (and taking ε sufficiently small). Conditioning on M and I_1, I_2 such that $\mathcal{B}''$ does not hold, with $\mathrm{Flats}(v)$ as in (3.87) we have

$$\begin{aligned}
\mathbb{E}_\pi \left| \mathrm{Flats}(v) \right| &= \sum_{j \in \mathrm{Ex}_M(I_1,I_2)} \mathbb{P}_\pi \left(\pi(j) \in v^{-1}(v(j)) \right) \\
&= \sum_{j \in \mathrm{Ex}_M(I_1,I_2)} \frac{\left| v^{-1}(v(j)) \cap \mathrm{Ex}_M(I_2, I_1) \right|}{\left| \mathrm{Ex}_M(I_2, I_1) \right|} \\
&= \frac{1}{\mathrm{ex}_M(I_1, I_2)} \sum_{\lambda : v^{-1}(\lambda) \neq \phi} \left| v^{-1}(\lambda) \cap \mathrm{Ex}_M(I_1, I_2) \right| \left| v^{-1}(\lambda) \cap \mathrm{Ex}_M(I_2, I_1) \right| \\
&\le \frac{1}{\mathrm{ex}_M(I_1, I_2)} \sum_{\lambda : v^{-1}(\lambda) \neq \phi} \frac{1}{4} \left| v^{-1}(\lambda) \cap \left(\mathrm{Ex}_M(I_1, I_2) \cup \mathrm{Ex}_M(I_2, I_1) \right) \right|^2 \\
&\le \frac{(1+2\varepsilon)p(1-p)n}{4\,\mathrm{ex}_M(I_1, I_2)} \sum_{\lambda : v^{-1}(\lambda) \neq \phi} \left| v^{-1}(\lambda) \cap \left(\mathrm{Ex}_M(I_1, I_2) \cup \mathrm{Ex}_M(I_2, I_1) \right) \right| \\
&\le \frac{1+2\varepsilon}{2} p(1-p)n.
\end{aligned}$$

We can then argue exactly as in (3.89) that

$$\mathbb{P}_\pi \left(\left| \mathrm{Flats}(v) \right| \ge \left(\frac{1}{2} + 2\varepsilon \right) d \left(1 - \frac{d}{n} \right) \right) \le \exp\left(-c \frac{\varepsilon^2}{1+\varepsilon} d \right) \tag{3.106}$$

(substituting $p = d/n$). On the other hand, on $\mathcal{G}^{\mathrm{ex}}(\delta)$ we have

$$\left| \mathrm{Ex}_M(I_1, I_2) \right| \ge (1-\delta) d \left(1 - \frac{d}{n} \right)$$

so that if we take ε and δ sufficiently small,

$$\left| \mathrm{Steps}(v) \right| = \left| \mathrm{Ex}_M(I_1, I_2) \right| - \left| \mathrm{Flats}(v) \right| \gg d \tag{3.107}$$

(again using our assumption $d \leq n/2$). Applying Lemma 3.4.4,

$$\mathbb{P}(\{W(v) = 0\} \wedge \mathcal{G}^{\mathrm{ex}}(\delta) \setminus \mathcal{B}'') \ll e^{-cd} + d^{-1/2}$$

which combines with (3.97) and (3.103) to give

$$\mathbb{P}(\mathcal{R}_1 \wedge \mathcal{R}_2^c \wedge \mathcal{G}^{\mathrm{sls}}(\eta)) \ll \eta + d^{-1/2}$$

as desired. □

3.5 Structured null vectors

Our aim in the section is to prove Propositions 3.2.3 and 3.2.4. The proof of the former outlines the proof of the latter and serves as a warmup. For the proof of Proposition 3.2.4 we will use Lemma 3.3.2 to inject random walks as in the previous section. We will also make heavier use of the discrepancy properties from Section 3.3.2. We remark that the proof of Proposition 3.2.3 is not needed for the proof of Proposition 3.2.4, so the reader who is only interested in the proof of the main theorem can begin at Section 3.5.2.

3.5.1 Warmup: no sparse null vectors for H

In this section we prove Proposition 3.2.3. We restrict the sample space to the event $\mathcal{G}(d)$ as defined in Theorem 3.1.13. Recall that $\mathcal{G}(d)$ is the event that for some constants $c_1, c_2, C_2 > 0$ and a parameter $\kappa_3 \in [1, \infty)$ (possibly depending on n), the following conditions on Σ hold:

0. (Minimum degree) For all $i \in [n]$, $|\mathcal{N}_\Sigma(i)|, |\mathcal{N}_{\Sigma^\intercal}(i)| \geq d$.

1. (Expansion of small sets) For all $\gamma \in (0, c_1]$, for all $S \subset [n]$ such that $|S| \leq \frac{\log n}{2\gamma}\frac{n}{d}$, we have $|\mathcal{N}_\Sigma(S)|, |\mathcal{N}_{\Sigma^\intercal}(S)| \geq \frac{\gamma}{\log n} d|S|$.

2. (No large sparse minors) For all $A, B \subset [n]$ such that $|A|, |B| \geq C_2 \frac{n}{d} \log n$, we have $e_\Sigma(A, B) \geq c_2 \frac{d}{n}|A||B|$.

3. (No thin dense minors) For any $S, B \subset [n]$, $e_\Sigma(S,B), e_\Sigma(B,S) \le \kappa_3 d|S|$.

We assume $d \ge C_0' \log^2 n$ for some $C_0' > 0$ to be taken sufficiently large depending on c_1, c_2, C_2. As in Section 3.4.1, for $0 \le i \le 3$ we let $\mathcal{G}_i(d)$ denote the event that condition i above holds for Σ, so that $\mathcal{G}(d) = \bigwedge_{i=0}^3 \mathcal{G}_i(d)$. We continue to denote the rows of Σ by r_i and the rows of Ξ by Y_i, so that the ith row of H is $R_i = r_i \circ Y_i$.

Since the event $\mathcal{G}(d)$ is the same if we replace Σ with Σ^T, it suffices to consider only right null vectors. For $k \in [n]$, let

$$\mathcal{E}_k = \{\exists x \in \mathbb{R}^n : 0 < |\operatorname{spt}(x)| \le k,\ Hx = 0\}.$$

Our goal is to show that $\mathcal{E}_{(1-\eta)n}^c$ holds with probability $1 - O(n^{-100})$. We have

$$\mathbb{P}\big(\mathcal{E}_{(1-\eta)n}\big) = \sum_{k=2}^{\lfloor (1-\eta)n \rfloor} \mathbb{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1}) \tag{3.108}$$

(noting that $\mathcal{E}_1$ is empty). Fix $1 \le k \le (1-\eta)n$. We can follow the same lines establishing (3.27) in the proof of Proposition 3.2.2 to bound

$$\mathbb{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1}) \le \sum_{S \in \binom{[n]}{k-1}} \sum_{T \in \binom{[n]}{k}} \mathbb{P}(\mathcal{E}_{S,T}) \tag{3.109}$$

where

$$\mathcal{E}_{S,T} := \Big\{ \exists x \in \mathbb{R}^n :\ Hx = 0,\ \operatorname{spt}(x) = T,\ \{R_i\}_{i \in S} \text{ are linearly independent} \Big\}.$$

(We have (3.109) instead of (3.27) since the rows and columns of H are not exchangeable.)

Now we fix arbitrary $S, T \subset [n]$ of respective sizes $k-1$, k. Fix also an arbitrary $v \in \mathbb{R}^n$ with support T. Since conditioning on $(R_i)_{i\in S}$ fixes x on $\mathcal{E}_{S,T}$, it suffices to bound

$$\mathbb{P}\Big(H_{S^c \times [n]} v = 0 \,\Big|\, \Sigma, (Y_i)_{i \in S} \Big)$$

uniformly in v.

Our approach is different depending on whether k is small or large. In both cases, we use the fact that the rows R_i decouple after conditioning on Σ:

$$\begin{aligned}\mathbb{P}\Big(H_{S^c\times[n]}v = 0 \,\Big|\, \Sigma\Big) &= \prod_{i\in S^c} \mathbb{P}\Big(R_i\cdot v = 0 \,\Big|\, \Sigma\Big)\\ &= \prod_{i\in S^c} \mathbb{P}\Big(Y_i\cdot (r_i\circ v) = 0 \,\Big|\, \Sigma\Big).\end{aligned}$$

Now under this conditioning, the random variables $Y_i\cdot(r_i\circ v)$ are random walks (in the sense of Theorem 3.2.1). For small k, it will be enough to show that there are many $i\in S^c$ such that

$$\Big|\operatorname{spt}(R_i)\cap T\Big| = \Big|\operatorname{spt}(r_i\circ v)\Big| \geq 1. \tag{3.110}$$

For such i, the random walk takes at least 1 nonzero step since $v(j)\neq 0$ for all $j\in T$, so we have

$$\mathbb{P}\left(Y_i\cdot(r_i\circ v) = 0 \middle| \Sigma\right) \leq 1/2$$

in this case. To lower bound the number of rows satisfying (3.110) we will use our restriction to the "expansion of small sets" event $\mathcal{G}_1(d)$.

For larger k we will need to argue that the random walks $Y_i\cdot(r_i\circ v)\Big|\Sigma$ take more steps. For this we prove a consequence of our restriction to $\mathcal{G}_2(d)$ (Lemma 3.5.1 below), which essentially guarantees that for most $i\in S^c$, the intersection of any sufficiently large set B with the neighborhood $\mathcal{N}_\Sigma(i) = \operatorname{spt}(r_i)$ has roughly its expected size, which by our restriction to $\mathcal{G}_0(d)$ is at least $p|B|$ (where we continue to denote $p := d/n$). Applying this with $B=T$ gives $|\operatorname{spt}(r_i)\cap T| \gg pk$ for most $i\in S^c$. We will build on this idea in the proof of Proposition 3.2.4 for the unsigned rrd matrix M, where we will also need that a large set B "sees" roughly the expected portion of the sets $\operatorname{Ex}_M(i_1,i_2)$.

We turn to the details. Let $\gamma \in (0, c_1]$ to be chosen later. First assume $k \le \frac{1}{2\gamma}\frac{n\log n}{d}$. Let $A_0 = \{i \in S^c : r_i 1_T \neq 0\}$. By our restriction to $\mathcal{G}_1(d)$ we have

$$|A_0| = |\mathcal{N}_{\Sigma^\mathsf{T}}(T) \setminus S| \ge \gamma \frac{dk}{\log n} - k. \tag{3.111}$$

Since $v(j) \neq 0$ for all $j \in T$, we have that for $i \in A_0$,

$$\mathbb{P}\left(Y_i \cdot (r_i \circ v) = 0 \middle| \Sigma, (Y_i)_{i\in S}\right) \le 1/2$$

whence

$$\mathbb{P}\Big(H_{S^c\times[n]} v = 0 \,\Big|\, \Sigma, (Y_i)_{i\in S}\Big) \le \prod_{i\in A_0} \mathbb{P}\Big(Y_i \cdot (r_i \circ v) = 0 \,\Big|\, \Sigma, (Y_i)_{i\in S}\Big) \le \left(\frac{1}{2}\right)^{\gamma\frac{dk}{\log n} - k}.$$

Since this bound is uniform in $v, \{r_i \circ Y_i\}_{i\in S}$, we conclude from (3.109) that

$$\begin{aligned}\mathbb{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1}) &\le \binom{n}{k}\binom{n}{k-1}\left(\frac{1}{2}\right)^{\gamma\frac{dk}{\log n} - k} \\ &\le \exp\left(Ck\log n - c\gamma \frac{dk}{\log n}\right) \\ &\le n^{-100k} \end{aligned} \tag{3.112}$$

where we have taken $C_0' > C'/\gamma$ for a sufficiently large constant C' (we will later fix γ depending on C_2), so that $d \ge \frac{C'}{\gamma}\log^2 n$. Summing the bounds (3.112) gives

$$\mathbb{P}(\mathcal{E}_k) \ll n^{-100} \tag{3.113}$$

for any $k \le \frac{1}{2\gamma}\frac{n\log n}{d}$.

Now assume $k > \frac{1}{2\gamma}\frac{n\log n}{d}$. For this case we apply the following consequence of our restriction to $\mathcal{G}_2(d)$. (Recall that $\mathcal{G}_2(d)$ is the event that condition 2 from Theorem 3.1.13 holds. Below we also make use of the constants C_2, c_2 defined there.)

Lemma 3.5.1. *For $A, B \subset [n]$ let*

$$A' = \{i \in A : \; |B(i)| \geq c_2 p|B|\}$$

where we use the shorthand $B(i) := B \cap \mathcal{N}_\Sigma(i)$, and denote $p := d/n$. On $\mathcal{G}_2(d)$, we have

$$|A \setminus A'| \ll p^{-1} \log n$$

if $|B| \geq \frac{\log n}{2\gamma p}$ with γ sufficiently small depending on C_2.

Proof. Define
$$\mathcal{F} = \left\{(A, B) : \; A, B \subset [n], \; \min(|A|, |B|) \geq C_2 \frac{\log n}{p}\right\}$$
so that on $\mathcal{G}_2(d)$ we have $e_\Sigma(A, B) \geq c_2 p|A||B|$ for all $(A, B) \in \mathcal{F}$.

Denote $S = A \setminus A'$. We claim $(S, B) \notin \mathcal{F}$. Indeed, if this were not the case we would have

$$\begin{aligned} c_2 p|S||B| &\leq e_\Sigma(S, B) \\ &= \sum_{i \in S} |B(i)| \\ &< c_2 p|S||B| \end{aligned}$$

a contradiction.

Suppose $|B| \leq |S|$. Since $(S, B) \notin \mathcal{F}$ we have

$$\frac{1}{2\gamma} \frac{n \log n}{d} \leq |B| \leq C_2 \frac{n \log n}{d}.$$

Taking γ sufficiently small depending on C_2 we obtain a contradiction, and so $|S| \leq |B|$, and by the definition of $\mathcal{F}$ we must have $|S| \ll \frac{\log n}{p}$. $\square$

Applying the lemma with $A = S^c$, $B = T$, we have that for all $i \in A'$, $|\operatorname{spt}(r_i) \cap T| \geq c_2 pk$, and so by Theorem 3.2.1,

$$\mathbb{P}\Big(Y_i \cdot (r_i \circ v) = 0 \,\Big|\, \Sigma, (Y_i)_{i \in S}\Big) \ll (pk)^{-1/2}. \tag{3.114}$$

It follows that

$$\begin{aligned}\mathbb{P}\Big(H_{S^c \times [n]} v = 0 \,\Big|\, \Sigma, (Y_i)_{i \in S}\Big) &\leq \prod_{i \in A'} \mathbb{P}\Big(Y_i \cdot (r_i \circ v) = 0 \,\Big|\, \Sigma, (Y_i)_{i \in S}\Big) \\ &\leq \left[\frac{C}{\sqrt{pk}}\right]^{n-k-O\left(\frac{n \log n}{d}\right)}. \end{aligned} \tag{3.115}$$

For $\frac{n \log n}{d} \ll k \leq \frac{n}{2}$ this expression is bounded by $O\big(\exp\left(-cn \log\log n\right)\big)$, which combines with (3.109) to give

$$\begin{aligned}\mathbb{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1}) &\ll 4^n \exp\left(-cn \log\log n\right) \\ &= O\left(\exp\left(-cn \log\log n\right)\right). \end{aligned} \tag{3.116}$$

For $\frac{n}{2} \leq k \leq (1-\eta)n$ we instead bound (3.115) by

$$O\left(\exp\left(-\frac{1}{2}(n-k)\log(pk)\right)\right) = O\left(\exp\left(-\frac{1}{2}(n-k)\log d\right)\right)$$

assuming $\eta \geq C\frac{\log n}{d}$ for $C > 0$ sufficiently large. With (3.109) we conclude

$$\begin{aligned}\mathbb{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1}) &\ll \binom{n}{n-k}^2 \exp\left(-\frac{1}{2}(n-k)\log d\right) \\ &\leq \left(\frac{en}{n-k}\right)^{2(n-k)} d^{-(n-k)/2} \\ &\leq \left(\frac{e}{\eta d^{1/4}}\right)^{2\eta n} \\ &\ll \exp\left(-c\eta n\right) \end{aligned} \tag{3.117}$$

assuming η is at least a sufficiently large multiple of $d^{-1/4}$.

Summing the bounds (3.112), (3.116), (3.117) over $1 \le k \le (1-\eta)n$ completes the proof. $\square$

3.5.2 Preliminary reductions

We now turn to the unsigned rrd matrix M and the proof of Proposition 3.2.4. Recall our notation for the level sets of a vector $x \in \mathbb{R}^n$:

$$x^{-1}(\lambda) := \{i \in [n] : x(i) = \lambda\}$$

for $\lambda \in \mathbb{R}$. Our aim is to show that the good event

$$\mathcal{G}^{\mathrm{sls}}(\eta) := \Big\{\forall\lambda \in \mathbb{R} \text{ and } \forall\, 0 \neq x \in \mathbb{R}^n : Mx = 0 \text{ or } M^{\mathsf{T}}x = 0, \text{ we have } |x^{-1}(\lambda)| \le \eta n\Big\}$$

holds with probability $1 - O(n^{-100})$ for any $\eta \in [C_1 d^{-c_0}, 1]$, for some constants $C_1, c_0 > 0$. Let

$$\mathcal{G}_R(\eta) = \Big\{\, \forall\lambda \in \mathbb{R},\ \forall\, 0 \neq x \in \mathbb{R}^n \text{ such that } Mx = 0\ ,\big|x^{-1}(\lambda)\big| \le \eta n \,\Big\}. \tag{3.118}$$

Since $M \overset{d}{=} M^{\mathsf{T}}$, by a union bound it suffices to show that $\mathcal{G}_R(\eta)$ holds with probability $1 - O(n^{-100})$. The following claim recasts $\mathcal{G}_R(\eta)$ as the event that there is a sparse vector that is mapped by M to a constant vector.

Claim 3.5.2. For any $\eta \in (0, 1]$, we have

$$\mathcal{G}_R(\eta)^c = \Big\{\, \exists\, 0 \neq y \in \mathbb{R}^n : \big|\operatorname{spt}(y)\big| \le (1-\eta)n,\ My \in \{0, \mathbf{1}\}\, \Big\} \tag{3.119}$$

where we recall that $\mathbf{1} \in \mathbb{R}^n$ is the vector with all components equal to 1.

(We actually only need the containment $\subset$ in (3.119).)

Proof. Let us denote the right hand side of (3.119) by $\mathcal{B}(\eta)$. Suppose that $\mathcal{G}_R(\eta)$ fails. Then there exists a nontrivial null vector $x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$ such that $|x^{-1}(\lambda)| > \eta n$. Let

$y = \lambda \mathbf{1} - x$. Then y is nontrivial and $|\operatorname{spt}(y)| < (1-\eta)n$. Moreover,

$$My = \lambda M \mathbf{1} - Mx = \lambda d \mathbf{1} \in \langle \mathbf{1} \rangle$$

so by dilating y we see that $\mathcal{B}(\eta)$ holds.

Conversely, suppose that $\mathcal{B}(\eta)$ holds. Then there exists a nontrivial vector y supported on at most $(1-\eta)n$ coordinates such that My is either 0 or $\mathbf{1}$. If $My = 0$ then we are in $\mathcal{G}_R(\eta)^c$ (simply taking $x = y$ and $\lambda = 0$). So assume $My = \mathbf{1}$. Now letting $x = y - \frac{1}{d}\mathbf{1}$, we have that x is a right null vector of M with $|x^{-1}(1/d)| > \eta n$, so we are in $\mathcal{G}_R(\eta)^c$. $\square$

It remains to show that $\mathcal{G}_R(\eta)$ holds with probability $1 - O(n^{-100})$. Letting

$$\mathcal{E}_k := \left\{ \exists y \in \mathbb{R}^n : |\operatorname{spt}(y)| = k, \; My \in \{0, \mathbf{1}\} \right\} \tag{3.120}$$

we have

$$\mathbb{P}\left(\mathcal{G}_L(\eta)^c\right) = \sum_{k=2}^{\lfloor (1-\eta)n \rfloor} \mathbb{P}\left(\mathcal{E}_k \setminus \mathcal{E}_{k-1}\right) \tag{3.121}$$

(note that $\mathcal{E}_1$ is empty since no column can be parallel to 0 or $\mathbf{1}$).

The following lemma is analogous to the bound (3.109) from the proof of Proposition 3.2.3. The proof is lengthier but follows similar reasoning.

Lemma 3.5.3 (Passing to a large minor)**.** *For $k \in [n]$, let*

$$\mathcal{W}_k = \left\{ \hat{v} \in \mathbb{R}^k : \; v(j) \neq 0 \; \forall j \in [k] \right\} \tag{3.122}$$

be the set of vectors in $\mathbb{R}^k$ with full support. Suppose that for some $Q_k \geq 0$ we have a bound

$$\mathbb{P}\Big(M_{[k+1,n]\times[k]} \hat{v} = \alpha \mathbf{1} \;\Big|\; R_1, \dots, R_k \Big) \leq Q_k \tag{3.123}$$

that is uniform in the choice of $\hat{v} \in \mathcal{W}_k$, $\alpha \in \{0,1\}$ and the realization $R_1, \dots, R_k$ of the first

k rows of M. Then we have

$$\mathbb{P}\left(\mathcal{E}_k \setminus \mathcal{E}_{k-1}\right) \ll \binom{n}{k}^2 Q_k. \tag{3.124}$$

Proof. By column exchangeability and a union bound, we have

$$\mathbb{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1}) \leq \binom{n}{k} \mathbb{P}\big(\mathcal{E}_{[k]} \setminus \mathcal{E}_{k-1}\big) \tag{3.125}$$

where $\mathcal{E}_{[k]} = \mathcal{E}^0_{[k]} \vee \mathcal{E}^1_{[k]}$, with

$$\mathcal{E}^0_{[k]} := \big\{\exists x \in \mathbb{R}^n : \ \mathrm{spt}(x) = [k],\ Mx = 0\big\}$$

and

$$\mathcal{E}^1_{[k]} := \big\{\exists x \in \mathbb{R}^n : \ \mathrm{spt}(x) = [k],\ Mx = \mathbf{1}\big\}.$$

From $\mathcal{E}_{[k]} = \mathcal{E}^0_{[k]} \vee (\mathcal{E}^1_{[k]} \setminus \mathcal{E}^0_{[k]})$ we may bound

$$\mathbb{P}\big(\mathcal{E}_{[k]} \setminus \mathcal{E}_{k-1}\big) \leq \mathbb{P}\big(\mathcal{E}^0_{[k]} \setminus \mathcal{E}_{k-1}\big) + \mathbb{P}\big(\mathcal{E}^1_{[k]} \setminus (\mathcal{E}^0_{[k]} \vee \mathcal{E}_{k-1})\big). \tag{3.126}$$

For the first term on the right hand side, note that on $\mathcal{E}^0_{[k]} \setminus \mathcal{E}_{k-1}$ the minor $M_{[n]\times[k]}$ has $k-1$ linearly independent rows. Indeed, if this were not the case we would have $\mathrm{rank}(M_{[n]\times[k]}) \leq k-2$, so that $M_{[n]\times[k]}$ has 2 linearly independent right null vectors $x_1, x_2 \in \mathbb{R}^k$. But there is a $k-1$-sparse linear combination of x_1, x_2, putting us in $\mathcal{E}_{k-1}$.

For the second term in (3.126), note that on the complement of $\mathcal{E}^0_{[k]} \vee \mathcal{E}_{k-1}$ the minor $M_{[n]\times[k]}$ has full rank, and hence has k linearly independent rows.

Now we spend some symmetry to fix the linearly independent rows. Let $\mathcal{L}_i$ denote the event that $R_1, \ldots, R_i$ are linearly independent. By row exchangeability we have

$$\mathbb{P}\big(\mathcal{E}^0_{[k]} \setminus \mathcal{E}_{k-1}\big) \leq \binom{n}{k-1} \mathbb{P}\big(\big(\mathcal{E}^0_{[k]} \setminus \mathcal{E}_{k-1}\big) \wedge \mathcal{L}_{k-1}\big) \tag{3.127}$$

and

$$\mathbb{P}\big(\mathcal{E}^1_{[k]} \setminus (\mathcal{E}^0_{[k]} \vee \mathcal{E}_{k-1})\big) \leq \binom{n}{k} \mathbb{P}\big(\big(\mathcal{E}^1_{[k]} \setminus (\mathcal{E}^0_{[k]} \vee \mathcal{E}_{k-1})\big) \wedge \mathcal{L}_k\big). \tag{3.128}$$

In (3.127), $(\mathcal{E}^0_{[k]} \setminus \mathcal{E}_{k-1}) \wedge \mathcal{L}_{k-1}$ is the event that the first $k-1$ rows of M are linearly independent, that there is a null vector x of M supported on $[k]$, and that there are no $k-1$-sparse null vectors of M. Now on this event there is actually only one possibility for x up to dilation. Indeed, on $\mathcal{L}_{k-1}$ the system

$$M_{[k-1]\times[k]} z = 0 \tag{3.129}$$

has a unique solution up to dilation, by the linear independence of the first $k-1$ rows. Let us pick a nontrivial solution $\hat{x} \in \mathbb{R}^k$ of (3.129) arbitrarily, and set $x^* = (\hat{x} \quad 0)^\mathsf{T} \in \mathbb{R}^n$. On the complement of $\mathcal{E}_{k-1}$, each component of $\hat{x}$ is nonzero. Hence, $(\mathcal{E}^0_{[k]} \setminus \mathcal{E}_{k-1}) \wedge \mathcal{L}_{k-1}$ is contained in the event

$$\begin{aligned} \mathcal{E}'_k &:= \mathcal{L}_{k-1} \wedge \big\{\hat{x}(j) \neq 0 \text{ for all } j \in [k]\big\} \wedge \big\{Mx^* = 0\big\} \\ &= \mathcal{L}'_{k-1} \wedge \big\{M_{[k,n]\times[k]}\hat{x} = 0\big\} \end{aligned}$$

where we have let $\mathcal{L}'_{k-1} = \mathcal{L}_{k-1} \wedge \{\hat{x}(j) \neq 0 \text{ for all } j \in [k]\}$. We emphasize that $\hat{x}$ is a random vector in $\mathbb{R}^k$, defined only on the event $\mathcal{L}_{k-1}$, and fixed by conditioning on the first $k-1$ rows of M through (3.129).

We may similarly fix the vector in the preimage of $\mathbf{1}$ on the event $(\mathcal{E}^1_{[k]} \setminus (\mathcal{E}^0_{[k]} \vee \mathcal{E}_{k-1})) \wedge \mathcal{L}_k$ from (3.128). This event is disjoint from the event $(\mathcal{E}^0_{[k]} \setminus \mathcal{E}_{k-1}) \wedge \mathcal{L}_{k-1}$ from (3.127), and on it we may define $\hat{y} \in \mathbb{R}^k$ as the unique solution of

$$M_{[k]\times[k]} y = \mathbf{1}\,. \tag{3.130}$$

Setting

$$\mathcal{E}''_k := \mathcal{L}''_k \wedge \big\{M_{[k+1,n]\times[k]}\hat{y} = \mathbf{1}\big\} \tag{3.131}$$

where

$$\mathcal{L}''_k := \mathcal{L}_k \wedge \big\{\hat{y}(j) \neq 0 \text{ for all } j \in [k]\big\}$$

we similarly conclude that

$$(\mathcal{E}^1_{[k]} \setminus (\mathcal{E}^0_{[k]} \vee \mathcal{E}_{k-1})) \wedge \mathcal{L}_k \subset \mathcal{E}''_k.$$

Here also, $\hat{y} \in \mathbb{R}^k$ is a random vector defined only on the event $\mathcal{L}_k$ via (3.130), fixed by conditioning on the first k rows of M.

Combined with (3.127), (3.128), (3.126) and (3.125), we have

$$\mathbb{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1}) \leq \binom{n}{k}\binom{n}{k-1}\mathbb{P}(\mathcal{E}'_k) + \binom{n}{k}\binom{n}{k}\mathbb{P}(\mathcal{E}''_k). \tag{3.132}$$

By conditioning on a realization of $R_1, \dots, R_k$ such that $\mathcal{L}'_k$ holds, which fixes $\hat{x}$, we see that $\mathbb{P}(\mathcal{E}'_k) \leq Q_k$, with Q_k as in (3.123). We similarly have that $\mathbb{P}(\mathcal{E}''_k) \leq Q_k$, and the result follows. □

The bound Q_k will play the same role as bounds on $\mathbb{P}(\mathcal{E}_{S,T})$ did in the proof of Proposition 3.2.3 in Section 3.5.1. As in that proof, our approach will be different depending on the size of k. We want to control the event

$$\{\, M_{[k+1,n]\times[k]}\hat{v} = \alpha\,\mathbf{1} \,\} = \bigwedge_{i=k+1}^{n} \{\, R_i \cdot v = \alpha \,\} \tag{3.133}$$

where $v = (\hat{v} \quad 0)$ extends $\hat{v}$ to a vector in $\mathbb{R}^n$. In Section 3.5.1 we did this by conditioning on M and using the randomness of the signs. We then viewed (3.133) as the event that several independent random walks all landed at the same point, and used the expansion properties enjoyed by Σ on the good events $\mathcal{G}_1(d), \mathcal{G}_2(d)$ to argue that a large number of the walks took a large number of steps.

Here we will "inject" random walks into the distribution of the dot products $R_i \cdot v$ by applying the shuffling couplings of Lemmas 3.3.2 and 3.3.3. For small k, we will apply shufflings to pairs of *columns*, which will be chosen so that the number of rows altered by the switchings is large. Conditioning on M, in the randomness of the switchings we will have that the events on the right hand side of (3.133) are independent, and have probability at most $1/2$ for the affected rows. For large k we will apply shufflings independently to several non-overlapping pairs of rows, and use Lemma 3.4.4 to bound the probabilities of the events in (3.133).

By Corollaries 3.3.6 and 3.3.7 we may restrict to $\mathcal{G}^{\mathrm{edge}}(\varepsilon)$ and $\mathcal{B}(\varepsilon_0,\gamma)^c$ for some $\varepsilon,\varepsilon_0,\gamma>0$ to be chosen sufficiently small and independent of n – these events will play similar roles to the events $\mathcal{G}_2(d),\mathcal{G}_1(d)$, respectively, in the previous section. By Theorem 3.3.4 we may also restrict to $\mathcal{G}^{\mathrm{ex}}(\delta)$ for some $\delta>0$ to be chosen small independent of n. For now let $\eta\in(0,1]$ possibly depending on d. We will put restrictions on the range of η as the proof develops, ultimately taking $\eta\ge C_1d^{-c_0}$ for some constants $C_1,c_0>0$.

3.5.3 High sparsity

Fix $k\le\frac{1}{2\gamma}\frac{n\log n}{d}$. Towards an application of Lemma 3.5.3, we fix $\hat{v}\in\mathcal{W}_k$ and $\alpha\in\{0,1\}$. Pair off the first k columns of M with the last k columns according to some bijection

$$\sigma:[k]\to[n-k+1,n]$$

chosen in some arbitrary fashion, say uniformly at random and independently of M.

The following lemma shows that we can locate a large number of pairs of columns $(j,\sigma(j))$ on which we can perform independent restricted shufflings (see Lemma 3.3.3). We use restricted shufflings rather than Lemma 3.3.2 in order to "spread out" the switching modifications to M. Specifically, we want to ensure that each row of M is only affected by at most one random sign, in order to decouple the events in (3.133).

Lemma 3.5.4 (Locating disjoint patches of row indices for shufflings)**.** *Let $\varepsilon_0\in(0,1)$, and assume $\varepsilon_0,\gamma,\delta$ are sufficiently small. Then on the event $\mathcal{G}^{\mathrm{ex}}(\delta)\wedge\mathcal{B}(\varepsilon_0,\gamma)^c$ (the former event was defined in Theorem 3.3.4 and the latter in Corollary 3.3.7), for some $m\gg k/\log n$ there exists an increasing sequence of column indices*

$$1\le j_1<j_2<\cdots<j_m\le k \tag{3.134}$$

and an increasing sequence of sets of row indices

$$[k] = \mathrm{Frozen}(1) \subset \mathrm{Frozen}(2) \subset \cdots \subset \mathrm{Frozen}(m) \subset [n] \tag{3.135}$$

such that the following properties hold:

1. *(Patches are large) For each $\ell \in [m]$, letting*

$$A_\ell^+ := \mathrm{Ex}_{M^\intercal}(j_\ell, \sigma(j_\ell)) \setminus \mathrm{Frozen}(\ell - 1)$$
$$A_\ell^- := \mathrm{Ex}_{M^\intercal}(\sigma(j_\ell), j_\ell) \setminus \mathrm{Frozen}(\ell - 1)$$

   we have

$$|A_\ell^+|,\, |A_\ell^-| \geq .01d. \tag{3.136}$$

2. *(Disjointness) The m sets $\left\{A_\ell^+ \cup A_\ell^-\right\}$ are pairwise disjoint.*

3. *(Conditioning) For each $\ell \in [m]$, $\mathrm{Frozen}(\ell)$ is fixed by conditioning on the $2(\ell - 1)$ columns $\bigcup_{\ell' < \ell} \{j_{\ell'}, \sigma(j_{\ell'})\}$.*

We defer the proof of this lemma for now and use it to bound bound $\mathbb{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1})$. As we have already restricted to $\mathcal{G}^{\mathrm{ex}}(\delta) \wedge \mathcal{B}(\varepsilon_0, \gamma)^c$, let $m \gg k/\log n$ and the sequences $(j_\ell)_{\ell=1}^m$, $(A_\ell^\pm)_{\ell=1}^m$ and $(\mathrm{Frozen}(\ell))_{\ell=1}^m$ as in the lemma. We can form a coupling $(M, \widetilde{M})$ of rrd matrices using Lemma 3.3.3 by performing independent restricted shufflings on M at the columns $(j_\ell, \sigma(j_\ell))$. Specifically, letting $s := \lceil .01d \rceil$, for each $\ell \in [m]$ we draw $S_\ell^+ \subset A_\ell^+$, $S_\ell^- \subset A_\ell^-$ of size s independently and uniformly at random, and conditional on these $2m$ sets we draw m independent uniform random bijections $\pi_\ell : S_\ell^+ \to S_\ell^-$. We let $\xi : [n] \to \{\pm 1\}$ be a sequence of iid uniform signs independent of all other random variables. Then for each $\ell \in [m]$ and each $i \in S_\ell^+$, we replace the minor $M_{(i,\pi_\ell(i)) \times (j_\ell, \sigma(j_\ell))}$ with the random 2×2 matrix

$$\mathbf{I}_2\, \mathbb{1}(\xi(i) = +1) + \mathbf{J}_2\, \mathbb{1}(\xi(i) = -1).$$

By the independence of the signs $\xi(i)$ and the fact that the $2m$ sets $\left\{S_\ell^+, S_\ell^-\right\}_{\ell \in [m]}$ are pairwise disjoint, we have

$$\begin{aligned}\mathbb{P}\Big(\widetilde{M}_{[k+1,n]\times[k]}\hat{v} = \alpha\,\mathbf{1}\;\Big|\;M\Big) &\le \prod_{\ell=1}^{m}\prod_{i\in S_\ell^+}\mathbb{P}\left(\sum_{j\in[k]}\widetilde{M}(i,j)v(j) = \alpha\;\middle|\;M\right)\\ &\le \left(\frac{1}{2}\right)^{.01dm}\\ &\le \exp\left(-cdk/\log n\right).\end{aligned}$$

Since $M \overset{d}{=} \widetilde{M}$, by Lemma 3.5.3, we conclude

$$\begin{aligned}\mathbb{P}(\mathcal{E}_k\setminus\mathcal{E}_{k-1}) &\ll \binom{n}{k}^2\exp\left(-c\frac{dk}{\log n}\right)\\ &\le \exp\left(2k\left(\log n - c\frac{d}{\log n}\right)\right)\\ &\le \exp\left(-c\frac{dk}{\log n}\right)\\ &\le n^{-100k}\end{aligned}$$

by our assumption $d \ge C_0\log^2 n$, taking C_0 sufficiently large. Summing the bounds (3.112) gives

$$\mathbb{P}(\mathcal{E}_k) \ll n^{-100} \tag{3.137}$$

for any $k \le \frac{1}{2\gamma}\frac{n\log n}{d}$.

Proof. of Lemma 3.5.4. Set $A_0 = [k]$. We build the sequences $(j_\ell)_{\ell=1}^m$ and $(\mathrm{Frozen}(\ell))_{\ell=1}^m$ by a simple greedy procedure. For each $\ell \ge 1$, we inductively define j_ℓ to be the smallest $j \in [k]$ such that both of the sets

$$\mathrm{Ex}_{M^\intercal}(j,\sigma(j))\setminus\mathrm{Frozen}(\ell-1),\quad \mathrm{Ex}_{M^\intercal}(\sigma(j),j)\setminus\mathrm{Frozen}(\ell-1)$$

are of size at least $0.1d$. Then with A_ℓ^+, A_ℓ^- as in the statement of the lemma, we set

$$\mathrm{Frozen}(\ell) = \mathrm{Frozen}(\ell-1) \cup A_\ell^+ \cup A_\ell^-.$$

If no such j_ℓ exists, we set $m := \ell - 1$ and STOP.

The resulting sequences $(j_\ell)_{\ell=1}^m$ and $(\mathrm{Frozen}(\ell))_{\ell=1}^m$ clearly satisfy the three properties in the statement of the lemma. It only remains to show that the halting time m of the greedy procedure is of size $\Omega(k/\log n)$ if we take ε_0, γ sufficiently small.

We abbreviate

$$\mathrm{Ex}^+(j) := \mathrm{Ex}_{M^\intercal}(j, \sigma(j)), \quad \mathrm{Ex}^-(j) := \mathrm{Ex}_{M^\intercal}(\sigma(j), j).$$

We have that for all $j \in [k]$, either $|\,\mathrm{Ex}^+(j) \setminus \mathrm{Frozen}(m)|$ or $|\,\mathrm{Ex}^-(j) \setminus \mathrm{Frozen}(m)|$ is $< .01d$. For each $j \in [k]$, put $j \in S$ if $|\,\mathrm{Ex}^+(j) \setminus \mathrm{Frozen}(m)| < .01d$ and otherwise put $\sigma(j) \in S$, so that $|S| = k$.

Taking δ sufficiently small, by our restriction to $\mathcal{G}^{\mathrm{ex}}(\delta)$ we may assume

$$|\,\mathrm{Ex}^+(j)|, |\,\mathrm{Ex}^-(j)| \geq .1d$$

for all $j \in [k]$. It follows that at least one of $\mathrm{Ex}^+(j) \cap \mathrm{Frozen}(m)$, $\mathrm{Ex}^-(j) \cap \mathrm{Frozen}(m)$ is of size at least $.09d$. Since $\mathrm{Ex}^+(j) \subset \mathcal{N}_{M^\intercal}(j)$, $\mathrm{Ex}^-(j) \subset \mathcal{N}_{M^\intercal}(\sigma(j))$, we have

$$|\mathcal{N}_M(j) \cap \mathrm{Frozen}(m)| \geq .09d$$

for all $j \in S$. Now

$$e_M(\mathrm{Frozen}(m), S) = \sum_{j \in S} |\mathcal{N}_M(j) \cap \mathrm{Frozen}(m)| \geq .09d|S|$$

so taking $\varepsilon_0 < .09$ and γ sufficiently small, by our restriction to $\mathcal{B}(\varepsilon_0, \gamma)^c$ we must have

$$|\operatorname{Frozen}(m)| \geq \frac{\varepsilon_0 \gamma}{\log n} dk.$$

But by the inductive procedure to produce Frozen(m) we have $|\operatorname{Frozen}(m)| \leq k + 2md$, from which it follows that

$$m \gg \frac{k}{d}\left(\frac{d}{\log n} - 1\right) \gg \frac{k}{\log n}$$

by our assumption $d \geq C_0 \log^2 n$ (here we only need $d \geq 2 \log n$, say). □

3.5.4 Moderate sparsity

Now we fix k in the range $\left[\frac{1}{2\gamma}\frac{n \log n}{d}, (1-\eta)n\right]$.

The proof mirrors the proof for large k for the Hadamard product $H = \Sigma \circ \Xi$ in Section 3.5.1. The general idea is to express the event that $M_{[k+1,n]\times[k]}\hat{v} = \alpha\,\mathbf{1}$ as the event that several independent random walks all land at α. Without the iid signs enjoyed by H we must use the shuffling coupling of Lemma 3.3.2 to create random walks. We use the discrepancy property enforced by our restriction to the event $\mathcal{G}^{\text{edge}}(\varepsilon)$ (from Corollary 3.3.6) to argue that these walks take many steps (in particular we will need an extension of Lemma 3.5.1 used in Section 3.5.1), at which point we can apply the anti-concentration bound from Theorem 3.2.1 to each walk.

More precisely, we will fix disjoint sets of row indices $A_1, A_2 \subset A := [k+1, n]$ of equal size $a_1 = |A_1| = |A_2| \gg n - k$, and pair off the elements of A_1 with those of A_2 according to a bijection $\sigma : A_1 \to A_2$. For each $i \in A_1$, we perform a shuffling on M at the row pair $(i, \sigma(i))$; we do this independently for each $i \in A_1$ and denote the new matrix by $\widetilde{M}$. We have

$$\mathbb{P}\Big(M_{[k+1,n]\times[k]}\hat{v} = 0 \,\Big|\, R_1, \dots, R_k\Big) = \mathbb{E}_{R_{k+1},\dots,R_n}\, \mathbb{P}\Big(\widetilde{M}_{[k+1,n]\times[k]}\hat{v} = \alpha\,\mathbf{1} \,\Big|\, M\Big)$$

so it suffices to bound

$$\mathbb{P}\Big(\widetilde{M}_{[k+1,n]\times[k]}\hat{v} = \alpha\,\mathbf{1}\,\Big|\,M\Big) \le \prod_{i\in A_1}\mathbb{P}\big(\widetilde{R}_i\cdot v = \alpha\big|M\big). \tag{3.138}$$

As in Section 3.4, in order to bound the probabilities in (3.138) using Theorem 3.2.1, we will need to argue that many of these random walks take many steps. For this we take the pairing σ to be *random* – it is then possible to show using our restriction to the edge discrepancy event $\mathcal{G}^{\mathrm{edge}}(\varepsilon)$ that with overwhelming probability most of the pairs $(i,\sigma(i))$ give walks that take a large number of steps.

We turn to the details. Fix disjoint sets $A_1, A_2 \subset A := [k+1,n]$ with $a_1 := |A_1| = |A_2| \gg n-k$. We create a new rrd matrix $\widetilde{M}$ coupled to M from three additional sources of randomness:

1. a uniform random bijection $\sigma : A_1 \to A_2$ independent of all other variables;
2. a sequence $(\pi_i)_{i\in A_1}$ of uniform random bijections

$$\pi_i : \mathrm{Ex}_M(i,\sigma(i)) \to \mathrm{Ex}_M(\sigma(i),i)$$

   which are jointly independent conditional on M and σ;
3. an array $\Xi : [n]^2 \to \{\pm1\}$ of iid uniform random signs independent of all other variables.

Let $\xi_i = \Xi(i,\cdot)$ denote the ith row of the array of signs. We form $\widetilde{M}$ by performing a shuffling on M at $(i,\sigma(i))$ according to π_i and ξ_i for each $i \in A_1$. We have $\widetilde{M} \stackrel{d}{=} M$ by Lemma 3.3.2 and independence.

Recall the notation Steps from (3.81), and for fixed $i \in A_1$ denote

$$\begin{aligned}\mathrm{Steps}_i(\hat{v}) &:= \mathrm{Steps}_{M,\pi_i}^{(i,\sigma(i))}(v)\\ &= \{j \in \mathrm{Ex}_M(i,\sigma(i)) : v(j) \neq v(\sigma(j))\}\end{aligned}$$

where we recall $v = (\hat{v} \quad 0) \in \mathbb{R}^n$ with $\hat{v} \in \mathbb{R}^k$. Now since $\mathrm{spt}(v) = [k]$, we have that for each $i \in A_1$,

$$|\mathrm{Steps}_i(\hat{v})| \geq |\mathrm{Cross}_i(k)| \tag{3.139}$$

where we define

$$\begin{aligned}\mathrm{Cross}_i(k) = \mathrm{Cross}_{M,\pi_i}^{(i,\sigma(i))}(k) \\ := \{j \in \mathrm{Ex}_M(i,\sigma(i)) \cap [k] : \pi_i(j) \in [k+1,n]\}\end{aligned} \tag{3.140}$$

the number of pairs $(j, \pi_i(j))$ which are in $[k] \times [k+1,n]$, i.e. pairs which cross the partition $[n] = [k] \cup [k+1,n]$ going from left to right. (We could also include pairs crossing right to left, but this will tend to improve the lower bound (3.139) by only a constant factor.)

Hence, for $m \geq 1$, defining the good events

$$\mathcal{G}_i(m) := \big\{ \big| \mathrm{Cross}_i(k) \big| \geq m \big\} \tag{3.141}$$

for each $i \in A_1$, by Lemma 3.4.4 we have

$$\mathbb{P}_{\xi_i} \left\{ \widetilde{R}_i \cdot v = \alpha \right\} 1_{\mathcal{G}_i(m)^c} = O(m^{-1/2}). \tag{3.142}$$

In the remainder of the proof, we show that with overwhelming probability in the randomness of the bijections σ and $(\pi_i)_{i \in A_1}$, for most $i \in A_1$ and for a reasonably large value of m, $\mathcal{G}_i(m)$ holds except on an exponentially small event. (Hence we are done with the iid signs Ξ.) The randomness of M will only enter through our restriction to the events $\mathcal{G}^{\mathrm{edge}}(\varepsilon)$ and $\mathcal{G}^{\mathrm{ex}}(\delta)$.

Lemma 3.5.5 below summarizes what we need from the discrepancy property enforced on $\mathcal{G}^{\mathrm{edge}}(\varepsilon)$ – it is an extension of Lemma 3.5.1 from the proof for H. While for H it was enough to know that the intersections $B(i)$ of a large set B with the neighborhoods $\mathcal{N}_M(i)$ were of size roughly $p|B|$, here we will need intersections of B with the sets $\mathrm{Ex}_M(i_1,i_2)$, $\mathrm{Ex}_M(i_2,i_1)$ to be at least a constant factor of their expected size.

For $\varepsilon \in (0,1)$ and a set of column indices $B \subset [n]$, say that an ordered pair (i_1, i_2) of distinct row indices in A is *ε-bad for B* if either

$$|\operatorname{Ex}_M(i_1,i_2) \cap B| \le \varepsilon p|B| \quad \text{or} \quad |\operatorname{Ex}_M(i_2,i_1) \cap B^c| \le \varepsilon p(n-|B|). \tag{3.143}$$

The following lemma shows that on $\mathcal{G}^{\mathrm{edge}}(\varepsilon)$ with ε sufficiently small, only a small number of pairs of elements of $A = [k+1, n]$ are ε-bad for $[k]$.

Lemma 3.5.5. *Let $B \subset [n]$, and continue to denote $A = [k+1, n]$. For $i \in [n]$, denote $B(i) := \mathcal{N}_M(i) \cap B$. For $\varepsilon \in (0,1)$, define*

$$A_\varepsilon = \left\{ i \in A : \left| \frac{|B(i)|}{p|B|} - 1 \right| \le \varepsilon,\ \left| \frac{|B^c(i)|}{p(n-|B|)} - 1 \right| \le \varepsilon \right\}, \tag{3.144}$$

and for $i \in A$, let

$$S_\varepsilon(i) = \{ i' \in A_\varepsilon : (i,i') \text{ is } \varepsilon\text{-bad for } B \}. \tag{3.145}$$

On the event $\mathcal{G}^{\mathrm{edge}}(\varepsilon)$ from Corollary 3.3.6 we have

$$|A \setminus A_\varepsilon| \ll_\varepsilon p^{-1} \log n \tag{3.146}$$

and for every $i \in A_\varepsilon$,

$$|S_\varepsilon(i)| \ll_\varepsilon p^{-1} \log n \tag{3.147}$$

assuming $|B|, |B^c| \ge \frac{1}{2\gamma} p^{-1} \log n$ for γ sufficiently small depending on ε.

Proof. We begin with (3.146).

Define the sets

$$S_1 = \{ i \in A : |B(i)| < (1-\varepsilon)p|B| \}$$
$$S_2 = \{ i \in A : |B(i)| > (1+\varepsilon)p|B| \}$$
$$S_3 = \{ i \in A : |B^c(i)| < (1-\varepsilon)p(n-|B|) \}$$
$$S_4 = \{ i \in A : |B^c(i)| > (1+\varepsilon)p(n-|B|) \}$$

so that $A \setminus A_\varepsilon = \bigcup_{k=1}^4 S_k$. By the same lines as the proof of Lemma 3.5.1 we have $|S_1| \ll_\varepsilon p^{-1} \log n$. By replacing B with B^c we obtain the same bound on $|S_3|$. $|S_2|$ and $|S_4|$ are bounded similarly.

We turn to the estimate (3.147). Fix $i \in A_\varepsilon$. We can write $S_\varepsilon(i) = S_\varepsilon^1(i) \cup S_\varepsilon^2(i)$ where

$$\begin{aligned} S_\varepsilon^1(i) &= \big\{ i' \in A_\varepsilon : |\operatorname{Ex}(i,i') \cap B| \le \varepsilon p |B| \big\} \\ S_\varepsilon^2(i) &= \big\{ i' \in A_\varepsilon : |\operatorname{Ex}(i',i) \cap B^c| \le \varepsilon p (n - |B|) \big\}. \end{aligned}$$

We first bound $|S_\varepsilon^1(i)|$. For $i' \in S_\varepsilon^1(i)$, we have

$$|\operatorname{Ex}(i,i') \cap B| \le \varepsilon p |B| \le \frac{\varepsilon}{1-\varepsilon} |B(i)| \tag{3.148}$$

since $i \in A_\varepsilon$. It follows that

$$\begin{aligned} e_M\big(S_\varepsilon^1(i), B(i)\big) &= \sum_{i' \in S_\varepsilon^1(i)} \big|\operatorname{Co}(i,i') \cap B\big| \\ &= \sum_{i' \in S_\varepsilon^1(i)} |B(i)| - |\operatorname{Ex}(i,i') \cap B| \\ &\ge |S_\varepsilon^1(i)| \left(1 - \frac{\varepsilon}{1-\varepsilon}\right) |B(i)| \\ &= \frac{1-2\varepsilon}{1-\varepsilon} |S_\varepsilon^1(i)||B(i)|. \end{aligned} \tag{3.149}$$

Now we show this contradicts our restriction to the event $\mathcal{G}^{\text{edge}}(\varepsilon)$ if ε is sufficiently small. Recall the family $\mathcal{F}(\varepsilon)$ of pairs of subsets of $[n]$ defined in Corollary 3.3.6. If $\big(S_\varepsilon^1(i), B(i)\big) \in \mathcal{F}(\varepsilon)$ we have

$$e_M(S_\varepsilon^1(i), B(i)) \le (1+\varepsilon) p |S_\varepsilon^1(i)||B(i)|. \tag{3.150}$$

From (3.150) it follows that

$$e_M(S_\varepsilon^1(i), B(i)) \le \frac{(1+\varepsilon)}{2} |S_\varepsilon^1(i)||B(i)|$$

which contradicts (3.149) if ε is a sufficiently small absolute constant. We may hence assume

$(S^1_\varepsilon(i), B(i)) \notin \mathcal{F}(\varepsilon)$. Similarly to how we argued in the bound for $|S_1|$, we can deduce from the lower bound

$$|B(i)| \gg p|B| \gg \gamma^{-1} \log n$$

(since $i \in A_\varepsilon$) that taking γ sufficiently small, we must have $|S^1_\varepsilon(i)| \leq |B(i)|$ (for n sufficiently large), and hence

$$|S^1_\varepsilon(i)| \ll_\varepsilon p^{-1} \log n.$$

The proof that $|S^2_\varepsilon(i)| = O_\varepsilon(p^{-1} \log n)$ follows similar lines and is omitted. □

We define the subset of A_1 of "good" row indices to be

$$A_1' = \Big\{ i \in A_1 \cap A_\varepsilon : \sigma(i) \in A_\varepsilon \setminus S_\varepsilon(i) \Big\} \tag{3.151}$$

where $S_\varepsilon(i)$ is as in (3.145) with $B = [k]$. That is, A_1' is the set of $i \in A_1$ such that i and $\sigma(i)$ are both in A_ε, and such that the pair $(i, \sigma(i))$ is not bad for $[k]$. Note that this is a random set depending on M and σ. We can now use Lemma 3.5.5 and the randomness of σ to show that with overwhelming probability, A_1' constitutes most of A_1.

Let

$$\mathcal{B}' = \Big\{ \, |A_1 \setminus A_1'| \geq |A_1|/2 \, \Big\}. \tag{3.152}$$

Now for arbitrary $m \geq 1$ we have

$$\begin{aligned} \mathbb{P}\Big(\widetilde{M}_{[k+1,n]\times[k]} \hat{v} = \alpha \, \mathbf{1} \, \Big| \, M \Big) &\leq \mathbb{P}_\sigma(\mathcal{B}') + \mathbb{E}_\sigma \, 1_{\mathcal{B}'^c} \mathbb{P}\Big(\widetilde{M}_{[k+1,n]\times[k]} \hat{v} = \alpha \, \mathbf{1} \, \Big| \, M, \sigma \Big) \\ &\leq \mathbb{P}_\sigma(\mathcal{B}') + \mathbb{E}_\sigma \, 1_{\mathcal{B}'^c} \prod_{i \in A_1'} \mathbb{P}\big(\widetilde{R}_i \cdot v = \alpha \big| M, \sigma \big) \\ &\leq \mathbb{P}_\sigma(\mathcal{B}') + \mathbb{E}_\sigma \, 1_{\mathcal{B}'^c} \prod_{i \in A_1'} \Big[\mathbb{P}_{\pi_i}(\mathcal{G}_i(m)^c) + \mathbb{P}_{\xi_i} \Big\{ \widetilde{R}_i \cdot v = \alpha \Big\} 1_{\mathcal{G}_i(m)} \Big] \, . \end{aligned} \tag{3.153}$$

The term $\mathbb{P}_{\xi_i} \Big\{ \widetilde{R}_i \cdot v = \alpha \Big\} 1_{\mathcal{G}_i(m)}$ is $O(m^{-1/2})$ by (3.142). It remains to bound $\mathbb{P}_\sigma(\mathcal{B}')$ and $\mathbb{P}_{\pi_i}(\mathcal{G}_i(m)^c)$ (for some large value of m).

From Lemma 3.5.5 with $B = [k]$ we have

$$|A \setminus A_\varepsilon|\,, \max_{i \in A_1 \cap A_\varepsilon} |S_\varepsilon(i)| \le s_0 \tag{3.154}$$

for some $s_0 = O(p^{-1} \log n)$ (assuming $\eta \ge \frac{1}{2\gamma} \frac{\log n}{d}$). By crudely estimating the number of bad realizations of σ, we can bound

$$\mathbb{P}_\sigma\left(\mathcal{B}'\right) \le \binom{a_1}{\lfloor a_1/2 \rfloor} \frac{s_0^{a_1/2} \lfloor a_1/2 \rfloor!}{a_1!}$$

(first fixing the $\lfloor a_1/2 \rfloor$ elements of $A_1 \setminus A_1'$, then choosing from the at most s_0 options for $\sigma(i)$ for each $i \in A_1 \setminus A_1'$). Simplifying this expression and applying the inequality $n! \ge (n/e)^n$, true for all $n \in \mathbb{N}$,

$$\mathbb{P}_\sigma\left(\mathcal{B}'\right) \le \frac{s_0^{a_1/2}}{\lfloor a_1/2 \rfloor!} \le \left(\frac{C s_0}{a_1}\right)^{a_1/2}. \tag{3.155}$$

Now we estimate the terms $\mathbb{P}_{\pi_i}(\mathcal{G}_i(m)^c)$ (see (3.141) for the definition of these events). For fixed $i \in A_1'$ we have

$$\mathbb{E}_{\pi_i} |\operatorname{Cross}_i(k)| = \frac{\left|\operatorname{Ex}_M(i, \sigma(i)) \cap [k]\right| \left|\operatorname{Ex}_M(\sigma(i), i) \cap [k+1, n]\right|}{\left|\operatorname{Ex}_M(i, \sigma(i))\right|}. \tag{3.156}$$

From our restriction to $\mathcal{G}^{\mathrm{ex}}(\delta)$ we know the denominator is of size $\Theta_\delta(d)$, and since $i \in A_1'$ the numerator is of size $\Omega\left(p^2 k(n-k)\right)$ whence,

$$\mathbb{E}_{\pi_i} |\operatorname{Cross}_i(k)| \gg \frac{d}{n} \frac{k(n-k)}{n} \gg p \min(k, n-k). \tag{3.157}$$

From Lemma 3.3.9 it follows that

$$|\operatorname{Cross}_i(k)| \gg p \min(k, n-k) \tag{3.158}$$

except with probability at most $\exp\left(-cp \min(k, n-k)\right)$ in the randomness of π_i. We have hence shown that for $i \in A_1'$,

$$\mathbb{P}_{\pi_i}(\mathcal{G}_i(m)^c) \leq \exp\big(-cp\min(k, n-k)\big) \tag{3.159}$$

where set $m := cp\min(k, n-k)$, and $c > 0$ is a sufficiently small absolute constant. In particular, this bound is of lower order than the bound $\mathbb{P}_{\xi_i}\left\{\widetilde{R}_i \cdot v = \alpha\right\} 1_{\mathcal{G}_i(m)} = O(m^{-1/2})$.

Substituting our bounds (3.142), (3.155) and (3.159) into (3.153), we have

$$\begin{aligned}\mathbb{P}\Big(\widetilde{M}_{[k+1,n]\times[k]}\hat{v} = \alpha\,\mathbf{1} \,\Big|\, M\Big) &\leq \mathbb{P}_\sigma\left(\mathcal{B}'\right) + \mathbb{E}_\sigma\, 1_{\mathcal{B}'^c} O(m^{-1/2})^{|A_1'|} \\ &\leq \left(\frac{Cn\log n}{d(n-k)}\right)^{\kappa a_1} + m^{-(1-\kappa+o(1))a_1/2}.\end{aligned}$$

Applying Lemma 3.5.3 we have

$$\mathbb{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1}) \ll (\mathrm{I})_k + (\mathrm{II})_k \tag{3.160}$$

where

$$\begin{aligned}(\mathrm{I})_k &= \binom{n}{k}^2 \left(\frac{Cn\log n}{d(n-k)}\right)^{\kappa a_1} \\ (\mathrm{II})_k &= \binom{n}{k}^2 m^{-(1-\kappa+o(1))a_1/2}.\end{aligned}$$

First assume $\frac{n\log n}{d} \ll k \leq \frac{n}{2}$. In this case we have

$$m = cpk \gg \log n. \tag{3.161}$$

and

$$a_1 \gg n - k \geq n/2 \tag{3.162}$$

so

$$\begin{aligned}(\mathrm{I})_k &\leq 4^n \left(\frac{Cn\log n}{d(n-k)}\right)^{a_1/2} \\ &\leq 4^n \left(\frac{C\log n}{d}\right)^{cn} \\ &\leq e^{-cn}\end{aligned}$$

if $d \geq C' \log n$ for some C' sufficiently large. For the second term,

$$(\mathrm{II})_k \leq 4^n m^{-(1/2+o(1))a_1/2} = (\log n)^{-\Omega(n)}$$

by the lower bounds (3.161), (3.162). From these bounds and (3.160) we conclude

$$\mathbb{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1}) \leq e^{-cn} \tag{3.163}$$

for $\frac{n\log n}{d} \ll k \leq n/2$.

Now assume $n/2 \leq k \leq (1-\eta)n$. In this case we have

$$m = cp(n-k) = cd\frac{n-k}{n} \gg \eta d. \tag{3.164}$$

Since $a_1 = |A_1| = |A_2|$, $A_1, A_2 \subset A$ are arbitrary disjoint subsets, and $|A| = n-k$, we may take $a_1 = (\frac{1}{2} - o(1))(n-k)$. We then have

$$\begin{aligned}(\mathrm{I})_k &\leq \left(\frac{en}{n-k}\right)^{2(n-k)} \left(\frac{Cn\log n}{d(n-k)}\right)^{a_1/2} \\ &\leq \left[\frac{C\log n}{d}\left(\frac{n}{n-k}\right)^9\right]^{(\frac{1}{4}-o(1))(n-k)} \\ &\leq \left(\frac{C\log n}{d\eta^9}\right)^{c(n-k)}.\end{aligned}$$

By our assumption $d \geq C_0 \log^2 n$ we conclude

$$(\mathrm{I})_k \leq \left(\frac{C}{d^{1/2}\eta^9}\right)^{c(n-k)} \leq \left(\frac{C}{d\eta^{18}}\right)^{c'(n-k)}. \tag{3.165}$$

For the other term:

$$\begin{aligned}(\mathrm{II})_k &\leq \left(\frac{en}{n-k}\right)^{2(n-k)} m^{-(1/2+o(1))a_1/2} \\ &\leq \left(\frac{C}{d\eta^{17+o(1)}}\right)^{c(n-k)}. \end{aligned} \tag{3.166}$$

Combining the bounds (3.165) and (3.166), we have that for $k \in [\frac{n}{2}, (1-\eta)n]$,

$$\mathbb{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1}) \leq e^{-c\eta n} \tag{3.167}$$

if we assume $\eta \geq C_1 d^{-1/18}$ for a sufficiently large constant $C_1 > 0$.

Summing the bounds (3.113), (3.163), (3.167) over their respective ranges of k, we conclude

$$\mathbb{P}\big(\mathcal{E}_{\lfloor (1-\eta)n \rfloor}\big) \ll n^{-100}$$

as desired.

CHAPTER 4

The smallest singular value for structured random matrices

4.1 Introduction

In this chapter we turn from the random regular digraph matrix model considered in Chapters 2 and 3 to random matrices with independent but non-identically distributed entries. While in Chapter 3 our aim was to obtain a quantitative bound on the probability that a random matrix is singular, here we aim for quantitative lower tail estimates on the smallest singular value. Recall that the singular values of an $n \times n$ matrix M are the eigenvalues of $\sqrt{M^*M}$, which we arrange in non-increasing order:

$$s_1(M) \geq \cdots \geq s_n(M) \geq 0.$$

Note that $s_n(M) > 0$ if and only if M is invertible. Informally we say that M is "well-invertible" if $s_n(M)$ is well-separated from zero. We have $s_1(M) = \|M\|$, and if M is invertible then $s_n(M) = \|M^{-1}\|^{-1}$ (here and throughout this chapter we write $\|M\|$ for the operator norm of a matrix M). Of particular interest in numerical linear algebra is the ratio $\kappa(M) = s_1(M)/s_n(M) = \|M\|\|M^{-1}\|$, called the *condition number*, which is useful for quantifying the sensitivity of matrix algorithms to small changes in the inputs.

The problem of estimating the singular values and condition number for a random matrix was considered by von Neumann and Goldstine in the 1940s [vNG47]. In an influential survey on the efficiency of algorithms, Smale asked for the asymptotic size of the condition number of a random $n \times n$ Gaussian matrix G [Sma85]. This problem was addressed by Edelman in his PhD thesis [Ede89, Ede88], who among other things proved

$$\mathbb{E}\log\kappa(G) = \log n + a + o(1) \tag{4.1}$$

for an explicit numerical constant a. He also obtained the limiting distribution for the rescaled smallest singular value $\sqrt{n}s_n(G)$.

Since then there has been much work on the singular values of random matrices with more general entry distributions. In the sequel we write $X = (\xi_{ij})$ for an $n \times n$ matrix with entries identically distributed to a real or complex random variable ξ which is centered and has unit variance, and call such X an "iid matrix". From the works [BSY88, YBK88] it is known that the rescaled largest singular value $\frac{1}{\sqrt{n}}s_1(X)$ has an almost sure limit if and only if ξ has finite fourth moment (strictly speaking this result requires a coupling of the matrices across n, where one first draws an infinite array of iid copies of the variable ξ, then takes each member of the sequence X_n to be the top-left $n \times n$ corner of the array). Soshnikov proved that if one further assumes that ξ is *sub-Gaussian*, i.e.

$$(\mathbb{E}\,|\xi|^p)^{1/p} \le C\sqrt{p} \quad \forall p \ge 1 \tag{4.2}$$

and some constant C, then $s_1(X)$, appropriately rescaled, converges in law to the Tracy–Widom distribution [Sos02]. More recently, Tao and Vu proved a universality theorem for the smallest singular value, showing that the limiting distribution proved by Edelman extends to hold for all iid matrices with entries ξ_{ij} having finite C_0-th moment for some sufficiently large constant C_0 [TV10a].

In connection with problems in computer science and the theory of Banach spaces, there has been considerable interest in obtaining non-asymptotic bounds on the largest and smallest singular values of random matrices with independent entries. Seginer [Seg00] proved that if X is an $n \times m$ matrix with iid centered entries, then

$$\mathbb{E}\,\|X\| \ll \mathbb{E}\max_{i\in[n]}\sqrt{\sum_{j\in[m]}\xi_{ij}^2} + \mathbb{E}\max_{j\in[m]}\sqrt{\sum_{i\in[n]}\xi_{ij}^2}. \tag{4.3}$$

Latała [Lat05] subsequently gave a bound (Theorem 4.6.7 below) that holds for matrices with independent but not necessarily identically distributed entries, with entries having finite fourth moment.

Lower tail estimates for the smallest singular value of square iid matrices proved to be more difficult. The first breakthrough was made by Rudelson [Rud08], who showed that if X has iid real-valued sub-Gaussian entries, then

$$\mathbb{P}\big(s_n(X) \leq tn^{-3/2}\big) \ll t + n^{-1/2} \tag{4.4}$$

for all $t \geq 0$, where the implied constant depends on the constant C in 4.2. This result was subsequently improved in work of Rudelson and Vershynin [RV08], who showed that under the same setup,

$$\mathbb{P}\big(s_n(X) \leq tn^{-1/2}\big) \ll t + e^{-cn} \tag{4.5}$$

where the implied constant and $c > 0$ depend on the constant C in (4.2). Around the same time, in their work on the *circular law* for iid matrices (discussed in Chapter 5) Tao and Vu [TV08] proved bounds of arbitrary polynomial order for complex iid matrices, assuming only finiteness of the second moment – see Theorem 4.1.5 below.

In this chapter we consider the problem of bounding the smallest singular value for *structured* random matrices, which we define as follows. (While all of our results concern square random matrices, we give the definition for the general rectangular case as we will often need to consider rectangular submatrices in the proofs.)

Definition 4.1.1 (Structured random matrix)**.** Let $A = (a_{ij})$ and $B = (b_{ij})$ be deterministic $n \times m$ matrices with $a_{ij} \in [0, 1]$ and $b_{ij} \in \mathbb{C}$ for all i, j. Let $X = (\xi_{ij})$ be an $n \times m$ matrix with independent entries, all identically distributed to a complex random variable ξ with mean zero and variance one. We refer to ξ as the *atom variable*. Put

$$M = A \circ X + B = (a_{ij}\xi_{ij} + b_{ij})_{i,j=1}^n \tag{4.6}$$

where $\circ$ denotes the matrix Hadamard product. We refer to A as the *standard deviation* *profile* for M, and B as the *mean profile* for M.

Remark 4.1.2. The assumption that the entries of M are shifted scalings of random variables ξ_{ij} having a common distribution is made for convenience. We expect the proofs can be modified to cover general matrices with independent entries having specified means and variances (possibly with some additional moment hypotheses), but we do not pursue this here.

As a concrete example one can consider a centered non-Hermitian band matrix, where one sets $a_{ij} \equiv 0$ for $|i-j|$ exceeding some bandwidth parameter w.

The singular value distributions for structured random matrices have been studied in connection with wireless MIMO networks [TV04, HLN07]. The limiting spectral distributions and spectral radius for certain structured random matrices have been used to model the dynamical properties of neural networks [RA06, ARS15]. The limiting spectral distribution for a general class of centered structured random matrices is investigated the forthcoming paper [CHNR] with Hachem, Najim and Renfrew. This work required bounds on the smallest singular value for shifts of the random matrix by scalar multiples of the identity, which motivated us to prove the results in this chapter (in particular, Corollary 4.1.21 below is a key input for the results in [CHNR]).

Recent works of Bandeira–van Handel [BvH] and van Handel [vH] have provided non-asymptotic bounds on the largest singular value of centered structured random matrices which are sharp in many cases. In [vH] van Handel puts forward the following conjecture, which he attributes to Latała. The conjecture remains open as of this writing.

Conjecture 4.1.3. *The bound* (4.3) *holds for arbitrary centered matrices with independent entries having finite second moment.*

The picture for the smallest singular value of structured random matrices is far less complete – in particular, it is not clear what is the correct analogue of Conjecture 4.1.3. Here we content ourselves with identifying sufficient conditions on the matrices A, B and the distribution of ξ for a random matrix M as in (4.6) to be well-invertible with high probability. More specifically, we seek to address the following:

Problem 4.1.4. *Let M be an $n \times n$ random matrix as in Definition 4.1.1. Under what assumptions on the standard deviation and mean profiles A, B and the distribution of the atom variable ξ do we have*

$$\mathbb{P}\left(s_n(M) \leq n^{-\beta}\right) = O(n^{-\alpha}) \tag{4.7}$$

for some $\alpha, \beta > 0$?

For applications to the limiting spectral distribution of centered random matrices with a standard deviation profile, the case that B is a multiple $-z\sqrt{n}$ of the identity for some fixed $z \in \mathbb{C}$ is of particular interest.

4.1.1 Previous results

In this subsection we give an overview of what is currently known about Problem 4.1.4.

For the case of a constant standard deviation profile A and essentially arbitrary mean profile B we have the following result of Tao and Vu, which was the key technical component of their proof of the circular law for iid matrices [TV08, TV10b].

Theorem 4.1.5 (Shifted iid matrix [TV08])**.** *Let X be an $n \times n$ matrix with iid entries $\xi_{ij} \in \mathbb{C}$ having mean zero and variance one. For any $\alpha, \gamma > 0$ there exists $\beta > 0$ such that for any fixed (deterministic) $n \times n$ matrix B with $\|B\| \leq n^{\gamma}$,*

$$\mathbb{P}\left(s_n(X + B) \leq n^{-\beta}\right) = O_{\alpha,\gamma}(n^{-\alpha}). \tag{4.8}$$

A stronger version of the above bound was established earlier by Sankar, Spielman and Teng for the case that X has iid standard Gaussian entries [SST06]. For the case that $B = 0$, the bound (4.5) of Rudelson and Vershynin is superior to (4.8), but requires the stronger assumption that the entries are real-valued and sub-Gaussian (we remark that their proof extends in a routine manner to allow an arbitrary shift B with $\|B\| = O(\sqrt{n})$).

When the entries of M have bounded density the problem is much simpler. The following is easily obtained from the argument in [BC12, Section 4.4].

Proposition 4.1.6 (Matrix with entries having bounded density [BC12])**.** *Let M be an $n \times n$ random matrix with independent entries having density on $\mathbb{C}$ or $\mathbb{R}$ uniformly bounded by $\varphi > 1$. For every $\alpha > 0$ there is a $\beta = \beta(\alpha, \varphi) > 0$ such that*

$$\mathbb{P}\left(s_n(M) \le n^{-\beta}\right) = O(n^{-\alpha}). \tag{4.9}$$

Note that above we make no assumptions on the moments of the entries of M – in particular, they may have heavy tails. The following result of Bordenave and Chafaï (Lemma A.1 in [BC12]) relaxes the hypothesis of continuous distributions from Proposition 4.1.6 while still allowing for heavy tails, but comes at the cost of a worse probability bound.

Proposition 4.1.7 (Heavy-tailed matrix with non-degenerate entries [BC12])**.** *Let Y be an $n \times n$ random matrix with independent entries $\eta_{ij} \in \mathbb{C}$. Suppose that for some $p, r, \sigma_0 > 0$ we have that for all $i, j \in [n]$,*

$$\mathbb{P}(|\eta_{ij}| \le r) \ge p, \qquad \operatorname{Var}(\eta_{ij}\, \mathbb{1}(|\eta_{ij}| \le r)) \ge \sigma_0^2. \tag{4.10}$$

For any $s \ge 1$, $t \ge 0$, and any fixed $n \times n$ matrix B we have

$$\mathbb{P}\left(s_n(Y+B) \le \frac{t}{\sqrt{n}},\ \|Y+B\| \le s\right) \ll_{p,r,\sigma_0} \sqrt{\log s}\left(ts^2 + \frac{1}{\sqrt{n}}\right). \tag{4.11}$$

The non-degeneracy conditions (4.10) do not allow for some entries to be deterministic. Litvak and Rivasplata [LR12] obtained a lower tail estimate of the form (4.7) for centered random matrices having a sufficiently small constant proportion of entries equal to zero deterministically. Later we will state results (Theorems 4.1.14 and 4.1.24 below) which give such bounds for matrices having all but an arbitrarily small proportion of entries deterministic.

Here we pause to remark that all of the above results give lower tail bounds for $s_n(M)$ that are either uniform in the shift B under the size constraint $\|B\| = n^{O(1)}$, or extend in a routine way to hold uniformly for $\|B\| = O(\sqrt{n})$, i.e.

$$\sup_{B\in\mathcal{M}_n(\mathbb{C}):\, \|B\|\le n^C} \mathbb{P}\left(s_n(A\circ X + B) \le n^{-\beta}\right) = O(n^{-\alpha}). \tag{4.12}$$

for some constant $C > 0$. Such bounds can be viewed as matrix analogues of classical anti-concentration (or "small ball") bounds of the form

$$\sup_{z\in\mathbb{C}} \mathbb{P}(|S_n - z| \le r) = o_{r\to 0}(1) + o_{n\to\infty}(1) \tag{4.13}$$

where S_n is a sequence of scalar random variables (such as the normalized partial sums of an infinite sequence of iid variables). In fact, bounds of the form (4.13) are a central ingredient in the proofs of estimates (4.12). Roughly speaking, the translation invariance of (4.13) causes the uniformity in the shift B in (4.12) to come for free once one can handle the centered case $B = 0$ (the assumption $\|B\| = n^{O(1)}$ is needed to have some continuity of the map $u \mapsto \|Mu\|$ on the unit sphere in order to apply a discretization argument).

In light of this we may pose the following:

Sub-problem 4.1.8. *Let M be an $n\times n$ random matrix as in Definition 4.1.1, and let $\gamma > 0$. Under what assumptions on the standard deviation profile A and the distribution of the atom variable ξ does a lower tail estimate of the form*

$$\mathbb{P}\left(s_n(M) \le n^{-\beta}\right) = O(n^{-\alpha}) \tag{4.14}$$

hold for some $\alpha, \beta > 0$, uniformly for $\|B\| \le n^\gamma$?

The following simple observiation puts a clear limitation on the standard deviation profiles A for which we can expect to have (4.14).

Observation 4.1.9. *Suppose that $A = (a_{ij})$ has a $k\times m$ sub-matrix of zeros for some k, m with $k + m > n$. Then $A\circ X$ is singular with probability 1. Thus,* (4.14) *fails (by taking $B = 0$) for any $\alpha, \beta > 0$.*

The following result of Rudelson and Zeitouni [RZ16] shows that, for the case of Gaussian matrices, Observation 4.1.9 is essentially the only obstruction to obtaining (4.14). To state their result we need to set up some graph theoretic notation, which is only a slightly generalized form of notation introduced in Section 2.1.5. We will use this notation repeatedly throughout the chapter.

To a non-negative $n \times m$ matrix $A = (a_{ij})$ we associate a bipartite graph $\Gamma_A = ([n], [m], E_A)$, with $(i, j) \in E_A$ if and only if $a_{ij} > 0$. For a row index $i \in [n]$ we denote by

$$\mathcal{N}_A(i) = \{j \in [m] : a_{ij} > 0\} \tag{4.15}$$

its neighborhood in Γ_A. Thus, the neighborhood of a column index $j \in [m]$ is denoted $\mathcal{N}_{A^\mathsf{T}}(j)$. Given sets of row and column indices $I \subset [n], J \subset [m]$, we define the associated *edge count*

$$e_A(I, J) := |\{(i, j) \in [n] \times [m] : a_{ij} > 0\}|. \tag{4.16}$$

We will generally want to work with the graph that only puts an edge (i, j) when a_{ij} exceeds some fixed cutoff parameter $\sigma_0 > 0$. Thus, we denote by

$$A(\sigma_0) = (a_{ij} 1_{a_{ij} \geq \sigma_0}) \tag{4.17}$$

the matrix which thresholds out entries smaller than σ_0.

Rudelson and Zeitouni obtain control on the smallest singular value of Gaussian matrices whose associated matrix of standard deviations $A = (a_{ij})$ satisfies the following expansion-type condition.

Definition 4.1.10 (Broad connectivity)**.** Let $A = (a_{ij})$ be an $n \times m$ matrix with non-negative entries. For $\delta, \nu \in (0, 1)$, we say that A is (δ, ν)-*broadly connected* if

1. $|\mathcal{N}_A(i)| \geq \delta m$ for all $i \in [n]$;

2. $|\mathcal{N}_{A^\mathsf{T}}(j)| \geq \delta n$ for all $j \in [m]$;

3. $|\mathcal{N}^{(\delta)}_{A^\mathsf{T}}(J)| \geq \min(n, (1 + \nu)|J|)$ for all $J \subset [m]$

where we define the set of δ-*broadly connected* neighbors of a set $J \subset [m]$ to be

$$\mathcal{N}_{A^\mathsf{T}}^{(\delta)}(J) = \{i \in [n] : |\mathcal{N}_A(i) \cap J| \geq \delta |J|\}. \tag{4.18}$$

Theorem 4.1.11 (Gaussian matrix with broadly connected profile [RZ16])**.** *Let G be an $n \times n$ matrix with iid standard real Gaussian entries, and let A be an $n \times n$ matrix with entries $a_{ij} \in [0,1]$ for all i, j. With notation as in* (4.17)*, assume that $A(\sigma_0)$ is (δ, ν)-broadly connected for some $\sigma_0, \delta, \nu \in (0,1)$. Let $K \geq 1$, and let B be a fixed $n \times n$ matrix with $\|B\| \leq K\sqrt{n}$. Then for any $t \geq 0$,*

$$\mathbb{P}\left(s_n(A \circ G + B) \leq tn^{-1/2}\right) \ll K^{O(1)}t + e^{-cn} \tag{4.19}$$

where the implied constants and $c > 0$ depend only on δ, ν, σ_0.

Note that the assumption of broad connectivity in a sense gives us an "epsilon of separation" from the bad example of Observation 4.1.9. Thus, except for the assumption that the entries are Gaussian, Theorem 4.1.11 provides a satisfactory answer to Sub-problem 4.1.8.

Definition 4.1.10 includes many standard deviation profiles of interest, such as band matrices:

Proposition 4.1.12 (Band profile)**.** *Let $\sigma_0 \in (0,1)$ and $0 \leq w \leq n-1$, and suppose $A = (a_{ij})_{i,j=1}^n$ is such that $a_{ij} \geq \sigma_0$ for all i, j with* $\min(|i-j|, n-|i-j|) \leq w$*. Then $A(\sigma_0)$ is (δ, ν)-broadly connected with $\delta = \nu = w/4n$.*

Remark 4.1.13. The same holds with slightly worse parameters if we only require $a_{ij} \geq \sigma_0$ when $|i-j| \leq w$.

Proof. For all $i \in [n]$ we have $|\mathcal{N}_{A(\sigma_0)}(i)|, |\mathcal{N}_{A^\mathsf{T}(\sigma_0)}(i)| \geq 2w+1$, so it only remains to verify the third property in Definition 4.1.10. From here on we abbreviate $\mathcal{N}(i) = \mathcal{N}_{A(\sigma_0)}(i)$ and $\mathcal{N}^{(\delta)}(J) = \mathcal{N}_{A^\mathsf{T}(\sigma_0)}^{(\delta)}(J)$.

Define $f : [n] \to [0,1]$ by $f(i) = |J \cap \mathcal{N}(i)|/(2w+1)$. By double counting we have

$$\sum_{i=1}^{n} f(i) = |J|. \tag{4.20}$$

On the other hand, we have the discrete derivative bound

$$|f(i) - f(i-1)| \leq 1/(2w+1) \tag{4.21}$$

(here is where we use the band structure of the matrix). Suppose that

$$|\mathcal{N}^{(\delta)}(J)| = |\{i : f(i) \geq \delta|J|/n\}| < (1+\nu)|J|.$$

From (4.21) we can bound

$$\sum_{i \in \mathcal{N}^{(\delta)}(J)} f(i) \leq (2w+1)\left(\frac{\delta|J|}{n} + \frac{1}{2}\right) + \Big((1+\nu)|J| - (2w+1)\Big).$$

Together with (4.20) we find

$$\begin{aligned} |J| &= \sum_{i \notin \mathcal{N}^{(\delta)}(J)} f(i) + \sum_{i \in \mathcal{N}^{(\delta)}(J)} f(i) \\ &< |J|(1+\delta+\nu) - \left(\frac{1}{2} - \delta\right)2w \end{aligned}$$

and taking $\delta = \nu = w/4n$ leads to a contradiction. $\square$

4.1.2 New results

Our first result removes the Gaussian assumption from Theorem 4.1.11, though at the cost of a worse probability bound.

Theorem 4.1.14 (General matrix with broadly connected profile)**.** *Let $M = A \circ X + B$ be an $n \times n$ matrix as in Definition 4.1.1, and assume that $A(\sigma_0)$ is (δ, ν)-broadly connected for* *some $\sigma_0, \delta, \nu \in (0,1)$. Let $K \geq 1$. For any $t \geq 0$,*

$$\mathbb{P}\left(s_n(M) \leq \frac{t}{\sqrt{n}}, \, \|M\| \leq K\sqrt{n} \right) \ll t + \frac{1}{\sqrt{n}} \tag{4.22}$$

where the implied constant depends on K, δ, ν, σ_0, and the distribution of ξ.

Remark 4.1.15. While we have stated no additional moment assumptions on the atom variable ξ other than finite variance, the restriction to the event $\{\|M\| \leq K\sqrt{n}\}$ implicitly requires us to assume at least four finite moments to deduce $\mathbb{P}(s_n(M) \leq t/\sqrt{n}) \ll t + o(1)$. The dependence of the implied constant in (4.22) on the distribution of ξ comes from a certain Fourier-analytic reduction we make in Section 4.2. Specifically, the dependence is through the parameter κ in Lemma 4.2.6.

Remark 4.1.16. We expect that the probability bound in (4.22) can be improved by making use more advanced tools of Littlewood–Offord theory (developed in [TV08, RV08]), though it appears that these tools cannot be applied in a straightforward manner. In the interest of keeping this chapter of reasonable length we have not pursued this matter here. We also note that under the assumption that the atom variable ξ has a bounded density on $\mathbb{C}$ or $\mathbb{R}$ a probability bound similar to the one in Theorem 4.1.11 can be obtained by following the argument from [RZ16] with minor modifications.

In light of Observation 4.1.9, Theorem 4.1.14 gives an essentially optimal answer to Sub-problem 4.1.8 for *dense* random matrices. It would be interesting to establish a version of this result that allows for only a proportion $o(1)$ of the entries to be random. (As the proof of Theorem 4.1.14 is quantitative, one could obtain such a result simply by making the dependence on parameters in the bound (4.22) more explicit. However, this would only allow for a slight polynomial decay in the density of the matrix A, while we expect a version of the above theorem to hold when A has density as small $(\log^{O(1)} n)/n$.)

Having addressed Sub-problem 4.1.8, we now ask whether we can further relax the assumptions on the standard deviation profile A by assuming more about the mean profile B. In particular, can we make assumptions on B that give (4.7) while allowing $A \circ X$ to be singular deterministically (as in Observation 4.1.9)?

Of course, a trivial example where this is the case is to take $A = 0$ and B any invertible matrix. Another easy example is to take take B to be *very well-invertible*, with $s_n(B) \geq K_0\sqrt{n}$ for a large constant $K_0 > 0$ (for instance, take $B = K_0\sqrt{n}\,\mathrm{I}$, where I is the identity matrix). Indeed, standard estimates for the operator norm of random matrices with centered entries (cf. Section 4.6.2) give $\|A \circ X\| = O(\sqrt{n})$ with high probability provided the atom variable ξ satisfies some additional moment hypotheses. It then follows from the triangle inequality that $s_n(M) \gg \sqrt{n}$ with high probability if K_0 is sufficiently large.

The problem becomes non-trivial when we allow B to have singular values of size $\varepsilon\sqrt{n}$ for small $\varepsilon > 0$ and A as in Observation 4.1.9. In this case any proof of a lower tail estimate of the form (4.7) must depart significantly from the proofs of the results in the previous section by making use of arguments which are not translation invariant.

Our main result shows that when the mean profile B is a *diagonal* matrix with smallest entry at least an arbitrarily small (fixed) multiple of $\sqrt{n}$, then we do not need to assume anything further about the standard deviation profile A. In the following we denote the L^p norm of the atom variable ξ in Definition 4.1.1 by

$$\mu_p := (\mathbb{E}\,|\xi|^p)^{1/p}. \tag{4.23}$$

Theorem 4.1.17 (Centered random matrix with small diagonal perturbation)**.** *Fix arbitrary $r_0 \in (0, \frac{1}{2}]$, $K_0 \in [1, \infty)$, and let $Z = \mathrm{diag}(z_i)_{i=1}^n$ be any complex diagonal matrix satisfying*

$$|z_i| \in [r_0, K_0] \quad \forall i \in [n].$$

Let M be an $n \times n$ random matrix as in Definition 4.1.1 with $B = Z\sqrt{n}$, and assume $\mu_{4+\eta} < \infty$. There are $\alpha(\eta) > 0$ and $\beta(r_0, \eta, \mu_{4+\eta}) > 0$ such that

$$\mathbb{P}\left(s_n(M) \leq n^{-\beta}\right) = O(n^{-\alpha}) \tag{4.24}$$

where the implied constant depends only on r_0, K_0 and the distribution of ξ.

Remark 4.1.18. As in Theorem 4.1.14, the implied constant in (4.24) depends on the distribution ξ through the value of κ from Lemma 4.2.6 below. Thus, to be more specific, the implied constant depends only on r_0, K_0, κ and $\mu_{4+\eta}$.

Remark 4.1.19 (Moment assumption). The reason for our assumption of $4+\eta$ finite moments is so that any submatrix of M of dimensions linear in n will have operator norm $O(\sqrt{n})$ with high probability, which is crucial for our argument. We also make use of a result of Vershynin (Theorem 4.6.8 below) on the operator norm of products of random matrices which assumes $4+\eta$ moments. The dependence of α on η is of the form $\alpha = c\min(1,\eta)$ for a constant $c > 0$ (which we take to be 1/9 in the proof). If we were to assume η is a sufficiently large constant, i.e. if ξ has finite pth moment for a sufficiently large constant p, then our proof would allow us to take any fixed $\alpha < 1/2$.

Remark 4.1.20 (Dependence on r_0). The dependence of β on $\nu_{4+\eta}$ and r_0 given by our proof is very bad: a tower exponential (i.e. a tower of twos $2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}$) of height

$$O_\eta(1)\exp(O(\mu_{4+\eta}/r_0)^{O(1)}). \tag{4.25}$$

(The factor $O_\eta(1)$ comes from Vershynin's bound mentioned in the previous remark – we do not know the precise dependence on η, but we expect it is relatively mild.) This is due to our use of Szemerédi's regularity lemma (specifically, a version for directed graphs due to Alon and Shapira – see Lemma 4.6.2). It would be interesting to obtain a version of Theorem 4.1.17 with a better dependence of β on the parameters.

As we remarked above, the case of a diagonal mean profile is of special interest for the problem of proving convergence of the empirical spectral distribution of centered random matrices with a variance profile.

Corollary 4.1.21 (Scalar shift of a centered random matrix). *Let $X = (\xi_{ij})$ be an $n \times n$ matrix whose entries are iid copies of a centered complex random variable ξ having unit variance and $\mu_{4+\eta} < \infty$ for some fixed $\eta > 0$. Let $A = (a_{ij})$ be a fixed $n \times n$ non-negative matrix with entries uniformly bounded by $\sigma_{\max} < \infty$. Put $Y = \frac{1}{\sqrt{n}}A \circ X$, and fix an arbitrary*

$z \in \mathbb{C} \setminus \{0\}$. *There are constants* $\alpha = \alpha(\eta) > 0$ *and* $\beta = \beta(z, \eta, \mu_{4+\eta}, \sigma_{\max}) > 0$ *such that*

$$\mathbb{P}\left(s_n(Y - z) \leq n^{-\beta}\right) = O(n^{-\alpha}) \tag{4.26}$$

where the implied constant depends only on z, $\sigma_{\max}$ *and the distribution of* ξ.

(In (4.26) we identify z with a scalar multiple of the identity matrix.) While our main motivation was to handle diagonal perturbations of centered random matrices, we conjecture that Theorem 4.1.17 extends to matrices as in Definition 4.1.1 with more general mean profiles B:

Conjecture 4.1.22. *Theorem 4.1.17 continues to hold for* $B \in \mathcal{M}_n(\mathbb{C})$ *not necessarily diagonal, assuming* $\frac{1}{\sqrt{n}} s_i(B) \in [r_0, K_0]$ *for all* $1 \leq i \leq n$.

4.1.3 Ideas of the proofs

Regular partitions of graphs, and the Schur complement formula

We start with a brief description of the main ideas in the proof of Theorem 4.1.17. As with Theorem 4.1.14, the key is to associate the standard deviation profile A with a graph. Since we want the diagonal of M to be preserved under relabeling of vertices will will associate A with a directed graph (digraph) which puts an edge $i \to j$ whenever a_{ij} exceeds some small threshold $\sigma_0 > 0$. Since A has no special connectivity structure *a priori*, we will apply a version of Szemerédi's regularity lemma for digraphs (Lemma 4.6.2) to partition the vertex set $[n]$ into a bounded number of parts of equal size $I_1, \dots, I_m$, together with a small set of "bad" vertices I_{bad}, such that for most $(k, l) \in [m]^2$ the subgraph on $I_k \cup I_l$ enjoys certain "pseudorandomness" properties. These properties will not be quite strong enough to control the smallest singular value of the corresponding submatrix M_{I_k,I_l} of M, but we can apply a "cleaning" procedure (as it is called in the extremal combinatorics literature) to remove a small number of bad vertices from each part in the partition (which we add to I_{bad}), after which we will be able to control $s_{\min}(M_{I_k,I_l})$ for most $(k, l) \in [m]^2$. The specific pseudorandomness properties and corresponding bound on the smallest singular value are given in Definition 4.1.23 and Theorem 4.1.24 below.

The task is then to lift this control on the invertibility of submatrices to the whole matrix M. The key tool here is the *Schur complement formula* (see Lemma 4.6.4) which allows us to control the smallest singular value of a block matrix

$$\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$$

assuming some control on the smallest singular values of (perturbations of) the diagonal block submatrices M_{11}, M_{22} and on the operator norm of the off-diagonal submatrices M_{12}, M_{21}. The bound on the smallest singular value of the whole matrix is somewhat degraded, but this is acceptable as we will only apply this a bounded number of times. Thus, we will want to find a generalized diagonal of block submatrices which are well-invertible under additive perturbations.

At this point it is best to think of the partition $I_1, \dots, I_m$ as inducing a "macroscopic scale" digraph $\mathcal{R} = ([m], E)$ (which is called the *reduced digraph* in extremal combinatorics) that puts an edge $(k, l) \in E$ whenever the corresponding submatrix A_{I_k, I_l} is pseudorandom and sufficiently dense. If we can cover the vertices of $\mathcal{R}$ with vertex-disjoint directed cycles, then we have found a generalized diagonal of submatrices of M with the desired properties.

Of course, it may be the case that $\mathcal{R}$ cannot be covered by disjoint cycles. For instance, if A were to have all ones in the first $n/2$ columns and all zeros in the last $n/2$ columns then roughly half of the vertices of $\mathcal{R}$ would have no incoming edges. This is where we make crucial use of the diagonal perturbation $Z\sqrt{n}$ (indeed, without this perturbation M would be singular in this example). The top left $n/2 \times n/2$ submatrix of M is dense, and we can apply Theorem 4.1.24 to control its smallest singular vale. The bottom right $n/2 \times n/2$ submatrix is a diagonal matrix with diagonal entries of size at least $r_0\sqrt{n}$, and hence its smallest singular value is at least $r_0\sqrt{n}$. This argument even allows for the bottom right submatrix of A to be nonzero but sufficiently sparse: we can use the triangle inequality and standard bounds on the operator norm of sparse random matrices to argue that the smallest singular value of the bottom right submatrix is still of order $\gg r_0\sqrt{n}$.

We handle the general case as follows. We greedily cover as many of the vertices of $\mathcal{R}$ as we can with disjoint cycles – call this set of vertices $U_{\text{cyc}} \subset [m]$. At this point we have either covered the whole graph (and we are done) or the graph on the remaining vertices U_{free} is cycle-free. This means that the vertices of $\mathcal{R}$ can be relabeled so that its adjacency matrix is upper-triangular on $U_{\text{free}} \times U_{\text{free}}$. Write $J_{\text{cyc}} = \bigcup_{k\in U_{\text{cyc}}} I_k$, $J_{\text{free}} = \bigcup_{k\in U_{\text{free}}} I_k$ and denote the corresponding submatrices of A on the diagonal by $A_{\text{cyc}}, A_{\text{free}}$, and likewise for M. We thus have a relabeling of $[n]$ under which A_{free} is upper triangular. Crucially, this relabeling has preserved the diagonal, so the submatrix M_{free} is a diagonal perturbation of an upper-triangular random matrix. We then show that such a matrix has smallest singular value of order $\gg_{r_0} \sqrt{n}$ with high probability. With another application of the Schur complement bound we can combine the control on the submatrices $M_{\text{cyc}}, M_{\text{free}}$ (along with standard bounds on the operator norm for the off-diagonal blocks) to conclude the proof. (Actually, the bad set I_{bad} of rows and columns requires some further arguments, but we do not discuss this issue here.)

This concludes the high level description of the proof of Theorem 4.1.17. We only remark that the above partitioning and cleaning procedures generate various error terms and residual submatrices (such as the vertices in I_{bad}, or the small proportion of pairs (I_k, I_l) which are not pseudorandom enough). As the smallest singular value is notoriously sensitive to perturbations, it will take some care to control these terms. We will use some high-powered tools such as bounds on the operator norm of sparse random matrices and products of random matrices due to Latała and Vershynin – see Section 4.6.2.

Well-invertibility from connectivity assumptions

Now we state the specific pseudorandomness condition on a standard deviation profile under which we have good control on the smallest singular value. While "pseudorandom" generally means that the edge distribution in a graph is close to uniform on a range of scales, we will only need control from below on the edge densities (morally speaking, we want the matrix

A to be as far as possible from the zero matrix, the most poorly invertible matrix). The following one-sided pseudorandomness condition is taken from the combinatorics literature (see [KS96, Definition 1.6]).

Definition 4.1.23 (Super-regularity)**.** Let A be an $n \times m$ matrix with non-negative entries. For $\delta, \varepsilon \in (0,1)$, we say that A is (δ,ε)-*super-regular* if the following hold:

1. $|\mathcal{N}_A(i)| \geq \delta m$ for all $i \in [n]$;

2. $|\mathcal{N}_{A^\mathsf{T}}(j)| \geq \delta n$ for all $j \in [m]$;

3. $e_A(I,J) \geq \delta |I||J|$ for all $I \subset [n], J \subset [m]$ such that $|I| \geq \varepsilon n$ and $|J| \geq \varepsilon m$.

The reader should compare this condition with Definition 4.1.10. Conditions (1) and (2) are are the same in both definitions, while it is not hard to see that condition (3) above implies

$$|\mathcal{N}_{A^\mathsf{T}}^{(\delta)}(J)| \geq (1-\varepsilon)n \tag{4.27}$$

whenever $|J| \geq \varepsilon n$ (with notation as in (4.18)), which is stronger than condition (3) in Definition 4.1.10 for such J. On the other hand, conditions (1) and (2) imply that $|\mathcal{N}_{A^\mathsf{T}}^{(\sqrt{\delta}/2)}(J)| \geq \frac{1}{2}\delta n$ for any $J \subset [m]$ (see Lemma 4.3.6), so super-regularity is stronger than broad connectivity for ε, η sufficiently small depending on δ.

Theorem 4.1.24 (Matrix with super-regular profile)**.** *Let $M = A \circ X + B$ be an $n \times n$ matrix as in Definition 4.1.1. Assume that $A(\sigma_0)$ is (δ,ε)-super-regular for some $\delta, \sigma_0 \in (0,1)$ and $0 < \varepsilon < c_1 \delta \sigma_0^2$ with $c_1 > 0$ a sufficiently small constant. For any $\gamma \geq 1/2$ there exists $\beta = O(\gamma^2)$ such that*

$$\mathbb{P}\left(s_n(M) \leq n^{-\beta},\, \|M\| \leq n^{\gamma}\right) \ll \sqrt{\frac{\log n}{n}} \tag{4.28}$$

where the implied constant depends on γ, δ, σ_0 and the distribution of ξ.

Note that Theorem 4.1.24 allows for a mean profile B of arbitrary polynomial size in operator norm, whereas in Theorem 4.1.14 we could only allow $\|B\| = O(\sqrt{n})$. The ability to handle such large perturbations will be crucial in the proof of Theorem 4.1.17, as the iterative application of the Schur complement bound discussed above will lead to perturbations of increasingly large polynomial order.

We defer discussion of the key technical ideas for Theorem 4.1.14 and Theorem 4.1.24 to Sections 4.3 and 4.4. We only mention here that our proof of Theorem 4.1.24 makes crucial use of a new "entropy reduction" argument, which allows us to control the event that $\|Mu\|$ is small for some u in certain portions of the sphere S^{n-1} by the event that this holds for some u in a random net of relatively low cardinality. The argument uses an improvement by Spielman and Srivastava [SS12] of the classic Restricted Invertibility Theorem due to Bourgain and Tzafriri [BT87] – see Section 4.3 for details.

4.1.4 Organization of the chapter

The rest of this chapter is organized as follows.

Sections 4.2, 4.3 and 4.4 are devoted to the proofs of Theorems 4.1.14 and 4.1.24. We prove these theorems in parallel as they involve many similar ideas. In Section 4.2 we collect some standard lemmas on anti-concentration for random walks and products of random matrices with fixed vectors, along with some facts about ρ-nets. In Section 4.3 show that random matrices as in Theorems 4.1.14 and 4.1.24 are well-invertible over sets of "compressible" vectors in the unit sphere, and in Section 4.4 we establish control over the complementary set of "incompressible" vectors to complete the proofs.

Section 4.5 builds on the results in Section 4.3 for broadly connected profiles to establish lower bounds for other "moderately small" singular values holding with high probability – see Theorem 4.5.1. This result will be a key ingredient for the proof in Chapter 5 of the circular law for signed rrd matrices. Theorem 4.5.1 and its application to the proof of the circular law follows the analogous work of Tao and Vu for iid matrices in [TV10b].

In Section 4.6 we prove Theorem 4.1.17.

4.1.5 Notation

In addition to the notation listed in Section 1.2 we will use the following conventions in the the present chapter.

For a matrix $A = (a_{ij}) \in \mathcal{M}_{n,m}(\mathbb{C})$ we will sometimes use the notation $A(i,j) = a_{ij}$. For $I \subset [n], J \subset [m]$ with elements $i_1 < \cdots < i_{|I|}$ and $j_1 < \cdots j_{|J|}$, respectively, $A_{I,J}$ denotes the $|I| \times |J|$ submatrix $(a_{i_k,j_l})_{1\le k\le |I|, 1\le l\le |J|}$. We often abbreviate $A_J := A_{J,J}$.

$\|A\|_{\mathrm{HS}}$ denotes the Hilbert–Schmidt (or Frobenius) norm of a matrix A. We will sometimes denote the smallest singular value of a square matrix M by $s_{\min}(M)$ (in situations where M is a submatrix of a larger matrix this will often be clearer than writing the index).

For $J \subset [n]$, we denote by $\mathbb{C}^J \subset \mathbb{C}^n$ (resp. $S^J \subset S^{n-1}$) the set of vectors (resp. unit vectors) in $\mathbb{C}^n$ supported on J. Given a vector $v \in \mathbb{C}^n$, we denote by $v_J \in \mathbb{C}^n$ the projection of v to the coordinate subspace $\mathbb{C}^J$. $\binom{[m]}{x}$ denotes the family of subsets of $[m]$ of size $\lfloor x \rfloor$.

When considering a random matrix M as in Definition 4.1.1, we write $\mathcal{F}_{I,J} := \langle \{\xi_{ij}\}_{i\in I, j\in J}\rangle$ for the sigma algebra of events generated by the entries $\{\xi_{ij}\}_{i\in I,j\in J}$ of X. For $I \subset [n]$ we write $\mathbb{P}_I(\cdot)$ for probability conditional on $\mathcal{F}_{[n]\setminus I,[n]}$.

Finally, we note that we will on occasion use the asymptotic notation $f = o(g)$ defined in Section 1.2, but we only do this for the sake of brevity, as all of the arguments in this chapter can be made quantitative.

4.2 Invertibility from connectivity: Preliminaries

4.2.1 Partitioning and discretizing the sphere

For the proofs of Theorems 4.1.14 and 4.1.24 we will make heavy use of ideas and notation developed in [LPRTJ05, LPR+05, Rud08, RV08] and related ideas from the local theory of Banach spaces. In particular, in order to lower bound

$$s_n(M) = \inf_{u\in S^{n-1}} \|Mu\|$$

we will partition the sphere into sets of vectors of different levels of "compressibility", which we presently define, and separately obtain control in the infimum of $\|Mu\|$ over each set.

Recall from Section 4.1.5 our notation $S^J \subset S^{m-1}$ for the set of unit vectors supported on $J \subset [m]$. For a set $T \subset \mathbb{C}^n$ and $\rho > 0$ we write T_ρ for the set of points within Euclidean distance ρ of T. We recall also the following definitions from [RZ16]. For $\theta, \rho \in (0,1)$, we define the set of *compressible vectors*

$$\mathrm{Comp}(\theta,\rho) := S^{m-1} \cap \bigcup_{J \in \binom{[m]}{\theta m}} (S^J)_\rho \tag{4.29}$$

and the complementary set of *incompressible vectors*

$$\mathrm{Incomp}(\theta,\rho) := S^{m-1} \setminus \mathrm{Comp}(\theta,\rho). \tag{4.30}$$

In words, $\mathrm{Comp}(\theta,\rho)$ is the set of unit vectors within (Euclidean) distance ρ of a vector supported on at most θm coordinates. On the other hand, incompressible vectors enjoy the following property which will lead to good anti-concentration properties for an associated random walk.

Lemma 4.2.1 (Incompressible vectors are spread, cf. [RV08, Lemma 3.4])**.** *Fix $\theta, \rho \in (0,1)$ and let $v \in \mathrm{Incomp}(\theta,\rho)$. There is a set $L^+ \subset [m]$ with $|L^+| \geq \theta m$ such that $|v_j| \geq \rho/\sqrt{m}$ for all $j \in L^+$. Moreover, for all $\lambda \geq 1$ there is a set $L \subset [m]$ with $|L| \geq (1 - \frac{1}{\lambda^2})\theta m$ such that for all $j \in L$,*

$$\frac{\rho}{\sqrt{m}} \leq |v_j| \leq \frac{\lambda}{\sqrt{\theta m}}.$$

Proof. Take $L^+ = \{j : |v_j| \geq \rho/\sqrt{m}\}$ and denote $L^- = \{j : |v_j| \leq \lambda/\sqrt{\theta m}\}$. Since v lies a distance at least ρ from any vector supported on at most θm coordinates we must have $|L^+| \geq \theta m$, which gives the first claim. On the other hand, since $v \in S^{m-1}$, by Markov's inequality we have $|L^-| \geq \theta m/\lambda^2$, so taking $L = L^+ \cap L^-$ we have $|L| \geq (1 - \frac{1}{\lambda^2})\theta m$. $\square$

We will informally refer to the coordinates of $v \in \mathrm{Incomp}(\theta,\rho)$ where $|v_j| \geq \rho/\sqrt{n}$ as the *essential support of v*.

In the remainder of this subsection we recall a couple of standard facts about nets of the sphere. For $\rho > 0$, recall that a ρ-net of a set $T \subset \mathbb{C}^m$ is a finite subset $\Sigma \subset T$ such that for all $v \in T$ there exists $v' \in \Sigma$ with $\|v - v'\| \leq \rho$. The following lemma provides a nets of reasonable cardinality for subsets of the sphere.

Lemma 4.2.2 (Metric entropy of the sphere)**.** *Let $V \subset \mathbb{C}^m$ be a k-dimensional subspace, let $T \subset V \cap S^{m-1}$, and let $\rho \in (0,1)$. Then T has a ρ-net $\Sigma \subset T$ of cardinality $|\Sigma| \leq (3/\rho)^{2k}$.*

Proof. Let $\Sigma \subset T$ be a ρ-separated (in Euclidean distance) subset that is maximal under set inclusion. It follows from maximality that Σ is a ρ-net of T. Let $\Sigma_{\rho/2}$ denote the $\rho/2$ neighborhood of Σ in V. Noting that $\Sigma_{\rho/2}$ is a disjoint union of k-dimensional Euclidean balls of radius $\rho/2$, we have

$$|\Sigma| c_k (\rho/2)^{2k} \leq \mathrm{vol}_k(\Sigma_{\rho/2}) \leq c_k (1 + \rho/2)^{2k}$$

where vol_k denotes the k-dimensional Lebesgue measure on V and c_k is the volume of the Euclidean unit ball in $\mathbb{C}^k$. The desired bound follows by rearranging. $\square$

The existence of ρ-nets of controlled size allows us to obtain uniform control from below on $\|Mu\|$ for all u drawn from a low-dimensional subset of the sphere, off a small event. The following is essentially Lemma 3.3 from [RZ16], adjusted to the current setting of matrices over $\mathbb{C}$ rather than $\mathbb{R}$.

Corollary 4.2.3 (Control by a net)**.** *Let M be an $n \times m$ random matrix with complex entries. Let $V \subset \mathbb{C}^m$ be a k-dimensional subspace, and let $T \subset V \cap S^{m-1}$. Let $K \geq 1$, and assume that for some $\rho, p_0 \in (0,1)$, for any $u \in T$ we have*

$$\mathbb{P}\left(\|Mu\| \leq \rho K \sqrt{n}\right) \leq p_0. \tag{4.31}$$

Then

$$\mathbb{P}\Big(\mathcal{B}(K) \wedge \big\{\, \exists u \in T_{\rho/4} : \|Mu\| \leq (\rho/2) K \sqrt{n} \,\big\}\Big) \leq \left(\frac{12}{\rho}\right)^{2k} \cdot p_0. \tag{4.32}$$

Proof. By Lemma 4.2.2, T has a $\rho/4$-net Σ of size at most $(12/\rho)^{2k}$. On the event that $\|M\| \le K\sqrt{n}$ and there exists $u \in T_\rho$ such that $\|Mu\| \le (\rho/2)K\sqrt{n}$, we fix such a u and let $u' \in \Sigma$ be such that $\|u - u'\| < \rho/2$ (which exists by the triangle inequality). It follows that

$$\|Mu'\| \le \|Mu\| + \|M(u-u')\| \le (\rho/2)K\sqrt{n} + K\sqrt{n} \cdot (\rho/2) = \rho K\sqrt{n}.$$

We have shown that

$$\left\{\|M\| \le K\sqrt{n} \ \text{ and } \ \exists u \in T_{\rho/4} : \|Mu\| \le (\rho/2)K\sqrt{n}\right\} \subset \left\{\exists u \in \Sigma : \|Mu\| \le \rho K\sqrt{n}\right\}.$$

The desired result now follows from taking a union bound over the at most $(12K/\rho)^{2k}$ choices of $u \in \Sigma$ and (4.31). $\square$

4.2.2 Anti-concentration for scalar random walks

In this subsection we collect some standard anti-concentration estimates for scalar random walks, which are perhaps the most central tool for proving that random matrices are (well-)invertible with high probability. In the next subsection we apply these to obtain anti-concentration estimates for products Mv of a random matrix M with a fixed vector v.

Definition 4.2.4 (Concentration probability)**.** Let ξ be a complex-valued random variable. For $v \in \mathbb{C}^n$ we let

$$S_\xi(v) = \sum_{j=1}^{n} \xi_j v_j \tag{4.33}$$

where $\xi_1, \dots, \xi_n$ are iid copies of ξ. For $r \ge 0$ we define the *concentration probability*

$$p_{\xi,v}(r) = \sup_{z \in \mathbb{C}} \mathbb{P}\left(|S_\xi(v) - z| \le r\right). \tag{4.34}$$

Throughout this section operate under the following distributional assumption on ξ.

Definition 4.2.5 (Controlled second moment (cf. [TV08, Definition 2.2]))**.** Let $\kappa \ge 1$. A complex random variable ξ is said to have *κ-controlled second moment* if one has the upper bound

$$\mathbb{E}\,|\xi|^2 \le \kappa$$

(in particular, $|\,\mathbb{E}\,\xi| \le \kappa^{1/2}$), and the lower bound

$$\mathbb{E}[\mathrm{Re}(z\xi) - a]^2\, \mathbb{1}(|\xi| \le \kappa) \ge \frac{1}{\kappa}[\mathrm{Re}(z)]^2$$

for all $z \in \mathbb{C}, a \in \mathbb{R}$.

Roughly speaking, a complex random variable ξ has controlled second moment if its distribution has a one-(real-)dimensional marginal with fairly large variance on some compact set. It turns out that we lose no generality in any of our theorems by assuming ξ has controlled second moment, due to the following lemma of Tao and Vu.

Lemma 4.2.6 (cf. [TV08, Lemma 2.4]). *Let ξ be a complex random variable with finite non-zero variance. Then there exist $\theta \in [0,1)$ and $\kappa \ge 1$ such that $e^{i\theta}\xi$ has κ-controlled second moment.*

Below we give two standard bounds on the concentration function $p_{\xi,v}(r)$ when ξ is a κ-controlled random variable and $v \in S^{n-1}$. The first gives a crude constant order bound that is uniform in $v \in S^{n-1}$:

Lemma 4.2.7 (Crude anti-concentration, cf. [TV10c, Corollary 6.3]). *Let ξ be a complex random variable with κ-controlled second moment. There is a constant $r_0 > 0$ depending only on κ such that $p_{\xi,v}(r_0) \le 1 - r_0$ for all $v \in S^{n-1}$.*

Note that Lemma 4.2.7 is sharp for the case that v is a standard basis vector. The following lemma gives an improved bound when v has small ℓ_∞ norm.

Lemma 4.2.8 (Improved anti-concentration). *Let ξ be a complex random variable that is κ-controlled for some $\kappa > 0$, and let $v \in S^{n-1}$. Then for all $r \ge 0$,*

$$p_{\xi,v}(r) \ll_\kappa r + \|v\|_\infty. \tag{4.35}$$

Remark 4.2.9. Since ξ and the coefficients of v are complex one would expect a bound with r^2 in place of r on the right hand side of (4.35), provided the the set of phases of the components of v is in some sense "genuinely two dimensional". Indeed, by a modification of the proof below one can show

$$p_{\xi,v}(r) \ll_\kappa \frac{1}{\alpha^2}(r^2 + \|v\|_\infty^2) \tag{4.36}$$

for all $r \geq 0$, provided v satisfies

$$\inf_{\theta \in \mathbb{R}} \| \operatorname{Re}(e^{i\theta} v)\| \geq \alpha \|v\| \tag{4.37}$$

for some $\alpha > 0$. The condition (4.37) was used in a more sophisticated anti-concentration bound of Rudelson and Vershynin – see [RV09, Theorem 3.3]. We will not need such improvements in the present work.

Lemma 4.2.8 can be deduced from the Berry–Esséen theorem (which is the approach taken in [LPRTJ05], for instance), but this would require ξ to have finite third moment, which we do not assume. (Generally speaking, higher moment assumptions should only be needed for proving concentration bounds as opposed to anti-concentration.) Since we could not locate a proof in the literature for the case that ξ and and the coefficients of v take values in $\mathbb{C}$ we give a proof below. We will first need to recall a couple of lemmas from [TV10c].

Lemma 4.2.10 (Fourier-analytic bound, cf. [TV10c, Lemma 6.1]). *Let ξ be a complex-valued random variable. For all $r > 0$ and any $v \in S^{n-1}$ we have*

$$p_{\xi,v}(r) \ll r^2 \int_{w \in \mathbb{C}: |w| \leq 1/r} \exp\Big(- c \sum_{j=1}^n \|wv_j\|_\xi^2 \Big) \mathrm{d}w \tag{4.38}$$

where

$$\|z\|_\xi^2 := \mathbb{E} \,\| \operatorname{Re}(z(\xi - \xi'))\|_{\mathbb{R}/\mathbb{Z}}^2, \tag{4.39}$$

ξ' is an independent copy of ξ, and $\|x\|_{\mathbb{R}/\mathbb{Z}}$ denotes the distance from x to the nearest integer.

The next lemma gives an important property enjoyed by the "norm" $\|\cdot\|_\xi$ from Lemma 4.2.10

under the assumption that ξ has κ-controlled second moment.

Lemma 4.2.11 (cf. [TV08, Lemma 5.3]). *For any $\kappa > 0$ there are constants $c_1, c_2 > 0$ such that if ξ is κ-controlled, then $\|z\|_\xi \geq c_1|\operatorname{Re}(z)|$ whenever $|z| \leq c_2$.*

Now we are ready to prove Lemma 4.2.8. We remark that similar lines were used to prove Lemma 4.2.7 in [TV10c].

Proof of Lemma 4.2.7. Let $r \geq 0$. We may assume $r \geq C_0\|v\|_\infty$ for any fixed constant $C_0 > 0$ depending only on κ. From Lemma 4.2.10,

$$p_{\xi,v}(r) \ll r^2 \int_{|w|\leq 1/r} \exp\Big(- c\sum_{j=1}^n \|wv_j\|_\xi^2\Big) \mathrm{d}w.$$

If C_0 is sufficiently large depending on κ, it follows from Lemma 4.2.11 that whenever $|w| \leq 1/r$, $\|wv_j\|_\xi \geq c_1|\operatorname{Re}(wv_j)|$, giving

$$p_{\xi,v}(r) \ll r^2 \int_{|w|\leq 1/r} \exp\Big(- c'\sum_{j=1}^n (\operatorname{Re}(wv_j))^2\Big) \mathrm{d}w$$

where c' depends only on κ. By change of variable,

$$p_{\xi,v}(r) \ll \int_{|w|\leq 1} \exp\Big(- \frac{c'}{r^2}\sum_{j=1}^n (\operatorname{Re}(wv_j))^2\Big) \mathrm{d}w. \tag{4.40}$$

Write $v_j = r_j e^{i\theta_j}$ for each $j \in [n]$. Since $v \in S^{n-1}$ we have $\sum_{j=1}^n r_j^2 = 1$. By Jensen's inequality,

$$\begin{aligned} p_{\xi,v}(r) &\ll \int_{|w|\leq 1} \exp\Big(- \frac{c'}{r^2}\sum_{j=1}^n r_j^2\big(\operatorname{Re}(we^{i\theta_j})\big)^2\Big) \mathrm{d}w \\ &\leq \int_{|w|\leq 1} \sum_{j=1}^n r_j^2 \exp\Big(- \frac{c'}{r^2}\big(\operatorname{Re}(we^{i\theta_j})\big)^2\Big) \mathrm{d}w. \end{aligned}$$

By rotational invariance the last expression is equal to

$$\sum_{j=1}^{n} r_j^2 \int_{|w|\le 1} \exp\Big(- \frac{c'}{r^2}(\mathrm{Re}(w))^2\Big)\mathrm{d}w = \int_{|w|\le 1} \exp\Big(- \frac{c'}{r^2}(\mathrm{Re}(w))^2\Big)\mathrm{d}w$$

which by direct computation is seen to be of size $O(r)$ (with implied constant depending on κ). Together with our assumption that $r \ge C_0\|v\|_\infty$ this gives (4.35). $\square$

4.2.3 Anti-concentration for the image of a fixed vector

In this subsection we boost the anti-concentration bounds for scalar random variables from the previous sections to anti-concentration for the image of a fixed vector under a random matrix. The following lemma of Rudelson and Vershynin is convenient for this task.

Lemma 4.2.12 (Tensorization, cf. [RV08, Lemma 2.2]). *Let $\zeta_1, \dots, \zeta_n$ be independent non-negative random variables.*

(a) Suppose that for some $\varepsilon_0, p_0 > 0$ and all $j \in [n]$, $\mathbb{P}(\zeta_j \le \varepsilon_0) \le p_0$. There are $c_1, p_1 \in (0,1)$ depending only on p_0 such that

$$\mathbb{P}\left(\sum_{j=1}^{n} \zeta_j^2 \le c_1 \varepsilon_0^2 n \right) \le p_1^n. \tag{4.41}$$

(b) Suppose that for some $K, \varepsilon_0 \ge 0$ and all $j \in [n]$, $\mathbb{P}(\zeta_j \le \varepsilon) \le K\varepsilon$ for all $\varepsilon \ge \varepsilon_0$. Then

$$\mathbb{P}\left(\sum_{j=1}^{n} \zeta_j^2 \le \varepsilon^2 n \right) \le (CK\varepsilon)^n \tag{4.42}$$

for all $\varepsilon \ge \varepsilon_0$.

Proof. We only prove part (a), as part (b) is given in [RV08] (note that in part (a) we have given more specific dependencies on the parameters than in [RV08]). Let $c_1 > 0$ to be taken sufficiently small depending on p_0, and let $\alpha > 0$ a sufficiently small constant to be chosen later. We have

$$\mathbb{P}\left(\sum_{j=1}^{n}|\zeta_j|^2 \le c_1\varepsilon_0^2 n\right) = \mathbb{P}\left(n - \frac{1}{c_1\varepsilon_0^2}\sum_{j=1}^{n}|\zeta_j|^2 \ge 0\right)$$
$$\le \mathbb{E}\exp\left(c_1\tau n - \frac{\alpha}{\varepsilon_0^2}\sum_{j=1}^{n}|\zeta_j|^2\right)$$
$$= e^{c_1\alpha n}\prod_{j=1}^{n}\mathbb{E}\exp\left(-\alpha|\zeta_j|^2/\varepsilon_0^2\right). \tag{4.43}$$

For arbitrary $j \in [n]$ we have

$$\begin{aligned}
\mathbb{E}\exp\left(-\alpha|\zeta_j|^2/\varepsilon_0^2\right) &= \int_0^1 \mathbb{P}\big(\exp\left(-\alpha|\zeta_j|^2/\varepsilon_0^2\right) \ge u\big)\mathrm{d}u \\
&= \int_0^\infty \mathbb{P}\big(|\zeta_j| \le s\varepsilon_0/\sqrt{\alpha}\big)\mathrm{d}(e^{-s^2}) \\
&\le p_0\int_0^{\sqrt{\alpha}} \mathrm{d}(e^{-s^2}) + \int_{\sqrt{\alpha}}^\infty \mathrm{d}(e^{-s^2}) \\
&= p_0(1-e^{-\alpha}) + e^{-\alpha} \\
&= 1-(1-p_0)(1-e^{-\alpha}).
\end{aligned}$$

Inserting this in (4.43), we obtain

$$\begin{aligned}
\mathbb{P}\left(\sum_{j=1}^{n}|\zeta_j|^2 \le c_1\varepsilon_0^2 n\right) &\le e^{c_1\alpha n}\big[1-(1-p_0)(1-e^{-\alpha})\big]^n \\
&\le \exp\Big(n\big(c_1\alpha - (1-p_0)(1-e^{-\alpha})\big)\Big).
\end{aligned}$$

The claim now follows by setting $c_1 = (1-p_0)/2$ (for instance) and taking α a sufficiently small constant. $\square$

Let $M = A \circ X + B$ be as in Definition 4.1.1. In the following lemmas we assume that the atom variable ξ has κ-controlled second moment for some $\kappa \ge 1$. For $v \in \mathbb{C}^m$ and $i \in [n]$ we write

$$v^i := (v_j a_{ij})_{j=1}^m \tag{4.44}$$

For $\alpha > 0$, we denote

$$I_\alpha(v) := \{i \in [n] : \|v^i\| \geq \alpha\}. \tag{4.45}$$

Lemma 4.2.13 (Crude anti-concentration for the image of a fixed vector)**.** *Fix $v \in \mathbb{C}^m$ and let $\alpha > 0$ such that $I_\alpha(v) \neq \varnothing$. For all $I_0 \subset I_\alpha(v)$,*

$$\sup_{w \in \mathbb{C}^n} \mathbb{P}_{I_0}\left(\|Mv - w\| \leq c_0 \alpha |I_0|^{1/2}\right) \leq e^{-c_0|I_0|} \tag{4.46}$$

where $c_0 > 0$ is a constant depending only on κ (recall that $\mathbb{P}_{I_0}(\,\cdot\,)$ denotes probability conditional on all rows of M with indices $i \notin I_0$).

Proof. Fix $w \in \mathbb{C}^n$ arbitrarily. For any $i \in I_\alpha(v)$ and any $t \geq 0$ we have

$$\begin{aligned}
\mathbb{P}(|R_i \cdot v - w_i| \leq t) &\leq p_{\xi, v^i}(t) \\
&= p_{\xi, v^i/\|v^i\|}(t/\|v^i\|) \\
&\leq p_{\xi, v^i/\|v^i\|}(t/\alpha).
\end{aligned}$$

Taking $t = \alpha r_0$, by Lemma 4.2.7 we have

$$\mathbb{P}(|R_i \cdot v - w_i| \leq \alpha r_0) \leq 1 - r_0 \tag{4.47}$$

where $r_0 > 0$ depends only on κ.

Fix $I_0 \subset I_\alpha(v)$ arbitrarily. We may assume without loss of generality that I_0 is non-empty. By Lemma 4.2.12(a) there exists $c_1 > 0$ depending only on κ such that

$$\mathbb{P}_{I_0}\left(\sum_{i \in I_0} |R_i \cdot v - w_i|^2 \leq c_1 r_0^2 \alpha^2 |I_0|\right) \leq e^{-c_1|I_0|}. \tag{4.48}$$

Now for any $\tau \geq 0$,

$$\mathbb{P}_{I_0}\left(\|Mv - w\| \leq \tau |I_0|^{1/2}\right) = \mathbb{P}_{I_0}\left(\sum_{i=1}^n |R_i \cdot v - w_i|^2 \leq \tau^2 |I_0|\right)$$
$$\leq \mathbb{P}_{I_0}\left(\sum_{i \in I_0} |R_i \cdot v - w_i|^2 \leq \tau^2 |I_0|\right)$$

and the claim follows by taking $\tau = c_1^{1/2} r_0 \alpha =: c_0 \alpha$ and applying (4.48). $\square$

By following similar lines, but using Lemma 4.2.12(b) in place of Lemma 4.2.12(a) and Lemma 4.2.8 in place of Lemma 4.2.7, one obtains the following, which is superior to Lemma 4.2.13 for vectors v with small ℓ_∞ norm. The proof is omitted.

Lemma 4.2.14 (Improved anti-concentration for the image of a fixed vector)**.** *Fix $v \in \mathbb{C}^m$. Let $\alpha > 0$ such that $I_\alpha(v) \neq \varnothing$ and fix $I_0 \subset I_\alpha(v)$. For all $t \geq 0$,*

$$\sup_{w \in \mathbb{C}^n} \mathbb{P}_{I_0}\left(\|Mv - w\| \leq t |I_0|^{1/2}\right) \leq O_\kappa\left(\frac{1}{\alpha}\big(t + \|v\|_\infty\big)\right)^{|I_0|}. \tag{4.49}$$

4.3 Invertibility from connectivity: Compressible vectors

In this section we combine the anti-concentration estimates from the previous section with union bounds over nets of the sphere to prove that with high probability, a random matrix M as in Theorem 4.1.14 or Theorem 4.1.24 is well-invertible on the set of compressible vectors $\mathrm{Comp}(\theta, \rho)$ (as defined in (4.29)) for appropriate choices of θ, ρ. In applying the union bound there is a competition between the cardinality of the net and the quality of the anti-concentration estimates for fixed vectors. For small values of θ we can use nets of small cardinality, but only have poor anti-concentration bounds (namely, Lemma 4.2.13), while for large θ the nets are very large, but we have access to the improved anti-concentration of Lemma 4.2.14.

In both cases we start with a crude result, Lemma 4.3.5, giving control for the vectors in $\mathrm{Comp}(\theta_0, \rho_0)$ for some small value of θ_0. We then use an iterative argument argument to obtain control for larger values of θ while lowering the parameter ρ. For M as in Theorem 4.1.14, where the standard deviation profile A is broadly connected and $\|B\| = O(\sqrt{n})$, we can essentially follow the approach taken by Rudelson–Zeitouni for the Gaussian case, which allows us to take θ very close to one.

New ingredients are required to handle matrices as in Theorem 4.1.24, however. The main source of difficulty is that the perturbation B is allowed to be of arbitrary polynomial order in the operator norm. As a consequence, the starting point θ_0 for our iterative argument will be of size $o(1)$. This prevents us from being able to use the third condition of the super-regularity hypothesis (see Definition 4.1.23), which only "sees" vectors that are essentially supported on more than εn coordinates.

We deal with this by *reducing the entropy cost* of the nets over which we take union bounds. In Section 4.3.2 we prove Lemma 4.3.7 which shows, roughly speaking, that if we have already established control on vectors in $\mathrm{Comp}(\theta, \rho)$ for some θ, ρ, then we can control the vectors in $\mathrm{Comp}(\theta+\Delta, \rho')$ for some small Δ, ρ' using a *random* net of significantly smaller cardinality than what we would get from Lemma 4.2.2. We can then increment θ from θ_0 up to size $\gg n$, taking steps of size Δ. It turns out that this is all the control we need on compressible vectors for Theorem 4.1.24 in order to apply an averaging argument for the remaining incompressible vectors, which we do in Section 4.4.

The entropy reduction argument for Lemma 4.3.7 makes use of a strong version of the well-known Restricted Invertibility Theorem due to Spielman and Srivastava – see Theorem 4.3.9.

We now state the main results of this section. For $K \geq 1$ we denote the *boundedness event*

$$\mathcal{B}(K) := \big\{ \|M\| \leq K\sqrt{n} \,\big\}. \tag{4.50}$$

With a fixed choice of K we write

$$\mathcal{E}(\theta, \rho) := \mathcal{B}(K) \wedge \big\{ \exists u \in \mathrm{Comp}(\theta, \rho) : \|Mu\| \leq \rho K \sqrt{n} \,\big\}. \tag{4.51}$$

Proposition 4.3.1 (Compressible vectors: broadly connected profile)**.** *Let $M = A \circ X + B$*

be as in Definition 4.1.1 with $n/2 \le m \le 2n$, and assume that ξ has κ-controlled second moment for some $\kappa \ge 1$ (see Definition 4.2.5). Let $K \ge 1$ and $\sigma_0, \delta, \nu \in (0,1)$. There exist $\theta_0(\kappa, \sigma_0, \delta, K) > 0$ and $\rho(\kappa, \sigma_0, \delta, \nu, K) > 0$ and sufficiently small such that the following holds. Assume

1. $|\mathcal{N}_{A(\sigma_0)^\intercal}(j)| \ge \delta n$ *for all* $j \in [m]$;

2. $|\mathcal{N}^{(\delta)}_{A(\sigma_0)^\intercal}(J)| \ge \min((1+\nu)|J|, n)$ *for all* $J \subset [m]$ *with* $|J| \ge \theta_0 m$.

Then for any $0 < \theta \le (1 - \frac{\delta}{4}) \min(\frac{n}{m}, 1)$,

$$\mathbb{P}(\mathcal{E}(\theta, \rho)) \ll \exp\left(-c_\kappa \delta \sigma_0^2 n\right) \tag{4.52}$$

where $c_\kappa > 0$ depends only on κ and the implied constant depends only on $\kappa, \sigma_0, \delta, \nu$ and K.

Remark 4.3.2. The proof shows we may take $\theta_0 = c_\kappa \delta \sigma_0^2 / \log(K/\delta\sigma_0^2)$.

Remark 4.3.3. The assumptions (1) and (2) above imply that $A(\sigma_0)$ is (δ, ν)-broadly connected. In the proof of Theorem 4.5.1 below we will want to apply the above proposition to a rectangular matrix obtained by removing a small number of rows from M. The resulting standard deviation profile may no longer be broadly connected, but it will still satisfy (2) as above if number of rows removed is at most an for some $a > 0$ sufficiently small depending on $\kappa, \delta, \nu, \sigma_0$ and K.

The following gives control of compressible vectors for more general profiles than in Proposition 4.3.1 (essentially removing the condition (2)). However, we have to take the parameter ρ much smaller, and we only cover vectors that are essentially supported on a small (linear) proportion of the coordinates, rather than a proportion close to one.

Proposition 4.3.4 (Compressible vectors: general profile with large perturbation)**.** *Let $M = A \circ X + B$ be as in Definition 4.1.1 with $n/2 \le m \le 2n$. Assume ξ has κ-controlled second moment for some $\kappa \ge 1$, and that for some $a_0 > 0$ we have*

$$\sum_{i=1}^{n} a_{ij}^2 \ge a_0^2 n \quad \textit{for all } j \in [m]. \tag{4.53}$$

Fix $\gamma \geq 1/2$ and let $1 \leq K = O(n^{\gamma-1/2})$. Then for some $\rho = \rho(\gamma, a_0, \kappa, n) \gg_{\gamma,a_0,\kappa} n^{-O(\gamma^2)}$ and a sufficiently small constant $c_0 > 0$ we have

$$\mathbb{P}\big(\mathcal{E}(c_0 a_0^2, \rho)\big) \ll_{\gamma,a_0,\kappa} \exp\left(-c_\kappa a_0^2 n\right) \tag{4.54}$$

where $c_\kappa > 0$ depends only on κ.

The remainder of this section is organized as follows. In Section 4.3.1 we apply Corollary 4.2.3 and the crude anti-concentration bound of Lemma 4.2.7 to obtain a crude form of Proposition 4.3.4 (which also applies to matrices as in Proposition 4.3.1). In Section 4.3.2 we present a key technical result, Lemma 4.3.7, which allows one to control the invertibility of a random matrix over a portion of the sphere using a *random net* of small cardinality, assuming one has already established some control over compressible vectors. In the remaining sections Lemma 4.3.7 is applied with the improved anti-concentration estimate of Lemma 4.2.8 in an iterative argument to upgrade the crude control from Lemma 4.3.5 to obtain Propositions 4.3.1 and 4.3.4.

4.3.1 Highly compressible vectors

In this subsection we establish the following crude version of Proposition 4.3.4, giving control on vectors in $\mathrm{Comp}(\theta_0, \rho_0)$ with θ_0 depending on both a_0 and K. This will be used in the proofs of both Proposition 4.3.1 and Proposition 4.3.4. In the former we will seek to increase θ_0 to be close to one, while in the latter we will seek to eliminate the dependence on K (which may depend on n).

Lemma 4.3.5 (Highly compressible vectors)**.** *Let $M = A \circ X + B$ be as in Definition 4.1.1 with $m \leq 2n$. Assume that ξ has κ-controlled second moment for some $\kappa \geq 1$. Suppose also that there is a constant $a_0 > 0$ such that for all $j \in [m]$, $\sum_{i=1}^n a_{ij}^2 \geq a_0^2 n$. Let $K \geq 1$. Then with notation as in* (4.51) *we have*

$$\mathbb{P}(\mathcal{E}(\theta_0, \rho_0)) \leq e^{-c_\kappa a_0^2 n} \tag{4.55}$$

where $\theta_0 = c_\kappa a_0^2/\log(K/a_0^2)$ and $\rho_0 = c_\kappa a_0^2/K$ for a sufficiently small $c_\kappa > 0$ depending only on κ.

We will need the following lemma, which ensures that the set $I_\alpha(v)$ from (4.45) is reasonably large when the columns of A have large ℓ_2 norm. A similar argument has been used in [LR12] and [RZ16].

Lemma 4.3.6 (Many good rows)**.** *Let A be an $n \times m$ matrix as in Definition 4.1.1, and assume that for some $a_0 > 0$ we have $\sum_{i=1}^n a_{ij}^2 \ge a_0^2 n$ for all $j \in [m]$. Then for any $v \in S^{m-1}$ we have $|I_{a_0/2}(v)| \ge \frac{1}{2}a_0^2 n$.*

Proof. Writing $\alpha = a_0/\sqrt{2}$, we have

$$\begin{aligned}
a_0^2 n &\le \sum_{i=1}^n \sum_{j=1}^m |v_j|^2 a_{ij}^2 \\
&= \sum_{i\in I_\alpha(v)} \sum_{j=1}^m |v_j|^2 a_{ij}^2 + \sum_{i\notin I_\alpha(v)} \sum_{j=1}^m |v_j|^2 a_{ij}^2 \\
&\le \sum_{i\in I_\alpha(v)} \sum_{j=1}^m |v_j|^2 + \sum_{i\notin I_\alpha(v)} \frac{1}{2}a_0^2 \\
&\le |I_\alpha(v)| + \frac{1}{2}a_0^2 n
\end{aligned}$$

and rearranging gives the claim. □

Proof of Lemma 4.3.5. Fix $J \subset [m]$ of size $\lfloor \theta_0 m \rfloor$ and let $u \in S^J$ be arbitrary. Writing $\alpha = a_0/\sqrt{2}$, by Lemma 4.2.13 and our choice of ρ_0 (with $c_\kappa > 0$ sufficiently small depending on κ) we have

$$\mathbb{P}\big(\|Mu\| \le \rho_0 K\sqrt{n}\big) \le \mathbb{P}\big(\|Mu\| \le c_\kappa a_0 |I_\alpha(u)|^{1/2}\big) \le e^{-c_\kappa |I_\alpha(u)|}.$$

Applying Lemma 4.3.6,

$$\mathbb{P}\big(\|Mu\| \le \rho_0 K\sqrt{n}\big) \le e^{-c_\kappa a_0^2 n} \tag{4.56}$$

(adjusting c_κ). By Corollary 4.2.3 and shrinking c_κ again,

$$\mathbb{P}\big(\exists u \in (S^J)_{\rho_0} : \|Mu\| \le \rho_0 K\sqrt{n}\big) \le (C/\rho_0)^{\theta_0 m} e^{-c_\kappa a_0^2 n}.$$

Applying the union bound over all choice of $J \in \binom{[m]}{\theta_0 m}$

$$\begin{aligned}\mathbb{P}(\mathcal{E}(\theta_0,\rho_0)) &\le (C/\theta_0)^{\theta_0 m}(C/\rho_0)^{2\theta_0 m} e^{-c_\kappa a_0^2 n}\\ &\le \left(\frac{C}{a_0\rho_0^2}\right)^{2\theta_0 n} e^{-c_\kappa a_0^2 n}\end{aligned}$$

where we applied our assumption $m \le 2n$. The desired bound now follows from substituting our choices of θ_0, ρ_0. □

4.3.2 An entropy reduction lemma

Lemma 4.3.7 (Control by a random net of small cardinality)**.** *For every $I \subset [n], J \subset [m]$, $\varepsilon > 0$ there is a random finite set $\Sigma_{I,J}(\varepsilon) \subset S^J$, measurable with respect to $\mathcal{F}_{I.J} = \langle \{\xi_{ij}\}_{i\in I, j\in J}\rangle$, such that the following holds. Let $\rho \in (0,1)$, $K > 0$ and $0 < \theta < \frac{n}{m}$. On $\mathcal{B}(K) \wedge \mathcal{E}(\theta,\rho)^c$, for all $J \subset [m]$ with $|J| > \theta m$ and all $\beta, \rho' \in (0,1)$, putting*

$$\rho'' = \frac{4\rho'}{\beta\rho}\left(\frac{n}{\lfloor \theta m\rfloor}\right)^{1/2} \tag{4.57}$$

there exists $I \subset [n]$ with $|I| = \lfloor (1-\beta)^2 \lfloor \theta m\rfloor \rfloor$ such that

1. *$|\Sigma_{I,J}(\rho'')| \le (C/\rho'')^{2(|J|-|I|)}$ for an absolute constant $C > 0$, and*
2. *for any $u \in (S^J)_{\rho'}$ such that $\|Mu\| \le \rho' K\sqrt{n}$, we have $\operatorname{dist}(u, \Sigma_{I,J}(\rho'')) \le 3\rho''$.*

As a consequence, writing

$$\mathcal{G}_{I,J}(\rho'') := \left\{ \big|\Sigma_{I,J}(\rho'')\big| \le \left(\frac{C}{\rho''}\right)^{2(|J|-|I|)} \right\} \tag{4.58}$$

we have that for any $\theta' \in (\theta, 1]$,

$$\mathcal{E}(\theta,\rho)^c \wedge \mathcal{E}(\theta',\rho') \subset \bigvee_{J\in\binom{[m]}{\theta' m}} \bigvee_{I\in\binom{[n]}{(1-\beta)^2\lfloor\theta m\rfloor}} \left(\mathcal{G}_{I,J}(\rho'') \wedge \left\{\exists u \in \Sigma_{I,J}(\rho'') : \|Mu\| \le 4\rho'' K\sqrt{n}\right\}\right). \tag{4.59}$$

Remark 4.3.8. We will obtain the random set $\Sigma_{I,J}(\varepsilon)$ as the intersection of the sphere S^J with an ε-net of the kernel of the sub-matrix $M_{I\times J}$. However, for our purposes it only matters that it is fixed by conditioning on the rows $\{R_i\}_{i\in I}$, has small cardinality, and serves as a net for almost-null vectors of M that are supported on J.

For the proof of Lemma 4.3.7 we will use the following version of the Restricted Invertibility Theorem due to Spielman and Srivastava [SS12] (the precise version we use is stated in [MSS14]), which strengthens the classic result of Bourgain and Tzafriri [BT87].

Theorem 4.3.9 (Restricted Invertibility Theorem (cf. [MSS14, Theorem 3.1])**.** *Suppose $v_1, \dots, v_n \in \mathbb{C}^m$ are such that $\sum_{i=1}^n v_i v_i^* = I_m$. For any $\beta \in (0,1)$, there is a subset $I \subset [n]$ of size $|I| = \lfloor (1-\beta)^2 m\rfloor$ for which*

$$\lambda_{|I|}\left(\sum_{i\in I} v_i v_i^*\right) \ge \beta^2 m/n \tag{4.60}$$

where $\lambda_k(A)$ denotes the kth largest eigenvalue of a Hermitian matrix A.

This has the following consequence, which can be seen as a quantitative version of the basic fact from linear algebra that the row rank of a matrix is equal to its column rank.

Corollary 4.3.10. *Let M be an $n\times m$ matrix with $n \ge m$, and assume $s_m(M) \ge \varepsilon_0\sqrt{n}$ for some $\varepsilon_0 > 0$. For any $\beta \in (0,1)$ there exists $I \subset [n]$ with $|I| = \lfloor (1-\beta)^2 m\rfloor$ such that*

$$s_{|I|}(M_{I\times[m]}) \ge \beta\varepsilon_0\sqrt{m}.$$

Remark 4.3.11. The original Restricted Invertibility Theorem of Bourgain and Tzafriri [BT87] only gives $|I| \ge cm$ and $s_{|I|}(M_{I\times[m]}) \ge c\varepsilon_0\sqrt{m}$ for some (small) absolute constant $c > 0$, while it will be important for our purposes to be able to take I of size close to m.

Proof of Corollary 4.3.10. By the singular value decomposition it suffices to consider M of the form

$$M = U\Sigma$$

where U is an $n \times m$ matrix with orthonormal columns and Σ is an $m \times m$ diagonal matrix with entries bounded below by $\varepsilon_0\sqrt{n}$. Fix $\alpha \in (0,1)$. Letting $v_1^*, \dots, v_n^* \in \mathbb{C}^m$ denote the rows of U, it follows from orthonormality that

$$I_m = U^*U = \sum_{i=1}^n v_i v_i^*.$$

Hence, we can apply Theorem 4.3.9 to obtain a subset $I \subset [n]$ with $|I| = \lfloor (1-\beta)^2 m \rfloor$ such that

$$s_{|I|}(U_{I\times[m]})^2 = \lambda_{|I|}\left(\sum_{i\in I} v_i v_i^*\right) \geq \beta^2 m/n.$$

Now we have

$$s_{|I|}(M_{I\times m}) \geq s_{|I|}(U_{I\times m}) s_m(\Sigma) \geq \beta\sqrt{\frac{m}{n}}\varepsilon_0\sqrt{n} = \beta\varepsilon_0\sqrt{m}. \qquad \square$$

Proof of Lemma 4.3.7. Let $I \subset [n]$, $J \subset [m]$, and write $V_{I,J} = \mathbb{C}^J \cap \ker(M_{I,J})$. Conditional on $\mathcal{F}_{I,J}$, for $\varepsilon > 0$ we let $\Sigma_{I,J}(\varepsilon)$ be an ε-net of $S^{m-1} \cap V_{I,J}$. By Lemma 4.2.2 we may take

$$|\Sigma_{I,J}(\varepsilon)| = O(1/\varepsilon)^{2\dim(V_{I,J})}. \tag{4.61}$$

Let $\rho, \rho' \in (0,1)$, $K > 0$ and $0 < \theta < \frac{n}{m}$. Fix $\beta \in (0,1)$ and $J \subset [m]$ with $|J| > \theta m$. On $\mathcal{E}(\theta,\rho)^c$, for all $J_0 \subset J$ with $|J_0| = \lfloor \theta m \rfloor$ we have

$$s_{\lfloor \theta m \rfloor}(M_{[n]\times J_0}) \geq \rho K \sqrt{n}.$$

By Corollary 4.3.10 there exists $I \subset [n]$ with $|I| = \lfloor (1-\beta)^2 \lfloor \theta m \rfloor \rfloor$ such that

$$s_{|I|}(M_{I\times J_0}) \geq \beta\rho K\sqrt{\lfloor \theta m \rfloor}.$$

By the Cauchy interlacing law,

$$s_{|I|}(M_{I\times J}) \geq \beta\rho K\sqrt{\lfloor \theta m \rfloor}. \tag{4.62}$$

In particular, the sub-matrix $(y_{ij})_{i\in I, j\in J}$ has full row-rank, which implies $\dim(V_{I,J}) = |J| - |I|$. From (4.61) we conclude

$$|\Sigma_{I,J}(\varepsilon)| = O(1/\varepsilon)^{2(|J|-|I|)} \tag{4.63}$$

for any $\varepsilon > 0$.

Now suppose there exists $u \in (S^J)_{\rho'}$ such that

$$\|Mu\| \leq \rho' K\sqrt{n}. \tag{4.64}$$

Letting $v \in S^J$ such that $\|u - v\| \leq \rho'$, by the triangle inequality we have

$$\|Mv\| \leq \|Mu\| + \|M\|\|u - v\| \leq 2\rho' K\sqrt{n}. \tag{4.65}$$

On the other hand,

$$\|Mv\| \geq \|M_{I\times[m]}v\| = \|M_{I\times[m]}(\mathrm{I} - P_{V_{I,J}})v\|$$

where $P_{V_{I,J}}$ is the matrix for orthogonal projection to the subspace $V_{I,J}$. Applying (4.62),

$$\|Mv\| \geq \|(\mathrm{I} - P_{V_{I,J}})v\|\beta\rho K\sqrt{\lfloor \theta m \rfloor}.$$

Together with (4.65) this implies that v lies within distance

$$\frac{2\rho'\sqrt{n}}{\beta\rho\sqrt{\lfloor \theta m \rfloor}} = \rho''/2 \tag{4.66}$$

of the subspace $V_{I,J}$. By the triangle inequality,

$$\begin{aligned}\mathrm{dist}(u, \Sigma_{I,J}(\rho'')) &\le \|u-v\| + \rho'' + \mathrm{dist}(v, S^{m-1} \cap V_{I,J}) \\ &\le \rho' + 2\rho'' \le 3\rho''\end{aligned}$$

as desired.

Now to prove (4.59), let $\theta' \in (\theta, 1]$. Intersecting with $\mathcal{E}(\theta,\rho)^c$ and applying the first part of the lemma,

$$\begin{aligned}&\mathcal{E}(\theta,\rho)^c \wedge \mathcal{E}(\theta',\rho') \\ &\qquad = \mathcal{B}(K) \wedge \mathcal{E}(\theta,\rho)^c \wedge \bigvee_{J \in \binom{[m]}{\theta' m}} \left\{ \exists v \in (S^J)_{\rho'} : \|Mv\| \le \rho' K\sqrt{n} \right\} \\ &\qquad \subset \bigvee_{J \in \binom{[m]}{\theta' m}} \bigvee_{I \in \binom{[n]}{(1-\beta)^2 \lfloor \theta m \rfloor}} \left(\mathcal{G}_{I,J}(\rho'') \wedge \left\{ \exists u \in \Sigma_{I,J}(\rho'') : \|Mu\| \le 4\rho'' K\sqrt{n} \right\} \right) \qquad (4.67)\end{aligned}$$

where in the last line we noted that for $v \in (S^J)_{\rho'}, u \in \Sigma_{I,J}(\rho'')$ such that $\|u-v\| \le 3\rho''$, we have

$$\|Mu\| \le \|Mv\| + 3\rho'' K\sqrt{n} \le (\rho' + 3\rho'')K\sqrt{n} \le 4\rho'' K\sqrt{n}. \qquad \square$$

4.3.3 Broadly connected profile: Proof of Proposition 4.3.1

We will obtain Proposition 4.3.1 from an iterative application of the following lemma:

Lemma 4.3.12 (Incrementing compressibility: broadly connected profile)**.** *Let $M = A \circ X + B$ be as in Definition 4.1.1 with $m \ge n/2$. Assume ξ has κ-controlled second moment for some $\kappa \ge 1$, and that for some $\sigma_0, \delta, \nu, \theta_1 \in (0,1)$ we have*

1. *$|\mathcal{N}_{A(\sigma_0)}(j)| \ge \delta n$ for all $j \in [m]$;*
2. *$|\mathcal{N}^{(\delta)}_{A(\sigma_0)}(J)| \ge \min((1+\nu)|J|, n)$ for all $J \subset [m]$ with $|J| \ge (\theta_1/2)m$.*

Let $K \ge 1$, $\rho \in (0,1)$, and $\theta \in [\theta_1, 1)$ such that $(1+\frac{\nu}{2})\theta m < n$. There exists $\rho' =$

$\rho'(\kappa, \sigma_0, \delta, \nu, \rho, \theta, K) > 0$ *such that*

$$\mathbb{P}\Big(\mathcal{E}(\theta,\rho)^c \wedge \mathcal{E}\Big(\Big(1+\frac{\nu}{10}\Big)\theta, \rho'\Big)\Big) = O(e^{-n}) \tag{4.68}$$

where the implied constant depends only on $\kappa, \sigma_0, \delta, \nu, \rho, \theta,$ *and* K.

Proof. We may assume n is sufficiently large depending on $\kappa, \sigma_0, \delta, \nu, \rho, \theta, K$. Write $\theta' = \big(1+\frac{\nu}{10}\big)\theta$ and take $\beta = \frac{\nu}{10}$. Let $\rho' > 0$ to be taken sufficiently small depending on $\kappa, \sigma_0, \delta, \nu, \rho, \theta, K$, and let ρ'' be as in (4.57). Intersecting the right hand side of (4.59) with $\mathcal{E}(\theta,\rho)^c$, we have

$$\begin{aligned}
&\mathcal{E}(\theta,\rho)^c \wedge \mathcal{E}(\theta',\rho') \subset \\
&\quad \bigvee_{J\in\binom{[m]}{\theta' m}} \bigvee_{I\in\binom{[n]}{(1-\beta)^2\lfloor\theta m\rfloor}} \Big(\mathcal{G}_{I,J}(\rho'') \wedge \mathcal{E}(\theta,\rho)^c \wedge \Big\{\exists u \in \Sigma_{I,J}(\rho'') : \|Mu\| \le 4\rho'' K\sqrt{n}\Big\}\Big) \\
&\quad \subset \bigvee_{J\in\binom{[m]}{\theta' m}} \bigvee_{I\in\binom{[n]}{(1-\beta)^2\lfloor\theta m\rfloor}} \Big(\mathcal{G}_{I,J}(\rho'') \wedge \Big\{\exists u \in \Sigma_{I,J}(\rho'') \setminus \mathrm{Comp}(\theta,\rho) : \|Mu\| \le 4\rho'' K\sqrt{n}\Big\}\Big)
\end{aligned} \tag{4.69}$$

where the second line follows by taking ρ' small enough that $4\rho'' < \rho$.

Fix $J \subset [m]$ and $I \subset [n]$ of sizes $\lfloor\theta' m\rfloor, \lfloor(1-\beta)^2\lfloor\theta m\rfloor\rfloor$, respectively, and condition on $\mathcal{F}_I$ to fix $\Sigma_{I,J}(\rho'')$. Consider an arbitrary element $u \in \Sigma_{I,J}(\rho'') \setminus \mathrm{Comp}(\theta,\rho)$. By Lemma 4.2.1, there is a set $L \subset [m]$ with $|L| \ge (1 - \frac{\nu}{C_0^2})\theta m$ and

$$\frac{\rho}{\sqrt{m}} \le |u_j| \le \frac{C_0}{\sqrt{\nu\theta m}} \tag{4.70}$$

for all $j \in L$, where $C_0 > 0$ is an absolute constant to be taken sufficiently large. For any $i \in \mathcal{N}^{(\delta)}(L)$, we have

$$\|(u_L)^i\|^2 \ge \sum_{i\in L: a_{ij}\ge\sigma_0} |u_j|^2 a_{ij}^2 \ge \frac{\rho^2}{m}\sigma_0^2\delta|L| \ge \frac{1}{2}\rho^2\sigma_0^2\delta\theta =: \alpha^2 \tag{4.71}$$

where in the last inequality we took C_0 sufficiently large. Hence,

$$|I_\alpha(u_L)| \geq |\mathcal{N}^{(\delta)}(L)| \geq \min\left(n, (1+\nu)(1-\nu/C_0^2)\theta m\right) \geq \left(1+\frac{\nu}{2}\right)\theta m \tag{4.72}$$

taking C_0 larger if necessary, where in the second inequality we used our assumption $\theta \geq \theta_1$, and in the third inequality we used our assumption $(1+\frac{\nu}{2})\theta m < n$.

Fix $I_0 \subset I_\alpha(u_L) \setminus I$ of size $n_0 := \lfloor (1+\frac{\nu}{2})\theta m \rfloor - |I|$. In particular,

$$\begin{aligned}\frac{\nu}{2}\theta m \leq n_0 &\leq \left(1+\frac{\nu}{2}\right)\theta m - (1-2\beta)\theta m \\ &\leq \nu\theta m \end{aligned} \tag{4.73}$$

and

$$\begin{aligned} n_0 + 2|I| - 2|J| &\geq \left(1+\frac{\nu}{2}\right)\theta m + (1-2\beta)\theta m - 2\left(1+\frac{\nu}{10}\right)\theta m - O(1) \\ &= \frac{1}{10}\nu\theta m - O(1). \end{aligned} \tag{4.74}$$

by our choice of β. By Lemma 4.2.14,

$$\mathbb{P}_{I_0}\left(\|Mu\| \leq 4\rho'' K\sqrt{n}\right) \leq O_\kappa\left(\frac{1}{\alpha}\left(\frac{\rho'' K\sqrt{n}}{\sqrt{|I_0|}} + \frac{1}{\sqrt{\nu\theta m}}\right)\right)^{n_0} \leq O_\kappa\left(\frac{\rho'' K}{\alpha\theta^{1/2}}\right)^{n_0} \tag{4.75}$$

where in the second inequality we applied the assumption $m \geq n/2$ and assumed that n is sufficiently large that $\rho'' \gg 1/K\sqrt{n}$ (it follows from (4.57) and our assumption that ρ' is independent of n that ρ'' is bounded below independent of n).

Suppose that $\mathcal{G}_{I,J}(\rho'')$ holds. Since the bound (4.75) is uniform in the choice of I_0, we can undo the conditioning and apply the union bound over elements of $\Sigma_{I,J}(\rho'') \setminus \mathrm{Comp}(\theta, \rho)$

to find

$$\begin{aligned}\mathbb{P}\left(\exists u \in \Sigma_{I,J}(\rho'') \setminus \mathrm{Comp}(\theta,\rho) : \|Mu\| \le 4\rho'' K\sqrt{n}\right) &\le O\left(\frac{1}{\rho''}\right)^{2(|J|-|I|)} O_\kappa\left(\frac{\rho'' K}{\alpha\theta^{1/2}}\right)^{n_0} \\ &= O_\kappa\left(\frac{K}{\alpha\theta^{1/2}}\right)^{n_0} O(\rho'')^{n_0+2|I|-2|J|} \\ &= O_\kappa\left(\frac{K}{\alpha\theta^{1/2}}\right)^{\nu\theta m} O(\rho'')^{\frac{1}{10}\nu\theta m - O(1)}\end{aligned}$$

where in the last line we applied the bounds (4.73) and (4.74). Since this is uniform in I, J, we can undo the conditioning on $\mathcal{F}_I$ and apply (4.69) with another union bound over the choices of I, J to obtain

$$\mathbb{P}(\mathcal{E}(\theta,\rho)^c \wedge \mathcal{E}(\theta',\rho')) \le 2^{m+n} O_\kappa\left(\frac{K}{\alpha\theta^{1/2}}\right)^{\nu\theta m} O\left(\frac{\rho'}{\nu\rho\theta^{1/2}}\right)^{\frac{1}{10}\nu\theta m - O(1)} \tag{4.76}$$

where we have substituted the definition of ρ''. The result now follows by taking ρ' sufficiently small. □

Now we conclude the proof of Proposition 4.3.1. From our assumptions it follows that for all $j \in [m]$ we have $\sum_{i=1}^n a_{ij}^2 \ge \delta\sigma_0^2 n$. Together with our assumption $m \le 2n$, this means we can apply Lemma 4.3.5 to find that

$$\mathbb{P}(\mathcal{E}(\theta_0,\rho_0)) \le e^{-c_\kappa \delta\sigma_0^2 n} \tag{4.77}$$

where $\theta_0 = c_\kappa\delta\sigma_0^2/\log(K/\delta\sigma_0^2)$ and $\rho_0 = c_\kappa\delta\sigma_0^2/K$.

We may assume without loss of generality that $\nu \le \delta/2$. For $l \ge 1$ set $\theta_l = (1+\frac{\nu}{10})^l\theta_0$, and let k be the smallest l such that $\theta_l \ge \theta$. We have

$$\left(1+\frac{\nu}{2}\right)\theta_{k-1} m \le \left(1+\frac{\nu}{2}\right)\theta m \le \left(1-\frac{\delta^2}{16}\right)\min(m,n) < n.$$

In particular, $(1+\nu/10)^k\theta_0 \le (1+\nu/10)\theta \le 1$, so

$$k \le \frac{\log \frac{1}{\theta_0}}{\log\left(1+\frac{\nu}{10}\right)} \ll_{\kappa,\sigma_0,\delta,\nu,K} 1. \tag{4.78}$$

Applying Lemma 4.3.12 inductively, we have that for every $1 \le l \le k$ there is $\rho_l > 0$ depending only on $\kappa, \sigma_0, \delta, \nu$ and K such that

$$\mathbb{P}(\mathcal{E}(\theta_l,\rho_l) \setminus \mathcal{E}(\theta_{l-1},\rho_{l-1})) = O_{\kappa,\sigma_0,\delta,\nu,K}(e^{-n}). \tag{4.79}$$

Together with (4.77) and the union bound,

$$\begin{aligned}\mathbb{P}(\mathcal{E}(\theta,\rho)) &\le \mathbb{P}(\mathcal{E}(\theta_0,\rho_0)) + \sum_{l=1}^{k} \mathbb{P}(\mathcal{E}(\theta_l,\rho_l) \setminus \mathcal{E}(\theta_{l-1},\rho_{l-1}))\\ &\le e^{-c_\kappa \delta \sigma_0^2 n} + O_{\kappa,\sigma_0,\delta,\nu,K}(e^{-n}) = O_{\kappa,\sigma_0,\delta,\nu,K}(e^{-c_\kappa\delta\sigma_0^2 n}).\end{aligned}$$

4.3.4 General profile: Proof of Proposition 4.3.4

For technical reasons (essentially due to the fact that we want to allow the operator norm to have arbitrary polynomial size) the anti-concentration argument from the previous section will not suffice here, and we will need the following substitute. Roughly speaking, while previously we argued by isolating a large set of coordinates on which the vector u is "flat" (see (4.70)), here we will need to locate a set on which u is *very flat*, only fluctuating by a constant factor. This is done by a simple dyadic decomposition of the range of u, which is responsible for the loss of a logarithmic factor in the probability bound. A similar argument will be necessary in the conclusion of the proof of Theorem 4.1.24 in Section 4.4.2.

Lemma 4.3.13 (Anti-concentration for the image of an incompressible vector)**.** *Let M be as in Proposition 4.3.4. Let $v \in \mathrm{Incomp}(\theta,\rho)$ for some $\theta,\rho \in (0,1)$, and fix $I_0 \subset [n]$ with $|I_0| \le \frac{1}{4}a_0^2 n$. Then*

$$\sup_{w\in\mathbb{C}^n} \mathbb{P}_{[n]\setminus I_0}\left(\|Mv-w\| \le t\sqrt{n}\right) = O_\kappa\left(\frac{t\log^{1/2}(\frac{\sqrt{m}}{\rho})}{a_0^2\rho\theta^{1/2}}\right)^{\frac14 a_0^2 n} \qquad \text{for all } t \ge \frac{a_0\rho}{\sqrt{m}}. \tag{4.80}$$

Remark 4.3.14. Proceeding as in the proof of Lemma 4.3.12 would yield

$$\sup_{w\in\mathbb{C}^n} \mathbb{P}_{[n]\setminus I_0}\left(\|Mv-w\|\le t\sqrt{n}\right) = O_\kappa\left(\frac{t}{a_0^2\rho\theta^{1/2}}\right)^{\frac{1}{4}a_0^2 n} \qquad \text{for all } \; t\ge \frac{a_0}{\sqrt{\theta m}}. \tag{4.81}$$

The ability to take t down to the scale $\sim \rho/\sqrt{m}$ will be crucial in the proof of Lemma 4.3.15 below.

Proof. We begin by finding a set of indices on which v varies by at most a factor of 2. For $k\ge 0$ let $L_k = \{j\in[m] : 2^{-(k+1)} < |v_j| \le 2^{-k}\}$. Since $v\in \mathrm{Incomp}(\theta,\rho)$, we have

$$|L^+| := |\{j\in[m] : |v_j|\ge \rho/\sqrt{m}\}| \ge \theta m.$$

Indeed, were this not the case then v would be within distance ρ of the vector v_{L^+} whose support is smaller than θm, implying $v\in\mathrm{Comp}(\theta,\rho)$. Thus,

$$L^+\subset \bigcup_{k=0}^{\ell} L_k$$

for some $\ell \ll \log(\frac{\sqrt{m}}{\rho})$. By the pigeonhole principle there exists $k^*\le \ell$ such that $L^* := L_{k^*}$ satisfies

$$|L^*| \ge \frac{\theta n}{\ell} \gg \frac{\theta m}{\log(\frac{\sqrt{m}}{\rho})}. \tag{4.82}$$

Denote $I^* := I_{\frac{a_0}{2}\|v_{L^*}\|}(v_{L^*})$. By Lemma 4.3.6,

$$|I^*| \ge \frac{1}{2}a_0^2 n. \tag{4.83}$$

Fix $i\in I^*$. By definition of I^*,

$$\|(v^i)_{L^*}\| \ge \frac{1}{2}a_0\|v_{L^*}\| \tag{4.84}$$

and since $|v_j|\gg \rho/\sqrt{m}$ on L^*,

$$\|v_{L^*}\| \gg \frac{\rho}{\sqrt{m}}|L^*|^{1/2}. \tag{4.85}$$

Furthermore, since $a_{ij} \le 1$ for all $j \in [m]$ and v varies by a factor at most 2 on L^*,

$$\|(v^i)_{L^*}\|_\infty \le \|v_{L^*}\|_\infty \le 2\frac{\|v_{L^*}\|}{|L^*|^{1/2}}. \tag{4.86}$$

Fix $w \in \mathbb{C}^n$ arbitrarily. By Lemma 4.2.8 and the above estimates, for all $t \ge 0$ we have

$$\begin{aligned}
\mathbb{P}(|R_i \cdot v - w_i| \le t) &\ll_\kappa \frac{t + \|(v^i)_{L^*}\|_\infty}{\|(v^i)_{L^*}\|} \\
&\ll \frac{1}{a_0}\left(\frac{t}{\|v_{L^*}\|} + \frac{\|(v^i)_{L^*}\|_\infty}{\|v_{L^*}\|}\right) \\
&\ll \frac{1}{a_0}\left(\frac{t}{\rho}\left(\frac{m}{|L^*|}\right)^{1/2} + \frac{1}{|L^*|^{1/2}}\right) \\
&= \frac{1}{a_0}\left(\frac{m}{|L^*|}\right)^{1/2}\left(\frac{t}{\rho} + \frac{1}{\sqrt{m}}\right).
\end{aligned}$$

By Lemma 4.2.12,

$$\begin{aligned}
\mathbb{P}_{I^*\setminus I_0}\left(\|Mv - w\| \le t|I^* \setminus I_0|^{1/2}\right) &\le \mathbb{P}_{I^*\setminus I_0}\left(\sum_{i\in I^*\setminus I_0} |R_i \cdot v - w_i|^2 \le t^2|I^* \setminus I_0|\right) \\
&= O_\kappa\left(\frac{t\sqrt{m}}{a_0\rho|L^*|^{1/2}}\right)^{|I^*\setminus I_0|}
\end{aligned}$$

for all $t \ge \rho/\sqrt{m}$. Substituting the lower bounds (4.82), (4.83) on $|L^*|$ and $|I^*|$ and our assumption $|I_0| \le \frac{1}{4}a_0^2 n$,

$$\mathbb{P}_{I^*\setminus I_0}\left(\|Mv - w\| \le \frac{1}{2}ta_0\sqrt{n}\right) = O_\kappa\left(\frac{t\log^{1/2}(\frac{\sqrt{m}}{\rho})}{a_0\rho\theta^{1/2}}\right)^{\frac{1}{4}a_0^2 n}$$

for all $t \ge \rho/\sqrt{m}$. The result now follows by replacing t with $2t/a_0$ as undoing the conditioning on the remaining rows in $[n] \setminus I_0$. $\square$

Now we are ready to prove the analogue of Lemma 4.3.12 for general profiles. Whereas in the broadly connected case we obtained control on vectors in $\mathrm{Comp}((1+\beta)\theta, \rho')$ after restricting to the event that we have control on $\mathrm{Comp}(\theta, \rho)$, for small $\beta > 0$, here we will also need to assume control on $\mathrm{Comp}(\theta_0, \rho_0)$ for a fixed small θ_0 at each step. The control on $\mathrm{Comp}(\theta, \rho)$ will be used to obtain a net of low cardinality using Lemma 4.3.7, while the control on $\mathrm{Comp}(\theta_0, \rho_0)$ will be used to obtain good anti-concentration estimates using Lemma 4.3.13. In the broadly connected case the control on $\mathrm{Comp}(\theta, \rho)$ was enough to serve both purposes.

Lemma 4.3.15 (Incrementing compressibility: general profile)**.** *Let M be as in Proposition 4.3.4, fix $\gamma > 1/2$ and put $K = n^{\gamma-1/2}$. Let θ_0, ρ_0 be as in Lemma 4.3.5, and fix $\theta \in [\theta_0, c_0 a_0^2]$, where c_0 is a sufficiently small constant (we may assume the constant c in Lemma 4.3.5 is sufficiently small so that this interval is non-empty). We have*

$$\mathbb{P}\Big(\mathcal{E}(\theta_0, \rho_0)^c \wedge \mathcal{E}(\theta, \rho)^c \wedge \mathcal{E}(\theta + \beta a_0^2, \rho')\Big) = O_{\gamma, a_0, \kappa}(e^{-n}) \tag{4.87}$$

for some $\rho' \gg_{\gamma, a_0, \kappa} n^{-O(\gamma)}\rho$, where we set

$$\beta = c_1 \min\left(1, \frac{1}{\gamma - 1/2}\right) \tag{4.88}$$

for a sufficiently small constant $c_1 > 0$.

Proof. Let $\rho' > 0$ to be taken sufficiently small, and let ρ'' be as in (4.57). We denote $\theta' = \theta + \beta a_0^2$. Intersecting both sides of (4.59) with $\mathcal{E}(\theta_0, \rho_0)^c$, we have

$$\begin{aligned} &\mathcal{E}(\theta_0, \rho_0)^c \wedge \mathcal{E}(\theta, \rho)^c \wedge \mathcal{E}(\theta', \rho') \subset \\ &\quad \subset \bigvee_{J \in \binom{[m]}{\theta' m}} \bigvee_{I \in \binom{[n]}{(1-\beta)^2 \lfloor \theta m \rfloor}} \left(\mathcal{G}_{I,J}(\rho'') \wedge \left\{ \exists u \in \Sigma_{I,J}(\rho'') \setminus \mathrm{Comp}(\theta_0, \rho_0) : \|Mu\| \le 4\rho'' K \sqrt{n} \right\} \right) \end{aligned} \tag{4.89}$$

where we have assumed ρ' is small enough that $4\rho'' < \rho_0$.

Fix $J \subset [m]$ and $I \subset [n]$ of size $\lfloor \theta' m \rfloor$, $\lfloor (1-\beta)^2 \lfloor \theta m \rfloor \rfloor$, respectively, and condition on $\mathcal{F}_I$ to fix $\Sigma_{I,J}(\rho'')$. Fix an arbitrary $u \in \Sigma_{I,J}(\rho'') \setminus \mathrm{Comp}(\theta_0, \rho_0)$. From Lemma 4.3.13 we have

$$\mathbb{P}_{[n]\setminus I}\left(\|Mu\| \le 4\rho'' K \sqrt{n} \right) = O_\kappa \left(\frac{\rho'' K \log^{1/2}(\frac{\sqrt{n}}{\rho_0})}{a_0^2 \rho_0 \theta_0^{1/2}} \right)^{\frac{1}{4} a_0^2 n} \tag{4.90}$$

provided

$$\rho'' \geq \frac{c a_0 \rho_0}{K\sqrt{n}} \tag{4.91}$$

for some small constant $c > 0$ (note that we used our assumption $n/2 \leq m \leq 2n$).

Applying the union bound over the choices of $u \in \Sigma_{I,J}(\rho'') \setminus \mathrm{Comp}(\theta_0, \rho_0)$, on the event $\mathcal{G}_{I,J}(\rho'')$ we have

$$\begin{aligned}
&\mathbb{P}\left(\exists u \in \Sigma_{I,J}(\rho'') \setminus \mathrm{Comp}(\theta_0, \rho_0) : \|Mu\| \leq 4\rho'' K\sqrt{n}\right) \\
&\qquad \leq O\left(\frac{1}{\rho''}\right)^{2(|J|-|I|)} O_\kappa\left(\frac{\rho'' K \log^{1/2}(\frac{\sqrt{n}}{\rho_0})}{a_0^2 \rho_0 \theta_0^{1/2}}\right)^{\frac{1}{4}a_0^2 n} \\
&\qquad = O\left(\frac{1}{\rho''}\right)^{2(|J|-|I|)} O_{\kappa,a_0}\left(\rho'' K^2 \log(K\sqrt{n})\right)^{\frac{1}{4}a_0^2 n}
\end{aligned}$$

where in the second line we substituted the expressions for ρ_0, θ_0 from Lemma 4.3.5. Denoting $\varepsilon = \rho'' K^2$, the above bound rearranges to

$$O_{\kappa,a_0}(\log n)^n n^{O(\gamma)} n^{O(\gamma-1/2)(|J|-|I|)} \varepsilon^{\frac{1}{4}a_0^2 n - 2(|J|-|I|)}. \tag{4.92}$$

We can bound

$$\begin{aligned}
|J| - |I| &= \theta m + \beta a_0^2 m - (1-\beta)^2 \theta m + O(1) \\
&\leq \beta a_0^2 m + 2\beta\theta m + O(1) \\
&= O(\beta a_0^2 m) + O(1)
\end{aligned}$$

where we used our assumption that $\theta \leq c_0 a_0^2$. In particular, $|J| - |I| \leq \frac{1}{8}a_0^2 n + O(1)$ if the constant c_1 in (4.88) is sufficiently small, and (4.92) is bounded by

$$O_{\kappa,a_0}(\log n)^n n^{O(\gamma)} n^{O(\gamma-1/2)\beta a_0^2 m} \varepsilon^{\frac{1}{8}a_0^2 n - O(1)}. \tag{4.93}$$

Applying the union bound over the choices of I, J in (4.89), which incurs a harmless factor of $2^{m+n} = O(1)^n$, and substituting the expression (4.88) for β we have

$$\mathbb{P}\Big(\mathcal{E}(\theta_0,\rho_0)^c \wedge \mathcal{E}(\theta,\rho)^c \wedge \mathcal{E}(\theta+\beta a_0^2,\rho')\Big) = O_{\kappa,a_0}(\log n)^n n^{O(\gamma)}\varepsilon^{-O(1)}(n^{O(c_1)}\varepsilon^{1/8})^{a_0^2 n}. \quad (4.94)$$

It only remains to check that we can take ε sufficiently small to obtain (4.87). From (4.91) we are constrained to take

$$\varepsilon = \rho'' K^2 \geq \frac{c a_0 \rho_0 K}{\sqrt{n}} = \frac{c' a_0^3}{\sqrt{n}}$$

for some constant $c' \in (0,1)$ sufficiently small. Taking $\varepsilon = a_0^3/\sqrt{n}$ and c_1 sufficiently small we have

$$\mathbb{P}\Big(\mathcal{E}(\theta_0,\rho_0)^c \wedge \mathcal{E}(\theta,\rho)^c \wedge \mathcal{E}(\theta+\beta a_0^2,\rho')\Big) \leq O_{\kappa,a_0}(1)^n n^{O(\gamma)} n^{-.01 a_0^2 n} \quad (4.95)$$

which yields (4.87) as desired. With this choice of ε,

$$\rho' \gg \rho''\beta\rho\theta \geq \rho''\beta\rho\theta_0 \gg_{\kappa,a_0,\gamma} \rho n^{-2\gamma+1/2-o(1)}$$

as desired (recall that $\theta_0 \gg_\kappa a_0^2/\log(K/a_0) \gg_{\gamma,a_0,\kappa} 1/\log n$). $\square$

Now we conclude the proof of Proposition 4.3.4. Since the event $\mathcal{B}(K)$ is monotone under increasing K, by perturbing γ and assuming n is sufficiently large we may take $K = n^{\gamma-1/2}$ with $\gamma > 1/2$. Let ρ_0, θ_0 be as in Lemma 4.3.5, and for $l \geq 1$ we let $\theta_l = \theta_0 + l\beta a_0^2$ with $\beta = \beta(\gamma)$ as in (4.88). By Lemma 4.3.15 we can inductively define a sequence ρ_l such that for each $l \geq 1$ such that $\theta_l \leq c_0 a_0^2$,

$$\rho_l \gg_{\gamma,a_0,\kappa} n^{-O(\gamma)}\rho_{l-1}$$

and

$$\mathbb{P}(\mathcal{E}(\theta_0,\rho_0)^c \wedge \mathcal{E}(\theta_{l-1},\rho_{l-1})^c \wedge \mathcal{E}(\theta_l,\rho_l)) = O_{\gamma,a_0,\kappa}(e^{-n}).$$

Applying the union bound, for some $k = O(\gamma)$ we have

$$\begin{aligned}\mathbb{P}\big(\mathcal{E}(c_0 a_0^2, \rho)\big) &\le \mathbb{P}(\mathcal{E}(\theta_0, \rho_0)) + \sum_{l=1}^{k} \mathbb{P}(\mathcal{E}(\theta_0, \rho_0)^c \wedge \mathcal{E}(\theta_{l-1}, \rho_{l-1})^c \wedge \mathcal{E}(\theta_l, \rho_l)) \\ &\le e^{-c_\kappa a_0^2 n} + O_{\gamma, a_0, \kappa}(e^{-n}) \\ &= O_{\gamma, a_0, \kappa}(e^{-c_\kappa a_0^2 n})\end{aligned}$$

and

$$\rho \gg_{\gamma, a_0, \kappa} n^{-O(\gamma^2)}.$$

This concludes the proof of Proposition 4.3.4.

4.4 Invertibility from connectivity: Incompressible vectors

In this section we conclude the proofs of Theorems 4.1.14 and 4.1.24 by bounding the event that $\|Mu\|$ is small for some incompressible vector u. We follow the by now standard approach of reducing to the event that a fixed row R_i of M lies close to the span of the remaining rows, an idea which goes back to the work of Komlós on the singularity probability for Bernoulli matrices [Kom67, Kom68, Kom]. This can in turn be controlled by the event that a random walk $R_i \cdot v$ concentrates near a particular point, where v is a fixed unit vector in the orthocomplement of the remaining rows. Independence of the rows allows us to condition on v, and our results from the previous section allow us to argue that v is incompressible. For the case that the entries of R_i have variances uniformly bounded below, we could then complete the proof by applying the anti-concentration estimate from Lemma 4.2.8.

In our case, however, a proportion $1 - \delta$ of the entries of R_i may have zero variance. For the case of broadly connected profile we follow the argument of Rudelson and Zeitouni [RZ16] and use Proposition 4.3.1 to show v has essential support of size $(1 - \delta/2)n$, so that R_i and v must have non-trivial overlap.

For the case of a super-regular profile, Proposition 4.3.4 only gives that v has essential support of size $\gg \delta\sigma_0^2$. In Lemma 4.4.1 we make use of a double counting argument to show that if we choose the row R_i at random, on average it will have good overlap with the corresponding normal vector $v^{(i)}$ (which also depends on i). Here is where we make crucial use of the super-regularity hypothesis on A. Lemma 4.4.1 is a natural extension of a double counting argument used by Komlós in his work on the singularity probability for Bernoulli matrices, and which was applied to bound the smallest singular value of iid matrices by Rudelson and Vershynin in [RV08]. We were also inspired by a similar refinement of the double counting argument from the recent paper [LLT+] on the singularity probability for adjacency matrices of random regular digraphs.

4.4.1 Proof of Theorem 4.1.14

By Lemma 4.2.6 and multiplying X and B by a phase (which does not affect our hypotheses) we may assume that ξ has κ-controlled second moment for some $\kappa \geq 1$. Fix $K \geq 1$, and let $\rho = \rho(\kappa, \sigma_0, \delta, \nu, K)$ be as in Proposition 4.3.1. We may assume n is sufficiently large depending on $\kappa, \sigma_0, \delta, \nu, K$. For the remainder of the proof we restrict to the event $\mathcal{B}(K) = \{\|M\| \leq K\sqrt{n}\}$.

For $j \in [n]$ let $M^{(i)}$ denote the $n-1 \times n$ matrix obtained by removing the ith row from M. Define the good event

$$\mathcal{G} = \Big\{\forall i \in [n], \forall u \in \mathrm{Comp}(1-\delta/2, \rho),\ \|u^*M\|, \|M^{(i)}u\| > \rho K\sqrt{n}\Big\}. \tag{4.96}$$

Applying Proposition 4.3.1 to M^* and $M^{(i)}$ for each $i \in [n]$ (using our restriction to $\mathcal{B}(K)$) and the union bound we have

$$\mathbb{P}(\mathcal{G}) = 1 - O_{\kappa,\sigma_0,\delta,\nu,K}(ne^{-c_\kappa \delta \sigma_0^2 n}) = 1 - O_{\sigma_0,\delta,\nu,K}(e^{-c_\kappa \delta \sigma_0^2 n}) \tag{4.97}$$

adjusting c_κ slightly. Let $t \leq 1$, and define the event

$$\mathcal{E}(t) = \mathcal{G} \wedge \big\{\exists u \in \mathrm{Incomp}(1/10, \rho) : \|u^*M\| \leq t/\sqrt{n}\big\}. \tag{4.98}$$

For n sufficiently large (larger than $1/\rho K$) it suffices to show

$$\mathbb{P}(\mathcal{E}(t)) \ll_{\kappa,\sigma_0,\delta,\nu,K} t + n^{-1/2}. \tag{4.99}$$

Recalling that R_i denotes the ith row of M, we denote

$$R_{-i} = \mathrm{span}(R_j : j \in [n] \setminus \{i\}) \tag{4.100}$$

and let

$$\mathcal{E}_i(t) = \mathcal{G} \wedge \{\mathrm{dist}(R_i, R_{-i}) \leq t/\rho\}. \tag{4.101}$$

We now use a double counting argument of Rudelson and Vershynin from [RV08] to control $\mathcal{E}(t)$ in terms of the events $\mathcal{E}_i(t)$. Suppose that $\mathcal{E}(t)$ holds, and let $u \in \mathrm{Incomp}(1/10, \rho)$ such that $\|u^*M\| \leq t/\sqrt{n}$. Then we must have $|u_i| \geq \rho/\sqrt{n}$ for at least $n/10$ elements $i \in [n]$. For each such i we have

$$\frac{t}{\sqrt{n}} \geq \|u^*M\| = \left\| \sum_{j=1}^n \overline{u_j} R_j \right\| \geq \left\| P_{R_{-i}^\perp} \sum_{j=1}^n \overline{u_j} R_j \right\| = |u_i| \left\| P_{R_{-i}^\perp} R_i \right\| \geq \frac{\rho}{\sqrt{n}} \mathrm{dist}(R_i, R_{-i})$$

where we denote by P_W the orthogonal projection to a subspace W. Thus, on $\mathcal{E}(t)$ we have that $\mathcal{E}_i(t)$ holds for at least $n/10$ values of $i \in [n]$, so by double counting,

$$\mathbb{P}(\mathcal{E}(t)) \leq \frac{10}{n} \sum_{i=1}^n \mathbb{P}(\mathcal{E}_i(t)). \tag{4.102}$$

Now it suffices to show that for arbitrary fixed $i \in [n]$,

$$\mathbb{P}(\mathcal{E}_i(t)) \ll_{\kappa,\sigma_0,\delta,\nu,K} t + n^{-1/2}. \tag{4.103}$$

Fix $i \in [n]$ and condition on $\{R_j : j \in [n] \setminus \{i\}\}$. Draw a unit vector $u \in R_{-i}^\perp$ independent of R_i, according to Haar measure (say). Since $\mathrm{dist}(R_i, R_{-i}) \leq |R_i \cdot u|$, it suffices to show

$$\mathbb{P}(|R_i \cdot u| \leq t/\rho) \ll_{\kappa,\sigma_0,\delta,\nu,K} t + n^{-1/2}. \tag{4.104}$$

Since $u \in \ker(M^{(i)})$, on $\mathcal{G}$ we have that $u \in \text{Incomp}(1-\frac{\delta}{2},\rho)$. By Lemma 4.2.1 there exists $L \subset [n]$ of size $|L| \geq (1-\frac{3}{4}\delta)n$ such that

$$\frac{\rho}{\sqrt{n}} \leq |u_j| \leq \frac{10}{\sqrt{\delta n}}$$

for all $j \in L$. By assumption we have $|\mathcal{N}_{A(\sigma_0)}(i)| = |\{j \in [n] : a_{ij} \geq \sigma_0\}| \geq \delta n$, so letting $J = \mathcal{N}_{A(\sigma_0)}(i) \cap L$ we have $|J| \geq \delta n/4$. Denoting $v = (u^i)_J = (a_{ij}u_j 1_{j\in J})_j$, we have

$$\|v\|^2 = \sum_{j\in J} a_{ij}^2 |u_j|^2 \geq |J|\sigma_0^2\rho^2/n \geq \delta\sigma_0^2\rho^2/4$$

and

$$\|v\|_\infty \leq \|u_J\|_\infty \leq \frac{10}{\sqrt{\delta n}}$$

(recall that $a_{ij} \leq 1$ for all $i,j \in [n]$). Conditioning on u and $\{\xi_{ij}\}_{j\notin J}$, we apply Lemma 4.2.8 to conclude

$$\mathbb{P}(|R_i \cdot u| \leq t/\rho) \ll_\kappa \frac{1}{\|v\|}\left(\frac{t}{\rho} + \|v\|_\infty\right) \ll \frac{1}{\rho\sigma_0\delta^{1/2}}\left(\frac{t}{\rho} + \frac{1}{\sqrt{\delta n}}\right)$$

which gives (4.104) as desired.

4.4.2 Proof of Theorem 4.1.24

By Lemma 4.2.6 and multiplying X and B by a phase (which does not affect our hypotheses) we may assume that ξ has κ-controlled second moment for some $\kappa \geq 1$. Fix $\gamma \geq 1/2$ let $K = O(n^{\gamma-1/2})$. We will show that for all $\tau \geq 0$,

$$\mathbb{P}\left(s_n(M) \leq \frac{\tau}{\sqrt{n}},\ \|M\| \leq K\sqrt{n}\right) \ll_{\gamma,\sigma_0,\delta,\kappa} n^{O(\gamma^2)}\tau + \sqrt{\frac{\log n}{n}}. \tag{4.105}$$

For the remainder of the proof we restrict to the boundedness event

$$\mathcal{B}(K) = \{\|M\| \leq K\sqrt{n}\}. \tag{4.106}$$

By the assumption that $A(\sigma_0)$ is (δ, ε)-super-regular we have

$$\sum_{i=1}^{n} a_{ij}^2 \geq \delta \sigma_0^2 n$$

for all $j \in [n]$. Let $a_0 = \delta^{1/2}\sigma_0$, and let $\rho = \rho(\gamma, a_0, \kappa n)$ and c_0 be as in Proposition 4.3.4. In particular,

$$\rho \gg_{\gamma,\delta,\sigma_0} n^{-O(\gamma^2)}. \tag{4.107}$$

Denoting $\theta = c_0 \delta \sigma_0^2$, for $\tau > 0$ we define the good event

$$\mathcal{G}(\tau) = \Big\{ \forall u \in \mathrm{Comp}(\theta, \rho),\ \|Mu\|, \|u^* M\| > \tau/\sqrt{n} \Big\}. \tag{4.108}$$

Applying Proposition 4.3.4 to M and M^*, along with the union bound, we have

$$\mathbb{P}(\mathcal{G}(\tau)) = 1 - O_{\gamma,\delta,\sigma_0,\kappa}(e^{-c_\kappa \delta \sigma_0^2 n}) \tag{4.109}$$

as long as $\tau \leq \rho K n$.

Let $0 < \tau \leq 1$ to be chosen later. Recall our notation $M^{(i)}$ from Section 4.4.1, we define the sets

$$S_i(\tau) = \left\{ u \in S^{n-1} : \|M^{(i)}\| \leq \frac{\tau}{\sqrt{n}} \right\}. \tag{4.110}$$

Informally, for small τ this is the set of unit almost-normal vectors to the subspace R_{-i} spanned by the rows of $M^{(i)}$. Writing $\mathcal{N}(i) = \mathcal{N}_{A(\sigma_0)}(i)$, we define the *good overlap events*

$$\mathcal{O}_i(\tau) = \big\{ \exists u \in S_i(\tau) : |\mathcal{N}(i) \cap L^+(u, \rho)| \geq \delta\theta n \big\} \tag{4.111}$$

where

$$L^+(u) = \{ j \in [n] : |u_j| \geq \rho/\sqrt{n} \}. \tag{4.112}$$

On $\mathcal{O}_i(\tau)$ we fix a vector $u^{(i)} = u^{(i)}(M^{(i)}, \tau) \in S_i(\tau)$, chosen measurably with respect to $M^{(i)}$, satisfying $|\mathcal{N}(i) \cap L^+(u, \rho)| \geq \delta\theta n$.

Lemma 4.4.1 (Good overlap on average)**.** *Assume $\varepsilon \le \theta/2$. Then*

$$\mathbb{P}\Big(\mathcal{G}(\tau) \wedge \Big\{s_n(M) \le \frac{\tau}{\sqrt{n}}\Big\}\Big) \le \frac{2}{\theta n}\sum_{i=1}^n \mathbb{P}\Big(\mathcal{O}_i(\tau) \wedge \Big\{|R_i \cdot u^{(i)}| \le \frac{2\tau}{\rho}\Big\}\Big). \tag{4.113}$$

Proof. Suppose $\mathcal{G}(\tau) \wedge \{s_n(M) \le \tau/\sqrt{n}\}$ holds. Then there exist $u, v \in S^{n-1}$ such that $\|Mu\|, \|M^*v\| \le \tau/\sqrt{n}$. By our restriction to $\mathcal{G}(\tau)$ we must have $u, v \in \mathrm{Incomp}(\theta, \rho)$. With notation as in (4.112) we have $|L^+(u)|, |L^+(v)| \ge \theta n$. In particular, $|L^+(u)| \ge \varepsilon n$, so

$$|\mathcal{N}(i) \cap L^+(u)| \ge \delta |L^+(u)| \ge \delta\theta n \tag{4.114}$$

for at least $(1-\varepsilon)n$ elements $i \in [n]$. Indeed, otherwise we would have

$$e_{A(\sigma_0)}(I, L^+(u)) = \sum_{i \in I} |\mathcal{N}(i) \cap L^+(u)| < \delta |I||L^+(u)|$$

for some $I \subset [n]$ with $|I| > \varepsilon n$, which contradicts our assumption that $A(\sigma_0)$ is (δ, ε)-super-regular. Since

$$\|M^{(i)}u\| \le \|Mu\| \le \frac{\tau}{\sqrt{n}}$$

for all $i \in [n]$, we have that $u \in S_i(\tau)$ for all $i \in [n]$. Thus,

$$\big|\big\{i \in L^+(v) : \mathcal{O}_i(\tau) \text{ holds}\big\}\big| \ge \theta n - \varepsilon n \ge \theta n/2. \tag{4.115}$$

Fix $i \in L^+(v)$ such that $\mathcal{O}_i(\tau)$ holds. We have

$$\frac{\tau}{\sqrt{n}} \ge \|v^*M\| \ge |v^*Mu^{(i)}| \ge |v_i||R_i \cdot u^{(i)}| - \Big|\sum_{j \ne i} \overline{v_j} R_j \cdot u^{(i)}\Big|.$$

The first term on the right hand side is bounded below by $\frac{\rho}{\sqrt{n}}|R_i \cdot u^{(i)}|$ since $i \in L^+(v)$. By Cauchy–Schwarz the second term is bounded above by $\|M^{(i)}u^{(i)}\| \le \tau/\sqrt{n}$, since $u^{(i)} \in S_i(\tau)$. Rearranging we conclude

$$|R_i \cdot u^{(i)}| \le \frac{2\tau}{\rho}$$

for all $i \in L^+(v)$ such that $\mathcal{O}_i(\tau)$ holds, and the claim follows from (4.115) and double counting. $\square$

Fix $i \in [n]$ arbitrarily, and suppose that $\mathcal{O}_i(\tau)$ holds. We condition on the rows $\{R_j\}_{j\in[n]\setminus\{i\}}$ to fix $u^{(i)}$. We begin by finding a large set on which $u^{(i)}$ is flat, following a similar dyadic pigeonholing argument as in the proof of Lemma 4.3.13. Letting $L_k = \{j \in [n] : 2^{-(k+1)} < |u_j^{(i)}| \le 2^{-k}$, since

$$\delta\theta n \le |\mathcal{N}(i) \cap L^+(u^{(i)})| \le \left| \bigcup_{k=0}^{\ell} \mathcal{N}(i) \cap L_k \right|$$

for some $\ell \ll \log(\sqrt{n}/\rho)$, by the pigeonhole principle there exists $k^* \le \ell$ such that $J := \mathcal{N}(i) \cap L_{k^*}$ satisfies

$$|J| \ge \delta\theta n/\ell \gg \frac{\delta\theta n}{\log(\sqrt{n}/\rho)}. \tag{4.116}$$

Let us denote $v = (a_{ij}u_j^{(i)} 1_{j\in J})_j$. Since $a_{ij} \ge \sigma_0$ for $j \in \mathcal{N}(i)$ and $|u_j^{(i)}| \gg \rho/\sqrt{n}$ for $j \in L_{k^*}$,

$$\|v\| \ge \sigma_0 \|(u^{(i)})_J\| \gg \sigma_0\rho(|J|/n)^{1/2} \tag{4.117}$$

and since $u^{(i)}$ varies by at most a factor of 2 on J,

$$\|v\|_\infty \le \|u^{(i)}1_J\|_\infty \le 2\|u^{(i)}\|/|J|^{1/2}. \tag{4.118}$$

By further conditioning on the variables $\{\xi_{ij}\}_{j\notin J}$ and applying Lemma 4.2.8 along with the estimates (4.117), (4.118) we have

$$\begin{aligned}
\mathbb{P}\big(|R_i \cdot u^{(i)}| \le 2\tau/\rho\big) &\ll_\kappa \frac{\tau/\rho + \|v\|_\infty}{\|v\|} \\
&\ll \frac{1}{\sigma_0}\left(\frac{\tau/\rho}{\rho(|J|/n)^{1/2}} + \frac{1}{|J|^{1/2}}\right) \\
&= \frac{1}{\sigma_0}\left(\frac{n}{|J|}\right)^{1/2}\left(\frac{\tau}{\rho^2} + \frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Inserting the bound (4.116) and undoing all of the conditioning, we have shown

$$\mathbb{P}\bigg(\mathcal{O}_i(\tau) \wedge \left\{|R_i \cdot u^{(i)}| \le \frac{2\tau}{\rho}\right\}\bigg) \ll_\kappa \frac{1}{\sigma_0\sqrt{\delta\theta}}\left(\frac{\tau}{\rho^2} + \frac{1}{\sqrt{n}}\right)\log^{1/2}(\sqrt{n}/\rho).$$

Since the right hand side is uniform in i, applying Lemma 4.4.1 (taking $c_1 = c_0/2$) and substituting the expression for θ we have

$$\mathbb{P}\bigg(\mathcal{G}(\tau) \wedge \left\{s_n(M) \le \frac{\tau}{\sqrt{n}}\right\}\bigg) \ll_\kappa \frac{1}{\sigma_0^4\delta^2}\left(\frac{\tau}{\rho^2} + \frac{1}{\sqrt{n}}\right)\log^{1/2}(\sqrt{n}/\rho) \tag{4.119}$$

for all $\tau \ge 0$ (note that this bound is only nontrivial when $\tau \le \rho^2$, in which case our constraint $\tau \le \rho K n$ from (4.109) holds). The bound (4.105) now follows by substituting the lower bound (4.107) on ρ and the the bound (4.109) on $\mathcal{G}(\tau)^c$ (which is dominated by the $O(n^{-1/2}\log^{1/2} n)$ term). This concludes the proof of Theorem 4.1.24.

4.5 Control of moderately small singular values

Recall the boundedness event

$$\mathcal{B}(K) = \big\{\|M\| \le K\sqrt{n}\big\}. \tag{4.120}$$

In this section we establish the following:

Theorem 4.5.1 (Control of moderately small singular values, broadly connected profile)**.** *Let $M = A \circ X + B$ be an $n \times n$ matrix as in Definition 4.1.1. Assume that ξ has κ-controlled second moment for some $\kappa \ge 1$, and that $A(\sigma_0)$ is (δ, ν)-broadly connected for some $\sigma_0, \delta, \nu \in (0,1)$. Let $K \ge 1$ and $\varepsilon \in (0,1)$. There are constants $a_0, a_1, a_2 > 0$ depending (polynomially) on $\kappa, \sigma_0, \delta, \nu, K$ such that*

$$\mathbb{P}\bigg(\mathcal{B}(K) \wedge \left\{\exists k \in [n^{2\varepsilon}, a_1 n] : s_{n-k}(M) < a_2\frac{k}{\sqrt{n}}\right\}\bigg) \ll \exp\left(-a_0 n^\varepsilon\right) \tag{4.121}$$

where the implied constant depends on $\kappa, \sigma_0, \delta, \nu, K$ and ε.

We follow the approach introduced by Tao and Vu in [TV10b]. The key idea is to reduce the problem of controlling small singular values to the problem of bounding the distance of a row of M to the span of all but a small number of the remaining rows – we make this reduction in Lemma 4.5.2 below. The distance between a row and a subspace can then be controlled using standard concentration tools (specifically, Talagrand's inequality). The main new difficulty in the present work stems from the fact that the entries of M are not identically distributed, which in general can cause the *expected* distance between a row and a fixed subspace to be quite small. However, in Lemma 4.5.3 we get around this obstruction by assuming that the subspace has *incompressible normal vectors*. Finally, under the assumption that the standard deviation profile A is broadly connected we can use Proposition 4.3.1 to argue that with high probability, any normal vector to the span of all but a small number of rows is highly incompressible.

Lemma 4.5.2. *Let $M \in \mathcal{M}_n(\mathbb{C})$ and $0 \le k \le n-1$. Put $m = n - \lceil k/2 \rceil$. Denote the rows of M by $R_1, \dots, R_n$, and for $i \in [m]$ denote*

$$R_{-i} := \operatorname{span}\big\{R_j : j \in [m] \setminus \{i\}\big\}. \tag{4.122}$$

We have

$$s_{n-k}(M) \gg \sqrt{\frac{k}{n}} \min_{i\in[m]} \operatorname{dist}(R_i, R_{-i}). \tag{4.123}$$

Proof. We follow the argument from [TV10b]. Denote $M' = M_{[m]\times[n]}$, the matrix obtained by removing the last $\lceil k/2 \rceil$ rows from M. By the Cauchy interlacing law,

$$s_{n-k}(M) \ge s_{n-k}(M'). \tag{4.124}$$

On the other hand, from the inverse second moment identity (cf. [TV10b, Lemma A.4]) we have

$$\sum_{i=1}^{n} s_i(M')^{-2} = \sum_{i=1}^{M} \operatorname{dist}(R_i, R_{-i})^{-2} \tag{4.125}$$

and so

$$
\begin{aligned}
m\Big(\min_{i\in[m]} \operatorname{dist}(R_i, R_{-i})\Big)^{-2} &\geq \sum_{i=1}^{m} \operatorname{dist}(R_i, R_{-i})^{-2} \\
&\geq \sum_{i=n-k}^{n-\lceil k/2\rceil} s_i(M')^{-2} \\
&\geq \frac{k}{2} s_{n-k}(M')^{-2}.
\end{aligned}
$$

(4.123) now follows from the above and (4.124) (noting that $m \geq n/2$). $\square$

Lemma 4.5.3 (Distance of a random vector to an incompressible subspace)**.** *Let ξ be a centered complex-valued random variable with unit variance, and let $X = (\xi_1, \dots, \xi_n)$ be a vector of iid copies of ξ. Let $a = (\sigma_1, \dots, \sigma_n) \in [0,1]^n$, and put $R = X \circ a = (\sigma_j \xi_j)_{j=1}^n$. Suppose that for some $\delta, \sigma_0 \in (0,1)$ we have*

$$|L| := |\{j \in [n] : \sigma_j \geq \sigma_0\}| \geq \delta n.$$

Let $\varepsilon \in (0,1)$ and let $V \subset \mathbb{C}^n$ be a subspace of dimension $n-k$. Suppose that for some $\rho > 0$ and any unit vector $u \in V^\perp$ we have $u \in \operatorname{Incomp}((1-\frac{\delta}{2}), \rho)$. Then for any fixed $v \in \mathbb{C}^n$ we have

$$\mathbb{P}\Big(\operatorname{dist}(R+v, V) \leq c\rho\sqrt{\delta k}\Big) = O_\varepsilon(\exp(-c\sigma_0^2\rho^2\delta k/n^\varepsilon)) \tag{4.126}$$

if

$$\frac{Cn^\varepsilon}{\sigma_0^2\rho^2\delta} \leq k \leq n-1 \tag{4.127}$$

for a sufficiently large constant $C > 0$.

Proof. Fix $\varepsilon \in (0,1)$. We may assume n is sufficiently large depending on ε. By replacing V with $\operatorname{span}(V, v)$ and k with $k-1$ we may take $v = 0$.

Towards an application of concentration of measure, we first perform a truncation. By Chebyshev's inequality, for all $j \in \mathbb{P}(|\xi_j| \geq n^{\varepsilon/2}) \leq n^{-\varepsilon}$. It follows from Hoeffding's inequality that

$$\mathbb{P}\big(|\{j \in [n] : |\xi_j| \leq n^{\varepsilon/2}\}| \geq n - n^{1-\varepsilon/2}\big) \geq 1 - \exp(-cn^{1-\varepsilon}). \tag{4.128}$$

Put $m = \lceil n - n^{1-\varepsilon/2} \rceil$, and for $J \in \binom{[n]}{m}$ denote the event

$$\mathcal{E}_J := \{|\xi_j| \le n^{\varepsilon/2} \; \forall j \in J\}. \tag{4.129}$$

It suffices to obtain control of the lower tail of $\operatorname{dist}(R, V)$ conditional on $\mathcal{E}_J$ that is uniform in the choice J. For $j \in [n]$ let

$$\lambda := \mathbb{E}\left(\xi_j \,\middle|\, |\xi_j| \le n^{\varepsilon/2}\right), \qquad \xi_j' := \xi_j - \lambda$$

and write $R' = (\xi_1', \dots, \xi_n')$. We have

$$\tau^2 := \mathbb{E}\left(|\xi_j'|^2 \,\middle|\, |\xi_j| \le n^{\varepsilon/2}\right) \ge 1/2.$$

Fix $J \in \binom{[n]}{m}$. Condition on a realization of $\{\xi_j\}_{j \notin J}$, and write $\mathbb{P}_J(\,\cdot\,)$ and $\mathbb{E}_J(\,\cdot\,)$ for probability and expectation conditional on $\{\xi_j\}_{j \notin J}$. Let $W = \operatorname{span}(V, R_{[n]\setminus J}, \lambda a_J)$; note that W is deterministic under the conditioning on $\{\xi_j\}_{j \notin J}$. Then $\dim(W) \le \dim(V) + 2$ and

$$\operatorname{dist}(R, V) \ge \operatorname{dist}(R', W). \tag{4.130}$$

Note that $\operatorname{dist}(R', W) = \|P_{W^\perp} R\|$, where $P_{W^\perp}$ is the orthogonal projection to $W^\perp$. Letting $u^1, \dots, u^{k-2}$ be an arbitrary set of orthonormal vectors in $W^\perp$, we have

$$\mathbb{E}_J \operatorname{dist}(R', W)^2 \ge \sum_{i=1}^{k-2} \mathbb{E}\,|R' \cdot u^i|^2 = \tau^2 \sum_{i=1}^{k-2} \sum_{j=1}^{n} |u_j^i \sigma_j|^2 \ge \frac{1}{2}\sigma_0^2 \sum_{i=1}^{k-2} \|(u^i)_L\|^2. \tag{4.131}$$

For each $1 \le i \le k-2$, since $u^i \in \operatorname{Incomp}((1-\frac{\delta}{2}), \rho)$ there is a set $J_i \subset [n]$ with $|J_i| \ge (1-\frac{\delta}{2})n$ and $|u_j^i| \ge \rho/\sqrt{n}$ for all $j \in J_i$. Then we have $\|(u^i)_L\|^2 \ge \|(u^i)_{L \cap J_i}\|^2 \ge \frac{1}{2}\delta\rho^2$. Hence,

$$\mathbb{E}_J \operatorname{dist}(R', W)^2 \ge c\delta\rho^2\sigma_0^2 k \tag{4.132}$$

for some absolute constant $c > 0$.

It remains to obtain a lower tail bound for $\operatorname{dist}(R', W)^2$. Note that $v \mapsto \operatorname{dist}(\,\cdot\,, W)$ is a convex 1-Lipschitz function on $\mathbb{C}^J$. Since the components of R' are bounded in magnitude by $2n^{\varepsilon/2}$, by Talagrand's concentration inequality [Tal96, Theorem 6.6] (see also [AGZ10, Corollary 4.4.11]) we have

$$\mathbb{P}_J(|\operatorname{dist}(R', W) - d| \geq t) = O(\exp(-ct^2/n^\varepsilon)) \tag{4.133}$$

for any $t \geq 0$, where d is any median for $\operatorname{dist}(R', W)$ conditional on $\{\xi_j\}_{j \notin J}$. In particular, $|\mathbb{E}_J \operatorname{dist}(R', W) - d| = O(n^{\varepsilon/2})$ and $\operatorname{Var}(\operatorname{dist}(R', W)) = O(n^\varepsilon)$, so

$$d = \sqrt{\mathbb{E} \operatorname{dist}(R', W)^2} + O(n^{\varepsilon/2}).$$

Together with (4.132) these estimates imply

$$\mathbb{P}_J\left(\operatorname{dist}(R', W) \leq c\sigma_0\rho\sqrt{\delta k} - Cn^{\varepsilon/2}\right) = O\big(\exp(-c\sigma_0^2\rho^2\delta k/n^\varepsilon)\big) \tag{4.134}$$

for some absolute constants $C, c > 0$. Undoing the conditioning on $\{\xi_j\}_{j \notin J}$, from the above and (4.130) we have

$$\mathbb{P}\left(\operatorname{dist}(R, V) \leq c\sigma_0\rho\sqrt{\delta k} - Cn^{\varepsilon/2} \,\middle|\, \mathcal{E}_J\right) = O\big(\exp(-c\sigma_0^2\rho^2\delta k/n^\varepsilon)\big). \tag{4.135}$$

Summing over J and using (4.128), we conclude

$$\mathbb{P}\Big(\operatorname{dist}(R, V) \leq c\sigma_0\rho\sqrt{\delta k} - Cn^{\varepsilon/2}\Big) = O(\exp(-c\sigma_0^2\rho^2\delta k/n^\varepsilon))$$

and the result follows from the lower bound in (4.127). $\square$

Proof of Theorem 4.5.1. For the duration of the proof we restrict to the event $\mathcal{B}(K)$. We may assume that n is sufficiently large depending on $\kappa, \varepsilon, \sigma_0, \delta, \nu$, and K. By the union bound it suffices to show that for a_2 sufficiently small depending on $\kappa, \sigma_0, \delta, \nu, K$,

$$\mathbb{P}\big(s_{n-k}(M) \leq a_2 k/\sqrt{n}\big) = O(n \exp(-a_0 n^\varepsilon)) \tag{4.136}$$

for arbitrary fixed $k \in [n^{2\varepsilon}, a_1 n]$. By Lemma 4.5.2 and another application of the union bound, after modifying a_2 by a constant factor it suffices to show

$$\mathbb{P}\Big(\mathrm{dist}(R_i, V_I) \le a_2\sqrt{k}\Big) = O(\exp(-a_0 n^{\varepsilon})) \tag{4.137}$$

where $V_I = \mathrm{span}\{R_i, i \in I\}$ for an arbitrary fixed subset $I \subset [n]$ with $|I| = n - \lfloor k/2 \rfloor - 1 =: n'$ and arbitrary fixed $i \in [n] \setminus I$.

Fix such $I \subset [n]$ and $i \in [n] \setminus I$. Let θ_0 be as in Proposition 4.3.1 and let $J \subset [n]$ with $|J| \ge \theta_0 n$. Denoting $\widetilde{A} := A(\sigma_0)_{I\times[n]}$, by the assumption that $A(\sigma_0)$ is (δ, ν)-broadly connected we have

$$|\mathcal{N}^{(\delta)}_{\widetilde{A}^{\mathsf{T}}}(J)| \ge \min\big(n', (1+\nu)|J| - \lceil k/2 \rceil - 1\big) \ge \min\big(n', (1+\nu/2)|J|\big)$$

taking $a_1 \le c\nu\theta_0$ for a sufficiently small constant $c > 0$. Applying Proposition 4.3.1 to $M_{I\times[n]}$ (with (n', n) in place of (n, m) and $(\delta/2, \nu/2)$ in place of (δ, ν)) we have that except with probability $O(\exp(-\frac{1}{2}c_\kappa \delta \sigma_0^2 n))$,

$$\|M_{I\times[n]}u\| \ge \rho K\sqrt{n} \quad \forall u \in \mathrm{Comp}\left(\left(1 - \frac{\delta}{4}\right)\frac{n'}{n}, \rho\right) \tag{4.138}$$

for some $\rho = \rho(\kappa, \sigma_0, \delta, \nu, K) > 0$. In particular, taking a_1 smaller if necessary, we have

$$\left(1 - \frac{\delta}{4}\right)\frac{n'}{n} \ge \left(1 - \frac{\delta}{4}\right)\left(1 - \frac{a_1}{2}\right) \ge 1 - \frac{\delta}{2}$$

and so

$$S^{n-1} \cap V_I^{\perp} = \ker(M_{I\times[n]}) \subset \mathrm{Incomp}(1 - \delta/2, \rho). \tag{4.139}$$

We may condition on the event that (4.138) holds. Conditional on $M_{I\times[n]}$, we apply Lemma 4.5.3 with $R + v = R_i$ and $V = V_I$ to obtain (4.137) as desired. □

4.6 Invertibility under diagonal perturbation: Proof of Theorem 4.1.17

4.6.1 Preliminary Tools

In this subsection we state two key lemmas: a version of Szemerédi's regularity lemma for directed graphs, and a lower bound for the smallest singular value given control on the singular values of block submatrices using Schur complements.

Recall that in Theorem 4.1.14 we associated the standard deviation profile A with a bipartite graph. Here it will be more convenient to associate A with a directed graph. That is, to a non-negative square matrix $A = (\sigma_{ij})_{1\le i,j\le n}$ we associate a directed graph Γ_A on vertex set $[n]$ having an edge $i \to j$ when $\sigma_{ij} > 0$ (note that we allow Γ_A to have self-loops, though the diagonal of A will have a negligible effect on our arguments). We continue to use the notation (4.15)–(4.16) (while we introduced this notation with a bipartite graph in mind, it applies equally well in the present setting). Additionally, we denote the *density* of the pair (I, J)

$$\rho_A(I, J) := \frac{e_A(I, J)}{|I||J|}.$$

Definition 4.6.1 (Regular pair)**.** Let A be an $n \times n$ matrix with non-negative entries. For $\varepsilon > 0$, we say that a pair of vertex subsets $I, J \subset [n]$ is *ε-regular for A* if for every $I' \subset I, J' \subset J$ satisfying

$$|I'| > \varepsilon|I|, \quad |J'| > \varepsilon|J|$$

we have

$$|\rho_A(I', J') - \rho_A(I, J)| < \varepsilon.$$

The following is a version of the regularity lemma for directed graphs suitable for our purposes, and follows quickly from a stronger result of Alon and Shapira [AS04, Lemma 3.1].We will apply this in Section 4.6.3 to partition the standard deviation profile into a bounded number of manageable pieces (see Section 4.1.3 for more motivation).

Lemma 4.6.2 (Regularity Lemma)**.** *Let $\varepsilon > 0$. There exists $m_0 \in \mathbb{N}$ with $\varepsilon^{-1} \le m_0 = O_\varepsilon(1)$ such that for all n sufficiently large depending on ε, for every $n \times n$ non-negative matrix A*

there is a partition of $[n]$ into m_0+1 sets $I_0, I_1, \ldots, I_{m_0}$ with the following properties:

1. $|I_0| < \varepsilon n$;

2. $|I_1| = |I_2| = \cdots = |I_{m_0}|$;

3. *all but at most εm_0^2 of the pairs (I_k, I_l) are ε-regular for A.*

Remark 4.6.3. The dependence on ε of the bound $m_0 \leq O_\varepsilon(1)$ is very bad: a tower of exponentials of height $O(\varepsilon^{-c})$. Indeed, as in Szemerédi's proof for the setting of bipartite graphs [Sze78], the proof in [AS04] gives such a bound with $c = 5$. It was shown by Gowers that for undirected graphs one cannot do better than $c = 1/16$ in general [Gow97]. As remarked in [AS04], his argument carries over to give a similar result for directed graphs.

Lemma 4.6.4 (Schur complement bound)**.** *Let $M \in \mathcal{M}_{N+n}(\mathbb{C})$, which we write in block form as*

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

for $A \in \mathcal{M}_N(\mathbb{C}), B \in \mathcal{M}_{N,n}(\mathbb{C}), C \in \mathcal{M}_{n,N}(\mathbb{C}), D \in \mathcal{M}_n(\mathbb{C})$. Assume that D is invertible. Then

$$s_{N+n}(M) \geq \left(1 + \frac{\|B\|}{s_n(D)}\right)^{-1} \left(1 + \frac{\|C\|}{s_n(D)}\right)^{-1} \min\Big(s_n(D),\ s_N(A - BD^{-1}C)\Big). \tag{4.140}$$

Proof. From the identity

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} I_N & BD^{-1} \\ 0 & I_n \end{pmatrix} \begin{pmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} I_N & 0 \\ D^{-1}C & I_n \end{pmatrix}$$

we have

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} I_N & 0 \\ -D^{-1}C & I_n \end{pmatrix} \begin{pmatrix} (A - BD^{-1}C)^{-1} & 0 \\ 0 & D^{-1} \end{pmatrix} \begin{pmatrix} I_N & -BD^{-1} \\ 0 & I_n \end{pmatrix}.$$

We can use the triangle inequality to bound the operator norm of the first and third matrices on the right hand side by $1 + \|BD^{-1}\|$ and $1 + \|CD^{-1}\|$, respectively. Now by submultiplicativity of the operator norm,

$$\begin{aligned}\|M^{-1}\| &\leq (1+\|BD^{-1}\|)(1+\|D^{-1}C\|)\max(\|(A-BD^{-1}C)^{-1}\|, \|D^{-1}\|)\\ &\leq \left(1+\frac{\|B\|}{s_n(D)}\right)\left(1+\frac{\|C\|}{s_n(D)}\right)\max(\|(A-BD^{-1}C)^{-1}\|, \|D^{-1}\|).\end{aligned}$$

The bound (4.140) follows after taking reciprocals. □

4.6.2 Control on the operator norm

The following lemma summarizes the control we will need on the operator norm of submatrices and products of submatrices of M.

Lemma 4.6.5 (Control on the operator norm)**.** *Let $\xi \in \mathbb{C}$ be a centered random variable with $\mathbb{E}|\xi|^{4+\eta} \leq 1$ for some $\eta \in (0,1)$. Let $\theta \in (0,1)$. Then the following hold for all $n \geq 1$:*

(a) (Control for sparse matrices) If $A \in \mathcal{M}_n([0,1])$ is a fixed matrix and $X = (\xi_{ij})$ is an $n \times n$ matrix of iid copies of ξ, then

$$\|A \circ X\| \ll \tau\sqrt{n} \tag{4.141}$$

except with probability $O_\tau(n^{-\eta/8})$, where $\tau = \tau(A) \in [0,1]$ is any number such that

$$\sum_{k=1}^n a_{ik}^2, \sum_{k=1}^n a_{kj}^2 \leq \tau^2 n \tag{4.142}$$

for all $i, j \in [n]$, and

$$\sum_{i,j=1}^n a_{ij}^4 \leq \tau^4 n. \tag{4.143}$$

(b) (Control for matrix products) Let $m \in [\theta n, n]$. If $A \in \mathcal{M}_{n,m}([0,1])$ and $D \in \mathcal{M}_{m,n}(\mathbb{C})$ are fixed matrices with $\|D\| \leq 1$, and $X = (\xi_{ij})$ is an $n \times m$ matrix of iid copies of ξ, then

$$\|D(A \circ X)\| \ll_\eta \sqrt{m} \tag{4.144}$$

except with probability $O_\theta(n^{-\eta/8})$.

Remark 4.6.6. The probability bounds in the above lemma can be improved under higher moment assumptions on ξ, and improve to exponential bounds under the assumption that ξ has finite sub-Gaussian moment, i.e. $\mathbb{E}\exp(t|\xi|^2) \le 2$ for some $t > 0$.

We will use standard truncation arguments to deduce Lemma 4.6.5 from the following bounds on the expected operator norm of random matrices due to Latała and Vershynin.

Theorem 4.6.7 (Latała [Lat05]). *Let n, m be sufficiently large and let Y be an $n \times m$ random matrix with independent, centered entries $Y_{ij} \in \mathbb{R}$ having finite fourth moment. Then*

$$\mathbb{E}\|Y\| \ll \max_{i\in[n]} \left(\sum_{j=1}^{m} \mathbb{E}\, Y_{ij}^2\right)^{1/2} + \max_{j\in[m]} \left(\sum_{i=1}^{n} \mathbb{E}\, Y_{ij}^2\right)^{1/2} + \left(\sum_{i=1}^{n}\sum_{j=1}^{m} \mathbb{E}\, Y_{ij}^4\right)^{1/4}. \tag{4.145}$$

Theorem 4.6.8 (Vershynin [Ver11]). *Let $\eta \in (0,1)$ and n, m, N sufficiently large natural numbers. Let $D \in \mathcal{M}_{m,N}(\mathbb{R})$ be a deterministic matrix satisfying $\|D\| \le 1$ and $Y \in \mathcal{M}_{N,n}(\mathbb{R})$ be a random matrix with independent centered entries Y_{ij} satisfying $\mathbb{E}|Y_{ij}|^{4+\eta} \le 1$. Then*

$$\mathbb{E}\|DY\| \ll_\eta \sqrt{n} + \sqrt{m}. \tag{4.146}$$

Proof of Lemma 4.6.5. We begin with (a). By splitting X into real and imaginary parts and applying the triangle inequality we may assume ξ is a real-valued random variable. Set $\eta_0 = \min(1/4, \eta/32)$ and define the product event

$$\mathcal{E} = \bigwedge_{i,j=1}^{n} \mathcal{E}_{ij}; \qquad \mathcal{E}_{ij} = \left\{|\xi_{ij}| \le n^{1/2-\eta_0}\right\}. \tag{4.147}$$

By Markov's inequality,

$$\mathbb{P}(\mathcal{E}_{ij}^c) \le n^{-(4+\eta)(1/2-\eta_0)} \le n^{-1} \tag{4.148}$$

for all $i, j \in [n]$. By the union bound,

$$\mathbb{P}(\mathcal{E}^c) \le n^2 n^{-(4+\eta)(1/2-\eta_0)} \le n^{-\eta/8}. \tag{4.149}$$

We denote

$$X' = (\xi'_{ij}) = (\xi_{ij} - \mathbb{E}\,\xi_{ij}\,\mathbb{1}_{\mathcal{E}_{ij}}) = X - \mathbb{E}(X\,\mathbb{1}_{\mathcal{E}}).$$

First we show

$$\|A \circ \mathbb{E}(X\,\mathbb{1}_{\mathcal{E}})\| \ll \tau\sqrt{n}. \tag{4.150}$$

Since the variables ξ_{ij} are centered,

$$|\,\mathbb{E}(\xi_{ij}\,\mathbb{1}_{\mathcal{E}_{ij}})| = |\,\mathbb{E}(\xi_{ij}\,\mathbb{1}_{\mathcal{E}_{ij}^c})|. \tag{4.151}$$

By two applications of Hölder's inequality and (4.148),

$$|\,\mathbb{E}(\xi_{ij}\,\mathbb{1}_{\mathcal{E}_{ij}^c})| \le (\mathbb{E}\,|\xi_{ij}|^4)^{1/4}\,\mathbb{P}(\mathcal{E}_{ij}^c)^{3/4} \le n^{-3/4}.$$

Thus,

$$\|A \circ \mathbb{E}(X\,\mathbb{1}_{\mathcal{E}})\| \le \|A \circ \mathbb{E}(X\,\mathbb{1}_{\mathcal{E}})\|_{\mathrm{HS}} \le n^{-3/4}\|A\|_{\mathrm{HS}} \le \tau n^{1/4} \tag{4.152}$$

which yields (4.150) with room to spare.

Now from (4.149), (4.150) and the triangle inequality it is enough to show

$$\mathbb{P}\big(\mathcal{E} \wedge \big\{\|A \circ X'\| \ge C\tau\sqrt{n}\big\}\big) = O_\tau(n^{-\eta/8}) \tag{4.153}$$

for a sufficiently large constant $C > 0$ (we will actually show an exponential bound). First note that the variables $\xi'_{ij}\,\mathbb{1}_{\mathcal{E}_{ij}}$ are centered and satisfy $\mathbb{E}\,|\xi'_{ij}\,\mathbb{1}_{\mathcal{E}_{ij}}|^4 = O(1)$. It follows from Theorem 4.6.7 that

$$\begin{aligned}\mathbb{E}\,\mathbb{1}_{\mathcal{E}}\,\|A \circ X'\| &\ll \max_{i\in[n]}\left(\sum_{j=1}^n a_{ij}^2\right)^{1/2} + \max_{j\in[n]}\left(\sum_{i=1}^n a_{ij}^2\right)^{1/2} + \left(\sum_{i,j=1}^n a_{ij}^4\right)^{1/4}\\ &\ll \tau\sqrt{n}.\end{aligned}$$

Thus, (4.153) will follow if we can show

$$\mathbb{P}\big(\|A\circ X'\|\,\mathbb{1}_{\mathcal{E}} - \mathbb{E}\,\|A\circ X'\|\,\mathbb{1}_{\mathcal{E}} \ge \tau\sqrt{n}\big) = O_\tau(n^{-\eta/8}). \tag{4.154}$$

This in turn follows in a routine manner from Talagrand's inequality [Tal96, Theorem 6.6] (see also [AGZ10, Corollary 4.4.11]): Observe that $X \mapsto \|A\circ X\|$ is a convex and 1-Lipschitz function on the space $\mathcal{M}_n(\mathbb{R})$ equipped with the (Euclidean) Hilbert–Schmidt metric. Since the matrix $X'\,\mathbb{1}_{\mathcal{E}}$ has centered entries that are bounded by $O(n^{1/2-\eta_0})$, Talagrand's inequality gives that the left hand side of (4.154) is bounded by

$$O\big(\exp(-c\tau^2 n/(n^{1/2-\eta_0})^2)\big) = O\big(\exp(-c\tau^2 n^{2\eta_0})\big) \tag{4.155}$$

which gives (4.154) with plenty of room.

Now we turn to part (b). The proof follows a very similar truncation argument to the one in part (a), so we only indicate the necessary modifications. As before, by splitting D and X into real and imaginary parts and applying the triangle inequality we may assume D and X are real matrices. We define $\mathcal{E}$ as in (4.147), with

$$\mathcal{E}_{ij} = \big\{|\xi_{ij}| \le (n\sqrt{m})^{1/3-\eta_1}\big\} \tag{4.156}$$

and

$$\eta_1 = \frac{1}{4}\frac{\eta}{4+\eta}. \tag{4.157}$$

With this choice of η_1, Markov's inequality and the union bound give $\mathbb{P}(\mathcal{E}^c) = O_\theta(n^{-\eta/8})$. Taking $X' = X - \mathbb{E}(X\,\mathbb{1}_{\mathcal{E}})$ as before, we can bound

$$\|D(A\circ\mathbb{E}(X\,\mathbb{1}_{\mathcal{E}}))\| \le \|A\circ\mathbb{E}(X\,\mathbb{1}_{\mathcal{E}})\|$$

by submultiplicativity of the operator norm, and the same argument as before gives

$$\|A\circ\mathbb{E}(X\,\mathbb{1}_{\mathcal{E}})\| \le nm(n\sqrt{m})^{-\frac{3}{4}(4+\eta)(1/3-\eta_1)} = m^{1/2-\eta/32} = o(\sqrt{m}). \tag{4.158}$$

Since $X' \mathbb{1}_{\mathcal{E}}$ has centered entries with finite moments of order $4+\eta$, by Theorem 4.6.8 we have

$$\mathbb{E}\, \|D(A \circ X' \mathbb{1}_{\mathcal{E}})\| \ll_\eta \sqrt{m}. \tag{4.159}$$

The mapping $X \mapsto \|D(A \circ X)\|$ is convex and 1-Lipschitz with respect to the Hilbert–Schmidt metric on $\mathcal{M}_n(\mathbb{R})$ (since $\|D\| \le 1$) so using Talagrand's inequality as in part (a) we find that

$$\begin{aligned}\mathbb{P}\big(\|D(A \circ X' \mathbb{1}_{\mathcal{E}})\| - \mathbb{E}\, \|D(A \circ X' \mathbb{1}_{\mathcal{E}})\| \ge \sqrt{m}\big) &\ll \exp\left(-cm/(n\sqrt{m})^{2/3-2\eta_1}\right)\\ &\le \exp\left(-c'(\theta) n^{c\eta}\right)\end{aligned}$$

for some constant $c > 0$ and $c'(\theta) > 0$ sufficiently small depending on θ. As the last line is bounded by $O_\theta(n^{-\eta/8})$, the result follows from the above, (4.158), (4.159) and the triangle inequality by the same argument as for part (a). $\square$

4.6.3 Decomposition of the standard deviation profile

We now begin the proof of Theorem 4.1.17, which occupies the remainder of the chapter. In the present subsection we prove Lemma 4.6.9 below, which shows that the standard deviation profile A can be partitioned into a bounded collection of submatrices with certain nice properties. For the motivation behind this lemma (and the notation $J_{\mathrm{free}}, J_{\mathrm{cyc}}$) see Section 4.1.3.

Lemma 4.6.9. *Let A be an $n \times n$ matrix with entries $a_{ij} \in [0,1]$. Let $\varepsilon, \delta, \sigma_0 \in (0,1)$, and assume ε is sufficiently small depending on δ. There exists $0 \le m = O_\varepsilon(1)$, a partition*

$$\begin{aligned}[n] &= J_{\mathrm{bad}} \cup J_{\mathrm{free}} \cup J_{\mathrm{cyc}}\\ &= J_{\mathrm{bad}} \cup J_{\mathrm{free}} \cup J_1 \cup \cdots \cup J_m\end{aligned}$$

and a set $F \subset [n]^2$ satisfying the following properties:

1. *$\varepsilon n \ll |J_{\mathrm{bad}}| \ll \delta^{1/2} n$.*

2. $|F| \ll \delta n^2$*, and for all* $i \in J_{\mathrm{free}}$*,*

$$|\{j \in J_{\mathrm{free}} : (i,j) \in F\}|,\ |\{j \in J_{\mathrm{free}} : (j,i) \in F\}| \le \delta^{1/2} n. \tag{4.160}$$

3. If $J_{\mathrm{free}} \neq \varnothing$ *then there is a permutation* $\tau : J_{\mathrm{free}} \to J_{\mathrm{free}}$ *such that for all* $(i,j) \in J_{\mathrm{free}} \times J_{\mathrm{free}} \setminus F$ *with* $\tau(i) \ge \tau(j)$*,* $a_{ij} < \sigma_0$*.*

4. If $m \ge 1$ *then*

$$|J_1| = \cdots = |J_m| \gg_\varepsilon n \tag{4.161}$$

and there is a permutation $\pi : [m] \to [m]$ *such that for all* $1 \le k \le m$*,* $A(\sigma_0)_{J_k, J_{\pi(k)}}$ *is* $(2\delta, 2\varepsilon)$*-super-regular.*

Proof. We begin by applying Lemma 4.6.2 to $A(\sigma_0)$ to obtain $m_0 \in \mathbb{N}$ with $\varepsilon^{-1} \le m_0 = O_\varepsilon(1)$ and a partition $[n] = I_0 \cup \cdots \cup I_{m_0}$ satisfying the properties in that lemma.

The partition $I_0, \ldots, I_{m_0}$ is almost what we need. In the remainder of the proof we perform a "cleaning" procedure (as it is commonly referred to in the extremal combinatorics literature) to obtain a partition $J_0, \ldots, J_{m_0}$ with improved properties, where $J_k \subset I_k$ for each $1 \le k \le m_0$, and $J_0 \supset I_0$ collects the leftover elements.

We start by forming a *reduced digraph* $\mathcal{R} = ([m_0], E)$ on the vertex set $[m_0]$ with directed edge set

$$E := \Big\{(k,l) \in [m_0]^2 : (I_k, I_l) \text{ is } \varepsilon\text{-regular and } \rho_{A(\sigma_0)}(I_k, I_l) > 5\delta\Big\}. \tag{4.162}$$

Next we find a (possibly empty) set $T \subset [m_0]$ such that the induced subgraph $\mathcal{R}(T)$ is covered by vertex-disjoint directed cycles, and the induced subgraph $\mathcal{R}([m_0] \setminus t)$ is cycle-free. Such a set can be obtained by greedily removing cycles and the associated vertices from $\mathcal{R}$ until the remaining graph has no more directed cycles. By relabeling $I_1, \ldots, I_{m_0}$ we may take $T = [m]$, where $m \in [0, m_0]$.

Assuming $m \neq 0$, the fact that $\mathcal{R}([m])$ is cycle-free is equivalent to the existence of a permutation $\pi : [m] \to [m]$ such that $(k, \pi(k)) \in E$ for all $1 \le k \le m$. Now we will obtain the sets $J_1, \ldots, J_m$ obeying the properties in part (4) of the lemma. Let $1 \leq k \leq m$. Since $(I_k, I_{\pi(k)})$ is ε-regular with density $\rho_k := \rho_{A(\sigma_0)}(I_k, I_{\pi(k)}) > 5\delta$, if we assume $\varepsilon \leq \delta$ then for every $I \subset I_k, J \subset I_{\pi(k)}$ with $|I|, |J| \geq \varepsilon |I_k|$,

$$e_{A(\sigma_0)}(I, J) \geq (\rho_k - \varepsilon)|I||J| \geq 4\delta |I||J|. \tag{4.163}$$

Letting

$$I'_k = \left\{ i \in I_k : |\mathcal{N}_{A(\sigma_0)}(i) \cap I_{\pi(k)}| < 4\delta |I_k| \right\}$$

we have $e_{A(\sigma_0)}(I'_k, I_{\pi(k)}) < 4\delta |I'_k||I_{\pi(k)}|$, and it follows that $|I'_k| \leq \varepsilon |I_k|$. Similarly, letting

$$I''_k = \left\{ i \in I_k : |\mathcal{N}_{A(\sigma_0)^\mathsf{T}}(i) \cap I_{\pi^{-1}(k)}| < 4\delta |I_k| \right\}$$

we have $|I''_k| \leq \varepsilon |I_k|$. Letting $I^*_k \subset I_k$ be a set of size $\lfloor 2\varepsilon |I_k| \rfloor$ containing $I'_k \cup I''_k$, we take

$$J_k = I_k \setminus I^*_k. \tag{4.164}$$

With this definition we have $|J_1| = \cdots |J_m|$, and for each $1 \leq k \leq m, i \in J_k$,

$$|\mathcal{N}_{A(\sigma_0)}(i) \cup J_{\pi(k)}|, \, |\mathcal{N}_{A(\sigma_0)^\mathsf{T}}(i) \cap J_{\pi^{-1}(k)}| \geq (4\delta - 2\varepsilon)|I_k| \geq 2\delta |J_k|. \tag{4.165}$$

Furthermore, for each $1 \leq k \leq m$ and $I \subset J_k, J \subset J_{\pi(k)}$ with $|I|, |J| \geq 2\varepsilon |J_k|$, if we assume $\varepsilon \leq 1/4$ then $|I|, |J| \geq \varepsilon |I_k|$, so by (4.163)

$$e_{A(\sigma_0)}(I, J) \geq 4\delta |I||J|. \tag{4.166}$$

It follows that for every $1 \leq k \leq m$ the sub-matrix $A(\sigma_0)_{J_k, J_{\pi(k)}}$ is $(2\delta, 2\varepsilon)$-super-regular, which concludes the proof of part (4) of the lemma.

Now we prove parts (2) and (3). We will obtain J_{free} by removing a small number of bad elements from $I_{\text{free}} := \bigcup_{k=m+1}^{m_0} I_k$. Since the induced subgraph $\mathcal{R}([m+1, m_0])$ is cycle-free we may relabel $I_{m+1}, \ldots, I_{m_0}$ so that

$$(k, l) \notin E \ \text{ for all } m < l \leq k \leq m_0. \tag{4.167}$$

We take

$$F = \big\{(i,j) \in [n]^2 : (i,j) \in I_k \times I_l \text{ for some } (k,l) \notin E\big\}. \tag{4.168}$$

The contribution to F from irregular pairs (I_k, I_l) is at most εn^2 by the regularity of the partition $I_0, \ldots, I_{m_0}$, and the contribution from pairs (I_k, I_l) with density less than 5δ is at most $5\delta n^2$. Hence,

$$|F| \leq \varepsilon n^2 + 5\delta n^2 \leq 6\delta n^2 \tag{4.169}$$

giving the first estimate in (2) (recall that we assumed $\varepsilon \leq \delta$). Setting

$$I'_{\text{free}} = \Big\{i \in I_{\text{free}} : \max\big(|\{j \in [n] : (i,j) \in F\}|, |\{j \in [n] : (j,i) \in F\}|\big) \geq \delta^{1/2} n\Big\} \tag{4.170}$$

it follows from (4.169) that

$$|I'_{\text{free}}| \leq 12\delta^{1/2} n. \tag{4.171}$$

Let $I^*_{\text{free}} \subset I_{\text{free}}$ be any set containing I'_{free} of size $\min(|I_{\text{free}}|, \lfloor 12\delta^{1/2} n\rfloor)$ and take $J_{\text{free}} = I_{\text{free}} \setminus I^*_{\text{free}}$. The bounds (4.160) now follow immediately from (4.170). For part (3), from (4.167) we may take for τ any ordering of the elements of J_{free} that respects the order of the sets $J_k := I_k \setminus I^*_{\text{free}}$, i.e. so that $\tau(j) \geq \tau(i)$ for all $i \in J_k, j \in J_l$ and all $m < l \leq k \leq m_0$.

Finally, taking

$$J_{\text{bad}} = I_0 \cup I^*_{\text{free}} \cup \bigcup_{k=1}^{m} I^*_k. \tag{4.172}$$

we have

$$|J_{\text{bad}}| \leq \varepsilon n + 12\delta^{1/2} n + 2\varepsilon n \leq 15\delta^{1/2} n$$

giving the upper bound in part (1). Now recalling that we took $|I^*_{\text{free}}| = 12\delta^{1/2}$ and $|I^*_k| =$

$\lfloor 2\varepsilon |I_k| \rfloor$ for all $1 \le k \le m$, we also have the lower bound

$$\begin{aligned} |J_{\mathrm{bad}}| &\ge \min\left(|I^*_{\mathrm{free}}|, \, \left| \bigcup_{k=1}^m I^*_k \right| \right) \\ &\ge \min\left(\lfloor 12\delta^{1/2} n \rfloor, \, |I_{\mathrm{free}}|, \, 2\varepsilon \left| \bigcup_{k=1}^m I_k \right| - m \right) \\ &= \min\left(\lfloor 12\delta^{1/2} n \rfloor, \, \left| \bigcup_{k=m+1}^{m_0} I_k \right|, \, 2\varepsilon \left| \bigcup_{k=1}^m I_k \right| - m \right) \\ &\gg \varepsilon n \end{aligned}$$

where we used that at least one of the sets $I_{\mathrm{free}} = \bigcup_{k=m+1}^{m_0} I_k$, $I_{\mathrm{cyc}} = \bigcup_{k=1}^{m} I_k$ must be of size at least $n/4$, say. This gives the lower bound in part (1) and completes the proof. $\square$

4.6.4 High level proof of Theorem 4.1.17

In this subsection we prove Theorem 4.1.17 on two lemmas (Lemmas 4.6.10 and 4.6.11) which give control on the smallest singular values of the submatrices $M_{J_{\mathrm{free}}}$ and (perturbations of) $M_{J_{\mathrm{cyc}}}$, with $J_{\mathrm{free}}, J_{\mathrm{cyc}}$ as in Lemma 4.6.9. The proofs of these lemmas are deferred to the remaining subsections.

As we noted in Remark 4.1.18, the dependence of the implied constant in (4.24) on the distribution of ξ is through its $(4+\eta)$-th moment $\mu_{4+\eta}$ and the value of κ provided by Lemma 4.2.6. For the sake of clarity, in the sequel we indicate dependence of constants on these two parameters. (By Lemma 4.2.6 and multiplying X and B by a phase we may assume ξ has κ-controlled second moment for some $\kappa \ge 1$.)

Without loss of generality we may assume $\eta < 1$. We introduce parameters $\sigma_0, \delta, \varepsilon \in (0,1)$, where σ_0 will be chosen sufficiently small depending on $r_0, \mu_{4+\eta}$, δ will be taken small depending on $r_0, \eta, \mu_{4+\eta}$, and ε will be taken sufficiently small depending on σ_0, δ. Below we introduce an additional free parameter $\theta \in (0,1)$ to be chosen later sufficiently small depending on ε. For the remainder of the proof we assume that n is sufficiently large depending on all parameters (which will only depend on $r_0, K_0, \eta, \mu_{4+\eta}$ and κ).

We begin by summarizing the control we have on the operator norm of submatrices of $A \circ X$ due to Lemma 4.6.5. By Lemma 4.6.5(a) with $\tau = 1$ we may restrict the sample space to the boundedness event

$$\mathcal{B}(K) = \left\{ \|A \circ X\| \le K\sqrt{n} \right\} \tag{4.173}$$

for some $K = O(\mu_{4+\eta})$. In particular, for all $I, J \subset [n]$ we now have the crude bound

$$\|(A \circ X)_{I,J}\| \le K\sqrt{n}. \tag{4.174}$$

We also note that by Lemma 4.6.5(a) we have that for any fixed $B = (b_{ij}) \in \mathcal{M}_n([0,1])$,

$$\mathbb{P}\Big(\|(B \circ X)_{I,J}\| \le \tau K \sqrt{|I|} \Big) = 1 - O_\tau(n^{-\eta/8}) \tag{4.175}$$

for all $I, J \subset [n]$ with $|I| = |J|$, and any $\tau \le 1$ satisfying

$$\tau \ge \max\left(\max_{i\in I} \left(\frac{1}{n} \sum_{j \in J} b_{ij}^2 \right)^{1/2}, \max_{j \in J} \left(\frac{1}{n} \sum_{i\in I} b_{ij}^2 \right)^{1/2}, \left(\frac{1}{n} \sum_{i,j=1}^n b_{ij}^4 \right)^{1/4} \right). \tag{4.176}$$

Note that for $B = A$, (4.175) gives improved control on $\|(A \circ X)_{I,J}\|$ for submatrices that are either small or sparse, but only holding with high probability for fixed I, J. (We state (4.175) for general $B \in \mathcal{M}_n([0,1])$ as at one point we will apply this to a residual matrix obtained by subtracting off a collection of "bad" entries from A.)

Finally, let $\theta \in (0,1)$ to be chosen later depending on ε. From Lemma 4.6.5(b) we have that for every $I, J \subset [n]$ with $|I| \ge |J| \ge \theta n$ and every deterministic $|J| \times |I|$ matrix D,

$$\mathbb{P}\Big(\|D(A \circ X)_{I,J}\| \le K' \|D\| \sqrt{|J|} \Big) = 1 - O_\theta(|I|^{-\eta/8}) \tag{4.177}$$

for some K' sufficiently large depending on η and $\mu_{4+\eta}$.

We now apply Lemma 4.6.9 (assuming ε is sufficiently small depending on δ) to obtain a partition $[n] = J_{\text{bad}} \cup J_{\text{free}} \cup J_{\text{cyc}}$ and a set $F \subset [n]^2$ satisfying the properties (1)–(4) in the lemma. In the following we abbreviate $M_{\text{free}} := M_{J_{\text{free}}}$ and $M_{\text{cyc}} := M_{J_{\text{cyc}}}$.

Lemma 4.6.10. *Assume $n_1 := |J_{\mathrm{free}}| \geq \delta^{1/2} n$. If σ_0, δ are sufficiently small depending on r_0 and K (from* (4.173)*), then*

$$s_{n_1}(M_{\mathrm{free}}) \gg_{K,r_0} \sqrt{n} \tag{4.178}$$

except with probability $O_{K,r_0,\delta}(n^{-\eta/9})$.

Lemma 4.6.11. *Assume $n_2 := |J_{\mathrm{cyc}}| \geq \delta^{1/2} n$. Fix $\gamma > 1/2$ and let $W \in \mathcal{M}_{n_2}(\mathbb{C})$ be a deterministic matrix with $\|W\| \leq n^{\gamma}$. There exists $\beta = \beta(\gamma, \sigma_0, \delta)$ such that if $\varepsilon = \varepsilon(\sigma_0, \delta)$ is sufficiently small,*

$$\mathbb{P}\big(s_{n_2}(M_{\mathrm{cyc}} + W) \leq n^{-\beta}\big) \ll_{K_0,K,\gamma,\delta,\sigma_0,\kappa} \sqrt{\frac{\log n}{n}}. \tag{4.179}$$

We defer the proofs of Lemmas 4.6.10 and 4.6.11 to subsequent sections, and conclude the proof of Theorem 4.1.17. We proceed in the following steps:

Step 1: Bound the smallest singular value of M_{free} using Lemma 4.6.10.

Step 2: Bound the smallest singular value of

$$M_1 := M_{J_{\mathrm{free}} \cup J_{\mathrm{bad}},\, J_{\mathrm{free}} \cup J_{\mathrm{bad}}} = \begin{pmatrix} M_{\mathrm{free}} & B_1 \\ C_1 & M_0 \end{pmatrix}. \tag{4.180}$$

using the result of Step 1, the Schur complement bound of Lemma 4.6.4, and the bounds (4.175), (4.177).

Step 3: Bound the smallest singular value of

$$M = \begin{pmatrix} M_{\mathrm{cyc}} & B_2 \\ C_2 & M_1 \end{pmatrix}. \tag{4.181}$$

using the result of Step 2, the Schur complement bound of Lemma 4.6.4, and

Lemma 4.6.11.

The case that one of J_{free} or J_{cyc} is small (or empty) can be handled essentially by skipping either Step 1 or Step 3. We will begin by assuming

$$|J_{\text{free}}|,\ |J_{\text{cyc}}| \geq \delta^{1/2} n \tag{4.182}$$

and address the case that this does not hold at the end.

Step 1

By Lemma 4.6.10 and the assumption (4.182), we can take σ_0 and δ sufficiently small depending on r_0 and K (where K depends only $\mu_{4+\eta}$) such that

$$s_{\min}(M_{\text{free}}) \gg_{K,r_0} \sqrt{n} \tag{4.183}$$

except with probability $O_{K,r_0,\delta}(n^{-\eta/9})$. We now fix $\sigma_0 = \sigma_0(r_0, \mu_{4+\eta})$ once and for all, but leave δ free to be taken smaller if necessary. By independence of the entries of M we may now condition on a realization of M_{free} such that (4.183) holds.

Step 2

From Lemma 4.6.4, (4.174) and (4.183),

$$\begin{aligned} s_{\min}(M_1) &\gg \left(1 + \frac{K\sqrt{n}}{s_{\min}(M_{\text{free}})}\right)^{-2} \min\left[s_{\min}(M_{\text{free}}), s_{\min}(M_0 - C_1 M_{\text{free}}^{-1} B_1)\right] \\ &\gg_{K,r_0} \min\left[\sqrt{n}, s_{\min}(M_0 - C_1 M_{\text{free}}^{-1} B_1)\right]. \end{aligned} \tag{4.184}$$

Again by (4.174) and (4.183),

$$\|C_1 M_{\text{free}}^{-1}\| \leq \frac{\|C_1\|}{s_{\min}(M_{\text{free}})} \ll_{K,r_0} 1. \tag{4.185}$$

Since C_1 is independent of B_1, we may condition on it (recall that we already conditioned on M_{free}) and apply (4.177) to conclude

$$\|C_1 M_{\text{free}}^{-1} B_1\| \leq K' \|C_1 M_{\text{free}}^{-1}\| |J_{\text{bad}}|^{1/2} \ll_{K,K',r_0} |J_{\text{bad}}|^{1/2} \tag{4.186}$$

except with probability $O_\theta(n_1^{-\eta/8}) = O_{\delta,\varepsilon}(n^{-\eta/9})$, where we have used the lower bound $|J_{\text{bad}}| \gg \varepsilon n$ from Lemma 4.6.9(1) and fixed θ sufficiently small depending on ε. On the other hand, by from the triangle inequality and the bound (4.175) with $\tau = 1$,

$$s_{\min}(M_0) = s_{\min}(Z_{J_{\text{bad}}}\sqrt{n} + (A \circ X)_{J_{\text{bad}}}) \geq r_0\sqrt{n} - K|J_{\text{bad}}|^{1/2} \tag{4.187}$$

except with probability $O(|J_{\text{bad}}|^{-\eta/8}) = O_\varepsilon(n^{-\eta/9})$. Combining the previous two displays,

$$s_{\min}(M_0 - C_1 M_{\text{free}}^{-1} B_1) \geq r_0\sqrt{n} - O_{K,K',r_0}(|J_{\text{bad}}|^{1/2}) \tag{4.188}$$

except with probability $O_{\delta,\varepsilon}(n^{-\eta/9})$. Since $|J_{\text{bad}}| \ll \delta^{1/2} n$ we can take δ smaller, if necessary, depending on r_0, K, K' to conclude that

$$s_{\min}(M_0 - C_1 M_{\text{free}}^{-1} B_1) \geq (r_0/2)\sqrt{n} \tag{4.189}$$

except with probability $O_{\delta,\varepsilon}(n^{-\eta/9})$. We may henceforth condition on the event that (4.189) holds. Together with (4.184) this implies

$$s_{\min}(M_1) \gg_{K,r_0} \sqrt{n}. \tag{4.190}$$

At this point we fix $\delta = \delta(r_0, \eta, \mu_{4+\eta})$.

Step 3

Condition on a realization of M_1 such that (4.190) holds, and also on the matrices B_2, C_2 in (4.181). By (4.174) we have $\|B_2\|, \|C_2\| \leq K\sqrt{n}$. Applying Lemma 4.6.4,

$$\begin{aligned} s_n(M) &\gg \left(1 + \frac{K\sqrt{n}}{s_{\min}(M_1)}\right)^{-2} \min\left[s_{\min}(M_1),\, s_{\min}(M_{\mathrm{cyc}} - B_2 M_1^{-1} C_2)\right] \\ &\gg_{K,r_0} \min\left[\sqrt{n}, s_{\min}(M_{\mathrm{cyc}} - B_2 M_1^{-1} C_2)\right]. \end{aligned} \tag{4.191}$$

Again by (4.174) and (4.190),

$$\|B_2 M_1^{-1} C_2\| \leq \frac{K^2 n}{s_{\min}(M_1)} \ll_{K,r_0} \sqrt{n} \tag{4.192}$$

(unlike in Step 2, here we did not need the stronger control on matrix products provided by (4.144) from the matrix boundedness property). Now since M_2 is independent of M_1, B_2, C_2, we can apply Lemma 4.6.11 with $\gamma = 0.51$ (say), fixing ε sufficiently small depending on $\sigma_0(r_0, \mu_{4+\eta})$ and $\delta(r_0, \eta, \mu_{4+\eta})$, to obtain

$$\mathbb{P}\big(s_{\min}(M_{\mathrm{cyc}} - B_2 M_1^{-1} C_2) \leq n^{-\beta}\big) \ll_{K_0, r_0, \eta, \mu_{4+\eta}, \kappa} \sqrt{\frac{\log n}{n}} \tag{4.193}$$

for some $\beta = \beta(r_0, \eta, \mu_{4+\eta}) > 0$, where we have substituted $K = O(\mu_{4+\eta})$. The result now follows from the above and (4.191), taking $\alpha = \min(\eta/9, 1/4)$, say.

It only remains to address the case that the assumption (4.182) fails. We may assume that δ is small enough that only one of these bounds fails. In this case we simply redefine J_{bad} to include the smaller of $J_{\mathrm{cyc}}, J_{\mathrm{free}}$. Note that we still have $|J_{\mathrm{bad}}| = O(\delta^{1/2} n)$. If $|J_{\mathrm{cyc}}| < \delta^{1/2} n$, then with this new definition of J_{bad} we have $M = M_1$, and the desired bound on $s_n(M)$ follows from (4.190) (with plenty of room). If $|J_{\mathrm{free}}| < \delta^{1/2} n$ then we skip Step 2, proceeding with Step 3 using M_0 in place of M_1. The bound (4.190) in this case follows from (4.187) and the bound $|J_{\mathrm{bad}}| \ll \delta^{1/2} n$, taking δ sufficiently small depending on K, r_0. This concludes the proof of Theorem 4.1.17.

4.6.5 Proof of Lemma 4.6.10

We denote

$$A_{\text{bad}} = (a_{ij} 1_{(i,j)\in F}). \tag{4.194}$$

By the estimates on F in Lemma 4.6.9 we can apply (4.175) with $\tau = O(\delta^{1/4})$ to obtain

$$\|(A_{\text{bad}}(\sigma_0) \circ X)_{J_{\text{free}}}\| \ll K\delta^{1/4}\sqrt{n} \tag{4.195}$$

except with probability at most $O_\delta(n_1^{-\eta/8}) = O_\delta(n^{-\eta/9})$. By another application of (4.175) with $\tau = 1$,

$$\left\|\big((A - A(\sigma_0)) \circ X\big)_{J_{\text{free}}}\right\| \ll K\sigma_0\sqrt{n} \tag{4.196}$$

except with probability at most $O_\delta(n^{-\eta/9})$. Let

$$\widetilde{M}_{\text{free}} := (\widetilde{A} \circ X)_{J_{\text{free}}} + Z_{J_{\text{free}}}\sqrt{n}, \qquad \widetilde{A} := A(\sigma_0) - A_{\text{bad}}(\sigma_0). \tag{4.197}$$

By the above estimates and the triangle inequality,

$$s_{\min}(M_{\text{free}}) \geq s_{\min}(\widetilde{M}_{\text{free}}) - \|((A - \widetilde{A}) \circ X)_{J_{\text{free}}}\| \tag{4.198}$$

$$\geq s_{\min}(\widetilde{M}_{\text{free}}) - O(\delta^{1/4} + \sigma_0)K\sqrt{n} \tag{4.199}$$

except with probability $O_\delta(n^{-\eta/9})$. Thus, it suffices to show

$$s_{\min}(\widetilde{M}_{\text{free}}) \gg_{K,r_0} \sqrt{n}. \tag{4.200}$$

except with probability $O_{K,r_0,\delta}(n^{-\eta/9})$ – the result will then follow from (4.200) and (4.199) by taking δ, σ_0 sufficiently small depending on K, r_0. Furthermore, by Lemma 4.6.9(3) and conjugating M_{free} by a permutation matrix we may assume that $\widetilde{A}$ is (strictly) upper triangular. Now it suffices to prove the following:

Lemma 4.6.12. *Let $M = A \circ X + B$ be an $n \times n$ matrix as in Definition 4.1.1, and further assume that for some $r_0 > 0, K \geq 1, \alpha > 0$,*

- *A is upper triangular;*
- *$B = Z\sqrt{n} = \mathrm{diag}(z_i\sqrt{n})_{i=1}^n$ with $|z_i| \geq r_0$ for all $1 \leq i \leq n$;*
- *ξ is such that for all $n' \geq 1$ and any fixed $A' \in \mathcal{M}_{n'}([0,1])$, $\|A' \circ X'\| \leq K\sqrt{n'}$ except with probability $O((n')^{-\alpha})$.*

Then $s_n(M) \gg_{K,r_0} \sqrt{n}$ except with probability $O_{K,r_0}(1)^\alpha n^{-\alpha}$.

Remark 4.6.13. The proof gives an implied constant of order $\exp(-O(K/r_0)^{O(1)})$ in the lower bound on $s_n(M)$.

To deduce Lemma 4.6.10 we apply the above lemma with $M = \widetilde{M}_{\mathrm{free}}$, $\alpha = \eta/8$, K as in (4.173) and $n_1 \gg_\delta n$ in place of n, which gives that (4.200) holds with probability

$$1 - O_{K,r_0}(n_1^{-\eta/8}) = 1 - O_{K,r_0,\delta}(n^{-n/9}) \tag{4.201}$$

where in the first bound we applied our assumption that $\eta < 1$.

Proof. First we note that we may take n to be a dyadic integer, i.e. $n = 2^q$ for some $q \in \mathbb{N}$. Indeed, if this is not the case, then letting 2^q be the smallest dyadic integer larger than n we can increase the dimension of M to 2^q by padding A out with rows and columns of zeros, adding additional rows and columns of iid copies of ξ to X, and extending the diagonal of Z with entries $z_i \equiv r_0$ for $n < i \leq 2^q$. The hypotheses on A and Z in the lemma are still satisfied, and the smallest singular value of the new matrix is a lower bound for that of the original matrix (since the original matrix is a submatrix of the new matrix).

Now fix an arbitrary dyadic filtration $\mathcal{F} = \big\{\{J_s\})_{s\in\{0,1\}^p}\big\}_{p\geq 0}$ of $[n]$, where we view $\{0,1\}^0$ as labeling the trivial partition of $[n]$, consisting only of the empty string $\varnothing$, so that $J_\varnothing = [n]$. Thus, for every $0 \leq p < q$ and every binary string $s \in \{0,1\}^p$, J_s has cardinality $n2^{-p}$ and is evenly partitioned by J_{s0}, J_{s1}. For a binary string s we abbreviate $M_s := M_{J_s}$ and similarly define A_s, X_s, Z_s. We also write $B_s = M_{J_{s0},J_{s1}}$, so that we have the block decomposition

$$M_s = \begin{pmatrix} M_{s0} & B_s \\ 0 & M_{s1} \end{pmatrix}. \tag{4.202}$$

For $p \geq 1$ define the boundedness event

$$\mathcal{B}^*(p) = \left\{ \|A \circ X\| \leq K\sqrt{n} \right\} \wedge \left\{ \forall s \in \{0,1\}^p,\ \|A_s \circ X_s\| \leq K\sqrt{n2^{-p}} \right\}. \tag{4.203}$$

By our assumption on ξ we have

$$\mathbb{P}(\mathcal{B}^*(p)) \geq 1 - O(n^{-\alpha}) - 2^p O((n2^{-p})^{-\alpha}) = 1 - O(2^{(1+\alpha)p} n^{-\alpha}). \tag{4.204}$$

For arbitrary $s \in \{0,1\}^p$, by the triangle inequality we have that on $\mathcal{B}^*(p)$,

$$s_{\min}(M_s) \geq s_{\min}(Z_s) - \|A_s \circ X_s\| \geq (r_0 - K2^{-p/2})\sqrt{n}.$$

Setting $p_0 = \lfloor 2\log(2K/r_0) \rfloor + 1$ we have that on $\mathcal{B}^*(p_0)$,

$$s_{\min}(M_s) \geq (r_0/2)\sqrt{n} \tag{4.205}$$

for all $s \in \{0,1\}^{p_0}$. For the remainder of the proof we restrict the sample space to the event $\mathcal{B}^*(p_0)$ and will use the Schur complement bound (Lemma 4.6.4) to show that the desired lower bound on $s_{\min}(M)$ holds deterministically (note that by (4.204) and our choice of p_0, $\mathcal{B}^*(p_0)$ holds with probability $1 - O_{K,r_0}(n^{-\alpha/2})$).

For $0 \leq p \leq p_0$ let

$$\lambda_p = \min_{s \in \{0,1\}^p} \frac{1}{\sqrt{n}} s_{\min}(M_s). \tag{4.206}$$

From (4.205) we have

$$\lambda_{p_0} \geq r_0/2 \tag{4.207}$$

Now let $1 \leq p \leq p_0$ and $s \in \{0,1\}^{p-1}$. By the block decomposition (4.202) and Lemma 4.6.4,

$$\begin{aligned} s_{\min}(M_s) &\gg \left(1 + \frac{\|B_s\|}{s_{\min}(M_{s0})}\right)^{-1} \min\big(s_{\min}(M_{s0}), s_{\min}(M_{s1})\big) \\ &\geq (1 + K/\lambda_p)^{-1}\lambda_p\sqrt{n} \end{aligned}$$

so $\lambda_{p-1} \gg (1+K/\lambda_p)^{-1}\lambda_p\sqrt{n}$ for all $0 \leq p \leq p_0$. Applying this iteratively along with (4.207) we conclude $\lambda_0 \gg_{K,r_0} 1$, i.e.

$$s_{\min}(M) \gg_{K,r_0} \sqrt{n} \tag{4.208}$$

as desired. □

4.6.6 Proof of Lemma 4.6.11

By multiplying M_{cyc} by a permutation matrix we may assume that $A_k := A_{J_k}$ is $(2\delta, 2\varepsilon)$-super-regular for $1 \leq k \leq m$ (unlike in the proof of Lemma 4.6.10 the diagonal matrix $Z\sqrt{n}$ will play no special role here). We denote $J_{\leq k} = J_1 \cup \cdots \cup J_k$, and for any matrix W of dimension at least $|J_{\leq k}|$ we abbreviate

$$W_k = W_{J_k}, \quad W_{\leq k} = W_{J_{\leq k}}, \quad W_{\leq k-1,k} = W_{J_{\leq k-1},J_k}, \quad W_{k,\leq k-1} = W_{J_k,J_{\leq k-1}} \tag{4.209}$$

so that for $2 \leq k \leq m$ we have the block decomposition

$$W_{\leq k} = \begin{pmatrix} W_{\leq k-1} & W_{\leq k-1,k} \\ W_{k,\leq k-1} & W_k \end{pmatrix}. \tag{4.210}$$

Let us denote

$$n' = |J_1| = \cdots = |J_m| \gg_\varepsilon n. \tag{4.211}$$

For $1 \leq k \leq m-1$, $\beta > 0$ and a fixed $kn' \times kn'$ matrix W, we denote the event

$$\mathcal{E}_k(\beta, W) := \big\{s_{kn'}(M_{\leq k} + W) > n^{-\beta}\big\}. \tag{4.212}$$

Let $\gamma > 1/2$ and fix an arbitrary matrix $W \in \mathcal{M}_{n',n'}(\mathbb{C})$ with $\|W\| \le n^\gamma$. By our restriction to $\mathcal{B}(K)$ and (4.174) we have

$$\|M_1 + W\| \le (K_0 + K)\sqrt{n} + n^\gamma \le 2n^\gamma \tag{4.213}$$

if n is sufficiently large depending on K_0, K and γ. By Theorem 4.1.24 there exists $\beta_1(\gamma) = O(\gamma^2)$ such that if ε is sufficiently small depending on σ_0, δ, then

$$\mathbb{P}\left(\mathcal{E}_1(\beta_1, W)^c\right) \ll_{K_0,K,\gamma,\delta,\sigma_0,\varepsilon,\kappa} \sqrt{\frac{\log n}{n}} \tag{4.214}$$

where we have used (4.211) to write n in $n^{-\beta_1}$ rather than n'. (Strictly speaking we are using (4.105) instead of Theorem 4.1.24 to say that the implied constant depends on the distribution of ξ only through κ.)

Now let $2 \le k \le m$, and suppose we have found a function $\beta_{k-1}(\gamma)$ such that for any $\gamma > 1/2$ and any fixed $(k-1)n' \times (k-1)n'$ matrix W with $\|W\| \le n^\gamma$,

$$\mathbb{P}(\mathcal{E}_{k-1}(\beta_{k-1}(\gamma), W)^c) \ll_{K_0,K,\gamma,\delta,\sigma_0,\varepsilon,\kappa} \sqrt{\frac{\log n}{n}}. \tag{4.215}$$

Fix a $kn' \times kn'$ matrix W with $\|W\| \le n^\gamma$. By Lemma 4.6.4 we have

$$\begin{aligned} s_{kn'}(M_{\le k} + W) \gg &\left(1 + \frac{\|(M+W)_{\le k-1,k}\|}{s_{(k-1)n'}(M_{\le k-1} + W_{\le k-1})}\right)^{-1} \left(1 + \frac{\|(M+W)_{k,\le k-1}\|}{s_{(k-1)n'}(M_{\le k-1} + W_{\le k-1})}\right)^{-1} \\ &\times \min\left[s_{(k-1)n'}(M_{\le k-1} + W_{\le k-1}), s_{n'}\left(M_k + B_k\right)\right] \end{aligned} \tag{4.216}$$

where we have abbreviated

$$B_k := W_k - (M+W)_{k,\le k-1}(M_{\le k-1} + W_{\le k-1})^{-1}(M+W)_{\le k-1,k} \tag{4.217}$$

Suppose that the event $\mathcal{E}_{k-1}(\beta_{k-1}(\gamma), W_{\le k-1})$ holds. We condition on realizations of the submatrices $M_{\leq k-1}, M_{\leq k-1,k}, M_{k,\leq k-1}$ satisfying

$$\|M_{\leq k-1,k}\|, \|M_{k,\leq k-1}\| \leq (K_0 + K)\sqrt{n} \tag{4.218}$$

and

$$s_{(k-1)n'}(M_{\leq k-1} + W_{\leq k-1}) \geq n^{-\beta_{k-1}(\gamma)}. \tag{4.219}$$

In particular,

$$\|(M + W)_{\leq k-1,k}\|, \|(M + W)_{k,\leq k-1}\| \leq 2n^{\gamma} \tag{4.220}$$

and by our restriction to $\mathcal{B}(K)$ and (4.174),

$$\|M_k + B_k\| \leq (K_0 + K_1)\sqrt{n} + n^{\gamma} + 4n^{\gamma+\beta_{k-1}(\gamma)} \leq 6n^{\gamma+\beta_{k-1}(\gamma)}. \tag{4.221}$$

By Theorem 4.1.24 and independence of M_k from $M_{\leq k-1}, M_{k,\leq k-1}, M_{k,\leq k-1}$, there exists $\beta_k' = O(\gamma^2 + \beta_{k-1}(\gamma)^2)$ such that

$$\mathbb{P}\Big(s_{n'}(M_k + B_k) \leq n^{-\beta_k'}\Big) \ll_{K_0,K,\gamma,\delta,\sigma_0,\varepsilon,\kappa} \sqrt{\frac{\log n}{n}}. \tag{4.222}$$

Restricting further to the event that this bound holds and substituting the above estimates into (4.216), we have

$$s_{kn'}(M_{\leq k} + W) \gg n^{-2\gamma-2\beta_{k-1}(\gamma)} \min(n^{-\beta_{k-1}(\gamma)}, n^{-\beta_k'}) \geq n^{-\beta_k(\gamma)} \tag{4.223}$$

for some $\beta_k(\gamma) = O(\gamma^2 + \beta_{k-1}(\gamma)^2)$. With this choice of $\beta_k(\gamma)$ we have shown

$$\mathbb{P}(\mathcal{E}_k(\beta_k(\gamma), W_{\leq k})^c \wedge \mathcal{E}_{k-1}(\beta_{k-1}(\gamma), W_{\leq k-1})) \ll_{K_0,K,\gamma,\delta,\sigma_0,\varepsilon,\kappa} \sqrt{\frac{\log n}{n}}. \tag{4.224}$$

Applying this bound for all $2 \leq k' \leq k$ together with (4.214) and Bayes' rule we conclude that for any fixed k and any square matrix W of dimension at least kn' and operator norm at most n^γ,

$$\mathbb{P}(\mathcal{E}_k(\beta_k(\gamma), W_{\leq k})^c) \ll_{K_0,K,\gamma,\delta,\sigma_0,\varepsilon,\kappa} k\sqrt{\frac{\log n}{n}}. \tag{4.225}$$

The result now follows by taking $k = m$ and recalling that $m = O_\varepsilon(1)$.

CHAPTER 5

The circular law for signed random regular digraphs

5.1 Introduction

5.1.1 Background

For an $n \times n$ matrix M with complex entries, denote by

$$\mu_M = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(M)} \tag{5.1}$$

its empirical spectral distribution (ESD), where $\lambda_1(M), \ldots, \lambda_n(M) \in \mathbb{C}$ are the (algebraic) eigenvalues of M, counted with multiplicity, and labeled in some arbitrary fashion. When M is a random matrix, μ_M is a random probability measure.

For random matrices with iid entries having finite second moment, the global distribution of the spectrum is asymptotically governed by the *circular law*:

Theorem 5.1.1 (Circular law). *Let ξ be a complex-valued random variable with mean zero and variance one, and for each n let $X_n = (\xi_{ij})_{1 \le i,j \le n}$ be a matrix whose entries are iid copies of ξ. Then with probability one, the rescaled empirical spectral measures $\mu_{\frac{1}{\sqrt{n}} X_n}$ converge weakly to μ_{circ}, the uniform measure on $B_{\mathbb{C}}(0,1)$.*

In the above form Theorem 5.1.1 was established by Tao and Vu in [TV10b], building on important advances of Girko [Gir84] and Bai [Bai97]; earlier versions of the circular law were proved in [Bai97], [BS10], [GT10], [PZ10], [TV08] under additional hypotheses on the distribution of ξ. Theorem 5.1.1 can be viewed as an instance of the *universality phenomenon* in random matrix theory: in the limit as $n \to \infty$, the global distribution of the spectrum of X_n is determined completely by the first two moments of the entries ξ_{ij}. The theorem hence identifies a wide class of matrix ensembles lying in the *circular law universality class*. For additional background and history on the circular law we point the reader to the survey [BC12].

One may then ask if this universality class extends to include some ensembles not covered by Theorem 5.1.1, that is, if we can relax the independence, identical distribution, or moment hypotheses on the entries. Informally, one might expect an ensemble $M_n = (\xi_{ij})$ to exhibit the circular law if the entries ξ_{ij} are centered around a common value, have reasonably bounded tails, are not too degenerate (i.e. their distributions do not concentrate too much near deterministic values), and are only weakly correlated.

For iid matrices, the second moment hypothesis is sharp, and the limiting spectral distribution for certain classes of heavy tail matrices has been described in [BCC11]. Nevertheless, in [Woo12], Wood has shown that the circular law is robust under *sparsification*: letting $X_n = (\xi_{ij})$ be an iid matrix and $B_n = (b_{ij})$ a matrix of iid Bernoulli(p) indicator variables, independent of X_n, it is shown that the empirical spectral distribution of $M_n := \frac{1}{\sqrt{np}} B_n \circ X_n$ converges weakly in probability (see Definition 5.1.3 below) to the circular law, provided that $p = p(n) \gg n^{\varepsilon - 1}$ for any fixed $\varepsilon > 0$. (Here $\circ$ denotes the matrix Hadamard product, i.e. $B_n \circ X_n = (b_{ij}\xi_{ij})$.) The lower bound on p is near-optimal, as it is easy to see that for $p \asymp 1/n$ the matrix M_n will have a linear proportion of trivial columns, so that the limiting spectral distribution will have an atom at zero.

There has been considerable interest in extending the circular law to cover models with dependent entries. In [Ngu14], Nguyen has shown that the circular law holds for matrices drawn uniformly (according to volume measure) from the Birkhoff polytope of doubly stochastic matrices. This followed an analogous result of Bordenave, Caputo and Chafaï on random stochastic matrices, which still enjoy joint independence of the matrix rows. In [AC15], Adamczak and Chafaï have established the circular law for matrices whose distribution over $\mathcal{M}_n(\mathbb{R})$ is a log-concave and unconditional measure, generalizing the circular law for iid real Gaussian matrices due to Edelman [Ede97]. Together with Wolff, the same authors have shown in [ACW16] that the circular law holds for a matrix whose entries are exchangeable and have a finite moments of order $20 + \varepsilon$.

A particularly motivating challenge in random matrix theory is to extend universality results to include random regular graph models. In the non-Hermitian setting we have the rrd matrices considered in Chapters 2 and 3 (see Definition 2.1.15). In this chapter we denote an rrd matrix by A. One can view A as a discrete analogue of the doubly stochastic matrix considered by Nguyen in [Ngu14].

For rrd matrices we have the following augmented version of a conjecture of Bordenave and Chafaï from [BC12] (only case (2) below is stated there).

Conjecture 5.1.2 (Bordenave–Chafaï [BC12])**.** *Let A_n be an rrd matrix.*

1. *If $d = d_n \leq n/2$ satisfies $d \to \infty$ as $n \to \infty$, then $\mu_{\frac{1}{\sqrt{d}}A_n}$ converges to the circular law.*
2. *If $d \geq 3$ is fixed independent of n, then μ_{A_n} converges to the* oriented Kesten-McKay law *on $\mathbb{C}$, with density given by*

$$f_{KM}(w) = \frac{1}{\pi} \frac{d^2(d-1)}{(d^2 - |w|^2)^2} \mathbb{1}_{\{|w| \leq \sqrt{d}\}} \,. \tag{5.2}$$

From the above one obtains results for $n/2 < d \leq n-1$ by considering the complementary matrix with entries $1 - a_{ij}$, which has the same limiting spectral distribution. See Figure 5.1 for some numerical evidence supporting Conjecture 5.1.2.

The explicit density in (5.2) can be computed as the *Brown measure* of the *free sum* of d Haar unitary operators — see [HL00, Example 5.5]. We note that in (5.2), if one rescales w by $\sqrt{d}$ and sends $d \to \infty$ the expression converges to the normalized indicator for the unit disk, which gives some evidence for part (1) of the conjecture. By contiguity results (see for instance [Jan95]), to establish Conjecture 5.1.2 for fixed d it suffices to consider a different measure, the sum of d iid uniform permutation matrices. It was shown by Basak and Dembo in [BD13] that the sum of d iid Haar unitary or orthogonal matrices has limiting spectral distribution given by (5.2), so Conjecture 5.1.2 posits that their result should hold if the unitaries are restricted to permutation matrices.

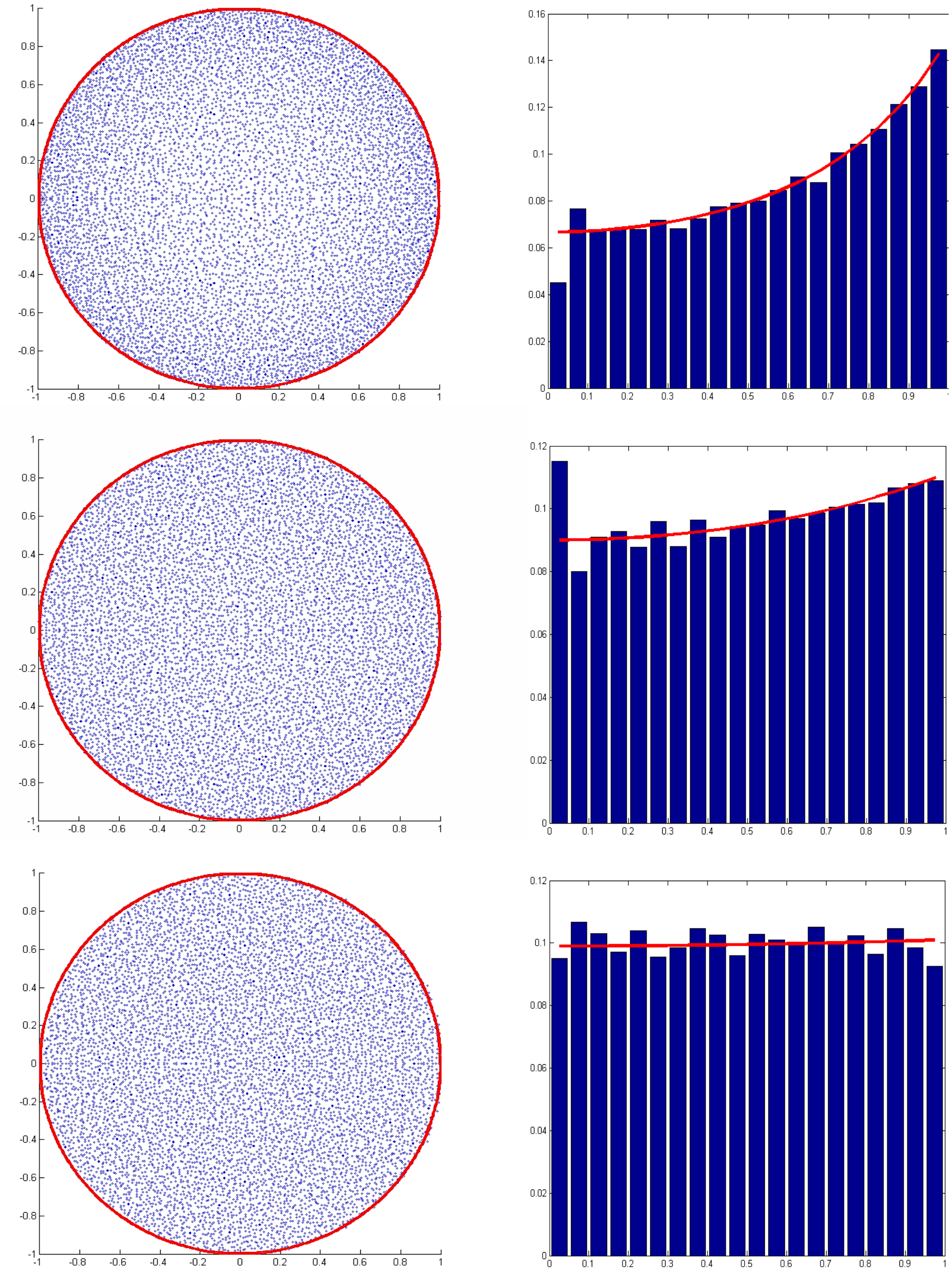

Figure 5.1: Empirical eigenvalue distributions for simulated 8000×8000 rescaled rrd matrices $\frac{1}{\sqrt{d}}A$ for $d = 3$ (top), 10 (middle), and 100 (bottom). Left: scatterplots of eigenvalues, with the unit circle plotted in red for reference. Right: histograms for eigenvalue moduli, with each bin count normalized by 2π times the distance to the origin. The curves predicted by (5.2) are plotted in red. While the eigenvalue distribution is noticeably more dense near the edge of the support for $d = 3$, for $d = 100$ it is indistinguishable from the uniform distribution on the disk.

The two cases of Conjecture 5.1.2 parallel known results for undirected regular graphs. Namely, it was shown by McKay in [McK81a] that for $d \geq 3$ fixed, the limiting spectral distribution of adjacency matrices A_n^{Sym} of d-regular undirected graphs is given by the (explicit) Kesten-McKay law. More recently, it has been shown in [DP12], [TVW13] and [BKY] that if $d \to \infty$ at certain speeds, the empirical spectral distribution of A_n^{Sym} converges to the *semicircular law*, matching the asymptotic behavior of Wigner matrices (the Hermitian analogue of iid matrices). Moreover, these results show that the convergence of ESDs holds on *mesoscopic scales*, i.e. on intervals with length shrinking to zero relative to the limiting support of the spectrum, but growing to infinity relative to the mean spacing of eigenvalues. In particular, the work [BKY] has shown convergence at the optimal mesoscopic scale, provided d grows in the range $f(n) \ll d \ll \left(\frac{n}{f(n)}\right)^{2/3}$ for some $f(n)$ growing poly-logarithmically.

We also highlight an interesting *edge-universality* result of Sodin for undirected random regular graphs [Sod09]. Sodin considers *signed random regular graphs*, i.e. symmetric random matrices of the form $W_n = A_n^{\mathrm{Sym}} \circ H_n$, where H_n is a symmetric matrix with iid uniform Bernoulli signs above the diagonal, independent of the adjacency matrix A_n^{Sym}. Assuming $3 \leq d \leq o(n^{2/3})$, he establishes the Tracy-Widom law for the (appropriately rescaled) largest eigenvalue of W_n, showing these matrices lie in the same class as Wigner matrices for edge universality.

5.1.2 Main result

In the present work we obtain some partial progress towards Conjecture 5.1.2 by considering *signed* random regular digraph matrices of the form

$$Y_n = A_n \circ X_n$$

where X_n is a matrix of iid Bernoulli signs (i.e. $\mathbb{P}(\xi_{ij} = +1) = \mathbb{P}(\xi_{ij} = -1) = 1/2$) independent of the rrd matrix A_n. One can view this as an rrd matrix with some additional randomness, or alternatively as an iid matrix that has been randomly "sparsified" to have each row and column supported on d entries. This is to be compared with the work of Wood

[Woo12] discussed above, where the sparsification is performed by iid Bernoulli indicators. As in that work, we do not obtain convergence to the circular law in the almost sure sense, as in Theorem 5.1.1, but in the following weaker sense:

Definition 5.1.3 (Convergence in probability of random measures)**.** Let $(\mu_n)_{n\geq 1}$ be a sequence of random Borel probability measures supported on the complex plane, and let μ be another Borel probability measure on $\mathbb{C}$. We say that μ_n converges to μ *weakly in probability* if for all bounded continuous functions $f : \mathbb{C} \to \mathbb{R}$ and any $\varepsilon > 0$,

$$\mathbb{P}\left(\left| \int_{\mathbb{C}} f d\mu_n - \int_{\mathbb{C}} f d\mu \right| > \varepsilon \right) \to 0 \qquad \text{as } n \to \infty. \tag{5.3}$$

Now we can state our main result for the chapter:

Theorem 5.1.4 (Circular law for signed rrd matrices)**.** *Fix $p \in (0,1)$, and for each n, let A_n be a uniform random element of $\mathcal{M}_n(d)$, where we write $d = \lfloor pn \rfloor$. Let $\xi \in \{+1, -1\}$ be a centered Bernoulli variable, and for each n let $X_n = (\xi_{ij})_{1\leq i,j\leq n}$ be a matrix of iid copies of ξ, independent of A_n. Denoting $Y_n = A_n \circ X_n$, we have that the rescaled empirical spectral measures $\mu_{\frac{1}{\sqrt{np}}Y_n}$ converge weakly in probability to μ_{circ}, the uniform measure on $B_{\mathbb{C}}(0,1)$.*

Remark 5.1.5 (Extension to sparse digraphs)*.* The above result concerns dense rrd matrices, where a proportion p of the entries are non-zero with p independent of n. An extension to sparse models, with p allowed to decrease at a rate $n^{\varepsilon-1}$ for any fixed $\varepsilon > 0$, is possible by our approach, but requires significantly more work in several places. In the interest of keeping the present chapter clear and of reasonable length, we defer this investigation to a later work.

Remark 5.1.6 (Generalizing the distribution of X_n)*.* The assumption that the entries of X_n are Bernoulli signs is mostly for concreteness. We remark that the entire proof applies if the entries ξ_{ij} are independent and uniformly bounded (and centered with variance one). The proof extends with some additional (standard) truncation arguments to that case that the entries have finite moments of order $4 + \varepsilon$ for any fixed $\varepsilon > 0$.

5.1.3 Outline

The rest of the chapter is organized as follows. In Section 5.2 we give a high level proof of Theorem 5.1.4 following Girko's well-known *Hermitization* approach, reducing our task to establishing (1) the convergence of empirical singular value distributions of scalar perturbations of Y_n, and (2) lower bounds on small singular values of these perturbed matrices. Step (1) is completed in Section 5.3. In Section 5.4, we use Theorem 2.1.5 from Chapter 2 to show that dense random regular digraphs are broadly connected with high probability (see Definition 4.1.10 in Chapter 4), at which point we can apply the results in Chapter 4 on small singular values of random matrices with broadly connected standard deviation profile.

5.1.4 Notation

In addition to the notation in Section 1.2 we use the following conventions. We will generally suppress the subscript n from the matrices X_n, A_n, Y_n etc. For a bounded continuous function f and a Borel probability measure μ supported on $\mathbb{C}$ or $\mathbb{R}$, we will often use the notation $\mu(f) := \int f d\mu$ when the domain of f is clear.

5.2 Reduction to estimates on singular values

We follow Girko's *Hermitization method*, introduced in [Gir84], and which we briefly recap below; see [TV10b] or the survey [BC12] for more background. For a probability measure μ on $\mathbb{C}$ we define the *log-potential*

$$U_\mu(z) := \int_{\mathbb{C}} \log|w - z| d\mu(w). \tag{5.4}$$

Let us abbreviate $\mu_n := \mu_{\frac{1}{\sqrt{np}}Y}$. The following is taken from [TV10b, Theorem 1.20]:

Lemma 5.2.1 (Hermitization). *Let μ be a probability measure on $\mathbb{C}$ satisfying $\int_{\mathbb{C}} |z|^2 d\mu(z) < \infty$. The following are equivalent:*

(i) μ_n converges weakly in probability to μ.

(ii) For almost every $z \in \mathbb{C}$, $U_{\mu_n}(z)$ converges in probability to $U_\mu(z)$.

Denoting the ordered singular values of a matrix $M \in \mathcal{M}_n(\mathbb{C})$ by $s_1(M) \geq \cdots \geq s_n(M)$, from the well-known identity

$$\prod_{i=1}^n |\lambda_i(M)| = |\det(M)| = \prod_{i=1}^n s_i(M)$$

we have

$$\begin{aligned} U_{\mu_n}(z) &= \frac{1}{n} \log \left| \det \left(\frac{1}{\sqrt{np}} Y - zI \right) \right| \\ &= \frac{1}{n} \sum_{i=1}^n \log s_i \left(\frac{1}{\sqrt{np}} Y - zI \right) \\ &= \int_{\mathbb{R}_+} \log(s) d\nu_{n,z}(s) \end{aligned}$$

where we have defined

$$\nu_{n,z} := \frac{1}{n} \sum_{i=1}^n \delta_{s_i\left(\frac{1}{\sqrt{np}} Y - zI\right)}. \tag{5.5}$$

The log-potential for the uniform measure on the unit disk is

$$U(z) := \frac{1}{\pi} \int_{B_{\mathbb{C}}(0,1)} \log |w - z| dw = \begin{cases} \log |z| & \text{if } |z| > 1 \\ -\frac{1}{2}(1 - |z|^2) & \text{otherwise.} \end{cases} \tag{5.6}$$

From the above lines and Lemma 5.2.1, to establish Theorem 5.1.4 it suffices to show that for almost every $z \in \mathbb{C}$, $\int_{\mathbb{R}_+} \log(s) d\nu_{n,z}(s)$ converges in probability to $U(z)$.

As a first step we will establish the following:

Proposition 5.2.2 (Weak convergence of $\nu_{n,z}$)**.** *For all $z \in \mathbb{C}$, $\nu_{n,z}$ converges vaguely in probability to a deterministic probability measure ν_z on $\mathbb{R}_+$. That is, for any $f \in C_c(\mathbb{R}_+)$ and any $\varepsilon > 0$,*

$$\mathbb{P}(|\nu_{n,z}(f) - \nu_z(f)| > \varepsilon) = o(1) \tag{5.7}$$

where the rate of decay on the right hand side can depend on ε, f. Furthermore, for all $z \in \mathbb{C}$

the measures ν_z *satisfy*

$$\int_{\mathbb{R}_+} \log(s)d\nu_z(s) = U(z) \tag{5.8}$$

where $U(z)$ *was defined in* (5.6).

Remark 5.2.3. For the $o(1)$ term in (5.7), in the proof we actually obtain the bound $O(\exp(-cpn^2))$, with c and the implied constant depending on f, ε and z. With the Borel-Cantelli lemma this implies that the measures $\nu_{n,z}$ converge weakly to ν_z with probability one. However, towards the proof of Theorem 5.1.4 we do not need this refinement.

In order to deduce from this the convergence of log-potentials $U_{\mu_n}(z)$, we will apply the above proposition to a truncation $f \in C_c(\mathbb{R}_+)$ of the function $s \mapsto \log(s)$. To show that the truncated integral $\int_{\mathbb{R}_+} f d\nu_{n,z}$ is a good approximation of $U_{\mu_n}(z)$, we must prove that the measures $\nu_{n,z}$ uniformly integrate the singularities of $s \mapsto \log(s)$.

To handle the singularity at $+\infty$ we need control of the largest singular value, i.e. the operator norm. For this we have the following:

Lemma 5.2.4 (Operator norm bound)**.** *Let* $M = (\xi_{ij})$ *be an* $n \times n$ *matrix whose entries are jointly independent, centered, with* $\xi_{ij} = O(1)$ *uniformly in* i, j*. Except with probability* $O(e^{-cn})$*, we have* $\|M\| = O(\sqrt{n})$*.*

The bounds above are stronger than what is needed for our present purposes, but will be useful in subsequent sections when we seek control of the smallest singular value $s_n(\frac{1}{\sqrt{np}}Y - zI)$.

Proof. From [Lat05, Theorem 2] we have that $\mathbb{E}\|M\| = O(\sqrt{n})$. Furthermore, as the function $M \mapsto \|M\|$ is a convex, 1-Lipschitz (with respect to the Euclidean (Hilbert-Schmidt) norm) function on $\mathcal{M}_n(\mathbb{C})$, from Talagrand's isoperimetric inequality [Tal96, Theorem 6.6] (see also [AGZ10, Corollary 4.4.11]), it is easily shown that $\|M\| = \mathbb{E}\|M\| + O(1)$ except with probability $O(e^{-cn})$, and the result follows. $\square$

As a consequence, by the triangle inequality for operator norm we have

Corollary 5.2.5. *Except with probability* $O(e^{-cn})$ *we have* $s_1(\frac{1}{\sqrt{np}}Y - zI) = O_{z,p}(1)$*.*

In Section 5.4 we will deduce the following two propositions from results in Chapters 2 and 4.

Proposition 5.2.6 (Smallest singular value)**.** *For any $t > 0$,*

$$\mathbb{P}\big(s_n(Y - z\sqrt{np}I) \leq tn^{-1/2}\big) \ll t + n^{-1/2} \tag{5.9}$$

where the implied constant depends only on z and p.

Proposition 5.2.7 (Local Wegner estimate)**.** *There are constants $a_0, a_1, a_2 > 0$ depending only on p, z such that for all $\alpha \in (0,1)$, except with probability $O(\exp(-a_0 n^\alpha))$, for all $i \in [n^\alpha, a_1 n]$ we have*

$$s_{n-i}\left(\frac{1}{\sqrt{np}}Y - zI\right) \geq a_2 \frac{i}{n}. \tag{5.10}$$

Now we conclude the proof of Theorem 5.1.4 on Propositions 5.2.2, 5.2.6, 5.2.7 and Corollary 5.2.5. Fix $z \in \mathbb{C}$. Here we will allow implied constants to depend on p and z. Our aim is to show that for any $\varepsilon > 0$,

$$\mathbb{P}\big(\big|\nu_{n,z}(\log) - \nu_z(\log)\big| > \varepsilon\big) = o(1) \tag{5.11}$$

with ν_z as in Proposition 5.2.2. From Corollary 5.2.5 we have

$$\mathbb{P}(\nu_{n,z}([C_0, \infty)) = 0) = 1 - o(1) \tag{5.12}$$

for some $C_0 > 0$ sufficiently large depending only on z. In particular, with Proposition 5.2.2 this implies that ν_z is supported on $[0, C_0)$. For $\eta > 0$ small, let $f_\eta \in C_c(\mathbb{R}_+)$ satisfy

$$f_\eta(s) = \begin{cases} 0 & s \in [0, \eta/2] \cup [2C_0, \infty) \\ \log(s) & s \in [\eta, C_0] \end{cases}$$

and take f_η to be linearly increasing and decreasing on the intervals $[\eta/2, \eta]$ and $[C_0, 2C_0]$, respectively. From Proposition 5.2.2 we have that for any fixed $\varepsilon, \eta > 0$,

$$\mathbb{P}(|\nu_{n,z}(f_\eta) - \nu_z(f_\eta)| > \varepsilon) = o(1) \tag{5.13}$$

so from (5.12) it only remains to show

$$\int_0^\eta |\log(s)| d\nu_{n,z}(s) = O(\eta^{0.9}) \tag{5.14}$$

(say) for all $\eta > 0$ sufficiently small.

For ease of notation we write s_i for $s_i(\frac{1}{\sqrt{np}}Y - zI)$. Fix $\alpha \in (0,1)$ arbitrarily. We write left hand side of (5.14) as

$$\frac{1}{n}\sum_{i=0}^{n-1} |\log(s_{n-i})| \, \mathbb{1}(s_{n-i} \leq \eta) = \frac{1}{n}\sum_{i=0}^{\lfloor n^\alpha \rfloor} |\log(s_{n-i})| + \frac{1}{n}\sum_{i > n^\alpha} |\log(s_{n-i})| \, \mathbb{1}(s_{n-i} \leq \eta). \tag{5.15}$$

By Proposition 5.2.6, except with probability $O(n^{-1/2})$, the first term on the right hand side is bounded by $O(n^{1-\alpha}\log n) = o(1)$. By Proposition 5.2.7, taking $\eta < a_1 a_2$ we have except with probability $O(\exp(-a_0 n^\alpha))$, the second term is bounded by

$$\frac{1}{n}\sum_{1 \leq i < \eta n/a_2} \left|\log\left(a_2 \frac{i}{n}\right)\right| \ll \frac{1}{n}\sum_{1 \leq i < \eta n/a_2} \left(a_2 \frac{i}{n}\right)^{-0.1} \ll \int_0^{2\eta/a_2} s^{-0.1} ds = O(\eta^{0.9})$$

which completes the proof.

5.3 Weak convergence of singular value distributions

In this section we prove Proposition 5.2.2. Following an approach of Tran, Vu and Wang from [TVW13], we compare the signed rrd matrix Y with an iid matrix taking the form

$$\widetilde{Y} = (b_{ij}\xi_{ij})_{1 \leq i,j \leq n} = B \circ X$$

where the variables b_{ij} are Bernoulli(p) indicator variables (recall that $d = \lfloor pn \rfloor$), jointly independent of each other and of the matrix X. Note that the entries of $\frac{1}{\sqrt{p}}\widetilde{Y}$ are iid and centered with unit variance. Writing $\widetilde{\nu}_{n,z} := \frac{1}{n}\sum_{i=1}^{n} \delta_{s_i(\frac{1}{\sqrt{np}}\widetilde{Y}-zI)}$, it follows from the Tao–Vu circular law Theorem 5.1.1 and Lemma 5.2.1 that for all $z \in \mathbb{C}$ the measures $\widetilde{\nu}_{n,z}$ converge weakly in probability to a deterministic probability measure ν_z satisfying (5.8). (Alternatively one can deduce the weak convergence of $\widetilde{\nu}_{n,z}$ from the result of Dozier and Silverstein [DS07]; the limit ν_z must satisfy (5.8) as it this holds for the case of real Gaussian entries by the work of Edelman [Ede97].) In particular, we have that for any $f \in BC(\mathbb{R}_+)$,

$$|\nu_z(f) - \mathbb{E}\,\widetilde{\nu}_{n,z}(f)| = o(1) \tag{5.16}$$

so it suffices to show

$$\mathbb{P}(|\nu_{n,z}(f) - \mathbb{E}\,\widetilde{\nu}_{n,z}(f)| > \varepsilon) = o(1) \tag{5.17}$$

for any $\varepsilon > 0$ and any $f \in C_c(\mathbb{R}_+)$.

Now we describe the comparison method from [TVW13]. Define the event

$$\mathcal{E}_{n,d} := \big\{B \in \mathcal{M}_n(d)\big\} \tag{5.18}$$

that the Bernoulli matrix B is the adjacency matrix of a d-regular digraph. Note that $B|\mathcal{E}_{n,d} \overset{d}{=} A$. Hence, to bound an event $\{A \in \mathcal{M}_0\}$ for some set $\mathcal{M}_0 \subset \mathcal{M}_n(d)$, we can bound the corresponding event $\{B \in \mathcal{M}_0\}$ and apply the identity

$$\mathbb{P}(A \in \mathcal{M}_0) = \frac{\mathbb{P}(B \in \mathcal{M}_0)}{\mathbb{P}(\mathcal{E}_{n,d})} \tag{5.19}$$

together with a lower bound on $\mathbb{P}(\mathcal{E}_{n,d})$. For this we have the following lemma of Tran:

Lemma 5.3.1 (Tran [Tra]). $\mathbb{P}(\mathcal{E}_{n,d}) = \exp\big(-O(nd^{1/2})\big)$.

By the above lemma, (5.19) and (5.17), it suffices to show that for any $f \in C_c(\mathbb{R}_+)$ and any $\varepsilon > 0$,

$$\mathbb{P}(|\,\widetilde{\nu}_{n,z}(f) - \mathbb{E}\,\widetilde{\nu}_{n,z}(f)| > \varepsilon) = \exp\big(-\omega\big(p^{1/2}n^{3/2}\big)\big). \tag{5.20}$$

We have hence reduced our task to proving a concentration bound for linear statistics of the singular value distribution of the perturbed iid matrix $\frac{1}{\sqrt{np}}\widetilde{Y} - zI$. For this task we have the following lemma, which follows from the work of Guionnet and Zeitouni in [GZ00]:

Lemma 5.3.2 (Concentration of linear statistics)**.** *Let $H = (h_{ij})_{1\le i,j\le n}$ be a Hermitian random matrix, and assume the variables on and above the diagonal are jointly independent and that $|h_{ij}| \le K/\sqrt{n}$ uniformly in i, j for some $K \in (0, \infty)$. Let $f : \mathbb{R} \to \mathbb{R}$ be an L-Lipschitz function supported on a compact interval $I \subset \mathbb{R}$, and let H_0 be an arbitrary deterministic $n \times n$ Hermitian matrix. Then for any $\delta > 0$,*

$$\mathbb{P}\left(|\mu_{H+H_0}(f) - \mathbb{E}\,\mu_{H+H_0}(f)| \ge \delta \right) \le \frac{C|I|}{\delta} \exp\left(-\frac{cn^2\delta^4}{K^2L^2|I|^2} \right) \tag{5.21}$$

for some absolute constants $C, c > 0$.

Proof. For the case that $H_0 = 0$, this follows directly from [GZ00, Theorem 1.3(a)]. For the general case, the only part of the argument that needs modification is in the proof of their Theorem 1.1(a), where we need to show that for $f : \mathbb{R} \to \mathbb{R}$ convex and L-Lipschitz, the function $H \mapsto \mu_{H+H_0}(f)$ is a convex and $O(L)$-Lipschitz function on the space of Hermitian matrices. However, this follows directly from their Lemma 1.2 and the fact that the convexity and Lipschitz properties are invariant under translations $H \mapsto H + H_0$. The rest of the proofs of [GZ00, Theorem 1.1(a)] and [GZ00, Theorem 1.3(a)] apply with no modification. $\square$

To apply this concentration estimate to the measures $\widetilde{\nu}_{n,z}$, we recall the linearization approach to the study of singular value distributions of random matrices. From $\widetilde{Y}$ we form a $2n \times 2n$ Hermitian matrix

$$H(z) = \frac{1}{\sqrt{np}} \begin{pmatrix} 0 & \widetilde{Y} - z\sqrt{np}I \\ (\widetilde{Y} - z\sqrt{np}I)^* & 0 \end{pmatrix}. \tag{5.22}$$

It is routine to verify that the $2n$ eigenvalues of $H(z)$, counted with multiplicity, are $\{\pm s_i(\frac{1}{\sqrt{np}}\widetilde{Y} - zI)\}_{i=1}^n$. In terms of empirical spectral measures, $\mu_{H(z)}$ is the symmetrization across the origin of the measure $\widetilde{\nu}_{n,z}$ on $\mathbb{R}_+$. From (5.20), it now suffices to show

$$\mathbb{P}\big(|\mu_{H(z)}(f) - \mathbb{E}\,\mu_{H(z)}(f)| > \varepsilon\big) = \exp\big(-\omega\big(p^{1/2}n^{3/2}\big)\big). \tag{5.23}$$

for any $\varepsilon > 0$ and any $f \in C_c(\mathbb{R})$.

Fix such ε and f. As f is uniformly continuous, there exists $\delta = \delta(\varepsilon) > 0$ such that $|f(s) - f(t)| \leq \varepsilon/10$ whenever $|s-t| \leq \delta$. By linear interpolation on a δ-mesh of the support of f we may find a function $f_\varepsilon \in C_c(\mathbb{R})$ with Lipschitz constant $O(\varepsilon/\delta) = O_\varepsilon(1)$, and such that $\|f - f_\varepsilon\|_\infty \leq \varepsilon/10$. It now suffices to show

$$\mathbb{P}\big(|\mu_{H(z)}(f_\varepsilon) - \mathbb{E}\,\mu_{H(z)}(f_\varepsilon)| > \varepsilon/2\big) = \exp\big(-\omega\big(p^{1/2}n^{3/2}\big)\big). \tag{5.24}$$

Finally, we note that $H(z)$ has the form $H + H_0$ as in Lemma 5.3.2, with $K = O(1/\sqrt{p})$, $H = H(0)$ and

$$H_0 = \begin{pmatrix} 0 & -zI \\ -z^*I & 0 \end{pmatrix}.$$

Applying that lemma we have

$$\mathbb{P}\big(|\mu_{H(z)}(f_\varepsilon) - \mathbb{E}\,\mu_{H(z)}(f_\varepsilon)| > \varepsilon/2\big) = O_\varepsilon\Big(\exp\big(-c_\varepsilon n^2 p\big)\Big) \tag{5.25}$$

which gives (5.24) (with room to spare), to complete the proof of (5.7).

Remark 5.3.3 (Extension to sparse matrices). In the present work we are leaving $p \in (0,1)$ as a fixed parameter, but we note in passing that for the final bound (5.25) to beat the desired bound (5.24), we simply need $np \to \infty$. However, for the above proof to apply with $p = o(1)$ we need a substitute for the convergence result (5.16), which we deduced from a result concerning matrices taken whose entries are iid copies of a fixed random variable with distribution independent of n. We will address this issue in a subsequent work.

5.4 Bounds on small singular values

In this section we deduce Propositions 5.2.6 and 5.2.7 from Theorems 4.1.14 and 4.5.1 in Chapter 4 on the singular values of random matrices with a broadly connected standard deviation profile. First we use Theorem 2.1.5 in Chapter 2 to show that dense random regular digraphs are broadly connected with high probability.

Proposition 5.4.1 (Random regular digraphs are broadly connected)**.** *Let $p \in (0,1)$ and let A be a uniform random element of $\mathcal{M}_n(\lfloor np \rfloor)$. Then A is $(p/2, p/8)$-broadly connected with probability $1 - O_p(e^{-cn})$ for some $c > 0$ depending only on p.*

Proof. We may assume that n is sufficiently large depending on p. A clearly satisfies conditions (1) and (2) in Definition 4.1.10 with $\delta = p/2$ since every row and column has support $\lfloor np \rfloor$. Now we verify condition (3). Let $J \subset [n]$, and abbreviate $I_0 = I_0(J) = \mathcal{N}_{A^\mathsf{T}}^{(p/2)}(J)$. Since each column of A has support $\lfloor np \rfloor$, we have

$$
\begin{aligned}
\lfloor np \rfloor |J| &= \sum_{i=1}^{n} \sum_{j \in J} a_{ij} \\
&= \sum_{i \in I_0} |\mathcal{N}(i) \cap J| + \sum_{i \notin I_0} |\mathcal{N}(i) \cap J| \\
&< |I_0||J| + (p/2)n|J|
\end{aligned}
$$

which rearranges to

$$|I_0| > \frac{1}{2}pn - O(1) \gg pn.$$

Thus, condition (3) is satisfied (deterministically) for any J of size at most cpn for a sufficiently small absolute constant $c > 0$. On the other hand, if $|J| > n(1 - p/4)$ then we must have $|\mathcal{N}(i) \cap J| \geq (p/2)n$ for all $i \in [n]$, so condition (3) also holds deterministically for any J of size at least $n(1 - p/4)$. So we may assume $cpn \leq |J| \leq (1 - p/4)n$. In particular,

$$(1 + 2\nu)|J| = (1 + p/4)|J| \leq n \tag{5.26}$$

Suppose that $|I_0(J)| < (1+\nu)|J|$. Then

$$|I_0^c| > n - (1+\nu)|J| = \nu|J| + n - (1+2\nu)|J| \geq \nu|J|$$

and

$$e_A(I_0^c, J) = \sum_{i \in I_0^c} |\mathcal{N}(i) \cap J| < \frac{1}{2} p |I_0^c||J|.$$

Thus, on the event that $|I_0(J)| < (1+\nu)|J|$ for some $J \subset [n]$, there exists $I \subset [n]$ such that $|I| \gg p|J|$ and $e_A(I,J) < \frac{1}{2}p|I||J|$. But by Theorem 2.1.5 from Chapter 2, if we restrict to the event $\mathcal{G}^{\mathrm{co}}(\eta)$ for $\eta > 0$ a sufficiently small constant then this occurs for fixed I, J with probability at most e^{-cn^2} for some $c > 0$ sufficiently small depending only on p and all n sufficiently large. Applying the union bound over all choices of $I, J \subset [n]$ of the relevant sizes we have

$$\mathbb{P}(\exists J \subset [n] : |I_0(J)| < (1+\nu)|J|) \leq \mathbb{P}(\mathcal{G}^{\mathrm{co}}(\eta)^c) + 4^n e^{-cn^2} \leq e^{-c'n} \tag{5.27}$$

for some $c' > 0$ sufficiently small depending only on p. □

Now we can prove Propositions 5.2.6 and 5.2.7 using the results from Chapter 4. Fix $p \in (0,1)$, $z \in \mathbb{C}$, let A, X, Y be as in Proposition 5.2.6, and write $B := -z\sqrt{np}I$, $M = Y + B$. Conditional on any realization of A, by Corollary 5.2.5 we may restrict to the event $\{\|M\| \leq K\sqrt{n}\}$ for some $K = O_{p,z}(1)$. By Proposition 5.4.1 we may further condition on A lying in the event that it is $(p/2, p/8)$-broadly connected. Propositions 5.2.6 and 5.2.7 now follow from Theorems 4.1.14 and 4.5.1.

Bibliography

[AC15] Radosław Adamczak and Djalil Chafaï. Circular law for random matrices with unconditional log-concave distribution. *Commun. Contemp. Math.*, 17(4):1550020 (22 pages), 2015.

[ACW16] Radosław Adamczak, Djalil Chafaï, and Paweł Wolff. Circular law for random matrices with exchangeable entries. *Random Structures Algorithms*, 48(3):454–479, 2016.

[AGZ10] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[AKS99] Noga Alon, Michael Krivelevich, and Benny Sudakov. List coloring of random and pseudo-random graphs. *Combinatorica*, 19(4):453–472, 1999.

[ARS15] Johnatan Aljadeff, David Renfrew, and Merav Stern. Eigenvalues of block structured asymmetric random matrices. *J. Math. Phys.*, 56(10):103502, 14, 2015.

[AS04] Noga Alon and Asaf Shapira. Testing subgraphs in directed graphs. *J. Comput. System Sci.*, 69(3):353–382, 2004.

[AS08] Noga Alon and Joel H. Spencer. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, third edition, 2008. With an appendix on the life and work of Paul Erdős.

[Bai97] Z. D. Bai. Circular law. *Ann. Probab.*, 25(1):494–529, 1997.

[Bar93] Ya M Barzdin. On the realization of networks in three-dimensional space. In *Selected Works of AN Kolmogorov*, pages 194–202. Springer, 1993.

[BC78] Edward A. Bender and E. Rodney Canfield. The asymptotic number of labeled graphs with given degree sequences. *J. Combinatorial Theory Ser. A*, 24(3):296–307, 1978.

[BC12] Charles Bordenave and Djalil Chafaï. Around the circular law. *Probab. Surv.*, 9:1–89, 2012.

[BCC11] Charles Bordenave, Pietro Caputo, and Djalil Chafaï. Spectrum of non-Hermitian heavy tailed random matrices. *Comm. Math. Phys.*, 307(2):513–560, 2011.

[BCC12] Charles Bordenave, Pietro Caputo, and Djalil Chafaï. Circular law theorem for random Markov matrices. *Probab. Theory Related Fields*, 152(3-4):751–779, 2012.

[BD13] Anirban Basak and Amir Dembo. Limiting spectral distribution of sums of unitary and orthogonal matrices. *Electron. Commun. Probab.*, 18:no. 69, 19, 2013.

[BFSU99] Andrei Z. Broder, Alan M. Frieze, Stephen Suen, and Eli Upfal. Optimal construction of edge-disjoint paths in random graphs. *SIAM J. Comput.*, 28(2):541–573 (electronic), 1999.

[BKY] Roland Bauerschmidt, Antti Knowles, and Horng-Tzer Yau. Local semicircle law for random regular graphs. Preprint available at arXiv:1503.08702.

[Bol80] Béla Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.*, 1(4):311–316, 1980.

[BS10] Zhidong Bai and Jack W. Silverstein. *Spectral analysis of large dimensional random matrices.* Springer Series in Statistics. Springer, New York, second edition, 2010.

[BSY88] Z. D. Bai, Jack W. Silverstein, and Y. Q. Yin. A note on the largest eigenvalue of a large-dimensional sample covariance matrix. *J. Multivariate Anal.*, 26(2):166–168, 1988.

[BT87] J. Bourgain and L. Tzafriri. Invertibility of "large" submatrices with applications to the geometry of Banach spaces and harmonic analysis. *Israel J. Math.*, 57(2):137–224, 1987.

[BvH] Afonso S. Bandeira and Ramon van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. Preprint available at arXiv:1408.6185.

[BVW10] Jean Bourgain, Van H. Vu, and Philip Matchett Wood. On the singularity probability of discrete random matrices. *J. Funct. Anal.*, 258(2):559–603, 2010.

[CDS] Sourav Chatterjee, Persi Diaconis, and Allan Sly. On properties of random doubly stochastic matrices. To appear in Ann. de l'Inst. Henri Poincaré.

[CGJ] Nicholas A. Cook, Larry Goldstein, and Tobias Johnson. Size biased couplings and the spectral gap for random regular graphs. Preprint available at arXiv:1510.06013.

[CGS11] Louis H. Y. Chen, Larry Goldstein, and Qi-Man Shao. *Normal approximation by Stein's method*. Probability and its Applications (New York). Springer, Heidelberg, 2011.

[CGW89] F. R. K. Chung, R. L. Graham, and R. M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, 1989.

[Cha05] Sourav Chatterjee. *Concentration inequalities with exchangeable pairs*. PhD thesis, Stanford University, 2005.

[Cha07] Sourav Chatterjee. Stein's method for concentration inequalities. *Probab. Theory Related Fields*, 138(1-2):305–321, 2007.

[CHNR] Nicholas A. Cook, Walid Hachem, Jamal Najim, and David Renfrew. Limiting spectral distribution for non-hermitian random matrices with a variance profile. In preparation.

[CM05] E. Rodney Canfield and Brendan D. McKay. Asymptotic enumeration of dense 0-1 matrices with equal row sums and equal column sums. *Electron. J. Combin.*, 12:Research Paper 29, 31 pp. (electronic), 2005.

[CM10] E. Rodney Canfield and Brendan D. McKay. Asymptotic enumeration of integer matrices with large equal row and column sums. *Combinatorica*, 30(6):655–680, 2010.

[Coo15] Nicholas A. Cook. On the singularity of adjacency matrices for random regular digraphs. *Probab. Theory Related Fields*, pages 1–58, 2015. doi:10.1007/s00440-015-0679-8.

[Coo16] Nicholas A. Cook. Discrepancy properties for random regular digraphs. *Random Structures Algorithms*, 2016. doi:10.1002/rsa.20643.

[DJ16] Ioana Dumitriu and Tobias Johnson. The Marčenko-Pastur law for sparse random bipartite biregular graphs. *Random Structures Algorithms*, 48(2):313–340, 2016.

[DJPP13] Ioana Dumitriu, Tobias Johnson, Soumik Pal, and Elliot Paquette. Functional limit theorems for random regular graphs. *Probab. Theory Related Fields*, 156(3-4):921–975, 2013.

[DP12] Ioana Dumitriu and Soumik Pal. Sparse regular random graphs: spectral density and eigenvectors. *Ann. Probab.*, 40(5):2197–2235, 2012.

[DS07] R. Brent Dozier and Jack W. Silverstein. On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices. *J. Multivariate Anal.*, 98(4):678–694, 2007.

[Ede88] Alan Edelman. Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.*, 9(4):543–560, 1988.

[Ede89] Alan Stuart Edelman. *Eigenvalues and condition numbers of random matrices.* PhD thesis, Massachusetts Institute of Technology, 1989.

[Ede97] Alan Edelman. The probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular law. *J. Multivariate Anal.*, 60(2):203–232, 1997.

[Erd45] Paul Erdős. On a lemma of Littlewood and Offord. *Bull. Amer. Math. Soc.*, 51:898–902, 1945.

[Fel68] William Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons, Inc., New York-London-Sydney, 1968.

[FKS89] Joel Friedman, Jeff Kahn, and Endre Szemeredi. On the second eigenvalue of random regular graphs. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 587–598. ACM, 1989.

[Fri08] Joel Friedman. A proof of Alon's second eigenvalue conjecture and related problems. *Mem. Amer. Math. Soc.*, 195(910):viii+100, 2008.

[GG11] Subhankar Ghosh and Larry Goldstein. Concentration of measures via size-biased couplings. *Probab. Theory Related Fields*, 149(1-2):271–278, 2011.

[Gir84] V. L. Girko. The circular law. *Teor. Veroyatnost. i Primenen.*, 29(4):669–679, 1984.

[GKZ11] Alice Guionnet, Manjunath Krishnapur, and Ofer Zeitouni. The single ring theorem. *Ann. of Math. (2)*, 174(2):1189–1217, 2011.

[Gow97] W. T. Gowers. Lower bounds of tower type for Szemerédi's uniformity lemma. *Geom. Funct. Anal.*, 7(2):322–337, 1997.

[GR96] Larry Goldstein and Yosef Rinott. Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.*, 33(1):1–17, 1996.

[GT10] Friedrich Götze and Alexander Tikhomirov. The circular law for random matrices. *Ann. Probab.*, 38(4):1444–1491, 2010.

[GZ00] A. Guionnet and O. Zeitouni. Concentration of the spectral measure for large matrices. *Electron. Comm. Probab.*, 5:119–136 (electronic), 2000.

[HL00] Uffe Haagerup and Flemming Larsen. Brown's spectral distribution measure for R-diagonal elements in finite von Neumann algebras. *J. Funct. Anal.*, 176(2):331–367, 2000.

[HLN07] Walid Hachem, Philippe Loubaton, and Jamal Najim. Deterministic equivalents for certain functionals of large random matrices. *Ann. Appl. Probab.*, 17(3):875–930, 2007.

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561 (electronic), 2006.

[Jan95] Svante Janson. Random regular graphs: asymptotic distributions and contiguity. *Combin. Probab. Comput.*, 4(4):369–405, 1995.

[KKS95] Jeff Kahn, János Komlós, and Endre Szemerédi. On the probability that a random ± 1-matrix is singular. *J. Amer. Math. Soc.*, 8(1):223–240, 1995.

[Kom] János Komlós. Circulated manuscript, 1977. Edited version available online at: http://www.math.rutgers.edu/∼komlos/01short.pdf.

[Kom67] J. Komlós. On the determinant of (0, 1) matrices. *Studia Sci. Math. Hungar*, 2:7–21, 1967.

[Kom68] J. Komlós. On the determinant of random matrices. *Studia Sci. Math. Hungar.*, 3:387–399, 1968.

[KS96] J. Komlós and M. Simonovits. Szemerédi's regularity lemma and its applications in graph theory. In *Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993)*, volume 2 of *Bolyai Soc. Math. Stud.*, pages 295–352. János Bolyai Math. Soc., Budapest, 1996.

[KS06] M. Krivelevich and B. Sudakov. Pseudo-random graphs. In *More sets, graphs and numbers*, volume 15 of *Bolyai Soc. Math. Stud.*, pages 199–262. Springer, Berlin, 2006.

[KSVW01] Michael Krivelevich, Benny Sudakov, Van H. Vu, and Nicholas C. Wormald. Random regular graphs of high degree. *Random Structures Algorithms*, 18(4):346–363, 2001.

[Lat05] Rafał Latała. Some estimates of norms of random matrices. *Proc. Amer. Math. Soc.*, 133(5):1273–1282 (electronic), 2005.

[LLT$^+$] Alexander E. Litvak, Anna Lytova, Konstantin Tikhomirov, Nicole Tomczak-Jaegermann, and Pierre Youssef. Adjacency matrices of random digraphs: singularity and anti-concentration. Preprint available at arXiv:math/1511.00113.

[LPR$^+$05] A. E. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, and R. Vershynin. Euclidean embeddings in spaces of finite volume ratio via random matrices. *J. Reine Angew. Math.*, 589:1–19, 2005.

[LPRTJ05] A. E. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann. Smallest singular value of random matrices and geometry of random polytopes. *Adv. Math.*, 195(2):491–523, 2005.

[LR12] Alexander E. Litvak and Omar Rivasplata. Smallest singular value of sparse random matrices. *Studia Math.*, 212(3):195–218, 2012.

[McD89] Colin McDiarmid. On the method of bounded differences. In *Surveys in combinatorics, 1989 (Norwich, 1989)*, volume 141 of *London Math. Soc. Lecture Note Ser.*, pages 148–188. Cambridge Univ. Press, Cambridge, 1989.

[McK] Brendan D. McKay. Non-enumerative proof that there are many derangements? MathOverflow. URL: http://mathoverflow.net/q/86202 (version: 2012-01-23).

[McK81a] Brendan D. McKay. The expected eigenvalue distribution of a large regular graph. *Linear Algebra Appl.*, 40:203–216, 1981.

[McK81b] Brendan D. McKay. Subgraphs of random graphs with specified degrees. In *Proceedings of the Twelfth Southeastern Conference on Combinatorics, Graph Theory and Computing, Vol. II (Baton Rouge, La., 1981)*, volume 33, pages 213–223, 1981.

[MRRW97] M. S. O. Molloy, H. Robalewska, R. W. Robinson, and N. C. Wormald. 1-factorizations of random regular graphs. *Random Structures Algorithms*, 10(3):305–321, 1997.

[MSS14] A.W. Marcus, D.A. Spielman, and N. Srivastava. Ramanujan graphs and the solution of the kadison–singer problem. In *Proc. ICM, Vol III*, pages 375–386, 2014.

[MW90] Brendan D. McKay and Nicholas C. Wormald. Asymptotic enumeration by degree sequence of graphs of high degree. *European J. Combin.*, 11(6):565–580, 1990.

[MW91] Brendan D. McKay and Nicholas C. Wormald. Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$. *Combinatorica*, 11(4):369–382, 1991.

[MW03] Brendan D. McKay and Xiaoji Wang. Asymptotic enumeration of 0-1 matrices with equal row sums and equal column sums. *Linear Algebra Appl.*, 373:273–287, 2003. Special issue on the Combinatorial Matrix Theory Conference (Pohang, 2002).

[Ngu13] Hoi H. Nguyen. On the singularity of random combinatorial matrices. *SIAM J. Discrete Math.*, 27(1):447–458, 2013.

[Ngu14] Hoi H. Nguyen. Random doubly stochastic matrices: the circular law. *Ann. Probab.*, 42(3):1161–1196, 2014.

[NV13] Hoi H. Nguyen and Van H. Vu. Circular law for random discrete matrices of given row sum. *J. Comb.*, 4(1):1–30, 2013.

[Pin73] Mark S Pinsker. On the complexity of a concentrator. In *7th International Telegraffic Conference*, volume 4, pages 1–318. Citeseer, 1973.

[PZ10] Guangming Pan and Wang Zhou. Circular law, extreme singular values and potential theory. *J. Multivariate Anal.*, 101(3):645–656, 2010.

[RA06] Kanaka Rajan and LF Abbott. Eigenvalue spectra of random matrices for neural networks. *Physical review letters*, 97(18):188104, 2006.

[Rud06] Mark Rudelson. Lower estimates for the singular values of random matrices. *C. R. Math. Acad. Sci. Paris*, 342(4):247–252, 2006.

[Rud08] Mark Rudelson. Invertibility of random matrices: norm of the inverse. *Ann. of Math. (2)*, 168(2):575–600, 2008.

[RV08] Mark Rudelson and Roman Vershynin. The Littlewood-Offord problem and invertibility of random matrices. *Adv. Math.*, 218(2):600–633, 2008.

[RV09] Mark Rudelson and Roman Vershynin. Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.*, 62(12):1707–1739, 2009.

[RV14] Mark Rudelson and Roman Vershynin. Invertibility of random matrices: unitary and orthogonal perturbations. *J. Amer. Math. Soc.*, 27(2):293–338, 2014.

[RZ16] Mark Rudelson and Ofer Zeitouni. Singular values of Gaussian matrices and permanent estimators. *Random Structures Algorithms*, 48(1):183–212, 2016.

[Seg00] Yoav Seginer. The expected norm of random matrices. *Combin. Probab. Comput.*, 9(2):149–166, 2000.

[Sma85] Steve Smale. On the efficiency of algorithms of analysis. *Bull. Amer. Math. Soc. (N.S.)*, 13(2):87–121, 1985.

[Sod09] Sasha Sodin. The Tracy-Widom law for some sparse random matrices. *J. Stat. Phys.*, 136(5):834–841, 2009.

[Sos02] Alexander Soshnikov. A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Statist. Phys.*, 108(5-6):1033–1056, 2002. Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays.

[SS12] Daniel A. Spielman and Nikhil Srivastava. An elementary proof of the restricted invertibility theorem. *Israel J. Math.*, 190:83–91, 2012.

[SST06] Arvind Sankar, Daniel A. Spielman, and Shang-Hua Teng. Smoothed analysis of the condition numbers and growth factors of matrices. *SIAM J. Matrix Anal. Appl.*, 28(2):446–476 (electronic), 2006.

[Sta99] Richard P. Stanley. *Enumerative combinatorics. Vol. 2*, volume 62 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999. With a foreword by Gian-Carlo Rota and appendix 1 by Sergey Fomin.

[Ste72] Charles Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pages 583–602. Univ. California Press, Berkeley, Calif., 1972.

[SU84] E. Shamir and E. Upfal. Large regular factors in random graphs. In *Convexity and graph theory (Jerusalem, 1981)*, volume 87 of *North-Holland Math. Stud.*, pages 271–282. North-Holland, Amsterdam, 1984.

[Sze78] Endre Szemerédi. Regular partitions of graphs. In *Problèmes combinatoires et* *théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, volume 260 of *Colloq. Internat. CNRS*, pages 399–401. CNRS, Paris, 1978.

[Tal96] Michel Talagrand. A new look at independence. *Ann. Probab.*, 24(1):1–34, 1996.

[Tho85] Andrew Thomason. Pseudo-random graphs. In M. Karoński, editor, *Proceedings of Random Graphs*, number 33, pages 307–331, 1985.

[Tho87] Andrew Thomason. Random graphs, strongly regular graphs and pseudorandom graphs. In *Surveys in combinatorics 1987 (New Cross, 1987)*, volume 123 of *London Math. Soc. Lecture Note Ser.*, pages 173–195. Cambridge Univ. Press, Cambridge, 1987.

[Tra] Linh V. Tran. Local law for eigenvalues of random regular bipartite graphs. Preprint available at arXiv:math/1310.4606.

[TV04] Antonia M Tulino and Sergio Verdú. *Random matrix theory and wireless communications*, volume 1. Now Publishers Inc, 2004.

[TV07] Terence Tao and Van H. Vu. On the singularity probability of random Bernoulli matrices. *J. Amer. Math. Soc.*, 20(3):603–628, 2007.

[TV08] Terence Tao and Van H. Vu. Random matrices: the circular law. *Commun. Contemp. Math.*, 10(2):261–307, 2008.

[TV09] Terence Tao and Van H. Vu. Inverse Littlewood-Offord theorems and the condition number of random discrete matrices. *Ann. of Math. (2)*, 169(2):595–632, 2009.

[TV10a] Terence Tao and Van Vu. Random matrices: the distribution of the smallest singular values. *Geom. Funct. Anal.*, 20(1):260–297, 2010.

[TV10b] Terence Tao and Van H. Vu. Random matrices: universality of ESDs and the circular law. *Ann. Probab.*, 38(5):2023–2065, 2010. With an appendix by Manjunath Krishnapur.

[TV10c] Terence Tao and Van H. Vu. Smooth analysis of the condition number and the least singular value. *Math. Comp.*, 79(272):2333–2352, 2010.

[TVW13] Linh V. Tran, Van H. Vu, and Ke Wang. Sparse random graphs: eigenvalues and eigenvectors. *Random Structures Algorithms*, 42(1):110–134, 2013.

[Ver11] Roman Vershynin. Spectral norm of products of random and deterministic matrices. *Probab. Theory Related Fields*, 150(3-4):471–509, 2011.

[vH] Ramon van Handel. On the spectral norm of gaussian random matrices. Preprint available at arXiv:1502.05003.

[vNG47] John von Neumann and H. H. Goldstine. Numerical inverting of matrices of high order. *Bull. Amer. Math. Soc.*, 53:1021–1099, 1947.

[Vu08] Van H. Vu. Random discrete matrices. In *Horizons of combinatorics*, volume 17 of *Bolyai Soc. Math. Stud.*, pages 257–280. Springer, Berlin, 2008.

[Woo12] Philip Matchett Wood. Universality and the circular law for sparse random matrices. *Ann. Appl. Probab.*, 22(3):1266–1300, 2012.

[Wor80] Nicholas C Wormald. Some problems in the enumeration of labelled graphs. *Bulletin of the Australian Mathematical Society*, 21(01):159–160, 1980.

[Wor99] N. C. Wormald. Models of random regular graphs. In *Surveys in combinatorics, 1999 (Canterbury)*, volume 267 of *London Math. Soc. Lecture Note Ser.*, pages 239–298. Cambridge Univ. Press, Cambridge, 1999.

[YBK88] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah. On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields*, 78(4):509–521, 1988.