Structure of eigenvectors of random regular digraphs

Alexander E. Litvak Anna Lytova Konstantin Tikhomirov
Nicole Tomczak-Jaegermann Pierre Youssef

Abstract

Let d and n be integers satisfying $C \leq d \leq \exp(c\sqrt{\ln n})$ for some universal constants $c, C > 0$, and let $z \in \mathbb{C}$. Denote by M the adjacency matrix of a random d-regular directed graph on n vertices. In this paper, we study the structure of the kernel of submatrices of $M - z\,\mathrm{Id}$, formed by removing a subset of rows. We show that with large probability the kernel consists of two non-intersecting types of vectors, which we call *very steep* and *gradual with many levels.* As a corollary, we show, in particular, that every eigenvector of M, except for constant multiples of $(1, 1, \ldots, 1)$, possesses a weak delocalization property: its level sets have cardinality less than $Cn\ln^2 d/\ln n$. For a large constant d this provides a principally new structural information on eigenvectors, implying that the number of their level sets grows to infinity with n. As a key technical ingredient of our proofs we introduce a decomposition of $\mathbb{C}^n$ into vectors of different degrees of "structuredness," which is an alternative to the decomposition based on the *least common denominator* in the regime when the underlying random matrix is very sparse.

AMS 2010 Classification: primary: 60B20, 15B52, 46B06, 05C80; secondary: 46B09, 60C05
Keywords: Delocalization of eigenvectors, Littlewood–Offord theory, random graphs, random matrices, regular graphs, sparse matrices, structure of the kernel.

Contents

1 Introduction

Fix a large integer n and an integer d in the range $\{3, 4, \dots, n-3\}$. Denote by $\mathcal{M}_{n,d}$ the collection of all $n \times n$ matrices with entries taking values in $\{0, 1\}$ such that in any row and any column there are exactly d ones. Every matrix from this set can be viewed as the adjacency matrix of a d-regular directed graph on n-vertices, where we allow loops but no multiple edges. For a subset $K \subset [n] := \{1, 2, \dots, n\}$ and a matrix B, by B^K we denote the submatrix of B formed by rows $R_i(B)$, $i \in K$. In this paper, we consider the structure of the kernel of random linear operators of the form $(M - z\mathrm{Id})^K$, where M is a random element of $\mathcal{M}_{n,d}$ (with respect to the uniform measure) and z is a fixed complex number.

Our motivation for this study is multifold. We obtain new results regarding delocalization properties of approximate eigenvectors for very sparse random matrices. Apart of being of an independent interest, these results provide new insights into spectral properties of random graphs. Furthermore, as is shown in [39], our results are key to understanding the intermediate singular values of the matrix $M - z\mathrm{Id}$, which in turn are crucial for establishing the limiting spectral distribution of appropriately rescaled adjacency matrices, when the dimension n tends to infinity.

Spectral properties of random graphs, in particular, graphs with predefined degree sequences, have been an object of active research. In the case of d-regular undirected graphs, the magnitude of the second largest eigenvalue as well as the limiting spectral distribution of the adjacency matrix have been considered in various regimes and for different models of randomness (uniform, permutation, and configuration models). In particular, the study of the second largest eigenvalue has been motivated by the well known relation between the magnitude of the spectral gap and the graph expansion properties [1, 19, 32]. We refer, in particular, to [13, 25, 24, 21, 46, 10, 17, 57] and references therein as well as to the survey [32] for more information on spectral expanders. The limiting spectral distribution of an (appropriately rescaled) adjacency matrix of an undirected d-regular graph follows the Kesten–McKay law [34, 43] which, for degree d converging to infinity with n, coincides with the classical semi-circle law [58]. We refer, in particular, to [20, 9, 8] for recent results in this direction.

In the case of directed d-regular graphs, establishing the limiting spectral distribution for constant d presents a major challenge not resolved as of this writing. It is conjectured that the limiting spectral distribution of the appropriately rescaled adjacency matrix follows the *oriented Kesten–McKay law* (see, for example, [11, p. 52]). For $\min(d, n-d) \to \infty$, the limiting distribution has been conjectured to follow the circular law, thus matching (up to rescaling) the standard i.i.d. non-Hermitian models. Very recently, this conjecture has been partially resolved in the uniform and permutation models of randomness under the assumption that the degree d grows with n at least poly-logarithmically [16, 5]. However, the case of very slowly growing d has remained open. This (very sparse) regime is in certain respects fundamentally different as it requires special handling of not only the smallest but also the *intermediate* singular values of the shifted adjacency matrix $M - z\mathrm{Id}$.

In [38], building upon arguments in [14, 15, 37, 36], we have established lower bounds for the smallest singular value of the matrix $M - z\mathrm{Id}$ which work for all d larger than a large absolute constant (with probability estimates depending on d). In this paper, we consider a more technical (and more difficult) aspect of the study by establishing a structural theorem for the kernel of random operators $(M - z\mathrm{Id})^K$.

This paper is an autonomous part in the series of works in which we resolve the conjecture for the limiting spectral distribution for any function $d = d(n)$ growing to infinity with the dimension n. In [39], we use the main result of this paper together with additional probabilistic arguments to derive bounds for intermediate singular values of $M - z\mathrm{Id}$ and, applying the standard argument of Girko

[26], to establish the circular law for the spectrum in the regime $d \to \infty$, resolving the corresponding conjecture (see [16, p. 5]). In order to avoid repetitions, we leave further discussion of Girko's approach and, more generally, the historical overview of the circular law to [39]. As mentioned before, while being very useful in proving the limiting law, the structural theorem is of interest on its own and in the case $K = [n]$ can be viewed as a *delocalization* statement about approximate eigenvectors of very sparse random matrices.

We start with formulating a "soft" version of the main result. In what follows, given a vector $x \in \mathbb{C}^n$, by $x^* = (x_i^*)_i$ we denote the non-increasing rearrangement of $(|x_i|)_i$. We also recall that for an $n \times n$ matrix B and a subset $K \subset [n]$, B^K denotes the $|K| \times n$ matrix with rows $R_i(B)$, $i \in K$.

Theorem 1.1 (Structural theorem). *There are universal constants $c, c', C > 0$ with the following property. Let $n \geq C$, $C \leq d \leq \exp(c\sqrt{\ln n})$, and let M be uniformly distributed on $\mathcal{M}_{n,d}$. Let $z \in \mathbb{C}$ be such that $|z| \leq \sqrt{d}\,\ln d$. Fix $d^{-1/2} \leq a \leq 1$ and any subset $K \subset [n]$ with $0 \leq |K^c| \leq n/d^3$. Set*

$$\rho := \max\Big(n^{-c}, \exp\Big(-\big(n/(1+|K^c|)\big)^{\frac{c\ln\ln d}{\ln d}}\Big)\Big),\ \delta := \frac{C\ln^2 d}{\ln(1/\rho)},\ q := \max(a|K^c|, 1).$$

Then with probability at least $1 - 1/n$ every unit complex vector x such that

$$\|(M - z\,\mathrm{Id})^K\, x\|_2 \leq |K^c|^3\, n^{-6},$$

satisfies one of the following two conditions:

- *(Gradual with many levels) One has* $x_i^* \leq (n/i)^3 x_q^*$ *for all* $i \leq q$,

  $$x_i^* \leq d^3 (n/i)^6 x^*_{\lfloor c'n \rfloor} \qquad \text{for all} \quad q \leq i \leq \lfloor c'n \rfloor,$$

  and
  $$\Big|\Big\{i \leq n :\, |x_i - \lambda| \leq \rho x^*_{\lfloor c'n \rfloor}\Big\}\Big| \leq \delta n \qquad \text{for all} \quad \lambda \in \mathbb{C}.$$

- *(Very steep)* $x_i^* > 0.9(n/i)^3 x_q^*$ *for some* $i \leq q$.

We note that the probability bound in the theorem can be replaced with $1 - n^{-r}$ for any fixed $r \geq 1$ by adjusting the absolute constants. The terms "steep" and "gradual" vectors will be discussed in detail below. In full generality, the theorem will be given at the end of the paper, see Theorem 7.10. The above statement asserts that, given a vector $x \in \mathbb{C}^n$ which is close to the kernel of $(M - z\mathrm{Id})^K$, either the coordinates of x^* decrease fast for small indices (x is "very steep") or, if this is not the case, the vector x is spread (has many non-zero components) and, moreover, for any complex λ very few coordinates of x are concentrated around λ. Note that when $|K^c| = 1$, i.e., when we consider normal vectors to a linear subspace spanned by $n-1$ matrix rows, the second assertion never holds (since $q = 1$), and the theorem says that the normal vectors are all spread and have many levels. Moreover, for $K = \{1, 2, \ldots, n\}$, combining the theorem with a simple covering argument, we obtain the following delocalization property.

Corollary 1.2 (Delocalization properties of eigenvectors). *There are universal constants $c, C > 0$ such that the following holds. Let n, d and M be as in Theorem 1.1. Then with probability at least $1 - 1/n$ any eigenvector x of M, which is not parallel to $(1, 1, \ldots, 1)$, satisfies $x_i^* \leq d^3 (n/i)^6 x^*_{\lfloor cn \rfloor}$ for all $1 \leq i \leq \lfloor cn \rfloor$, and, moreover,*

$$\forall \lambda \in \mathbb{C} \qquad \Big|\Big\{i \leq n :\, |x_i - \lambda| \leq n^{-c} x^*_{\lfloor cn \rfloor}\Big\}\Big| \leq Cn\ln^2 d/\ln n.$$

For completeness, we give a proof of Corollary 1.2 at the end of the paper. We would like to notice that delocalization properties of eigenvectors for various models have been a focus of active research. In the case of sparse matrices we refer to [12] for eigenvector statistics for Erdős–Rényi graphs and to

[2, 3, 4, 18] for delocalization properties of eigenvectors (and almost eigenvectors) of undirected regular graphs. In the non-Hermitian setting (relevant to our present work) we refer to [52, 53]. The term *delocalization* usually refers to upper bounds on the ℓ_∞-norm of a random vector or, more generally, upper bounds for the scalar product with a fixed unit vector. The delocalization statement provided by Corollary 1.2, is closer to the concept of *no-gaps delocalization*, introduced in [53], which bounds the ℓ_2-mass of the vector supported on every subset of coordinates of a given size. At the same time, Corollary 1.2 not only provides lower bounds for the order statistics of eigenvectors but also measures cardinalities of sets of almost equal coordinates, thus giving an additional structural information (very important in our context). Corollary 1.2, to our best knowledge, is the first statement which provides quantitative information on the delocalization for non-Hermitian random matrices with a *constant* number of non-zero elements in rows/columns.

A fundamental feature of the main theorem is that it provides information on the kernel of the matrix for large constant d, with the "unstructuredness" measured in terms of the dimension n. Here, "unstructuredness" refers to the allowed number of approximately equal components of a vector. For example, we show that for $M^{\{2,\dots,n\}}$ with large probability any vector in the kernel can have at most $O(n/\ln n)$ equal components. We expect that the theorem and the argument used in its proof will turn to be useful in the study of the spectrum of random directed graphs of constant degree. In fact, combined with some known arguments, our result implies that the random adjacency matrix M has rank *at least* $n-1$ with probability going to one with n (d being a large constant) [40].

Theorem 1.1 can be interpreted as follows: with probability very close to one we have

$$\inf_{x\in T} \|(M - z\mathrm{Id})^K x\|_2 > 0,$$

where the infimum is taken over the set $T \subset \mathbb{C}^n$ of all non-zero vectors, which are neither "very steep" and not "gradual with many levels." Estimates of this type fall into a large body of research dealing with matrix singularity and structural properties of null vectors for various models of randomness. A possible strategy in estimating the infimum $\inf\limits_{x\in S} \|Bx\|_2$ (for a random matrix B and a subset S, say, the unit Euclidean sphere) consists in representing S as a union of subsets $\bigcup_\alpha S_\alpha$ grouping together vectors with a similar structure, and then combining bounds for $\inf\limits_{x\in S_\alpha} \|Bx\|_2$. In turn, each of the infima is bounded using the structural information about vectors in S_α and may involve, depending on the problem, a discretization of the subset (i.e., a version of a covering argument). A very incomplete list of works involving this approach in the study of square non-Hermitian or "almost square" matrices is [33, 54, 58] (singularity of random Bernoulli matrices), [41, 49, 50, 47] (matrices with i.i.d. entries with a tail decay condition), [55, 56, 27, 6, 7] (sparse matrices with i.i.d. entries, see also [42] for the non-i.i.d. case), [15, 37, 16, 5] (adjacency matrices of directed d-regular graphs). For a detailed exposition of this method we refer to [51] and [59]. The decomposition into subsets differs significantly, depending on the randomness model. In particular, estimating the singularity probability for Bernoulli matrices in [33, 54, 58] involved defining the *combinatorial dimension* of certain discrete vectors. Further, the idea from [41] of splitting the Euclidean sphere into sets of "close to sparse" and "far from sparse" vectors was developed in [49, 50], where the notion of *compressible* and *incompressible* vectors appeared. In [49, 50], building upon earlier works on the Littlewood–Offord theory (see, in particular, [56]), the concept of the *least common denominator* (LCD) of a vector was introduced, and the Euclidean sphere was partitioned into subsets according to the magnitude of the LCD. Further, in works [16, 5, 38, 6, 7] dealing with adjacency matrices of sparse directed random graphs, the crucial structural property of a vector was statistics of "jumps" in its non-increasing rearrangement, i.e., the magnitude of ratios of the form x_i^*/x_{Li}^*, where $i \le n$ and $L > 0$ is a scaling factor. In works [6, 7], this analysis of jumps in the rearrangement was combined with the LCD-based approach of [49, 50].

Despite the progress in this research direction in the past years, an efficient estimate of the smallest singular value and, more generally, of quantities of the form $\inf\limits_{x\in T} \|Bx\|_2$ for a very sparse random matrix B (with a constant average number of non-zero elements in a row/column) seems to require essentially new arguments. In this work, we propose such a new argument for 0/1 random matrices with prescribed row/column sums. We believe that our approach can be extended to more general sparse models.

A crucial new ingredient of our paper is the concept of the *ℓ-decomposition* of a vector which is a partition of $[n]$ into subsets encoding useful structural properties of a complex n-dimensional vector x. We combine the ℓ-decomposition with new tensorization arguments (which allow to pass from individual small ball probability estimates for $\langle R_i(M - z\,\mathrm{Id}), x\rangle$ to the matrix-vector product $(M - z\,\mathrm{Id})^K x)$ and a discretization (covering) procedure to get a characterization of *gradual* vectors in the kernel. On the other hand, the *steep* vectors are treated by a combination of covering arguments and a procedure utilizing expansion properties of the graph, thus augmenting the approach in [38]. In the remainder of the introduction, we will discuss in detail the three main features of the proof: steep and gradual vectors, ℓ-decomposition, and tensorization.

Steep, almost constant, and gradual vectors. The notions of steep, almost constant, and gradual vectors appeared in [38] in the context of bounding the smallest singular value of the shifted adjacency matrix. Naturally, these notions play an important role in the present paper as well. For technical reasons, we slightly modified the definitions, compared to [38].

A full description of the class of steep vectors in $\mathbb{C}^n$ is provided in Subsection 4.1. Since the precise formulas are long and involve many parameters, we omit them in the introduction. We just loosely describe this class as the collection of vectors $x \in \mathbb{C}^n$ such that for some indices $1 \le i < j \ll n$, the ratio x_i^*/x_j^* is *very large* compared to the ratio j/i. The basic example of a steep vector is $(1, 1, \dots, 1, 0, 0, \dots, 0)$, with less than cn ones (for a small constant $c > 0$). At the same time, the steep vectors are not necessarily close to sparse in the Euclidean metric – in particular, the steep vectors cannot be identified with *compressible* vectors introduced in [41, 49, 50].

The second class of vectors which we call *almost constant*, is much easier to describe explicitly – those are all the vectors $x \in \mathbb{C}^n$ such that

$$\exists \lambda \in \mathbb{C} \qquad |\{i \le n \,:\, |x_i - \lambda| \le \theta\, x_{cn}^*\}| > n - cn,$$

where θ is a negative constant power of the degree d and $c > 0$ is a small universal constant. Vectors which are neither steep nor almost constant are called *gradual*. Thus, a gradual vector has many pairs of distinct coordinates and controlled ratios x_i^*/x_j^* for all indices $1 \le i < j \ll n$.

In our study of the kernel of the random operator $(M - z\,\mathrm{Id})^K$, we consider its intersection with the classes of steep and almost constant vectors in Section 4 and with gradual vectors – in Sections 5–7. Theorem 1.1 combines the information about the intersections.

For large enough K^c and small d, simultaneous existence of very steep and gradual (with many levels) vectors in the kernel of the matrix M^K is an objective fact. We can consider the following informal argument. For an integer $p \ge 1$, the kernel of $M^{\{p+1,\dots,n\}}$ contains $\ker M^{\{2,\dots,n\}}$, which, in view of Theorem 1.1 and the above remark, typically consists of gradual vectors with many levels. At the same time, the columns of M, $C_i(M)$, are "locally" almost independent, in the sense that for every small subset $Q \subset [n]$, the joint distribution of $C_i(M)$, $i \in Q$, is "close" to the joint distribution of independent vectors uniform on the set $\{y \in \{0,1\}^n \,:\, |\mathrm{supp} y| = d\}$ (in order not to expand the paper we prefer not to discuss quantitative aspects of this observation). Thus, for fixed integers $p \gg d$ and $r \ll n$, we have

$$\begin{aligned}
&\mathbb{P}\big\{\exists i \le r \,:\, \mathrm{supp}\, C_i(M) \cap \{p+1, \dots, n\} = \emptyset\big\} \\
&\quad \approx 1 - \prod_{i=1}^{r} \mathbb{P}\big\{\mathrm{supp}\, C_i(M) \cap \{p+1, \dots, n\} \neq \emptyset\big\} \\
&\quad = 1 - \bigg(1 - \binom{p}{d}\binom{n}{d}^{-1}\bigg)^r \approx 1 - \exp\big(-(p/n)^d r\big).
\end{aligned}$$

Accordingly, when $r \gg (n/p)^d$, with large probability there exists a null-column in $M^{\{p+1,\dots,n\}}$, so that a typical realization of $M^{\{p+1,\dots,n\}}$ contains a coordinate vector in its kernel, which is "very steep."

The analysis of the null steep and almost constant vectors, which occupies Section 4, is a development of an argument from [38] with some important technical additions. It combines deterministic estimates for Mx (assuming certain expansion properties of the underlying graph) with covering arguments for non-constant vectors with large support. For almost constant vectors x, a satisfactory bound for the Lévy concentration function of the product Mx is generally impossible. For example, if $x = (1, 1, \ldots, 1)$ then $Mx = (d, d, \ldots, d)$ *deterministically.* A key step in bounding the Euclidean norm of $(M - z\,\mathrm{Id})x$ (or, more generally, $(M - z\,\mathrm{Id})^K x$) from below for almost constant vectors is a non-probabilistic argument which utilizes d-regularity (see Lemma 4.10). A related method is used for steep vectors if the jump occurs at the beginning of the non-increasing rearrangement (Lemma 4.4). For the remaining vectors x, anti-concentration estimates for individual row-vector products are tensorized to obtain individual estimates for Mx, which are combined with a covering argument, using specially constructed nets (see Subsection 4.3).

The least common denominator (LCD), and the ℓ-decomposition. The correspondence between small ball probability and arithmetic properties of the vector of coefficients goes back to Halász [29, 30], and was refined in the work [54] on the singularity probability, before being developed for the least singular value problem in [48, 56]. Then in [49, 50], the LCD with parameters $\gamma, \alpha > 0$ of a vector $x \in \mathbb{R}^n$ was defined as

$$LCD(x) := \inf\big\{\theta > 0,\ \mathrm{dist}(\theta x, \mathbb{Z}^n) < \min(\gamma\theta\|x\|_2, \alpha)\big\}.$$

It was used there with γ being a small positive constant and α being a small constant multiple of $\sqrt{n}$. Thus, LCD encapsulates information on the distance of a rescaled vector to the integer lattice. The fundamental correspondence between the magnitude of the LCD of a vector x and the small ball probability for the random sum $\sum_{i=1}^n x_i\xi_i$ (for independent sufficiently "spread" random variables ξ_i) was the key ingredient employed in [49, 50]. Specifically, it was shown in [49] that a normal vector to the hyperplane spanned by $n-1$ columns typically has LCD which is exponentially large in dimension. This fact was then applied to estimate the small ball probability for the least singular value of the random matrix. However, in the sparse regime this approach presents certain challenges. Note that the definition of the LCD does not allow to distinguish a vector having a (small) proportion of coordinates of the same value from a vector with almost all components distinct. For example, the vectors $x^1 := (1/n, 2/n, 3/n, \ldots, n/n)$ and $x^2 := (0, 0, \ldots, 0, 0, 0.01 + 1/n, 0.01 + 2/n, \ldots, 0.01 + 0.99n/n)$ (with the first $0.01n$ components equal to zero) would have comparable LCDs, whereas behaviour of the corresponding scalar products with a row of a random d-regular matrix is completely different. Specifically, if $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ is the random 0/1 vector uniformly distributed on

$$\{y \in \{0,1\}^n :\ |\mathrm{supp}\, y| = d\}$$

then for any number $r \in \mathbb{R}$ the scalar product $\langle \xi, x^1\rangle$ is equal to r with probability $O(1/n)$ while $\langle \xi, x^2\rangle = 0$ with probability *at least* 0.01^d. Indeed, it is the absence (or presence) of large constant blocks of coordinates which becomes crucial in the sparse regime.

The notion of the ℓ-decomposition developed in this paper is in particular designed to deal with the above issue. More importantly, the ℓ-decomposition carries very detailed information about the structure of a vector. This information turns out extremely useful in estimating cardinalities of coverings as well as in bounding small ball probabilities.

The precise definition of the ℓ-decomposition is rather involved and includes an iterative procedure for constructing a partition of $[n]$ associated to a vector. We skip the technical details in the introduction (see Subsection 5.2 for the actual construction procedure) and describe the ℓ-decomposition of a vector y in the lattice $\frac{1}{k}\big(\mathbb{Z}^2\big)^n$ (where k is a large integer) as a partition of $[n]$ into non-empty subsets $(\mathcal{L}^{(q)})_{q=1}^m$ (called *ℓ-parts*) which satisfy, among some other conditions, the following: For any $q \le m$ and any numbers $a, b \in \frac{1}{k}\mathbb{Z}^2$, setting

$$J_{a,q} := \{i \in \mathcal{L}^{(q)} :\ y_i = a\} \quad \text{and} \quad J_{b,q} := \{i \in \mathcal{L}^{(q)} :\ y_i = b\},$$

one has that either one of the sets is empty or $|J_{a,q}| \leq 4|J_{b,q}|$. In other words, considering the levels (blocks of coordinates having the same value) of the vector y restricted to $\mathcal{L}^{(q)}$, all the blocks have approximately the same cardinality. If the vector y is real and non-increasing, its restriction to $\mathcal{L}^{(q)}$ can be viewed as a ladder or staircase, where all the stairs are of about the same size, whereas the gaps between the stairs are allowed to differ significantly. The number of "stairs" within the ℓ-part is called *the height* of the ℓ-part. Note that the height is *not* determined by the magnitude of y_i, $i \in \mathcal{L}^{(q)}$, but instead by the number of levels in $\mathcal{L}^{(q)}$. The position of the ℓ-parts and their heights provide essential information on anti-concentration properties of the vector, specifically, on anti-concentration of the scalar product with a row of our random matrix. The information contained in the ℓ-decomposition allows us to compute *conditional* small ball probability with imposed restrictions on the distribution of the row. This becomes useful when studying anti-concentration for the matrix-vector product $M^K y$ (see Section 6). Observe that the ℓ-decomposition is defined for a discrete subset of $\mathbb{C}^n$; in fact, given a gradual vector x we construct its approximations by vectors in $\frac{1}{k}\left(\mathbb{Z}^2\right)^n$ for various values of k (*k-approximations*) and consider the ℓ-decomposition of each of the approximations.

At a high level, the way we apply the ℓ-decomposition to the original problem is similar to the way the least common denominator was used in [49]. In [49], the set of unit incompressible vectors was split into subsets according to the magnitude of the LCD. Then, with the help of a covering argument combined with small ball probability estimates, the subsets of vectors with small (subexponential in dimension) LCD were excluded from the set of potentially null vectors of the matrix, leaving only those with very large LCD.

In our setting, we partition the collection of gradual vectors according to properties of the ℓ-decompositions of their lattice approximations. Using a combination of small ball probability estimates and coverings, we exclude those gradual vectors with "not that many" levels, leaving only those satisfying the first condition in Theorem 1.1. The partitioning scheme is rather complicated because of the "multidimensional" nature of the ℓ-decomposition, i.e., due to the necessity to take into consideration a set of parameters rather than a single number. The crucial notions used in the partitioning are those of *regular* and *spread* ℓ-parts. Given a vector $y \in \frac{1}{k}\left(\mathbb{Z}^2\right)^n$ with the ℓ-decomposition $(\mathcal{L}^{(q)})_{q=1}^m$, we say that the ℓ-part $\mathcal{L}^{(q)}$ is *spread* if it contains at least two distinct levels and for every pair of coordinates y_i, y_j $(i, j \in \mathcal{L}^{(q)},\ y_i \neq y_j)$, we have $|y_i - y_j| \geq d/k$. In the case of a vector with ordered real components, we may say that the gaps between the "stairs" within the spread ℓ-parts are d times larger than their absolute minimum $1/k$. The ℓ-parts which are not spread are called *regular*. Naturally, the matrix-vector products $M^K y$, with y having spread ℓ-parts of large cardinality, enjoy relatively better anti-concentration properties. At a more detailed level, the following procedure is applied:

- We isolate the set of gradual vectors, whose k-approximation (for some k within a specific range) has large spread ℓ-parts. The sets of vectors are denoted by $\mathcal{K}_u$ in the text (see Subsection 5.3), where $u \geq 5$ is related to k by $k = d^u$.
- For every vector from $\bigcup \mathcal{K}_u$, its k-approximation y has the structure which guarantees strong bounds for the small ball probability

$$\mathbb{P}\{\|(M - z\,\mathrm{Id})^K\, y\|_2 \leq t\}.$$

  These bounds, combined with a covering procedure for the vectors in $\mathcal{K}_u$, implies that with probability close to one no vector in $\bigcup \mathcal{K}_u$ is approximately a null vector.
- Thus, it remains to deal with vectors in the complement $\mathcal{S} \setminus \bigcup \mathcal{K}_u$, where $\mathcal{S}$ denotes the set of gradual vectors in $\mathbb{C}^n$ (we emphasize again that the union is taken over u within a specific range determined by n, d, and the cardinality of the set K). We show (see Subsection 5.3) that the complement $\mathcal{S} \setminus \bigcup \mathcal{K}_u$ consists of gradual vectors, whose k-approximation (for a specially chosen k) has ℓ-parts of very large heights (see the definition of sets $\mathcal{P}_v$ and Propositon 5.6). At an elementary level, the coordinates of those k-approximations y take many distinct values.

Naturally, this property provides a fine small ball probability estimate for $(M - z\,\mathrm{Id})^K\, y$, however, for a different reason than in the case of large spread ℓ-parts.

- As the final step, we make the following observation: the set of gradual vectors from $\mathcal{P}_v$, which have relatively large constant blocks of coordinates, has much smaller complexity than the entire $\mathcal{P}_v$. In a sense, it is possible to construct a net on the set of such vectors with cardinality balanced by individual small ball probabilities, thus excluding the set from the collection of potentially null vectors. The remaining set – gradual vectors without large constant blocks – are the "gradual vectors with many levels" from the first assertion of Theorem 1.1.

The procedure described above occupies Section 5 and a considerable part of Section 7. Passing from small ball probability bounds for scalar products with individual matrix rows to the entire matrix-vector product is a crucial step, with technical complexities arising because of the lack of independence between the rows. This *tensorization* part of the argument spans Section 6 and partially continues into Section 7. We would like to describe it in more detail.

Tensorization. Our goal is to represent the small ball probability

$$\mathbb{P}\{\|(M - z\,\mathrm{Id})^K y\|_2 \le t\}$$

for a given k-vector y (i.e., $y \in \frac{1}{k}\big(\mathbb{Z}^2\big)^n$) and $t > 0$ in terms of the structure of y, i.e, in terms of the ℓ-decomposition with respect to y. A bound is obtained via a series of reduction steps. At each step, we replace our random model or quantities with objects that are simpler to analyze.

As the first step, given a vector y with its ℓ-decomposition $(\mathcal{L}^{(q)})_{q=1}^m$, we condition on realizations of $\sum_{j\in\mathcal{L}^{(q)}} C_j(M)$, $q \le m$ (recall that $C_i(M)$ denote the i-th column of M). Specifically, we define a collection of $n \times m$ matrices Q with integer valued entries and study the small ball probability within the event

$$\mathcal{E}_Q = \Big\{ \sum_{j\in\mathcal{L}^{(q)}} C_j(M) = C_q(Q), \ \ q \le m \Big\}$$

for some fixed matrix Q. In particular, this forces the distributions of the columns of M from different ℓ-parts to be independent. Moreover, using the expansion properties of the underlying graph, we impose additional assumptions on Q thus getting statistics of the number of non-zero entries of the matrix M "restricted" to each ℓ-part. The conditioning on $\mathcal{E}_Q$ is described at the beginning of Section 6, however, additional structural assumptions on matrices Q are introduced later in Section 7.

The next – crucial – step consists in replacing the given random model with another one having independent components. More precisely, we define a random n-dimensional vector Z with *jointly independent* components such that, conditioned on a certain event, its distribution matches the conditional distribution of My given $\mathcal{E}_Q$. In its essence, every component Z_i is a sum of independent random variables, where each variable indicates the level of y "hit" by a non-zero entry in the i-th row of M. We connect the (conditional) distribution of Z with the distribution of My by introducing an intermediate random model involving bipartite multigraphs. We start by showing that the distribution of the adjacency matrix of that multigraph, conditioned on the event that the graph is simple, coincides with the distribution of M. Results of this type are known in the random graph literature. In our setting we apply a result from [44] to get the correspondence. In its turn, the distribution of the adjacency matrix of the multigraph is directly related to the distribution of Z conditioned on a certain event that can be viewed as a sort of "d-regularity" property.

Anti-concentration estimates for My thus can be obtained by multiplying bounds for individual components Z_i. Those, in turn, can be written in terms of the ℓ-decomposition of y and the structure of the i-th row of the matrix Q. The functions which encapsulate this information are called *the small ball probability estimators*. As the main result of Section 6, we estimate $\mathbb{P}\{\|(M - z\,\mathrm{Id})^K y\|_2 \le t\}$ in terms of the product $\prod_{i=1}^n \mathrm{SB}_i$, where SB_i is the small ball probability estimator for the i-th row/i-th component (see Theorem 6.1). Computing the product $\prod_{i=1}^n \mathrm{SB}_i$ is not straightforward as it involves analysis of both the ℓ-decomposition of y and the matrix Q. In the first part of Section 7, we introduce other *estimators* related to SB_i, which are easier to study. Once we obtain an explicit upper bound for the small ball probability, we combine it with covering arguments, which were briefly mentioned above.

The arguments in the paper are largely self-contained, although we employ several external results. This includes (mostly standard) bounds on concentration and anti-concentration of the sum of independent variables; certain expansion properties of the underlying random d-regular digraph; an upper bound on the second largest singular value of the adjacency matrix; some estimates regarding the configuration model for random graphs with predefined degree sequences.

2 Preliminaries

By *universal* or *absolute* constants we always mean numbers independent of all involved parameters, in particular independent of d and n. Given positive integers $\ell < k$ we denote sets $\{1, 2, \ldots, \ell\}$ and $\{\ell, \ell+1, \ldots, k\}$ by $[\ell]$ and $[\ell, k]$ correspondingly. Having two functions f and g we write $f \approx g$ if there are two absolute positive constants c and C such that $cf \leq g \leq Cf$. Given $z \in \mathbb{C}$, we denote by $\operatorname{Re} z$ (resp., $\operatorname{Im} z$) the real (resp., imaginary) part of z. We define a lexicographical order on $\mathbb{C}$ in the following way. Given $x, y \in \mathbb{C}$, we have $x \geq y$ if either $\operatorname{Re} x > \operatorname{Re} y$ or $\operatorname{Re} x = \operatorname{Re} y$ and $\operatorname{Im} x \geq \operatorname{Im} y$. The lexicographical ordering will be useful when defining maximum over a finite subset of the complex plane.

By Id we denote the identity $n \times n$ matrix, we use $\mathbf{1}$ for a vector of ones. For $I \subset [n]$, by P_I we denote the orthogonal projection on the coordinate subspace $\mathbb{R}^I$ (or $\mathbb{C}^I$), and denote the complement of I inside $[n]$ by I^c (n is always clear from the context). Given a vector $x = (x_i)_{i=1}^n \in \mathbb{C}^n$, we denote $x^\dagger = \bar{x} = (\bar{x}_i)_{i=1}^n$, where $\bar{z}$ is the complex conjugate of $z \in \mathbb{C}$, and by $(x_i^*)_{i=1}^n$ we denote the non-increasing rearrangement of the sequence $(|x_i|)_{i=1}^n$. We use $\langle \cdot, \cdot \rangle$ for the standard inner product on $\mathbb{C}^n$, that is $\langle x, y \rangle = \sum_{i=1}^n x_i \bar{y}_i$. Further, we write $\|x\|_\infty = \max_i |x_i|$ for the ℓ_∞-norm of x. For a list of notation related to matrices and graphs we refer to the beginning of Section 3.

Working with classes of vectors, we often consider the Minkowski sum for two subsets V and W of $\mathbb{C}^n$, which is defined as
$$V + W = \{v + w \,:\, v \in V,\, w \in W\}.$$

To obtain probability bounds we often consider certain relations between sets and use the following simple claims to estimate their probabilities. Let A, B be sets, and $R \subset A \times B$ be a relation. Given $a \in A$ and $b \in B$, the image of a and preimage of b are defined by
$$R(a) = \{y \in B \,:\, (a, y) \in R\} \quad \text{ and } \quad R^{-1}(b) = \{x \in A \,:\, (x, b) \in R\}.$$
We also set $R(A) = \cup_{a \in A} R(a)$. We use the following standard estimate (see e.g. Claim 2.1 in [37]).

Claim 2.1. *Let $s, t > 0$. Let R be a relation between two finite sets A and B such that for every $a \in A$ and every $b \in B$ one has $|R(a)| \geq s$ and $|R^{-1}(b)| \leq t$. Then $s|A| \leq t|B|$.*

2.1 Anti-concentration

For a random vector X distributed over a real or complex inner product space E, its *Lévy concentration function* $\mathcal{Q}(X, t)$ is defined as
$$\mathcal{Q}(X, t) := \sup_{\lambda \in E} \mathbb{P}\big\{\|X - \lambda\|_2 \leq t\big\}, \;\; t > 0.$$
In particular, if X is a complex random variable then
$$\mathcal{Q}(X, t) = \sup_{\lambda \in \mathbb{C}} \mathbb{P}\big\{|X - \lambda| \leq t\big\}.$$

Dealing with the Lévy concentration function of a complex random variable, we often identify $\mathbb{C}$ with $\mathbb{R}^2$, which allows us to apply Propositions 2.2 and 2.3 formulated below for $E = \mathbb{R}^2$.

Upper bounds on the concentration function of a sum of independent random variables is a standard subject, with many results available in the literature. In our setting, we primarily deal with complex-valued random variables, which in some situations require more delicate arguments. In this subsection, we combine classical estimates of the concentration function with some (not complicated) computations for vector-valued variables.

We will use (a particular version of) a theorem of Esseen [23] for sums of random vectors (Corollary 1 of Theorem 6.1 in [23] applied with $\rho_i = t_0$, $\rho = t$).

Proposition 2.2 (Esseen)**.** *Let $m \geq 1$ and $\xi_1, \ldots, \xi_m$ be independent random vectors in $\mathbb{R}^2$. Then for any $t \geq t_0 > 0$ one has*

$$\mathcal{Q}\Big(\sum_{i=1}^{m} \xi_i, t\Big) \leq \frac{C_{2.2}(t/t_0)^2}{\sqrt{m - \sum_{i=1}^{m} \mathcal{Q}(\xi_i, t_0)}},$$

where $C_{2.2} > 0$ is a universal constant. In particular, if $\alpha \geq \max_{i \leq m} \mathcal{Q}(\xi_i, t_0)$ then

$$\mathcal{Q}\Big(\sum_{i=1}^{m} \xi_i, t_0\Big) \leq \frac{C_{2.2}}{\sqrt{m(1-\alpha)}}.$$

We will also need a result of Miroshnikov [45], which extends estimates on the concentration function due to Kesten [35] to the multi-dimensional setting. We state below the two dimensional version of the Corollary following Theorem 1 in [45] (note that the letter E in that paper stands for the two-dimensional cube B_∞^2, while we deal with the unit disc B_2^2, so that $B_2^2 \subset E = B_\infty^2 \subset \sqrt{2} B_2^2$).

Proposition 2.3 (Miroshnikov)**.** *Let $m \geq 1$ and $\xi_1, \ldots, \xi_m$ be independent random vectors in $\mathbb{R}^2$. Let $t_0 > 0$ be such that $\max_{i \leq m} \mathcal{Q}(\xi_i, t_0) \leq 1/2$. Then for any $t \geq t_0$ one has*

$$\mathcal{Q}\Big(\sum_{i=1}^{m} \xi_i, t\Big) \leq \frac{C\, t}{t_0 \sqrt{m}} \max_{i \leq m} \mathcal{Q}(\xi_i, \sqrt{2}\, t),$$

where C is a positive universal constant.

In general, the factor $1/\sqrt{m}$ in the above estimates is the best possible and is attained for example on ξ_i's with $\operatorname{Re} \xi_i = \operatorname{Im} \xi_i$ being Bernoulli random variables. But if for example for every $i \leq m$, $\operatorname{Re} \xi_i$ and $\operatorname{Im} \xi_i$ are independent Bernoulli random variables, or if ξ_i is uniformly distributed over the unit square, then Theorem 2 from [35] implies respectively bounds C/m and Ct^2/m for all $t \in (0, 1/2]$. Some other cases when the factor $1/\sqrt{m}$ can be improved to $1/m$ were considered in [22] (for distributions satisfying a certain symmetry condition) and in [23] (for, in a sense, well spread distributions). We will need the following statement which is known to specialists. We provide its proof for the sake of completeness at the end of this section.

Proposition 2.4. *Let $m \geq 1$ and $\xi_1, \ldots, \xi_m$ be independent random vectors in $\mathbb{R}^2$ with densities bounded by 1. Then the density of $\xi_1 + \ldots + \xi_m$ is bounded by C/m, where C is a universal constant.*

Our next proposition is another case where the factor $1/\sqrt{m}$ can be improved.

Proposition 2.5. *Let $u, \varepsilon > 0$, $m \geq 1$, and let $\xi_1, \xi_2, \ldots, \xi_m$ be i.i.d. discrete complex random variables taking values on an ε-separated (in the Euclidean metric) subset of the complex plane satisfying $\sup_{a \in \mathbb{C}} \mathbb{P}\{\xi_j = a\} \leq u$. Then for any $t > 0$ one has*

$$\mathcal{Q}\Big(\sum_{j=1}^{m} \xi_j, t\Big) \leq C \max\big(ut^2/(m\varepsilon^2), u\big),$$

where C is a positive absolute constant.

Proof. As above, in this proof it will be convenient to identify $\mathbb{C}$ with $\mathbb{R}^2$ and to present complex random variables as two dimensional real random vectors. We also denote the Euclidean unit ball (centered at 0) on $\mathbb{R}^2$ by B. Let η_j, $j \le m$, be i.i.d. random vectors in $\mathbb{R}^2$ uniformly distributed on $(\varepsilon/2)B$ and jointly independent with ξ_j's. Note that by standard concentration estimates (e.g., one can apply Hoeffding's inequality [31] for the first and second coordinates), we have for a large enough absolute positive constant C,

$$q := \mathbb{P}\Big\{\Big|\sum_{j=1}^m \eta_j\Big| \le C\sqrt{m}\varepsilon\Big\} \ge 1/2.$$

Define smoothed i.i.d. random variables ξ_j', $j \le m$, by setting $\xi_j' := \xi_j + \eta_j$. Since interiors of the discs of radius $\varepsilon/2$ centered at atoms of ξ_j's are disjoint, we get that the densities of ξ_j', $j \le m$, are uniformly bounded by $4u/(\pi\varepsilon^2)$. By independence of η_j and ξ_j we observe that for every $t > 0$,

$$\begin{aligned}\mathcal{Q}\Big(\sum_{j=1}^m \xi_j, t\Big) &= \max_{a\in\mathbb{C}} \ \frac{1}{q}\ \mathbb{P}\Big\{\Big|\sum_{j=1}^m \xi_j - a\Big| \le t \ \text{ and } \Big|\sum_{j=1}^m \eta_j\Big| \le C\sqrt{m}\varepsilon\Big\}\\ &\le 2\mathcal{Q}\Big(\sum_{j=1}^m \xi_j', t + C\sqrt{m}\varepsilon\Big) = 2\mathcal{Q}\Big(\sum_{j=1}^m \xi_j'', \tau\Big),\end{aligned}$$

where $\xi_j'' := (2/\varepsilon)\xi_j'\sqrt{u/\pi}$ and $\tau = (2/\varepsilon)t\sqrt{u/\pi} + 2C\sqrt{um/\pi}$. Note that the densities of ξ_j'', $j \le m$, are uniformly bounded by 1. Thus Proposition 2.4 implies the desired result. □

To prove Proposition 2.4 we need the following lemma (which, in a sense, similar to the proof of Theorem 1 in [22]). In this lemma $\langle\cdot,\cdot\rangle$ denotes the canonical inner product on $\mathbb{R}^2$.

Lemma 2.6. *Let p be a probability density on $\mathbb{R}^2$ bounded by 1. Let $f = \hat{p}$, that is*

$$f(x) = \hat{p}(x) = \int_{\mathbb{R}^2} \exp(-2\pi i \langle x, y\rangle) p(y) dy.$$

Then for every $q \ge 2$ one has $\int_{\mathbb{R}^2} |f(x)|^q dx \le 47/q$.

Proof. Denote $\tilde{p}(x) = p(-x)$, $P(x) := p * \tilde{p}$, where $*$ denotes the convolution, and $F(x) = |f(x)|^2$. Then,

$$F = f \cdot \bar{f} = \hat{P}.$$

Observe that the function P satisfies $P(x) \le 1$, $P(x) = P(-x)$ for every $x \in \mathbb{R}^2$, and $\int_{\mathbb{R}^2} P(x)dx = 1$. Therefore, for every $x \in \mathbb{R}^2$,

$$F(x) = \int_{\mathbb{R}^2} \cos(2\pi\langle x,y\rangle)P(y)dy = \int_{\mathbb{R}^2}(1 - 2\sin^2(\pi\langle x,y\rangle)P(y)dy = 1 - 2\int_{\mathbb{R}^2} \sin^2(\pi\langle x,y\rangle)P(y)dy.$$

Consider the sets

$$A_\delta := \{x \in \mathbb{R}^2 \ : \ F(x) \ge 1 - \delta^2\}$$

for $\delta \in (0, 1/2]$. Note that for every integer $k \ge 1$ one has $k^2 \sin^2 t \ge \sin^2(kt)$. Given $\delta \in (0, 1/2]$, let $k = \lfloor 1/(2\delta)\rfloor$. Then for every $x \in A_\delta$ we have

$$F(kx) \ge 1 - 2k^2 \int_{\mathbb{R}^2} \sin^2(\pi\langle x,y\rangle)P(y)dy = 1 - k^2(1 - F(x)) \ge 1 - (k\delta)^2 \ge 3/4,$$

that is on the set $kA_\delta = \{kx \ : \ x \in A_\delta\}$ we have $F \ge 3/4$. On the other hand, by the Plancherel theorem we have

$$\int_{\mathbb{R}^2} F(x)dx = \int_{\mathbb{R}^2} |f(x)|^2 dx = \int_{\mathbb{R}^2} p^2(x)dx \le \int_{\mathbb{R}^2} p(x)dx = 1.$$

This implies $|kA_\delta| \le 4/3$, hence $|A_\delta| \le 4/(3k^2) \le 64\delta^2/3$, in particular, $|A_{1/2}| \le 4/3$. Finally we estimate
$$\int_{\mathbb{R}^2} |f(x)|^q dx = \int_{\mathbb{R}^2} (F(x))^{q/2} dx.$$
Then for $q \ge 2$ we have
$$I_1 := \int_{A_{1/2}^c} (F(x))^{q/2} dx \le (3/4)^{q/2-1} \int_{\mathbb{R}^2} F(x) dx \le (3/4)^{q/2-1} \le 3/q,$$
and
$$\begin{aligned} I_2 &:= \int_{A_{1/2}} (F(x))^{q/2} dx = \int_0^1 (q/2) s^{q/2-1} |\{F \ge \max(s, 3/4)\}| ds \\ &= \int_0^{3/4} |\{F \ge 3/4\}| ds^{q/2} + \int_{3/4}^1 (q/2) s^{q/2-1} |\{F \ge s\}| ds \\ &= (3/4)^{q/2} |A_{1/2}| + \int_{3/4}^1 (q/2) s^{q/2-1} |A_{\sqrt{1-s}}|\, ds \le 9/(16q) + (64/3) \int_{3/4}^1 (q/2) s^{q/2-1} (1-s) ds \end{aligned}$$
Using integration by parts, we have
$$\int_{3/4}^1 (q/2) s^{q/2-1} (1-s)\, ds \le \int_0^1 (1-s)\, ds^{q/2} = \int_0^1 s^{q/2}\, ds = \frac{2}{q+2}.$$
Therefore,
$$I_2 \le 9/(16q) + 128/(3q) \le 44/q.$$
Since $\int_{\mathbb{R}^2} |f(x)|^q dx = I_1 + I_2$, this completes the proof. □

Proof of Proposition 2.4. The case $m = 1$ is trivial, so we assume $m \ge 2$. As in Lemma 2.6, set $f_i = \hat{p}_i$, and denote the density of $\xi_1, \xi_2, \ldots, \xi_m$ by p. Then, applying the Hölder inequality and Lemma 2.6, we obtain
$$p(x) = \left| \int_{\mathbb{R}^2} \prod_{i=1}^m f_i(y) \exp(2\pi i \langle x, y \rangle) dy \right| \le \int_{\mathbb{R}^2} \prod_{i=1}^m |f_i(y)| dy \le \prod_{i=1}^m \left(\int_{\mathbb{R}^2} |f_i(y)|^m dy \right)^{1/m} \le C/m.$$
□

2.2 Concentration

The next lemma is the tensorization argument. It is a variant of Lemma 2.2 in [49] and its proof follows the same lines. We include it for the sake of completeness.

Lemma 2.7. *Let $\xi_1, \ldots, \xi_n$ be independent complex random variables and $\varepsilon_0, p_1, \ldots, p_n$ be non-negative real numbers. Assume that for every $i \le n$ and every $\varepsilon \ge \varepsilon_0$ one has*
$$\mathbb{P}\{|\xi_i| \le \varepsilon\} \le \varepsilon^2 p_i.$$
Then for every $\varepsilon \ge \varepsilon_0$ one has
$$\mathbb{P}\Big\{ \sum_{i=1}^n |\xi_i|^2 \le \varepsilon^2 n \Big\} \le (C_{2.7}\varepsilon)^{2n} \prod_{i=1}^n p_i,$$
where $C_{2.7} > 0$ is a universal constant.

Proof. Let $\varepsilon \geq \varepsilon_0$. Using the hypothesis of the lemma and the distribution integral formula, we have

$$\mathbb{E}\exp(-|\xi_i|^2/\varepsilon^2) = \int_0^1 \mathbb{P}\{\exp(-|\xi_i|^2/\varepsilon^2) > s\}\, ds = \int_0^\infty 2ue^{-u^2}\mathbb{P}\{|\xi_i| < u\varepsilon\}\, du$$
$$\leq p_i\varepsilon^2 \int_{\varepsilon_0/\varepsilon}^\infty 2u^3 e^{-u^2}\, du + p_i\varepsilon_0^2 \int_0^{\varepsilon_0/\varepsilon} 2ue^{-u^2}\, du \leq Cp_i\varepsilon^2.$$

By Markov's inequality, we obtain

$$\mathbb{P}\Big\{\sum_{i=1}^n |\xi_i|^2 \leq \varepsilon^2 n\Big\} = \mathbb{P}\Big\{\exp\Big(-\frac{1}{\varepsilon^2}\sum_{i=1}^n |\xi_i|^2\Big) \geq e^{-n}\Big\}$$
$$\leq e^n\mathbb{E}\exp\Big(-\frac{1}{\varepsilon^2}\sum_{i=1}^n |\xi_i|^2\Big) = e^n\prod_{i=1}^n \mathbb{E}\exp(-|\xi_i|^2/\varepsilon^2),$$

which implies the desired result. □

The following statement is a non-Hermitian counterpart of the spectral gap estimates for undirected random d-regular graphs – a characteristic of major importance in connection with the graph expansion properties. We refer to [1, 10, 13, 17, 19, 21, 24, 25, 32, 46, 57] for more information on random expanders. The following statement was first proved in [13] for $d \leq C\sqrt{n}$ (which is enough in this paper), then it was extended to the range $d \leq Cn^{2/3}$ in [17] and to $d \leq n/2$ in [57].

Theorem 2.8. *There exists a universal constant $C_{2.8} > 0$ such that for $1 \leq d \leq \frac{n}{2}$ one has $\mathbb{P}(\mathcal{E}_{2.8}) \geq 1 - 1/n^{100}$, where*

$$\mathcal{E}_{2.8} := \Big\{M \in \mathcal{M}_{n,d} : \Big\|M - \frac{d}{n}\mathbf{1}\mathbf{1}^t\Big\| \leq C_{2.8}\sqrt{d}\Big\}.$$

In fact one can replace the term $1/n^{100}$ in the above probability bound by any negative power of n at the expense of increasing the constant $C_{2.8}$.

3 Edge count statistics of d-regular digraphs

A combination of probabilistic arguments shows that the edge counting statistics of random d-regular digraphs (i.e., the number of edges connecting subsets of vertices of given cardinalities) concentrate around their average values. In this section, we collect some estimates of the number of edges connecting given subsets of vertices (equivalently, the number of non-zero elements in a given submatrix) and of the number of in- or out-neighbors of a given vertex subset. While some of the statements are borrowed from earlier works, others are new. We would like to note that properties of this type were considered in the random setting in [14, 37, 38].

First we introduce some notation. Denote by $\mathcal{D}_{n,d}$ the set of directed d-regular graphs on n vertices, where we allow loops but no multiple edges. This way, there is a natural bijection between $\mathcal{D}_{n,d}$ and $\mathcal{M}_{n,d}$. We endow $\mathcal{D}_{n,d}$ with the uniform probability measure also denoted by $\mathbb{P}$. Given a graph $G \in \mathcal{D}_{n,d}$ with an edge set E and a subset $I \subset [n]$ of its vertices, define sets of out- and in-neighbors as

$$\mathcal{N}_G^{out}(I) = \big\{v \leq n : \exists i \in I\ (i,v) \in E\big\} \quad \text{and} \quad \mathcal{N}_G^{in}(I) = \big\{v \leq n : \exists i \in I\ (v,i) \in E\big\}.$$

Similarly, we define the out-edges and the in-edges as

$$E_G^{out}(I) = \big\{e \in E : e = (i,j) \text{ for some } i \in I \text{ and } j \leq n\big\}$$

and

$$E_G^{in}(I) = \big\{e \in E : e = (i,j) \text{ for some } i \leq n \text{ and } j \in I\big\}.$$

If $I=\{i\}$ we use lighter notation $\mathcal{N}_G^{out}(i)$, $\mathcal{N}_G^{in}(i)$, $E_G^{out}(i)$, and $E_G^{in}(i)$. Given a graph $G=([n],E)$, for every $I,J\subset[n]$ the set of all edges departing from I and landing in J is denoted by

$$E_G(I,J):=\big\{e\in E:\ e=(i,j)\text{ for some }i\in I\text{ and }j\in J\big\}.$$

Let $M=\{\mu_{ij}\}\in\mathcal{M}_{n,d}$ and let $R_i=R_i(M)$ be the i's row of M, $i=1,...,n$. For every subset $J\subset[n]$, let

$$S_J:=\{i\le n\ :\ \operatorname{supp}R_i\cap J\neq\emptyset\}$$

be the union of supports of columns indexed by J (the matrix will be clear from the context). Given an $n\times n$ matrix M and a set $K\subset[n]$, we use notation M^K for a $|K|\times n$ matrix obtained from M by removing rows $R_i(M)$ with indices $i\notin K$.

We start with the statement which essentially says that given a typical d-regular digraph and a set of vertices J, which is not too large, the set of all in-neighbors of J has cardinality close to the largest possible, i.e., $d|J|$. To formulate the statement, given $k\le n$ and $\varepsilon\in(0,1)$ we introduce the set

$$\Omega_{k,\varepsilon}:=\Big\{M\in\mathcal{M}_{n,d}\ :\ \forall J\subset[n]\ \text{ with }\ |J|=k\ \text{ one has }\ |S_J|\ge(1-\varepsilon)dk\Big\}. \tag{1}$$

Clearly, if $k=1$ then $\Omega_{k,\varepsilon}=\mathcal{M}_{n,d}$. The following theorem is essentially Theorem 2.2 of [37] (see also Theorem 3.1 there).

Theorem 3.1. *Let $e^8<d\le n$, $\varepsilon_0=\sqrt{\ln d/d}$, and $\varepsilon\in[\varepsilon_0,1)$. Let $k\le c_{3.1}\varepsilon n/d$, where $c_{3.1}\in(0,1)$ is a sufficiently small absolute positive constant. Then*

$$\mathbb{P}(\Omega_{k,\varepsilon})\ge1-\exp\left(-\frac{\varepsilon^2dk}{8}\ln\left(\frac{e\varepsilon c_{3.1}n}{kd}\right)\right).$$

In particular,

$$\mathbb{P}\Big(\bigcap_{k=1}^{\lfloor c_{3.1}\varepsilon n/d\rfloor}\Omega_{k,\varepsilon}\Big)\ge1-(Cd/\varepsilon n)^{\varepsilon^2d/8},$$

where C is an absolute positive constant.

In this paper, we prove the following auxiliary theorem, which states that given a large set of columns (of a typical matrix from $\mathcal{M}_{n,d}$), there are many rows having many ones in this set.

Theorem 3.2. *Let $d\le n$ be large enough integers and let $\ell_0\ge d+24en/d$. For every $k\ge\ell_0$, denote*

$$\alpha_k:=\frac{d(k-d)}{8en}-1\qquad\text{and}\qquad\beta_k:=\max\Big(en\,\exp\left(-\alpha_k/2\right),\frac{4k\ln(en/k)}{\alpha_k}\Big).$$

Let $\mathcal{E}_{3.2}$ be the set of all $M\in\mathcal{M}_{n,d}$ such that for every $J\subset[n]$ with $|J|\ge\ell_0$ one has

$$\big|\big\{i\le n:\ |\operatorname{supp}R_i(M)\cap J|<\alpha_{|J|}\big\}\big|\le\beta_{|J|}.$$

Then

$$\mathbb{P}(\mathcal{E}_{3.2})\ge1-4e^{-\ell_0}.$$

Before passing to the proof of Theorem 3.2 we mention an immediate corollary which will be used in Section 7.

Corollary 3.3. *There exist positive absolute constants $C_{3.3},c_{3.3}$ such that the following holds. Let $C_{3.3}\le d\le c_{3.3}\sqrt{n}$ and let $\mathcal{E}_{3.3}$ be the set of all $M\in\mathcal{M}_{n,d}$ such that for every $J\subset[n]$ with $|J|\ge n/\sqrt{d}$ one has*

$$\big|\big\{i\le n:\ |\operatorname{supp}R_i(M)\cap J|<c_{3.3}d|J|/n\big\}\big|\le n/\sqrt{d}.$$

Then

$$\mathbb{P}(\mathcal{E}_{3.3})\ge1-4\exp\big(-n/\sqrt{d}\big).$$

Proof of Corollary 3.3. Let $J \subset [n]$ be such that $k := |J| \geq n/\sqrt{d}$. By the conditions on n and d we have $k \geq 2d$ so that, with α_k, β_k defined as in Theorem 3.2, we have $\alpha_k \geq dk/(32en) \geq \sqrt{d}/(32e)$ and $\beta_k \leq Cn \ln d/d$, for some positive constant C. Adjusting the choice of the constant $c_{3.3}$ and applying Theorem 3.2 with $\ell_0 = \lceil n/\sqrt{d} \rceil \geq d + 24en/d$ we obtain the result. $\square$

In order to prove Theorem 3.2, we need the lemma below (it will be more convenient for us to formulate it in the graph language). For every $S, J \subset [n]$ we introduce

$$\Gamma_S^J := \{G \in \mathcal{D}_{n,d} : \forall i \in S \text{ one has } |E_G(i, J)| < \alpha_{|J|}\},$$

where α_k's were defined in Theorem 3.2. When $S = [s]$ and $J = [k]$ (we postulate that $[0] = \emptyset$) we will denote the above set by Γ_s^k. For every $\ell \leq d$, denote

$$\Gamma_{s,\ell}^k := \Gamma_{s-1}^k \cap \{G \in \mathcal{D}_{n,d} : |E_G(s, [k])| = \ell\}.$$

With these notations, we have $\Gamma_0^k = \mathcal{D}_{n,d}$ for every k. Clearly

$$\Gamma_s^k \subseteq \Gamma_{s-1}^k \qquad \text{and} \qquad \Gamma_s^k = \bigsqcup_{\ell=0}^{\lfloor \alpha_k \rfloor} \Gamma_{s,\ell}^k. \tag{2}$$

Lemma 3.4. *Let d, n, ℓ_0, and α_k, $k \geq \ell_0$, be as in Theorem 3.2 and let $S, J \subset [n]$ be such that $|J| \geq \ell_0$. Then*

$$\mathbb{P}(\Gamma_S^J) \leq \exp(-\alpha_{|J|}|S|).$$

Proof. Without loss of generality we may assume that $S = [s]$ and $J = [k]$ for some $s \geq 1$ and $k \geq \ell_0$. Let q be a parameter in the interval $\alpha_k < q \leq d$, which will be chosen later. We first compare the cardinalities of $\Gamma_{s,\ell}^k$ and $\Gamma_{s,\ell+1}^k$ for every $\ell < q$. To this end we construct a relation $R_\ell \subset \Gamma_{s,\ell}^k \times \Gamma_{s,\ell+1}^k$. Let $G \in \Gamma_{s,\ell}^k$. For $j > k$ denote

$$E_j := E_G([s]^c, [k]) \setminus \Big(E_G^{in}\big(\mathcal{N}_G^{out}(s) \cap [k]\big) \cup E_G^{out}\big(\mathcal{N}_G^{in}(j)\big)\Big).$$

Since $G \in \Gamma_{s,\ell}^k$,

$$|E_G([s]^c, [k])| = kd - |E_G([s], [k])| \geq kd - \alpha_k(s-1) - \ell.$$

On the other hand, since $|\mathcal{N}_G^{out}(s) \cap [k]| = \ell$, then $|E_G\big([s]^c, \mathcal{N}_G^{out}(s) \cap [k]\big)| \leq \ell(d-1)$. Using that

$$|E_G\big(\mathcal{N}_G^{in}(j) \setminus [s], [k]\big)| \leq d(d-1)$$

we obtain

$$|E_j| \geq kd - \alpha_k(s-1) - \ell d - d(d-1) \geq (k-d+1)d - q(s+d-1). \tag{3}$$

Now we are ready to define the relation R_ℓ. We let a pair (G, G') belong to R_ℓ for some $G' \in \Gamma_{s,\ell+1}^k$ if G' can be obtained from G in the following way. Choose $j \in \mathcal{N}_G^{out}(s) \cap [k]^c$ and an edge $(u, v) \in E_j$. We destroy the edge (s, j) and create the edge (s, v), then we destroy the edge (u, v) and create the edge (u, j) (in other words, we perform the simple switching on the vertices s, u, j, v). Note that the conditions $u \notin \mathcal{N}_G^{in}(j)$ and $v \notin \mathcal{N}_G^{out}(s)$, which are implied by the definition of E_j, guarantee that the simple switching does not create multiple edges, and we obtain a valid graph in $\Gamma_{s,\ell+1}^k$. Using (3) and assuming

$$q \leq \frac{d(k-d)}{2(n+d)}, \tag{4}$$

we deduce that for every $G \in \Gamma_{s,\ell}^k$ one has

$$|R_\ell(G)| \geq (d-\ell)\Big[(k-d+1)d - q(s+d-1)\Big] \geq \frac{d(k-d)(d-q)}{2}. \tag{5}$$

Now we estimate the cardinalities of preimages. Let $G' \in R_\ell(\Gamma^k_{s,\ell})$. In order to reconstruct a graph G for which $(G, G') \in R_\ell$ we need to perform a simple switching which destroys an edge in $E_{G'}(s, [k])$ and adds an edge in $E_{G'}(s, [k]^c)$. To this end, choose

$$v \in \mathcal{N}^{out}_{G'}(s) \cap [k] \quad \text{and} \quad j \in [k]^c \setminus \mathcal{N}^{out}_{G'}(s).$$

Since $|\mathcal{N}^{in}_{G'}(j)| = d$, there are at most d simple switchings which destroy the edge (s, v) and create the edge (s, j). Using that $|\mathcal{N}^{out}_{G'}(s) \cap [k]| = \ell + 1$, we observe

$$|R_\ell^{-1}(G')| \leq d(\ell + 1)(n - k - (d - \ell - 1)) \leq dq(n - k). \tag{6}$$

Using Claim 2.1 together with inequalities (5) and (6), we obtain that for every $\ell < q$,

$$|\Gamma^k_{s,\ell}| \leq \frac{2q(n-k)}{(k-d)(d-q)} |\Gamma^k_{s,\ell+1}|.$$

Therefore

$$|\Gamma^k_{s,\ell}| \leq \left(\frac{2q(n-k)}{(k-d)(d-q)} \right)^{q-\ell} |\Gamma^k_{s,q}| \leq \left(\frac{2q(n-k)}{(k-d)(d-q)} \right)^{q-\ell} |\Gamma^k_{s-1}|.$$

This together with (2) implies that

$$|\Gamma^k_s| \leq \sum_{\ell=0}^{\lfloor \alpha_k \rfloor} \left(\frac{2q(n-k)}{(k-d)(d-q)} \right)^{q-\ell} |\Gamma^k_{s-1}| \leq e^{\lfloor \alpha_k \rfloor + 1 - q} \, |\Gamma^k_{s-1}|,$$

provided that

$$\frac{2q(n-k)}{(k-d)(d-q)} \leq e^{-1}.$$

We choose $q = 2\lfloor \alpha_k \rfloor + 2$ which satisfies the above condition and the condition (4) by the definition of α_k. Therefore we have

$$|\Gamma^k_s| \leq e^{-\alpha_k} |\Gamma^k_{s-1}|.$$

Since we do not impose any restrictions on s, we conclude that

$$|\Gamma^k_s| = |\mathcal{D}_{n,d}| \prod_{p=1}^{s} \frac{|\Gamma^k_p|}{|\Gamma^k_{p-1}|} \leq e^{-s\alpha_k} \, |\mathcal{D}_{n,d}|.$$

$\square$

Proof of Theorem 3.2. We start by defining

$$\Gamma := \Big\{ G \in \mathcal{D}_{n,d} : \, \exists J \subset [n], \, |J| \geq \ell_0 \text{ with } \big| \{ i \leq n : \, |E_G(i, J)| < \alpha_{|J|} \} \big| > \beta_{|J|} \Big\}.$$

It is not difficult to see that Γ is the graph counterpart of event $\mathcal{E}_{3.2}$, in particular $\mathbb{P}(\mathcal{E}_{3.2}) = \mathbb{P}(\Gamma)$. Note that

$$\Gamma = \bigcup_{\substack{J \subset [n] \\ |J| \geq \ell_0}} \bigcup_{\substack{S \subset [n] \\ |S| > \beta_{|J|}}} \Gamma^J_S.$$

Therefore, applying Lemma 3.4 and taking the union bound, we obtain

$$\mathbb{P}(\Gamma) \leq \sum_{\substack{J \subset [n] \\ |J| \geq \ell_0}} \sum_{\substack{S \subset [n] \\ |S| > \beta_{|J|}}} \mathbb{P}(\Gamma^J_S) \leq \sum_{\substack{J \subset [n] \\ |J| \geq \ell_0}} \sum_{\substack{S \subset [n] \\ |S| \geq \beta_{|J|}}} e^{-\alpha_{|J|} |S|} = \sum_{k \geq \ell_0} \sum_{s \geq \beta_k} \binom{n}{k} \binom{n}{s} e^{-\alpha_k s}.$$

Further, by the choice of β_k, we have $\binom{n}{s}e^{-\alpha_k s} \leq (en/s)^s e^{-\alpha_k s} \leq e^{-\alpha_k s/2}$ for all $s \geq \beta_k$ and since $\alpha_k \geq 2$, then

$$\sum_{s\geq\beta_k} \binom{n}{s} e^{-\alpha_k s} \leq 2e^{-\alpha_k\beta_k/2}.$$

By the choice of β_k's, this implies

$$\mathbb{P}(\Gamma) \leq 2\sum_{k\geq\ell_0} \binom{n}{k} e^{-\alpha_k\beta_k/2} \leq 2\sum_{k\geq\ell_0} \left(\frac{k}{en}\right)^k \leq 4e^{-\ell_0},$$

which completes the proof. □

Combining Theorems 3.1 and 3.2, we prove the following proposition.

Proposition 3.5. *There exist absolute positive constants C and $c_{3.5}$ such that the following holds. Let $C \leq d \leq c_{3.5}\sqrt{n}/\ln n$ and let $\mathcal{E}_{3.5}$ be the set of all $M \in \mathcal{M}_{n,d}$ such that for all $J \subset [n]$ one has*

$$\Big|\big\{i \leq n : |\operatorname{supp} R_i(M) \cap J| \geq c_{3.5}d|J|/n \quad \text{and} \\ |\operatorname{supp} R_i(M) \cap J^c| \geq c_{3.5}d|J^c|/n\big\}\Big| \geq c_{3.5}\min(d|J|, d|J^c|, n).$$

Then

$$\mathbb{P}(\mathcal{E}_{3.5}) \geq 1 - n^{-c_{3.5}d}.$$

Proof. Let $c > 0$ be small enough absolute constant and let $d \leq c\sqrt{n}$. We first treat subsets J satisfying $|J| < 2d$. Recall that S_J denotes the union of supports of columns indexed by J. Therefore, using $d \leq c\sqrt{n}$ and applying Theorem 3.1, we observe that there exists an absolute constant $c' > 0$ such that with probability at least $1 - n^{-c'd}$ one has $|S_J| \geq 0.9d|J|$ for all J with $|J| < 2d$. Now fix $J \subset [n]$ satisfying both $|J| < 2d$ and $|S_J| \geq 0.9d|J|$. Since $d \leq c\sqrt{n}$, then $cd|J|/n \leq 2c^3 < 1$ for small enough c. Therefore, the condition

$$|\operatorname{supp} R_i(M) \cap J| \geq cd|J|/n$$

means that $i \in S_J$. Note also that in this case $\min(d|J|, d|J^c|, n) = d|J|$. Thus, it is enough to show that

$$|\{i \in S_J : |\operatorname{supp} R_i(M) \cap J^c| \geq cd\}| \geq cd|J|.$$

Let ℓ denote the number of rows having at least $d-1$ ones in J^c. Counting ones in the columns indexed by J we have

$$d|J^c| = dn - d|J| \leq d\ell + (d-2)(n-\ell) = 2\ell + dn - 2n.$$

This implies $\ell \geq n - d|J|/2$. Therefore there are at least $\ell + |S_J| - n \geq 0.4d|J|$ rows indexed by S_J and having at least $d-1$ ones in J^c. This proves that the set of all matrices in $\mathcal{M}_{n,d}$ satisfying the condition of the proposition for subsets $J \subset [n]$ with $|J| < 2d$, has measure at least $1 - n^{-c'd}$. Interchanging the role of J and J^c we obtain the same bound for subsets J satisfying $|J| > n - 2d$.

For the rest of the proof we deal only with sets J satisfying $|J| \in [2d, n-2d]$ for which the quantities $d(|J|-d)/n$ and $d|J|/n$ are equivalent up to a constant multiple (and similarly for J^c). We will prove a more precise relation, which is convenient to formulate in the graph language. Namely, denoting by $\mathcal{E}$ the set of all digraphs $G \in \mathcal{D}_{n,d}$ such that for every $J \subset [n]$ with $|J| \in [2d, n-2d]$

$$\left|\left\{i \leq n : |E_G(i,J)| \geq \frac{c_1 d(|J|-d)}{n} \;\text{ and }\; |E_G(i,J^c)| \geq \frac{c_1 d(|J^c|-d)}{n}\right\}\right| \\ \geq c_1\min(d|J|, d|J^c|, n),$$

we show that the event $\mathcal{E}$ has probability at least $1 - n^{-c_1 d}$, where $c_1 > 0$ is a sufficiently small universal constant.

Given $\ell_0 \geq d + 24en/d$ and $k \geq \ell_0$, we consider parameters α_k and β_k introduced in Theorem 3.2. Additionally, for $k < \ell_0$ we set $\alpha_k := \alpha_{\ell_0}$ and let β_k be defined by the same formula as β_k for $k \geq \ell_0$. Note that $(\alpha_k)_k$ is a non-decreasing sequence, while $(\beta_k)_k$ is a non-increasing sequence. Now set $\ell_0 = \lfloor d + Cn/d \rfloor$, where $C \geq 24e$ is a sufficiently large universal constant chosen so that $\beta_{\ell_0} \leq n/4$ (then $\beta_i + \beta_j \leq n/2$ for every $i, j \geq \ell_0$). Note that $2d < \ell_0 < n/2$. Define the event

$$\begin{aligned} \mathcal{E}_1 := \Big\{ G \in \mathcal{D}_{n,d} :\, & \forall J \subset [n], |J| \in [\ell_0, n - \ell_0]), \quad \text{one has} \\ & \big| \big\{ i \leq n :\, |E_G(i, J)| \geq \alpha_{|J|} \text{ and } |E_G(i, J^c)| \geq \alpha_{|J^c|} \big\} \big| \geq n - \beta_{|J|} - \beta_{|J^c|} \Big\} \end{aligned}$$

and, for $m = 2, 3$, the events

$$\begin{aligned} \mathcal{E}_m := \Big\{ G \in \mathcal{D}_{n,d} :\, & \forall J \subset [n], |J| \in S_m, \quad \text{one has} \\ & \Big| \Big\{ i \leq n :\, |E_G(i, J)| \geq \frac{\alpha_{|J|}}{\alpha_{\ell_0}} \text{ and } |E_G(i, J^c)| \geq \frac{\alpha_{|J^c|}}{\alpha_{\ell_0}} \Big\} \Big| \geq c_1 \, \min(d|J|, d|J^c|, n) \Big\}. \end{aligned}$$

where $S_2 = [\ell_0 - 1]$, $S_3 = [n - \ell_0 + 1, n]$. Then we clearly have $\mathbb{P}(\mathcal{E}_2) = \mathbb{P}(\mathcal{E}_3)$ and, moreover, $\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3 \subset \mathcal{E}$, provided that c_1 is sufficiently small. Therefore,

$$\mathbb{P}(\mathcal{E}) \geq \mathbb{P}(\mathcal{E}_1) + \mathbb{P}(\mathcal{E}_2) + \mathbb{P}(\mathcal{E}_3) - 2 = \mathbb{P}(\mathcal{E}_1) + 2\mathbb{P}(\mathcal{E}_2) - 2. \tag{7}$$

First, we estimate probability of $\mathcal{E}_1$. For any set J with $|J| \geq \ell_0$ and $|J^c| \geq \ell_0$, the condition

$$\big| \big\{ i \leq n :\, |E_G(i, J)| \geq \alpha_{|J|} \big\} \big| \geq n - \beta_{|J|} \text{ and } \big| \big\{ i \leq n :\, |E_G(i, J^c)| \geq \alpha_{|J^c|} \big\} \big| \geq n - \beta_{|J^c|}$$

for a graph $G \in \mathcal{D}_{n,d}$ implies

$$\big| \big\{ i \leq n :\, |E_G(i, J)| \geq \alpha_{|J|} \quad \text{and} \quad |E_G(i, J^c)| \geq \alpha_{|J^c|} \big\} \big| \geq n - \beta_{|J|} - \beta_{|J^c|}.$$

Therefore by Theorem 3.2 we obtain $\mathbb{P}(\mathcal{E}_1) \geq 1 - 8e^{-\ell_0}$.

We now turn to the probability of $\mathcal{E}_2$. Note that for every J with $|J| < \ell_0$, we have $\alpha_{|J|} = \alpha_{\ell_0}$ and $\alpha_{|J^c|} \leq \alpha_n$. Further, for every $i \leq n$,

$$|E_G(i, J)| + |E_G(i, J^c)| = d.$$

Therefore, for every graph $G \in \mathcal{D}_{n,d}$ we have

$$\begin{aligned} \Big\{ i \leq n :\, |E_G(i, J)| \geq \frac{\alpha_{|J|}}{\alpha_{\ell_0}} \quad \text{and} \quad |E_G(i, J^c)| \geq \frac{\alpha_{|J^c|}}{\alpha_{\ell_0}} \Big\} & \\ \supset \Big\{ i \leq n :\, |E_G(i, J)| \geq 1 \quad \text{and} \quad |E_G(i, J^c)| \geq \frac{\alpha_n}{\alpha_{\ell_0}} \Big\} & \\ = \mathcal{N}_G^{in}(J) \setminus \Big\{ i \leq n :\, |E_G(i, J)| > d - \frac{\alpha_n}{\alpha_{\ell_0}} \Big\}. & \end{aligned}$$

Since $\alpha_{\ell_0} \geq 2$, $\alpha_n \leq d/(8e)$, we observe

$$\Big| \Big\{ i \leq n :\, |E_G(i, J)| > d - \frac{\alpha_n}{\alpha_{\ell_0}} \Big\} \Big| \leq \frac{d|J|}{d - \alpha_n/\alpha_{\ell_0}} \leq 2|J|.$$

Therefore,

$$\mathcal{E}_2 \supset \Big\{ G \in \mathcal{D}_{n,d} :\, \forall J \subset [n], |J| < \ell_0, \, |\mathcal{N}_G^{in}(J)| \geq c_1 \min(d|J|, n) + 2|J| \Big\}$$

We apply Theorem 3.1 with $\varepsilon = 0.1$. Recall that c_1 is small enough. Theorem 3.1 implies that there exists a universal constant $c_1' > 0$, such that with probability at least $1 - n^{-c_1' d}$ for every J with $|J| \le c_{3.1} n/(10d)$ one has

$$|\mathcal{N}_G^{in}(J)| \ge 0.9d|J| \ge c_1 \min(d|J|, n) + 2|J|$$

and for every J with $c_{3.1} n/(10d) \le |J| < \ell_0$, passing to a subset $J_0 \subset J$ with $|J_0| = \lfloor c_{3.1} n/(10d) \rfloor$ and using $\ell_0 \le 2Cn/d$, one has

$$|\mathcal{N}_G^{in}(J)| \ge |\mathcal{N}_G^{in}(J_0)| \ge 0.9d|J_0| \ge 9c_{3.1} d|J|/(400C) \ge c_1 \min(d|J|, n) + 2|J|.$$

Thus $\mathbb{P}(\mathcal{E}_2) \ge 1 - n^{-c_1' d}$. By (7) this implies $\mathbb{P}(\mathcal{E}) \ge 1 - 8e^{\ell_0} - 2n^{-c_1' d}$, which together with the bounds obtained at the beginning of the proof implies the desired result. $\square$

Finally, we need the following deterministic statement dealing with sets of in-neighbours of two disjoint sets of vertices. Given two disjoint subsets J^ℓ, $J^r \subset [n]$ and a matrix $M \in \mathcal{M}_{n,d}$, denote

$$I^\ell = I^\ell(M) := \{i \le n : |\mathrm{supp} R_i \cap J^\ell| = 1 \ \text{ and } \ \mathrm{supp} R_i \cap J^r = \emptyset\},$$

and

$$I^r = I^r(M) := \{i \le n : \mathrm{supp} R_i \cap J^\ell = \emptyset \ \text{ and } \ |\mathrm{supp} R_i \cap J^r| = 1\}.$$

Here the upper indices ℓ and r refer to *left* and *right*, since later for a given vector $x \in \mathbb{C}^n$, denoting by σ a permutation of $[n]$ satisfying $x_i^* = |x_{\sigma(i)}|$ for all $i \le n$, we will choose $J^\ell = \sigma([k_1])$ and $J^r = \sigma([k_2, n])$ for some $k_1 < k_2$. The following statement is Lemma 2.7 from [38].

Lemma 3.6. *Let d and ε be as in Theorem 3.1. Let $p \ge 2$, $m \ge 1$ be integers satisfying $pm \le c_{3.1} \varepsilon n/d$ and let $J^\ell, J^r \subset [n]$ be such that $J^\ell \cap J^r = \emptyset$, $|J^\ell| = m$, $|J^r| = (p-1)m$. Let $M \in \Omega_{pm,\varepsilon}$. Then*

$$|I^\ell| \ge (1 - 2\varepsilon p) d |J^\ell|.$$

In particular, if $|J^r| = |J^\ell| = m$ with $m \le c_{3.1} \varepsilon n/(2d)$ then

$$(1 - 4\varepsilon) dm \le \min(|I^\ell|, |I^r|) \le \max(|I^\ell|, |I^r|) \le dm.$$

4 Almost constant and steep vectors

As in [38] we split $\mathbb{C}^n$ into three classes of vectors which we call *steep*, *gradual*, and *almost constant vectors*. This section is devoted to steep and almost constant vectors. The definition of steep vectors is similar to the one given in [38], with slight modifications one of which is quite important. Note that if a subset $K \subset [n]$ is such that $|K^c|$ is much larger than $n^{1-1/d}$, then the submatrix M^K will contain null columns with large probability, hence the kernel of M^K will contain very sparse vectors. Therefore, when studying the kernel of M^K (or $(M - z\,\mathrm{Id})^K$), very sparse vectors and those "close" to very sparse should be handled separately, see the definition of $\mathcal{T}_3$ below. The set $\mathcal{T}_3$ – the set of *very steep* vectors – can be viewed as an enlargement of the set of very sparse vectors; in this sense, our construction is related to the definition of compressible vectors,

$$\mathrm{Comp}(m, \rho) := \{x \in \mathbb{C}^n \ : \ \exists \ m\text{-sparse vector } y \in \mathbb{C}^n \ \text{ such that } \ \|x - y\|_2 \le \rho \|x\|_2\},$$

introduced in [49] following ideas from [41] (as usual, m-sparse means that a vector has at most m non-zero coordinates). Both classes, $\mathcal{T}_3$ and Comp, are introduced as classes of vectors close to m-sparse vectors (for an appropriate m). An important difference between the two lies in how the distance to the set of sparse vectors is measured – instead of the Euclidean distance used for compressible vectors, we estimate the ℓ_∞-norm after some normalization related to a variant of the weak $\ell_{1/3}$-norm.

After introducing and eliminating very steep vectors we consider other vectors with a jump in their non-increasing rearrangement. The goal is to show that such vectors are far from the kernel of M^K. We will distinguish two types of jumps. The first one occurs at the beginning of the non-increasing rearrangement and is of order $4d$, that is for certain $m < k$, we have $x^*_m > 4dx^*_k$, see the definition of $\mathcal{T}_0$ below. To treat such vectors x we use Lemma 4.4, which yields that with large probability the random matrix distributed in $\mathcal{M}_{n,d}$ has many rows with exactly one 1 in coordinates corresponding to m largest components of x, and all zero coordinates in places corresponding to $m+1$-st to k-th largest component of x. Since the total numbers of ones in every row is d we have that the inner product of every such row with x is separated from zero. Unfortunately, since this procedure relies on graph expansion properties given by Theorem 3.1, it works only when k is not too large, namely when $k \leq c_{3.1}\varepsilon n/d$. For larger values of k we use a different technique, which requires considering a jump of order $d^{3/2}$, see the definitions of $\mathcal{T}_1$ and $\mathcal{T}_2$. For vectors with such jumps, using the switching technique, we estimate the probability that a fixed vector is close to the kernel, construct a special net of rather small cardinality in the set of such vectors, then use the union bound. This scheme is similar to the one used in [38] with an important difference in the class $\mathcal{T}_2$ – we use a number of coordinates proportional to n in the definition (meaning n_3 is proportional to n), while in [38] we used only $n/\ln d$ coordinates. This makes bounding the probability for individual vectors much more difficult and involved (the method of [38] does not work). The main novelty of our new method is splitting the set $\mathcal{M}_{n,d}$ into many equivalence classes and working in each class separately. The construction of nets, which is also rather delicate, comes from [38].

Finally, we treat almost constant vectors, i.e., vectors having many coordinates which are almost equal to each other. Having excluded vectors with jumps it only remains to treat those without any. However, if a vector x with no jumps has many almost equal coordinates (say on a set J), then its inner product with a row having many ones inside J cannot be close to zero (and we show that there are many such rows). In view of this observation, it is important to find a balance between quantitative characteristics of the level of jumps and the places where these jumps occur. Due to technical reasons, such a balance cannot be achieved directly, in particular, to treat almost constant vectors, one needs to consider constant jumps (not a power of d). Fortunately, it turns out that every almost constant vector without a constant jump can be represented as the sum of a vector with a big jump and a constant vector, i.e., a vector whose coordinates are equal to each other. Moreover, our proof for vectors with big jumps is stable under shifts by constant vectors which makes our treatment of almost constant vectors a lot easier.

We now introduce the following parameters, which will be used throughout this section. First fix $1 \leq L \leq n/d^3$ (we always assume that $n \geq d^3$). When considering the minor M^K, L will be responsible for the size of the set K^c. In order to use Theorem 3.1, we fix ε_0 and a related parameter p as follows:

$$\varepsilon_0 = \sqrt{(\ln d)/d}, \qquad p = \lfloor 1/(5\varepsilon_0) \rfloor = \left\lfloor \tfrac{1}{5}\sqrt{d/\ln d} \right\rfloor$$

(the choice of p comes from $\varepsilon_0 p < 1$ needed in Lemma 4.4 in order to apply Lemma 3.6). Furthermore, we fix a sufficiently small positive absolute constant a_3 (we don't try to estimate the actual value of a_3, the conditions on how small it is appear in the corresponding proofs). Set

$$n_1 := \lceil n/d^{3/2} \rceil, \quad n_2 := \lfloor n/d^{2/3} \rfloor, \quad \text{and} \quad n_3 := \lfloor a_3 n \rfloor.$$

We also fix two positive integers r and $r_0 = r_0(L)$ such that $p^r < n_1 \leq p^{r+1}$ and r_0 is the smallest non-negative integer satisfying $p^{r_0} \geq 20L/d$. Note that $0 \leq r_0 < r$. Indeed,

$$p^{r-1} \geq n_1/p^2 \geq 25n \ln d/d^{5/2} > 20L/d,$$

which implies that $r_0 \leq r - 1$.

Finally, denote the class of constant vectors by

$$\mathcal{K} := \{x = (x_i)_{i=1}^n \in \mathbb{C}^n \, : \, x_1 = x_2 = ... = x_n\}.$$

4.1 Steep vectors

The definition of the class of steep vectors consists of few steps at which we define the sets $\mathcal{T}_0$, $\mathcal{T}_1$, $\mathcal{T}_2$, and $\mathcal{T}_3$. We start with $\mathcal{T}_3$, the class of *very steep* vectors. Set

$$\mathcal{T}_3 = \mathcal{T}_3(L) := \{x \in \mathbb{C}^n \,:\, \exists i \le p^{r_0} \text{ such that } x_i^* > (n/i)^3 x_{p^{r_0}}^*\}.$$

Note that one can relate this class to the class of vectors close to m-sparse vectors with $m = p^{r_0} - 1$. Indeed, consider the following variant of the weak $\ell_{1/3}$-norm,

$$|||x||| = \frac{1}{n^3} \sup_{i \le m} i^3 x_i^*.$$

Then

$$\mathcal{T}_3 = \{x \in \mathbb{C}^n \,:\, \exists \; m\text{-sparse vector } y \in \mathbb{C}^n \text{ such that } \|x - y\|_\infty < |||x|||\}.$$

We now define the set $\mathcal{T}_0$. For $r_0 \le i \le r-1$ set

$$\mathcal{T}_{0,i} = \mathcal{T}_{0,i}(L) := \{x \in \mathbb{C}^n \,:\, x \notin \mathcal{T}_3 \cup \bigcup_{j=r_0}^{i-1} \mathcal{T}_{0,j} \text{ and } x_{p^i}^* > 4d x_{p^{i+1}}^*\},$$

where $\cup_{j=r_0}^{r_0-1} \mathcal{T}_{0,j}$ means $\emptyset$, and for $i = r$ let

$$\mathcal{T}_{0,r} = \mathcal{T}_{0,r}(L) := \{x \in \mathbb{C}^n \,:\, x \notin \mathcal{T}_3 \cup \bigcup_{j=r_0}^{r-1} \mathcal{T}_{0,j} \text{ and } x_{\lceil n_1/p \rceil}^* > 4d x_{n_1}^*\},$$

Let

$$\mathcal{T}_0 = \mathcal{T}_0(L) := \bigcup_{i=r_0}^{r} \mathcal{T}_{0,i}.$$

Next we define $\mathcal{T}_1$ and $\mathcal{T}_2$ as

$$\mathcal{T}_1 = \mathcal{T}_1(L) := \{x \in \mathbb{C}^n \,:\, x \notin \mathcal{T}_0 \cup \mathcal{T}_3 \text{ and } x_{n_1}^* > d^{3/2}\, x_{n_2}^*\}$$

and

$$\mathcal{T}_2 = \mathcal{T}_2(L) := \{x \in \mathbb{C}^n \,:\, x \notin \mathcal{T}_0 \cup \mathcal{T}_1 \cup \mathcal{T}_3 \text{ and } x_{n_2}^* > d^{3/2}\, x_{n_3}^*\}.$$

Below we work with constant shifts of steep vectors, so we also introduce the following sets for $0 \le i \le 3$,

$$\mathcal{T}_i^{\mathcal{K}} := \{v \in \mathbb{C}^n \,:\, v = x + y \quad \text{for some} \quad x \in \mathcal{T}_i \text{ and } y \in \mathcal{K} \text{ with } |y_1| \le x_{n_1}^*/10\}.$$

Note that

$$\mathcal{T}_3^{\mathcal{K}} \subset \{v \in \mathbb{C}^n \,:\, \exists i \le p^{r_0} \text{ such that } v_i^* > 0.9(n/i)^3 v_{p^{r_0}}^*\}. \tag{8}$$

Finally we define sets of *steep* and *shifted steep* vectors as

$$\mathcal{T} := \mathcal{T}_0 \cup \mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}_3 \quad \text{and} \quad \mathcal{T}_{\mathcal{K}} := \mathcal{T}_0^{\mathcal{K}} \cup \mathcal{T}_1^{\mathcal{K}} \cup \mathcal{T}_2^{\mathcal{K}}.$$

Note that the set of steep vectors contains *very steep* vectors.

Our first goal in this section is to prove the following theorem.

Theorem 4.1. *Let $d \ge 1$ be large enough, $n \ge d^3$, $1 \le L \le n/d^3$, $K \subset [n]$ with $|K^c| \le L$, and $z \in \mathbb{C}$ be such that $|z| \le d/2$. Let*

$$\mathcal{E}_{steep} := \Big\{M \in \mathcal{M}_{n,d} \,:\, \exists\, v \in \mathcal{T}_{\mathcal{K}} \quad \textit{such that} \quad \|(M - z\mathrm{Id})^K v\|_2 < \frac{L^3 d}{n^6}\, \|v\|_2\Big\}.$$

Then

$$\mathbb{P}(\mathcal{E}_{steep}) \le \min\big(\exp(-L/d)\,, \exp\big(-(\ln d)(\ln n)/20\big)\big)\,.$$

We will now formulate simple properties of steep vectors which will be used later. The following lemma shows that the vectors from the complement of $\mathcal{T}$ have a rather regular decay of their coordinates.

Lemma 4.2. *Let $d \geq 1$ be large enough, $n \geq d^3$, and $x \notin \mathcal{T}$. Then*

$$x_m^* \leq \begin{cases} (n/m)^6 \ x_{n_3}^* & \text{if } 1 \leq m \leq p^{r_0}, \\ d\,(n/m)^3 \ x_{n_3}^*, & \text{if } p^{r_0} \leq m \leq n_1. \end{cases}$$

Furthermore, for every $n_1 \leq j \leq i \leq n_3$ one has

$$x_j^* \leq x_{n_1}^* \leq d^3 x_{n_3}^* \leq d^3 x_i^*.$$

Proof. Since $p^r \geq \lceil n_1/p \rceil$ and $x \notin \mathcal{T}$, we have $x_{p^r}^* \leq 4d x_{n_1}^* \leq 4d^4 x_{n_3}^*$. Therefore, for every $r_0 \leq j \leq r$,

$$x_{p^j}^* \leq (4d) x_{p^{j+1}}^* \leq \ldots \leq (4d)^{r-j} x_{p^r}^* \leq 4d^4 (4d)^{r-j} x_{n_3}^*.$$

Since for large d one has $4d < p^3$ and $p^r \leq n_1 \leq n/d^{3/2} + 1$, we deduce for $j = r_0$ that

$$x_{p^{r_0}}^* \leq 4d^4 (4d)^{r-r_0} x_{n_3}^* \leq 4d^4 p^{3(r-r_0)} x_{n_3}^* \leq 4d^4 (n_1/p^{r_0})^3 x_{n_3}^* \leq (n/p^{r_0})^3 x_{n_3}^*.$$

Since $x \notin \mathcal{T}_3$, this implies the bound for every $1 \leq m \leq p^{r_0}$.

Now let $p^j \leq m < p^{j+1}$ for some $r_0 \leq j < r$. Then

$$x_m^* \leq x_{p^j}^* \leq 4d^4 p^{3(r-j)} x_{n_3}^* \leq 4d^4 (n_1 p/m)^3 x_{n_3}^* \leq d\,(n/m)^3 \ x_{n_3}^*,$$

which proves the case $m < p^r$. For $p^r \leq m \leq n_1$ we have $n/m \geq n/n_1 \geq d^{3/2}/2$, hence

$$x_m^* \leq 4d x_{n_1}^* \leq 4d^4 x_{n_3}^* \leq (n/m)^3 \ x_{n_3}^*.$$

The last inequality is trivial. $\square$

The next lemma provides a comparison of the ℓ_2-norm of a given vector with one of its coordinates. It is similar to Lemma 3.5 from [38]. Since our choice of parameters as well as the definition of steep vectors is slightly different we provide the proof for the sake of completeness.

Lemma 4.3. *Let $d \geq 1$ be large enough, $n \geq d^3$, $1 \leq L \leq n/d^3$, and $x \in \mathbb{C}^n \setminus \mathcal{T}_3$. Then*

$$\|x\|_2 \leq \frac{n^6}{100 L^3 d^{3/2}} \, x_m^*,$$

where $m = p^i$ if $x \in \mathcal{T}_{0,i}$ for some $r_0 \leq i \leq r$ and $m = n_1$ if $x \notin \mathcal{T}_3 \cup \mathcal{T}_0$.

Proof. By the definition of $\mathcal{T}_3$, for $x \notin \mathcal{T}_3$ one has

$$\sum_{i=1}^{p^{r_0}} (x_i^*)^2 \leq \sum_{i=1}^{p^{r_0}} (n/i)^6 \, (x_{p^{r_0}}^*)^2 \leq \frac{4}{3} n^6 \, (x_{p^{r_0}}^*)^2.$$

If $x \in \mathcal{T}_{0,r_0}$, then

$$\|x\|_2^2 = \sum_{j=1}^{p^{r_0}-1} (x_j^*)^2 + \sum_{j=p^{r_0}}^{n} (x_j^*)^2 \leq \frac{4}{3} n^6 \, (x_{p^{r_0}}^*)^2 + n (x_{p^{r_0}}^*)^2 \leq 2\, n^6 \, (x_{p^{r_0}}^*)^2,$$

which implies the bound in the case $i = r_0$. Let $x \in \mathcal{T}_{0,i}$ for some $r_0 < i \leq r$. Then $x \notin \mathcal{T}_3$ and for every $j < i$ one has $x \notin \mathcal{T}_{0,j}$. Therefore, assuming without loss of generality that $x_{p^i}^* = 1$, as in the previous lemma we observe

$$x_{p^{r_0}}^* \leq (4d)^{i-r_0} x_{p^i}^* = (4d)^{i-r_0} \leq p^{3(i-r_0)}.$$

Therefore, using again that $4d \le p^3$, we observe

$$
\begin{aligned}
\|x\|_2^2 &= \sum_{j=1}^{p^{r_0}} (x_j^*)^2 + \sum_{j=p^{r_0}+1}^{p^{r_0+1}} (x_j^*)^2 + \sum_{j=p^{r_0+1}+1}^{p^{r_0+2}} (x_j^*)^2 + \dots \\
&\le 2\, n^6\, (x_{p^{r_0}}^*)^2 + p^{r_0+1}(4d)^{2(i-r_0)} + p^{r_0+2}(4d)^{2(i-r_0-1)} + \dots + p^i(4d)^2 + n \\
&\le 3n^6 p^{6(i-r_0)} + (i-r_0)p^{6i+1-5r_0} \\
&\le 4n^6 p^{6(r-r_0)}.
\end{aligned}
$$

Recalling that $20L/d \le p^{r_0} \le 20Lp/d$ and $p^r \le n_1 = \lceil n/d^{3/2} \rceil \le p^{r+1}$, we have

$$
p^{r-r_0} \le \frac{n_1 d}{20L} \le \frac{n}{10L\sqrt{d}},
$$

which, together with the above, implies the desired bound in the case $r_0 < i \le r$. Repeating the above scheme and using that $p^r \le n_1 < p^{r+1}$, we obtain the result for $x \notin \mathcal{T}_3 \cup \mathcal{T}_0$. $\square$

4.2 Lower bounds on $\|Mx\|_2$ for vectors from $\mathcal{T}_0$

Here we provide lower bounds on the ratio $\|Mx\|_2/\|x\|_2$ for vectors x from $\mathcal{T}_0$. Recall that given ε and k the set $\Omega_{k,\varepsilon}$ was introduced before Theorem 3.1 (see (1)).

Lemma 4.4. *There exists an absolute positive constant C such that the following holds. Let $d \ge C$, $n \ge d^3$, $1 \le L \le n/d^3$, $K \subset [n]$ with $|K^c| \le L$, and let $z \in \mathbb{C}$ be such that $|z| \le d/2$. Then for every $v \in \mathcal{T}_0^{\mathcal{K}}$ and every*

$$
M \in \Omega_{n_1,\varepsilon_0} \cap \bigcap_{j=r_0+1}^{r} \Omega_{p^j,\varepsilon_0}
$$

one has

$$
\|(M - z\mathrm{Id})^K v\|_2 \ge \frac{p^{r_0/2}\, L^3\, d^2}{n^6}\, \|v\|_2.
$$

Proof. Let $v = x + y$, where $x \in \mathcal{T}_0$ and $y \in \mathcal{K}$ with $|y_1| \le x_{n_1}^*/10$. Fix $r_0 \le i \le r$ such that $x \in \mathcal{T}_{0,i}$. If $i < r$ set $m = p^i$, if $i = r$ set $m = \lceil n_1/p \rceil$. Then $x_m^* > 4dx_{pm}^*$. Fix a permutation $\sigma = \sigma_x$ of $[n]$ such that $x_i^* = |x_{\sigma(i)}|$ for $i \le n$. Let

$$
J^\ell = \sigma([m]), \quad J^r = \sigma([pm] \setminus [m]), \quad \text{and} \quad J = (J^\ell \cup J^r)^c.
$$

Then, for sufficiently large d,

$$
|J^\ell \cup J^r| = pm \le p\lceil n_1/p \rceil \le c_{3.1}\varepsilon_0 n/d \qquad \text{and} \qquad |J^r| = (p-1)|J^\ell| = (p-1)m.
$$

Denote by I_ℓ the set of rows having exactly one 1 in J^ℓ and no 1's in J^r. Lemma 3.6 implies that

$$
|I_\ell| \ge (1 - 2p\varepsilon_0)md \ge 3md/5.
$$

Let $I = (I_\ell \setminus (J^\ell \cup J^r)) \cap K$ (so that the minor $I \times (J^\ell \cup J^r)$ does not intersect the main diagonal and only rows indexed by K are considered). Since $|K^c| \le L$ and $m \ge p^{r_0} \ge 20L/d$, we have

$$
|I| \ge 3md/5 - pm - L \ge md(3/5 - p/d - 1/20) \ge md/2
$$

provided that d is large enough. By definition, for every $s \in I$ there exists $j(s) \in J^\ell$ such that

$$
\mathrm{supp} R_s \cap J^\ell = \{j(s)\}, \quad \mathrm{supp} R_s \cap J^r = \emptyset, \quad \text{and} \quad \max_{i \in J} |x_i| \le x_{mp}^*.
$$

Since $|y_1| \le x^*_{n_1}/10 \le x^*_{pm}/10$, $s \notin J^\ell \cup J^r$ (which implies $|x_s| \le x^*_{pm}$), and $j(s) \in J^\ell$ (which implies $|x_{j(s)}| \ge x^*_m > 4dx^*_{mp}$), we obtain

$$\begin{aligned}|\langle R_s(M - z\mathrm{Id}),\, (x+y)^\dagger\rangle| &= \Big| x_{j(s)} + \sum_{j \in J \cap \mathrm{supp} R_s} x_j - z x_s + d y_1 - z y_1 \Big| \\ &\ge |x_{j(s)}| - (d-1)\, x^*_{mp} - |z|\, x^*_{mp} - (d + |z|)|y_1| \ge x^*_m/2.\end{aligned}$$

Since the number of such rows is $|I| \ge md/2$ and $I \subset K$, we obtain

$$\|(M - z\mathrm{Id})^K (x+y)\|_2 \ge \sqrt{md}\, x^*_m / 2\sqrt{2}.$$

Using Lemma 4.3, we have

$$\|x+y\|_2 \le \|x\|_2 + \|y\|_2 \le \frac{n^6}{100 L^3 d^{3/2}}\, x^*_m + \sqrt{n}|y_1| \le \left(\frac{n^6}{100 L^3 d^{3/2}} + \frac{n^{1/2}}{10} \right) x^*_m,$$

which implies the desired result. □

4.3 Bounds for vectors from $\mathcal{T}_1^{\mathcal{K}} \cup \mathcal{T}_2^{\mathcal{K}}$

For the vectors from $\mathcal{T}_1^{\mathcal{K}} \cup \mathcal{T}_2^{\mathcal{K}}$ we will use the union bound together with a covering argument. We first construct nets for "normalized" versions of the sets $\mathcal{T}_i^{\mathcal{K}}$ and then provide individual probability bounds for elements of the nets. The natural normalization for "non-shifted" component would be $x^*_{n_1} = 1$, which we use for $\mathcal{T}_1^{\mathcal{K}}$. However, for individual probability bounds below and to have the same level of approximation, it is more convenient to use a slightly different normalization for $\mathcal{T}_2^{\mathcal{K}}$. We construct nets for the sets

$$\mathcal{T}_i' = \{x + y \;:\; x \in \mathcal{T}_i \;:\; x^*_{n_i} = 1 \quad \text{and} \quad y \in \mathcal{K} \;:\; |y_1| \le x^*_{n_1}/10\}, \;\; i = 1, 2.$$

Then, repeating the proof of Lemma 3.8 from [38] with slight adjustments, we obtain the following lemma.

Lemma 4.5 (Cardinalities of nets)**.** *Let $d \le n^{1/3}$ be large enough and $i = 1, 2$. There exists a set $\mathcal{N}_i = \mathcal{N}_i' + \mathcal{N}_i''$, $\mathcal{N}_i' \subset \mathbb{C}^n$, $\mathcal{N}_i'' \subset \mathcal{K}$, with the following properties. The cardinality*

$$|\mathcal{N}_i| \le \exp\left(7 n_{i+1} \ln d\right).$$

For every $u \in \mathcal{N}_i'$ one has $u^*_j = 0$ for all $j \ge n_{i+1}$. For every $x \in \mathcal{T}_i$ with $x^*_{n_i} = 1$ and every $y \in \mathcal{K}$ with $|y_1| \le x^*_{n_1}/10$ there are $u \in \mathcal{N}_i'$ and $w \in \mathcal{N}_i''$ satisfying

$$\|x - u\|_\infty \le 1/d^{3/2} \quad \textit{and} \quad \|y - w\|_\infty \le 1/d^{3/2}.$$

We now turn to the individual probability bounds where we will work in a more general setting by considering any $n \times n$ complex matrix W instead of the shift $z\mathrm{Id}$. To obtain the lower bounds on $\|(M + W)x\|_2$ for vectors x from our nets, we investigate the behavior of the inner products $\langle R_i(M+W), x^\dagger \rangle$. One of the tools that we use is the Lévi concentration function for $\langle R_i(M+W), x^\dagger \rangle$. To estimate this function we, in particular, will use Theorem 3.1 for $2m$ columns of M corresponding to the m biggest and m smallest (in modulus) coordinates of x, where $m = n_1$ or $m = n_2$. The main difficulty in this scheme comes from the restriction $2m \le c_{3.1}\varepsilon n/d$ in Theorem 3.1, which is not satisfied for $m = n_2$. To resolve this problem we split the set of $2m$ columns into smaller subsets of columns of size at most $c_{3.1}\varepsilon n/d$, and create independent random variables corresponding to this splitting and such that their sum is $\langle R_i(M+W), x^\dagger \rangle$ up to a constant. Then we apply Proposition 2.2, allowing to deal with Lévy concentration function for sums of independent random variables.

We first describe subdivisions of $\mathcal{M}_{n,d}$ needed for our construction. Given $J \subset [n]$ and $M \in \mathcal{M}_{n,d}$ denote

$$I(J, M) = \{i \le n \,:\, |\text{supp} R_i(M) \cap J| = 1\}$$

(*cf.*, the definition of $I^\ell(M)$, $I^r(M)$ before Lemma 3.6 – clearly, if we split J into J^ℓ and J^r, then $I(J, M) = I^\ell(M) \cup I^r(M)$). By $\mathcal{M}_J$ we denote the set of $n \times |J|$ matrices obtained from matrices $M \in \mathcal{M}_{n,d}$ by taking columns with indices in J, i.e.,

$$\mathcal{M}_J = \{V = \{v_{ij}\}_{i \le n, j \in J} \,:\, \exists M \in \mathcal{M}_{n,d} \text{ such that } \forall i \le n \, \forall j \in J \;\; v_{ij} = \mu_{ij}\}.$$

Now we fix $q_0 \le n$ and a partition J_0, J_1, ..., J_{q_0} of $[n]$. Given subsets I_1, ..., I_{q_0} of $[n]$ and $V \in \mathcal{M}_{J_0}$, denote $\mathcal{I} = (I_1, \ldots, I_{q_0})$ and consider the class

$$\mathcal{F}(\mathcal{I}, V) = \{M \in \mathcal{M}_{n,d} \,:\, \forall q \in [q_0] \quad I(J_q, M) = I_q \;\text{ and }\; \forall i \le n \, \forall j \in J_0 \;\; \mu_{ij} = v_{ij}\}$$

(depending on the choice of $\mathcal{I}$ such a class can be empty). In words, we fix the columns indexed by J_0 and for each $q \in [q_0]$ we fix the rows having exactly one 1 in columns indexed by J_q. Then $\mathcal{M}_{n,d}$ is the disjoint union of classes $\mathcal{F}(\mathcal{I}, V)$ over all $V \in \mathcal{M}_{J_0}$ and all $\mathcal{I} \in (\mathcal{P}([n]))^{q_0}$, where $\mathcal{P}(\cdot)$ denotes the power set.

Furthermore, given V and $\mathcal{I}$ as above, we split each class $\mathcal{F}(\mathcal{I}, V)$ into smaller equivalence classes using the following equivalence relation. Fix $i \le n$ and $A \subset [q_0]$. Denote $A_0 := \{0\} \cup ([q_0] \setminus A)$. We say that two matrices $M, \widetilde{M} \in \mathcal{F}(\mathcal{I}, V)$ are equivalent if

$$\forall s < i \;\forall j \le n \qquad \mu_{sj} = \tilde{\mu}_{sj},$$

$$\forall s \le n \;\forall j \in J' := \bigcup_{q \in A_0} J_q \qquad \mu_{sj} = \tilde{\mu}_{sj},$$

and

$$\forall s \le n \;\forall q \in A \qquad \sum_{j \in J_q} \mu_{sj} = \sum_{j \in J_q} \tilde{\mu}_{sj}.$$

The collection of equivalence classes corresponding to this relation will be denoted by

$$\mathcal{H} = \mathcal{H}(\mathcal{F}(\mathcal{I}, V), i, A), \quad \text{ in particular } \quad \mathcal{F}(\mathcal{I}, V) = \bigcup_{H \in \mathcal{H}} H.$$

Note that for matrices in a given class H, the rows R_1, ..., R_{i-1} are fixed and every block $[n] \times J_q$ has a prescribed sum in each row, thus, in a sense, these blocks are independent of each other on H.

Finally, given a vector $x \in \mathbb{C}^n$, an index $i \le n$, a class $H \in \mathcal{H}$ (in particular, $V, \mathcal{I}, i, A$ are fixed), and $q \in A$, we introduce a random variable ξ_q on H by

$$\xi_q = \xi_q(M) := \sum_{j \in J_q} \mu_{ij} x_j.$$

In words, ξ_q represents the dot product of x with the restriction of the i-th row to J_q. Later we use this construction in the case when $i \in I_q$ for all $q \in A$, that is for a specific choice of parameters defining our classes (recall here that for $M \in H$, $|\text{supp} R_i(M) \cap J_q| = 1$ provided that $i \in I_q$). As we have already mentioned, by construction, for matrices in the class H every block $[n] \times J_q$ has a prescribed sum in each row, therefore the random variables ξ_q, $q \in A$, are independent. Thus, using that for a fixed matrix $W = \{w_{ij}\}$ and a fixed constant vector $y \in \mathcal{K}$, the function

$$\xi' = \xi'(M) := \sum_{j \in J'} \mu_{ij} x_j + \sum_{j=1}^{n} w_{ij} x_j + y_1 d + y_1 \sum_{j=1}^{n} w_{ij}$$

is a constant on H, we may apply Proposition 2.2 (in which we identify $\mathbb{C}$ with $\mathbb{R}^2$) to

$$\left|\left\langle R_i(M+W),(x+y)^\dagger\right\rangle\right| = \Big|\sum_{q\in A}\xi_q + \xi'\Big|$$

with some $\alpha > 0$ satisfying $\mathcal{Q}(\xi_q, 1/3) \le \alpha$ for every $q \in A$. This gives

$$\mathbb{P}\left(\left|\left\langle R_i(M+W),(x+y)^\dagger\right\rangle\right| \le 1/3\right) \le \frac{C_0}{\sqrt{(1-\alpha)|A|}}, \tag{9}$$

where C_0 is a positive absolute constant.

We are ready now to estimate individual probabilities.

Lemma 4.6 (Individual probability). *There exist absolute constants $C, C' > 1 > c_1 > 0$ such that the following holds. Let $C < d < n$, $K \subset [n]$, $\varepsilon \in [\varepsilon_0, 0.01]$. Set $m_0 = \lfloor c_{3.1}\varepsilon n/(2d)\rfloor$ and let m_1 and m_2 be such that $m_1 < m_2 \le n - m_1$. Assume that $x \in \mathbb{C}^n$ satisfies*

$$x^*_{m_1} > 2/3 \quad \text{and} \quad x^*_i = 0 \quad \text{for every} \quad i > m_2.$$

Let W be a complex $n\times n$ matrix, $y \in \mathcal{K}$, and denote $m = \min(m_0, m_1)$ and

$$E = E(x) = \left\{M \in \mathcal{M}_{n,d} \,:\, \|(M+W)^K(x+y)\|_2 \le \sqrt{c_1 m d}\right\}.$$

Then if $m_1 \le m_0$ and $|K^c| \le 3m_1 d/5$

$$\mathbb{P}(E \cap \Omega_{2m_1,\varepsilon}) \le (5/6)^{m_1 d/2},$$

if $m_1 > C' m_0$, $\varepsilon = 0.01$, and $|K^c| \le 3m_0 d/5$

$$\mathbb{P}(E \cap \Omega_{2m_0,\varepsilon}) \le \left(\frac{Cn}{m_1 d}\right)^{m_0 d/4}.$$

Remark 4.7. *We apply this lemma below twice: first with $m_1 = n_1 < m_0$, $m_2 = n_2$, $\varepsilon = 0.01$, obtaining*

$$\mathbb{P}(E \cap \Omega_{2n_1,0.01}) \le (5/6)^{n_1 d/2};$$

then with $m_1 = n_2 > m_0$, $m_2 = n_3$, $\varepsilon = 0.01$, obtaining

$$\mathbb{P}(E \cap \Omega_{2m_0,0.01}) \le \left(\frac{Cn}{dn_2}\right)^{0.01\,c_{3.1}n/8} \le \left(\frac{C_1}{d}\right)^{cn},$$

where $C_1 = 8C^3$ and $c = c_{3.1}n/2400$ are positive absolute constants.

Proof. Fix $\gamma = 3md/5n$. Fix $x \in \mathbb{C}^n$ and $y \in \mathcal{K}$ satisfying the condition of the lemma. Let σ be a permutation of $[n]$ such that $x^*_i = |x_{\sigma(i)}|$ for all $i \le n$. Denote $q_0 = m_1/m$ and without loss of generality assume that either $q_0 = 1$ or that q_0 is a large enough integer. Let $J_1^\ell, J_2^\ell, \dots, J_{q_0}^\ell$ be a partition of $\sigma([m_1])$ into sets of cardinality m. Let $J_1^r, J_2^r, \dots, J_{q_0}^r$ be a partition of $\sigma([n-m_1+1, n])$ into sets of cardinality m. Denote

$$J_q := J_q^\ell \cup J_q^r \quad \text{for} \quad q \in [q_0] \qquad \text{and} \qquad J_0 := [n] \setminus \bigcup_{q=1}^{q_0} J_q.$$

Then J_0, J_1, ..., J_{q_0} is a partition of $[n]$, which we fix in this proof. Let $M \in \Omega_{2m,\varepsilon}$. For every pair J_q^ℓ, J_q^r, let the sets $I_q^\ell(M)$ and $I_q^r(M)$ be defined as before Lemma 3.6 and let $I_q = I_q(M) = I_q^\ell(M) \cup I_q^r(M)$. Since

$$|J_q| = 2m \le 2m_0 \le c_{3.1}\varepsilon n/d,$$

Lemma 3.6 implies that

$$|I_q^\ell(M)|,\, |I_q^r(M)| \in [(1-4\varepsilon)md,\, md],$$

in particular,

$$|I_q| \in [2(1-4\varepsilon)md,\, 2md]. \tag{10}$$

Now we split $\mathcal{M}_{n,d}$ into a disjoint union of classes $\mathcal{F}(\mathcal{I}, V)$ defined at the beginning of this subsection with $V \in \mathcal{M}_{J_0}$ and $\mathcal{I} = (I_1, \ldots, I_q)$ and note that $\Omega_{2m,\varepsilon} \cap \mathcal{F}(\mathcal{I}, V) \neq \emptyset$ implies that I_q satisfies (10) for every q. Thus, to prove our lemma it is enough to prove a uniform upper bound for such classes, indeed,

$$\mathbb{P}(E(x) \cap \Omega_{2m,\varepsilon}) \leq \max \mathbb{P}(E(x) \cap \Omega_{2m,\varepsilon} \,|\, \mathcal{F}(\mathcal{I}, V)) \leq \max \mathbb{P}(E(x) |\, \mathcal{F}(\mathcal{I}, V))$$

where the first maximum is taken over all $\mathcal{F}(\mathcal{I}, V)$ with $\Omega_{2m,\varepsilon} \cap \mathcal{F}(\mathcal{I}, V) \neq \emptyset$ and the second maximum is taken over $\mathcal{F}(\mathcal{I}, V)$ with I_q's satisfying (10).

Fix such a class $\mathcal{F}(\mathcal{I}, V)$ and denote the uniform probability on it just by $\mathbb{P}_{\mathcal{F}}$, that is

$$\mathbb{P}_{\mathcal{F}}(\cdot) = \mathbb{P}(\cdot \,|\, \mathcal{F}(\mathcal{I}, V)).$$

Let

$$I := \bigcup_{q=1}^{q_0} I_q.$$

Note that $|I| \leq 2q_0 md$. We first show that the set of i's belonging to many I_q's is rather large. More precisely, given $i \in [n]$ denote

$$A_i = \{q \in [q_0] \,:\, i \in I_q\}, \quad I_{00} = \{i \in I \,:\, |A_i| \geq \gamma q_0\}, \quad \text{and} \quad I_0 = I_{00} \cap K.$$

Then, using bounds on cardinalities of I_q's, one has

$$2(1-4\varepsilon)mdq_0 \leq \sum_{q=1}^{q_0} |I_q| = \sum_{i=1}^{n} |A_i| \leq |I_{00}|q_0 + (n - |I_{00}|)\gamma q_0 \leq |I_{00}|q_0 + n\gamma q_0.$$

Since $\varepsilon \leq 0.01$, $\gamma = 3md/(5n)$ and $|K^c| \leq 3md/5$, we get

$$|I_0| \geq |I_{00}| - |K^c| \geq 2(1-4\varepsilon)md - 6md/5 \geq 2md/3.$$

Without loss of generality we assume that $I_0 = \{1, 2, \ldots |I_0|\}$ and only consider the first $k := \lceil 2md/3 \rceil$ indices from it. Then $[k] \subset I_0 \subset K$.

Now, by definition, for matrices $M \in E(x)$ we have

$$\|(M+W)^K(x+y)\|_2^2 = \sum_{i \in K} |\left\langle R_i(M+W), (x+y)^\dagger \right\rangle|^2 \leq c_1\, md.$$

Therefore there are at most $9c_1 md$ rows with $|\langle R_i(M+W), (x+y)^\dagger \rangle| \geq 1/3$. Hence,

$$|\{i \leq k \,:\, |\langle R_i(M+W), (x+y)^\dagger \rangle| < 1/3\}| \geq 2md/3 - 9c_1 md \geq 2(1-14c_1)md/3$$

(we used that $I_0 \subset K$). Let $k_0 := \lceil 2(1-14c_1)md/3 \rceil$ and for every $i \leq k$ denote

$$\Omega_i := \{M \in \mathcal{F}(\mathcal{I}, V) \,:\, |\left\langle R_i(M+W), (x+y)^\dagger \right\rangle| < 1/3\} \quad \text{and} \quad \Omega_0 = \mathcal{F}(\mathcal{I}, V).$$

Then

$$\mathbb{P}_{\mathcal{F}}(E(x)) \leq \sum_{\substack{B \subset [k] \\ |B| = k_0}} \mathbb{P}_{\mathcal{F}}\Big(\bigcap_{i \in B} \Omega_i\Big) \leq \binom{k}{k_0} \max_{\substack{B \subset [k] \\ |B| = k_0}} \mathbb{P}_{\mathcal{F}}\Big(\bigcap_{i \in B} \Omega_i\Big).$$

Without loss of generality we assume that the maximum above is attained at $B=[k_0]$. Then

$$\mathbb{P}_{\mathcal{F}}(E(x)) \leq (1/c_1)^{10c_1 md} \prod_{i=1}^{k_0} \mathbb{P}_{\mathcal{F}}(\Omega_i |\, \Omega_1 \cap \ldots \cap \Omega_{i-1}). \tag{11}$$

Next we estimate the factors in the product. Fix i and $A_i = \{q \,:\, i \in I_q\}$. Since $i \in I_0$, we have $|A_i| \geq \gamma q_0$. Consider the splitting of $\mathcal{F}(\mathcal{I}, V)$ into classes $H \in \mathcal{H} = \mathcal{H}(\mathcal{F}(\mathcal{I}, V), i, A_i)$ as described before the statement of the lemma and let $\mathbb{P}_H$ denote the uniform probability on a class H, i.e., $\mathbb{P}_H(\cdot) = \mathbb{P}(\cdot|H)$. Since in every class H all matrices have their first $i-1$ rows fixed, for every H the intersection $H_i := H \cap \Omega_1 \cap \ldots \cap \Omega_{i-1}$ is either H or $\emptyset$. Thus

$$\mathbb{P}_{\mathcal{F}}(\Omega_i |\, \Omega_1 \cap \ldots \cap \Omega_{i-1}) \leq \max_{H: H_i \neq \emptyset} \mathbb{P}_H(\Omega_i).$$

Fix H such that $H_i \neq \emptyset$ and consider the random variables ξ_q, $q \in A_i$, defined above. Then by (9) we have

$$\mathbb{P}_H(\Omega_i) = \mathbb{P}_H(|\left\langle R_i(M+W), (x+y)^\dagger \right\rangle| \leq 1/3) \leq \frac{C_0 \alpha}{\sqrt{(1-\alpha)|A_i|}} \leq \frac{C_0 \alpha}{\sqrt{(1-\alpha)\gamma q_0}}$$

where $\alpha = \max_{q \in A_i} \mathcal{Q}(\xi_q(M), 1/3)$. Note that in the case $q_0 = 1$ we just have

$$\mathbb{P}_H(\Omega_i) = \alpha = \mathcal{Q}(\xi_1(M), 1/3).$$

Thus it remains to estimate $\mathcal{Q}(\xi_q, 1/3)$ for $q \in A_i$. Fix $q \in A_i$, so that $i \in I_q$. Recall that, by construction, the intersection of the support of $R_i(M)$ with J_q is a singleton. Denote the corresponding index by $j(q)$. Then

$$\xi_q = \xi_q(M) = \sum_{j \in J_q} \mu_{ij} x_j = x_{j(q)}$$

and note that $|x_{j(q)}| > 2/3$ whenever $j(q) \in J_q^\ell$ and $x_{j(q)} = 0$ whenever $j(q) \in J_q^r$. Denote

$$H^\ell = \{M \in H \,:\, j(q) \in J_q^\ell\} \quad \text{ and } \quad H^r = \{M \in H \,:\, j(q) \in J_q^r\}.$$

Since $|\xi_q|$ is either larger than $2/3$ or equals 0 we observe that

$$\mathcal{Q}(\xi_q(M), 1/3) \leq \max\{\mathbb{P}_H(H^\ell),\, \mathbb{P}_H(H^r)\}.$$

Claim 4.8. *For $i \leq \lceil 2md/3 \rceil$ one has*

$$\max\{\mathbb{P}_H(H^\ell),\, \mathbb{P}_H(H^r)\} \leq 4/5.$$

Combining the probability estimates starting with (11), using that $\gamma = 3md/5n$, and applying Claim 4.8, we obtain in the case $q_0 = m_1/m \geq C'$,

$$\begin{aligned}\mathbb{P}_{\mathcal{F}}(E(z)) &\leq \left(\frac{1}{c_1}\right)^{10c_1 md} \left(\frac{4C_0}{\sqrt{5\gamma q_0}}\right)^{2(1-14c_1)md/3} \\ &= \left(\frac{1}{c_1}\right)^{10c_1 md} \left(\frac{4C_0\sqrt{n}}{\sqrt{3m_1 d}}\right)^{2(1-14c_1)md/3} \leq \left(\frac{C_1 n}{m_1 d}\right)^{md/4},\end{aligned}$$

provided that c_1 is small enough and C' is large enough. In the case $q_0 = 1$ we have

$$\mathbb{P}_{\mathcal{F}}(E(z)) \leq \left(\frac{1}{c_1}\right)^{10c_1 md} \left(\frac{4}{5}\right)^{2(1-14c_1)md/3} \leq \left(\frac{5}{6}\right)^{md/2}$$

provided that c_1 is small enough. This completes the proof. □

Proof of Claim 4.8. We show the bound for $\mathbb{P}_H(H^\ell)$, the other bound is similar. Note that for matrices in H we have $I(J_q, M) = I_q$ and I_q satisfies (10). Since $|J_q^\ell| = |J_q^r| = m$, we observe that on H one has

$$|I_q^\ell(M)| \geq |I_q| - md \geq (1-8\varepsilon)md \quad \text{ and } \quad |I_q^r(M)| \leq md.$$

To compare cardinalities, define a relation $R \in H^\ell \times H^r$ by $(M,\, M') \in R$ iff $M \in H^\ell$, $M' \in H^r$, and M' can be obtained from M by a simple switching in

$$(I_q \setminus [i-1]) \times J_q$$

(note that the i-th row is necessarily involved in the switching). It is easy to check that for every $M \in H^\ell$ and every $M' \in H^r$ one has

$$|R(M)| = |I_q^r(M) \setminus [i-1]| \quad \text{ and } \quad |R^{-1}(M')| = |I_q^\ell(M') \setminus [i-1]|,$$

hence $|R(M)| \geq (1-8\varepsilon)md - i + 1$ and $|R^{-1}(M')| \leq md$. Claim 2.1 yields

$$|H^\ell|/|H^r| \leq \frac{md}{(1-8\varepsilon)md - i + 1} \leq \frac{1}{1/3 - 8\varepsilon}.$$

Therefore,

$$|H|/|H^\ell| = (|H^\ell| + |H^r|)/|H^\ell| \geq 4/3 - 8\varepsilon,$$

which completes the proof since $\varepsilon \leq 0.01$. $\square$

4.4 Proof of Theorem 4.1

We are ready to complete the proof.

Proof of Theorem 4.1. Recall that d is large enough, $\varepsilon_0 = \sqrt{(\ln d)/d}$, $p = \lfloor 1/(5\varepsilon_0) \rfloor$, and let $\varepsilon = 0.01$. Denote $m = m_0 = \lfloor c_{3.1}\varepsilon n/(2d) \rfloor$ and note that $n/d^{3/2} \leq n_1 \leq m_0 \leq n_2$ and that $|K^c| \leq L \leq 3n_1 d/5 \leq 3m_0 d/5$. Below we deal with matrices from

$$\Omega_0 = \Omega_{2n_1,\varepsilon} \cap \Omega_{2m_0,\varepsilon} \cap \Omega_{n_1,\varepsilon_0} \cap \bigcap_{j=r_0}^{r} \Omega_{p^j,\varepsilon_0}.$$

If $v \in \mathcal{T}_0^{\mathcal{K}}$ and $M \in \Omega_0$ then Lemma 4.4 implies that

$$\|(M - z\mathrm{Id})^K v\|_2 \geq \frac{L^3\, d^2}{n^6}\, \|v\|_2.$$

We turn now to the case $v \in \mathcal{T}_i^{\mathcal{K}}$ for $i = 1, 2$. Let

$$\mathcal{E}_i := \Big\{ M \in \mathcal{M}_{n,d} \,:\, \exists\, v \in \mathcal{T}_i^{\mathcal{K}} \text{ such that } \|(M - z\mathrm{Id})^K v\|_2 \leq \frac{\sqrt{c_1 md}}{2\, b_i}\, \|v\|_2 \Big\},$$

where c_1 is the constant from Lemma 4.6, $b_1 = n^6/(L^3 d^{3/2})$, and $b_2 = d^{3/2} b_1 = n^6/L^3$. For a matrix $M \in \mathcal{E}_i$ there exists $v = v(M) \in \mathcal{T}_i^{\mathcal{K}}$

$$\|(M - z\mathrm{Id})^K v)\|_2 \leq \frac{\sqrt{c_1 md}}{2\, b_i}\, \|v\|_2.$$

Write $v = x + y$, where $x \in \mathcal{T}_i$ and $y \in \mathcal{K}$ such that $|y_1| \leq x^*_{n_1}/10$. Normalize v so that $x^*_{n_i} = 1$ (that is, $v \in \mathcal{T}_i'$). By Lemma 4.3 we have

$$\|v\|_2 = \|x + y\|_2 \leq \frac{n^6}{100 L^3 d^{3/2}}\, x^*_{n_1} + \frac{\sqrt{n}\, x^*_{n_1}}{10} \leq \frac{n^6}{L^3 d^{3/2}}\, x^*_{n_1} \leq b_i x^*_{n_i} = b_i.$$

Let $\mathcal{N}_i = \mathcal{N}_i' + \mathcal{N}_i''$ be the net constructed in Lemma 4.5. Then there exist $u \in \mathcal{N}_i'$ with $u^*_{n_i} \geq 1 - 1/2d^{3/2} > 2/3$ and $u^*_j = 0$ for $j > n_{i+1}$, and $w \in \mathcal{N}_i'' \subset \mathcal{K}$, such that for large enough d,

$$\|v - (u + w)\|_2 \leq \sqrt{n}\left(\|x - u\|_\infty + \|y - w\|_\infty\right) \leq 2\sqrt{n}\, d^{-3/2} \leq \frac{\sqrt{c_1 md}}{4d}.$$

Therefore, using that $\|M\| = d$ and $|z| \leq d$, we obtain that for every matrix $M \in \mathcal{E}_i$ there exist $u = u(M) \in \mathcal{N}_i'$ and $w = w(M) \in \mathcal{N}_i'' \subset \mathcal{K}$ with

$$\|(M - z\mathrm{Id})^K (u + w)\|_2 \leq \|(M - z\mathrm{Id})^K v\|_2 + (\|M\| + |z|)\|v - (u + w)\|_2 \leq \sqrt{c_1 md}.$$

Using union bound, our choice of n_1, n_2, n_3, Lemma 4.5, and Lemma 4.6 twice – first with $m_1 = n_1 < m_0$, $m_2 = n_2$, $\varepsilon = 0.01$, then with $m_1 = n_2 > m_0$, $m_2 = n_3$, $\varepsilon = 0.01$ (see Remark 4.7), we obtain for small enough a_3 and large enough d,

$$\mathbb{P}\left(\mathcal{E}_1 \cap \Omega_0\right) \leq \exp\left(-(n_1 d/2)\ln(6/5) + 7n_2 \ln d\right) \leq \exp\left(-n_1 d/20\right)$$

and

$$\mathbb{P}\left(\mathcal{E}_2 \cap \Omega_0\right) \leq \exp\left(-cn\ln d + 7n_3 \ln d\right) \leq \exp\left(-c_0 n \ln d\right),$$

where $c_0 > 0$ is an absolute constant.

Combining all cases we obtain that for $x \in \mathcal{T}_{\mathcal{K}}$ one has $\|(M - z\mathrm{Id})^K x\|_2 \leq \beta\|x\|_2$, where

$$\beta := \min\left(\frac{L^3\, d^2}{n^6},\ \frac{L^3\sqrt{c_1 m_0 d}}{2n^6}\right) \geq \frac{L^3 d^2}{n^6}\ \min\left(1,\ \frac{\sqrt{c_1 c_{3.1}\varepsilon n/2}}{2d^2}\right) \geq \frac{L^3 d}{n^6}$$

with probability at most

$$p_0 := \mathbb{P}\left(\Omega_0^c\right) + \exp\left(-n_1 d/20\right) + \exp\left(-c_0 n \ln d\right).$$

We now estimate the probability p_0. Using Theorem 3.1 and that $n_1 \geq n/d^{3/2}$, $\varepsilon_0^2 d = \ln d$, $\varepsilon = 0.01$, we obtain for large enough d,

$$p_1 := \mathbb{P}\left(\Omega^c_{2n_1,\varepsilon}\right) + \mathbb{P}\left(\Omega^c_{2n_1,\varepsilon_0}\right) + \exp\left(-n_1 d/20\right)$$

$$\leq \exp\left(-\varepsilon^2 dn_1/4\right) + \exp\left(-\varepsilon_0^2 dn_1/8\right) + \exp\left(-n_1 d/20\right) \leq \exp\left(-n/d^{3/2}\right);$$

$$p_2 := \mathbb{P}\left(\Omega^c_{2m_0,\varepsilon}\right) + \exp\left(-c_0 n \ln d\right) \leq \exp\left(-\varepsilon^2 dm_0/4\right) + \exp\left(-c_0 n \ln d\right) \leq \exp\left(-c_3 n\right);$$

and

$$p_3 := \sum_{i=r_0}^{r} \mathbb{P}\Big(\Omega^c_{p^j,\varepsilon_0}\Big) \leq \sum_{i=r_0}^{r} \exp\Big(-\frac{p^i \ln d}{8}\ln\Big(\frac{n}{p^i d^{3/2}}\Big)\Big)$$
$$\leq \exp\Big(-\frac{p^{r_0}\ln d}{9}\ln\Big(\frac{n}{p^{r_0} d^{3/2}}\Big)\Big),$$

where c_3 is a positive absolute constant. Since $r_0 \geq 0$ and $n \geq d^3$, we have

$$p_3 \leq \exp\left(-\frac{\ln d}{9}\ln\Big(\frac{n}{d^{3/2}}\Big)\right) \leq \exp\left(-\frac{(\ln d)\ln n}{18}\right).$$

Since $p^{r_0} \geq 20L/d$, we also have $p_3 \leq \exp(-2L(\ln d/d))$. Since $p_0 \leq p_1 + p_2 + p_3$, the desired estimate follows.

$\square$

4.5 Almost constant vectors

Given $\theta > 0$, we introduce a class of almost constant vectors by

$$\mathcal{B}(\theta) = \{x \in \mathbb{C}^n \,:\, \exists \lambda \in \mathbb{C} \text{ such that } |\{i \leq n \,:\, |x_i - \lambda| \leq \theta\, x^*_{n_3}\}| > n - n_3\}.$$

Note that this class slightly differs from the class considered in [38] – there we compared the error in terms of $\|x\|_2$ instead of $x^*_{n_3}$.

Remark 4.9. *Let $x \in \mathcal{B}(\theta)$. Fix a permutation $\sigma = \sigma_x$ of $[n]$ such that $x^*_i = |x_{\sigma(i)}|$ for $i \leq n$. Fix $\lambda_0 = \lambda_0(x) \in \mathbb{C}$ such that the cardinality of*

$$J_1 := \{i \leq n \,:\, |x_i - \lambda_0| \leq \theta\, x^*_{n_3}\}$$

is at least $n - n_3 + 1$. Then there exist positive integers k, ℓ with $k \leq n_3 < \ell$ such that $\sigma(k), \sigma(\ell) \in J_1$ and

$$x^*_k - \theta x^*_{n_3} \leq |x_{\sigma(k)}| - |x_{\sigma(k)} - \lambda_0| \leq |\lambda_0| \leq |\lambda_0 - x_{\sigma(\ell)}| + |x_{\sigma(\ell)}| \leq \theta x^*_{n_3} + x^*_\ell,$$

which implies

$$(1-\theta)x^*_{n_3} \leq |\lambda_0| \leq (1+\theta)x^*_{n_3}. \tag{12}$$

Define $n_0 = \lfloor n/16d \rfloor$. Given $t > 0$, consider the following class of vectors

$$S(t) := \{x \in \mathbb{C}^n \,:\, 0 < x^*_{n_0} \leq t x^*_{n_3}\}.$$

The proof of the next lemma is similar to that of Theorem 3.1 from [38]. We provide it at the end of the section for the sake of completeness.

Lemma 4.10. *Let $\theta \in (0, 1/20]$ and $t \geq 12$ be such that $a_3 t \leq 1/100$. Let $K \subset [n]$ with $|K^c| \leq n/4$ and $z \in \mathbb{C}$ with $|z| \leq d/5$. Then for every $x \in \mathcal{B}(\theta) \cap S(t)$ and every $M \in \mathcal{M}_{n,d}$ one has*

$$\|(M - z\mathrm{Id})^K x\|_2 \geq \frac{d\sqrt{n}}{2\sqrt{2}}\, x^*_{n_3}.$$

We also need the following simple lemma about almost constant vectors not covered by Lemma 4.10.

Lemma 4.11. *Let $d \geq 3$, $0 < \theta \leq 10/d^3$ and $t \geq 12$ be such that $a_3 t \leq 1/100$. Then every $x \in \mathcal{B}(\theta) \setminus S(t)$ can be represented as $x = w + y$ with $w \in \mathcal{T}$ and $y \in \mathcal{K}$ with $|y_1| \leq w^*_{n_1}/10$.*

Proof. Fix $x \in \mathcal{B}(\theta) \setminus S(t)$. Then $x^*_{n_0} > t x^*_{n_3}$. Let σ, J_1, and λ_0 be as in Remark 4.9. Consider $y = (\lambda_0, \lambda_0, ..., \lambda_0) \in \mathcal{K}$ and $w = x - y$. Since $x \notin S(t)$ and by (12), for every $i \leq n_0$ we have

$$|w_{\sigma(i)}| \geq |x_{\sigma(i)}| - |\lambda_0| \geq x^*_{n_0} - (1+\theta)x^*_{n_3} > (t - 1 - \theta)x^*_{n_3} > 10 x^*_{n_3}.$$

This implies $w^*_{n_0} > 10 x^*_{n_3}$. On the other hand, for every $i \in J_1$, one has $|w_i| \leq \theta x^*_{n_3}$. Since $|J_1| > n - n_3$, this implies $w^*_{n_3} \leq \theta x^*_{n_3}$. Using that $\theta \leq 10/d^3$, we obtain

$$w^*_{n_1} \geq w^*_{n_0} > d^3 w^*_{n_3},$$

which shows $w \in \mathcal{T}$.

Using again that $x \notin S(t)$ and the inequality (12), we observe that for every $i \leq n_0$,

$$|w_{\sigma(i)}| \geq |x_{\sigma(i)}| - |\lambda_0| \geq |x_{\sigma(n_0)}| - |\lambda_0| \geq t x^*_{n_3} - |\lambda_0| > (t/(1+\theta) - 1)|\lambda_0| > 10|\lambda_0|,$$

which implies $|y_1| = |\lambda_0| \leq w^*_{n_1}/10$ and completes the proof. □

As a consequence of Theorem 4.1 and Lemmas 4.10, 4.11, and 4.3, we obtain the main theorem of this section.

Theorem 4.12. *Let $a_3 \le 1/1200$, $d \ge 1$ be large enough, $n \ge d^3$, $1 \le L \le n/d^3$, and $0 < \theta \le 10/d^3$. Let $K \subset [n]$ with $|K^c| \le L$ and $z \in \mathbb{C}$ with $|z| \le d/5$. Then with probability at least*

$$1 - \min(\exp(-L/d), \exp(-(\ln d)(\ln n)/20),$$

one has that for every $x \in (\mathcal{B}(\theta) \setminus \mathcal{T}_3^{\mathcal{K}}) \cup \mathcal{T}_{\mathcal{K}}$

$$\|(M - z\mathrm{Id})^K x\|_2 \ge \frac{L^3 d}{n^6} \|x\|_2.$$

Proof. Fix $t = 12$. Fix $x \in \mathcal{B}(\theta) \setminus \mathcal{T}_3^{\mathcal{K}}$. If $x \in \mathcal{T}_{\mathcal{K}}$ then the result follows by Theorem 4.1. Therefore we assume that $x \notin \mathcal{T}_{\mathcal{K}} \cup \mathcal{T}_3^{\mathcal{K}}$. Then, in particular, $x \notin \mathcal{T}_0 \cup \mathcal{T}_3$ and $x \notin \mathcal{T}_1 \cup \mathcal{T}_2$, hence, by Lemma 4.3 we have

$$x_{n_3}^* \ge x_{n_1}^*/d^3 \ge \frac{100L^3}{n^6 d^{3/2}} \|x\|_2.$$

Since $n \ge d^3$, this and Lemma 4.10 implies the case when $x \in S(t)$. Note that Lemma 4.11 says that

$$\mathcal{B}(\theta) \setminus S(t) \subset \mathcal{T}_{\mathcal{K}} \cup \mathcal{T}_3^{\mathcal{K}},$$

therefore we are done. □

Proof of Lemma 4.10. Since $x \in S(t)$, we have $x_{n_3}^* \ne 0$. Let σ, J_1, and λ_0 be as in Remark 4.9 and set

$$J_2 = \sigma([n_0]) \setminus J_1, \quad J_3 = \sigma([n_3]) \setminus (J_1 \cup J_2), \quad \text{and} \quad J_4 = [n] \setminus (J_1 \cup \sigma([n_3])).$$

Then $|J_3|, |J_4| \le n_3$, $[n] = J_1 \cup J_2 \cup J_3 \cup J_4$, and

$$\forall j \in J_4 \;\; |x_j| \le x_{n_3}^* \quad \text{and} \quad \forall j \in J_3 \;\; |x_j| \le x_{n_0}^* \le t x_{n_3}^* \tag{13}$$

Now, given a matrix $M \in \mathcal{M}_{n,d}$, consider

$$I_2 = \{i \le n \, : \, \mathrm{supp} R_i(M) \cap J_2 \ne \emptyset\} \text{ and } I_\ell = \{i \le n \, : \, |\mathrm{supp} R_i(M) \cap J_\ell| \ge 16 n_3 d/n\},$$

for $\ell = 3, 4$. Since $M \in \mathcal{M}_{n,d}$, we have $|I_2| \le d\, n_0$ and $(16 n_3 d/n)|I_\ell| \le d|J_\ell|$, hence

$$|I_2| \le n/16 \qquad \text{and} \qquad |I_\ell| \le n/16 \;\text{ for }\; \ell = 3, 4.$$

Set $I := [n] \setminus (I_2 \cup I_3 \cup I_4 \cup \sigma([n_3]) \cup K^c)$. Then for small enough a_3,

$$|I| \ge n - 3n/16 - n_3 - n/4 \ge n/2 \quad \text{and} \quad \forall i \in I \;\; |x_i| \le x_{n_3}^* \le |\lambda_0|/(1-\theta).$$

Moreover, for every $i \in I$, denote $\overline{J_\ell} = \overline{J_\ell}(i) = J_\ell \cap \mathrm{supp} R_i(M)$ for $1 \le \ell \le 4$, and note that $\overline{J_2} = \emptyset$ since $i \notin I_2$. Using the triangle inequality, we observe for every $i \in I$,

$$|\langle R_i(M - z\mathrm{Id}), x^\dagger \rangle| \ge \Big| \sum_{j \in \overline{J_1}} x_j \Big| - \sum_{j \in \overline{J_3}} |x_j| - \sum_{j \in \overline{J_4}} |x_j| - |z x_i|.$$

We estimate each of the terms on the right hand side separately. By the definition of J_1, we have

$$\Big| \sum_{j \in \overline{J_1}} x_j \Big| \ge |\lambda_0|\, |\overline{J_1}| - \sum_{j \in \overline{J_1}} \Big| x_j - \lambda_0 \Big| \ge |\overline{J_1}|\, (|\lambda_0| - \theta x_{n_3}^*) \ge (d - 32 n_3 d/n)\, (1 - 2\theta)\, x_{n_3}^*,$$

where for the last inequality we used (12) and that for $i \notin I_2 \cup I_3 \cup I_4$ one has

$$|\overline{J_1}| = d - |\overline{J_2}| - |\overline{J_3}| - |\overline{J_4}| \ge d - 32 n_3 d/n.$$

Using (13),

$$\sum_{j\in\overline{J_3}}|x_j|+\sum_{j\in\overline{J_4}}|x_j|\leq|\overline{J_3}|\,x^*_{n_0}+|\overline{J_4}|\,x^*_{n_3}\leq 16(1+t)n_3dx^*_{n_3}/n.$$

Putting together the above estimates, we obtain for large enough d

$$\begin{aligned}|\langle R_i(M-z\mathrm{Id}),x^\dagger\rangle|&\geq((d-32n_3d/n)(1-2\theta)-16(1+t)n_3d/n-|z|)\,x^*_{n_3}\\&\geq(1-2\theta-16a_3(3+t)-|z|/d)\,dx^*_{n_3}\geq dx^*_{n_3}/2,\end{aligned}$$

where we used $\theta\leq 1/20$, $t+3\leq 5t/4$, $a_3t\leq 1/100$, and $n_3/n\leq a_3$, and $|z|\leq d/5$. This implies

$$\|(M-z\mathrm{Id})x\|_2\geq\frac{dx^*_{n_3}}{2}\,|I|^{1/2}\geq\frac{dx^*_{n_3}}{2}\sqrt{\frac{n}{2}},$$

and completes the proof. □

5 Gradual vectors

In this section we introduce the notion of k-vectors, which provide a discretization of the set of gradual vectors, and discuss their properties. We will use notations of Section 4, in particular, ε_0, p, r, n_1, n_2, and n_3.

We first define the set of gradual vectors as the set of all vectors which are not almost constant and not steep. Note that any gradual vector x satisfies $x^*_{n_3}\neq 0$. We will use the following normalization of gradual vectors,

$$\mathcal{S}:=\big\{x\in\mathbb{C}^n\setminus(\mathcal{T}\cup\mathcal{B})\,:\,x^*_{n_3}=1\big\},$$

where $\mathcal{B}=\mathcal{B}(\theta_0)$ with $\theta_0=10/d^3$ (the set $\mathcal{B}(\theta)$ was introduced at the beginning of Section 4.5). Note that, by the definition of the almost constant vectors, we have for any $x\in\mathcal{S}$ that

$$\forall\,\lambda\in\mathbb{C}\quad|\{i\leq n\,:\,|x_i-\lambda|\leq\theta_0\}|\leq n-n_3,$$

and by the definition of the steep vectors,

$$\begin{aligned}&\forall\,0\leq i\leq r_0: && x^*_{p^i}\leq(n/p^i)^3(4d)^{r-r_0+1}d^3,\\&\forall\,r_0<i\leq r: && x^*_{p^i}\leq(4d)^{r-i+1}d^3,\\&x^*_{\lceil n_1/p\rceil}\leq 4d^4,\quad x^*_{n_1}\leq d^3,\quad\text{and} && x^*_{n_2}\leq d^{3/2}.\end{aligned}$$

5.1 Gradual k-vectors

For every positive integer k we define *k-vectors* as vectors in $\mathbb{C}^n$ with coordinates taking values in the set $\mathbb{Z}^2/k=\{\omega/k\,:\,\omega\in\mathbb{Z}^2\}$. Let $x=(x_1,x_2,\ldots,x_n)\in\mathbb{C}^n$ and $k\in\mathbb{N}$. The *k-approximation* of x is defined as the k-vector $y\in\mathbb{C}^n$ such that $\mathrm{Re}\,y_i=\lfloor k\mathrm{Re}\,x_i\rfloor/k$ and $\mathrm{Im}\,y_i=\lfloor k\mathrm{Im}\,x_i\rfloor/k$ for all $i\leq n$. Clearly, $\|x-y\|_\infty\leq\sqrt{2}/k$.

Below we split gradual vectors into classes of vectors, such that every pair of vectors from a given class has the same coordinates up to some permutation. We formalize it as follows. Let $x=\{x_i\}_i\in\mathbb{C}^n$. By $x^\sharp=\{x^\sharp_i\}_i$ denote the vector $(x_{\sigma(1)},x_{\sigma(2)},\ldots,x_{\sigma(n)})$, where the permutation σ is chosen so that $x_{\sigma(1)}\geq x_{\sigma(2)}\geq\ldots\geq x_{\sigma(n)}$ in the sense of lexicographical order introduced in Section 2. Recall that $x^*=\{x^*_i\}_i$ denotes the non-increasing rearrangement of $\{|x_i|\}_i$. Consider the following subset of "normalized" k-vectors,

$$\mathcal{A}_k:=\big\{y\in\mathbb{C}^n\,:\,y\text{ is a }k\text{-approximation of a vector in }\mathcal{S}\big\}.$$

Observe that for every $y \in \mathcal{A}_k$,

$$1 - \sqrt{2}/k \le y^*_{n_3} \le 1 + \sqrt{2}/k. \tag{14}$$

Next consider the equivalence relation on $\mathcal{A}_k$ defined by $x \overset{\sharp}{\sim} y$ iff $x^\sharp = y^\sharp$ for two k-vectors x and y. This relation partitions the set $\mathcal{A}_k$ into the equivalence classes. We first estimate how many classes we have.

Lemma 5.1. *Let $d \le n^{1/3}$ be large enough and $1 \le k \le \sqrt{n}/d^{3/2}$. Then the number of the equivalence classes (with respect to the relation $\overset{\sharp}{\sim}$) in $\mathcal{A}_k$ does not exceed e^n.*

Proof. From every equivalence class choose exactly one representative x, satisfying $|x_1| \ge |x_2| \ge \ldots \ge |x_n|$, multiply it by k and consider the set $\mathcal{A}'_k$ of such elements. Note that by definitions every element of $\mathcal{A}'_k$ has integer coordinates and, moreover, $\mathcal{A}'_k \subset k\mathcal{A}_k$.

Define a partition of $[n]$ into following $r+4$ sets. Let $I_0 = [n] \setminus [n_3]$. Set $I_1 = [p]$. Then for every $1 < i \le r$, set $I_i = [p^i] \setminus [p^{i-1}]$. Finally, set

$$I_{r+1} = [n_1] \setminus [p^r], \qquad I_{r+2} = [n_2] \setminus [n_1], \qquad \text{and} \qquad I_{r+3} = [n_3] \setminus [n_2].$$

The cardinalities of I_i's, $0 \le i \le r+3$, we denote by N_i's. Clearly, $N_0 \le n$, $N_{r+j} \le n_j$ for $j = 1, 2, 3$, and $N_i \le p^i$ for $1 \le i \le r$.

By the normalization of vectors in $\mathcal{S}$ and by (14), for every $x \in k\mathcal{A}_k$, we have $x^*_{n_3} \le k + \sqrt{2} \le 2.5k$. Therefore, by the definition of gradual vectors we have that for every $x \in \mathcal{A}'_k$ and every $r_0 \le i \le r$,

$$x^*_{n_3} \le 2.5k, \quad x^*_{n_2} \le 2.5kd^{3/2}, \quad x^*_{p^{r+1}} \le x^*_{n_1} \le 2.5kd^3, \quad x^*_{p^i} \le 2.5kd^3\,(4d)^{r+1-i} \tag{15}$$

and, using that $n \le n_1 d^{3/2} \le p^{r+1} d^{3/2}$ and $p^2 \le d$, for $0 \le i < r_0$,

$$x^*_{p^i} \le 2.5k(n/p^i)^3 x^*_{p^{r_0}} \le 2.5kp^{3(r+1-i)}\,d^{4.5}\,(4d)^{r+1-r_0} \le 2.5kd^{4.5}\,(4d)^{2.5(r+1-i)}. \tag{16}$$

For $0 \le i \le r+3$ let ν_i be the number of possible distinct coordinates of the projection of $y \in \mathcal{A}'_k$ on $\mathbb{C}^{I_i}$. Recall that every element of $\mathcal{A}'_k$ has integer coordinates. Note that if a complex number $z = a + \mathbf{i}b$ with integer a and b satisfy $|z| \le A$ for some $A \ge 2.5$ then $-A \le a, b \le A$, so there are at most $(2A+1)^2 \le 6A^2$ such numbers z. Therefore, by (15) and (16), we have for $1 \le i \le r+1$,

$$\nu_0 \le 40k^2, \quad \nu_{r+3} \le 40k^2 d^3, \quad \nu_{r+2} \le 40k^2 d^6, \quad \text{and} \quad \nu_i \le 40k^2 d^9\,(4d)^{5(r+2-i)}.$$

The number of sequences $\{x_i\}_{i=1}^N$ in $\mathbb{C}^N$ taking values in a set of cardinality ν, where we don't distinguish between sequences which can be obtained one from another by a permutation, equals $\binom{N+\nu-1}{N}$ (indeed, after introducing an order, this corresponds to the number of non-increasing sequences $\{y_i\}_{i=1}^N \subset [\nu]$ and we can pass to the strictly decreasing sequences $\{z_i\}_{i=1}^N \subset [\nu + N - 1]$, where $z_i = y_i + N - i$, hence this number is the same as the number of N elements subsets of $[\nu + N - 1]$). This leads to

$$|\mathcal{A}'_k| \le \prod_{i=0}^{r+3} \binom{N_i + \nu_i - 1}{N_i}.$$

Using bounds for ν_i and N_i, the standard estimate $\binom{m}{\ell} \le (em/\ell)^\ell$, and that $d^3k^2 \le n$, $p^{r_0} \le n/d^{3/2}$, we get

$$B_1 := \binom{N_0 + \nu_0 - 1}{N_0} \le \binom{N_0 + \nu_0}{\nu_0} \le \binom{n + 40k^2}{40k^2} \le \binom{n + \lfloor 40n/d^3 \rfloor}{\lfloor 40n/d^3 \rfloor} \le \left(ed^3/20\right)^{40n/d^3};$$

$$B_2 := \binom{N_{r+2} + \nu_{r+2} - 1}{N_{r+2}} \le \binom{n_2 + 40k^2 d^6}{n_2} \le \left(\frac{41end^3}{n_2}\right)^{n_2} \le d^{4n/d^{2/3}};$$

$$B_3 := \binom{N_{r+3}+\nu_{r+3}-1}{N_{r+3}-1} \le \binom{n_3+40k^2d^3}{n_3} \le (41e/a_3)^{a_3 n};$$

and, for $1 \le i \le r+1$,

$$B_{4,i} := \binom{N_i+\nu_i-1}{N_i} \le \binom{p^i+40k^2d^9\,(4d)^{5(r+2-i)}}{p^i}.$$

If $p^i > 40k^2d^9\,(4d)^{5(r+2-i)}$ then $B_{4,i} \le 4^{p^i}$, otherwise, using

$$k^2d^3p^{-i} \le np^{-i} \le n_1 d^{3/2}p^{-i} \le p^{r+1-i}d^{3/2},$$

we have

$$B_{4,i} \le \left(\frac{80ek^2d^9\,(4d)^{5(r+2-i)}}{p^i}\right)^{p^i} \le \left(d^8\,(4d)^{6(r+2-i)}\right)^{p^i}.$$

Denoting $B_4 = \prod_{i=1}^{r+1} B_{4,i}$, using that d is large enough, and passing to sums of logarithms, we have

$$\begin{aligned}\ln B_4 &\le \sum_{i=1}^{r+1} p^i\,\ln\left(d^8\,(4d)^{6(r+2-i)}\right) \le \sum_{\substack{\ell=1\\(\ell=r+2-i)}}^{r+1} p^{r+2-\ell}\,(6\ell\,\ln(4d)+8\ln d)\\ &\le 20p^{r+1}\ln d \le 20pn_1\ln d \le n(\ln d)/d.\end{aligned}$$

Combining all bounds we obtain

$$|\mathcal{A}'_k| \le B_1B_2B_3B_4 \le e^n,$$

provided that a_3 is small enough and d is large enough. □

5.2 The ℓ-decomposition with respect to k-vectors

In this subsection, we introduce one of the most important technical ingredients of the paper – the ℓ-decomposition with respect to k-vectors, which is a special way to structure a k-vector y as a collection of two-dimensional "stairs" or "ladders" which ultimately determine the anti-concentration properties of the product My (with a random matrix M uniformly distributed in $\mathcal{M}_{n,d}$).

Let $y = (y_i)_{i=1}^n \in \mathbb{C}^n$ be a k-vector. We will construct a partition of [n] into two sequences of subsets of $[n]$, $(\mathcal{LS}_j(y))_{j=0}^\infty$ and $(\mathcal{LR}_j(y))_{j=0}^\infty$, which we call *spread ℓ-parts* and *regular ℓ-parts*, respectively. Note that all but a finite number of the subsets are empty. When the vector y is clear from the context, we will simply write $\mathcal{LS}_j$ and $\mathcal{LR}_j$ for the corresponding ℓ-parts.

Our construction consists of a series of steps (indexed by j), and each step comprises a sequence of substeps. At j-th step (except $j=0$), we already have sets $(\mathcal{LS}_u)_{u=0}^{j-1}$ and $(\mathcal{LR}_u)_{u=0}^{j-1}$ constructed. If $j=0$ set $I_0 := [n]$ and $\Lambda_0 := \{y_i : i \in [n]\}$, otherwise set

$$I_j := [n] \setminus \Big(\bigcup_{u\le j-1}\mathcal{LS}_u \cup \bigcup_{u\le j-1}\mathcal{LR}_u\Big) \quad\text{and}\quad \Lambda_j := \{y_i : i \in I_j\}.$$

Now, for each $\lambda \in \Lambda_j$ such that $|\{i \in I_j : y_i = \lambda\}| < 2^{j+1}$ we let

$$L(j,\lambda) := \{i \in I_j : y_i = \lambda\},$$

and for every $\lambda \in \Lambda_j$ with $|\{i \in I_j : y_i = \lambda\}| \ge 2^{j+1}$ we let $L(j,\lambda)$ be the subset of $\{i \in I_j : y_i = \lambda\}$ of cardinality 2^j such that

$$L(j,\lambda) = I_j \cap [1, \sup L(j,\lambda)]$$

(that is, we choose $L(j,\lambda)$ as the "leftmost" subset of cardinality 2^j). Note that by construction for $j \geq 0$ we have

$$2^{j-1} \leq |L(j,\lambda)| < 2^{j+1}. \tag{17}$$

We refer to sets $(L(j,\lambda))_{\lambda\in\Lambda_j}$ as *level sets of order* j (with respect to y). The union of the level sets of order j will form the spread and regular parts, $\mathcal{LS}_j$ and $\mathcal{LR}_j$, i.e., we define $\mathcal{LS}_j$ and $\mathcal{LR}_j$ so that

$$\mathcal{LS}_j \cup \mathcal{LR}_j = \bigcup_{\lambda\in\Lambda_j} L(j,\lambda).$$

To separate the spread part from the regular one of the same order, we apply an embedded procedure consisting of substeps. Our construction of spread vectors is based on extracting a maximal (d/k)-separated set subset from Λ_j, consisting of at least 2 elements, provided that such a set exists. Note, that we need to have at least 2 elements to be able to apply anti-concentration later. We construct a subset $\Lambda_j^S \subset \Lambda_j$ as follows.

Substep 1. If the diameter of Λ_j is strictly less than d/k then we set $\Lambda_j^S := \emptyset$ and terminate. Otherwise, note that there is at least one pair of numbers $\lambda, \lambda' \in \Lambda_j$ such that $|\lambda - \lambda'| \geq d/k$. Define λ_1 as the *largest* (with respect to the lexicographical order, see Section 2) number in Λ_j such that $|\lambda_1 - \lambda'| \geq d/k$ for some $\lambda' \in \Lambda_j$ and pass to the next substep.

Substep m ($m > 1$). We have already chosen numbers $\lambda_1, \lambda_2, \ldots, \lambda_{m-1}$ in Λ_j. If all $\lambda \in \Lambda_j$ are within a distance strictly less than d/k to $\{\lambda_1, \lambda_2, \ldots, \lambda_{m-1}\}$ then set $\Lambda_j^S := \{\lambda_1, \lambda_2, \ldots, \lambda_{m-1}\}$ and terminate (note, by the construction, this cannot happen if $m = 2$). Otherwise, let λ_m be the largest number in Λ_j with the distance to $\{\lambda_1, \lambda_2, \ldots, \lambda_{m-1}\}$ greater or equal to d/k and go to the next substep.

Note that by construction we have that the sequence $(\lambda_m)_{m\geq 1}$ is decreasing (with respect to the lexicographical order) and, moreover, $|\lambda_u - \lambda_v| \geq d/k$ for every admissible $u \neq v$. Now, by *the spread ℓ-part of order j* with respect to y, we call the union

$$\mathcal{LS}_j = \mathcal{LS}_j(y) := \bigcup_{\lambda\in\Lambda_j^S} L(j,\lambda)$$

and by *the regular ℓ-part of order j* with respect to y, we call the union

$$\mathcal{LR}_j = \mathcal{LR}_j(y) := \bigcup_{\lambda\in\Lambda_j\setminus\Lambda_j^S} L(j,\lambda).$$

The *height* $h(\cdot)$ of a regular (resp, spread) ℓ-part is the number of level sets it comprises (if the ℓ-part is empty then $h = 0$). In particular, by (17), if $\mathcal{L}_j$ is either $\mathcal{LS}_j$ or $\mathcal{LR}_j$, then

$$2^{j-1}h(\mathcal{L}_j) \leq |\mathcal{L}_j| \leq 2^{j+1}h(\mathcal{L}_j). \tag{18}$$

Note also that by the construction the height of a non-empty spread part is at least 2. We will often write $\mathcal{L}$ to denote an ℓ-part (of some order) with respect to y. Note also that the maximal number of steps (starting with the step $j = 0$) that we can have is the smallest $j + 1$ such that $n < 2^{j+1}$, i.e. $j + 1 = \lceil \log_2 n \rceil < 1.5 \ln n$ for large enough n. Therefore, the number of non-empty ℓ-parts, denoted below by $m(y)$ is at most $3 \ln n$.

Finally we introduce the ℓ-decomposition. Let y be a k-vector with the corresponding ℓ-parts $\{\mathcal{LS}_j, \mathcal{LR}_j\}_{j\geq 0}$. We will re-enumerate the non-empty spread and regular ℓ-parts and will write $(\mathcal{L}^{(q)})_{q=1}^m$ (i.e., suppressing the order and spreadness/regularity), where $m = m(y) \leq 3 \ln n$. To make this representation unique, we assume that within the sequence $(\mathcal{L}^{(q)})_{q=1}^m$, any spread ℓ-part precedes (by the index) any regular ℓ-part, and that for any two spread (resp. regular) parts, the one of smaller order precedes the other. In what follows, such a sequence will be called the *ℓ-decomposition* with respect to y.

Below, given a level set $L \subset [n]$, i.e., a set of coordinates where y_i preserves its value, we denote this value by $y(L)$.

To clarify our construction we would like to provide the following example.

Example. Let $n = 7$, $d = 2$, $k = 6$. Consider $y = (1/2, 1/3, 1/2, 1/6, 1/2, 1/3, -1/3)$. Note that y is a k-vector. According to the above procedure, at step $j = 0$ we have $\Lambda_0 = \{1/2, 1/3, 1/6, -1/3\}$ and construct level sets $L(0, 1/2) = \{1\}$, $L(0, 1/3) = \{2\}$, $L(0, 1/6) = \{4\}$, $L(0, -1/3) = \{7\}$. Since $d/k = 1/3$, we get that $\Lambda_0^S = \{1/2, 1/6, -1/3\}$. Thus $\{1, 4, 7\}$ is the spread ℓ-part of order 0, and $\{2\}$ is the regular ℓ-part of order 0. At step 1, we have $\Lambda_1 = \{1/2, 1/3\}$ and construct level sets $L(1, 1/2) = \{3, 5\}$ and $L(1, 1/3) = \{6\}$. Then $\Lambda_1^S = \emptyset$, therefore $\emptyset$ is the spread ℓ-part of order 1, and $\{3, 5, 6\}$ is the regular ℓ-part of order 1. Altogether, we have $m(y) = 3$ non-empty ℓ-parts – one spread ℓ-part of order 0 with the height 3, one regular ℓ-part of order 0 with the height 1, and one regular ℓ-part of order 1 with the height 2. The ℓ-decomposition with respect to y is $(\{1, 4, 7\}, \{2\}, \{3, 5, 6\})$.

A quick analysis of the construction procedure for the ℓ-parts gives the following properties, which we summarize into three lemmas. We leave the (rather straightforward) proofs to the reader.

Lemma 5.2. *Let y be a k-vector, $\lambda \in \mathbb{Z}/k$ and set $I = \{i \le n : y_i = \lambda\}$. Assume that $I \neq \emptyset$ and denote $u := \lfloor \log_2((|I| + 1)/3) \rfloor$. Then*

$$I = \bigcup_{j=0}^{u+1} L(j, \lambda),$$

$$2^u \le |L(u+1, \lambda)| = |I| - 2^{u+1} + 1 \le 2^{u+2} - 1, \qquad \text{and} \qquad \forall\, 0 \le j \le u : \ |L(j, \lambda)| = 2^j.$$

Lemma 5.3. *Let y be a k-vector, let $j \ge 1$ and assume that $\mathcal{LS}_j \cup \mathcal{LR}_j \neq \emptyset$. Then for all $0 \le m < j$ we have $\mathcal{LS}_m \cup \mathcal{LR}_m \neq \emptyset$,*

$$h(\mathcal{LS}_m) + h(\mathcal{LR}_m) \ge h(\mathcal{LS}_j) + h(\mathcal{LR}_j),$$

and

$$\{y_i : i \in \mathcal{LS}_j \cup \mathcal{LR}_j\} \subset \{y_i : i \in \mathcal{LS}_m \cup \mathcal{LR}_m\}.$$

Lemma 5.4. *Let y be a k-vector and let $\mathcal{LS}_j$, $\mathcal{LR}_j$, $j \ge 0$, be its ℓ-parts. Then*

- *The height of every non-empty spread ℓ-part is at least 2.*
- *For every non-empty spread ℓ-part $\mathcal{LS}$ and any $i_1, i_2 \in \mathcal{LS}$ with $y_{i_1} \neq y_{i_2}$ we have $|y_{i_1} - y_{i_2}| \ge d/k$.*
- *If $\widetilde{y}$ is a permutation of the vector y then necessarily the ℓ-parts of y and $\widetilde{y}$ agree up to a permutation of $[n]$; in particular, the heights and cardinalities of spread or regular ℓ-parts of a given order with respect to y and $\widetilde{y}$ are the same.*

The last property implies that with every equivalence class $\mathcal{C} \subset \mathcal{A}_k$ and every $j \ge 0$, we may associate four integers by fixing (an arbitrary) $y \in \mathcal{C}$ and setting

$$\mathrm{cs}_j(\mathcal{C}) := |\mathcal{LS}_j(y)|, \ \mathrm{cr}_j(\mathcal{C}) := |\mathcal{LR}_j(y)|, \ \mathrm{hs}_j(\mathcal{C}) := h(\mathcal{LS}_j(y)), \ \mathrm{hr}_j(\mathcal{C}) := h(\mathcal{LR}_j(y)).$$

The following lemma allows to estimate cardinalities of equivalence classes in terms of these quantities.

Lemma 5.5. *Let $k \ge 1$, $\mathcal{C}$ be an equivalence class in $\mathcal{A}_k$ with respect to the relation $\overset{\sharp}{\sim}$. Then the cardinality of the class $\mathcal{C}$ can be estimated as*

$$|\mathcal{C}| \le n! \prod_{j=0}^{\infty} \frac{\mathrm{hs}_j^{\mathrm{cs}_j}\, \mathrm{hr}_j^{\mathrm{cr}_j}}{\mathrm{cs}_j!\, \mathrm{cr}_j!},$$

where we adopt the notation $0^0 = 1$.

Proof. There are clearly $n!/\prod_{j=0}^{\infty}\mathrm{cs}_j!\mathrm{cr}_j!$ ways to "assign" ℓ-parts to specific locations within $[n]$. Fix for a moment $j\geq 0$ with $\mathrm{cs}_j\neq 0$ and let $\mathcal{LS}$ be a fixed subset of $[n]$ of cardinality cs_j. Recall that $P_{\mathcal{LS}}(y)\in\mathbb{C}^{\mathcal{LS}}$ denotes the coordinate projection of y onto $\mathbb{C}^{\mathcal{LS}}:=\mathrm{span}\,\{e_i:\, i\in\mathcal{LS}\}$. Consider the set

$$W_{\mathcal{LS}}:=\big\{P_{\mathcal{LS}}(y):\, y\in\mathcal{C} \text{ is such that } \mathcal{LS}_j(y)=\mathcal{LS}\big\}$$

Since all vectors within a given equivalence class share the same *levels*, the cardinality of $W_{\mathcal{LS}}$ can be estimated from above by ${\mathrm{hs}_j}^{\mathrm{cs}_j}$. Similarly, we can estimate the number of realizations of regular ℓ-parts. Combining this with the estimate for "location assignments," we obtain the desired bound. □

5.3 Decomposition of the set of gradual vectors

In this subsection, we define a way to partition the set of gradual vectors $\mathcal{S}$ in terms of structure of their k-approximations. Roughly speaking, we will observe the following dichotomy for a vector x in $\mathcal{S}$: either x possesses a k-approximation y (for a relatively small k) whose ℓ-decomposition contains many spread ℓ-parts (that is, the distance between the "stairs" in a graphical representation of y is often large), or, for an appropriately chosen k, the k-approximation of x contains ℓ-parts with large heights.

Given integer $u\geq 0$ we introduce two subsets of $\mathcal{S}$,

$$\begin{aligned}\mathcal{K}_u:=\big\{x\in\mathcal{S}:\, &\text{in the } \ell\text{-decomposition with respect to the } d^u\text{-approximation of } x,\\ &\text{the total cardinality of the spread } \ell\text{-parts is at least } c_{\mathcal{K}}n_3\big\}\end{aligned}$$

and

$$\begin{aligned}\mathcal{P}_u:=\big\{x\in\mathcal{S}:\, &\text{in the } \ell\text{-decomposition with respect to the } d^u\text{-approximation of } x,\\ &\text{the total cardinality of spread and regular } \ell\text{-parts with heights not smaller}\\ &\text{than } c_{\mathcal{P}}2^{c_{\mathcal{P}}(u-4)a_3}a_3 \text{ is at least } c_{\mathcal{P}}n_3\big\}.\end{aligned}$$

Here, by "total cardinality" we mean the cardinality of the union of the respective ℓ-parts, and $c_{\mathcal{K}}$, $c_{\mathcal{P}}\in(0,1)$ are two universal constants whose values can be derived from the proofs. Note that for small $u\geq 1$, we have $c_{\mathcal{P}}2^{c_{\mathcal{P}}(u-4)a_3}a_3\leq 1$, so the set $\mathcal{P}_u$ coincides with $\mathcal{S}$.

The next theorem is the main statement of the subsection, and one of the main technical ingredients of the paper.

Theorem 5.6 (Decomposition of $\mathcal{S}$)**.** *Let $v\geq 5$ be an integer. Then*

$$\mathcal{S}=\bigcup_{u=4}^{v}\mathcal{K}_u\cup\mathcal{P}_v.$$

Theorem 5.6 says that for any vector x in $\mathcal{S}$, either x belongs to $\mathcal{K}_u$ for some $u\leq v$ or $x\in\mathcal{P}_v$. To prove this theorem, we first consider more technical (yet more simple) ways to partition $\mathcal{S}$, and then gradually "replace" them with the conditions we are interested in.

The following lemma is a straightforward implication of Lemma 2.2 in [38].

Lemma 5.7. *Let $\theta_0=10/d^3$, $x\in\mathcal{S}$, $k\geq 5/\theta_0$, and let y be the k-approximation of x. Then there exist disjoint subsets $I,J\subset[n]$ such that $|I|,|J|\geq n_3/4$ and for any $i\in I$ and $j\in J$ we have $|y_i-y_j|\geq\theta_0/2$.*

We now prove a dichotomy lemma dealing with cardinalities of ℓ-parts.

Lemma 5.8. *Let $\theta_0=10/d^3$, $x\in\mathcal{S}$, $k\geq 2d/\theta_0$, and let y be the k-approximation of x. Then at least one of the following assertions holds.*

- *The cardinality of $\bigcup_j \mathcal{LS}_j \cup \mathcal{LR}_j$, where the union is taken over all $j \geq 0$ with $h(\mathcal{LS}_j)+h(\mathcal{LR}_j) \geq 10$, is at least $n_3/8$.*
- *The total cardinality of the spread ℓ-parts in the ℓ-decomposition with respect to y is at least $n_3/120$.*

Proof. By Lemma 5.7, we can find disjoint sets $I, J \subset [n]$ of cardinality at least $n_3/4$ such that for any $i \in I$ and any $j \in J$ one has $|y_i - y_j| \geq \theta_0/2 \geq d/k$. Let $(\mathcal{LS}_j, \mathcal{LR}_j)_{j=0}^{\infty}$ be the ℓ-parts of y, and let j_0 be the largest integer j such that

$$(\mathcal{LS}_j \cup \mathcal{LR}_j) \cap (I \cup J) \neq \emptyset.$$

For concreteness, assume that $(\mathcal{LS}_{j_0} \cup \mathcal{LR}_{j_0}) \cap I \neq \emptyset$ (the other case is treated similarly). By Lemma 5.3, $\{y_i : i \in \mathcal{LS}_j \cup \mathcal{LR}_j\} \cap \{y_i : i \in I\} \neq \emptyset$ for all $j \leq j_0$.

Consider two disjoint sets of indices,

$$U_1 = \{j \leq j_0 \,:\, (\mathcal{LS}_j \cup \mathcal{LR}_j) \cap J \neq \emptyset \;\text{ and }\; h(\mathcal{LS}_j) + h(\mathcal{LR}_j) \geq 10\}$$

and

$$U_2 = \{j \leq j_0 \,:\, (\mathcal{LS}_j \cup \mathcal{LR}_j) \cap J \neq \emptyset \;\text{ and }\; h(\mathcal{LS}_j) + h(\mathcal{LR}_j) \leq 9\}\,.$$

Clearly,

$$J \subset \bigcup_{j \in U_1 \cup U_2} (\mathcal{LS}_j \cup \mathcal{LR}_j),$$

hence either

$$\Big|\bigcup_{j \in U_1} (\mathcal{LS}_j \cup \mathcal{LR}_j)\Big| \geq n_3/8 \quad \text{ or } \quad \Big|\bigcup_{j \in U_2} (\mathcal{LS}_j \cup \mathcal{LR}_j)\Big| \geq n_3/8.$$

If the first bound holds we get that the total cardinality of spread or regular ℓ-parts of cumulative height at least 10 is at least $n_3/8$, i.e. the first alternative of the lemma holds. We now assume that the second bound holds. Note that for every $j \in U_2$ we have

$$\{y_i : i \in \mathcal{LS}_j \cup \mathcal{LR}_j\} \cap \{y_i : i \in I\} \neq \emptyset \text{ and } \{y_i : i \in \mathcal{LS}_j \cup \mathcal{LR}_j\} \cap \{y_i : i \in J\} \neq \emptyset.$$

Using that $|y_a - y_b| \geq d/k$ for all $a \in I$, $b \in J$, by the definition of the spread ℓ-part, we necessarily have $h(\mathcal{LS}_j) \geq 2$, hence $h(\mathcal{LR}_j) \leq 7$. By (18) this implies $|\mathcal{LS}_j| \geq |\mathcal{LR}_j|/14$ for every $j \in U_2$. Thus,

$$\Big|\bigcup_{j \in U_2} \mathcal{LS}_j\Big| \geq \frac{1}{15}\Big|\bigcup_{j \in U_2} (\mathcal{LS}_j \cup \mathcal{LR}_j)\Big| \geq n_3/120,$$

which implies the desired result. □

Lemma 5.8 allows us to prove a more elaborate dichotomy statement.

Lemma 5.9. *Let $x \in \mathcal{S}$, $u \geq 4$, and let y^u and y^{u+1} be the d^u- and d^{u+1}-approximations of x, respectively. For each $i \leq n$, set*

$$J^u(i) := \{j \leq n : y_j^u = y_i^u\} \quad \text{ and } \quad J^{u+1}(i) := \{j \leq n : y_j^{u+1} = y_i^{u+1}\}.$$

Then we have the following dichotomy.

- *Either $x \in \mathcal{K}_u \cup \mathcal{K}_{u+1}$,*
- *or $\big|\{i \leq n : 2|J^{u+1}(i)| \leq |J^u(i)|\}\big| \geq n_3/192$.*

Proof. Let x, y^u, y^{u+1} be as above and note that $d^u \geq 2d/\theta_0$ for any $u \geq 4$, in particular we may apply Lemma 5.8 witk $k = d^u$ to the vector y^u. Denote

$$I := \{i \leq n : 2|J^{u+1}(i)| \leq |J^u(i)|\}.$$

Assume that $x \notin \mathcal{K}_u$ and that $I < n_3/192$. We show that $x \in \mathcal{K}_{u+1}$.

For $m = u, u+1$ denote

$$\begin{aligned} U^m &:= \{J^m(i) : i \in I^c\} \quad \text{ and} \\ V^m &:= \{j \geq 0 : |\{J \in U^m : (\mathcal{LS}_j(y^m) \cup \mathcal{LR}_j(y^m)) \cap J \neq \emptyset\}| \geq 10\}. \end{aligned}$$

We first prove that

$$\Big| \bigcup_{j \in V^{u+1}} (\mathcal{LS}_j(y^{u+1}) \cup \mathcal{LR}_j(y^{u+1})) \cap \bigcup_{J \in U^{u+1}} J \Big| \geq n_3/144. \tag{19}$$

Note that by the definition of k-approximation, given j, Re $y_j^u = \ell/d^u$ for some integer ℓ if and only if $\mathrm{Re}\, y_j^{u+1} = \ell/d^u + m/d^{u+1}$ for some $0 \leq m < d$, and the same holds for the imaginary parts of y_j^u, y_j^{u+1}. This implies that $J^{u+1}(i) \subset J^u(i)$ for every i. Thus, there exists a bijection $\rho : U^u \to U^{u+1}$ such that each set $J \in U^u$ corresponds to $\rho(J) \in U^{u+1}$ with $\rho(J) \subset J$ and $2|\rho(J)| > |J|$. Since every J in U^u is a level set, Lemma 5.3 implies that the set V^u is an interval in $\mathbb{Z}$, that is, either $V^u = \emptyset$ or $V^u = \{0, \ldots, \sup V^u\}$. Similarly, V^{u+1} is an interval. Moreover, if $V^u \neq \emptyset$ then Lemma 5.2 together with the inequality $2|\rho(J)| > |J|$ implies

$$\sup V^{u+1} \geq \max(\sup V^u - 1, 0)$$

Consider the set

$$J_0 := \{j \geq 0 \, : \, j \notin V^u \quad \text{ and } \quad h(\mathcal{LS}_j(y^u)) + h(\mathcal{LR}_j(y^u)) \geq 10\}.$$

Observe that for every $j \in J_0$, the union $\mathcal{LS}_j(y^u) \cup \mathcal{LR}_j(y^u)$ has at least

$$h(\mathcal{LS}_j(y^u)) + h(\mathcal{LR}_j(y^u)) - 9 \geq (h(\mathcal{LS}_j(y^u)) + h(\mathcal{LR}_j(y^u)))/10$$

of its level sets contained entirely in I. Hence, by (17) (see also Lemma 5.2),

$$\forall j \in J_0 \quad \big|(\mathcal{LS}_j(y^u) \cup \mathcal{LR}_j(y^u)) \cap I\big| \geq \frac{1}{40} \big|\mathcal{LS}_j(y^u) \cup \mathcal{LR}_j(y^u)\big|.$$

Since $x \notin \mathcal{K}_u$, the total cardinality of the spread ℓ-parts in the ℓ-decomposition with respect to y^u is at most $c_{\mathcal{K}} n_3 < n_3/120$ provided that $c_{\mathcal{K}} < 1/120$. Therefore, by Lemma 5.8, the total cardinality of spread and regular parts of cumulative height 10 or more, is at least $n_3/8$. Then the last relation and the upper bound on the cardinality of I yield that

$$\Big| \bigcup_{j \in V^u} \mathcal{LS}_j(y^u) \cup \mathcal{LR}_j(y^u) \Big| \geq n_3/8 - 12|I| \geq n_3/16$$

(in particular, $V^u \neq \emptyset$). Using that $\bigcup_{J \in U^u} J \supset I^c$, we obtain

$$\Big| \bigcup_{j \in V^u} (\mathcal{LS}_j(y^u) \cup \mathcal{LR}_j(y^u)) \cap \bigcup_{J \in U^u} J \Big| \geq n_3/16 - |I| \geq n_3/18. \tag{20}$$

Next, consider a set $J \in U^u$ satisfying

$$L := \bigcup_{j \in V^u} (\mathcal{LS}_j(y^u) \cup \mathcal{LR}_j(y^u)) \cap J \neq \emptyset.$$

Then L is the union of level sets of y^u of all orders $0, ..., j_0$ for some $0 \le j_0 \le \sup V^u$. Since the set $\rho(J) \in U^{u+1}$ has cardinality greater than $|J|/2$, Lemma 5.2 implies that $\rho(J)$ must contain level sets of y^{u+1} of all orders $0, ..., \max(j_0 - 1, 0)$ (note that necessarily $\max(j_0 - 1, 0) \in V^{u+1}$). Applying Lemma 5.2 again, we obtain

$$\Big| \bigcup_{j \in V^{u+1}} (\mathcal{LS}_j(y^{u+1}) \cup \mathcal{LR}_j(y^{u+1})) \cap \rho(J) \Big| \ge \frac{1}{8} \Big| \bigcup_{j \in V^u} (\mathcal{LS}_j(y^u) \cup \mathcal{LR}_j(y^u)) \cap J \Big|.$$

This together with (20) implies (19).

Finally we show that (19) implies that $x \in \mathcal{K}_{u+1}$.

Fix $j \in V^{u+1}$ and let $J^1, J^2, \dots, J^b$ ($b \ge 10$) be (distinct) elements of U^{u+1}, which have a non-empty intersection with $\mathcal{LS}_j(y^{u+1}) \cup \mathcal{LR}_j(y^{u+1})$. Denote $z_a = y^{u+1}(J^a)$ and $w_a = y^u(\rho^{-1}(J^a))$, $a \le b$. Since ρ is a bijection, w_1, ..., w_b are also distinct. It will be convenient, to see elements of those two sequences as elements of lattices $\Lambda_u := (\mathbb{Z}/d^u)^2$ and $\Lambda_{u+1} := (\mathbb{Z}/d^{u+1})^2$. We also denote $D = [0, (d-1)/d^{u+1}] \times [0, (d-1)/d^{u+1}]$. As we noticed above, by construction, we have $z_a \in w_a + D$ for every $a \le b$. Now we split Λ_u into nine equivalence classes using the relation $(v_1, v_2) \sim (v_3, v_4)$ if and only if $d^u(v_1 - v_3)$ and $d^u(v_2 - v_4)$ are divisible by 3. Let Λ be an equivalence class such that $|\Lambda \cup \{w_a\}_{a \le b}| \ge b/9$. Note, if $w_a, w_\ell \in \Lambda$ then $\|z_a - z_\ell\|_\infty \ge 2/d^u$, in particular, $\mathcal{LS}_j(y^{u+1}) \ne \emptyset$. Let λ_1, ..., λ_m, $m \le b/9 - 1$, be as in the construction of $\mathcal{LS}_j(y^{u+1})$. Then for each $i \le m$, $\lambda_i \in \Lambda_{u+1}$ and $\lambda_i \in \mu_i + D$ for some $\mu_i \in \Lambda_{u+1}$. Let $\bar{\mu}_i$ be the closest (in ℓ_∞-metric) to μ_i point of Λ. Since $m \le b/9 - 1$, there exists $w_a \in \Lambda \setminus \{\mu_i\}_{i \le m}$. Then for each $i \le m$ we have

$$\|w_a - \mu_i\|_\infty \ge \|w_a - \bar{\mu}_i\|_\infty - \|\bar{\mu}_i - \mu_i\|_\infty \ge 2/d^u.$$

Since $z_a \in w_a + D$, $\lambda_i \in \mu_i + D$, we observe

$$|z_a - \lambda_i| \ge \|z_a - \lambda_i\|_\infty \ge 1/d^u.$$

This shows that the sequence $\{\lambda_i\}_{i \le m}$ can be continued. Thus, $\mathcal{LS}_j(y^{u+1})$, the spread ℓ-part of order j with respect to y^{u+1}, must comprise at least $b/9$ levels (i.e., its height is at least $b/9$). Then, applying estimates for cardinalities of individual level sets (17), we obtain

$$|\mathcal{LS}_j(y^{u+1})| \ge \frac{1}{4} \cdot \frac{1}{9} \Big| (\mathcal{LS}_j(y^{u+1}) \cup \mathcal{LR}_j(y^{u+1})) \cap \bigcup_{J \in U^{u+1}} J \Big|.$$

Taking the union over all $j \in V^{u+1}$ and choosing small enough $c_{\mathcal{K}}$, we obtain the desired result. $\square$

We are now ready to prove Theorem 5.6.

Proof of Theorem 5.6. Fix a vector $x \in \mathcal{S}$, and assume that $x \notin \bigcup_{u=4}^{v} \mathcal{K}_u$. We show that $x \in \mathcal{P}_v$. For every $u \ge 4$, let y^u be the d^u-approximation of x. For $i \le n$ and $u \ge 4$ let

$$J^u(i) := \{j \le n : y^u_j = y^u_i\} \quad \text{and} \quad I^u := \big\{j \le n : 2|J^{u+1}(j)| \le |J^u(j)|\big\}.$$

The assumption that $x \notin \bigcup_{u=4}^{v} \mathcal{K}_u$, together with Lemma 5.9, implies that $|I^u| \ge n_3/192$ for $4 \le u < v$. Define an auxiliary integer vector $a = (a_i)_{i=1}^n$ by setting for $i \le n$,

$$a_i := \big|\big\{4 \le u < v : i \in I^u\big\}\big|.$$

The lower bound on cardinalities of sets I^u implies that

$$\sum_{i=1}^{n} a_i \ge (v-4) n_3/192.$$

On the other hand, clearly $a_i \le v-4$ for all $i \le n$. Recall $n_3 = \lfloor a_3 n \rfloor$. Let

$$J := \{i \le n \,:\, a_i \ge (v-4)n_3/(384n)\}.$$

Then

$$(v-4)n_3/192 \le \sum_{i=1}^{n} a_i \le |J|(v-4) + (n-|J|)(v-4)a_3/384,$$

which implies

$$|J| \ge (n_3/192 - a_3 n/384)/(1-a_3/384) \ge n_3/400.$$

By the definitions of I^u and a_i's, we have for every $i \in J$, $|J^v(i)| \le 2^{-(v-4)a_3/384}\,n$, hence, by Lemma 5.3, in the ℓ-decomposition of y^v, any regular or spread ℓ-part of order

$$j > j_0 := \lfloor \log_2(2^{-(v-4)a_3/384}n) \rfloor + 1$$

does not have a non-empty intersection with J. Thus, we obtain

$$\Big|\bigcup_{j\ge 0} \mathcal{LS}_j(y^v) \cup \mathcal{LR}_j(y^v)\Big| = \Big|\bigcup_{j=0}^{j_0} \mathcal{LS}_j(y^v) \cup \mathcal{LR}_j(y^v)\Big| \ge |J| \ge n_3/400.$$

Finally, since by (18) any regular or spread ℓ-part of order j and of height at most h has cardinality at most $2^{j+1}h$, the last relation yields for every positive integer h,

$$\begin{aligned}
&\Big|\bigcup_{j\ge 0:\, h(\mathcal{LS}_j(y^v))\ge h} \mathcal{LS}_j(y^v) \;\cup \bigcup_{j\ge 0:\, h(\mathcal{LR}_j(y^v))\ge h} \mathcal{LR}_j(y^v)\Big| \\
&\ge n_3/400 - \Big|\bigcup_{j\le j_0:\, h(\mathcal{LS}_j(y^v))< h} \mathcal{LS}_j(y^v) \;\cup \bigcup_{j\le j_0:\, h(\mathcal{LR}_j(y^v))< h} \mathcal{LR}_j(y^v)\Big| \\
&\ge n_3/400 - 2\cdot 2^{j_0+2}(h-1) \ge n_3/400 - h\cdot 2^{4-(v-4)a_3/384}n.
\end{aligned}$$

Choosing $h = 2^{(v-4)a_3/384}a_3/(400\cdot 2^5)$, we get the result with $c_{\mathcal{P}} = 1/(400\cdot 2^5)$. $\square$

6 A small ball probability theorem

Let $K \subset [n]$ and M be the random matrix uniformly distributed on $\mathcal{M}_{n,d}$. The purpose of this section is to study anti-concentration properties of a random vector of the form $M^K y + V$, where y is a fixed k-vector and V is a fixed vector in $\mathbb{C}^{|K|}$. The high-level idea is to replace the random vector $M^K y$, whose distribution is difficult to describe due to dependencies within M^K, by a "simpler" random vector $Z = (Z_i)_{i\in K}$ whose anti-concentration properties can be studied with the help of standard tools. The construction of Z will be done in such a way that we will be able to pass from estimates for Z back to $M^K y$ by conditioning on a certain event of not too small probability. The actual proof is technical, and even stating the main result of the section requires some preparatory work. Instead of working with the probability space $\mathcal{M}_{n,d}$, we will split it into certain equivalence classes (the splitting will depend on the structure of the vector y, more precisely, on the partition of $[n]$ given by the ℓ-decomposition of y), and study the conditional anti-concentration. The probability estimate will be given as a function of the ℓ-decomposition. As we mentioned in the introduction, this argument is related to the LCD-based method of Rudelson and Vershynin [49] which in turn was strongly influenced by earlier works on singularity of discrete random matrices [33, 54, 56]. A principal difference of our approach is that the ℓ-decomposition, being a "multidimensional" characteristic of a vector, provides much more structural information than LCD. This structural information is heavily used in this part of the paper.

We start by introducing a structure on $\mathcal{M}_{n,d}$. For each $m \leq n$, let $\mathcal{R}_{n,m,d}$ be the set of $n \times m$ matrices with integers coefficients from the set $\{0, 1, \dots, d\}$ such that
1. the sum in each row is d, and
2. the sum in every column is a non-negative integer multiple of d.

Now, for every k-vector y with the ℓ-decomposition $(\mathcal{L}^{(q)})_{q=1}^m$, where

$$m = m(y) \leq 3 \ln n,$$

we define the mapping $\xrightarrow{(y)}$ from $\mathcal{M}_{n,d}$ into $\mathcal{R}_{n,m,d}$, which assigns to each matrix $M \in \mathcal{M}_{n,d}$ an $n \times m$ matrix $Q = (Q_{iq})_{iq} \in \mathcal{R}_{n,m,d}$ defined by

$$\forall i \leq n \ \ \forall q \leq m \quad Q_{iq} := \sum_{j \in \mathcal{L}^{(q)}} M_{ij},$$

that is, the matrix Q is obtained from M by summing up respective columns. This mapping defines an equivalence relation $\overset{y}{\sim}$ on $\mathcal{M}_{n,d}$, where $M \overset{y}{\sim} M'$ whenever both M and M' are mapped to the same element of $\mathcal{R}_{n,m,d}$. Further, a given matrix $Q \in \mathcal{R}_{n,m,d}$, we denote by $\mathcal{M}_{n,d}(Q, y)$ the equivalence class of matrices in $\mathcal{M}_{n,d}$, which are mapped to Q via the correspondence $\xrightarrow{(y)}$. If $\mathcal{M}_{n,d}(Q, y) \neq \emptyset$ then the uniform probability measure on $\mathcal{M}_{n,d}(Q, y)$ will be denoted by $\mathbb{P}_{Q,y}$.

For the rest of this section we fix integers $k \geq 1$, $m \geq 1$, and a vector $y \in \mathcal{A}_k$ with the ℓ-decomposition $(\mathcal{L}^{(q)})_{q=1}^m$. Let $Q \in \mathcal{R}_{n,m,d}$ be such that there exists $M \in \mathcal{M}_{n,d}$ which is mapped to Q by $\xrightarrow{(y)}$, in particular, for every $q \leq m$ one has

$$\sum_{i=1}^n Q_{iq} = d\, |\mathcal{L}^{(q)}|. \tag{21}$$

In what follows, such matrices Q will be called *y-admissible*. Denote $h_q := h(\mathcal{L}^{(q)})$. For all $i \leq n$ and $q \leq m$, we define *the weight w_{iq}* by

$$w_{iq} = w_{iq}(y, k, Q) := \begin{cases} h_q Q_{iq}/d, & \text{if } \mathcal{L}^{(q)} \text{ is regular,} \\ h_q \sqrt{Q_{iq}}, & \text{if } \mathcal{L}^{(q)} \text{ is spread.} \end{cases} \tag{22}$$

Now, given $i \leq n$, *the small ball probability estimator* SB_i is

$$\mathrm{SB}_i = \mathrm{SB}_i(y, k, Q) := \min\big(1, \min_{q \leq m} w_{iq}^{-1}\big),$$

where we adopt the convention $0^{-1} = \infty$. The estimators SB_i are designed to measure anti-concentration of inner products $\langle R_i(M), y^\dagger \rangle$, for M distributed in $\mathcal{M}_{n,d}(Q, y)$. We prove the following theorem.

Theorem 6.1 (Small ball probability). *Let d, n be large enough integers such that $d^3 \leq n$. Let $K \subset [n]$ be such that $|K^c| \leq n/(50 \ln d)$ and assume*

$$1 \leq k \leq \min\big(\sqrt{n}/(8 d^{3/2} \sqrt{\ln d}),\ d^{-10} e^{n/(5|K^c|)}\big).$$

Let y and Q be as above. Then for any non-random vector $V \in \mathbb{C}^{|K|}$ and any $\gamma \geq 1$ one has

$$\mathbb{P}_{Q,y}\big\{M \in \mathcal{M}_{n,d}(Q, y) :\ \|M^K y + V\|_2 \leq \gamma \sqrt{d|K|/k}\big\} \leq C^n \gamma^{2|K|} \prod_{i=1}^n \mathrm{SB}_i,$$

where $C > 0$ is a universal constant.

The main difficulty in proving Theorem 6.1 lies in the fact that the rows of the random matrix uniformly disctributed on $\mathcal{M}_{n,d}$ are dependent. To deal with this issue, we construct a special random vector Z in $\mathbb{C}^{|K|}$ with *independent* coordinates having a property that, conditioned on a certain event of not too small probability, it has the same distribution as $M^K y$.

Let y and K be as in the theorem. Recall that by the definition each ℓ-part $\mathcal{L}^{(q)}$ is representable as the union of level sets of y,

$$\mathcal{L}^{(q)} = \bigcup_{p=1}^{h_q} L_p^q, \tag{23}$$

where we assume for concreteness that $y(L_{p+1}^q) < y(L_p^q)$ (in lexicographical order) for all $p < h_q$. For each $q \le m$, we define the set of pairs

$$\Delta_q := \big\{(i,w) \;:\; 1 \le i \le n,\, 1 \le Q_{iq},\, 1 \le w \le Q_{iq}\big\}.$$

Then $|\Delta_q| = \sum\limits_{i=1}^{n} Q_{iq} = d|\mathcal{L}^{(q)}|$. Further, let

$$\big\{\xi_\delta^q \;:\; 1 \le q \le m,\, \delta \in \Delta_q\big\}$$

be a collection of jointly independent random variables, where each ξ_δ^q is distributed in the set $\{1,2,\dots,h_q\}$ in such a way that for all $p \le h_q$,

$$\mathbb{P}\big\{\xi_\delta^q = p\big\} = \frac{|L_p^q|}{|\mathcal{L}^{(q)}|}.$$

Define random variables Z_i, $i \in K$, as

$$Z_i := \sum_{q=1}^{m} \sum_{w=1}^{Q_{iq}} y\Big(L^q_{\xi^q_{(i,w)}}\Big) \tag{24}$$

and set $Z := (Z_i)_{i\in K}$. Note that each variable Z_i is a function of

$$\big\{\xi_{(i,w)}^q \;:\; 1 \le q \le m,\, 1 \le w \le Q_{iq}\big\},$$

and those sets of variables are clearly disjoint for distinct i's, hence $(Z_i)_{i\in K}$ are jointly independent. Since each Z_i is a sum of discrete complex-valued random variables, we can apply Proposition 2.5 to study its anti-concentration properties. As we show below, the conditional distribution of Z given a certain event of not too small probability, coincides with the distribution of $A^K y$, where A^K is a "multigraph" version of M^K in which we allow entries greater than one (i.e., multiple edges). This correspondence will be made precise later, as the first step we define and estimate the probability of the event to be conditioned on.

Claim 6.2. *Let h, $N, N_1, \dots, N_h$ be positive integers satisfying $\sum_{p=1}^h N_p = N$. Let $\xi_1, \xi_2, \dots, \xi_N$ be i.i.d. random variables taking values in the set $\{1,2,\dots,h\}$ with probabilities $\mathbb{P}\{\xi_i = p\} = N_p/N$ for all $p \le h$. Then*

$$\mathbb{P}\big\{\forall p \le h: \quad |\{i \le N \;:\; \xi_i = p\}| = N_p\big\} \ge (h/(Ne^2))^{h/2}.$$

Proof. Denote the event $\big\{\forall p \le h : |\{i \le N : \xi_i = p\}| = N_p\big\}$ by $\mathcal{E}$. Note that the random variables

$$\eta_p := |\{i \le N \;:\; \xi_i = p\}|,\; p \le h,$$

have a multinomial distribution. Since $(n/e)^n < n! \le e\sqrt{n}(n/e)^n$, we have

$$\mathbb{P}(\mathcal{E}) = \frac{N!}{N_1!\cdots N_h!}\prod_{p=1}^{h}\Big(\frac{N_p}{N}\Big)^{N_p} > \frac{(N/e)^N \prod_{p=1}^h {N_p}^{N_p}}{N^N \prod_{p=1}^h e\sqrt{N_p}(N_p/e)^{N_p}} = 1/\prod_{p=1}^{h} e\sqrt{N_p}.$$

The arithmetic-geometric mean inequality $\prod_{p=1}^h N_p \le (N/h)^h$ implies the bound. $\square$

Lemma 6.3. *Let $d \le n$ be large enough positive integers, $k \le \sqrt{n}/(8d^{3/2}\sqrt{\ln d})$, $y \in \mathcal{A}_k$, and $\{\xi_\delta^q\}$ be as above. Define the event*

$$\mathcal{E}_{6.3} := \bigcap_{q \le m} \big\{\forall p \le h_q : \quad |\{\delta \in \Delta_q : \xi_\delta^q = p\}| = d|L_p^q|\big\}.$$

Then $\mathbb{P}(\mathcal{E}_{6.3}) \ge e^{-n}$.

Proof. Let $H = \sum_{q=1}^m h_q$ be the total number of level sets of y. Since y is the k-approximation of a gradual vector, we have $y_{n_2}^* \le 2d^{\frac{3}{2}}$, with n_2 defined in Section 4. Therefore, using the assumption on k,

$$H \le n_2 + ((2k+1)\, 2d^{\frac{3}{2}})^2 \le n/(2\ln d).$$

By the independence of ξ_δ^q, $q \le m$, and by Claim 6.2 applied for every $q \le m$ with $N = d|\mathcal{L}^{(q)}|$ and $N_p = d|L_p^q|$, $p \le h_q$, we get

$$\begin{aligned}\mathbb{P}(\mathcal{E}_{6.3}) &\ge \prod_{q=1}^m \Big(\frac{h_q}{d|\mathcal{L}^{(q)}|e^2}\Big)^{h_q/2} = (e^2 d)^{-H/2} \prod_{q=1}^m \Big(\frac{h_q}{|\mathcal{L}^{(q)}|}\Big)^{h_q/2} \\ &\ge (e^2 d)^{-n/(4\ln d)} \prod_{q=1}^m e^{-|\mathcal{L}^{(q)}|/2e} = \exp\left(-n/(2\ln d) - n/4 - n/(2e)\right),\end{aligned}$$

where in the last inequality we used the bound on H and that $x^x \ge e^{-1/e}$ for all $x > 0$. This completes the proof for large enough d. $\square$

As the next step, we study anti-concentration properties of a single variable Z_i.

Lemma 6.4. *Given $k \ge 1$, let $y \in \mathcal{A}_k$ and let vector Z be defined as above. Then for every $i \in K$ and every $\tau \ge 1$ one has*

$$\mathcal{Q}(Z_i, \sqrt{d}\tau/k) \le C\tau^2\, \mathrm{SB}_i,$$

where $C \ge 1$ is a universal constant.

Proof. A simple estimate $\mathcal{Q}(\eta_1 + \eta_2, \gamma) \le \min\big(\mathcal{Q}(\eta_1, \gamma), \mathcal{Q}(\eta_2, \gamma)\big)$, which is valid for any pair η_1, η_2 of independent random variables and all $\gamma > 0$, together with the definitions of Z_i's and SB_i's, implies that it is sufficient to prove the relations

$$\mathcal{Q}\Big(\sum_{w=1}^{Q_{iq}} y\big(L^q_{\xi^q_{(i,w)}}\big), \frac{\sqrt{d}\tau}{k}\Big) \le \frac{C\tau^2}{w_{iq}}.$$

for all $i \in K$ and $q \le m$.

Fix $i \in K$ and $q \le m$ such that $w_{iq} \ne 0$, and denote the variables $y\big(L^q_{\xi^q_{(i,w)}}\big)$ by ψ_w, $1 \le w \le Q_{iq}$. Note that each ψ_w is a discrete random variable taking values in the set

$$B := \big\{y(L_p^q) \,:\, p \le h_q\big\}.$$

By (23) and by (17) one has

$$|\mathcal{L}^{(q)}| = \sum_{p=1}^{h_q} |L_p^q| \qquad \text{and} \qquad \max_{p \le h_q} |L_p^q| \le 4 \min_{p \le h_q} |L_p^q|.$$

Hence, for any $b \in B$

$$\mathbb{P}\{\psi_w = b\} \le \max_{p \le h_q} \frac{|L_p^q|}{|\mathcal{L}^{(q)}|} \le \frac{4}{h_q + 3}.$$

If the part $\mathcal{L}^{(q)}$ is regular then the set B is a $(1/k)$-separated subset of the complex plane. Applying Proposition 2.5 and using that $\tau \geq 1$ and $Q_{iq} \leq d$, we obtain

$$\mathcal{Q}\Big(\sum_{w=1}^{Q_{iq}} \psi_w, \frac{\sqrt{d}\tau}{k}\Big) \leq C_0 \max\Big(\frac{d\tau^2}{h_q Q_{iq}}, \frac{1}{h_q}\Big) = \frac{C_0\, d\tau^2}{h_q Q_{iq}} = \frac{C_0\,\tau^2}{w_{iq}},$$

where $C_0 > 0$ is a universal constant.

If the part $\mathcal{L}^{(q)}$ is spread then the set B is a (d/k)-separated subset of the complex plane. Note that $w_{iq} = h_q\sqrt{Q_{iq}} \leq h_q\sqrt{d}$. Without loss of generality, we can also assume that $1 \leq \tau^2 \leq w_{iq}$ (otherwise the probability estimate is trivial). Using that the number of (d/k)-separated points in a ball of radius λ is smaller than $\big(1+2\lambda k/d\big)^2$, we obtain for all $\lambda > 0$ and $w \leq Q_{iq}$,

$$\mathcal{Q}(\psi_w, \lambda) \leq \frac{4}{h_q+3}\big(1+2\lambda k/d\big)^2.$$

Assume first that $h_q \leq 32$. Using that the heights of non-empty spread parts h_q are at least 2, we have

$$\mathcal{Q}(\psi_w, \sqrt{d}/k) \leq \frac{4}{h_q+3}\big(1+2/\sqrt{d}\big)^2 \leq \frac{5}{6},$$

provided d is large enough. Therefore, applying Proposition 2.2 with $t = \sqrt{d}\tau/k$ and $t_0 = \sqrt{d}/k$, we get

$$\mathcal{Q}\Big(\sum_{w=1}^{Q_{iq}} \psi_w, \frac{\sqrt{d}\tau}{k}\Big) \leq \frac{C_1\,\tau^2}{\sqrt{Q_{iq}}} \leq \frac{32C_1\,\tau^2}{h_q\sqrt{Q_{iq}}} = \frac{32C_1\,\tau^2}{w_{iq}},$$

where $C_1 > 0$ is an absolute constant. Let now $h_q > 32$. Then, using $\tau^2 \leq h_q\sqrt{d}$ and that d is large enough,

$$\mathcal{Q}(\psi_w, \sqrt{d}\,\tau/k) \leq \mathcal{Q}(\psi_w, \sqrt{2d}\,\tau/k) \leq \frac{4}{h_q+3}\big(1+2\sqrt{2}\tau/\sqrt{d}\big)^2 \leq \frac{8+64\tau^2/d}{h_q} \leq 1/2.$$

Therefore, applying Proposition 2.3 with $t = t_0 = \sqrt{d}\tau/k$, we obtain

$$\mathcal{Q}\Big(\sum_{w=1}^{Q_{iq}} \psi_w, \frac{\sqrt{d}\tau}{k}\Big) \leq C_2\frac{8+64\tau^2/d}{h_q\sqrt{Q_{iq}}} \leq 72C_2\frac{\tau^2}{h_q\sqrt{Q_{iq}}} = 72C_2\frac{\tau^2}{w_{iq}},$$

where $C_2 > 0$ is an absolute constant. This completes the proof. $\square$

To complete the proof of Theorem 6.1 we tensorize the last lemma, i.e., we pass from the anti-concentration estimates for individual Z_i's to the vector Z, and then we tie the obtained estimates for Z with anti-concentration properties of $M^K y$. At this point, it will be convenient to introduce a new random object – a multigraph on $[n]$ which, in a certain sense, will correspond to the vector Z. This way, a direct relation between Z and $M^K y$ can be defined by conditioning on the event that the multigraph is simple, i.e., does not contain multiple edges.

Let y, $(\mathcal{L}^{(q)})_{q=1}^m$, and Q be as above. We construct the multigraph $\widehat{G}_Q$ on $[n]$ as the union of certain independent bipartite multigraphs $\widehat{G}_q$ on $([n], \mathcal{L}^{(q)})$, that is

$$\widehat{G}_Q = \bigcup_{q=1}^{m} \widehat{G}_q,$$

where to form $\widehat{G}_q$ we adapt the configuration model in the following way. For every $q \leq m$, define

$$\Delta_q' := \mathcal{L}^{(q)} \times [d] = \big\{(j, w') : j \in \mathcal{L}^{(q)}, 1 \leq w' \leq d\big\}.$$

Clearly, $|\Delta'_q| = d|\mathcal{L}^{(q)}| = |\Delta_q|$. Let g_q be a (fixed) bijection from Δ_q to Δ'_q, and let σ_q be a random uniform permutation on Δ'_q (σ_q does not respect the two-dimensional structure of Δ'_q and can be viewed as a uniform random element of the permutation group $\Pi_{|\Delta'_q|}$, also we suppose that σ_1, ..., σ_m are jointly independent). We define $\widehat{G}_q$ as a bipartite multigraph on $([n], \mathcal{L}^{(q)})$ with the edge multiset

$$E_q := \Big\{(i,j) \in [n] \times \mathcal{L}^{(q)} \ : \exists\, 1 \le w \le Q_{iq}\ , 1 \le w' \le d \text{ such that}$$
$$(i,w) \in \Delta_q,\ (j,w') \in \Delta'_q \text{ and } g_q(i,w) = \sigma_q(j,w')\Big\},$$

where the multiplicity $r_q(i,j)$ of each edge (i,j) in E_q is equal to the cardinality of the set

$$g_q(\{i\} \times [Q_{iq}]) \cap \sigma_q(\{j\} \times [d]).$$

Note that by construction, $\widehat{G}_q$ has degree sequence $(Q_{iq})_{i \le n}$ for vertices in $[n]$ and a constant degree d for vertices in $\mathcal{L}^{(q)}$. We define $\widehat{G}_Q$ as the union of $\widehat{G}_q$'s, $q \le m$, in particular, the edge multisets E of $\widehat{G}_Q$ is

$$E = \bigcup_{q=1}^m E_q.$$

Denote

$$p_q = p_q(i,j) := \sum_{(w,w') \in [Q_{iq}] \times [d]} \mathbb{P}\{g_q(i,w) = \sigma_q(j,w')\} \le Q_{iq}/|\mathcal{L}^{(q)}|.$$

Then $\mathbb{E}\{r_q(i,j) \mid (i,j) \in E_q\} = p_q$, and by the union bound $\mathbb{P}\{(i,j) \in E_q\} \le p_q$. Therefore, using that each Q_{iq} is at most d, we observe

$$\mathbb{E}\, r_q(i,j) = \mathbb{E}\{r_q(i,j) \mid (i,j) \in E_q\}\ \mathbb{P}\{(i,j) \in E_q\} \le p_q^2 \le Q_{iq}^2/|\mathcal{L}^{(q)}|^2 \le d\, Q_{iq}/|\mathcal{L}^{(q)}|^2.$$

Let N_q be the total number of multiple edges produced in the random bipartite multigraph $\widehat{G}_q$. Using Markov's inequality and (21), we obtain

$$\mathbb{P}\big\{N_q \ge 2d^2\big\} \le \frac{1}{2d^2}\, \mathbb{E}\, N_q \le \frac{1}{2d^2} \sum_{i \le n} \sum_{j \le |\mathcal{L}^{(q)}|} \mathbb{E}\, r_q(i,j) \le \frac{1}{2d} \sum_{i \le n} \frac{Q_{iq}}{|\mathcal{L}^{(q)}|} \le \frac{1}{2}. \tag{25}$$

Thus for every $q \le m$ at least half of realizations of the random bipartite multigraph $\widehat{G}_q$ have the number of multiple edges at most $2d^2$. In the sequel we will see that for every $q \le m$ a non-negligible part of realizations of $\widehat{G}_q$ have no multiple edges, so that a non-negligible proportion of realizations of $\widehat{G}_Q$ are simple.

One can check that any realization of $\widehat{G}_Q$ occurs with probability

$$\prod_{q=1}^m \frac{(d!)^{|\mathcal{L}^{(q)}|} \prod_{i=1}^n Q_{iq}!}{(d|\mathcal{L}^{(q)}|)!}.$$

Moreover, the realizations of $\widehat{G}_Q$ which are simple precisely correspond to the graphs whose adjacency matrices belong to $\mathcal{M}_{n,d}(Q,y)$. Therefore,

$$\mathbb{P}\{\widehat{G}_Q \text{ is simple}\} = |\mathcal{M}_{n,d}(Q,y)| \prod_{q=1}^m \frac{(d!)^{|\mathcal{L}^{(q)}|} \prod_{i=1}^n Q_{iq}!}{(d|\mathcal{L}^{(q)}|)!}. \tag{26}$$

Below we denote the adjacency matrix of $\widehat{G}_Q$ by A (with the entries of A respecting multiplicities). Then for any $M \in \mathcal{M}_{n,d}(Q,y)$ one has

$$\mathbb{P}\{A = M \mid \widehat{G}_Q \text{ is simple}\} = \Big(\mathbb{P}\{\widehat{G}_Q \text{ is simple}\}\Big)^{-1} \prod_{q=1}^m \frac{(d!)^{|\mathcal{L}^{(q)}|} \prod_{i=1}^n Q_{iq}!}{(d|\mathcal{L}^{(q)}|)!} = \frac{1}{|\mathcal{M}_{n,d}(Q,y)|}, \tag{27}$$

which means that conditioned on the event that $\widehat{G}_Q$ is simple, the matrix A is uniformly distributed on $\mathcal{M}_{n,d}(Q,y)$.

In the next proposition, we provide a lower bound on the probability that $\widehat{G}_Q$ is simple. Note that by our construction, this probability is equal to the product of the probabilities that each of the bipartite graphs $\widehat{G}_q$ is simple.

Proposition 6.5. *Let $d \le n$ be large enough and denote by $\mathcal{E}_{6.5}$ the event that $\widehat{G}_Q$ is simple. Then*

$$\mathbb{P}\big(\mathcal{E}_{6.5}\big) \ge \exp(-33d^2 \ln^2 n).$$

Proof. For every $q \le m$ such that $|\mathcal{L}^{(q)}| \le 5d$, we bound from below the probability that $\widehat{G}_q$ is simple by one over the number of realizations of such multigraphs, that is, by

$$\big(d|\mathcal{L}^{(q)}|\big)^{-d|\mathcal{L}^{(q)}|} \ge \exp(-5d^2 \ln(5d)^2) \ge \exp(-11d^2 \ln n).$$

Now we treat $q \le m$ such that $|\mathcal{L}^{(q)}| > d \ln n$. For such q we could use precise asymptotics obtained in [44] (see also [28, Theorem 1.1]), however, for the readers' convenience, we prefer to provide a simple self-contained argument (which leads also to a better bound). For every $q \le m$, denote by $\mathcal{M}_{n,d}(Q,y,\mathcal{L}^{(q)})$ the set of $n \times |\mathcal{L}^{(q)}|$ matrices corresponding to blocks of columns indexed by $\mathcal{L}^{(q)}$ of matrices from the equivalence class $\mathcal{M}_{n,d}(Q,y)$. With this notation, we have

$$|\mathcal{M}_{n,d}(Q,y)| = \prod_{q=1}^{m} |\mathcal{M}_{n,d}(Q,y,\mathcal{L}^{(q)})|.$$

Similarly to (26), the probability that the random multigraph $\widehat{G}_q$ on $([n],\mathcal{L}^{(q)})$ is simple is given by

$$|\mathcal{M}_{n,d}(Q,y,\mathcal{L}^{(q)})| \, \frac{(d!)^{|\mathcal{L}^{(q)}|} \prod_{i=1}^{n} Q_{iq}!}{(d|\mathcal{L}^{(q)}|)!}.$$

Therefore it is sufficient to estimate the cardinality of $\mathcal{M}_{n,d}(Q,y,\mathcal{L}^{(q)})$ for each $q \le m$. Let $\widehat{\mathcal{M}}_q$ be the set of all adjacency matrices corresponding to realizations of $\widehat{G}_q$ (with the entries respecting multiplicities). Moreover, let $\widehat{\mathcal{M}''}$ be the subset of $\widehat{\mathcal{M}}_q$ given by matrices such that the sum over entries exceeding 1 is bounded above by $2d^2$. The latter corresponds to multigraphs having at most $2d^2$ multiple edges. By (25), we have

$$|\widehat{\mathcal{M}''}| \ge \frac{1}{2}|\widehat{\mathcal{M}}_q|. \tag{28}$$

To estimate the cardinality of $\mathcal{M}_{n,d}(Q,y,\mathcal{L}^{(q)})$, we define a relation $R \in \widehat{\mathcal{M}''} \times \mathcal{M}_{n,d}(Q,y,\mathcal{L}^{(q)})$ as follows. We let a pair (M,M') belong to R if M' can be obtained from M by a sequence (of maximal length $2d^2$) of simple switching operations in the following way: for every (i,j) such that $M_{ij} > 1$, choose first (i',j') such that $M_{ij'} = M_{i'j} = 0$ and $M_{i'j'} \ge 1$, then operate the simple switching on i,j,i',j'. By regularity of our matrices and by the definition of $\widehat{\mathcal{M}''}$, the number of pairs (i',j') with $M_{i'j'} \ge 1$ is at least

$$\sum_{s,t} M_{s,t} - \sum_{s,t:\, M_{s,t} \ge 1} (M_{s,t} - 1) \ge d|\mathcal{L}^{(q)}| - 2d^2.$$

Moreover, the number of s and t such that either $M_{is} \ne 0$ or $M_{tj} \ne 0$ is at most $2d(d-1)$. Therefore, using that $|\mathcal{L}^{(q)}| \ge 5d$, we observe that there are at least $d|\mathcal{L}^{(q)}| - 4d^2 \ge d|\mathcal{L}^{(q)}|/5$ choices for a "good" pair (i',j'), hence

$$|R(M)| \ge d|\mathcal{L}^{(q)}|/5.$$

Note that after such a switching, the sum over entries exceeding 1 must decrease. Then we reiterate this procedure until the all entries becomes less than or equal to 1. Note that we do not need more than $2d^2$ steps (since the sum over entries exceeding 1 is bounded above by $2d^2$). Now we revert the procedure and start with $M' \in R(\widehat{\mathcal{M}''})$. Since at each step the number of non-zero elements is at most $d|\mathcal{L}^{(q)}|$, the number of possible switching operations is smaller than $d^2|\mathcal{L}^{(q)}|^2/2$. Since the number of steps is at most $2d^2$, we have

$$|R^{-1}(M')| \leq \big(d^2|\mathcal{L}^{(q)}|^2/2\big)^{2d^2}.$$

Claim 2.1 and the bound $d|\mathcal{L}^{(q)}| \leq n^2$ imply that

$$|\widehat{\mathcal{M}''}| \leq \big(5/d|\mathcal{L}^{(q)}|\big)\ \big(d^2|\mathcal{L}^{(q)}|^2/2\big)^{2d^2}\ |\mathcal{M}_{n,d}(Q,y,\mathcal{L}^{(q)})| \leq \big(1/2\big)\ \exp\big(8d^2\ln n\big)\ |\mathcal{M}_{n,d}(Q,y,\mathcal{L}^{(q)})|.$$

By (28) this yields

$$|\mathcal{M}_{n,d}(Q,y,\mathcal{L}^{(q)})| \geq \big(d|\mathcal{L}^{(q)}|/4\big)\ \big(\sqrt{2}/(d|\mathcal{L}^{(q)}|)\big)^{4d^2}\ \geq \exp\big(-8d^2\ln n\big)\ |\widehat{\mathcal{M}}_q|,$$

hence, the probability that the $\widehat{G}_q$ is simple is at least $\exp(-8d^2\ln n)$.

Finally, as we mentioned above, the probability that $\widehat{G}_Q$ is simple is equal to the product of the probabilities that each $\widehat{G}_q$, $q \leq m$, is simple. Thus, the probability that $\widehat{G}_Q$ is simple is at least $\exp\big(-11d^2\,m\ln n\big)$. Since by the construction of the ℓ-decomposition, $m \leq 3\ln n$, we obtain the desired estimate. □

We now verify that the adjacency matrix A of $\widehat{G}_Q$ satisfies a condition similar to that of Theorem 6.1.

Lemma 6.6. *Let $d \leq n$ be large enough, $m \geq 1$, and $k \leq \sqrt{n}/(8d^{3/2}\sqrt{\ln d})$. Let $y \in \mathcal{A}_k$, $(\mathcal{L}^{(q)})_{q=1}^m$ be its ℓ-decomposition, Q be a y-admissible matrix, and $\widehat{G}_Q$ be the random multigraph constructed above with the adjacency matrix A. Then for any $K \subset [n]$, any non-random vector $V \in \mathbb{C}^{|K|}$, and any $\gamma \geq 1$ one has*

$$\mathbb{P}\Big\{\|A^K y + V\|_2 \leq \gamma\,\frac{\sqrt{d|K|}}{k}\Big\} \leq C_{6.6}^n\gamma^{2|K|}\prod_{i\in K}\mathrm{SB}_i.$$

Proof. As before, for every $q \leq m$ we represent the ℓ-set $\mathcal{L}^{(q)}$ as the union $\bigcup_{p\leq h_q} L_p^q$. For every $j \in \mathcal{L}^{(q)}$ let $f^q(j)$ denote the index of the level set in $\mathcal{L}^{(q)}$ containing j, i.e., $f^q(j) = p$ whenever $j \in L_p^q$. It is convenient to have a representation for the multiset E_q in the form

$$E_q = \big\{e_\delta :\ \delta \in \Delta_q\big\},$$

where for every $\delta = (i,j) \in \Delta_q$ we have $e_\delta = (i, j_\delta^q)$ with $j_\delta^q \in \mathcal{L}^{(q)}$ being equal to the first component of the pair

$$\sigma_q^{-1}(g_q(\delta)) \in \Delta'_q.$$

Recall that random variables $\{\xi_\delta^q\}$, $q \leq m$, $\delta \in \Delta_q$, were introduced in the first part of this section. Observe that for any fixed $q \leq m$, the joint distribution of the variables $\{\xi_\delta^q\}$, $\delta \in \Delta_q$, conditioned on the event $\mathcal{E}_{6.3}$, coincides with the joint distribution of

$$\{f^q(j_\delta^q) :\ \delta \in \Delta_q\}.$$

Indeed, by the construction of the multigraph $\widehat{G}_Q$, the variables $\big(f^q(j_\delta^q)\big)_{\delta\in\Delta_q}$ take values in the set of sequences

$$\big\{(a_\delta)_{\delta\in\Delta_q} \in \mathbb{N}^{\Delta_q} :\qquad \forall p \leq h_q \qquad |\{\delta :\ a_\delta = p\}| = d|L_p^q|\big\}\,. \tag{29}$$

Note that the set of permutations of Δ_q acts transitively on the set in (29). Hence, taking into account that the distribution of $\big(f^q(j_\delta^q)\big)_{\delta\in\Delta_q}$ is invariant under permutations of Δ_q, we get that all realizations of the sequence are equi-probable. On the other hand, conditioned on $\mathcal{E}_{6.3}$, the sequence $(\xi_\delta^q)_{\delta\in\Delta_q}$ is distributed over (29), and it is not difficult to see that the conditional distribution is uniform. Thus, the distribution of $\big(f^q(j^q_\delta)\big)_{\delta\in\Delta_q}$ and $(\xi^q_\delta)_{\delta\in\Delta_q}$ given $\mathcal{E}_{6.3}$ must coincide.

We now relate the coordinates of the vector $A^K y$ to the variables Z_i. Denoting the entries of A by a_{ij}, note that for every pair (i,j) the entry a_{ij} is the multiplicity of the edge (i,j) in E_q (which can be zero if the edge does not belong to E_q). Therefore for all $i\in K$, we have

$$(Ay)_i=\sum_{j=1}^n a_{ij}y_j=\sum_{q=1}^m\sum_{j\in\mathcal{L}^{(q)}} a_{ij}y_j=\sum_{q=1}^m\sum_{j\in\mathcal{L}^{(q)}} y\big(L^q_{f^q(j)}\big)\,a_{ij}=\sum_{q=1}^m\sum_{w=1}^{Q_{iq}} y\big(L^q_{f^q(j^q_{(i,w)})}\big).$$

Hence,

$$A^K y\overset{d}{\sim} Z \text{ conditioned on } \mathcal{E}_{6.3}.$$

Applying Lemma 6.3, we get

$$\begin{aligned}\mathbb{P}\Big\{\|A^K y+V\|_2\le\gamma\,\frac{\sqrt{d|K|}}{k}\Big\}&=\mathbb{P}\Big\{\|Z+V\|_2\le\gamma\,\frac{\sqrt{d|K|}}{k}\;\Big|\;\mathcal{E}_{6.3}\Big\}\\&\le e^n\,\mathbb{P}\Big\{\|Z+V\|_2\le\gamma\,\frac{\sqrt{d|K|}}{k}\Big\}.\end{aligned}$$

By Lemma 6.4 there is an absolute constant $C\ge 1$ such that for all $\gamma\ge 1$,

$$\mathcal{Q}(Z_i,\gamma\sqrt{d}/k)\le C\gamma^2\,\mathrm{SB}_i.$$

Therefore, the random vector $Z+v$ satisfies the assumptions of Lemma 2.7 with $\varepsilon_0=\sqrt{d}/k$ and $p_i=Ck^2\,\mathrm{SB}_i/d$. Applying this lemma and taking $\varepsilon=\gamma\varepsilon_0$, we obtain

$$\mathbb{P}\Big\{\|Z+V\|_2\le\gamma\,\frac{\sqrt{d|K|}}{k}\Big\}\le C_1^{|K|}\gamma^{2|K|}\prod_{i\in K}\mathrm{SB}_i,$$

where $C_1\ge 1$ is an absolute constant. This completes the proof. □

Before we complete the proof of Theorem 6.1, we show that $\prod_{i=1}^n\mathrm{SB}_i$ and $\prod_{i\in K}\mathrm{SB}_i$ are comparable whenever K is not too small.

Lemma 6.7. *Let d,n be large enough integers with $d^3\le n$. Let $k\ge 1$ and $K\subset[n]$ be such that*

$$|K^c|\le n/(50\ln d)\quad\text{ and }\quad k\le d^{-10}e^{n/(5|K^c|)}$$

and let $y\in\mathcal{A}_k$. Then

$$\prod_{i\in K^c}\mathrm{SB}_i^{-1}\le e^n.$$

Proof. Since $Q_{iq}\le d$, we have $w_{iq}\le h_q\sqrt{d}$ for all $i\le n$ and $q\le m$. For each $b\ge 1$, denote

$$I_b:=\{q\le m:\,h_q\in[2^{b-1},\,2^b)\}\qquad\text{and}\qquad\mathcal{H}_b:=\bigcup_{q\in I_b}\mathcal{L}^{(q)}.$$

Let b_0 be such that

$$2^{b_0-1}\sqrt{d}\le e^{n/(2|K^c|)}<2^{b_0}\sqrt{d}.$$

Note that the assumption on the cardinality of K implies that $2^{b_0}\ge d^{24}$. Since for every

$$q\in I_0:=\bigcup_{b<b_0}I_b$$

we have $w_{iq} \le h_q\sqrt{d} \le e^{n/(2|K^c|)}$, then

$$S := \prod_{i\in K^c} \mathrm{SB}_i^{-1} \le \prod_{i\in K^c} e^{n/(2|K^c|)} \max(1, \max_{q\notin I_0} w_{iq}) \le e^{n/2} \prod_{b\ge b_0} \prod_{i\in K^c} \max(1, \max_{q\in I_b} w_{iq}). \tag{30}$$

Fix $b \ge b_0$. By the construction of the matrix Q (from a d-regular matrix M), there are at most $d|\mathcal{H}_b|$ indices i for which $Q_{iq} \ne 0$ for some $q \in I_b$. Therefore,

$$\prod_{i\in K^c} \max(1, \max_{q\in I_b} w_{iq}) \le (2^b\sqrt{d})^{d|\mathcal{H}_b|}. \tag{31}$$

Let j be the maximal order of ℓ-parts in the definition of $\mathcal{H}_b$ and $\mathcal{L}$ be a corresponding ℓ-part (if there are two of them, spread and regular, we choose and fix one). Denote $h := h(\mathcal{L})$. By the definition of $\mathcal{H}_b$, the construction of the ℓ-decomposition, (17), and (18), we have

$$2^{b_0-1} \le 2^{b-1} \le h < 2^b \quad \text{ and } \quad |\mathcal{H}_b| \le 2 \max_{q\in I_b} h_q \sum_{i=0}^{j} 2^{i+1} \le 2^{j+b+3}.$$

Moreover, the size of each level set in $\mathcal{L}$ is in the interval $[2^{j-1}, 2^{j+1}]$, in particular, $|\mathcal{L}| \in [2^{j-1}h,\ 2^{j+1}h]$. Since y is a k-vector, its levels are $1/k$-separated. Thus, using that for any $p > 0$ there are at most $(2p+1)^2$ integer complex numbers of absolute value less or equal p, we obtain for $s = \lceil 2^{j-3}h\rceil$:

$$y_s^* > \frac{\sqrt{h/8}}{k} \ge \frac{2^{b/2}}{4k} \ge \frac{2^{b_0/2}}{4k} \ge \frac{e^{n/(4|K^c|)}}{4kd^{1/4}} \ge 2,$$

provided that $k \le e^{n/(4|K^c|)}/(8kd^{1/4})$. Now we use that y is the k-approximation of a vector $x \in \mathcal{S}$. Since $x_{n_3}^* = 1$, we observe $s < n_3$. On the other hand, applying Lemma 4.2 to x,

$$y_s^* \le 2x_s^* \le d^3(n/s)^6,$$

which implies

$$2^{j+b} \le 2^{j+1}h \le 16s \le 16n\sqrt{d}\,(4k)^{1/6}\,2^{-b/12}.$$

Therefore,

$$|\mathcal{H}_b| \le 2^{j+b+3} \le C'n\sqrt{d}\,k^{1/6}\,2^{-b/12}$$

for a universal constant $C' > 0$. This, together with (30), (31), and $2^b \ge 2^{b_0} \ge e^{n/(2|K^c|)}/\sqrt{d} \ge d$, implies for an appropriate absolute positive constant C,

$$\begin{aligned} S &\le e^{n/2}\, \exp\Big(\sum_{b\ge b_0} C'nd^{3/2}k^{1/6}2^{-b/12}\ln(2^b\sqrt{d})\Big) \le e^{n/2}\, \exp\Big(Cnd^{3/2}k^{1/6}b_0\,2^{-b_0/12}\Big) \\ &\le e^{n/2}\, \exp\big((n/2)d^{3/2}k^{1/6}\big(\sqrt{d}e^{-n/(2|K^c|)}\big)^{1/13}\big) \le e^n, \end{aligned}$$

provided that $k \le d^{-10}e^{n/5|K^c|}$ and that d is large enough. □

Proof of Theorem 6.1. Following our configuration-type model construction, the law of the adjacency matrix A of the multigraph $\widehat{G}_Q$, conditioned on the event that $\widehat{G}_Q$ is simple, coincides with the uniform distribution on $\mathcal{M}_{n,d}(Q,y)$ (see (27)). Using this and applying Proposition 6.5 and Lemma 6.6, we obtain

$$\begin{aligned} &\mathbb{P}_{Q,y}\big\{M \in \mathcal{M}_{n,d}(Q,y) :\ \|M^K y + V\|_2 \le \gamma\sqrt{d|K|}/k\big\} \\ &= \mathbb{P}\Big\{\|A^K y + V\|_2 \le \gamma\,\sqrt{d|K|}/k \,\Big|\, \mathcal{E}_{6.5}\Big\} \\ &\le \mathbb{P}\Big\{\|A^K y + V\|_2 \le \gamma\,\sqrt{d|K|}/k\Big\}/\mathbb{P}(\mathcal{E}_{6.5}) \\ &\le \exp(33d^2\ln^2 n)\,C_{6.6}^n\gamma^{2|K|}\prod_{i\in K}\mathrm{SB}_i. \end{aligned}$$

Lemma 6.7 implies the desired result. □

7 Proof of Theorem 1.1

This section is devoted to the proof of the main result of the paper, obtained by a combination of the estimates for steep and almost constant vectors from Section 4, the structural information on the set of gradual vectors from Section 5 and the small ball probability theorem of Section 6. Setting aside technical details, the principal idea of the proof is to define a discrete structure on the set of gradual vectors with "not small" subsets of almost equal coordinates and, by a combination of small ball probability estimates for individual vectors (Lemmas 7.7 and 7.9), the union bound and an approximation argument (Proposition 7.6), to eliminate those vectors from the set of "potential" null vectors of our matrix. Construction of the discrete subset (which can be viewed as a collection of nets with respect to ℓ_∞-metric in $\mathbb{C}^n$) is quite involved – it uses rather complex information about the structure of a gradual vector (the ℓ-decompositions of its k-approximations) which affects both cardinalities of the nets and the probability estimates. As for the latter, to simplify analysis of the product $\prod_{i=1}^n \mathrm{SB}_i$, we introduce another set of estimators $\{\mathrm{TE}_i\}_{i=1}^n$, which we call *trivial estimators* (see Subsection 7.1 for definitions). The product $\prod_{i=1}^n \mathrm{SB}_i$ is then estimated in terms of $\prod_{i=1}^n \mathrm{TE}_i$ and an auxiliary functional η (also defined in Subsection 7.1). Note that, by the definition, the probability estimators SB_i depend both on the structure of the underlying k-vector and on statistics of the corresponding matrix Q. By introducing the estimators TE_i and the functional η, we "separate" these dependencies: the trivial estimators are entirely determined by the ℓ-decomposition of the related k-vector while η carries information about the matrix Q and the ℓ-decomposition in a much more convenient form compared to SB_i's.

The next informal argument, following the universality paradigm of the random matrix theory, may be useful as an illustration of our approach. Given the random matrix M, we may think that null vectors of $(M - z\,\mathrm{Id})^K$ behave essentially like Gaussian vectors (up to rescaling). In particular, this would imply that the ℓ-decompositions of k-approximations of the null vectors are comprised of ℓ-parts which are "mostly" regular and, moreover, there are very few ℓ-sets of comparable (up to a constant multiple) cardinalities. Accordingly, vectors whose ℓ-decompositions contain "many" spread ℓ-parts or many ℓ-parts of approximately equal cardinalities, should be typically in the complement of the matrix kernel. This imprecise observation can in fact be rigorously verified, in particular, we show that vectors with large spread ℓ-parts (from the sets $\mathcal{K}_u$) are not in the kernel with high probability.

Before we pass to the probability estimators and computing the union bound over a discrete subset of $\mathcal{S}$, we reformulate the main statement of Section 6. We first construct the following subset $\mathcal{R}_{n,m,d}^{ST}$ ("ST" stands for "standard") of $\mathcal{R}_{n,m,d}$. First consider the subset of matrices $Q = (Q_{iq}) \in \mathcal{R}_{n,m,d}$ such that

1. For every $q \le m$ with $\|C_q(Q)\|_1 = \sum_{i=1}^n Q_{iq} \ge \sqrt{d}n$ one has

$$\big|\{i \le n :\, Q_{iq} < c_{3.3}\|C_q(Q)\|_1/n\}\big| \le n/\sqrt{d};$$

2. For every non-empty subset $J \subset [m]$ and $\kappa := \sum\limits_{q\in J} \|C_q(Q)\|_1$ one has

$$\Big|\Big\{i \le n :\, \sum_{q\in J} Q_{iq} \ge c_{3.5}\frac{\kappa}{n} \text{ and } \sum_{q\in J^c} Q_{iq} \ge c_{3.5}\frac{dn-\kappa}{n}\Big\}\Big| \ge c_{3.5}\min\big(\kappa, (dn-\kappa), n\big).$$

We denote this subset by $\mathcal{R}_{n,m,d}^{ST_0}$. Note that in Corollary 3.3 and Proposition 3.5, we showed that the event

$$\mathcal{E} = \Big\{ M \in \mathcal{M}_{n,d} :\ \forall J \subset [n] \text{ one has}$$
$$\Big| \Big\{ i \le n :\ |\operatorname{supp} R_i(M) \cap J| \ge c_{3.5} \frac{d|J|}{n} \text{ and } |\operatorname{supp} R_i(M) \cap J^c| \ge c_{3.5} \frac{d|J^c|}{n} \Big\} \Big|$$
$$\ge c_{3.5} \min(d|J|, d(|J^c|), n) \quad \text{AND}$$
$$\text{if } |J| \ge n/\sqrt{d} \text{ one has } \Big| \Big\{ i \le n :\ |\operatorname{supp} R_i(M) \cap J| < \frac{c_{3.3} d|J|}{n} \Big\} \Big| \le n/\sqrt{d} \Big\}$$

has probability very close to one. Now, if $M \in \mathcal{E}$ and y is a k-vector with m non-empty ℓ-parts in its ℓ-decomposition then the correspondence $\xrightarrow{(y)}$ necessarily maps M into $\mathcal{R}^{ST_0}_{n,m,d}$. The image of $\mathcal{E}$ we denote by $\mathcal{R}^{ST}_{n,m,d}$. Thus, the preimage of $\mathcal{R}^{ST}_{n,m,d}$ with respect to $\xrightarrow{(y)}$ is almost the entire set $\mathcal{M}_{n,d}$. This fact combined with Theorem 6.1 gives the following proposition.

Proposition 7.1. *Let d, n be large enough integers such that $d^3 \le n$. Let $K \subset [n]$ be such that $|K^c| \le n/(50 \ln d)$ and assume*

$$1 \le k \le \min\big(\sqrt{n}/(8d^{3/2}\sqrt{\ln d}),\ d^{-10} e^{n/(5|K^c|)}\big).$$

Let $y \in \mathcal{A}_k$ and $A > 0$ be such that $\prod_{i=1}^{n} \mathrm{SB}_i \le A$ for every $Q \in \mathcal{R}^{ST}_{n,m,d}$. Then for any non-random vector $V \in \mathbb{C}^{|K|}$ and any $\gamma \ge 1$ we have

$$\mathbb{P}\big\{ M \in \mathcal{M}_{n,d} :\ \|M^K y + V\|_2 \le \gamma \sqrt{d|K|}/k \ \big|\ \mathcal{E} \big\} \le 2C^n \gamma^{2|K|} A,$$

where $C > 0$ is an absolute constant.

Proof. As we already noted, the event $\mathcal{E}$ is contained inside

$$\mathcal{E}_0 := \{ M :\ M \text{ is mapped into } \mathcal{R}^{ST}_{n,m,d} \text{ via } \xrightarrow{(y)} \}.$$

By Corollary 3.3 and Proposition 3.5, we have $\mathbb{P}(\mathcal{E}_0) \ge \mathbb{P}(\mathcal{E}) \ge 1/2$, in particular, $\mathbb{P}(\mathcal{E}_0)/\mathbb{P}(\mathcal{E}) \le 2$. This and Theorem 6.1 imply

$$\mathbb{P}\big\{ M \in \mathcal{M}_{n,d} :\ \|M^K y + V\|_2 \le \gamma \sqrt{d|K|}/k \ \big|\ \mathcal{E} \big\}$$
$$\le 2\mathbb{P}\big\{ M \in \mathcal{M}_{n,d} :\ \|M^K y + V\|_2 \le \gamma \sqrt{d|K|}/k \ \big|\ \mathcal{E}_0 \big\} \le 2C^n \gamma^{2|K|} A.$$

□

7.1 Rough estimators for small ball probability

We now introduce a "rougher" estimator than SB_i, which is easier to study. Note that conditioned on the event $\mathcal{E}$ defined above, if $|\mathcal{L}^{(q)}| \ge n/\sqrt{d}$ then for most rows, Q_{iq} is of the same order of magnitude as $d|\mathcal{L}^{(q)}|/n$. Having this in mind, we replace non-zero Q_{iq} in the definition of weights w_{iq} in (22) with $d|\mathcal{L}^{(q)}|/n$ if $|\mathcal{L}^{(q)}| \ge d^{-1/3}n$ and with 1 otherwise and come to the following definition. Given a vector $y \in \mathcal{A}_k$ with the corresponding ℓ-decomposition $(\mathcal{L}^{(q)})_{q=1}^m$, let

$$\widetilde{w}_q = \widetilde{w}_q(y,k) := \begin{cases} h_q |\mathcal{L}^{(q)}|/n, & \text{if } |\mathcal{L}^{(q)}| \ge d^{-1/3}n \text{ and } \mathcal{L}^{(q)} \text{ is regular,} \\ h_q/d, & \text{if } |\mathcal{L}^{(q)}| < d^{-1/3}n \text{ and } \mathcal{L}^{(q)} \text{ is regular,} \\ h_q \sqrt{d\,|\mathcal{L}^{(q)}|/n}, & \text{if } |\mathcal{L}^{(q)}| \ge d^{-1/3}n \text{ and } \mathcal{L}^{(q)} \text{ is spread,} \\ h_q & \text{if } |\mathcal{L}^{(q)}| < d^{-1/3}n \text{ and } \mathcal{L}^{(q)} \text{ is spread} \end{cases}$$

be *the truncated weight* ("truncated" because for small non-zero Q_{iq} we replace it with its minimal possible value 1). The reason why we truncate at level $d^{-1/3}n$ instead of $n/\sqrt{d}$ is that we want the weights $\widetilde{w}_q$ to be significantly weaker than w_{iq}. Specifically, this definition of the truncated weights will allow us to estimate the product $\prod_{i=1}^n \mathrm{SB}_i$ from above in terms of *trivial estimators* defined with respect to the truncated weights (see below). Note that for any $q \le m$, we have

$$\widetilde{w}_q \ge h_q/d \quad \text{if } \mathcal{L}^{(q)} \text{ is regular} \qquad \text{and} \qquad \widetilde{w}_q \ge h_q \quad \text{if } \mathcal{L}^{(q)} \text{ is spread.} \tag{32}$$

The principal difference between the weights $\widetilde{w}_q$ and w_{iq} is that $\widetilde{w}_q$ does not depend on Q_{iq}, which makes its analysis easier. We define also *the trivial estimator* TE_i, $i \le n$, as a weighted geometric mean of $\widetilde{w}_q$, $q \le m$,

$$\mathrm{TE}_i := \mathrm{TE}_i(y,k,Q) := \prod_{q\le m} (\widetilde{w}_q)^{-Q_{iq}/d}.$$

Note that by (21) and (32),

$$\prod_{i=1}^n \mathrm{TE}_i = \prod_{q\le m} (\widetilde{w}_q)^{-|\mathcal{L}^{(q)}|} \le d^n \prod_{q\le m} h_q{}^{-|\mathcal{L}^{(q)}|}. \tag{33}$$

In what follows, we usually do not mention explicitly dependency of the weights and the estimators on the vector y, which is assumed to be fixed throughout most of the subsection. To compare SB_i and TE_i, we introduce more notations. For each $b \in \mathbb{Z}$, let $\mathcal{W}_b$ be the union of all spread and regular ℓ-parts $\mathcal{L}^{(q)}$, whose truncated weights $\widetilde{w}_q$ lie in the interval $[2^b, 2^{b+1})$, that is, we set

$$I(b) := \{q \,:\, 2^b \le \widetilde{w}_q < 2^{b+1}\} \qquad \text{and} \qquad \mathcal{W}_b := \bigcup_{I(b)} \mathcal{L}^{(q)}$$

(we will call $\mathcal{W}_b$ *the w-set of order* b). For a fixed $i \le n$, define

$$b(i) := \max\{b \in \mathbb{Z} \,:\, \exists q \in I(b) \;\text{ such that }\; Q_{iq} \ne 0\}.$$

Put $b_{\min} := \lfloor \log_2 1/d \rfloor$, $b_{\max} := \max_{i\le n} b(i)$ and define for all $b_{\min} \le b \le b_{\max}$,

$$\mathcal{W}_b^1 := \bigcup_{s=b_{\min}}^{b} \mathcal{W}_s = \bigcup_{I_{\min}(b)} \mathcal{L}^{(q)}, \quad I_{\min}(b) = \{q \,:\, 2^{b_{\min}} \le \widetilde{w}_q < 2^{b+1}\}$$

$$\mathcal{W}_b^2 := \bigcup_{s=b+1}^{b_{\max}} \mathcal{W}_s = \bigcup_{I^c_{\min}(b)} \mathcal{L}^{(q)}, \quad I^c_{\min}(b) = \{q \,:\, 2^{b+1} \le \widetilde{w}_q < 2^{b_{\max}+1}\}.$$

Note that for every $b_{\min} \le b \le b_{\max}$ one has

$$I_{\min}(b) = \bigcup_{s=b_{\min}}^{b} I(b), \quad I^c_{\min}(b) = \bigcup_{s=b+1}^{b_{\max}} I(b), \qquad \text{and} \qquad \mathcal{W}_b^1 \cup \mathcal{W}_b^2 = [n].$$

The following quantities will play an important role below,

$$\eta_i := \frac{1}{d} \sum_{b=b_{\min}}^{b_{\max}} \min\Big(\sum_{q\in I_{\min}(b)} Q_{iq}, \sum_{q\in I^c_{\min}(b)} Q_{iq} \Big) \qquad \text{and} \qquad \eta = \sum_{i=1}^n \eta_i.$$

We start with a useful bound on cardinalities of ℓ-parts $\mathcal{L}^{(q)}$ inside $I(b)$ for a given b, in which we also use that for all positive integers N, N_i, $i \le \ell$, with $N = N_1 + \ldots + N_\ell$ one has

$$N!/\prod_{i=1}^{\ell} N_i! \le \prod_{i=1}^{\ell} (eN/N_i)^{N_i} \le (eN)^N \prod_{i=1}^{\ell} 1/(N_i)^{N_i} \tag{34}$$

(this follows by the standard inequality $\binom{N}{\ell} \le (eN/\ell)^\ell$).

Lemma 7.2. *Let $b_{\min} \le b \le b_{\max}$. Then the multiset $\{|\mathcal{L}^{(q)}| : q \in I(b)\}$, when arranged in the non-increasing order, can be majorized by the geometric sequence $\big(C|\mathcal{W}_b|\exp(-s/C)\big)_{s\ge 0}$ for a sufficiently large absolute constant $C > 0$. In particular,*

$$\prod_{I(b)} |\mathcal{L}^{(q)}|! \ge |\mathcal{W}_b|! \exp(-C'|\mathcal{W}_b|),$$

where $C' > 0$ is another absolute constant.

Proof. We apply (18), which roughly speaking says that an ℓ-part obtained at step j satisfies $|\mathcal{L}_q| \approx 2^j h_q$. Note also that at most two ℓ-parts can be obtained on a given step j.

Split the set $\{\mathcal{L}^{(q)} : q \in I(b)\}$ into four subsets U_1, U_2, U_3, U_4 determined by whether an ℓ-part is spread or regular and whether its cardinality is greater than $d^{-1/3}n$ or not. For concreteness, assume that U_1, U_2 contain regular ℓ-parts, with larger ℓ-parts in U_1, and U_3, U_4 include spread ℓ-parts, with larger ones in U_3. Within the set U_2, the heights of the respective ℓ-parts are equivalent up to multiple 2. Therefore their cardinalities, when arranged in non-increasing order, are majorized by an appropriate geometric sequence. The same argument works for ℓ-parts in U_4. For the ℓ-parts in U_1, the quantities $h_q|\mathcal{L}_q|$ are equivalent to each other, say to a number a. This means that for an ℓ-part obtained at the step j we have $h_q^2 \approx a2^{-j}$. This in turn implies $|\mathcal{L}_q| \approx 2^{j/2}\sqrt{a}$. Thus $|\mathcal{L}_q|$ (after a rearrangement) are geometrically decreasing. Finally, for set U_3, the quantities $h_q\sqrt{|\mathcal{L}_q|}$ are stable and a similar argument works. Combining the four decreasing sequences into one, we obtain a sequence that can be also majorized by a geometric series (with worse constants).

To prove the "in particular" part, let N_s, $s \ge 0$, corresponds to cardinalities of $\mathcal{L}^{(q)}$, $q \in I(b)$. Then $N = \sum_s N_s = |\mathcal{W}_b|$ and, by the first part, $N_s \le CN\exp(-s/C)$. Therefore, by (34),

$$\ln\Big(|\mathcal{W}_b|!/\prod_{I(b)}|\mathcal{L}^{(q)}|!\Big) \le \sum_{s\ge 0} N_s \ln(eN/N_s) \le N\sum_{s\ge 0}(s+1)e^{-s/C} \le C'N,$$

where $C' > 0$ is an absolute constant. This completes the proof. □

In the next lemma we relate SB_i and TE_i estimators using the parameter η introduced above.

Lemma 7.3. *Let $y \in \mathcal{A}_k$, $(\mathcal{L}^{(q)})_{q=1}^m$ be its ℓ-decomposition, and Q be a y-admissible matrix in $\mathcal{R}^{ST}_{n,m,d}$. Then*

$$\prod_{i=1}^n \mathrm{SB}_i \le C_{7.3}^n\, 2^{-\eta} \prod_{i=1}^n \mathrm{TE}_i,$$

where $C_{7.3} \ge 1$ is a universal constant.

Proof. For every $i \le n$, set

$$\widetilde{\mathrm{SB}}_i = \widetilde{\mathrm{SB}}_i(y,k,Q) := \min\big\{\widetilde{w}_q^{-1} : q \le m \text{ and } Q_{iq} \ne 0\big\}.$$

Since $Q_{iq} \ge 1$ in the above minimum, by the definition of the weights, we get for all $i \le n$,

$$\widetilde{\mathrm{SB}}_i \ge \mathrm{SB}_i/d. \tag{35}$$

We first show that

$$\prod_{i=1}^n \widetilde{\mathrm{SB}}_i \ge \exp(-Cn)\prod_{i=1}^n \mathrm{SB}_i, \tag{36}$$

where C is a positive universal constant. Let $\mathcal{L}^{(q)}$ be one of ℓ-parts from the ℓ-decomposition (no matter whether spread or regular), having cardinality at least $d^{-1/3}n$. Then, by the definition of $\mathcal{R}^{ST}_{n,m,d}$ (part 1), d-regularity of matrices in $\mathcal{M}_{n,d}$, and (21), there are at most $n/\sqrt{d}$ indices $i \le n$ such that $Q_{iq} < c_{3.3}d|\mathcal{L}^{(q)}|/n$. For all other i's, we have $Q_{iq} \ge c_{3.3}d|\mathcal{L}^{(q)}|/n$, implying together with the definitions of w_{iq} and $\widetilde{w}_q$ that $w_{iq} \geq c_{3.3}\, \widetilde{w}_q$. On the other hand, if the cardinality of $\mathcal{L}^{(q)}$ is less than $d^{-1/3}n$ and $Q_{iq} \neq 0$ (hence greater or equal to 1) then necessarily $w_{iq} \geq \widetilde{w}_q$. Denote

$$I := \{i \in [n] :\ w_{iq} < c_{3.3}\, \widetilde{w}_q \quad \text{for some} \quad q \leq m \quad \text{with} \quad Q_{iq} \neq 0\}.$$

By definitions of SB_i and $\widetilde{\mathrm{SB}}_i$, we have $\widetilde{\mathrm{SB}}_i \geq c_{3.3}\, \mathrm{SB}_i$ for all $i \in I^c$. Since there are at most $d^{1/3}$ ℓ-parts of cardinality at least $d^{-1/3}n$ each, then from the above we also have that $|I| \leq d^{1/3}d^{-1/2}n = d^{-1/6}n$. Therefore, using (35) for $i \in I$, we obtain

$$\prod_{i=1}^{n} \widetilde{\mathrm{SB}}_i \geq (c_{3.3})^{|I^c|}\, d^{-|I|} \prod_{i=1}^{n} \mathrm{SB}_i \geq c_{3.3}^{n} d^{-d^{-1/6}n} \prod_{i=1}^{n} \mathrm{SB}_i,$$

which leads to (36).

To complete the proof, it is sufficient to show that for every $i \leq n$

$$\widetilde{\mathrm{SB}}_i \leq 2^{-\eta_i + 1}\, \mathrm{TE}_i.$$

Fix $i \leq n$. By the definition of $\widetilde{\mathrm{SB}}_i$ and $b(i)$ we have $\widetilde{\mathrm{SB}}_i \leq 2^{-b(i)}$. Since

$$\sum_{b=b_{\min}}^{b(i)} \sum_{q \in I(b)} Q_{iq}/d = \sum_{q=1}^{m} Q_{iq}/d = 1, \tag{37}$$

the definition of TE_i implies

$$\begin{aligned}\mathrm{TE}_i &= \prod_{b=b_{\min}}^{b(i)} \prod_{q \in I(b)} (\widetilde{w}_q)^{-Q_{iq}/d} > \prod_{b=b_{\min}}^{b(i)} \prod_{q \in I(b)} 2^{-(b+1)Q_{iq}/d} \\ &= \frac{1}{2} \exp\Big(-\ln 2 \sum_{b=b_{\min}}^{b(i)} b \sum_{q \in I(b)} \frac{Q_{iq}}{d}\Big).\end{aligned}$$

Using (37) again and applying the simple identity

$$\sum_{j=j_0}^{j_1} (j_1 - j) a_j = \sum_{j=j_0}^{j_1 - 1} \sum_{k=j_0}^{j} a_k,$$

valid for any integers $j_0 < j_1$ and any numbers $a_{j_0},\ldots, a_{j_1}$, we get

$$\begin{aligned} db(i) - \sum_{b=b_{\min}}^{b(i)} b \sum_{q \in I(b)} Q_{iq} &= \sum_{b=b_{\min}}^{b(i)} (b(i) - b) \sum_{q \in I(b)} Q_{iq} = \sum_{b=b_{\min}}^{b(i)-1} \sum_{a=b_{\min}}^{b} \sum_{q \in I(a)} Q_{iq} \\ &= \sum_{b=b_{\min}}^{b(i)-1} \sum_{q \in I_{\min}(b)} Q_{iq} \geq \sum_{b=b_{\min}}^{b_{\max}} \min\Big(\sum_{q \in I_{\min}(b)} Q_{iq}, \sum_{q \in I^c_{\min}(b)} Q_{iq}\Big) = d\eta_i. \end{aligned}$$

where we used that if $b > b(i)$ then for every $q \in I(b)$ one has $Q_{iq} = 0$. Therefore,

$$\widetilde{\mathrm{SB}}_i \leq 2^{-b(i)} \leq 2\, \mathrm{TE}_i \exp\Big(-\ln 2 \,\Big(b(i) - \sum_{b=b_{\min}}^{b(i)} b \sum_{q \in I(b)} Q_{iq}/d\Big)\Big) \leq 2^{-\eta_i + 1}\mathrm{TE}_i,$$

This completes the proof. $\square$

In the next two lemmas, we estimate η in the case when Q belongs to $\mathcal{R}^{ST}_{n,m,d}$.

Lemma 7.4. *Let $y \in \mathcal{A}_k$, $(\mathcal{L}^{(q)})_{q=1}^m$ be its ℓ-decomposition and $\mathcal{W}_b$, $b \in \mathbb{Z}$, be its corresponding w-sets. Further, let $Q = (Q_{iq})$ be a y-admissible matrix in $\mathcal{R}_{n,m,d}^{ST}$. Then*

$$\eta \geq c_{7.4} \sum_{b=b_{\min}}^{b_{\max}} \min\left(|\mathcal{W}_b^1|,\, |\mathcal{W}_b^2|\right),$$

where $c_{7.4} > 0$ is a universal constant.

Proof. By the definition of η we have

$$\eta = \frac{1}{d} \sum_{b=b_{\min}}^{b_{\max}} \sum_{i=1}^{n} \min\Big(\sum_{q \in I_{\min}(b)} Q_{iq},\ \sum_{q \in I_{\min}^c(b)} Q_{iq} \Big).$$

To prove the lemma we prove the corresponding inequality for each summand in the first sum. To this end, for every (fixed) $b_{\min} \leq b < b_{\max}$ we apply the definition of $\mathcal{R}_{n,m,d}^{ST}$ (more precisely, part **2** of the definition of $\mathcal{R}_{n,m,d}^{ST_0}$), with

$$J = J(b) := I_{\min}(b) \quad \text{and} \quad \kappa = \kappa(b) := d|\mathcal{W}_b^1|.$$

Note that by (21) and d-regularity we have

$$\kappa = d|\mathcal{W}_b^1| = \sum_{q \in J} \|C_q(Q)\|_1 \quad \text{and} \quad dn - \kappa = d|\mathcal{W}_b^2|.$$

We distinguish two cases.
Case 1. $\min\big(|\mathcal{W}_b^1|, |\mathcal{W}_b^2|\big) \geq n/d$. In this case $\min\big(\kappa, dn-\kappa, n\big) = n$. Thus, the definition of $\mathcal{R}_{n,m,d}^{ST}$ yields that the cardinality of the set

$$I := \Big\{ i \leq n : \sum_{q \in I_{\min}(b)} Q_{iq} \geq c_{3.5} \frac{d|\mathcal{W}_b^1|}{n} \ \text{ and } \sum_{q \in I_{\min}^c(b)} Q_{iq} \geq c_{3.5} \frac{d|\mathcal{W}_b^2|}{n} \Big\}$$

is at least $c_{3.5} n$. Therefore,

$$\begin{aligned} \sum_{i=1}^{n} \min\Big(\sum_{q \in I_{\min}(b)} Q_{iq},\ \sum_{q \in I_{\min}^c(b)} Q_{iq} \Big) &\geq \sum_{i \in I} \min\Big(\sum_{q \in I_{\min}(b)} Q_{iq},\ \sum_{q \in I_{\min}^c(b)} Q_{iq} \Big) \\ &\geq c_{3.5}^2 d \min\big(|\mathcal{W}_b^1|, |\mathcal{W}_b^2|\big). \end{aligned}$$

Case 2. $1 \leq \min\big(|\mathcal{W}_b^1|, |\mathcal{W}_b^2|\big) < n/d$. In this case $\kappa/n < 1$ or $(dn-\kappa)/n < 1$. Using that Q_{iq} are non-negative integers, that $c_{3.5} < 1$, and the definition of $\mathcal{R}_{n,m,d}^{ST}$, we observe

$$\begin{aligned} &\Big| \Big\{ i \leq n : \sum_{q \in I_{\min}(b)} Q_{iq} \geq 1 \ \text{ and } \sum_{q \in I_{\min}^c(b)} Q_{iq} \geq 1 \Big\} \Big| \\ &\quad \geq \Big| \Big\{ i \leq n : \sum_{q \in I_{\min}(b)} Q_{iq} \geq c_{3.5} \frac{\kappa}{n} \ \text{ and } \sum_{q \in I_{\min}^c(b)} Q_{iq} \geq c_{3.5} \frac{dn-\kappa}{n} \Big\} \Big| \\ &\quad \geq c_{3.5}\, d \min\big(|\mathcal{W}_b^1|, |\mathcal{W}_b^2|\big). \end{aligned}$$

Therefore,

$$\sum_{i=1}^{n} \min\Big(\sum_{q \in I_{\min}(b)} Q_{iq},\ \sum_{q \in I_{\min}^c(b)} Q_{iq} \Big) \geq c_{3.5} d \min\big(|\mathcal{W}_b^1|, |\mathcal{W}_b^2|\big).$$

Since the case $\min\big(|\mathcal{W}_b^1|, |\mathcal{W}_b^2|\big) = 0$ is trivial, this completes the proof. $\square$

Lemma 7.5. *Let $y \in \mathcal{A}_k$, $(\mathcal{L}^{(q)})_{q=1}^m$ and $\mathcal{W}_b$, $b \in \mathbb{Z}$, be as above, and $Q = (Q_{iq})$ be a y-admissible matrix from $\mathcal{R}_{n,m,d}^{ST}$. Then*

$$n! \prod_{q \le m} \frac{1}{|\mathcal{L}^{(q)}|!} \le C^n \, 2^{\eta/2},$$

where C is a positive universal constant.

Proof. Denote $c := (c_{7.4} \ln 2)/2$. By Lemma 7.4, it is enough to show that

$$n! \prod_{q=1}^{m} \frac{1}{|\mathcal{L}^{(q)}|!} \exp\Big(- c \sum_{b=b_{\min}}^{b_{\max}} \min\big(|\mathcal{W}_b^1|, \, |\mathcal{W}_b^2|\big)\Big) \le C^n$$

for a universal constant $C > 0$. Denote

$$I := \Big\{ b \in \mathbb{Z} \, : \; \mathcal{W}_b \ne \emptyset \quad \text{and} \quad |\mathcal{W}_b| \ln(n/|\mathcal{W}_b|) \ge c \min\big(|\mathcal{W}_b^1|, \, |\mathcal{W}_b^2|\big) \Big\}$$

and for every integer $p \ge 0$,

$$I_p := \big\{ b \in \mathbb{Z} : \, |\mathcal{W}_b| \in (n2^{-p-1}, n2^{-p}] \big\} \cap I.$$

Note that if $b \in I_p$ then $\min\big(|\mathcal{W}_b^1|, |\mathcal{W}_b^2|\big) \le n(p+1)2^{-p}/c$. Hence,

$$\begin{aligned}
|I_p| &\le \Big| \Big\{ b \in \mathbb{Z} : \, |\mathcal{W}_b| \in (n2^{-p-1}, n2^{-p}] \quad \text{and} \quad \min\big(|\mathcal{W}_b^1|, |\mathcal{W}_b^2|\big) \le n(p+1)2^{-p}/c \Big\} \Big| \\
&\le \Big| \Big\{ b \in \mathbb{Z} : \, |\mathcal{W}_b| \in (n2^{-p-1}, n2^{-p}] \quad \text{and} \quad |\mathcal{W}_b^1| \le n(p+1)2^{-p}/c \Big\} \Big| \\
&\quad + \Big| \Big\{ b \in \mathbb{Z} : \, |\mathcal{W}_b| \in (n2^{-p-1}, n2^{-p}] \quad \text{and} \quad |\mathcal{W}_b^2| \le n(p+1)2^{-p}/c \Big\} \Big|.
\end{aligned}$$

Denote the cardinalities of the sets in the last inequality by $\alpha = \alpha(p)$ and $\beta = \beta(p)$ correspondingly. Let $b_1 < \ldots < b_\alpha$ and $b_1' < \ldots < b_\beta'$ be the elements of those set. Then

$$\alpha \, n2^{-p-1} \le \Big| \bigcup_{i=1}^{\alpha} \mathcal{W}_{b_i} \Big| \le \Big| \mathcal{W}_{b_\alpha}^1 \Big| \le n(p+1)2^{-p}/c$$

and

$$\beta \, n2^{-p-1} \le \Big| \bigcup_{i=1}^{\beta} \mathcal{W}_{b_i'} \Big| \le \Big| \mathcal{W}_{b_1'} \Big| + \Big| \mathcal{W}_{b_1'}^2 \Big| \le n2^{-p} + n(p+1)2^{-p}/c.$$

This implies that

$$|I_p| \le \alpha + \beta \le 6(p+1)/c.$$

Therefore using that $\sum_b |\mathcal{W}_b| = n$ and (34) with $N = n$ and $N_b = |\mathcal{W}_b|$, we obtain

$$\begin{aligned}
&n! \Big(\prod_{b=b_{\min}}^{b_{\max}} \frac{1}{|\mathcal{W}_b|!} \Big) \exp\Big(-c \sum_{b=b_{\min}}^{b_{\max}} \min\big(|\mathcal{W}_b^1|, |\mathcal{W}_b^2|\big) \Big) \\
&\le e^n \prod_{b : \mathcal{W}_b \ne \emptyset} \Big(\Big(\frac{n}{|\mathcal{W}_b|} \Big)^{|\mathcal{W}_b|} \exp\Big(-c \min\big(|\mathcal{W}_b^1|, |\mathcal{W}_b^2|\big) \Big) \Big) \\
&\le e^n \prod_{b \in I} \Big(\frac{n}{|\mathcal{W}_b|} \Big)^{|\mathcal{W}_b|} = e^n \prod_{p=0}^{\log_2 n} \prod_{b \in I_p} \Big(\frac{n}{|\mathcal{W}_b|} \Big)^{|\mathcal{W}_b|} \\
&\le e^n \prod_{p=0}^{\log_2 n} 2^{(p+1)n2^{-p} \, |I_p|} \le e^n \prod_{p=0}^{\log_2 n} 2^{6(p+1)^2 n 2^{-p}/c} \le \exp(\widetilde{C} n),
\end{aligned}$$

where $\widetilde{C} > 0$ is a sufficiently large absolute constant. Applying Lemma 7.2 and using $\sum_b |\mathcal{W}_b| = n$ again, we get

$$\Big(n! \prod_{q \le m} \frac{1}{|\mathcal{L}^{(q)}|!}\Big) \exp\Big(-c \sum_{b=b_{\min}}^{b_{\max}} \min\big(|\mathcal{W}_b^1|, |\mathcal{W}_b^2|\big)\Big) \le \exp\big((\widetilde{C} + C'')n\big),$$

which completes the proof. □

7.2 Completion of the proof

In the proof of Theorem 1.1 we use the results established in the previous sections. Recall our decomposition of $\mathbb{C}^n$ into three types of vectors: almost constant vectors, steep vectors, and gradual vectors. We treat each of these types separately. The former two types are treated in Section 4, where a lower bound on $\|(M - z\mathrm{Id})^K x\|_2$ is given. This leaves us with the case of gradual vectors which we approximate by k-vectors. First, one needs to check that it is sufficient to establish a lower bound for the action of M^K on the k-approximation of gradual vectors in order to deduce a similar bound for all such vectors. The next proposition provides such an approximation argument.

Proposition 7.6. *Let $d \ge 1$ be large enough, $n \ge d^3$, and $1 \le L \le n/d^3$. Let $K \subset [n]$ be such that $|K^c| \le L$ and z be such that $|z| \le r\sqrt{d}$ for some $r \ge 1$. Let X be a subset of the set of normalized gradual vectors $X \subset \mathcal{S}$ and $\mathcal{A}_k(X)$ be the set of k-approximations of vectors from X. Then*

$$\begin{aligned}
&\mathbb{P}\Big\{M \in \mathcal{M}_{n,d} : \exists x \in X,\ \|(M - z\mathrm{Id})^K x\|_2 \le \frac{L^3}{kn^{5.5}\,d}\|x\|_2\Big\} \\
&\le 2d^2 \sum_{y \in \mathcal{A}_k(X)} \sup_{V \in \mathbb{C}^{|K|}} \mathbb{P}\Big\{M \in \mathcal{M}_{n,d} :\ \|M^K y + V\|_2 \le C_{7.6}\, r \frac{\sqrt{dn}}{k}\Big\} + n^{-100},
\end{aligned}$$

where $C_{7.6} \ge 1$ is a universal constant.

Proof. By definitions we have $x^*_{n_1} \le d^3$ for $x \in \mathcal{S}$. Therefore, Lemma 4.3 implies

$$\begin{aligned}
&\mathbb{P}\Big\{M \in \mathcal{M}_{n,d} : \exists\, x \in X,\ \|(M - z\mathrm{Id})^K x\|_2 \le \frac{L^3}{kn^{5.5}\,d}\,\|x\|_2\Big\} \\
&\qquad\qquad \le \mathbb{P}\Big\{M \in \mathcal{M}_{n,d} :\ \exists\, x \in X,\ \|(M - z\mathrm{Id})^K x\|_2 \le \frac{\sqrt{dn}}{k}\Big\}.
\end{aligned}$$

Suppose that $M \in \mathcal{E}_{2.8}$. Let $x \in X$ and let $y \in \mathcal{A}_k(X)$ be its k-approximation. Define $a = (a_1, a_2)$, $\widetilde{a} = (\widetilde{a}_1, \widetilde{a}_2) \in \mathbb{C}$ by

$$a = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i) \qquad \text{and} \qquad \widetilde{a}_1 = \mathrm{Re}\,\widetilde{a} = \frac{\lceil kda_1 \rceil}{kd}, \quad \widetilde{a}_2 = \mathrm{Im}\,\widetilde{a} = \frac{\lceil kda_2 \rceil}{kd}.$$

Below we use the same notation $\mathbf{1}$ for the vectors in $\mathbb{C}^\ell$ for every $\ell \ge 1$. Note that since $M \in \mathcal{E}_{2.8}$ we have that $\|Mw\|_2 \le C_{2.8}\sqrt{d}\|v\|_2$ for every $v \in C^n$ which is orthogonal to $\mathbf{1}$. Denote

$$V(y, \widetilde{a}) := -zy^K + \widetilde{a}(d - z)\mathbf{1} \in \mathbb{C}^{|K|}.$$

Since y is the k-approximation of x, then for $i = 1, 2$ we have $0 \le ka_i \le 1$ and thus $\widetilde{a}_i \in \{0, \ldots, d\}/kd$. Using the triangle inequality, d-regularity, $|a - \widetilde{a}| \le \sqrt{2}/(kd)$, and that $x - y - a\,\mathbf{1}$ is orthogonal to $\mathbf{1}$, we observe

$$\begin{aligned}
&\|M^K y + V(y, \widetilde{a})\|_2 \\
&\quad \le \|(M - z\mathrm{Id})^K x\|_2 + \|(M - z\mathrm{Id})^K (x - y - a\,\mathbf{1})\|_2 + |a - \widetilde{a}|\ \|(M - z\mathrm{Id})^K \mathbf{1}\|_2 \\
&\quad \le \|(M - z\mathrm{Id})^K x\|_2 + C_{2.8}\Big(\sqrt{d} + |z|\Big)\|x - y - a\,\mathbf{1}\|_2 + \frac{\sqrt{2}|d - z|}{kd}\sqrt{|K|}.
\end{aligned}$$

Using that
$$\|x-y-a\,\mathbf{1}\|_2 \le \|x-y\|_2 + |a|\sqrt{n} \le 2\sqrt{2n}/k$$
together with $|z|\le r\sqrt{d}$ we get
$$\|M^K y + V(y,\widetilde{a})\|_2 \le \|(M-z\mathrm{Id})^K x\|_2 + (2C_{2.8}+1)\,(r+1)\sqrt{2dn}/k.$$
Theorem 2.8 and the union bound imply
$$\begin{aligned}
&\mathbb{P}\Big\{M\in\mathcal{M}_{n,d}:\ \exists\, x\in X,\ \|(M-z\mathrm{Id})^K x\|_2\le \sqrt{dn}/k\Big\}\\
&\quad\le \mathbb{P}\Big\{M\in\mathcal{M}_{n,d}:\ \mathcal{E}_{2.8}\quad\text{and}\quad \exists\, x\in X,\ \|(M-z\mathrm{Id})^K x\|_2\le \sqrt{dn}/k\Big\}+n^{-100}\\
&\quad\le \mathbb{P}\Big\{M\in\mathcal{M}_{n,d}:\ \mathcal{E}_{2.8}\quad\text{and}\quad \exists\, y\in\mathcal{A}_k(X), \exists\,\widetilde{a}\in\{0,\dots,d\}^2/(kd),\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad \|M^K y+V(y,\widetilde{a})\|_2\le C\,r\sqrt{dn}/k\Big\}+n^{-100}\\
&\quad\le 2d^2\sum_{y\in\mathcal{A}_k(X)}\ \sup_{V\in\mathbb{C}^{|K|}}\mathbb{P}\Big\{M\in\mathcal{M}_{n,d}:\ \|M^K y+V\|_2\le C\,r\sqrt{dn}/k\Big\}+n^{-100},
\end{aligned}$$
where $C>0$ is an appropriate large universal constant. □

Applying Theorem 5.6, we further decompose gradual vectors into two types, $\mathcal{K}_u$ and $\mathcal{P}_u$, depending on some properties satisfied by their k-approximation. Recall that the set of k-vectors is partitioned into equivalence classes and bounds on the cardinality of each class and on their total number are established in Lemmas 5.1 and 5.5. Therefore, in view of the previous proposition, we can concentrate our effort on bounding the probability that $\|M^K y+V\|_2$ is small for a fixed k-vector y satisfying the properties given in $\mathcal{K}_u$ or $\mathcal{P}_u$ and any vector $V\in\mathbb{C}^{|K|}$. Theorem 6.1 and Proposition 7.1 establish such bound in terms of the small ball estimators SB_i of the vector y. Therefore, it remains to estimate SB_i for these two types of vectors using all the tools developed in Section 7.1. We start with vectors in $\mathcal{K}_u$. Recall that for $x\in\mathcal{K}_u$, the total cardinality of the spread ℓ-parts in the ℓ-decomposition with respect to the d^u-approximation of x is at least $c_{\mathcal{K}} n_3$, where $c_{\mathcal{K}}\in(0,1)$ is an absolute constant.

Lemma 7.7. *Let $u\ge 2$ be an integer, y be the d^u-approximation with respect to a vector in $\mathcal{K}_u$, $(\mathcal{L}^{(q)})_{q=1}^m$ be its ℓ-decomposition, and $Q=(Q_{iq})$ be a y-admissible matrix in $\mathcal{R}^{ST}_{n,m,d}$. Then there exists a universal constant $c_{7.7}>0$ such that*
$$\prod_{i=1}^n \mathrm{SB}_i(y,k,Q)\le d^{-c_{7.7}n}(n!)^{-1}\prod_{q\le m}\frac{|\mathcal{L}^{(q)}|!}{h_q^{|\mathcal{L}^{(q)}|}}.$$

Proof. To prove the lemma, it is enough to show that
$$\beta := n!\prod_{q\le m}\frac{1}{|\mathcal{L}^{(q)}|!}\prod_{i=1}^n \mathrm{SB}_i \le d^{-c_{7.7}n}\prod_{q\le m} h_q^{-|\mathcal{L}^{(q)}|}. \tag{38}$$
Applying Lemmas 7.3 and 7.5 we get
$$\beta\le C^n 2^{-\eta/2}\prod_{i=1}^n \mathrm{TE}_i \le C^n\prod_{i=1}^n \mathrm{TE}_i,$$
where C is a positive absolute constant. Let $c:=c_{\mathcal{K}}a_3/16<1/3$, where a_3 comes from the definition of n_3 (recall $n_3=\lfloor a_3 n\rfloor$), and $I:=\{q\,:\,|\mathcal{L}^{(q)}|\ge d^{-c}n\}$. We consider two cases.

Case 1. $\big|\bigcup_{q\in I}\mathcal{L}^{(q)}\big|\ge n-c_{\mathcal{K}}n_3/4$. Denote by I_1 the set of all indices q corresponding to spread ℓ-parts of cardinality at least $d^{-c}n$ and by I_2 be the set of indices corresponding to regular ℓ-parts of cardinality at least $d^{-c}n$. Let I_3 be all the remaining indices, that is $I_3 = [m] \setminus (I_1 \cup I_2)$. Note that $\big|\bigcup_{q\in I_3} \mathcal{L}^{(q)}\big| \le c_{\mathcal{K}} n_3/4$. By (33) we have

$$\beta \le C^n \prod_{q\in I_1} \widetilde{w}_q^{-|\mathcal{L}^{(q)}|} \prod_{q\in I_2} \widetilde{w}_q^{-|\mathcal{L}^{(q)}|} \prod_{q\in I_3} \widetilde{w}_q^{-|\mathcal{L}^{(q)}|}.$$

By the definition of $\mathcal{K}_u$, the total cardinality of ℓ-parts with indices from I_1 is at least

$$c_{\mathcal{K}} n_3 - c_{\mathcal{K}} n_3/4 = 3c_{\mathcal{K}} n_3/4.$$

This together with the definition of the truncated weights for

$$|\mathcal{L}^{(q)}| \ge d^{-c}n \ge d^{-1/3}n,$$

implies that

$$\prod_{q\in I_1} (\widetilde{w}_q)^{-|\mathcal{L}^{(q)}|} \le \prod_{q\in I_1} \Big(\frac{n}{d|\mathcal{L}^{(q)}|}\Big)^{|\mathcal{L}^{(q)}|/2} h_q^{-|\mathcal{L}^{(q)}|} \le d^{-3c_{\mathcal{K}} n_3/8} \prod_{q\in I_1} \Big(\frac{n}{|\mathcal{L}^{(q)}|}\Big)^{|\mathcal{L}^{(q)}|} h_q^{-|\mathcal{L}^{(q)}|}.$$

Hence, using the definition of the truncated weights for $q \in I_2$ and the bounds (32) for $q \in I_3$, we get

$$\begin{aligned}\beta &\le C^n d^{-3c_{\mathcal{K}} n_3/8} \prod_{q\in I_1\cup I_2} (n/|\mathcal{L}^{(q)}|)^{|\mathcal{L}^{(q)}|} \prod_{q\in I_3} d^{|\mathcal{L}^{(q)}|} \prod_{q\le m} h_q^{-|\mathcal{L}^{(q)}|} \\ &\le C^n d^{-3c_{\mathcal{K}} n_3/8} d^{cn} d^{c_{\mathcal{K}} n_3/4} \prod_{q\le m} h_q^{-|\mathcal{L}^{(q)}|} \le d^{-c_{\mathcal{K}} n_3/16} \prod_{q\le m} h_q^{-|\mathcal{L}^{(q)}|},\end{aligned}$$

which leads to (38).

Case 2. $\big|\bigcup_{q\in I} \mathcal{L}^{(q)}\big| < n - c_{\mathcal{K}} n_3/4$. In this case $\big|\bigcup_{q\in I^c} \mathcal{L}^{(q)}\big| \ge c_{\mathcal{K}} n_3/4$. Using (33), we have

$$\beta \le C^n d^n 2^{-\eta/2} \prod_{q\le m} h_q^{-|\mathcal{L}^{(q)}|}.$$

Denote

$$J(b) := I^c \cap I(b) = \{q \,:\, |\mathcal{L}^{(q)}| < d^{-c}n \ \text{ and } \ 2^b \le \widetilde{w}_q < 2^{b+1}\}.$$

Arguing as in the proof of Lemma 7.2, we have

$$\Big|\bigcup_{q\in J(b)} \mathcal{L}^{(q)}\Big| \le C' d^{-c} n,$$

for a universal constant $C' > 0$. Define two integer numbers b_1 and b_2 by

$$b_1 := \min\Big\{b \in \mathbb{Z} : |\mathcal{W}_b^1| \ge \frac{c_{\mathcal{K}} n_3}{16}\Big\} \ \text{ and } \ b_2 := \max\Big\{b \in \mathbb{Z} : |\mathcal{W}_b \cup \mathcal{W}_b^2| \ge \frac{c_{\mathcal{K}} n_3}{16}\Big\}.$$

Clearly, $b_2 \ge b_1$. Denoting $J := I^c \cap \bigcup_{b_1\le b\le b_2} I(b)$, we observe

$$\Big|\bigcup_{q\in J} \mathcal{L}^{(q)}\Big| \le (b_2 - b_1 + 1)\, C' d^{-c} n.$$

On the other hand, using the definition of b_1, b_2 together with the condition of this case, we have

$$\Big|\bigcup_{q\in J} \mathcal{L}^{(q)}\Big| \ge \big|\bigcup_{q\in I^c} \mathcal{L}^{(q)}\big| - \big|\mathcal{W}_{b_1-1}^1\big| - \big|\mathcal{W}_{b_2}^2\big| \ge c_{\mathcal{K}} n_3/4 - c_{\mathcal{K}} n_3/8 = c_{\mathcal{K}} n_3/8.$$

Thus $b_2 - b_1 + 1 \geq c_{\mathcal{K}} a_3 d^c / 8C'$. Now applying Lemma 7.4, we get

$$\eta \geq c_{7.4} \sum_{b=b_1}^{b_2-1} \min\big(|\mathcal{W}_b^1|, |\mathcal{W}_b^2|\big) \geq c_{7.4}(b_2 - b_1) c_{\mathcal{K}} n_3/16 \geq c'\, d^c n,$$

where $c' > 0$ is an absolute constant. Since d is large enough, we obtain that $C^n d^n 2^{-\eta/2} \leq d^{-n}$, which completes the proof. □

We turn now to a particular class of vectors in $\mathcal{P}_v$ for some $v \geq 5$. Recall that for $x \in \mathcal{P}_u$, the total cardinality of spread and regular ℓ-parts in the ℓ-decomposition with respect to the d^u-approximation of x with heights not smaller than $c_{\mathcal{P}} 2^{c_{\mathcal{P}}(u-4)a_3} a_3$ is at least $c_{\mathcal{P}} n_3$, where $c_{\mathcal{P}} < 1$ is a universal constant. For every integer $v \geq 5$ and every positive numbers δ, ρ we define

$$\mathcal{P}_{v,\rho,\delta} := \big\{x \in \mathcal{P}_v : \exists \lambda \in \mathbb{C} \text{ such that } |\{i \leq n : |x_i - \lambda| \leq \rho\}| \geq \delta n\big\}.$$

We start with a useful property of vectors in this set.

Lemma 7.8. *Let $v \geq 5$, $\rho \leq d^{-v}$, and $\delta \in (0,1)$. Let $x \in \mathcal{P}_{v,\rho,\delta}$ and y be its d^v-approximation. Then in the ℓ-decomposition of y, there exists a w-set $\mathcal{W}_b$ of order $b \leq \log_2(72\sqrt{d}/\delta)$ and of cardinality at least $\delta n/36$.*

Proof. Let $x \in \mathcal{P}_{v,\rho,\delta}$ and y be its d^v-approximation. Let $\lambda = \lambda(x) \in \mathbb{C}$ be such that

$$|\{i \leq n : |x_i - \lambda| \leq \rho\}| \geq \delta n.$$

Note that if $|x_i - \lambda| \leq \rho$, then since $d^v \rho < 1$ we have

$$\mathrm{Re}(d^v y_i) \in \{\lfloor \mathrm{Re}(d^v \lambda)\rfloor - 1, \lfloor \mathrm{Re}(d^v \lambda)\rfloor, \lfloor \mathrm{Re}(d^v \lambda)\rfloor + 1\}$$

and

$$\mathrm{Im}(d^v y_i) \in \{\lfloor \mathrm{Im}(d^v \lambda)\rceil - 1, \lfloor \mathrm{Im}(d^v \lambda)\rfloor, \lfloor \mathrm{Im}(d^v \lambda)\rfloor + 1\},$$

which means that $d^v y_i$ can take at most 9 possible values. This implies the existence of a set $I \subset [n]$ of size at least $\delta n/9$ such that $y_i = y_j$ for all $i, j \in I$. Let $a = \lfloor \log_2(1 + \delta n/9)\rfloor - 1$ so that

$$\sum_{i=0}^{a} 2^i \leq |I|.$$

From the construction of the ℓ-decomposition of y, at the step $j = a$ the level set $L(j, y(I))$ is of size at least $2^a \geq \delta n/36$. This implies the existence of an ℓ-part of size at least $\delta n/36$ and, by (18), of height at most $n/2^{a-1} \leq 72/\delta$. Since for every q, $\widetilde{w}_q \leq h_q\sqrt{d}$, there exists a w-set of order at most $\log_2(72\sqrt{d}/\delta)$ and of cardinality at least $\delta n/36$. □

Next we estimate the product of SB_i for approximations of vectors from $\mathcal{P}_{v,\rho,\delta}$.

Lemma 7.9. *Let $v \geq 5$ be an integer, $0 \leq \rho \leq d^{-v}$ and $0 \leq \delta \leq 36 c_{\mathcal{P}} a_3$ be such that*

$$c_{\mathcal{P}} a_3 (v - 4) \geq 2\log_2 d + 2\log_2(72\sqrt{d}/\delta) + 2 - \log_2(c_{\mathcal{P}} a_3).$$

Further, let y be the d^v-approximation of a vector in $\mathcal{P}_{v,\rho,\delta}$ and $(\mathcal{L}^{(q)})_{q=1}^m$ be its ℓ-decomposition. Finally, let $Q = (Q_{iq})$ be a y-admissible matrix in $\mathcal{R}_{n,m,d}^{ST}$. Then

$$\prod_{i=1}^{n} \mathrm{SB}_i(y, k, Q) \leq (Cd)^n 2^{-c\delta n v} (n!)^{-1} \prod_{q \leq m} \frac{|\mathcal{L}^{(q)}|!}{h_q^{|\mathcal{L}^{(q)}|}},$$

where $C > c > 0$ are universal constants.

Proof. Applying Lemmas 7.3 and 7.5 together with (33), we get

$$\prod_{i=1}^{n} \mathrm{SB}_i(y,k,Q) \le (C\,d)^n 2^{-\eta/2}(n!)^{-1} \prod_{q\le m} \frac{|\mathcal{L}^{(q)}|!}{h_q^{|\mathcal{L}^{(q)}|}},$$

for a positive absolute constant C. Let

$$b_1 := \log_2(72\sqrt{d}/\delta) \quad \text{and} \quad b_2 := \log_2\left(c_{\mathcal{P}} a_3 2^{(v-4)c_{\mathcal{P}} a_3}\right) - \log_2 d.$$

By the assumptions of the lemma, $b_2 - b_1 \ge c_{\mathcal{P}} a_3 (v-4)/2$.

By Lemma 7.8, there exists a w-set of order at most b_1 and of cardinality at least $\delta n/36$. On the other hand, using the definition of $\mathcal{P}_v$ and (32), the total cardinality of w-sets of order at least b_2, is at least $c_{\mathcal{P}} a_3 n$. Therefore, for every integer b in the range $[b_1, b_2)$ we have

$$\min\big(|\mathcal{W}_b^1|, |\mathcal{W}_b^2|\big) \ge \delta n/36.$$

Now, we apply Lemma 7.4 to deduce that

$$\eta \ge c_{7.4} \sum_{b_1 \le b < b_2} \min\big(|\mathcal{W}_b^1|, |\mathcal{W}_b^2|\big) \ge c_{7.4}\delta n(b_2 - b_1)/36 \ge (c_{\mathcal{P}} c_{7.4} a_3/72)\, \delta\, n\, (v-4),$$

which implies the desired bound. □

We are now ready to state and complete the proof of a generalization of Theorem 1.1.

Theorem 7.10 (Structural theorem)**.** *There exist absolute positive constants c, c', and C such that the following holds. Let d, n be a large enough integers satisfying $d \le \exp(\sqrt{c' \ln n})$. Let $z \in \mathbb{C}$ be such that $|z| \le \sqrt{d}\ln d$. Let $1 \le L \le n/d^3$ and let r_0 be the smallest integer such that $p^{r_0} \ge 20L/d$, where $p = \lfloor (1/5)\sqrt{d/\ln d} \rfloor$. Let $K \subset [n]$ satisfy $|K^c| \le L$ and assume that*

$$\max(n^{-c}, e^{-cn/|K^c|}) \le \rho \le e^{-C\ln^2 d} \qquad \text{and} \qquad \delta = C\frac{\ln^2 d}{\ln(1/\rho)}.$$

Then with probability at least $1 - 1/n$ any non-zero vector $x \in \mathbb{C}^n$ with the property that

$$\|(M - z\mathrm{Id})^K x\|_2 \le L^3 n^{-6} \|x\|_2$$

satisfies one of the two conditions:

- *(Gradual with many levels) One has*

$$x_i^* \le \begin{cases} (n/i)^3\, x_{p^{r_0}}^* & \text{if } i \le p^{r_0}, \\ d(n/i)^3 x_{n_3}^* & \text{if } p^{r_0} \le i \le n_1, \\ d^3 x_{n_3}^* & \text{if } n_1 \le i \le n_3, \end{cases}$$

  and

$$\Big|\Big\{ i \le n : |x_i - \lambda| \le \rho x_{n_3}^* \Big\}\Big| \le \delta n \qquad \text{for all} \quad \lambda \in \mathbb{C}.$$

- *(Very steep) $x_i^* > 0.9(n/i)^3 x_{p^{r_0}}^*$ for some $i \le p^{r_0}$.*

Proof. Choose v from $\rho = d^{-v}$. Without loss of generality we assume that v is an integer. Then $\delta = Cv^{-1}\log_2 d$ and

$$C\ln d \le v \le c\,\min\big(\log_d n, \frac{n}{|K^c|\ln d}\big),$$

where C is a large enough absolute constant and $c \in (0,1/2)$ is a small enough absolute constant ($c = 1/6$ works). Note that the left hand side of this inequality is always smaller than the right hand side, provided that $d \le \exp(\sqrt{c' \ln n})$ with $c' = c/C$. Then, using that d is large enough we have

$$d^v \le \min\big(\sqrt{n}/(8d^{3/2}\sqrt{\ln d}), d^{-10}e^{n/5|K^c|}\big),$$

in particular, we may apply Lemma 5.1 and Proposition 7.1 with $k \le d^v$. Note also that the assumptions of Lemma 7.9 are satisfied as well.

Let $\Gamma_{\rho,\delta}$ be the set of non-zero vectors satisfying none of the two conditions in Theorem 7.10. For every $x \in \Gamma_{\rho,\delta}$, define

$$\mathcal{E}_x := \Big\{M \in \mathcal{M}_{n,d} : \ \|(M - z\mathrm{Id})^K x\|_2 \le \frac{L^3}{n^6}\|x\|_2\Big\}.$$

Since the event $\mathcal{E}_x$ is homogeneous in x, we may restrict $\Gamma_{\rho,\delta}$ to vectors satisfying $x^*_{n_3} = 1$ (if $x^*_{n_3} = 0$, we consider a slight perturbation of x).

Our goal is to show that $\mathbb{P}(\bigcup_{x\in\Gamma_{\rho,\delta}} \mathcal{E}_x) \le 1/n$. Recall that we decomposed $\mathbb{C}^n$ into the set of almost constant vectors, denoted by $\mathcal{B} := \mathcal{B}(\theta_0)$ with $\theta_0 = 10/d^3$, the set of steep vectors, denoted by $\mathcal{T}$ (note, an almost constant vector can be also steep), and the set of gradual vectors, denoted by $\mathcal{S} = \mathbb{C}^n \setminus (\mathcal{B} \cup \mathcal{T})$. If x doesn't satisfy the second condition of the theorem then, by (8), $x \notin \mathcal{T}_3^{\mathcal{K}}$, that is $\Gamma_{\rho,\delta} \cap \mathcal{T}_3^{\mathcal{K}} = \emptyset$. Note also that

$$\mathcal{S}^c \setminus \mathcal{T}_3^{\mathcal{K}} \subset \mathcal{B}_0 := (\mathcal{B} \setminus \mathcal{T}_3^{\mathcal{K}}) \cup \mathcal{T}_{\mathcal{K}}.$$

Therefore, applying Theorem 4.12 we obtain

$$\begin{aligned}\mathbb{P}\Big(\bigcup_{x\in\Gamma_{\rho,\delta}} \mathcal{E}_x\Big) &\le \mathbb{P}\Big(\bigcup_{x\in\mathcal{B}_0\cap\Gamma_{\rho,\delta}} \mathcal{E}_x\Big) + \mathbb{P}\Big(\bigcup_{x\in\mathcal{S}\cap\Gamma_{\rho,\delta}} \mathcal{E}_x\Big)\\ &\le \exp\big(-(\ln d)\,(\ln n)/20\big) + \mathbb{P}\Big(\bigcup_{x\in\mathcal{S}\cap\Gamma_{\rho,\delta}} \mathcal{E}_x\Big).\end{aligned}$$

We now show that $\Gamma_{\rho,\delta} \cap \mathcal{P}_v \subset \mathcal{P}_{v,\rho,\delta}$. Indeed, a vector in $\Gamma_{\rho,\delta}$ does not belong in particular to $\mathcal{T}_3$. Moreover, since $\mathcal{P}_v \subset \mathcal{S} \subset \mathcal{T}^c$, by Lemma 4.2 every $x \in \mathcal{P}_v$ satisfies $x^*_i \le d\big(n/i\big)^3 x^*_{n_3}$ for all $p^{r_0} \le i \le n_1$ and $x^*_i \le d^3 x^*_{n_3}$ for all $n_1 \le i \le n_3$. Therefore, if $x \in \Gamma_{\rho,\delta} \cap \mathcal{P}_v$ then it cannot satisfy the last condition in "gradual with many levels," which means that $x \in \mathcal{P}_{v,\rho,\delta}$

This together with Theorem 5.6 and the union bound gives

$$\mathbb{P}\Big(\bigcup_{x\in\mathcal{S}\cap\Gamma_{\rho,\delta}} \mathcal{E}_x\Big) \le \sum_{u=4}^{v} \mathbb{P}\Big(\bigcup_{x\in\mathcal{K}_u} \mathcal{E}_x\Big) + \mathbb{P}\Big(\bigcup_{x\in\mathcal{P}_{v,\rho,\delta}} \mathcal{E}_x\Big).$$

Applying Proposition 7.6 with $r = \ln d$ and $k = d^u$ (or $k = d^v$) and using $d^{1+u} \le d^{1+v} \le \sqrt{n}$, we get that there exists an absolute constant $C' > 0$ such that for any $5 \le u \le v$,

$$\mathbb{P}\Big(\bigcup_{x\in\mathcal{K}_u} \mathcal{E}_x\Big) \le 2d^2 \sum_{y\in\mathcal{A}_{d^u}(\mathcal{K}_u)} \sup_{V\in\mathbb{C}^{|K|}} \mathbb{P}\Big\{M \in \mathcal{M}_{n,d} : \|M^K y + V\|_2 \le C' \ln d\,\frac{\sqrt{dn}}{d^u}\Big\} + n^{-100},$$

and

$$\mathbb{P}\Big(\bigcup_{x\in\mathcal{P}_{v,\rho,\delta}} \mathcal{E}_x\Big) \le 2d^2 \sum_{y\in\mathcal{A}_{d^v}(\mathcal{P}_{v,\rho,\delta})} \sup_{\omega\in\mathbb{C}^{|K|}} \mathbb{P}\Big\{M \in \mathcal{M}_{n,d} : \|M^K y + V\|_2 \le C' \ln d\,\frac{\sqrt{dn}}{d^v}\Big\} + n^{-100}.$$

Take any $4 \le u \le v$ and fix for a moment $y \in \mathcal{A}_{d^u}(\mathcal{K}_u)$, the set of k-approximations of vectors in $\mathcal{K}_u$. Assume that its ℓ-decomposition consists of m sets $(\mathcal{L}^{(q)})_{q=1}^m$. Proposition 7.1 applied with $\gamma = C \ln d \sqrt{n/|K|}$ and Lemma 7.7 imply

$$\sup_{V \in \mathbb{C}^{|K|}} \mathbb{P}\big\{M \in \mathcal{M}_{n,d} : \|M^K y + V\|_2 \le C' (\ln d) \sqrt{dn}/d^u \,|\, \mathcal{E}_{3.5}\big\} \le \frac{e^{-2n}}{n!} \prod_{q \le m} \frac{|\mathcal{L}^{(q)}|!}{h_q^{|\mathcal{L}^{(q)}|}},$$

provided that d is large enough.

Let $\mathcal{C}$ be the equivalence class in $\mathcal{A}_{d^u}$ generated by y. By Lemma 5.5 we have

$$\sum_{\widetilde{y} \in \mathcal{C}} \sup_{V \in \mathbb{C}^{|K|}} \mathbb{P}\big\{M \in \mathcal{M}_{n,d} : \ \|M^K \widetilde{y} + V\|_2 \le C' (\ln d) \sqrt{dn}/d^u \,|\, \mathcal{E}_{3.5}\big\} \le \exp(-2n).$$

Finally, Lemma 5.1 implies

$$\sum_{y \in \mathcal{A}_{d^u}(\mathcal{K}_u)} \sup_{V \in \mathbb{C}^{|K|}} \mathbb{P}\big\{M \in \mathcal{M}_{n,d} : \ \|M^K y + V\|_2 \le C' (\ln d) \sqrt{dn}/d^u \,|\, \mathcal{E}_{3.5}\big\} \le e^{-n}.$$

Repeating the above argument for vectors in $\mathcal{A}_{d^v}(\mathcal{P}_{r,v,\rho})$ with Lemma 7.9 instead of Lemma 7.7 and using that $\delta = C(\log_2 d)/v$ with large enough C, we get

$$\sum_{y \in \mathcal{A}_{d^v}(\mathcal{P}_{v,\rho,\delta})} \sup_{V \in \mathbb{C}^{|K|}} \mathbb{P}\big\{M \in \mathcal{M}_{n,d} : \ \|M^K y + V\|_2 \le C (\ln d) \sqrt{dn}/d^v \,|\, \mathcal{E}_{3.5}\big\} \le e^{-n}.$$

Applying Proposition 3.5 to remove the conditioning from the two previous estimates and combining bounds, we obtain

$$\mathbb{P}\Big(\bigcup_{x \in \mathcal{S} \cap \Gamma_{\rho,\delta}} \mathcal{E}_x \Big) \le 2vd^2 e^{-n} + \frac{v}{n^{100}}.$$

The proof is finished by the choice of v. □

Note once more that the probability bound can be made $1 - n^{-\kappa}$ for any fixed $\kappa \ge 1$ at the expense of having worse constants.

Finally we prove Theorem 1.1 and Corollary 1.2.

Proof of Theorem 1.1. Set $L = \max(1, |K^c|)$ and

$$\rho = \max\Big(n^{-c}, \exp\Big(- \big(n/(1+|K^c|)\big)^{c \ln\ln d/\ln d}\Big)\Big)$$

for an appropriate positive constant c. Given $d^{-1/2} \le a \le 1$, let $q = \max(1, a|K^c|)$. Then $p^{r_0} \le q$ and $a|K^c| \le n_1$. Let x satisfy the dichotomy from Theorem 7.10. If x is very steep in the sense of Theorem 7.10 then, using $p^{r_0} \le q$, we get that x is very steep in the sense of Theorem 1.1. Assume now that x is not very steep in the sense of Theorem 1.1, i.e., assume that $x_i^* \le (n/i)^3 x_q^*$ for all $i \le q$. If in addition, x is gradual with many levels in the sense of Theorem 7.10 then it is not difficult to see that is gradual with many levels in the sense of Theorem 1.1 provided that $c' = a_3$. This proves the desired result. □

Proof of Corollary 1.2. Let n, d be as in Theorem 7.10. By Theorem 2.8, there is a universal constant $C > 0$ such that the event

$$\mathcal{E} := \Big\{M \in \mathcal{M}_{n,d} : \ \Big\|M - \frac{d}{n}\mathbf{1}\mathbf{1}^t\Big\| \ge C\sqrt{d}\Big\}$$

has probability less than n^{-2}. Denote by V the set of all vectors in $\mathbb{C}^n$ having sum of coordinates equal 0. Clearly, V is an invariant subspace of M. Let λ_i, $i \le n$, be eigenvalues of M arranged so that $|\lambda_1| \ge |\lambda_2| \ge \cdots \ge |\lambda_n|$. Since $\lambda_1(M) = d$ corresponds to the eigenvector $\mathbf{1}$, we observe that all eigenvectors corresponding to λ_i, $i \ge 2$, belong to V. This implies that, conditioned on $\mathcal{E}^c$, we have $|\lambda_i(M)| < C\sqrt{d}$, $i \ge 2$. Now, let $\mathcal{N}$ be a fixed $(1/(C\sqrt{d}n^6)$-net in the disk of radius $C\sqrt{d}$ of the complex plane (we assume the usual Euclidean metric on $\mathbb{C}$). Clearly, $\mathcal{N}$ can be chosen so that $|\mathcal{N}| \le n^{13}$. For any point $z' \in \mathcal{N}$, applying Theorem 7.10 with $K := [n]$, $L = 1$, $\rho := n^{-c}$, $\delta := \ln^2 d/\ln n$, we get that with probability at least $1 - n^{-15}$ every unit complex vector x satisfying $\|(M - z' \,\mathrm{Id})x\|_2 < n^{-6}$, is "gradual with many levels" (since $q = 1$, there are no "very steep" vectors). Now, observe that for any matrix $M \in \mathcal{M}_{n,d}$, any eigenvalue λ of M satisfying $|\lambda| \le C\sqrt{d}$ and a corresponding normalized eigenvector x, we necessarily have $\|(M - z' \,\mathrm{Id})x\|_2 < n^{-6}$ for some $z' = z'(\lambda) \in \mathcal{N}$. Combining this with the last remark, we get that the event

$$\mathcal{E}' := \Big\{ M \in \mathcal{M}_{n,d} : \text{any normalized eigenvector of } M \text{ with eigenvalue} \\ \text{in the disk of radius } C\sqrt{d} \text{ is ``gradual with many levels''} \Big\}$$

has probability at least $1 - n^{-2}$. Here, "gradual with many levels" means that the vector satisfies the conditions listed in the corollary. Finally note, that conditioned on $\mathcal{E}^c$ all eigenvectors except for $(1/\sqrt{n}, \dots, 1/\sqrt{n})$, have corresponding eigenvalues in the disk of radius $C\sqrt{d}$. Thus, all matrices in $\mathcal{E}^c \cap \mathcal{E}'$ satisfy the assertion of the statement. The result follows. $\square$

Acknowledgments

We are grateful for referees for careful reading and many suggestions, which help us to improve presentation. In particular, for showing us relatively short direct proofs of Propositions 2.5 and 6.5. A significant part of this work was completed while the last three named authors were in residence at the Mathematical Sciences Research Institute in Berkeley, California, supported by NSF grant DMS-1440140, and the first two named authors visited the institute. The hospitality of MSRI and of the organizers of the program on Geometric Functional Analysis and Applications is gratefully acknowledged. The research of the last named author was partially supported by grant ANR-16-CE40-0024-01.

References

[1] N. Alon and V. D. Milman, λ_1, isoperimetric inequalities for graphs, and superconcentrators, J. Combin. Theory Ser. B **38** (1985), no. 1, 73–88. MR0782626

[2] N. Anantharaman and E. Le Masson, Quantum ergodicity on large regular graphs, Duke Math. J. **164** (2015), no. 4, 723–765. MR3322309

[3] Á. Backhausz and B. Szegedy, On the almost eigenvectors of random regular graphs, arXiv:1607.04785.

[4] Z. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices*, second edition, Springer Series in Statistics, Springer, New York, 2010. MR2567175

[5] A. Basak, N. Cook and O. Zeitouni, Circular law for the sum of random permutation matrices, Electronic Journal of Probability, **23** (2018), Paper No. 33, 51 pp.

[6] A. Basak and M. Rudelson, Invertibility of sparse non-Hermitian matrices, Adv. Math. **310** (2017), 426–483. MR3620692

[7] A. Basak and M. Rudelson, The circular law for sparse non-Hermitian matrices, arXiv:1707.03675

[8] R. Bauerschmidt, J. Huang and H.-T. Yau, Local Kesten–McKay law for random regular graphs, arXiv:1609.09052

[9] R. Bauerschmidt, A. Knowles and H.-T. Yau, Local semicircle law for random regular graphs, Comm. Pure Appl. Math. **70** (2017), no. 10, 1898–1960. MR3688032

[10] C. Bordenave, A new proof of Friedman's second eigenvalue Theorem and its extension to random lifts, arXiv:1502.04482

[11] C. Bordenave and D. Chafaï, Around the circular law, Probab. Surv. **9** (2012), 1–89. MR2908617

[12] P. Bourgade, J. Huang and H.-T. Yau, Eigenvector statistics of sparse random matrices, Electron. J. Probab. **22** (2017), Paper No. 64, 38 pp. MR3690289

[13] A. Z. Broder, A. M. Frieze, S. Suen and E. Upfal, Optimal construction of edge-disjoint paths in random graphs, SIAM J. Comput. **28** (1999), no. 2, 541–573. MR1634360

[14] N. A. Cook, Discrepancy properties for random regular digraphs, Random Structures Algorithms **50** (2017), no. 1, 23–58. MR3583025

[15] N. A. Cook, On the singularity of adjacency matrices for random regular digraphs, Probab. Theory Related Fields **167** (2017), no. 1-2, 143–200. MR3602844

[16] N. Cook, The circular law for random regular digraphs, Ann. Inst. Henri Poincare Probab. Stat., to appear, arXiv:1703.05839

[17] N. Cook, L. Goldstein and T. Johnson, Size biased couplings and the spectral gap for random regular graphs, Ann. Probab. **46** (2018), no. 1, 72–125. MR3758727

[18] L. Geisinger, Convergence of the density of states and delocalization of eigenvectors on random regular graphs. J. Spectr. Theory **5** (2015), no. 4, 783827.

[19] J. Dodziuk, Difference equations, isoperimetric inequality and transience of certain random walks, Trans. Amer. Math. Soc. **284** (1984), no. 2, 787–794. MR0743744

[20] I. Dumitriu and S. Pal, Sparse regular random graphs: spectral density and eigenvectors, Ann. Probab. **40** (2012), no. 5, 2197–2235. MR3025715

[21] I. Dumitriu, T. Johnson, S. Pal, E. Paquette, Functional limit theorems for random regular graphs, Probab. Theory Related Fields 156 (2013), no. 3-4, 921–975. MR3078290

[22] C. G. Esseen, On the Kolmogorov-Rogozin inequality for the concentration function, Z. Wahrsch. Verw. Gebiete **5** (1966), 210–216. MR0205297

[23] C. G. Esseen, On the concentration function of a sum of independent random variables, Z. Wahrsch. Verw. Gebiete **9** (1968), 290–308. MR0231419

[24] J. Friedman, A proof of Alon's second eigenvalue conjecture and related problems, Mem. Amer. Math. Soc. **195** (2008), no. 910, viii+100 pp. MR2437174

[25] J. Friedman, J. Kahn and E. Szemerédi, On the second eigenvalue of random regular graphs, Proceedings of the twenty-first annual ACM symposium on Theory of computing (1989), 587–598.

[26] V. L. Girko, The circular law, Teor. Veroyatnost. i Primenen. **29** (1984), no. 4, 669–679. MR0773436

[27] F. Götze and A. Tikhomirov, The circular law for random matrices, Ann. Probab. **38** (2010), no. 4, 1444–1491. MR2663633

[28] C. Greenhill, B. D. McKay and X. Wang, Asymptotic enumeration of sparse 0-1 matrices with irregular row and column sums, J. Combin. Theory Ser. A **113** (2006), no. 2, 291–324. MR2199276

[29] G. Halász, On the distribution of additive arithmetic functons, Acta Arithmetica, **XXVII** (1975), 143–152.

[30] G. Halász, Estimates for the concentration function of combinatorial number theory and probability, Per. Math. Hung. **8**, (1977), 197–211.

[31] W. Hoeffding, Probability inequalities for sums of bounded random variables, J. Amer. Statist. Assoc. **58** (1963), 13–30. MR0144363

[32] S. Hoory, N. Linial and A. Wigderson, Expander graphs and their applications, Bull. Amer. Math. Soc. (N.S.) **43** (2006), no. 4, 439–561. MR2247919

[33] J. Kahn, J. Komlós and E. Szemerédi, On the probability that a random ± 1-matrix is singular, J. Amer. Math. Soc. **8** (1995), no. 1, 223–240. MR1260107

[34] H. Kesten, Symmetric random walks on groups, Trans. Amer. Math. Soc. **92** (1959), 336–354. MR0109367

[35] H. Kesten, A sharper form of the Doeblin-Lévy-Kolmogorov-Rogozin inequality for concentration functions, Math. Scand. **25** (1969), 133–144. MR0258095

[36] A. E. Litvak, A. Lytova, K. Tikhomirov, N. Tomczak-Jaegermann and P. Youssef, Anti-concentration property for random digraphs and invertibility of their adjacency matrices, C. R. Math. Acad. Sci. Paris **354** (2016), no. 2, 121–124. MR3456885

[37] A. E. Litvak, A. Lytova, K. Tikhomirov, N. Tomczak-Jaegermann and P. Youssef, Adjacency matrices of random digraphs: singularity and anti-concentration, J. Math. Anal. Appl. **445** (2017), no. 2, 1447–1491. MR3545253

[38] A. E. Litvak, A. Lytova, K. Tikhomirov, N. Tomczak-Jaegermann and P. Youssef, The smallest singular value of a shifted d-regular random square matrix, Probab. Theory Related Fields, to appear.

[39] A. E. Litvak, A. Lytova, K. Tikhomirov, N. Tomczak-Jaegermann and P. Youssef, The circular law for sparse random regular digraphs, submitted, arXiv:1801.05576.

[40] A. E. Litvak, A. Lytova, K. Tikhomirov, N. Tomczak-Jaegermann and P. Youssef, The rank of random regular digraphs of constant degree, J. of Complexity, **48** (2018), 103–110.

[41] 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** (2005), no. 2, 491–523. MR2146352

[42] A. E. Litvak and O. Rivasplata, Smallest singular value of sparse random matrices, Studia Math. **212** (2012), no. 3, 195–218. MR3009072

[43] B. D. McKay, The expected eigenvalue distribution of a large regular graph, Linear Algebra Appl. **40** (1981), 203–216. MR0629617

[44] B. D. McKay, Asymptotics for 0-1 matrices with prescribed line sums, in *Enumeration and design (Waterloo, Ont., 1982)*, 225–238, Academic Press, Toronto, ON. MR0782316

[45] A. L. Miroshnikov, Theory Probab. Appl. **34** (1989), no. 3, 535–540 (1990); translated from Teor. Veroyatnost. i Primenen. **34** (1989), no. 3, 593–598. MR1022646

[46] D. Puder, Expansion of random graphs: new proofs, new results, Invent. Math. **201** (2015), no. 3, 845–908. MR3385636

[47] E. Rebrova, K. Tikhomirov, Coverings of random ellipsoids, and invertibility of matrices with i.i.d. heavy-tailed entries, Israel J. Math., to appear.

[48] M. Rudelson, Invertibility of random matrices: norm of the inverse, Ann. of Math. **168** (2008), 575–600.

[49] M. Rudelson and R. Vershynin, The Littlewood-Offord problem and invertibility of random matrices, Adv. Math. **218** (2008), no. 2, 600–633. MR2407948

[50] M. Rudelson and R. Vershynin, Smallest singular value of a random rectangular matrix, Comm. Pure Appl. Math. **62** (2009), no. 12, 1707–1739. MR2569075

[51] M. Rudelson and R. Vershynin, Non-asymptotic theory of random matrices: extreme singular values, in *Proceedings of the International Congress of Mathematicians. Volume III*, 1576–1602, Hindustan Book Agency, New Delhi. MR2827856

[52] M. Rudelson and R. Vershynin, Delocalization of eigenvectors of random matrices with independent entries, Duke Math. J. **164** (2015), 2507–2538. MR3405592

[53] M. Rudelson and R. Vershynin, No-gaps delocalization for general random matrices, Geom. Funct. Anal. **26** (2016), no. 6, 1716–1776. MR3579707

[54] T. Tao and V. Vu, On the singularity probability of random Bernoulli matrices, J. Amer. Math. Soc. **20** (2007), no. 3, 603–628. MR2291914

[55] T. Tao and V. Vu, Random matrices: the circular law, Commun. Contemp. Math. **10** (2008), 261–307. MR2409368

[56] T. Tao and V. H. Vu, Inverse Littlewood-Offord theorems and the condition number of random discrete matrices, Ann. of Math. **169** (2009), 595–632. MR2480613

[57] K. Tikhomirov and P. Youssef, The spectral gap of dense random regular graphs, Annals of Prob., to appear, arXiv:1610.01765.

[58] L. V. Tran, V. H. Vu and K. Wang, Sparse random graphs: eigenvalues and eigenvectors, Random Structures Algorithms **42** (2013), no. 1, 110–134. MR2999215

[59] R. Vershynin, Introduction to the non-asymptotic analysis of random matrices, in *Compressed sensing*, 210–268, Cambridge Univ. Press, Cambridge. MR2963170

Alexander E. Litvak and Nicole Tomczak-Jaegermann,
Dept. of Math. and Stat. Sciences,
University of Alberta,
Edmonton, AB, Canada, T6G 2G1.
`e-mails: aelitvak@gmail.com and nicole.tomczak@ualberta.ca`

Anna Lytova,
Faculty of Math., Physics, and Comp. Science,
University of Opole,
plac Kopernika 11A, 45-040,
Opole, Poland.
`e-mail: alytova@math.uni.opole.pl`

Konstantin Tikhomirov,
Dept. of Math., Princeton University,
Fine Hall, Washington road,
Princeton, NJ 08544.
`e-mail: kt12@math.princeton.edu`

Pierre Youssef,
Université Paris Diderot,
Laboratoire de Probabilités, Statistiques et Modélisation,
75013 Paris, France.
`e-mail: youssef@lpsm.paris`