

FINDING A LARGE SUBMATRIX OF A GAUSSIAN RANDOM MATRIX

BY DAVID GAMARNIK¹ AND QUAN LI

Massachusetts Institute of Technology

We consider the problem of finding a $k \times k$ submatrix of an $n \times n$ matrix with i.i.d. standard Gaussian entries, which has a large average entry. It was shown in [Bhamidi, Dey and Nobel (2012)] using nonconstructive methods that the largest average value of a $k \times k$ submatrix is $2(1 + o(1))\sqrt{\log n/k}$, with high probability (w.h.p.), when $k = O(\log n / \log \log n)$. In the same paper, evidence was provided that a natural greedy algorithm called the Largest Average Submatrix (\mathcal{LAS}) for a constant k should produce a matrix with average entry at most $(1 + o(1))\sqrt{2} \log n/k$, namely approximately $\sqrt{2}$ smaller than the global optimum, though no formal proof of this fact was provided.

In this paper, we show that the average entry of the matrix produced by the \mathcal{LAS} algorithm is indeed $(1 + o(1))\sqrt{2} \log n/k$ w.h.p. when k is constant and n grows. Then, by drawing an analogy with the problem of finding cliques in random graphs, we propose a simple greedy algorithm which produces a $k \times k$ matrix with asymptotically the same average value $(1 + o(1))\sqrt{2} \log n/k$ w.h.p., for $k = o(\log n)$. Since the greedy algorithm is the best known algorithm for finding cliques in random graphs, it is tempting to believe that beating the factor $\sqrt{2}$ performance gap suffered by both algorithms might be very challenging. Surprisingly, we construct a very simple algorithm which produces a $k \times k$ matrix with average value $(1 + o_k(1) + o(1))(4/3)\sqrt{2} \log n/k$ for $k = o((\log n)^{1.5})$, that is, with the asymptotic factor $4/3$ when k grows.

To get an insight into the algorithmic hardness of this problem, and motivated by methods originating in the theory of spin glasses, we conduct the so-called expected overlap analysis of matrices with average value asymptotically $(1 + o(1))\alpha\sqrt{2} \log n/k$ for a fixed value $\alpha \in [1, \sqrt{2}]$. The overlap corresponds to the number of common rows and the number of common columns for pairs of matrices achieving this value (see the paper for details). We discover numerically an intriguing phase transition at $\alpha^* \triangleq 5\sqrt{2}/(3\sqrt{3}) \approx 1.3608 \dots \in [4/3, \sqrt{2}]$: when $\alpha < \alpha^*$ the space of overlaps is a continuous subset of $[0, 1]^2$, whereas $\alpha = \alpha^*$ marks the onset of discontinuity, and as a result the model exhibits the *Overlap Gap Property (OGP)* when $\alpha > \alpha^*$, appropriately defined. We conjecture that the *OGP* observed for $\alpha > \alpha^*$ also marks the onset of the algorithmic hardness—no polynomial time algorithm exists for finding matrices with average value at least $(1 + o(1))\alpha\sqrt{2} \log n/k$, when $\alpha > \alpha^*$ and k is a mildly growing function of n .

Received March 2016; revised June 2017.

¹Supported by the NSF Grant CMMI-1335155.

MSC2010 subject classifications. 68Q87, 97K50, 60C05, 68Q25.

Key words and phrases. Random matrix, random graphs, maximum clique, submatrix detection, computational complexity, overlap gap property.

CONTENTS

1. Introduction	2512
2. Main results	2516
3. Analysis of the \mathcal{IGP} algorithm	2524
4. The overlap gap property	2527
5. Analysis of the \mathcal{LAS} algorithm	2532
5.1. Preliminary results	2532
5.2. Conditional distribution of the row-dominant and column-dominant submatrices	2535
5.3. Bounding the number of steps of \mathcal{LAS} . Proof of Theorem 2.1	2544
6. Conclusions and open questions	2557
Appendix A: Proof of Lemma 3.1	2558
Appendix B: Derivation of two phase transition points $\alpha_1^* = \sqrt{3}/\sqrt{2}$ and $\alpha_2^* = 5\sqrt{2}/(3\sqrt{3})$	2559
References	2561

1. Introduction. We consider the algorithmic problem of finding a submatrix of a given random matrix such that the average value of the submatrix is appropriately large. Specifically, consider an $n \times n$ matrix \mathbf{C}^n with i.i.d. standard Gaussian entries. Given $k \leq n$, the goal is to find algorithmically a $k \times k$ submatrix \mathbf{A} of \mathbf{C}^n (not necessarily principal) with average entry as large as possible. The problem has motivations in several areas, including biomedicine, genomics and social networks [17], [14], [8]. The search of such matrices is called “biclustering” [14]. The problem of finding asymptotically the $k \times k$ submatrix with the largest average entry among all $k \times k$ submatrices of \mathbf{C}^n was recently studied by Bhamidi et al. [6] (see also [18] for a related study) and questions arising in this paper constitute the motivation for our work. It was shown in [6] using nonconstructive methods that the largest achievable average entry of a $k \times k$ submatrix of \mathbf{C}^n is asymptotically with high probability (w.h.p.) $(1 + o(1))2\sqrt{\log n/k}$ when n grows and $k = O(\log n / \log \log n)$ (a more refined distributional result is obtained). Here, $o(1)$ denotes a function converging to zero as $n \rightarrow \infty$. Furthermore, the authors consider the asymptotic value and the number of so-called locally maximum matrices. A $k \times k$ matrix \mathbf{A} is locally maximal if every $k \times k$ matrix of \mathbf{C}^n with the same set of rows as \mathbf{A} has a smaller average value than that of \mathbf{A} , and every $k \times k$ matrix of \mathbf{C}^n with the same set of columns as \mathbf{A} has a smaller average value than that of \mathbf{A} . Such local maxima are natural objects arising as terminal matrices produced by a simple iterative procedure called the Large Average Submatrix (\mathcal{LAS}), designed for finding a matrix with a large average entry. \mathcal{LAS} proceeds by starting from an arbitrary $k \times k$ submatrix \mathbf{A}_0 and finding a matrix \mathbf{A}_1 sharing the same set of rows with \mathbf{A}_0 , which has the largest average value. The procedure is then repeated for \mathbf{A}_1 by searching through columns of \mathbf{A}_1 and identifying the best matrix \mathbf{A}_2 . The iterations proceed while possible and at the end some locally maximum matrix $\mathbf{A}_{\mathcal{LAS}}$ is produced as the output. The authors show that when k is constant, the majority of locally maximum matrices of \mathbf{C}^n have the asymptotic value $(1 + o(1))\sqrt{2 \log n/k}$ w.h.p. as n grows, thus factor $\sqrt{2}$ smaller than the global optimum. Motivated by

this result, the authors suggest that the outcome of the \mathcal{LAS} algorithm should be also factor $\sqrt{2}$ smaller than the global optimum; however, one cannot deduce this from the result of [6], since it is not ruled out that \mathcal{LAS} might be clever enough to find a “rare” locally maximum matrix with a significantly larger average value than $\sqrt{2 \log n/k}$.

The main result of this paper is the confirmation of this conjecture for the case of constant k : the \mathcal{LAS} algorithm produces a matrix with asymptotic average value $(1 + o(1))\sqrt{2 \log n/k}$ w.h.p. We further establish that the number of iterations of the \mathcal{LAS} algorithm is stochastically bounded as n grows. The proof of this result is fairly involved and proceeds by a careful conditioning argument. In particular, we show that for fixed r , conditioned on the event that \mathcal{LAS} succeeded in iterating at least r steps, the probability distribution of the “new best matrix” which will be used in constructing the matrix for the next iteration is very close to the largest matrix in the $k \times n$ strip of \mathbf{C}^n , and which is known to have asymptotic average value of $\sqrt{2 \log n/k}$ due to result in [6]. Then we show that the matrix produced in step r and the best matrix in the $k \times n$ strip among the unseen entries are asymptotically independent. Using this, we show that, given that \mathcal{LAS} proceeded with r steps, the likelihood that it proceeds with the next $r + 2k + 4$ steps is at most some value $\psi < 1$, which is bounded away from 1 as n grows. As a result, the number of steps of \mathcal{LAS} is upper bounded by a geometrically decaying function, and thus is stochastically bounded as n grows. We use this as a key result in computing the average value produced by \mathcal{LAS} , again relying on the asymptotic independence and the average value of the $k \times n$ strip dominant submatrix.

As it was observed already in [6], the factor $\sqrt{2}$ gap between the global optimum and the performance of \mathcal{LAS} is reminiscent of a similar gap arising in studying the largest cliques of random graphs. Arguably, one of the oldest algorithmic open problems in the field of random graph is the problem of finding a largest clique (a fully connected subgraph) of a random Erdős–Rényi graph $\mathbb{G}(n, p)$, when p is at least $n^{-1+\delta}$ for some positive constant δ . It is known that the value is asymptotically $2 \log n / (-\log p)$ and a simple greedy procedure produces a clique with size $\log n / (-\log p)$, namely factor 2 smaller than the global optimum. A similar result holds for the bipartite Erdős–Rényi graph: the largest clique is asymptotically $2 \log n / (-\log p)$ and the greedy algorithm produces a (bipartite) clique of size asymptotically $\log n / (-\log p)$. Karp in his 1976 paper [12] challenged to find a better algorithm, leading to a clique with size say $(1 + \varepsilon) \log n / (-\log p)$, and this problem remains open. The factor $\sqrt{2}$ appearing in our context is then arguably an analogue of the factor 2 arising in the context of the clique problem in $\mathbb{G}(n, p)$. In order to further investigate the possible connection between the two problems, we propose the following simple algorithm for finding a submatrix of \mathbf{C}^n with a large average entry. Fix a positive threshold θ and consider the random 0, 1 matrix \mathbf{C}_θ^n obtained by thresholding each Gaussian entry of \mathbf{C}^n at θ . Clearly \mathbf{C}_θ^n is an adjacency matrix of a bipartite Erdős–Rényi graph $\mathbb{G}(n, p_\theta)$, where $p_\theta = \mathbb{P}(Z > \theta)$ and Z is a standard Gaussian random variable. Observe that any $k \times k$ clique of

$\mathbb{G}(n, p_\theta)$ corresponds to a $k \times k$ submatrix of \mathbb{C}^n with *each* entry at least θ . Thus any polynomial time algorithm which finds a $k \times k$ clique in $\mathbb{G}(n, p_\theta)$ w.h.p. immediately gives a matrix with the average value at least θ w.h.p. Consider the greedy algorithm and adjust θ so that the size of the clique is at least k on each side. Reverse engineering θ from such k , one can find that $\theta \approx \sqrt{2 \log n/k}$ with $p \approx \exp(-\theta^2/2) = n^{-\frac{1}{k}}$ (see the next section for a simple derivation of this fact). Namely, both \mathcal{LAS} and the greedy algorithm have the same asymptotic power. [Note, however, that this analysis extends beyond the $k = O(1)$ unlike our analysis of the \mathcal{LAS} algorithm.]

In light of these connections with studying cliques in random graphs and the apparent failure to bridge the factor 2 gaps for cliques, one might suspect that $\sqrt{2}$ is equally challenging to beat for the maximum submatrix problem. Perhaps surprisingly, we establish that this is not the case, and we construct a very simple algorithm, both in terms of the analysis and implementation, which constructs a submatrix with the average value asymptotically $(1 + o_k(1) + o(1))\sqrt{2 \log n/k}$, for $k = o((\log n)^{1.5})$. Here, $o_k(1)$ denotes a function decaying to zero as k increases. That is, the asymptotic factor $4/3$ is valid for growing k . The algorithm proceeds by starting from one entry and iteratively building a sequence of $r \times r$ and $r \times (r + 1)$ matrices for $r = 1, \dots, k$ in a simple greedy fashion. We call this algorithm the Incremental Greedy Procedure (\mathcal{IGP}), referring to the incremental increase of the matrix size. No immediate simple modifications of \mathcal{IGP} led to the improvement of the $4/3$ factor, unfortunately.

The discussion above raises the following question: where is the true algorithmic hardness threshold value for the maximum submatrix problem if such exists? Short of proving some formal hardness of this problem, which seems out of reach for the currently known techniques both for this problem and the clique problem in $\mathbb{G}(n, p)$, we propose an approach which indirectly suggests the hardness regime for this problem, and this is our last contribution. Specifically, our last contribution is the conjecture for this value based on the *Overlap Gap Property* (OGP) which originates in the theory of spin glasses and which we adopt here in the context of our problem in the following way. We fix $\alpha \in (1, \sqrt{2})$ and let $\mathcal{L}(\alpha)$ denote the set of matrices with average value asymptotically $\alpha\sqrt{2 \log n/k}$. Thus α conveniently parameterizes the range between the achievable value, namely $\alpha = 1$ for \mathcal{LAS} and the greedy algorithms, $\alpha = 4/3$ for the \mathcal{IGP} , and $\alpha = \sqrt{2}$ for the global optimum. For every pair of matrices $\mathbf{A}_1, \mathbf{A}_2 \in \mathcal{L}(\alpha)$ with row sets I_1, I_2 and column sets J_1, J_2 , respectively, let $x(\mathbf{A}_1, \mathbf{A}_2) = |I_1 \cap I_2|/k$, $y(\mathbf{A}_1, \mathbf{A}_2) = |J_1 \cap J_2|/k$. Namely, x and y are the normalized counts of the common rows and common columns for the two matrices. For every $(x, y) \in [0, 1]^2$, we consider the expected number of pairs $\mathbf{A}_1, \mathbf{A}_2$ such that $x(\mathbf{A}_1, \mathbf{A}_2) \approx x$, $y(\mathbf{A}_1, \mathbf{A}_2) \approx y$, in some appropriate sense to be made precise. We compute this expectation asymptotically. We define $R(x, y) = 0$ if such an expectation converges to zero as $n \rightarrow \infty$ and $= 1$ otherwise. Thus the set $\mathcal{R}(\alpha) \triangleq \{(x, y) : R(x, y) = 1\}$ describes the set of achievable

in expectation overlaps of pairs of matrices with average value $\alpha\sqrt{2\log n/k}$. At $\alpha^* \triangleq 5\sqrt{2}/(3\sqrt{3}) \approx 1.3608\dots$ we observe an interesting phase transition—the set $\mathcal{R}(\alpha)$ is connected for $\alpha < \alpha^*$, and is disconnected for $\alpha > \alpha^*$ (see Figure 5). Namely, for $\alpha > \alpha^*$ the model exhibits the OGP: the overlaps of two matrices belong to the points of a disconnected region.

Motivated by this observation, we conjecture that the problem of finding a matrix with the corresponding value $\alpha > \alpha^*$ is not polynomially solvable when k grows. In fact, by considering multi-overlaps instead of pairwise overlaps, (which we intend to research in future), we conjecture that this hardness threshold might be even lower than α^* . The link between OGP and algorithmic hardness has been suggested and partially established in the context of sparse random constraint satisfaction problems, such as random K-SAT problem, coloring of the sparse Erdős–Rényi problem and the problem of finding a largest independent set of a sparse Erdős–Rényi graph problem [2], [1], [7], [9], [16], [10], [15]. A more recent link between OGP and tractable algorithms was recently established in the context of the sparse high dimensional linear regression problem [11]. Many of these problems exhibit an apparent gap between the best existential values and the best values found by known algorithms, very similar in spirit to the gaps 2 , $\sqrt{2}$, etc. discussed above in our context. For example, the largest independent set of a random d -regular graph normalized by the number of nodes is known to be asymptotically $2\log d/d$ as d increases, while the best algorithm can produce independent sets of size only $\log d/d$, again as d increases. As shown in [7], [9] and [16], the threshold $\log d/d$ marks the onset of a certain version of OGP. Furthermore, [7], [9] show that the OGP is the bottleneck for a certain class of algorithms, namely local algorithms (appropriately defined). Roughly speaking, an algorithm is local if for each node of a graph, the decisions resulting in a construction of a solution are conducted by taking account only small neighborhood of a node. Then by a careful coupling construction, it can be shown that solutions resulting from local algorithm exhibit a certain continuity property, which is inconsistent with the OGP. Thus, such algorithms cannot overcome the OGP barrier. For the case of the sparse linear regression problem in [11], it is shown that local improvement type algorithms also exhibit a continuity property, also inconsistent with the OGP. It is further shown that the onset of the OGP occurs asymptotically around the threshold, below which the celebrated compressive sensing methods such as LASSO and Dantzig Selector fail.

A key step observed in [16] is that the threshold for multi-overlap version of the OGP, namely considering m -tuples of solutions as opposed to pairs of solutions as we do in this paper, lowers the OGP phase transition point. The multi-overlap version of OGP was also a key step in [10] in the context of the random Not-All-Equal-K-SAT (NAE-K-SAT) problem, which also exhibits a marked gap between the regime where the existence of a feasible solution is known, and the regime where such a solution can be found by known polynomial time algorithms. We

conjecture that the OGP threshold established in this paper can also be lowered by considering overlaps of $m \geq 3$ matrices.

The OGP for the largest submatrix problem thus adds to the growing class of optimization problems with random input which exhibit a significant gap between the value achievable by an optimal solution and values achievable by the currently known tractable algorithmic methods, and where the gap is evidenced by the phase transition associated with the OGP.

The remainder of the paper is structured as follows. In the next section, we formally state our four main results: the result regarding the performance of \mathcal{LAS} , the result regarding the performance of the greedy algorithm via a reduction to random bipartite graphs, the result regarding the performance of \mathcal{IGP} and finally the result regarding the OGP. The same section provides a short proof for the result regarding the greedy algorithm. Section 3 is devoted to the proof of the result regarding the performance of \mathcal{IGP} . Section 4 is devoted to the proof of the result regarding OGP, and Section 5 (which is the most technically involved part of the paper) is devoted to the proof of the result regarding the performance of the \mathcal{LAS} algorithm. We conclude in Section 6 with some open questions.

We close this section with some notational convention. We use standard notation $o(\cdot)$, $O(\cdot)$ and $\Theta(\cdot)$ with respect to $n \rightarrow \infty$. $o_k(1)$ denotes a function $f(k)$ satisfying $\lim_{k \rightarrow \infty} f(k) = 0$. Given a positive integer n , $[n]$ stands for the set of integers $1, \dots, n$. For any real value α , $\lfloor \alpha \rfloor$ denotes the largest integer that is less than or equal to α . Given a matrix A , A^T denotes its transpose. \Rightarrow denotes weak convergence. $\stackrel{d}{=}$ denotes equality in distribution. A complement of event \mathcal{A} is denoted by \mathcal{A}^c . For two events \mathcal{A} and \mathcal{B} , we write $\mathcal{A} \cap \mathcal{B}$ and $\mathcal{A} \cup \mathcal{B}$ for the intersection (conjunction) and the union (disjunction) of the two events, respectively. When conditioning on the event $\mathcal{A} \cap \mathcal{B}$, we will often write $\mathbb{P}(\cdot | \mathcal{A}, \mathcal{B})$ in place of $\mathbb{P}(\cdot | \mathcal{A} \cap \mathcal{B})$. For nonnegative integers b_1, b_2, \dots, b_l such that $\sum_{i=1}^l b_i = n$, the multinomial coefficient is

$$\binom{n}{b_1, b_2, \dots, b_l} = \frac{n!}{b_1! b_2! \dots b_l!}.$$

Let $\Phi(u)$ be the cumulative distribution function of the standard normal random variable. When u is large, the function $1 - \Phi(u)$ can be approximated by

$$(1) \quad \frac{1 - 2u^{-2}}{u\sqrt{2\pi}} \exp(-u^2/2) \leq 1 - \Phi(u) \leq \frac{1}{u\sqrt{2\pi}} \exp(-u^2/2).$$

2. Main results. In this section, we formally describe the algorithms analyzed in this paper and state our main results. Given an $n \times n$ matrix A and subsets $I \subset [n]$, $J \subset [n]$, we denote by $A_{I,J}$ the submatrix of A indexed by rows I and columns J . When I consists of a single row i , we use $A_{i,J}$ in place of a more proper notation $A_{\{i\},J}$. Given any $m_1 \times m_2$ matrix B , let $\text{Ave}(B) \triangleq \frac{1}{m_1 m_2} \sum_{i,j} B_{i,j}$ denote the average value of the entries of B .

Algorithm 1 Large average submatrix algorithm (\mathcal{LAS}).

Input: An $n \times n$ matrix A and a fixed integer $k \geq 1$.
Initialize: Select k rows I and k columns J arbitrarily.
Loop: (Iterate until no improvement is achieved)
 Find the set $\hat{J} \subset [n]$, $|\hat{J}| = k$ such that $\text{Ave}(A_{I, \hat{J}}) \geq \text{Ave}(A_{I, J'})$, for all $J' \subset [n]$, $|J'| = k$. Break ties arbitrarily.
 If $\hat{J} = J$, STOP. Otherwise, set $J = \hat{J}$.
 Find the set $\hat{I} \subset [n]$, $|\hat{I}| = k$ such that $\text{Ave}(A_{\hat{I}, J}) \geq \text{Ave}(A_{I', J})$, for all $I' \subset [n]$, $|I'| = k$. Break ties arbitrarily.
 If $\hat{I} = I$, STOP. Otherwise, Set $I = \hat{I}$.
Output: $A_{I, J}$.

Let $\mathbf{C} = (\mathbf{C}_{ij}, i, j \geq 1)$ denote an infinite two-dimensional array of independent standard normal random variables. Denote by $\mathbf{C}^{n \times m}$ the $n \times m$ upper left corner of \mathbf{C} . If $n = m$, we use \mathbf{C}^n instead.

The large average submatrix algorithm is defined as Algorithm 1.

Since the entries of \mathbf{C}^n are continuous independent random variables, the ties in the \mathcal{LAS} algorithm occur with zero probability. Each step of the \mathcal{LAS} algorithm is easy to perform, since given a fixed set of rows I , finding the corresponding set of columns \hat{J} which leads to the matrix with maximum average entry is easy: simply find k columns corresponding to k largest entry sums. Also the algorithm will stop after finitely many iterations since in each step the matrix sum (and the average) increases and the number of submatrices is finite. In fact, a major part of our analysis is to bound the number of steps of \mathcal{LAS} . Our convention is that in step zero, the \mathcal{LAS} algorithm sets $I_0 = I = \{1, \dots, k\}$ and $J_0 = J = \{1, \dots, k\}$. We denote by $T_{\mathcal{LAS}}$ the number of iterations of the \mathcal{LAS} algorithm applied to the $n \times n$ matrix \mathbf{C}^n with i.i.d. standard normal entries. For concreteness, searching for \hat{I} and \hat{J} are counted as two separate iterations. We denote by \mathbf{C}_r^n the matrix produced by \mathcal{LAS} in step (iteration) r , assuming $T_{\mathcal{LAS}} \geq r$. Thus our goal is to obtain asymptotic values of $\text{Ave}(\mathbf{C}_{T_{\mathcal{LAS}}}^n)$, as well as the number of iterations $T_{\mathcal{LAS}}$.

Our first main result concerns the performance of \mathcal{LAS} and is stated as follows. Let ω_n denote any positive function satisfying $\omega_n = o(\sqrt{\log n})$ and $\log \log n = O(\omega_n)$.

THEOREM 2.1. *Suppose a positive integer k is fixed. For every $\varepsilon > 0$, there is a positive integer N which depends on k and ε only, such that for all $n \geq N$, $\mathbb{P}(T_{\mathcal{LAS}} \geq N) \leq \varepsilon$. Furthermore,*

$$(2) \quad \lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \text{Ave}(\mathbf{C}_{T_{\mathcal{LAS}}}^n) - \sqrt{\frac{2 \log n}{k}} \right| \leq \omega_n \right) = 1.$$

Theorem 2.1 states that the average of the $k \times k$ submatrix produced by \mathcal{LAS} converges to the value $(1 + o(1))\sqrt{2 \log n/k}$, as n increases, and furthermore, the number of iterations is stochastically bounded in n . In fact, we will show the existence of a constant $0 < \psi < 1$, which depends on k and ε only, such that $\mathbb{P}(T_{\mathcal{LAS}} > t) \leq \psi^t$, $t \geq 1$ for all large enough n . Namely, $T_{\mathcal{LAS}}$ is bounded by a geometric random variable for all large enough n .

Next we turn to the performance of the greedy algorithm applied to the random graph produced from \mathbf{C}^n by first thresholding it at a certain level θ . Given \mathbf{C}^n , let $\mathbb{G}(n, n, p(\theta))$ denote the corresponding $n \times n$ bipartite graph, where the edge (i, j) , $i, j \in [n]$ is present if $\mathbf{C}_{i,j}^n > \theta$ and is absent otherwise. The edge probability is then $p(\theta) = \mathbb{P}(Z > \theta)$, where Z is a standard normal random variable. A pair of subsets $I \subset [n]$, $J \subset [n]$ is a clique in $\mathbb{G}(n, n, p(\theta))$ if the edge (i, j) exists for every $i \in I, j \in J$. In this case, we write $i \sim j$.

Consider the following simple algorithm for generating a clique in $\mathbb{G}(n, n, p(\theta))$, which we call a greedy algorithm for simplicity. Pick node $i_1 = 1$ on the left part of the graph and let $J_1 = \{j : 1 \sim j\}$. Pick any node $j_1 \in J_1$ and let $I_1 = \{i \in [n] : i \sim j_1\}$. Clearly, $i_1 \in I_1$. Pick any node $i_2 \in I_1$ different from i_1 and let $J_2 = \{j \in J_1 : i_2 \sim j\}$. Clearly, $j_1 \in J_2$. Pick any $j_2 \in J_2$ different from j_1 and let $I_2 = \{i \in I_1 : i \sim j_2\}$, and so on. Repeat this process for as many steps m as possible ending it on the right-hand side of the graph, so that the number of chosen nodes on the left and the right is the same. The end result I_m, J_m is clearly a clique. It is also immediate that $|I_m| = |J_m| = m$. The corresponding submatrix \mathbf{C}_{I_m, J_m}^n of \mathbf{C}^n indexed by rows I_m and columns J_m has every entry at least θ and, therefore, $\text{Ave}(\mathbf{C}_{I_m, J_m}^n) \geq \theta$. If we can guarantee that θ is small enough so that m is at least k , we obtain a simple algorithm for producing a $k \times k$ matrix with average entry at least θ . From the theory of a random graph, it is known (and easy to establish) that w.h.p. our greedy algorithm produces a clique of size asymptotically $\log n / \log(1/p)$, as n increases, provided that p is at least $n^{-1+\varepsilon}$ for some $\varepsilon > 0$. Since we need to produce a $k \times k$ clique, we obtain a requirement $\log n / \log(1/p) \geq k$ (provided of course the lower bound $n^{-1+\varepsilon}$ holds, which we will verify retroactively), leading to

$$p = \mathbb{P}(Z > \theta) \geq n^{-\frac{1}{k}},$$

and in particular $k \geq 2$ is enough to satisfy the $n^{-1+\varepsilon}$ lower bound requirement. Now suppose $k = o(\log n)$, implying $n^{-\frac{1}{k}} = o(1)$. Solving for θ_n defined by

$$\mathbb{P}(Z > \theta_n) = n^{-\frac{1}{k}},$$

and using the fact

$$\lim_{t \rightarrow \infty} t^{-2} \log \mathbb{P}(Z > t) = -\frac{1}{2},$$

which is an easy consequence of (1), we conclude that

$$\theta_n = (1 + o(1))\sqrt{\frac{2 \log n}{k}},$$

leading to the same average value as the \mathcal{LAS} algorithm. The two algorithms have asymptotically the same performance [though the greedy algorithm guarantees a *minimum* value of $(1 + o(1))\sqrt{\frac{2 \log n}{k}}$ as opposed to just the (same) average value]. We summarize our finding as follows.

THEOREM 2.2. *Setting $\theta_n = (1 + o(1))\sqrt{\frac{2 \log n}{k}}$, the greedy algorithm w.h.p. produces a $k \times k$ submatrix with minimum value θ_n for $k = o(\log n)$.*

Next we turn to an improved algorithm for finding a $k \times k$ submatrix with a large average entry, which we call Incremental Greedy Procedure (\mathcal{IGP}) and which achieves the $(1 + o_k(1))(4/3)\sqrt{2 \log n/k}$ asymptotics. We first provide a heuristic idea behind the algorithm which ignores certain dependencies, and then provide the appropriate fix for dealing with the dependency issue. The algorithm is described informally as follows. Fix an arbitrary $i_1 \in [n]$ and in the corresponding row $\mathbf{C}_{i_1, [n]}^n$ find the largest element \mathbf{C}_{i_1, j_1}^n . This term is asymptotically $\sqrt{2 \log n}$ as the largest of n i.i.d. standard normal random variables [see (21) in Section 5]. Then find the largest element \mathbf{C}_{i_2, j_1}^n in the column $\mathbf{C}_{[n], j_1}$ other than \mathbf{C}_{i_1, j_1} , which asymptotically is also $\sqrt{2 \log n}$. Next, in the $2 \times n$ matrix $\mathbf{C}_{\{i_1, i_2\}, [n]}^n$ find a column $j_2 \neq j_1$ such that the sum of the two elements of the column $\mathbf{C}_{\{i_1, i_2\}, j_2}^n$ is larger than the sum for all other columns $\mathbf{C}_{\{i_1, i_2\}, j}^n$ for all $j \neq j_1$. Ignoring the dependencies, this sum is asymptotically $\sqrt{2}\sqrt{2 \log n}$, though the dependence is present here since the original row $\mathbf{C}_{i_1, [n]}$ is a part of this computation. We have created a 2×2 matrix $(\mathbf{C}_{i, j}^n, i = i_1, i_2; j = j_1, j_2)$. Then we find a row $i_3 \neq i_1, i_2$ such that the sum of the two elements of the row $\mathbf{C}_{i_3, \{j_1, j_2\}}$ is larger than any other such sum of $\mathbf{C}_{i, \{j_1, j_2\}}$ for $i \neq i_1, i_2$. Again, ignoring the dependencies, this average is asymptotically $\sqrt{2}\sqrt{2 \log n}$. We continue in this fashion, greedily and incrementally expanding the matrix to a larger sizes, creating in alternation $r \times r$ and $(r + 1) \times r$ matrices, and stop when $r = k$ and we arrive at a $k \times k$ matrix. In each step, ignoring the dependencies, the sum of the elements of the added row and added column is $\sqrt{r}\sqrt{2 \log n}$ when the number of elements in the row and in the column is r , again ignoring the dependency. Thus we expect the total asymptotic size of the final matrix to be

$$2 \sum_{1 \leq r \leq k-1} \sqrt{r}\sqrt{2 \log n} + \sqrt{k}\sqrt{2 \log n}.$$

Approximating $2 \sum_{1 \leq r \leq k-1} \sqrt{r} + \sqrt{k}$ by $2 \int_1^k \sqrt{x} dx \approx 4k^{3/2}/3$ for growing k and then dividing the expression above by k^2 , we obtain the required asymptotics. The

flaw in the argument above comes from ignoring the dependencies: when $r \times 1$ row is chosen among the best such rows outside of the already created $r \times r$ matrix, the distribution of this row is dependent on the distribution of this matrix. A simple fix comes from partitioning the entire $n \times n$ matrix into $k \times k$ equal size groups, and only searching for the best $r \times 1$ row within the respective group. The sum of the elements of the r th added row is then $\sqrt{r} \sqrt{2 \log(n/k)}$, which is asymptotically the same as $\sqrt{r} \sqrt{2 \log n}$, provided k is small enough. The independence of entries between the groups is then used to estimate rigorously the performance of the algorithm.

We now formalize the approach and state our main result. The proof for the performance of the algorithm is in Section 3. Given $n \in \mathbb{Z}^+$ and $k \in [n]$, divide the set $[n]$ into $k + 1$ disjoint subsets, where the first k subsets are

$$P_i^n = \{(i - 1)\lfloor n/k \rfloor + 1, (i - 1)\lfloor n/k \rfloor + 2, \dots, i\lfloor n/k \rfloor\} \quad \text{for } i = 1, 2, \dots, k.$$

When n is a multiple of k , the last subset is by convention an empty set. Otherwise, it is $\{k\lfloor n/k \rfloor + 1, \dots, n\}$. The last subset is not used in the construction below. A detailed description of \mathcal{IGP} algorithm is as Algorithm 2.

As shown in Figure 1, the \mathcal{IGP} algorithm in step $2r$ adds a row of r entries (represented by symbol “ Δ ”) with the largest entry sum, to the previous $r \times r$ submatrix $\mathbf{C}_{\mathcal{IGP}}^{n, 2r-1}$. Similarly, as shown in Figure 2, the \mathcal{IGP} algorithm in step $2r + 1$ adds a column of $r + 1$ entries (represented by symbol “ Δ ”) with the largest entry sum to the previous $(r + 1) \times r$ submatrix $\mathbf{C}_{\mathcal{IGP}}^{n, 2r}$.

Just as for the \mathcal{LAS} algorithm, each step of \mathcal{IGP} algorithm is easy to perform: simply find one column (row) corresponding to the largest entry sum. The algorithm stops after $2k$ steps. We denote by $\mathbf{C}_{\mathcal{IGP}}^n$ the $k \times k$ submatrix produced by \mathcal{IGP} applied to \mathbf{C}^n . Our goal is to obtain the asymptotic value of $\text{Ave}(\mathbf{C}_{\mathcal{IGP}}^n)$.

Our main result regarding the performance of the \mathcal{IGP} algorithm is as follows.

Algorithm 2 \mathcal{IGP} algorithm.

Input: An $n \times n$ matrix A and a fixed integer $k \geq 1$.

Initialize: Select $i_1 \in P_1^n$ arbitrarily and set $I = \{i_1\}$, and let $J = \emptyset$.

Loop: Proceed until $|I| = |J| = k$

 Find the column $j \in P_{|I|}^n$ such that $\text{Ave}(A_{I,j}) \geq \text{Ave}(A_{I,j'})$ for all $j' \in P_{|I|}^n$.
 Set $J = J \cup \{j\}$.

 Find the $i \in P_{|I|+1}^n$ such that $\text{Ave}(A_{i,J}) \geq \text{Ave}(A_{i',J})$ for all $i' \in P_{|I|+1}^n$. Set $I = I \cup \{i\}$.

Output: $A_{I,J}$.

$$\begin{bmatrix} C_{11} & & \dots & & C_{1n} \\ C_{21} & & & r & C_{2n} \\ \vdots & r+1 & \begin{bmatrix} * & \dots & * \\ \vdots & \mathbf{C}_{IGP}^{n,2r} & \vdots \\ * & \dots & * \\ \Delta & \dots & \Delta \end{bmatrix} & & \vdots \\ C_{n1} & & \dots & & C_{nn} \end{bmatrix}$$

FIG. 1. Step $2r$ of IGP algorithm.

THEOREM 2.3. *Let $f(n)$ be any positive function such that $f(n) = o((\log n)^{1.5})$. Then*

$$(3) \quad \lim_{n \rightarrow \infty} \min_{1 \leq k \leq f(n)} \mathbb{P} \left(\left| \text{Ave}(\mathbf{C}_{IGP}^n) - \frac{4}{3} \sqrt{\frac{2 \log n}{k}} \right| \leq 3 \max \left(\frac{1}{k} \sqrt{\frac{\log n}{k}}, \frac{\log \log n}{\sqrt{\log n}} \right) \right) = 1.$$

The bound on the right-hand side is of the order magnitude $O(\sqrt{\log n})$ when k is constant and is $o(\sqrt{\log n/k})$ when k is a growing function of n . The asymptotics $(1 + o_k(1) + o(1)) \frac{4}{3} \sqrt{\frac{2 \log n}{k}}$ corresponds to the latter case.

Next we turn to the discussion of the OGP. Fix $\alpha \in (1, \sqrt{2})$, real values $0 < y_1, y_2 < 1$ and $\delta \in (0, \alpha)$. Let $\mathcal{O}_k(\alpha, y_1, y_2, \delta)$ denote the set of pairs of $k \times k$ submatrices $\mathbf{C}_{I_1, J_1}^n, \mathbf{C}_{I_2, J_2}^n$ with average value in the interval $[(\alpha - \delta)\sqrt{2 \log n/k}, (\alpha + \delta)\sqrt{2 \log n/k}]$, and which satisfy $|I_1 \cap I_2|/k \in (y_1 - \delta, y_1 + \delta), |J_1 \cap J_2|/k \in (y_2 - \delta, y_2 + \delta)$. Namely, $\mathcal{O}_k(\alpha, y_1, y_2, \delta)$ is the set of pairs of $k \times k$ matrices with average value approximately $\alpha \sqrt{2 \log n/k}$, and which share approximately

$$\begin{bmatrix} C_{11} & & \dots & & C_{1n} \\ C_{21} & & & r+1 & C_{2n} \\ \vdots & r+1 & \begin{bmatrix} * & \dots & * & \Delta \\ \vdots & \mathbf{C}_{IGP}^{n,2r+1} & \vdots & \vdots \\ * & \dots & * & \Delta \end{bmatrix} & & \vdots \\ C_{n1} & & \dots & & C_{nn} \end{bmatrix}$$

FIG. 2. Step $2r + 1$ of IGP algorithm.

y_1k rows and y_2k columns. Let

$$(4) \quad f(\alpha, y_1, y_2) \triangleq 4 - y_1 - y_2 - \frac{2}{1 + y_1y_2}\alpha^2.$$

The next result says that the expected cardinality of the set $\mathcal{O}_k(\alpha, y_1, y_2, \delta)$ is approximately $n^{kf(\alpha, y_1, y_2)}$ when $f(\alpha, y_1, y_2)$ is positive, and on the other hand, $\mathcal{O}_k(\alpha, y_1, y_2, \delta)$ is empty w.h.p. when $f(\alpha, y_1, y_2)$ is negative.

THEOREM 2.4. *Fix $\alpha \in (1, \sqrt{2})$. For every $\varepsilon > 0$, $c > 0$ and $y_1, y_2 \in (0, 1)$, there exists $\delta_0 \in (0, \alpha)$ and $n_0 > 0$ such that for all $n \geq n_0$, $k \leq c \log n$ and $\delta \in (0, \delta_0)$*

$$(5) \quad \left| \frac{\log \mathbb{E}[|\mathcal{O}_k(\alpha, y_1, y_2, \delta)|]}{k \log n} - f(\alpha, y_1, y_2) \right| < \varepsilon.$$

As a result, when $f(\alpha, y_1, y_2) < 0$, for every $\varepsilon > 0$ and $c > 0$, there exists $\delta \in (0, \alpha)$ and $n_0 > 0$ such that for all $n \geq n_0$ and $k \leq c \log n$

$$(6) \quad \mathbb{P}(\mathcal{O}_k(\alpha, y_1, y_2, \delta) \neq \emptyset) < \varepsilon.$$

We see that the region $\mathcal{R}(\alpha) \triangleq \{(y_1, y_2) : f(\alpha, y_1, y_2) \geq 0\}$ identifies the region of achievable in expectation overlaps for matrices with average values approximately $\alpha\sqrt{2\log n/k}$.

Viewing the region $\mathcal{R}(\alpha)$ as a function of α , we establish two phase transition points: the first one at $\alpha_1^* = \sqrt{3/2} = 1.2247\dots$, and the second one at $\alpha_2^* = 5\sqrt{2}/(3\sqrt{3}) = 1.3608\dots$. The derivation of these values is delayed till Section 4. Computing $\mathcal{R}(\alpha)$ numerically we will see that it exhibits three qualitatively different behaviors for $\alpha \in (0, \alpha_1^*)$, (α_1^*, α_2^*) and $(\alpha_2^*, \sqrt{2})$, respectively, as shown in Figures 3, 4 and 5. These figures are heat maps of f , where the darker color corresponds to the higher value of f , and the lighter color corresponds to the lower value of f .

(a) When $\alpha \in (1, \sqrt{3}/\sqrt{2})$, $\mathcal{R}(\alpha)$ coincides with the entire region $[0, 1]^2$; see Figure 3. In this figure, in particular, the part closer to the origin, we see that the bulk of the overlaps corresponds to the pairs (y_1, y_2) which are very close to the origin. In other words, the picture suggests that most matrices with average value approximately $\alpha\sqrt{2\log n/k}$ tend to be far from each other.

(b) When $\alpha \in (\sqrt{3}/\sqrt{2}, 5\sqrt{2}/(3\sqrt{3}))$, we see that $\mathcal{R}(\alpha)$ is a connected subset of $[0, 1]^2$ (Figure 4), but a nonachievable overlap region emerges (colored white on the figure) for pairs of matrices with this average value. At a critical value $\alpha = 5\sqrt{2}/(3\sqrt{3})$, the set is connected through a single point $(1/3, 1/3)$; see Figure 6.

(c) When $\alpha \in (5\sqrt{2}/(3\sqrt{3}), \sqrt{2})$, $\mathcal{R}(\alpha)$ is a disconnected subset of $[0, 1]^2$, and the OGP emerges; see Figure 5 for the case $\alpha = 1.364$. In this case, every pair of matrices has either approximately at least $0.4k$ common columns or approximately at most $0.28k$ common columns.

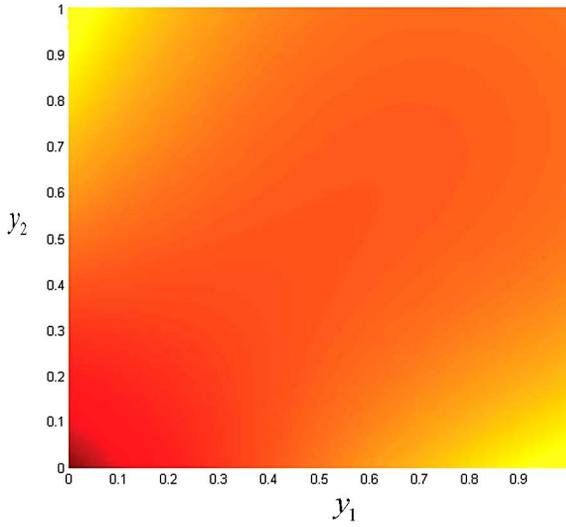


FIG. 3. $\mathcal{R}(\alpha)$ for $\alpha \in (0, \sqrt{3}/\sqrt{2})$.

We conjecture that the regime (c) described on Figure 5 corresponds to the hard on average case for which we predict that no polynomial time algorithm exists for nonconstant k . Since the OGP was analyzed based on overlaps of two matrices and the overlap of three matrices is likely to push the critical value of OGP even lower, we further conjecture that the hardness regime begins at a value strictly lower than our current estimate $5\sqrt{2}/(3\sqrt{3})$. An interesting open question is to conduct an overlap analysis of m -tuples of matrices and identify the critical value for the onset of disconnectedness.

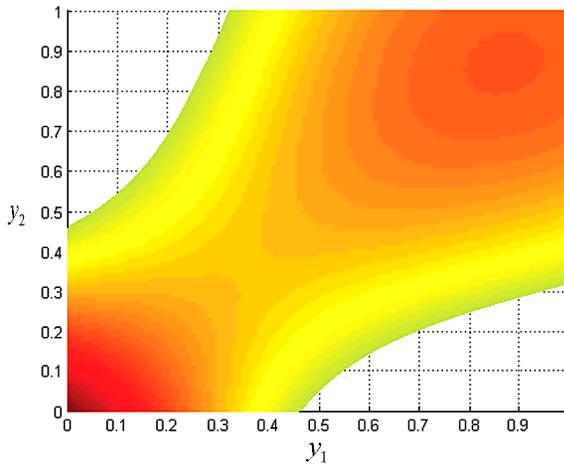


FIG. 4. $\mathcal{R}(\alpha)$ for $\alpha \in (\sqrt{3}/\sqrt{2}, 5\sqrt{2}/(3\sqrt{3}))$.

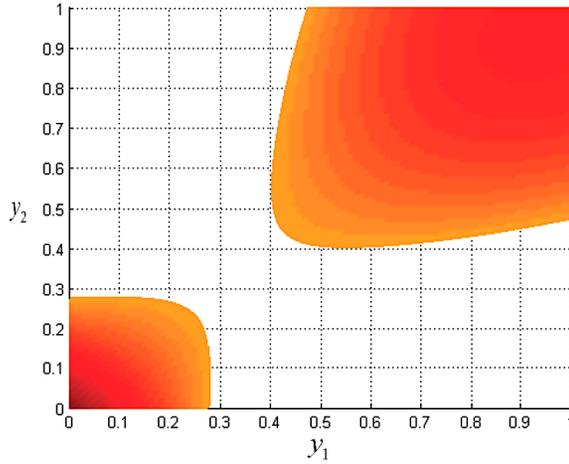


FIG. 5. $\mathcal{R}(\alpha)$ for $\alpha \in (5\sqrt{2}/(3\sqrt{3}), \sqrt{2})$.

3. Analysis of the \mathcal{IGP} algorithm. This section is devoted to the proof of Theorem 2.3. Denote by I_r^n the set of rows produced by \mathcal{IGP} algorithm in steps $2r$, $r = 0, 1, \dots, k - 1$ and by J_r^n the set of columns produced by \mathcal{IGP} algorithm in steps $2r - 1$, $r = 1, \dots, k$. Their cardinalities satisfy $|I_r^n| = r + 1$ for $r = 0, 1, \dots, k - 1$ and $|J_r^n| = r$ for $r = 1, \dots, k$. In particular, \mathcal{IGP} algorithm chooses $I_0^n = \{i_1\}$ arbitrarily from P_1^n and J_1^n is obtained by finding the column in \mathbf{C}_{i_1, P_1^n} corresponding to the largest entry. Let $M_i^n, i = 1, 2, \dots, 2k - 1$ be the entry sum of the row or column \mathcal{IGP} algorithm adds to the submatrix in the i th step,

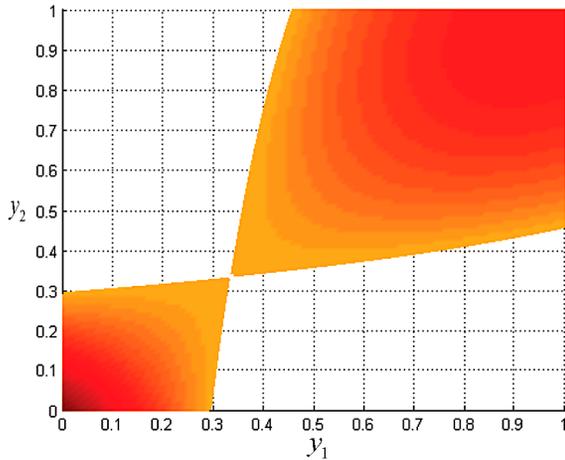


FIG. 6. $\mathcal{R}(5\sqrt{2}/(3\sqrt{3}))$.

namely

$$\begin{aligned}
 (7) \quad M_{2r-1}^n &\triangleq \max_{j \in P_{|I_{r-1}^n|}^n} \sum_{i \in I_{r-1}^n} C_{i,j} && \text{for } r = 1, 2, \dots, k, \\
 M_{2r}^n &\triangleq \max_{i \in P_{|J_r^n|+1}^n} \sum_{j \in J_r^n} C_{i,j} && \text{for } r = 1, 2, \dots, k-1.
 \end{aligned}$$

Introduce

$$(8) \quad b_n := \sqrt{2 \log n} - \frac{\log(4\pi \log n)}{2\sqrt{2 \log n}}.$$

In order to quantify $M_i^n, i = 1, 2, \dots, 2k - 1$, we now introduce a probabilistic bound on the maximum of n independent standard normal random variables.

LEMMA 3.1. *Let $Z_i, i = 1, 2, \dots, n$ be n independent i.i.d. standard normal random variables. There exists a positive integer N and a constant $c > 0$ such that for all $n > N$,*

$$\begin{aligned}
 (9) \quad &\mathbb{P}\left(\left|\sqrt{2 \log n} \left(\max_{1 \leq i \leq n} Z_i - b_n\right)\right| \leq \frac{3}{2} \log \log n\right) \\
 &\geq 1 - c \frac{1}{(\log n)^{1.5}}.
 \end{aligned}$$

Lemma 3.1 is a cruder version of the well-known fact described later in Section 5 as fact (21). For convenience, in what follows, we use n/k in place of $\lfloor n/k \rfloor$. We first establish Theorem 2.3 from the lemma above, the proof of which we delay to the Appendix A.

PROOF OF THEOREM 2.3. Denote by $E_{2r-1}^n, r = 1, 2, \dots, k$ the event that

$$(10) \quad \left|\sqrt{2 \log(n/k)} \left(\frac{M_{2r-1}^n}{\sqrt{|I_{r-1}^n|}} - b_{n/k}\right)\right| \leq \frac{3}{2} \log \log(n/k),$$

and by $E_{2r}^n, r = 1, 2, \dots, k - 1$ the event that

$$(11) \quad \left|\sqrt{2 \log(n/k)} \left(\frac{M_{2r}^n}{\sqrt{|J_r^n|}} - b_{n/k}\right)\right| \leq \frac{3}{2} \log \log(n/k).$$

By Lemma 3.1 and since $k \leq f(n) = o((\log n)^{1.5})$, we can choose a positive integer N_1 and $c > 0$ such that for all $n > N_1$,

$$(12) \quad \mathbb{P}(E_i^n) \geq 1 - c \frac{1}{(\log(n/k))^{1.5}}, \quad \forall 1 \leq i \leq 2k - 1.$$

Since $M_i^n, i = 1, 2, \dots, 2k - 1$ corresponds to nonoverlapping parts of C^n , they are mutually independent, and so are $E_i^n, i = 1, 2, \dots, 2k - 1$. Choose another positive integer N_2 such that for all $n > N_2$,

$$3 \frac{k}{(\log n)^{1.5}} \geq \frac{1}{(\log(n/k))^{1.5}}(2k - 1).$$

Let $N \triangleq \max(N_1, N_2)$. Then for all $n > N$ and $k \leq f(n) = o((\log n)^{1.5})$ we have

$$\begin{aligned} \mathbb{P}\left(\bigcap_{i=1}^{2k-1} E_i^n\right) &= \prod_{i=1}^{2k-1} \mathbb{P}(E_i^n) \\ &\geq \left(1 - c \frac{1}{(\log(n/k))^{1.5}}\right)^{2k-1} \\ &\geq 1 - c \frac{1}{(\log(n/k))^{1.5}}(2k - 1) \geq 1 - 3c \frac{k}{(\log n)^{1.5}}. \end{aligned}$$

As a result, $\bigcap_{i=1}^{2k-1} E_i^n$ occurs w.h.p.

Under the event $\bigcap_{i=1}^{2k-1} E_i^n$, we use (10) and (11) to estimate the average value of C_{IGP}^n :

$$\begin{aligned} \text{Ave}(C_{IGP}^n) &\leq \frac{1}{k^2} \left(\sum_{r=1}^k \left(\sqrt{|I_{r-1}^n|} b_{n/k} + \sqrt{|I_{r-1}^n|} \frac{\frac{3}{2} \log \log(n/k)}{\sqrt{2 \log(n/k)}} \right) \right. \\ &\quad \left. + \sum_{r=1}^{k-1} \left(\sqrt{|J_r^n|} b_{n/k} + \sqrt{|J_r^n|} \frac{\frac{3}{2} \log \log(n/k)}{\sqrt{2 \log(n/k)}} \right) \right) \\ &= \frac{\sum_{r=1}^k \sqrt{|I_{r-1}^n|} b_{n/k} + \sum_{r=1}^{k-1} \sqrt{|J_r^n|} b_{n/k}}{k^2} \\ &\quad + \frac{\sum_{r=1}^k \sqrt{|I_{r-1}^n|} + \sum_{r=1}^{k-1} \sqrt{|J_r^n|}}{k^2} \frac{\frac{3}{2} \log \log(n/k)}{\sqrt{2 \log(n/k)}}. \end{aligned}$$

Recall $|I_{r-1}^n| = r$ and $|J_r^n| = r$. We can choose a positive integer N_3 such that for all $n > N_3$ and $k \leq f(n) = o((\log n)^{1.5})$, $2 \log(n/k) \geq \log n$ holds. Also using $b_{n/k} < \sqrt{2 \log n}$ and $\log \log(n/k) < \log \log n$, the right-hand side of the last equation for all $n > N_3$,

$$\begin{aligned} &\leq \frac{\sum_{r=1}^k \sqrt{2 \log n} \sqrt{r} + \sum_{r=1}^{k-1} \sqrt{2 \log n} \sqrt{r}}{k^2} + \frac{\frac{3}{2} \log \log n}{\sqrt{\log n}} \\ &= 2 \sqrt{\frac{2 \log n}{k}} \sum_{r=1}^{k-1} \sqrt{\frac{r}{k}} + \frac{\sqrt{2 \log n}}{k^{3/2}} + \frac{\frac{3}{2} \log \log n}{\sqrt{\log n}} \end{aligned}$$

$$\begin{aligned} &\leq 2\sqrt{\frac{2\log n}{k}} \int_0^1 \sqrt{x} dx + 3 \max\left(\frac{1}{k}\sqrt{\frac{\log n}{k}}, \frac{\log \log n}{\sqrt{\log n}}\right) \\ &= \frac{4}{3}\sqrt{\frac{2\log n}{k}} + 3 \max\left(\frac{1}{k}\sqrt{\frac{\log n}{k}}, \frac{\log \log n}{\sqrt{\log n}}\right). \end{aligned}$$

Similarly, we establish that

$$\text{Ave}(\mathbf{C}_{\mathcal{IGP}}^n) \geq \frac{4}{3}\sqrt{\frac{2\log n}{k}} - 3 \max\left(\frac{1}{k}\sqrt{\frac{\log n}{k}}, \frac{\log \log n}{\sqrt{\log n}}\right).$$

Then (3) follows and the proof is completed. \square

4. The overlap gap property. We delay the derivation of the critical values for the two phase transition points $\alpha_1^* = \sqrt{3}/\sqrt{2}$ and $\alpha_2^* = 5\sqrt{2}/(3\sqrt{3})$ to Appendix B. Now we complete the proof of Theorem 2.4.

PROOF OF THEOREM 2.4. The rest of the section is devoted to establishing part (5) of Theorem 2.4. The second result (6) follows from the Markov inequality.

Fix positive integers k_1, k_2, k and n such that $0 < k_1 \leq k \leq n$ and $0 < k_2 \leq k \leq n$. Let X, Y_1 and Y_2 be three mutually independent normal random variables: $X \stackrel{d}{=} \mathcal{N}(0, k_1 k_2)$ and $Y_1 \stackrel{d}{=} Y_2 \stackrel{d}{=} \mathcal{N}(0, k^2 - k_1 k_2)$. Recall the definition of the multinomial coefficient at the end of the Introduction. Then

$$\begin{aligned} &\mathbb{E}(|\mathcal{O}_k(\alpha, y_1, y_2, \delta)|) \\ &= \sum_{\substack{k_1 \in ((y_1 - \delta)k, (y_1 + \delta)k) \\ k_2 \in ((y_2 - \delta)k, (y_2 + \delta)k)}} \binom{n}{k - k_1, k_1, k - k_1, n - 2k + k_1} \\ (13) \quad &\times \binom{n}{k - k_2, k_2, k - k_2, n - 2k + k_2} \\ &\times \mathbb{P}\left(X + Y_1, X + Y_2 \in \left[(\alpha - \delta)\sqrt{\frac{2\log n}{k}}, (\alpha + \delta)k^2 \sqrt{\frac{2\log n}{k}}\right]\right). \end{aligned}$$

For the rest of the proof, we will first estimate the last term in (13), then estimate the first two combinatorial terms in (13), and finally compute $\mathbb{E}(|\mathcal{O}_k(\alpha, y_1, y_2, \delta)|)$ by combining the two estimation results.

We estimate the last term in (13) for the cases $\min(k_1, k_2) < k$ and $k_1 = k_2 = k$, separately. Consider the case $\min(k_1, k_2) < k$. We let $\tau \triangleq (\alpha - \delta) \times \sqrt{2k_1 k_2 / (k^2 + k_1 k_2)}$ and write

$$\mathbb{P}\left(X + Y_1, X + Y_2 \in \left[(\alpha - \delta)k^2 \sqrt{\frac{2\log n}{k}}, (\alpha + \delta)k^2 \sqrt{\frac{2\log n}{k}}\right]\right) = I_1 + I_2,$$

where

$$\begin{aligned}
 I_1 &= \int_{-\infty}^{\tau k^2 \sqrt{\frac{2 \log n}{k}}} \mathbb{P}\left((\alpha + \delta)k^2 \sqrt{\frac{2 \log n}{k}} - x \geq Y_1 \geq (\alpha - \delta)k^2 \sqrt{\frac{2 \log n}{k}} - x\right)^2 \\
 &\quad \times \frac{1}{\sqrt{2\pi k_1 k_2}} \exp\left(-\frac{x^2}{2k_1 k_2}\right) dx, \\
 I_2 &= \int_{\tau k^2 \sqrt{\frac{2 \log n}{k}}}^{\infty} \mathbb{P}\left((\alpha + \delta)k^2 \sqrt{\frac{2 \log n}{k}} - x \geq Y_1 \geq (\alpha - \delta)k^2 \sqrt{\frac{2 \log n}{k}} - x\right)^2 \\
 &\quad \times \frac{1}{\sqrt{2\pi k_1 k_2}} \exp\left(-\frac{x^2}{2k_1 k_2}\right) dx.
 \end{aligned}$$

Now we estimate I_1 . Let

$$u(x) = \frac{(\alpha - \delta)k^2 \sqrt{\frac{2 \log n}{k}} - x}{\sqrt{k^2 - k_1 k_2}}.$$

We claim that for $x \leq \tau k^2 \sqrt{\frac{2 \log n}{k}}$, $u(x)$ diverges to infinity as $n \rightarrow \infty$. Namely, $\lim_{n \rightarrow \infty} \min u(x) = \infty$, where the minimum is over $x \leq \tau k^2 \sqrt{\frac{2 \log n}{k}}$. We have

$$\begin{aligned}
 u(x) &\geq \frac{(\alpha - \delta - \tau)k^2 \sqrt{\frac{2 \log n}{k}}}{\sqrt{k^2 - k_1 k_2}} \\
 &= \frac{1 - \sqrt{2k_1 k_2 / (k^2 + k_1 k_2)}}{\sqrt{k^2 - k_1 k_2}} (\alpha - \delta)k^2 \sqrt{\frac{2 \log n}{k}} \\
 &= \frac{1 - \sqrt{1 - (k^2 - k_1 k_2) / (k^2 + k_1 k_2)}}{\sqrt{k^2 - k_1 k_2}} (\alpha - \delta)k^2 \sqrt{\frac{2 \log n}{k}}.
 \end{aligned}$$

Using the fact $\sqrt{1 - a} \leq 1 - a/2$ for $a = (k^2 - k_1 k_2) / (k^2 + k_1 k_2) \in [0, 1]$, we have that the expression above is at least

$$\begin{aligned}
 \frac{\sqrt{k^2 - k_1 k_2}}{2(k^2 + k_1 k_2)} (\alpha - \delta)k^2 \sqrt{\frac{2 \log n}{k}} &\geq \frac{\sqrt{k^2 - k(k - 1)}}{4k^2} (\alpha - \delta)k^2 \sqrt{\frac{2 \log n}{k}} \\
 &= \frac{\alpha - \delta}{4} \sqrt{2 \log n},
 \end{aligned}$$

and the claim is verified. Then using (1) to approximate the first term in the integrand of I_1 , we further divide I_1 into two parts as follows:

$$\frac{1}{k \log n} \log I_1 = o(1) + \frac{1}{k \log n} \log(I_{11} + I_{12}),$$

where

$$\begin{aligned}
 I_{11} &= \int_{-k^2(\log n)^{2/3}}^{\tau k^2 \sqrt{\frac{2 \log n}{k}}} \frac{1}{2\pi u(x)^2} \frac{1}{\sqrt{2\pi k_1 k_2}} \\
 &\quad \times \exp\left(-\frac{((\alpha - \delta)k^2 \sqrt{2 \log n/k} - x)^2}{2(k^2 - k_1 k_2)} \times 2 - \frac{x^2}{2k_1 k_2}\right) dx, \\
 I_{12} &= \int_{-\infty}^{-k^2(\log n)^{2/3}} \frac{1}{2\pi u(x)^2} \frac{1}{\sqrt{2\pi k_1 k_2}} \\
 &\quad \times \exp\left(-\frac{((\alpha - \delta)k^2 \sqrt{2 \log n/k} - x)^2}{2(k^2 - k_1 k_2)} \times 2 - \frac{x^2}{2k_1 k_2}\right) dx.
 \end{aligned}$$

Since for any $x \in [-k^2(\log n)^{2/3}, \tau k^2 \sqrt{\frac{2 \log n}{k}}]$, we have

$$\frac{1}{k \log n} \log(u(x)^2) = o(1),$$

then

$$\begin{aligned}
 &\frac{1}{k \log n} \log I_{11} \\
 &= o(1) + \frac{1}{k \log n} \log \int_{-k^2(\log n)^{2/3}}^{\tau k^2 \sqrt{\frac{2 \log n}{k}}} \frac{1}{\sqrt{2\pi k_1 k_2}} \\
 &\quad \times \exp\left(-\frac{((\alpha - \delta)k^2 \sqrt{2 \log n/k} - x)^2}{2(k^2 - k_1 k_2)} \times 2 - \frac{x^2}{2k_1 k_2}\right) dx.
 \end{aligned}$$

Then we rewrite the integrand in the equation above in terms of a density function of a normal random variable

$$\begin{aligned}
 (14) \quad &\frac{1}{k \log n} \log I_{11} = o(1) - 2(\alpha - \delta)^2 \frac{k^2}{k^2 + k_1 k_2} \\
 &+ \frac{1}{k \log n} \log \int_{-k^2(\log n)^{2/3}}^{\tau k^2 \sqrt{\frac{2 \log n}{k}}} \frac{1}{\sqrt{2\pi \frac{k_1 k_2 (k^2 - k_1 k_2)}{k^2 + k_1 k_2}}} \\
 &\quad \times \exp\left(-\frac{(x - \frac{2k_1 k_2 k^2 (\alpha - \delta) \sqrt{2 \log n/k}}{k^2 + k_1 k_2})^2}{2 \frac{k_1 k_2 (k^2 - k_1 k_2)}{k^2 + k_1 k_2}}\right) dx.
 \end{aligned}$$

It follows from $\tau = (\alpha - \delta) \sqrt{2k_1 k_2 / (k^2 + k_1 k_2)}$ and $\sqrt{a} > a$ for $a \in (0, 1)$ that

$$(15) \quad \tau k^2 \sqrt{\frac{2 \log n}{k}} \geq \frac{2k_1 k_2 k^2 (\alpha - \delta) \sqrt{2 \log n/k}}{k^2 + k_1 k_2}.$$

Also we have as $n \rightarrow \infty$

$$(16) \quad \frac{-k^2(\log n)^{2/3} - \frac{2k_1k_2k^2(\alpha-\delta)\sqrt{2\log n/k}}{k^2+k_1k_2}}{\sqrt{\frac{k_1k_2(k^2-k_1k_2)}{k^2+k_1k_2}}} \rightarrow -\infty.$$

Since the integrand in (14) is a density function of a normal random variable, (15) and (16) implies that the integral in (14) is in $[1/2 + o(1), 1]$. The last term in (14) is $o(1)$, and thus

$$\frac{1}{k \log n} \log I_{11} = o(1) - 2(\alpha - \delta)^2 \frac{k^2}{k^2 + k_1k_2}.$$

Also we have

$$\frac{1}{k \log n} \log I_{12} \leq \frac{1}{k \log n} \log \int_{-\infty}^{-k^2(\log n)^{2/3}} \exp\left(-\frac{x^2}{2k_1k_2}\right) dx,$$

where the right-hand side goes to $-\infty$ as $n \rightarrow \infty$.

Now we estimate I_2 . We have

$$\begin{aligned} \frac{1}{k \log n} \log I_2 &\leq \frac{1}{k \log n} \log \int_{\tau k^2 \sqrt{\frac{2\log n}{k}}}^{\infty} \exp\left(-\frac{x^2}{2k_1k_2}\right) dx \\ &= o(1) - \tau^2 \frac{k^2}{k_1k_2} \\ &= o(1) - 2(\alpha - \delta)^2 \frac{k^2}{k^2 + k_1k_2}. \end{aligned}$$

Using $\log(\max(a, b)) \leq \log(a + b) \leq \log(2 \max(a, b))$ for $a, b > 0$, we conclude

$$(17) \quad \begin{aligned} &\frac{1}{k \log n} \log \mathbb{P}\left(X + Y_1, \right. \\ &\quad \left. X + Y_2 \in \left[(\alpha - \delta)k^2 \sqrt{\frac{2\log n}{k}}, (\alpha + \delta)k^2 \sqrt{\frac{2\log n}{k}} \right] \right) \\ &= \frac{1}{k \log n} \log(I_1 + I_2) \\ &= o(1) + \frac{1}{k \log n} \max(\log I_1, \log I_2) \\ &= o(1) + \frac{1}{k \log n} \max(\log I_{11}, \log I_{12}, \log I_2) \\ &= o(1) - 2(\alpha - \delta)^2 \frac{k^2}{k^2 + k_1k_2}. \end{aligned}$$

Consider the case $k_1 = k_2 = k$. Observing $Y_1 = Y_2 = 0$ and using (1), we obtain

$$\begin{aligned} & \frac{1}{k \log n} \log \mathbb{P} \left(X + Y_1, X + Y_2 \in \left[(\alpha - \delta)k^2 \sqrt{\frac{2 \log n}{k}}, (\alpha + \delta)k^2 \sqrt{\frac{2 \log n}{k}} \right] \right) \\ &= \frac{1}{k \log n} \log \mathbb{P} \left(X \in \left[(\alpha - \delta)k^2 \sqrt{\frac{2 \log n}{k}}, (\alpha + \delta)k^2 \sqrt{\frac{2 \log n}{k}} \right] \right) \\ &= o(1) - (\alpha - \delta)^2. \end{aligned}$$

Hence equation (17) still holds for the case $k_1 = k_2 = k$.

Now we estimate the first two terms in (13). It is for this part that the assumption $k \leq c \log n$ will be used. Let $\beta_1 \triangleq k_1/k$ and $\beta_2 \triangleq k_2/k$. Using Stirling's approximation $a! \approx \sqrt{2\pi a}(a/e)^a$, $(n - b) \log(n - b) = (n - b) \log n - b(1 + o(1))$ for $b = O(\log n)$ and $k \leq c \log n$, taking log of the first two terms in the right-hand side of (13) gives

$$\begin{aligned} & \log \left(\frac{n!}{(k - k_1)!k_1!(k - k_1)!(n - 2k + k_1)!} \frac{n!}{(k - k_2)!k_2!(k - k_2)!(n - 2k + k_2)!} \right) \\ &= O(1) + \log \left(\sqrt{2\pi n} n^n / (2\pi(k - k_1)(k - k_1)^{2(k-k_1)} \sqrt{2\pi k_1} k_1^{k_1} \right. \\ & \quad \times \sqrt{2\pi(n - 2k + k_1)}(n - 2k + k_1)^{n-2k+k_1}) \\ & \quad \times \sqrt{2\pi n} n^n / (2\pi(k - k_2)(k - k_2)^{2(k-k_2)} \sqrt{2\pi k_2} k_2^{k_2} \\ & \quad \times \sqrt{2\pi(n - 2k + k_2)}(n - 2k + k_2)^{n-2k+k_2}) \Big) \\ &= O(1) + (\log n + 2n \log n) - (\log(k - k_1) + 2(k - k_1) \log(k - k_1)) \\ & \quad - \left(\frac{1}{2} \log k_1 + k_1 \log k_1 \right) \\ (18) \quad & - \left(\frac{1}{2} \log(n - 2k + k_1) + (n - 2k + k_1) \log(n - 2k + k_1) \right) \\ & \quad - (\log(k - k_2) + 2(k - k_2) \log(k - k_2)) \\ & \quad - \left(\frac{1}{2} \log k_2 + k_2 \log k_2 \right) \\ & \quad - \left(\frac{1}{2} \log(n - 2k + k_2) + (n - 2k + k_2) \log(n - 2k + k_2) \right) \\ &= o(1)k \log n + (\log n + 2n \log n) \\ & \quad - \left(\frac{1}{2} \log(n - 2k + k_1) + (n - 2k + k_1) \log(n - 2k + k_1) \right) \\ & \quad - \left(\frac{1}{2} \log(n - 2k + k_2) + (n - 2k + k_2) \log(n - 2k + k_2) \right) \\ &= (4 - \beta_1 - \beta_2 + o(1))k \log n. \end{aligned}$$

Then it follows from (18) and (17) that

$$\begin{aligned} & \frac{1}{k \log n} \log \mathbb{E}(|\mathcal{O}_k(\alpha, y_1, y_2, \delta)|) \\ &= \sup_{\substack{\beta_1 \in (y_1 - \delta, y_1 + \delta) \\ \beta_2 \in (y_2 - \delta, y_2 + \delta)}} 4 - \beta_1 - \beta_2 - \frac{2}{1 + \beta_1 \beta_2} (\alpha - \delta)^2 + o(1) \\ &= \sup_{\substack{\beta_1 \in (y_1 - \delta, y_1 + \delta) \\ \beta_2 \in (y_2 - \delta, y_2 + \delta)}} f(\alpha - \delta, \beta_1, \beta_2) + o(1), \end{aligned}$$

where the region of (β_1, β_2) for the sup above comes from range of the sum in (13). Then (5) follows from the continuity of $f(\alpha, y_1, y_2)$. This completes the proof of Theorem 2.4. \square

5. Analysis of the \mathcal{LAS} algorithm.

5.1. *Preliminary results.* We denote by I_r^n the set of rows produced by the \mathcal{LAS} algorithm in iterations $2r$, $r = 0, 1, \dots$ and by J_r^n the set of columns produced by \mathcal{LAS} in iterations $2r - 1$, $r = 1, 2, \dots$. Without the loss of generality, we set $I_0 = \{1, \dots, k\}$. Then J_1 is obtained by searching the k columns with the largest sum of entries in the submatrix $\mathbf{C}^{k \times n}$. Furthermore, $\mathbf{C}_{2r+1}^n = \mathbf{C}_{I_r^n, J_{r+1}^n}^n$, $r \geq 0$ and $\mathbf{C}_{2r}^n = \mathbf{C}_{I_r^n, J_r^n}^n$, $r \geq 1$.

Next, for every $r \geq 1$, denote by \tilde{J}_r^n the set of k columns with the largest sum of entries in the $k \times (n - k)$ matrix $\mathbf{C}_{I_r^n, [n] \setminus J_r^n}$. In particular, in iteration $2r + 1$ the algorithm chooses the best k columns J_{r+1}^n (k columns with the largest entry sums) from the $2k$ columns, the k of which are the columns of $\mathbf{C}_{I_r^n, J_r^n}$, and the remaining k of which are columns of $\mathbf{C}_{I_r^n, \tilde{J}_r^n}$. Similarly, for $r \geq 0$ we define \tilde{I}_r^n to be the set of k rows with the largest sum of entries in the $(n - k) \times k$ matrix $\mathbf{C}_{[n] \setminus I_r^n, J_{r+1}^n}$.

The following definition was introduced in [6].

DEFINITION 5.1. Let I be a set of k rows and J be a set of k columns in \mathbf{C}^n . The submatrix $[\mathbf{C}_{ij}^n]_{i \in I, j \in J}$ is defined to be row dominant in \mathbf{C}^n if

$$\min_{i \in I} \left\{ \sum_{j \in J} \mathbf{C}_{ij}^n \right\} \geq \max_{i \in [n] \setminus I} \left\{ \sum_{j \in J} \mathbf{C}_{ij}^n \right\},$$

and is column dominant in \mathbf{C}^n if

$$\min_{j \in J} \left\{ \sum_{i \in I} \mathbf{C}_{ij}^n \right\} \geq \max_{j \in [n] \setminus J} \left\{ \sum_{i \in I} \mathbf{C}_{ij}^n \right\}.$$

A submatrix which is both row dominant and column dominant is called a locally maximum submatrix.

From the definition above, the $k \times k$ submatrix that \mathcal{LAS} returns in each iteration is either row dominant or column dominant, and the final submatrix the \mathcal{LAS} converges to is a locally maximum submatrix.

We now recall the Analysis of Variance (ANOVA) Decomposition of a matrix. Given any $k \times k$ matrix B , let $B_{i\cdot}$ be the average of the i th row, $B_{\cdot j}$ be the average of the j th column, and $B_{\cdot\cdot} := \text{avg}(B)$ be the average of the matrix B . Then the ANOVA decomposition $\text{ANOVA}(B)$ of the matrix B is defined as

$$(19) \quad \text{ANOVA}(B)_{ij} = B_{ij} - B_{i\cdot} - B_{\cdot j} + B_{\cdot\cdot}, \quad 1 \leq i, j \leq k.$$

The matrix B can then be rewritten as

$$(20) \quad B = \text{avg}(B)\mathbf{1}\mathbf{1}' + \text{Row}(B) + \text{Col}(B) + \text{ANOVA}(B),$$

where $\text{Row}(B)$ denotes the matrix with the i th row entries all equal to $B_{i\cdot} - B_{\cdot\cdot}$ for all $1 \leq i \leq k$, and similarly $\text{Col}(B)$ denotes the matrix with the j th column entries all equal to $B_{\cdot j} - B_{\cdot\cdot}$ for all $1 \leq j \leq k$. An essential property of ANOVA decomposition is that, if B consists of independent standard Gaussian variables, the random variables and matrices $B_{\cdot\cdot}$, $\text{Row}(B)$, $\text{Col}(B)$ and $\text{ANOVA}(B)$ are independent. This property is easily verified by establishing that the corresponding covariances are zero.

Recall the definition of b_n in (8). Let L_n be the maximum of n independent standard normal random variables. It is known that [13]

$$(21) \quad \sqrt{2 \log n}(L_n - b_n) \Rightarrow -\log G,$$

as $n \rightarrow \infty$, where G is an exponential random variable with parameter 1.

Let (S_1, S_2) be a pair of positive random variables with joint density

$$(22) \quad f(s_1, s_2) = C(\log(1 + s_2/s_1))^{k-1} s_1^{k-1} e^{-(s_1+s_2)},$$

where C is the normalizing constant to make $f(s_1, s_2)$ a density function. Let $\mathbf{U} = (U_1, \dots, U_k)$ be a random vector with the Dirichlet distribution with parameter 1. Namely, \mathbf{U} is uniformly distributed on the simplex $\{(x_1, \dots, x_k) \mid \sum_{i=1}^k x_i = 1, x_i \geq 0, 1 \leq i \leq k\}$. Let

$$\mathbf{C}_\infty^{\text{Row}} \triangleq (-\log G, \log(1 + S_1/S_2)(k\mathbf{U} - 1)\mathbf{1}^T, \text{Col}(\mathbf{C}^k), \text{ANOVA}(\mathbf{C}^k)),$$

and

$$\mathbf{C}_\infty^{\text{Col}} \triangleq (-\log G, \text{Row}(\mathbf{C}^k), \log(1 + S_1/S_2)\mathbf{1}(k\mathbf{U} - 1)^T, \text{ANOVA}(\mathbf{C}^k)),$$

where $G, (S_1, S_2), \mathbf{U}$ are independent and distributed as above, and as before \mathbf{C}^k is a $k \times k$ matrix of i.i.d. standard normal random variables independent from $G, (S_1, S_2), \mathbf{U}$.

Denote by \mathcal{RD}_n the event that the matrix \mathbf{C}^k (the top $k \times k$ matrix of \mathbf{C}^n) is row dominant. Similarly, denote by \mathcal{CD}_n the event that the same matrix is column

dominant. Let $\mathbf{D}_{\text{row}}^n$ be a random $k \times k$ matrix distributed as \mathbf{C}^k conditioned on the event \mathcal{RD}_n . Similarly, define $\mathbf{D}_{\text{col}}^n$.

Introduce the following two operators acting on $k \times k$ matrices A :

$$(23) \quad \Psi_n^{\text{Row}}(A) \triangleq (\sqrt{2 \log n}(\sqrt{k} \text{ave}(A) - b_n), \sqrt{2k \log n} \text{Row}(A), \text{Col}(A), \text{ANOVA}(A)) \in \mathbb{R} \times (\mathbb{R}^{k \times k})^3,$$

$$(24) \quad \Psi_n^{\text{Col}}(A) \triangleq (\sqrt{2 \log n}(\sqrt{k} \text{ave}(A) - b_n), \text{Row}(A), \sqrt{2k \log n} \text{Col}(A), \text{ANOVA}(A)) \in \mathbb{R} \times (\mathbb{R}^{k \times k})^3.$$

As a result, writing $\Psi_n^{\text{Row}}(A) = (\Psi_{n,j}^{\text{Row}}(A), 1 \leq j \leq 4)$ and applying (20), we have

$$(25) \quad A = \left(\frac{\Psi_{n,1}^{\text{Row}}(A)}{\sqrt{2k \log n}} + \frac{b_n}{\sqrt{k}} \right) \mathbf{1}\mathbf{1}' + \frac{\Psi_{n,2}^{\text{Row}}(A)}{\sqrt{2k \log n}} + \Psi_{n,3}^{\text{Row}}(A) + \Psi_{n,4}^{\text{Row}}(A).$$

A similar expression holds for A in terms of $\Psi_n^{\text{Col}}(A)$.

Bhamidi, Dey and Nobel ([6]) established the limiting distribution result for locally maximum submatrix. For row (column) dominant submatrix, the following result can be easily derived following a similar proof.

THEOREM 5.2. *For every $k > 0$, the following convergence in distribution takes place as $n \rightarrow \infty$:*

$$(26) \quad \Psi_n^{\text{Row}}(\mathbf{D}_{\text{row}}^n) \Rightarrow \mathbf{C}_{\infty}^{\text{Row}}.$$

Similarly,

$$(27) \quad \Psi_n^{\text{Col}}(\mathbf{D}_{\text{col}}^n) \Rightarrow \mathbf{C}_{\infty}^{\text{Col}}.$$

Applying ANOVA decomposition (20), the result can be interpreted loosely as follows. $\mathbf{D}_{\text{row}}^n$ is approximately

$$\mathbf{D}_{\text{row}}^n \approx \sqrt{\frac{2 \log n}{k}} \mathbf{1}\mathbf{1}' + \text{Col}(\mathbf{C}^k) + \text{ANOVA}(\mathbf{C}^k) + o\left(\frac{\log \log n}{\sqrt{\log n}}\right).$$

Indeed the first component of convergence (26) means

$$\text{avg}(\mathbf{D}_{\text{row}}^n) \approx \frac{b_n}{\sqrt{k}} - \frac{\log G}{\sqrt{2k \log n}} = \sqrt{\frac{2 \log n}{k}} + o\left(\frac{\log \log n}{\sqrt{\log n}}\right),$$

and the second component of the same convergence means

$$\text{Row}(\mathbf{D}_{\text{row}}^n) = o\left(\frac{1}{\sqrt{\log n}}\right).$$

5.2. *Conditional distribution of the row-dominant and column-dominant submatrices.* Our next goal is to establish a conditional version of Theorem 5.2. We begin with several preliminary steps.

LEMMA 5.3. *Fix a sequence Z_1, \dots, Z_n of i.i.d. standard normal random variables and r distinct subsets $I_1, \dots, I_r \subset [n], |I_\ell| = k, 1 \leq \ell \leq r$. Let $Y_\ell = k^{-\frac{1}{2}} \sum_{i \in I_\ell} Z_i$. Then there exists a lower triangular matrix*

$$(28) \quad L = \begin{pmatrix} L_{1,1} & 0 & 0 & \cdots & 0 \\ L_{2,1} & L_{2,2} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ L_{r,1} & L_{r,2} & L_{r,3} & \cdots & L_{r,r} \end{pmatrix},$$

such that

- (a) $(Y_1, \dots, Y_r)^T$ equals in distribution $L(Y_1, W_2, \dots, W_r)^T$, where W_2, \dots, W_r are i.i.d. standard normal random variables independent from Y_1 .
- (b) The values $L_{i,j}$ are determined completely by the cardinalities of the intersections $I_{\ell_1} \cap I_{\ell_2}, 1 \leq \ell_1, \ell_2 \leq k$.
- (c) $L_{i,1} \in \{0, 1/k, \dots, (k-1)/k, 1\}$ for all i , with $L_{1,1} = 1$, and $L_{i,1} \leq (k-1)/k$, for all $i = 2, \dots, r$,
- (d) $\sum_{1 \leq i \leq r} L_{\ell,i}^2 = 1$ for each $\ell = 1, \dots, r$.

Note that Y_1, \dots, Y_r are correlated standard normal random variables. The lemma effectively provides a representation of these variables as a linear operator acting on a vector of independent standard normal random variables, where since by condition (c) we have $L_{1,1} = 1$, the first component Y_1 is preserved.

PROOF. Let Σ be the covariance matrix of (Y_1, \dots, Y_r) and let $\Sigma = LL^T$ be its Cholesky factorization. We claim that L has the required property. Note that the elements of Σ are completely determined by the cardinalities of intersections $I_\ell \cap I_{\ell'}, 1 \leq \ell, \ell' \leq r$, and thus (b) holds. Since Σ is the covariance matrix of (Y_1, \dots, Y_r) , we obtain that this vector equals in distribution $L(W_1, \dots, W_r)^T$, where $W_i, 1 \leq i \leq r$ are i.i.d. standard normal random variables, and thus (a) holds. We can take W_1 to be Y_1 since Y_1 is also a standard normal random variable. Note that $L_{1,1}$ is the variance of Y_1 hence $L_{1,1} = 1$. The variance of Y_ℓ is $\sum_{1 \leq i \leq r} L_{\ell,i}^2$ which equals 1 since Y_ℓ is also standard normal, namely (d) holds. Finally, note that $L_{i,1}$ is the covariance of Y_1 with $Y_i, i = 2, \dots, r$, which takes one of the values $0, 1/k, \dots, (k-1)/k$, since I_ℓ are distinct subsets of $[n]$ with cardinality k . This establishes (c). \square

Recall that ω_n denotes any strictly increasing positive function satisfying $\omega_n = o(\sqrt{2 \log n})$ and $\log \log n = O(\omega_n)$. We now establish the following conditional version of (21).

LEMMA 5.4. Fix a positive integer $r \geq 2$ and $r \times r$ lower triangular matrix L satisfying $|L_{\ell,i}| \leq 1$ and $L_{\ell,1} \leq (k - 1)/k, \ell = 2, \dots, r$. Let $\mathbf{Z} = (Z_{i,\ell}, 1 \leq i \leq n, 1 \leq \ell \leq r)$ be a matrix of i.i.d. standard normal random variables. Given any $\bar{c} = (c_\ell, 1 \leq \ell \leq r - 1) \in \mathbb{R}^{r-1}$, for each $i = 1, \dots, n$, let $\mathcal{B}_i = \mathcal{B}_i(\bar{c})$ denote the event

$$[L(Z_{i,1}, Z_{i,2}, \dots, Z_{i,r})^T]_\ell \leq \sqrt{2 \log n} + c_{\ell-1}, \quad \forall 2 \leq \ell \leq r,$$

where $[\cdot]_\ell$ denotes the ℓ th component of the vector in the argument. Then for every $w \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \sup_{\bar{c}: \|\bar{c}\|_\infty \leq \omega_n} \left| \mathbb{P} \left(\sqrt{2 \log n} \left(\max_{1 \leq i \leq n} Z_{i,1} - b_n \right) \leq w \mid \bigcap_{1 \leq i \leq n} \mathcal{B}_i \right) - \exp(-\exp(-w)) \right| = 0.$$

Namely, the events \mathcal{B}_i have an asymptotically negligible effect on the weak convergence fact (21), namely that

$$\sqrt{2 \log n} \left(\max_{1 \leq i \leq n} Z_{i,1} - b_n \right) \Rightarrow -\log G.$$

PROOF. Note that the events $\mathcal{B}_i, 1 \leq i \leq n$ are independent. Thus we rewrite

$$\begin{aligned} & \mathbb{P} \left(\sqrt{2 \log n} \left(\max_{1 \leq i \leq n} Z_{i,1} - b_n \right) \leq w \mid \bigcap_{1 \leq i \leq n} \mathcal{B}_i \right) \\ &= \mathbb{P} \left(\max_{1 \leq i \leq n} Z_{i,1} \leq b_n + \frac{w}{\sqrt{2 \log n}} \mid \bigcap_{1 \leq i \leq n} \mathcal{B}_i \right) \\ (29) \quad &= \mathbb{P}(Z_{1,1} \leq b_n + w/\sqrt{2 \log n} \mid \mathcal{B}_1)^n \\ &= \left(1 - \frac{\mathbb{P}(Z_{1,1} > b_n + w/\sqrt{2 \log n}, \mathcal{B}_1)}{\mathbb{P}(\mathcal{B}_1)} \right)^n. \end{aligned}$$

Fix any $\delta_1, \delta_2 \in (0, 1/(2k))$. Let $\tilde{\mathcal{B}}_1 = \tilde{\mathcal{B}}_1(\delta_1, \delta_2)$ be the event that

$$Z_{1,1} \leq (1 + \delta_2)b_n \quad \text{and} \quad |Z_{1,\ell}| \leq \frac{\delta_1}{r-1} b_n, \quad \forall 2 \leq \ell \leq r.$$

We claim that $\tilde{\mathcal{B}}_1 \subset \mathcal{B}_1$ for all large enough n and any \bar{c} satisfying $\|\bar{c}\|_\infty \leq \omega_n$. Indeed, using $L_{\ell,1} \leq (k - 1)/k$ and $|L_{\ell,i}| \leq 1, \ell = 2, \dots, r$, the event $\tilde{\mathcal{B}}_1$ implies

$$L_{\ell,1}Z_{1,1} + \sum_{i=2}^{\ell} L_{\ell,i}Z_{1,i} \leq (1 - 1/k)(1 + \delta_2)b_n + \delta_1 b_n, \quad \forall 2 \leq \ell \leq r.$$

Then using $\delta_j \in (0, \frac{1}{2k}), \forall j = 1, 2$, we can choose sufficiently large n such that for any \bar{c} satisfying $\|\bar{c}\|_\infty \leq \omega_n$,

$$(1 - 1/k)(1 + \delta_2)b_n + \delta_1 b_n \leq \sqrt{2 \log n} + c_{\ell-1}, \quad \forall 2 \leq \ell \leq r,$$

from which the claim follows. Then we have

$$\begin{aligned} & 1 - \mathbb{P}(Z_{1,1} > b_n + w/\sqrt{2 \log n}, \tilde{\mathcal{B}}_1) \\ (30) \quad & \geq 1 - \frac{\mathbb{P}(Z_{1,1} > b_n + w/\sqrt{2 \log n}, \mathcal{B}_1)}{\mathbb{P}(\mathcal{B}_1)} \\ & \geq 1 - \frac{\mathbb{P}(Z_{1,1} > b_n + w/\sqrt{2 \log n})}{\mathbb{P}(\tilde{\mathcal{B}}_1)}. \end{aligned}$$

Using (1), we simplify

$$\begin{aligned} & \mathbb{P}(Z_{1,1} > b_n + w/\sqrt{2 \log n}, \tilde{\mathcal{B}}_1) \\ & = \mathbb{P}((1 + \delta_2)b_n \geq Z_{1,1} > b_n + w/\sqrt{2 \log n}) \mathbb{P}\left(|Z_{1,\ell}| \leq \frac{\delta_1}{r-1} b_n\right)^{r-1} \\ & = \frac{1}{(b_n + w/\sqrt{2 \log n})\sqrt{2\pi}} \exp\left(-\frac{(b_n + w/\sqrt{2 \log n})^2}{2}\right) (1 + o(1)). \end{aligned}$$

Also using $\lim_{n \rightarrow \infty} \mathbb{P}(\tilde{\mathcal{B}}_1) = 1$, we simplify

$$\begin{aligned} & \frac{\mathbb{P}(Z_{1,1} > b_n + w/\sqrt{2 \log n})}{\mathbb{P}(\tilde{\mathcal{B}}_1)} \\ & = \frac{1}{(b_n + w/\sqrt{2 \log n})\sqrt{2\pi}} \exp\left(-\frac{(b_n + w/\sqrt{2 \log n})^2}{2}\right) (1 + o(1)). \end{aligned}$$

The two equations above give the same asymptotics of the two sides in (30). Hence the term in the middle also has the same asymptotics

$$\begin{aligned} & 1 - \frac{\mathbb{P}(Z_{1,1} > b_n + w/\sqrt{2 \log n}, \mathcal{B}_1)}{\mathbb{P}(\mathcal{B}_1)} \\ (31) \quad & = 1 - \frac{1}{(b_n + w/\sqrt{2 \log n})\sqrt{2\pi}} \\ & \quad \times \exp\left(-\frac{(b_n + w/\sqrt{2 \log n})^2}{2}\right) (1 + o(1)) \\ & = 1 - \mathbb{P}(Z_{1,1} > b_n + w/\sqrt{2 \log n}) (1 + o(1)). \end{aligned}$$

Substituting (31) into (29), we have for any \bar{c} satisfying $\|\bar{c}\|_\infty \leq \omega_n$

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P}\left(\sqrt{2 \log n} \left(\max_{1 \leq i \leq n} Z_{i,1} - b_n\right) \leq w \mid \bigcap_{1 \leq i \leq n} \mathcal{B}_i\right) \\ &= \lim_{n \rightarrow \infty} \left(1 - \mathbb{P}(Z_{1,1} > b_n + w/\sqrt{2 \log n})\right)^n \\ &= \lim_{n \rightarrow \infty} \mathbb{P}\left(\sqrt{2 \log n} \left(\max_{1 \leq i \leq n} Z_{i,1} - b_n\right) \leq w\right). \end{aligned}$$

By the limiting distribution of the maximum of n independent standard Gaussians, namely (21),

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P}\left(\sqrt{2 \log n} \left(\max_{1 \leq i \leq n} Z_{i,1} - b_n\right) \leq w\right) \\ &= \exp(-\exp(-w)). \end{aligned}$$

Then the result follows. \square

We now state and prove the main result of this section—the conditional version of Theorem 5.2. By Portmanteau’s theorem, a weak convergence $X_n \Rightarrow X$ is established by showing $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ for every bounded continuous function f . We use this version in the theorem below.

THEOREM 5.5. *Fix a positive integer r and for each n fix any distinct subsets $I_0, \dots, I_{r-1} \subset [n], |I_\ell| = k, 0 \leq \ell \leq r - 1$, and distinct subsets $J_1, \dots, J_r \subset [n], |J_\ell| = k, 1 \leq \ell \leq r$. Fix any sequence C_1, \dots, C_{2r-1} of $k \times k$ matrices satisfying $\|C_\ell\|_\infty \leq \omega_n, 1 \leq \ell \leq 2r - 1$. Let*

$$\mathcal{E}_r = \mathcal{E}(I_i, 1 \leq i \leq r; J_j, 1 \leq j \leq r; C_\ell, 1 \leq \ell \leq 2r - 1)$$

be the event that $\mathbf{C}_{I_{\ell-1}, J_\ell}^n - \sqrt{\frac{2 \log n}{k}} \mathbf{1}\mathbf{1}' = C_{2\ell-1}$ for each $1 \leq \ell \leq r$, $\mathbf{C}_{I_\ell, J_\ell}^n - \sqrt{\frac{2 \log n}{k}} \mathbf{1}\mathbf{1}' = C_{2\ell}$ for each $1 \leq \ell \leq r - 1$ and, furthermore, $\sqrt{\frac{2 \log n}{k}} \mathbf{1}\mathbf{1}' + C_\ell$ is the ℓ th matrix returned by the algorithm \mathcal{LAS} for all $\ell = 1, \dots, 2r - 1$. Namely, $\mathbf{C}_\ell^n = \sqrt{\frac{2 \log n}{k}} \mathbf{1}\mathbf{1}' + C_\ell$. Fix any set of columns $J \subset [n], |J| = k$ such that $J \setminus (\bigcup_{1 \leq \ell \leq r-1} J_\ell) \neq \emptyset$, including possibly J_r , and let $\mathbf{D}_{\text{Row}}^n$ be the $k \times k$ submatrix of $\mathbf{C}_{([n] \setminus I_{r-1}), J}^n$ with the largest average value and $\hat{\mathbf{D}}_{\text{Row}}^n$ be the $k \times k$ submatrix of $\mathbf{C}_{([n] \setminus \bigcup_{0 \leq \ell \leq r-1} I_\ell), J}^n$ with the largest average value. Then the following holds:

(a)

$$(32) \quad \lim_{n \rightarrow \infty} \inf \mathbb{P}(\hat{\mathbf{D}}_{\text{Row}}^n = \mathbf{D}_{\text{Row}}^n | \mathcal{E}_r) = 1,$$

where \inf is over all I_ℓ, J_ℓ and $C_\ell, 1 \leq \ell \leq 2r - 1$ satisfying $\|C_\ell\|_\infty \leq \omega_n$.

(b) Conditional on \mathcal{E}_r , $\Psi_n^{\text{Row}}(\mathbf{D}_{\text{Row}}^n)$ converges to $\mathbf{C}_{\infty}^{\text{Row}}$ uniformly in $(C_\ell, 1 \leq \ell \leq 2r - 1)$. Specifically, for every bounded continuous function $f : \mathbb{R} \times (\mathbb{R}^{k \times k})^3 \rightarrow \mathbb{R}$ [and similarly to (26)] we have

$$(33) \quad \lim_{n \rightarrow \infty} \sup |\mathbb{E}[f(\Psi_n^{\text{Row}}(\mathbf{D}_{\text{Row}}^n)) | \mathcal{E}_r] - \mathbb{E}[f(\mathbf{C}_{\infty}^{\text{Row}})]| = 0,$$

where sup is over all I_ℓ, J_ℓ and $C_\ell, 1 \leq \ell \leq 2r - 1$ satisfying $\|C_\ell\|_\infty \leq \omega_n$.
(c)

$$\lim_{n \rightarrow \infty} \inf \mathbb{P} \left(\left\| \mathbf{D}_{\text{Row}}^n - \sqrt{\frac{2 \log n}{k}} \right\|_\infty \leq \omega_n | \mathcal{E}_r \right) = 1,$$

where inf is over all I_ℓ, J_ℓ and $C_\ell, 1 \leq \ell \leq 2r - 1$ satisfying $\|C_\ell\|_\infty \leq \omega_n$.

Similar results of (a), (b) and (c) hold for $\mathbf{D}_{\text{Col}}^n, \hat{\mathbf{D}}_{\text{Col}}^n$ and $\Psi_n^{\text{Col}}(\mathbf{D}_{\text{Col}}^n)$ when $I \subset [n], |I| = k$ is such that $I \setminus (\bigcup_{0 \leq \ell \leq r-1} I_\ell) \neq \emptyset$, $\mathbf{D}_{\text{Col}}^n$ is the $k \times k$ submatrix of $\mathbf{C}_{I, ([n] \setminus J_r)}^n$ with the largest average value and $\hat{\mathbf{D}}_{\text{Col}}^n$ is the $k \times k$ submatrix of $\mathbf{C}_{I, ([n] \setminus \bigcup_{1 \leq \ell \leq r} J_r)}^n$ with the largest average value.

Regarding the subset of columns J in the theorem above, primarily the special case $J = J_r$ will be used. Note that indeed $J_r \setminus (\bigcup_{1 \leq \ell \leq r-1} J_\ell) \neq \emptyset$ w.h.p., by applying part (a) of the theorem to the previous steps of the algorithm which claims the identity $\hat{\mathbf{D}}_{\text{Col}}^n = \mathbf{D}_{\text{Col}}^n$ w.h.p.

PROOF. Unlike for $\mathbf{D}_{\text{Row}}^n$, in the construction of $\hat{\mathbf{D}}_{\text{Row}}^n$ we only use rows $\mathbf{C}_{i,J}^n$ which are outside the rows $\bigcup_{0 \leq \ell \leq r-1} I_\ell$ already used in the previous iterations of the algorithm. The bulk of the proof of the theorem will be to establish that claims (b) and (c) of the theorem hold for this matrix instead. Assuming this is the case, (a) then implies that (b) and (c) hold for $\mathbf{D}_{\text{Row}}^n$ as well, completing the proof of theorem.

First, we prove part (a) assuming that (c) holds for $\hat{\mathbf{D}}_{\text{Row}}^n$. We fix any set of rows $I \subset [n] \setminus I_{r-1}$ with cardinality k satisfying $I \cap (\bigcup_{0 \leq \ell \leq r-2} I_\ell) \neq \emptyset$. For every $i \in I \cap (\bigcup_{0 \leq \ell \leq r-2} I_\ell)$ and $j \in J \cap (\bigcup_{1 \leq \ell \leq r-1} J_\ell^n)$, $\mathbf{C}_{i,j}^n$ is either included in some \mathbf{C}_ℓ^n , in which case $|\mathbf{C}_{i,j}^n - \sqrt{2 \log n / k}| \leq \omega_n$ holds under the event \mathcal{E}_r , or $\mathbf{C}_{i,j}^n$ is not included in any \mathbf{C}_ℓ^n , in which case $\mathbf{C}_{i,j}^n$ is $O(1)$ w.h.p. under \mathcal{E}_r . Then in both cases we have

$$\lim_{n \rightarrow \infty} \inf \mathbb{P} \left(\mathbf{C}_{i,j}^n - \sqrt{\frac{2 \log n}{k}} \leq \omega_n | \mathcal{E}_r \right) = 1,$$

where inf is over all I_ℓ, J_ℓ and $C_\ell, 1 \leq \ell \leq 2r - 1$ satisfying $\|C_\ell\|_\infty \leq \omega_n$. Since $|\bigcup_{0 \leq \ell \leq r-2} I_\ell| \leq (r - 1)k$ and r is fixed, by the union bound the same applies to all such elements $\mathbf{C}_{i,j}^n$. By part (c), which was assumed to hold for $\hat{\mathbf{D}}_{\text{Row}}^n$, we have

$$\lim_{n \rightarrow \infty} \inf \mathbb{P} \left(\sum_{j \in J} \mathbf{C}_{i,j}^n - k \sqrt{\frac{2 \log n}{k}} \leq k \omega_n, \forall i \in [n] \setminus \left(\bigcup_{0 \leq \ell \leq r-1} I_\ell \right) | \mathcal{E}_r \right) = 1,$$

where inf is over the same set of events as above. On the other hand for every $i \in I \cap (\bigcup_{0 \leq \ell \leq r-2} I_\ell)$ and $j \in J \setminus (\bigcup_{1 \leq \ell \leq r-1} J_\ell)$, $\mathbf{C}_{i,j}^n$ is not included in any \mathbf{C}_ℓ^n , $1 \leq \ell \leq 2r - 1$, and hence is $O(1)$ w.h.p. under the event \mathcal{E}_r , which gives

$$(34) \quad \lim_{n \rightarrow \infty} \sup \mathbb{P}(\mathbf{C}_{i,j}^n \geq (1/2)\sqrt{2 \log n/k} | \mathcal{E}_r) = 0.$$

Since $|\bigcup_{0 \leq \ell \leq r-2} I_\ell| \leq (r - 1)k$ and r is fixed, by the union bound the same applies to all such elements $\mathbf{C}_{i,j}^n$. It follows, that w.h.p. the average value of the matrix $\mathbf{C}_{I,J}^n$ for all sets of rows $I \in [n] \setminus I_{r-1}$ satisfying $I \cap (\bigcup_{0 \leq \ell \leq r-2} I_\ell) \neq \emptyset$ is at most $(1 - 1/(2k^2))\sqrt{2 \log n/k} + \omega_n$, since by assumption $J \setminus (\bigcup_{1 \leq \ell \leq r-1} J_\ell) \neq \emptyset$, and thus there exists at least one entry in $\mathbf{C}_{I,J}^n$ satisfying (34). On the other hand by part (c), the average value of $\hat{\mathbf{D}}_{\text{Row}}^n$ is at least $\sqrt{2 \log n/k} - \omega_n$, and thus (32) in (a) follows. The proof for $\hat{\mathbf{D}}_{\text{Col}}^n$ is similar.

Thus we now establish (b) and (c) for $\hat{\mathbf{D}}_{\text{Row}}^n$. In order to simplify the notation, we use $\mathbf{D}_{\text{Row}}^n$ in place of $\hat{\mathbf{D}}_{\text{Row}}^n$. We fix I_ℓ, J_ℓ, C_ℓ and J as described in the assumption of the theorem. Let $I^c = [n] \setminus (\bigcup_{0 \leq \ell \leq r-1} I_\ell)$. For each $i \in I^c$, consider the event denoted by $\mathcal{B}_i^{\text{Row}}$ that for each $\ell = 1, \dots, r - 1$ $\mathbf{C}_{I_\ell, J_\ell}^n = \sqrt{\frac{2 \log n}{k}} \mathbf{1}\mathbf{1}^T + C_{2\ell}$ and

$$(35) \quad \text{Ave}(\mathbf{C}_{i, J_\ell}^n) \leq \min_{i' \in I_\ell} \text{Ave}(\mathbf{C}_{i', J_\ell}^n).$$

Our key observation is that the distribution of the submatrix $\mathbf{C}_{I^c, J}^n$ conditional on the event \mathcal{E}_r is the same as the distribution of the same submatrix conditional on the event $\bigcap_{i \in I^c} \mathcal{B}_i^{\text{Row}}$. Thus in proving parts (b) and (c) we may replace conditioning on the event \mathcal{E}_r by conditioning on the event $\bigcap_{i \in I^c} \mathcal{B}_i^{\text{Row}}$. A similar observation holds for the column version of the statement which we skip.

Now fix any $i \in I^c$. Let $J_0 = J$, and consider the r -vector

$$(36) \quad (Y_\ell \triangleq k^{\frac{1}{2}} \text{Ave}(\mathbf{C}_{i, J_\ell}^n), 0 \leq \ell \leq r - 1).$$

Without any conditioning, the distribution of this vector is the distribution of standard normal random variables with correlation structure determined by the vector of cardinalities of intersections of the sets J_ℓ , namely vector $\sigma \triangleq (|J_\ell \cap J_{\ell'}|, 0 \leq \ell, \ell' \leq r - 1)$. By Lemma 5.3, there exists a $r \times r$ matrix L which depends on σ only and with properties (a)–(d) described in the lemma, such that the distribution of the vector (36) is the same as the one of LZ , where Z is the r -vector of i.i.d. standard normal random variables. We will establish Theorem 5.5 from the following proposition, which is an analogue of Lemma 5.4. We delay its proof for later.

PROPOSITION 5.6. *Let $\mathbf{Z} = (Z_{i,\ell}, 1 \leq i \leq n, 1 \leq \ell \leq r - 1)$ be a matrix of i.i.d. standard normal random variables independent from the $n \times k$ matrix $\mathbf{C}^{n \times k}$.*

Given any $\bar{c} = (c_\ell, 1 \leq \ell \leq r - 1) \in \mathbb{R}^{r-1}$, for each $i = 1, \dots, n$, let \mathcal{B}_i be the event

$$\begin{aligned} & \left[L(k^{\frac{1}{2}} \text{Ave}(\mathbf{C}_{i,[k]}^{n \times k}), Z_{i,1}, \dots, Z_{i,r-1})^T \right]_{\ell+1} \\ & \leq \sqrt{2 \log n} + \sqrt{k} c_\ell, \quad \forall 1 \leq \ell \leq r - 1, \end{aligned}$$

where $[\cdot]_\ell$ denotes the ℓ th component of the vector in the argument. For every bounded continuous function $f : \mathbb{R} \times (\mathbb{R}^{k \times k})^3 \rightarrow \mathbb{R}$,

$$(37) \quad \lim_{n \rightarrow \infty} \sup_{\bar{c}: \|\bar{c}\|_\infty \leq \omega_n} \left| \mathbb{E} \left[f(\Psi_n^{\text{Row}}(\mathbf{C}^k)) | \mathcal{RD}_n, \bigcap_{1 \leq i \leq n} \mathcal{B}_i \right] - \mathbb{E} [f(\mathbf{C}_\infty^{\text{Row}})] \right| = 0.$$

The proposition essentially says that the events \mathcal{B}_i have an asymptotically negligible effect on the distribution of the largest $k \times k$ submatrix of $\mathbf{C}^{n \times k}$.

First, we show how this proposition implies part (b) of Theorem 5.5. The event $\bigcap_{i \in I^c} \mathcal{B}_i^{\text{Row}}$ implies that $\|C_{2\ell}\|_\infty \leq \omega_n$, for all ℓ and, therefore,

$$-\omega_n \leq c_\ell \triangleq \min_{i' \in I^c} \text{Ave}(\mathbf{C}_{i', J_\ell}^n) - \sqrt{\frac{2 \log n}{k}} \leq \omega_n, \quad 1 \leq \ell \leq r - 1.$$

The events $\bigcap_{i \in I^c} \mathcal{B}_i^{\text{Row}}$ and $\bigcap_{1 \leq i \leq n} \mathcal{B}_i$ are then identical modulo the difference of cardinalities $|I^c|$ versus n . Since k is a constant, then $|I^c| = n - O(1)$, and the result is claimed in the limit $n \rightarrow \infty$. The assertion (b) holds.

We now establish (c). Recalling the representation (25) and the definition of b_n , we have

$$\begin{aligned} \mathbf{D}_{\text{Row}}^n - \sqrt{\frac{2 \log n}{k}} \mathbf{1}\mathbf{1}' &= \frac{\Psi_{n,1}^{\text{Row}}(\mathbf{D}_{\text{Row}}^n)}{\sqrt{2k \log n}} \mathbf{1}\mathbf{1}' + \frac{\Psi_{n,2}^{\text{Row}}(\mathbf{D}_{\text{Row}}^n)}{\sqrt{2k \log n}} + \Psi_{n,3}^{\text{Row}}(\mathbf{D}_{\text{Row}}^n) \\ &\quad + \Psi_{n,4}^{\text{Row}}(\mathbf{D}_{\text{Row}}^n) + O\left(\frac{\log \log n}{\sqrt{\log n}}\right). \end{aligned}$$

The claim then follows immediately from part (b), specifically from the uniform weak convergence $\Psi_n^{\text{Row}}(\mathbf{D}_{\text{Row}}^n) \Rightarrow \mathbf{C}_\infty^{\text{Row}}$. \square

PROOF OF PROPOSITION 5.6. According to Theorem 5.2, for every bounded continuous function f ,

$$(38) \quad \lim_{n \rightarrow \infty} \mathbb{E} [f(\Psi_n^{\text{Row}}(\mathbf{C}^k)) | \mathcal{RD}_n] = \mathbb{E} [f(\mathbf{C}_\infty^{\text{Row}})].$$

Our goal is to show

$$(39) \quad \begin{aligned} & \lim_{n \rightarrow \infty} \sup_{\bar{c}: \|\bar{c}\|_\infty \leq \omega_n} \left| \mathbb{E} \left[f(\Psi_n^{\text{Row}}(\mathbf{C}^k)) | \mathcal{RD}_n, \bigcap_{1 \leq i \leq n} \mathcal{B}_i \right] \right. \\ & \left. - \mathbb{E} [f(\Psi_n^{\text{Row}}(\mathbf{C}^k)) | \mathcal{RD}_n] \right| = 0. \end{aligned}$$

(37) follows from (38) and (39). We claim that if the following relation holds for any $W \in \mathbb{R} \times (\mathbb{R}^{k \times k})^3$

$$(40) \quad \lim_{n \rightarrow \infty} \sup_{\bar{c}: \|\bar{c}\|_\infty \leq \omega_n} \left| \frac{\mathbb{P}(\bigcap_{1 \leq i \leq n} \mathcal{B}_i | \Psi_n^{\text{Row}}(\mathbf{C}^k) = W, \mathcal{RD}_n)}{\mathbb{P}(\bigcap_{1 \leq i \leq n} \mathcal{B}_i)} - 1 \right| = 0,$$

then (39) follows. By symmetry,

$$\mathbb{P}\left(\mathcal{RD}_n | \bigcap_{1 \leq i \leq n} \mathcal{B}_i\right) = \binom{n}{k}^{-1} = \mathbb{P}(\mathcal{RD}_n).$$

Using the equation above, we compute

$$\begin{aligned} & \mathbb{E}\left[f(\Psi_n^{\text{Row}}(\mathbf{C}^k)) | \mathcal{RD}_n, \bigcap_{1 \leq i \leq n} \mathcal{B}_i\right] \\ &= \int f(W) \frac{d\mathbb{P}(\Psi_n^{\text{Row}}(\mathbf{C}^k) = W, \mathcal{RD}_n, \bigcap_{1 \leq i \leq n} \mathcal{B}_i)}{\mathbb{P}(\mathcal{RD}_n, \bigcap_{1 \leq i \leq n} \mathcal{B}_i)} \\ (41) \quad &= \int f(W) \frac{\mathbb{P}(\bigcap_{1 \leq i \leq n} \mathcal{B}_i | \Psi_n^{\text{Row}}(\mathbf{C}^k) = W, \mathcal{RD}_n)}{\mathbb{P}(\bigcap_{1 \leq i \leq n} \mathcal{B}_i)} \\ & \quad \times \frac{d\mathbb{P}(\Psi_n^{\text{Row}}(\mathbf{C}^k) = W, \mathcal{RD}_n)}{\mathbb{P}(\mathcal{RD}_n | \bigcap_{1 \leq i \leq n} \mathcal{B}_i)} \\ &= \int f(W) \frac{\mathbb{P}(\bigcap_{1 \leq i \leq n} \mathcal{B}_i | \Psi_n^{\text{Row}}(\mathbf{C}^k) = W, \mathcal{RD}_n)}{\mathbb{P}(\bigcap_{1 \leq i \leq n} \mathcal{B}_i)} d\mathbb{P}(\Psi_n^{\text{Row}}(\mathbf{C}^k)) \\ &= W | \mathcal{RD}_n). \end{aligned}$$

Substituting (41) into the left-hand side of (39) and then using (40) and the boundedness of f , we obtain (39).

The rest of the proof is to show that (40) holds for any $W \in \mathbb{R} \times (\mathbb{R}^{k \times k})^3$. Fix any $W \triangleq (w_1, W_2, W_3, W_4)$ where $w_1 \in \mathbb{R}$ and $W_2, W_3, W_4 \in \mathbb{R}^{k \times k}$. Conditional on $\Psi_n^{\text{Row}}(\mathbf{C}^k) = W$, and writing $W_2 = (W_{i,j}^2)$ the average value of the i th row of \mathbf{C}^k is

$$\mathbf{C}_{i \cdot}^k = \frac{W_{i,1}^2}{\sqrt{2k \log n}} + \frac{w_1}{\sqrt{2k \log n}} + \frac{b_n}{\sqrt{k}} \triangleq w_{i,n}, \quad i = 1, \dots, k,$$

where we used (25) and the fact that for any matrix $B \in \mathbb{R}^{k \times k}$, the average values of the i th rows of $\text{Col}(B)$ and $\text{ANOVA}(B)$ are both zero, for all $i \in [k]$. Let $c_n(W) = \min_{1 \leq i \leq k} w_{i,n}$. Note that

$$(42) \quad w_{i,n} = \sqrt{\frac{2 \log n}{k}} + o(1), \quad c_n(W) = \sqrt{\frac{2 \log n}{k}} + o(1).$$

The event \mathcal{RD}_n is equivalent to the event

$$\max_{k+1 \leq i \leq n} \text{Ave}(\mathbf{C}_i^{n \times k}) \leq c_n(W).$$

Now observe that by independence of rows of \mathbf{Z}

$$\begin{aligned} & \mathbb{P}\left(\bigcap_{1 \leq i \leq n} \mathcal{B}_i \mid \Psi_n^{\text{Row}}(\mathbf{C}^k) = W, \max_{k+1 \leq i \leq n} \text{Ave}(\mathbf{C}_i^{n \times k}) \leq c_n(W)\right) \\ (43) \quad &= \mathbb{P}\left(\bigcap_{1 \leq i \leq k} \mathcal{B}_i \mid \Psi_n^{\text{Row}}(\mathbf{C}^k) = W\right) \\ & \times \mathbb{P}\left(\bigcap_{k+1 \leq i \leq n} \mathcal{B}_i \mid \max_{k+1 \leq i \leq n} \text{Ave}(\mathbf{C}_i^{n \times k}) \leq c_n(W)\right). \end{aligned}$$

By (21), we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P}\left(\max_{k+1 \leq i \leq n} \text{Ave}(\mathbf{C}_i^{n \times k}) \leq c_n(W)\right) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}\left(\max_{k+1 \leq i \leq n} \sqrt{2 \log n} (\sqrt{k} \text{Ave}(\mathbf{C}_i^{n \times k}) - b_n) \leq w_1 + \min_{1 \leq i \leq k} W_{i,1}^2\right) \\ &= \exp\left(-\exp\left(-w_1 - \min_{1 \leq i \leq k} W_{i,1}^2\right)\right). \end{aligned}$$

Furthermore, by Lemma 5.4 we also have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sup_{\bar{c}: \|\bar{c}\|_\infty \leq \omega_n} \left| \mathbb{P}\left(\max_{k+1 \leq i \leq n} \text{Ave}(\mathbf{C}_i^{n \times k}) \leq c_n(W) \mid \bigcap_{k+1 \leq i \leq n} \mathcal{B}_i\right) \right. \\ & \left. - \exp\left(-\exp\left(-w_1 - \min_{1 \leq i \leq k} W_{i,1}^2\right)\right) \right| = 0. \end{aligned}$$

Applying Bayes' rule, we obtain

$$\begin{aligned} (44) \quad & \lim_{n \rightarrow \infty} \sup_{\bar{c}: \|\bar{c}\|_\infty \leq \omega_n} \left| \frac{\mathbb{P}\left(\bigcap_{k+1 \leq i \leq n} \mathcal{B}_i \mid \max_{k+1 \leq i \leq n} \text{Ave}(\mathbf{C}_i^{n \times k}) \leq c_n(W)\right)}{\mathbb{P}\left(\bigcap_{k+1 \leq i \leq n} \mathcal{B}_i\right)} - 1 \right| \\ &= 0. \end{aligned}$$

Now we claim that

$$(45) \quad \lim_{n \rightarrow \infty} \sup_{\bar{c}: \|\bar{c}\|_\infty \leq \omega_n} \left| \mathbb{P}\left(\bigcap_{1 \leq i \leq k} \mathcal{B}_i \mid \Psi_n^{\text{Row}}(\mathbf{C}^k) = W\right) - 1 \right| = 0.$$

Indeed the event $\mathcal{B}_i, i \leq k$ conditioned on $\Psi_n^{\text{Row}}(\mathbf{C}^k) = W$ is

$$L_{\ell+1,1} k^{\frac{1}{2}} w_{i,n} + L_{\ell+1,2} Z_{i,1} + \dots + L_{\ell+1,r+1} Z_{i,r} \leq \sqrt{2 \log n} + c_\ell, \quad 1 \leq \ell \leq r.$$

Now recall from Lemma 5.3 that $L_{\ell+1,1} \leq 1 - 1/k$. Then applying (42) we conclude

$$L_{\ell+1,1} k^{\frac{1}{2}} w_{i,n} \leq (1 - 1/k) \sqrt{2 \log n} + o(1).$$

Trivially, we have

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(L_{\ell,2} Z_{i,1} + \dots + L_{\ell,r+1} Z_{i,r} \leq \frac{1}{2k} \sqrt{2 \log n}, \forall 1 \leq i \leq k, 1 \leq \ell \leq r \right) = 1,$$

simply because $\sqrt{2 \log n}$ is a growing function and the elements of L are bounded by 1. The claim then follows since $|c_\ell| \leq \omega_n = o(\sqrt{2 \log n})$. Similar to the reasoning of (45), we also have

$$(46) \quad \lim_{n \rightarrow \infty} \sup_{\tilde{c}: \|\tilde{c}\| \leq \omega_n} \left| \mathbb{P} \left(\bigcap_{1 \leq i \leq k} \mathcal{B}_i \right) - 1 \right| = 0.$$

Then if we multiply the denominator of the first term in (44) by $\mathbb{P}(\bigcap_{1 \leq i \leq k} \mathcal{B}_i)$, we still have

$$(47) \quad \lim_{n \rightarrow \infty} \sup_{\tilde{c}: \|\tilde{c}\| \leq \omega_n} \left| \frac{\mathbb{P}(\bigcap_{k+1 \leq i \leq n} \mathcal{B}_i \mid \max_{k+1 \leq i \leq n} \text{Ave}(\mathbf{C}_i^{n \times k}) \leq c_n(W))}{\mathbb{P}(\bigcap_{1 \leq i \leq n} \mathcal{B}_i)} - 1 \right| = 0.$$

Applying (45) and (47) for (43), we obtain (40). \square

5.3. *Bounding the number of steps of \mathcal{LAS} . Proof of Theorem 2.1.* Next we obtain an upper bound on the number of steps taken by the \mathcal{LAS} algorithm as well as a bound on the average value of the matrix \mathbf{C}_r^n obtained by the \mathcal{LAS} algorithm in step r , when r is constant, and use these bounds to conclude the proof of Theorem 2.1. For this purpose, we will rely on a repeated application of Theorem 5.5.

We now introduce some additional notation. Fix r and consider the matrix $\mathbf{C}_{2r}^n = \mathbf{C}_{I_r^n, J_r^n}^n$ obtained in step $2r$ of \mathcal{LAS} , assuming $T_{\mathcal{LAS}} \geq 2r$. Recall that \tilde{I}_{r-1}^n is the set of k rows with the largest sum of entries in $\mathbf{C}_{[n] \setminus I_{r-1}^n, J_r^n}^n$. Then the matrix \mathbf{C}_{2r}^n is obtained by combining top rows of $\mathbf{C}_{2r-1}^n = \mathbf{C}_{I_{r-1}^n, J_r^n}^n$ and the top rows of $\mathbf{C}_{\tilde{I}_{r-1}^n, J_r^n}^n$. We denote the part of $\mathbf{C}_{2r}^n = \mathbf{C}_{I_r^n, J_r^n}^n$ coming from $\mathbf{C}_{I_{r-1}^n, J_r^n}^n$ by $\mathbf{C}_{2r,1}^n$ and the part coming from $\mathbf{C}_{\tilde{I}_{r-1}^n, J_r^n}^n$ by $\mathbf{C}_{2r,2}^n$. The rows of $\mathbf{C}_{I_{r-1}^n, J_r^n}^n$ leading to $\mathbf{C}_{2r,1}^n$ are denoted by $I_{r,1}^n \subset I_{r-1}^n$ with $|I_{r,1}^n| \triangleq K_1$ (a random variable), and the rows of $\mathbf{C}_{\tilde{I}_{r-1}^n, J_r^n}^n$ leading to $\mathbf{C}_{2r,2}^n$ are denoted by $I_{r,2}^n \subset \tilde{I}_{r-1}^n$ with $|I_{r,2}^n| \triangleq K_2 = k - K_1$. Thus $I_{r,1}^n \cup I_{r,2}^n = I_r^n$ and $\mathbf{C}_{2r,\ell}^n = \mathbf{C}_{I_{r,\ell}^n, J_r^n}^n$, $\ell = 1, 2$, as shown in Figure 7 where the symbol “ Δ ” represents the entries in \mathbf{C}_{2r}^n . Our first step is to show that starting from $r = 2$, for every positive real a the average value of \mathbf{C}_r^n is at least $\sqrt{\frac{2 \log n}{k}} + a$ with probability bounded away from zero as n increases. We will only show this result for odd r since by monotonicity we also have $\text{Ave}(\mathbf{C}_{r+1}^n) \geq \text{Ave}(\mathbf{C}_r^n)$.

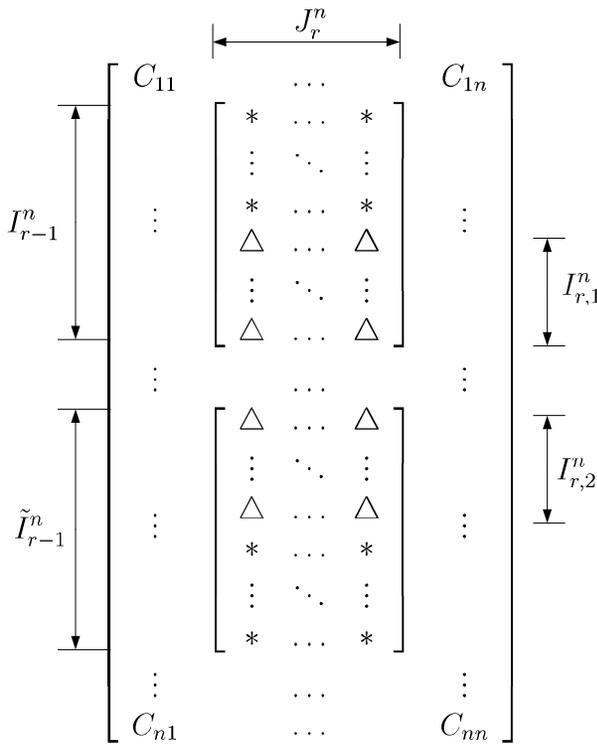


FIG. 7. Step $2r$ of \mathcal{LAS} algorithm.

PROPOSITION 5.7. *There exists a strictly positive function $\psi_1 : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ which depends only on k , such that for all $r > 0, a > 0$*

$$\liminf_n \mathbb{P} \left(\text{Ave}(\mathbf{C}_{2r+1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a \cup \{T_{\mathcal{LAS}} \leq 2r\} \mid T_{\mathcal{LAS}} \geq 2r - 1 \right) \geq \psi_1(a).$$

Namely, assuming the algorithm proceeds for $2r - 1$ steps, with probability at least approximately $\psi_1(a)$ either it stops in step $2r$ or proceeds to step $2r + 1$, producing a matrix with average at least $\sqrt{2 \log n / k} + a$.

PROOF. By Theorem 5.5, the distribution of $\Psi_r^{\text{Row}}(\mathbf{C}_{\tilde{I}_{r-1}, J_r^n}^n)$ conditional on the event $T_{\mathcal{LAS}} \geq 2r - 1$ is given by $\mathbf{C}_\infty^{\text{Row}}$ in the limit as $n \rightarrow \infty$. In particular, the row averages $\text{Ave}(\mathbf{C}_{i, J_r^n}^n), i \in \tilde{I}_{r-1}^n$ of this matrix are concentrated around $\sqrt{\frac{2 \log n}{k}}$ w.h.p. as $n \rightarrow \infty$. Motivated by this, we write the row averages of $\mathbf{C}_{\tilde{I}_{r-1}, J_r^n}^n$ as $\sqrt{\frac{2 \log n}{k}} + C_1 / (\sqrt{2k \log n}), \dots, \sqrt{\frac{2 \log n}{k}} + C_k / (\sqrt{2k \log n})$ for the appropriate random values C_1, \dots, C_k . Denote the event $\max_j |C_j| \leq \omega_n$ by \mathcal{L}_{2r} . Then by Theo-

rem 5.5, we have

$$(48) \quad \lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{L}_{2r}^c | T_{\mathcal{L}AS} \geq 2r - 1) = 0.$$

If the event $T_{\mathcal{L}AS} \leq 2r - 1$ takes place, then $T_{\mathcal{L}AS} \leq 2r$ and, therefore, the event

$$\left\{ \text{Ave}(\mathbf{C}_{2r+1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a \right\} \cup \{T_{\mathcal{L}AS} \leq 2r\}$$

also takes place. Now consider the event $T_{\mathcal{L}AS} \geq 2r$. On this event, the matrices $\mathbf{C}_{2r,1}^n$ and $\mathbf{C}_{2r,2}^n$ are well-defined. Recall the notation $I_{r,1}^n$ and $I_{r,2}^n$ for the row indices of $\mathbf{C}_{2r,1}^n$ and $\mathbf{C}_{2r,2}^n$, respectively, and $0 \leq K_1 \leq k - 1$ and $K_2 = k - K_1$ —their respective cardinalities. Suppose first that

$$(49) \quad \text{Sum}(\mathbf{C}_{2r,1}^n) > K_1 \sqrt{2k \log n} + 2k^2 a,$$

where $\text{Sum}(B)$ denotes the sum of all the entries in a matrix B . Then by the bound $\max_j |C_j| \leq \omega_n$ where we recall $\omega_n = o(\sqrt{\log n})$ we have

$$\begin{aligned} \text{Sum}(\mathbf{C}_{2r}^n) &\geq (K_1 + K_2) \sqrt{2k \log n} + 2k^2 a - K_2 k \omega_n / \sqrt{2k \log n} \\ &\geq k^2 \sqrt{\frac{2 \log n}{k}} + k^2 a, \end{aligned}$$

for large enough n , implying $\text{Ave}(\mathbf{C}_{2r}^n) \geq \sqrt{\frac{2 \log n}{k}} + a$ and, therefore, either $\text{Ave}(\mathbf{C}_{2r+1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a$, for large enough n or $T_{\mathcal{L}AS} \leq 2r$.

Now instead assume the event

$$(50) \quad \text{Sum}(\mathbf{C}_{2r,1}^n) \leq K_1 \sqrt{2k \log n} + 2k^2 a$$

takes place (including the possibility $K_1 = 0$) which we denote by \mathcal{H}_1 . Then there exists $j_0 \in J_r^n$ such that

$$\text{Sum}(\mathbf{C}_{I_{r,1}^n, j_0}^n) \leq K_1 \sqrt{\frac{2 \log n}{k}} + 2ka.$$

We pick any such column j_0 , for example the one which is the smallest index-wise. Consider the event

$$\text{Sum}(\mathbf{C}_{I_{r,2}^n, j_0}^n) \leq K_2 \sqrt{\frac{2 \log n}{k}} - 4k^2 a,$$

which we denote by \mathcal{H}_2 .

We claim that the probability of the event \mathcal{H}_2 conditioned on the events $T_{\mathcal{L}AS} \geq 2r$, \mathcal{L}_{2r} and \mathcal{H}_1 is bounded away from zero as n increases:

$$\liminf_n \mathbb{P}(\mathcal{H}_2 | T_{\mathcal{L}AS} \geq 2r, \mathcal{L}_{2r}, \mathcal{H}_1) > 0.$$

For this purpose, fix any realization of the matrix \mathbf{C}_{2r-1}^n which we write as $\sqrt{\frac{2\log n}{k}} + C$ for an appropriate $k \times k$ matrix C , the realizations c_1, \dots, c_k of C_1, \dots, C_k , and the realization $j_0 \in J_r^n$, which are all consistent with the events $T_{\mathcal{L}AS} \geq 2r, \mathcal{L}_{2r}, \mathcal{H}_1$. In particular, the row averages of $\mathbf{C}_{\tilde{I}_{r-1}^n, J_r^n}^n$ are $\sqrt{\frac{2\log n}{k}} + c_1/(\sqrt{2k \log n}), \dots, \sqrt{\frac{2\log n}{k}} + c_k/(\sqrt{2k \log n})$ and $\max_j |c_j| \leq \omega_n$. Note that C and c_1, \dots, c_k uniquely determine the subsets $I_{r,1}^n$ and $I_{r,2}^n$, and their cardinalities which we denote by I_1, I_2 and k_1, k_2 , respectively. Additionally, c_1, \dots, c_k uniquely determine $\text{Ave}(\mathbf{C}_{\tilde{I}_{r-1}^n, J_r^n}^n)$:

$$\text{Ave}(\mathbf{C}_{\tilde{I}_{r-1}^n, J_r^n}^n) = \sqrt{\frac{2\log n}{k}} + \frac{\sum c_j}{k\sqrt{2k \log n}},$$

which we can also write as $\text{Ave}(\mathbf{C}_{\tilde{I}_{r-1}^n, J_r^n}^n) = \bar{c}/(\sqrt{2k \log n}) + b_n/\sqrt{k}$ where $\bar{c} \triangleq \Psi_{n,1}^{\text{Row}}(\mathbf{C}_{\tilde{I}_{r-1}^n, J_r^n}^n)$. Note that $\max_j |c_j| \leq \omega_n = o(\sqrt{\log n})$ also implies $\bar{c} = o(\sqrt{\log n})$. Next we show that

$$(51) \quad \lim_{n \rightarrow \infty} \inf_{C, c_1, \dots, c_k} \mathbb{P}(\mathcal{H}_2 | C, c_1, \dots, c_k) \geq \psi_1(a),$$

for some strictly positive function ψ_1 which depends on k only, where $\mathbb{P}(\cdot | C, c_1, \dots, c_k)$ indicates conditioning on the realizations C, c_1, \dots, c_k and $\inf_{C, c_1, \dots, c_k}$ is taken over all choices of C, c_1, \dots, c_k consistent with the events $T_{\mathcal{L}AS} \geq 2r, \mathcal{L}_{2r}, \mathcal{H}_1$. These realizations imply

$$\Psi_{n,2}^{\text{Row}}(\mathbf{C}_{\tilde{I}_{r-1}^n, J_r^n}^n) = \begin{pmatrix} c_1 - \bar{c} \\ \vdots \\ c_k - \bar{c} \end{pmatrix} \mathbf{1}' + \frac{\log(4\pi \log n)}{2},$$

where the last term is simply $\sqrt{2 \log n}(\sqrt{2 \log n} - b_n)$. Thus by representation (25) and by $\bar{c}, c_j = o(\sqrt{\log n})$, we have

$$\begin{aligned} \mathbf{C}_{\tilde{I}_{r-1}^n, J_r^n}^n &= \frac{\bar{c}}{\sqrt{2k \log n}} + \frac{b_n}{\sqrt{k}} \mathbf{1}' + (\sqrt{2k \log n})^{-1} \begin{pmatrix} c_1 - \bar{c} \\ \vdots \\ c_k - \bar{c} \end{pmatrix} \mathbf{1}' + \frac{\log(4\pi \log n)}{2\sqrt{2k \log n}} \\ &\quad + \Psi_{n,3}^{\text{Row}}(\mathbf{C}_{\tilde{I}_{r-1}^n, J_r^n}^n) + \Psi_{n,4}^{\text{Row}}(\mathbf{C}_{\tilde{I}_{r-1}^n, J_r^n}^n) \\ &= \sqrt{\frac{2 \log n}{k}} \mathbf{1}' + \Psi_{n,3}^{\text{Row}}(\mathbf{C}_{\tilde{I}_{r-1}^n, J_r^n}^n) + \Psi_{n,4}^{\text{Row}}(\mathbf{C}_{\tilde{I}_{r-1}^n, J_r^n}^n) + O\left(\frac{\omega_n}{\sqrt{\log n}}\right) \end{aligned}$$

[recall that $\log \log n = O(\omega_n)$ and $\omega_n = o(\sqrt{\log n})$]. Then by Theorem 5.5, we have

$$\lim_{n \rightarrow \infty} \inf_{C, c_1, \dots, c_k} \mathbb{P}(\mathcal{H}_2 | C, c_1, \dots, c_k)$$

is the probability that the sum of the entries of $\text{Col}(\mathbf{C}^k) + \text{ANOVA}(\mathbf{C}^k)$ indexed by the subset I_2 and column j_0 is at most $-4k^2a$ which takes some value $\psi(a, |I_2|) > 0$ and depends only on a, k and the cardinality of I_2 . Let $\psi_1(a) \triangleq \min_{1 \leq |I_2| \leq k} \psi(a, |I_2|)$. Then the claim in (51) follows. We have established

$$\liminf_{n \rightarrow \infty} \mathbb{P}(\mathcal{H}_2 | T_{\mathcal{L}AS} \geq 2r, \mathcal{L}_{2r}, \mathcal{H}_1) \geq \psi_1(a).$$

The event \mathcal{H}_2 implies that for some column j_0

$$\begin{aligned} \text{Sum}(\mathbf{C}_{I_r^n, j_0}^n) &\leq K_1 \sqrt{\frac{2 \log n}{k}} + 2ka + K_2 \sqrt{\frac{2 \log n}{k}} - 4k^2a \\ &\leq \sqrt{2k \log n} - 3k^2a. \end{aligned}$$

By Theorem 5.5 conditional on all of the events $T_{\mathcal{L}AS} \geq 2r, \mathcal{L}_{2r}, \mathcal{H}_1, \mathcal{H}_2$, every column average of $\mathbf{C}_{I_r^n, \tilde{j}_r^n}$ is concentrated around $\sqrt{\frac{2 \log n}{k}}$ w.h.p., implying that the column sum is concentrated around $\sqrt{2k \log n}$ w.h.p. Thus w.h.p. the j_0 th column will be replaced by one of the column in $\mathbf{C}_{I_r^n, \tilde{j}_r^n}$ (and in particular $T_{\mathcal{L}AS} \geq 2r + 1$), and thus during the transition $\mathbf{C}_{2r}^n \rightarrow \mathbf{C}_{2r+1}^n$ the sum of the entries increases by $3k^2a - o(1)$, and thus the average value increases by at least $3a - o(1)$ w.h.p. Recall from Theorem 5.2 that w.h.p. $\text{Ave}(\mathbf{C}_{2r}^n) \geq \text{Ave}(\mathbf{C}_1^n) \geq \sqrt{\frac{2 \log n}{k}} - a$. Then we obtain $\text{Ave}(\mathbf{C}_{2r+1}^n) \geq \sqrt{\frac{2 \log n}{k}} + 2a - o(1) \geq \sqrt{\frac{2 \log n}{k}} + a$ w.h.p. We have obtained

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\text{Ave}(\mathbf{C}_{2r+1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a | T_{\mathcal{L}AS} \geq 2r, \mathcal{L}_{2r}, \mathcal{H}_1, \mathcal{H}_2\right) = 1.$$

By earlier derivation, we have

$$\liminf_{n \rightarrow \infty} \mathbb{P}(\mathcal{H}_2 | T_{\mathcal{L}AS} \geq 2r, \mathcal{L}_{2r}, \mathcal{H}_1) \geq \psi_1(a),$$

thus implying

$$\liminf_n \mathbb{P}\left(\text{Ave}(\mathbf{C}_{2r+1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a | T_{\mathcal{L}AS} \geq 2r, \mathcal{L}_{2r}, \mathcal{H}_1\right) \geq \psi_1(a).$$

Next recall that $\mathcal{H}_1^c \cap \mathcal{L}_{2r}$ implies either $T_{\mathcal{L}AS} \leq 2r$ or $\text{Ave}(\mathbf{C}_{2r+1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a$ for large enough n , from which we obtain

$$\begin{aligned} \liminf_n \mathbb{P}\left(\text{Ave}(\mathbf{C}_{2r+1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a \cup \{T_{\mathcal{L}AS} \leq 2r\} | T_{\mathcal{L}AS} \geq 2r - 1, \mathcal{L}_{2r}\right) \\ \geq \psi_1(a). \end{aligned}$$

Finally, recalling (48) we conclude

$$\begin{aligned} \liminf_n \mathbb{P} \left(\text{Ave}(\mathbf{C}_{2r+1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a \cup \{T_{\mathcal{L}AS} \leq 2r\} \mid T_{\mathcal{L}AS} \geq 2r - 1 \right) \\ \geq \psi_1(a). \end{aligned}$$

This concludes the proof of Proposition 5.7. \square

Now consider the event $T_{\mathcal{L}AS} \geq 2r$, and thus again $\mathbf{C}_{2r,1}^n$ and $\mathbf{C}_{2r,2}^n$ are well-defined. The definitions of $I_{r,1}^n, I_{r,2}^n$ and K_1, K_2 are as above. For any $a > 0$, consider the event that for every $j \in J_r^n$ the sum of entries of the column j in $\mathbf{C}_{2r,1}^n$ is at least $K_1 \sqrt{\frac{2 \log n}{k}} - a$. Denote this event by \mathcal{F}_{2r} . Next we show that provided that $\text{Ave}(\mathbf{C}_{2r-1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a$ with probability bounded away from zero as $n \rightarrow \infty$, for every fixed r , either the event \mathcal{F}_{2r+2t} takes place for some $t \leq k$ or the algorithm stops earlier. To be more precise, we have the following.

PROPOSITION 5.8. *There exists a strictly positive function $\psi_2 : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, which depends on k only such that for every $r > 0$ and $a > 0$,*

$$\begin{aligned} \liminf_{n \rightarrow \infty} \mathbb{P} \left(\bigcup_{0 \leq t \leq k} (\{T_{\mathcal{L}AS} \leq 2r + 2t - 1\} \cup \mathcal{F}_{2r+2t}) \mid T_{\mathcal{L}AS} \geq 2r - 1, \right. \\ \left. \text{Ave}(\mathbf{C}_{2r-1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a \right) \\ \geq \psi_2^{2(k+1)}(a). \end{aligned}$$

The conditioning on the event $\text{Ave}(\mathbf{C}_{2r-1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a$ will not be used explicitly below. The result just shows that even with this conditioning, the claim still holds, so that this result can be used together with Proposition 5.7.

PROOF. On the event $T_{\mathcal{L}AS} \geq 2r - 1$, consider the event \mathcal{G}_{2r} defined by

$$(52) \quad \mathcal{G}_{2r} \triangleq \left\| \mathbf{C}_{I_{r-1}^n, J_r^n}^n - \sqrt{\frac{2 \log n}{k}} \right\|_\infty \leq \frac{a}{4k}.$$

Applying Theorem 5.5, the distribution of $\mathbf{C}_{I_{r-1}^n, J_r^n}^n$ conditioned on $T_{\mathcal{L}AS} \geq 2r - 1$ and $\text{Ave}(\mathbf{C}_{2r-1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a$ is given asymptotically by $\mathbf{C}_\infty^{\text{Row}}$. Recalling the representation (25), we then have that for a certain strictly positive function ψ_2

$$(53) \quad \liminf_n \mathbb{P} \left(\mathcal{G}_{2r} \mid T_{\mathcal{L}AS} \geq 2r - 1, \text{Ave}(\mathbf{C}_{2r-1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a \right) \geq \psi_2(a).$$

If $T_{\mathcal{L}AS} \leq 2r - 1$, then the event $\bigcup_{0 \leq t \leq k} (\{T_{\mathcal{L}AS} \leq 2r + 2t - 1\} \cup \mathcal{F}_{2r+2t})$ holds as well. Otherwise assume the event $T_{\mathcal{L}AS} \geq 2r$ takes place and then the matrices $\mathbf{C}_{2r,1}^n$ and $\mathbf{C}_{2r,2}^n$ which constitute $\mathbf{C}_{2r}^n = \mathbf{C}_{I_r^n, J_r^n}^n$ are well-defined. If the event \mathcal{F}_{2r}^c holds, then there exists $j_0 \in J_r^n$, such that the sum of entries of the column $\mathbf{C}_{I_r^n, j_0}^n$ satisfies

$$(54) \quad \text{Sum}(\mathbf{C}_{I_r^n, j_0}^n) < |I_{r,1}^n| \sqrt{\frac{2 \log n}{k}} - a.$$

The event \mathcal{G}_{2r} implies that the sum of entries of the column $\mathbf{C}_{I_{r,2}^n, j_0}^n$ is at most $|I_{r,2}^n| \sqrt{\frac{2 \log n}{k}} + a/4$, implying that the sum of entries of the column $\mathbf{C}_{I_r^n, j_0}^n$ is at most

$$(55) \quad |I_{r,1}^n| \sqrt{\frac{2 \log n}{k}} - a + |I_{r,2}^n| \sqrt{\frac{2 \log n}{k}} + a/4 = \sqrt{2k \log n} - 3a/4.$$

Introduce now the event \mathcal{G}_{2r+1} as

$$(56) \quad \left\| \mathbf{C}_{I_r^n, \tilde{J}_r^n}^n - \sqrt{\frac{2 \log n}{k}} \right\|_{\infty} \leq \frac{a}{4k}.$$

Again applying Theorem 5.5, we have that

$$(57) \quad \liminf_n \mathbb{P} \left(\mathcal{G}_{2r+1} | \mathcal{G}_{2r}, T_{\mathcal{L}AS} \geq 2r, \mathcal{F}_{2r}^c, \text{Ave}(\mathbf{C}_{2r-1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a \right) \geq \psi_2(a),$$

for the same function ψ_2 . The event \mathcal{G}_{2r+1} implies that the sum of entries of every column in matrix $\mathbf{C}_{I_r^n, \tilde{J}_r^n}^n$ is in particular at least $\sqrt{2k \log n} - a/4$. Now recalling (55) this implies that every column $\mathbf{C}_{I_r^n, j_0}^n$ satisfying (54) will be replaced by a new column from $\mathbf{C}_{I_r^n, \tilde{J}_r^n}^n$ in the transition $\mathbf{C}_{2r}^n \rightarrow \mathbf{C}_{2r+1}^n$ (and in particular this transition takes place and $T_{\mathcal{L}AS} \geq 2r + 1$). The event \mathcal{G}_{2r+1} then implies that every column $\mathbf{C}_{I_r^n, j_0}^n$ possibly contributing to the event \mathcal{F}_{2r}^c is replaced by a new column in which

every entry belongs to the interval $[\sqrt{\frac{2 \log n}{k}} - a/(4k), \sqrt{\frac{2 \log n}{k}} + a/(4k)]$.

Now if $T_{\mathcal{L}AS} \leq 2r + 1$, then also $\bigcup_{0 \leq t \leq k} (\{T_{\mathcal{L}AS} \leq 2r + 2t - 1\} \cup \mathcal{F}_{2r+2t})$. Otherwise, consider $T_{\mathcal{L}AS} \geq 2r + 2$. In this case, we have a new matrix \mathbf{C}_{2r+2}^n consisting of $\mathbf{C}_{2r+2,1}^n$ and $\mathbf{C}_{2r+2,2}^n$. Note that the event \mathcal{G}_{2r+1} implies that for every subset $I \subset I_r^n$, and for every $j \in \tilde{J}_r^n$, the sum of entries of the sub-column $\mathbf{C}_{I,j}^n$ satisfies

$$\begin{aligned} \text{Sum}(\mathbf{C}_{I,j}^n) &\geq |I| \left(\sqrt{\frac{2 \log n}{k}} - a/(4k) \right) \\ &> |I| \sqrt{\frac{2 \log n}{k}} - a. \end{aligned}$$

In particular, this holds for $I = I_{r+1,1}^n$ and, therefore, j does not satisfy the property (54) with $r + 1$ replacing r . Thus the columns in $\mathbf{C}_{I_{r+1,1}^n, J_{r+1}^n}^n$ satisfying (54) with $r + 1$ replacing r can only be the columns which *were not* replaced in the transition $\mathbf{C}_{2r}^n \rightarrow \mathbf{C}_{2r+1}^n$. Therefore, if the event \mathcal{F}_{2r+2}^c takes place, the columns contributing to this event are from the original columns of \mathbf{C}_{2r}^n .

To complete the proof, we use a similar construction inductively and use the fact that the total number of original columns is at most k , and thus after $2(k + 1)$ iterations all of such columns will be replaced with columns for which (54) cannot occur. Thus assuming the events $\mathcal{G}_{2r}, \dots, \mathcal{G}_{2r+2t-1}$ are defined for some $t \geq 1$, on the event $T_{\mathcal{L}AS} \geq 2r + 2t - 1$ we let

$$\mathcal{G}_{2r+2t} \triangleq \left\| \mathbf{C}_{I_{r+t-1}^n, J_{r+t}^n}^n - \sqrt{\frac{2 \log n}{k}} \right\|_{\infty} \leq \frac{a}{4k},$$

and on the event $T_{\mathcal{L}AS} \geq 2r + 2t$

$$\mathcal{G}_{2r+2t+1} \triangleq \left\| \mathbf{C}_{I_{r+t}^n, \tilde{J}_{r+t}^n}^n - \sqrt{\frac{2 \log n}{k}} \right\|_{\infty} \leq \frac{a}{4k}.$$

Applying Theorem 5.5, we have for $t \geq 0$

$$(58) \quad \liminf_n \mathbb{P}(\mathcal{G}_{2r+2t} | \cdot) \geq \psi_2(a),$$

where \cdot stands for conditioning on $T_{\mathcal{L}AS} \geq 2r + 2t - 1$, $\text{Ave}(\mathbf{C}_{2r-1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a$, as well as

$$(\mathcal{G}_{2r} \cap \dots \cap \mathcal{G}_{2r+2t-1}) \cap (\mathcal{F}_{2r}^c \cap \dots \cap \mathcal{F}_{2r+2t}^c).$$

(Here, for the case $t = 0$ the event above is assumed to be the entire probability space and corresponds to the case considered above.) Similarly, for $t \geq 0$,

$$(59) \quad \liminf_n \mathbb{P}(\mathcal{G}_{2r+2t+1} | \cdot) \geq \psi_2(a),$$

where \cdot stands for the conditioning on $T_{\mathcal{L}AS} \geq 2r + 2t$, $\text{Ave}(\mathbf{C}_{2r-1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a$ as well as

$$(\mathcal{G}_{2r} \cap \dots \cap \mathcal{G}_{2r+2t}) \cap (\mathcal{F}_{2r}^c \cap \dots \cap \mathcal{F}_{2r+2t}^c).$$

By the observation above, since the total number of original columns of \mathbf{C}_{2r-1}^n is k , we have

$$(\mathcal{G}_{2r} \cap \dots \cap \mathcal{G}_{2r+2(k+1)}) \cap (\mathcal{F}_{2r}^c \cap \dots \cap \mathcal{F}_{2r+2(k+1)}^c) = \emptyset.$$

Iterating the relations (58), (59), we conclude that conditional on $T_{\mathcal{L}AS} \geq 2r - 1$ and $\text{Ave}(\mathbf{C}_{2r-1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a$, with probability at least $\psi_2^{2(k+1)}(a)$ the event $\bigcup_{0 \leq t \leq k} (\{T_{\mathcal{L}AS} \leq 2r + 2t - 1\} \cup \mathcal{F}_{2r+2t}^c)$ takes place. This concludes the proof of the proposition. \square

Our next step in proving Theorem 2.1 is to show that if the events $\text{Ave}(\mathbf{C}_{2r-1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a$ and \mathcal{F}_{2r} take place (and in particular $T_{\mathcal{L}AS} \geq 2r$) then with probability bounded away from zero as $n \rightarrow \infty$ the algorithm actually stops in step $2r$: $T_{\mathcal{L}AS} \leq 2r$.

On the event $T_{\mathcal{L}AS} \geq 2r - 1$, the matrix $\mathbf{C}_{I_r^n, J_r^n}^n$ is defined. As earlier, we write the row averages of $\mathbf{C}_{I_r^n, J_r^n}^n$ as

$$\sqrt{\frac{2 \log n}{k}} + C_1^n / (\sqrt{2k \log n}), \dots, \sqrt{\frac{2 \log n}{k}} + C_k^n / (\sqrt{2k \log n}),$$

for the appropriate values C_1^n, \dots, C_k^n . Denote the event $\max_j |C_j^n| \leq \omega_n$ by \mathcal{L}_{2r} . Then by Theorem 5.5,

$$(60) \quad \lim_{n \rightarrow \infty} \mathbb{P} \left(\mathcal{L}_{2r}^c | T_{\mathcal{L}AS} \geq 2r - 1, \text{Ave}(\mathbf{C}_{2r-1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a \right) = 0.$$

This observation will be used for our next result.

PROPOSITION 5.9. *There exists a strictly positive function $\psi_3 : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that for every $r > 0$ and $a > 0$,*

$$\begin{aligned} \liminf_n \mathbb{P} \left(T_{\mathcal{L}AS} \leq 2r | T_{\mathcal{L}AS} \geq 2r, \mathcal{F}_{2r}, \mathcal{L}_{2r}, \text{Ave}(\mathbf{C}_{2r-1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a \right) \\ \geq \psi_3(a). \end{aligned}$$

PROOF. Consider any $k \times k$ matrix C , which is a realization of the matrix $\mathbf{C}_{2r-1}^n - \sqrt{\frac{2 \log n}{k}}$ satisfying $\text{Ave}(C) \geq a$, namely consistent with the event $\text{Ave}(\mathbf{C}_{2r-1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a$. Note that the event $\text{Ave}(\mathbf{C}_{2r-1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a$ implies that at least one of the row averages of \mathbf{C}_{2r-1}^n is also at least $\sqrt{\frac{2 \log n}{k}} + a$. This event and the event \mathcal{L}_{2r} then imply that for large enough n , at least one row of \mathbf{C}_{2r-1}^n will survive till the next iteration $T_{\mathcal{L}AS} = 2r$, provided that this iteration takes place, taking into account the realizations of C_1^n, \dots, C_k^n corresponding to the row averages of $\mathbf{C}_{I_{r-1}^n, J_r^n}^n$.

Now we assume that all of the events $T_{\mathcal{L}AS} \geq 2r, \mathcal{F}_{2r}, \mathcal{L}_{2r}, \text{Ave}(\mathbf{C}_{2r-1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a$ indeed take place. Consider any constant $1 \leq k_1 < k$ and the subset $I \subset I_r^n$ with cardinality k_1 which corresponds to the k_1 largest rows of C with respect to row averages of C (and therefore of \mathbf{C}_{2r-1}^n as well). Let A_1, \dots, A_k be the column sums of the $k_1 \times k$ submatrix of C indexed by the rows I . Assume $A_1, \dots, A_k \geq -a$. Consider the event that $I = I_{2r,1}^n$ corresponds precisely to the rows of \mathbf{C}_{2r-1}^n which survive in the next iteration. Then the column sums of $\mathbf{C}_{2r,1}^n$

are $k_1\sqrt{\frac{2\log n}{k}} + A_j, 1 \leq j \leq k$ consistently with the event \mathcal{F}_{2r} . Note that the lower bound $\text{Ave}(C) \geq a$. The fact that the k_1 row selected are the largest $k_1 \geq 1$ rows in C^n implies

$$(61) \quad \sum_{1 \leq j \leq k} A_j^n \geq k_1 a \geq a.$$

In order for the event above to take place, it should be the case that indeed precisely $k_2 = k - k_1 < k$ rows of $C^n_{I_{r-1}^n, J_r^n}$ will be used in creating C^n_{2r} with the corresponding subset $I_{2r,2}^n, |I_{2r,2}^n| = k_2$. We denote this event by \mathcal{K}_{k_2} . Note that whether this event takes place is completely determined by the realization C corresponding to the matrix C^n_{2r-1} , in particular the realization of the row averages of this matrix, and the realizations C_1, \dots, C_k of C_1^n, \dots, C_k^n corresponding to the row averages of $C^n_{I_{r-1}^n, J_r^n}$. Furthermore, the realizations C, C_1, \dots, C_k determine the values A_1, \dots, A_k .

We write the k column sums of $C^n_{2r,2}$ as $k_2\sqrt{\frac{2\log n}{k}} + U_j^n, 1 \leq j \leq k$. Then the column sums of C^n_{2r} are $\sqrt{2k \log n} + U_j^n + A_j^n, 1 \leq j \leq k$. We claim that for a certain strictly positive function ψ_3 which depends on k only these column sums are all at least $\sqrt{2k \log n} + a/(2k)$:

$$\liminf_n \inf \mathbb{P}(\sqrt{2k \log n} + U_j^n + A_j^n \geq \sqrt{2k \log n} + a/(2k), \\ j = 1, \dots, k | C^n, C_1^n, \dots, C_k^n) \geq \psi_3(a),$$

where \inf is over all sequences C, C_1, \dots, C_k consistent with the events $T_{\mathcal{L}AS} \geq 2r, \mathcal{F}_{2r}, \mathcal{L}_{2r}, \text{Ave}(C^n_{2r-1}) \geq \sqrt{\frac{2\log n}{k}} + a$. We first show how this claim implies the claim of the proposition. The claim implies that conditional on the realizations of C, C_1, \dots, C_k these column sums are at least $\sqrt{2k \log n} + a/(2k)$ with probability $\psi_3(a) - o(1)$. By Theorem 5.5 conditional on C^n_{2r} , the column sums of $C^n_{I_r^n, \tilde{J}_r^n}$ are concentrated around $\sqrt{2k \log n}$ w.h.p. Thus with high probability all columns of C^n_{2r} dominate the columns of $C^n_{I_r^n, \tilde{J}_r^n}$ by at least an additive factor $a/(2k) - o(1)$ and, therefore, algorithm stops at $T_{\mathcal{L}AS} = 2r$. Integrating over $k_2 = 0, \dots, k - 1$ and realizations C, C_1, \dots, C_k consistent with the events $T_{\mathcal{L}AS} \geq 2r, \mathcal{F}_{2r}, \mathcal{L}_{2r}, \text{Ave}(C^n_{2r-1}) \geq \sqrt{\frac{2\log n}{k}} + a$, we obtain the result.

Thus it remains to establish the claim. We have

$$\mathbb{P}(\sqrt{2k \log n} + U_j^n + A_j^n \geq \sqrt{2k \log n} + a/(2k), j = 1, \dots, k | C, C_1, \dots, C_k) \\ = \mathbb{P}(U_j^n + A_j^n \geq a/(2k), j = 1, \dots, k | C, C_1, \dots, C_k).$$

Let $\hat{A}_j^n = \min(A_j^n, 2ka)$. Then

$$\mathbb{P}(U_j^n + A_j^n \geq a/(2k), j = 1, \dots, k | C, C_1, \dots, C_k) \\ \geq \mathbb{P}(U_j^n + \hat{A}_j^n \geq a/(2k), j = 1, \dots, k | C, C_1, \dots, C_k).$$

The event \mathcal{L}_{2r} implies that $\Psi_{n,1}^{\text{Row}}(\mathbf{C}_{\hat{I}_{r-1}, J_r^n}^n) = o(\sqrt{\log n})$, and thus $\Psi_{n,1}^{\text{Row}}(\mathbf{C}_{\hat{I}_{r-1}, J_r^n}^n) / \sqrt{2 \log n} = o(1)$. By a similar reason, $\Psi_{n,2}^{\text{Row}}(\mathbf{C}_{\hat{I}_{r-1}, J_r^n}^n) / \sqrt{2 \log n} = o(1)$ thus implies from (25) that

$$\mathbf{C}_{\hat{I}_{r-1}, J_r^n}^n = \sqrt{\frac{2 \log n}{k}} + \Psi_{n,3}^{\text{Row}}(\mathbf{C}_{\hat{I}_{r-1}, J_r^n}^n) + \Psi_{n,4}^{\text{Row}}(\mathbf{C}_{\hat{I}_{r-1}, J_r^n}^n) + o(1).$$

Then by Theorem 5.5, we have that

$$\lim_{n \rightarrow \infty} \sup_{C, C_1, \dots, C_k} |\mathbb{P}(U_j^n + \hat{A}_j^n \geq a/(2k), j = 1, \dots, k | C, C_1, \dots, C_k) - \mathbb{P}(U_j + \hat{A}_j^n \geq a/(2k), j = 1, \dots, k | \hat{A}_1^n, \dots, \hat{A}_k^n)| = 0,$$

where U_j is the j th column sum of the matrix of the $k_2 \times k$ submatrix of $\text{Col}(\mathbf{C}^k) + \text{ANOVA}(\mathbf{C}^k)$ indexed by $I_{r,2}^n$ and $\sup_{C, C_1, \dots, C_k}$ is over the realizations C, C_1, \dots, C_k consistent with $T_{\mathcal{L}AS} \geq 2r, \mathcal{F}_{2r}, \mathcal{L}_{2r}, \text{Ave}(\mathbf{C}_{2r-1}^n) \geq \sqrt{\frac{2 \log n}{k}} + a$. Thus it suffice to show that

$$\inf_{\hat{A}_1^n, \dots, \hat{A}_k^n} \mathbb{P}(U_j + \hat{A}_j^n \geq a/(2k), j = 1, \dots, k | \hat{A}_1^n, \dots, \hat{A}_k^n) \geq \psi_3(a),$$

for some strictly positive function ψ_3 which depends on k only, where the infimum is over $\hat{A}_1^n, \dots, \hat{A}_k^n$ satisfying $-a \leq \hat{A}_j^n \leq 2ka$ and (61). The joint distribution of $U_j, 1 \leq j \leq k$ is the one of $(\sqrt{k_2}(Z_j - \bar{Z}), 1 \leq j \leq k)$ where Z_1, \dots, Z_k are i.i.d. standard normal and $\bar{Z} = k^{-1} \sum_{1 \leq j \leq k} Z_j$. Thus our goal is to show that

$$\inf_{\hat{A}_1^n, \dots, \hat{A}_k^n} \mathbb{P}(\sqrt{k_2}(Z_j - \bar{Z}) + \hat{A}_j^n \geq a/(2k), 1 \leq j \leq k | \hat{A}_1^n, \dots, \hat{A}_k^n) \geq \psi_3(a),$$

for some ψ_3 . The distribution of the normal $(\sqrt{k_2}(Z_j - \bar{Z}), j = 1, \dots, k)$ vector has a full support on the set $\{x = (x_1, \dots, x_k) : \sum_j x_j = 0\}$.

Consider the set of such vectors $x \in \mathbb{R}^k$ satisfying $\sum_j x_j = 0$ and $x_j + \hat{A}_j^n \geq a/(2k)$. Denote this set by $X(\hat{A}_1^n, \dots, \hat{A}_k^n)$. By (61), we have $\sum_j (a/(2k) - A_j^n) \leq -a/2$. We claim that in fact

$$(62) \quad \sum_j (a/(2k) - \hat{A}_j^n) \leq -a/2 < 0,$$

and thus the set $X(\hat{A}_1^n, \dots, \hat{A}_k^n)$ is nonempty. Indeed, if $A_j^n \leq 2ka$, for all j then $\hat{A}_j^n = A_j^n$ and assertion holds from (61). Otherwise, if $A_{j_0}^n > 2ka$ for some j_0 , then since $A_j^n \geq -a$ and, therefore, $\hat{A}_j^n \geq -a$, we have

$$\sum_j (a/(2k) - \hat{A}_j^n) \leq a/2 - 2ka + (k - 1)a < -ka < -a/2 < 0.$$

In fact, since $a > 0$, the set $X(\hat{A}_1^n, \dots, \hat{A}_k^n)$ has a nonempty interior, and thus a positive measure with respect to the induced Lebesgue measure of the subset $\{x = (x_1, \dots, x_k) : \sum_j x_j = 0\} \subset \mathbb{R}^k$. As a result, the probability

$$\mathbb{P}((\sqrt{k_2}(Z_j - \bar{Z}), 1 \leq j \leq k) \in X(\hat{A}_1^n, \dots, \hat{A}_k^n) | \hat{A}_1^n, \dots, \hat{A}_k^n)$$

is strictly positive. This probability is a continuous function of $\hat{A}_1^n, \dots, \hat{A}_k^n$ which belongs to the bounded interval $[-a, 2ka]$. By the compactness argument, we then obtain

$$\inf \mathbb{P}((\sqrt{k_2}(Z_j - \bar{Z}), 1 \leq j \leq k) \in X(\hat{A}_1^n, \dots, \hat{A}_k^n) | A_1^n, \dots, A_k^n) > 0,$$

where the infimum is over $-a \leq \hat{A}_1^n, \dots, \hat{A}_k^n \leq 2ka$ satisfying (62). Denoting the infimum by $\psi_3(a)$, we obtain the result. \square

We now synthesize Propositions 5.7, 5.8 and 5.9 to obtain the following corollary.

COROLLARY 5.10. *There exists a strictly positive function ψ_4 , which depends on k only such that for every $r > k + 2$ and $a > 0$,*

$$\liminf_n \mathbb{P}(T_{\mathcal{L}AS} \leq 2r | T_{\mathcal{L}AS} \geq 2r - 2k - 3) \geq \psi_4(a).$$

PROOF. By Proposition 5.7, we have

$$\begin{aligned} \liminf_n \mathbb{P} \left(\text{Ave}(\mathbf{C}_{2r-2k-1}) \geq \sqrt{\frac{2 \log n}{k}} + a \right. \\ \left. \cup \{T_{\mathcal{L}AS} \leq 2r - 2k - 2\} | T_{\mathcal{L}AS} \geq 2r - 2k - 3 \right) \geq \psi_1(a). \end{aligned}$$

Combining with Proposition 5.8, we obtain that there exists $t, 0 \leq t \leq k$ such that

$$\begin{aligned} \liminf_{n \rightarrow \infty} \mathbb{P} \left(\{T_{\mathcal{L}AS} \leq 2r - 2t - 1\} \right. \\ \left. \cup \left(\mathcal{F}_{2r-2t} \cap \text{Ave}(\mathbf{C}_{2r-2t-1}) \geq \sqrt{\frac{2 \log n}{k}} + a \right) | T_{\mathcal{L}AS} \geq 2r - 2k - 3 \right) \\ \geq (k + 1)^{-1} \psi_1(a) \psi_2^{2(k+1)}(a). \end{aligned}$$

By observation (60), we also obtain

$$\begin{aligned} \liminf_{n \rightarrow \infty} \mathbb{P} \left(\{T_{\mathcal{L}AS} \leq 2r - 2t - 1\} \right. \\ \left. \cup \left(\mathcal{F}_{2r-2t} \cap \text{Ave}(\mathbf{C}_{2r-2t-1}) \geq \sqrt{\frac{2 \log n}{k}} + a \cap \mathcal{L}_{2r-2t} \right) | T_{\mathcal{L}AS} \right) \end{aligned}$$

$$\begin{aligned} &\geq 2r - 2k - 3 \Big) \\ &\geq (k + 1)^{-1} \psi_1(a) \psi_2^{2(k+1)}(a). \end{aligned}$$

Finally, applying Lemma 5.9 we obtain

$$\begin{aligned} &\liminf_{n \rightarrow \infty} \mathbb{P}(\{T_{\mathcal{L}AS} \leq 2r - 2t\} | T_{\mathcal{L}AS} \geq 2r - 2k - 3) \\ &\geq (k + 1)^{-1} \psi_1(a) \psi_2^{2(k+1)}(a) \psi_3(a), \end{aligned}$$

implying by monotonicity the same result for $T_{\mathcal{L}AS} \leq 2r$. Letting $\psi_4(a) \triangleq (k + 1)^{-1} \psi_1(a) \psi_2^{2(k+1)}(a) \psi_3(a)$, we obtain the result. \square

We are now ready to complete the proof of Theorem 2.1.

PROOF OF THEOREM 2.1. Given $\varepsilon > 0$, we fix arbitrary $a > 0$ and find $r = r(\varepsilon, a)$ large enough so that $(1 - \psi_4(a))^r < \varepsilon$. Applying Corollary 5.10, we obtain for $N = r(2k + 4)$,

$$\begin{aligned} \mathbb{P}(T_{\mathcal{L}AS} \geq N) &= \prod_{1 \leq t \leq r} \mathbb{P}(T_{\mathcal{L}AS} \geq t(2k + 4) | T_{\mathcal{L}AS} \geq (t - 1)(2k + 4)) \\ &\leq (1 - \psi_4(a))^r \\ &\leq \varepsilon, \end{aligned}$$

which gives the first part of Theorem 2.1. We now show (2). Fix $\varepsilon > 0$. We have

$$\begin{aligned} &\mathbb{P}\left(\left| \text{Ave}(\mathbf{C}_{T_{\mathcal{L}AS}}^n) - \sqrt{\frac{2 \log n}{k}} \right| > \omega_n\right) \\ &\leq \mathbb{P}\left(\left| \text{Ave}(\mathbf{C}_{T_{\mathcal{L}AS}}^n) - \sqrt{\frac{2 \log n}{k}} \right| > \omega_n, T_{\mathcal{L}AS}^n \leq N\right) + \mathbb{P}(T_{\mathcal{L}AS}^n > N) \\ &\leq \mathbb{P}\left(\left| \text{Ave}(\mathbf{C}_{T_{\mathcal{L}AS}}^n) - \sqrt{\frac{2 \log n}{k}} \right| > \omega_n, T_{\mathcal{L}AS}^n \leq N\right) + \varepsilon \\ &= \sum_{1 \leq r \leq N} \mathbb{P}\left(\left| \text{Ave}(\mathbf{C}_r^n) - \sqrt{\frac{2 \log n}{k}} \right| > \omega_n, T_{\mathcal{L}AS}^n = r\right) + \varepsilon \\ &\leq \sum_{1 \leq r \leq N} \mathbb{P}\left(\left| \text{Ave}(\mathbf{C}_r^n) - \sqrt{\frac{2 \log n}{k}} \right| > \omega_n, T_{\mathcal{L}AS}^n \geq r\right) + \varepsilon. \end{aligned}$$

By part (c) of Theorem 5.5, we have for every r

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left| \text{Ave}(\mathbf{C}_r^n) - \sqrt{\frac{2 \log n}{k}} \right| > \omega_n, T_{\mathcal{L}AS}^n \geq r\right) = 0.$$

We conclude that for every ε

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \text{Ave}(\mathbf{C}_{T_{\mathcal{L}AS}}^n) - \sqrt{\frac{2 \log n}{k}} \right| > \omega_n \right) \leq \varepsilon.$$

Since the left-hand side does not depend on ε , we obtain (2). This concludes the proof of Theorem 2.1. \square

6. Conclusions and open questions. We close the paper with several open questions for further research. In light of the new algorithm \mathcal{IGP} which improves upon the \mathcal{LAS} algorithm by a multiplicative factor $4/3$, a natural direction is to obtain a better performing polynomial time algorithm. It would be especially interesting if such an algorithm can improve upon the $5\sqrt{2}/3\sqrt{3}$ threshold since it would then indicate that the OGP is not an obstacle for polynomial time algorithms. Improving the $5\sqrt{2}/3\sqrt{3}$ threshold perhaps by considering multi-overlaps of matrices with a fixed asymptotic average value, is another important challenge. Based on such improvements obtained for the independent sets in sparse random graphs [16] and for the random satisfiability (random NAE-K-SAT) problem [10], it is very plausible that such an improvement is achievable.

Studying the maximum submatrix problem for non-Gaussian distribution is another interesting direction, especially for distributions with tail behavior different from the one of the normal distribution, namely for not sub-Gaussian distributions. Heavy tail distributions are of particular interest for this problem.

Finally, a very interesting version of the maximum submatrix problem is the sparse Principal Component Analysis (PCA) problem for sample covariance data. Suppose, X_i , $1 \leq i \leq n$ are p -dimensional uncorrelated random variables (say Gaussian), and let Σ be the corresponding sample covariance matrix. When the dimension p is comparable with n , the distribution of Σ exhibits a nontrivial behavior. For example, the limiting distribution of the spectrum is described by the Marcenko–Pastur law as opposed to the “true” underlying covariance matrix which is just the identity matrix. The sparse PCA problem is the maximization problem $\max \beta^T \Sigma \beta$ where the maximization is over p -dimensional vectors β with $\|\beta\|_2 = 1$ and $\|\beta\|_0 = k$, where $\|a\|_0$ is the number of nonzero components of the vector a (sparsity). What is the limiting distribution of the objective value and what is the algorithmic complexity of this problem? What is the solution space geometry of this problem and in particular, does it exhibit the OGP? The sparse PCA problem has received an attention recently in the hypothesis testing version [4], [5], where it was shown that for a certain parameter regime, detecting the sparse PCA signal is hard, provided the so-called hidden clique problem in the theory of random graphs is hard [3]. Here, we propose to study the problem from the estimation point of view—computing the distribution of the k -dominating principal components and studying the algorithmic hardness of this problem.

Finally, a bigger challenge is to either establish that the problems exhibiting the OGP are indeed algorithmically hard and do not admit polynomial time algorithms,

or constructing an example where this is not the case. In light of the repeated failure to improve upon the important special case of this problem—the maximum clique in the Erdős–Rényi graph $\mathbb{G}(n, p)$, this challenge might be out of reach for the existing methods of analysis.

APPENDIX A: PROOF OF LEMMA 3.1

We have

$$\begin{aligned}
 & \mathbb{P}\left(\left|\sqrt{2\log n}\left(\max_{1\leq i\leq n} Z_i - b_n\right)\right| \leq \frac{3}{2}\log\log n\right) \\
 &= \mathbb{P}\left(\max_{1\leq i\leq n} Z_i \leq \frac{3}{2}\log\log n/\sqrt{2\log n} + b_n\right) \\
 (63) \quad & - \mathbb{P}\left(\max_{1\leq i\leq n} Z_i < -\frac{3}{2}\log\log n/\sqrt{2\log n} + b_n\right) \\
 &= \mathbb{P}\left(Z_1 \leq \frac{3}{2}\log\log n/\sqrt{2\log n} + b_n\right)^n \\
 & - \mathbb{P}\left(Z_1 < -\frac{3}{2}\log\log n/\sqrt{2\log n} + b_n\right)^n.
 \end{aligned}$$

Next we use (1) to approximate

$$\begin{aligned}
 & \mathbb{P}\left(Z_1 \leq \frac{3}{2}\log\log n/\sqrt{2\log n} + b_n\right) \\
 &= 1 - (1 + o(1))\frac{1}{\left(\frac{3}{2}\log\log n/\sqrt{2\log n} + b_n\right)\sqrt{2\pi}} \\
 (64) \quad & \times \exp\left(-\frac{\left(\frac{3}{2}\log\log n/\sqrt{2\log n} + b_n\right)^2}{2}\right) \\
 &= 1 - \Theta\left(\frac{1}{n(\log n)^{3/2}}\right),
 \end{aligned}$$

and

$$\begin{aligned}
 & \mathbb{P}\left(Z_1 < -\frac{3}{2}\log\log n/\sqrt{2\log n} + b_n\right) \\
 &= 1 - (1 + o(1))\frac{1}{\left(-\frac{3}{2}\log\log n/\sqrt{2\log n} + b_n\right)\sqrt{2\pi}} \\
 (65) \quad & \times \exp\left(-\frac{\left(-\frac{3}{2}\log\log n/\sqrt{2\log n} + b_n\right)^2}{2}\right) \\
 &= 1 - \Theta\left(\frac{(\log n)^{3/2}}{n}\right).
 \end{aligned}$$

Now we substitute (64) and (65) into (63)

$$\begin{aligned} \mathbb{P}\left(\left|\sqrt{2 \log n}\left(\max_{1 \leq i \leq n} Z_i - b_n\right)\right| \leq \log \log n\right) &= \left(1 - \Theta\left(\frac{1}{n(\log n)^{3/2}}\right)\right)^n - \left(1 - \Theta\left(\frac{(\log n)^{3/2}}{n}\right)\right)^n \\ &= \left(1 - \Theta\left(\frac{1}{n(\log n)^{3/2}}\right)\right)^n - \exp(-\Theta((\log n)^{3/2})). \end{aligned}$$

Then the result follows from choosing a positive integer N and a constant $c > 0$ such that for all $n > N$ the following inequality holds:

$$\left(1 - \Theta\left(\frac{1}{n(\log n)^{3/2}}\right)\right)^n - \exp(-\Theta((\log n)^{3/2})) \geq 1 - c \frac{1}{(\log n)^{1.5}}.$$

APPENDIX B: DERIVATION OF TWO PHASE TRANSITION POINTS

$$\alpha_1^* = \sqrt{3}/\sqrt{2} \text{ AND } \alpha_2^* = 5\sqrt{2}/(3\sqrt{3})$$

We start with α_1^* , which we define as a critical point such that for any $\alpha > \alpha_1^*$ and $\alpha \in (0, \sqrt{2})$, $\mathcal{R}(\alpha)$ does not cover the whole region $[0, 1]^2$, that is, $[0, 1]^2 \setminus \mathcal{R}(\alpha) \neq \emptyset$. We formulate this as follows:

$$(66) \quad \alpha_1^* \triangleq \max\left\{\alpha \in (0, \sqrt{2}) : \min_{y_1, y_2 \in [0, 1]^2} f(\alpha, y_1, y_2) \geq 0\right\}.$$

Since $f(\alpha, y_1, y_2)$ is differentiable with respect to y_1 and y_2 , the minimum of $f(\alpha, y_1, y_2)$ for a fixed α appears either at the boundaries or the stationary points. Using the symmetry of y_1 and y_2 , we only need to consider the following boundaries:

$$\{(y_1, y_2) : y_1 = 0, y_2 \in [0, 1]\} \cup \{(y_1, y_2) : y_1 = 1, y_2 \in [0, 1]\}.$$

By inspection, $\min_{y_1=0, y_2 \in [0, 1]} f(\alpha, y_1, y_2) = 3 - 2\alpha^2$ and

$$\min_{y_1=1, y_2 \in [0, 1]} f(\alpha, y_1, y_2) = \min_{y_2 \in [0, 1]} \left\{3 - y_2 - \frac{2}{1 + y_2} \alpha^2\right\}.$$

Since the objective function above is a concave function with respect to y_2 , its minimum is obtained at $y_2 = 0$ or 1 , which is $3 - 2\alpha^2$ or $2 - \alpha^2$. Hence the minimum of $f(\alpha, y_1, y_2)$ at the boundaries above is either $3 - 2\alpha^2$ or $2 - \alpha^2$. Both of them being nonnegative requires

$$3 - 2\alpha^2 \geq 0 \quad \text{and} \quad 2 - \alpha^2 \geq 0 \quad \text{and} \quad \alpha \in (0, \sqrt{2}),$$

$$\text{which implies } \alpha \in (0, \sqrt{3}/\sqrt{2}].$$

Next we consider the stationary points of $f(\alpha, y_1, y_2)$ for a fixed α . The stationary points are determined by solving

$$\begin{aligned} \frac{\partial f(\alpha, y_1, y_2)}{\partial y_1} &= 0, & \text{which implies } -1 + \frac{2\alpha^2 y_2}{(1 + y_1 y_2)^2} &= 0, \\ \frac{\partial f(\alpha, y_1, y_2)}{\partial y_2} &= 0, & \text{which implies } -1 + \frac{2\alpha^2 y_1}{(1 + y_1 y_2)^2} &= 0. \end{aligned}$$

Observe from above that $y_1 = y_2$. Then we can simplify the equations above as follows:

$$(67) \quad y_1^4 + 2y_1^2 - 2\alpha^2 y_1 + 1 = 0.$$

Using ‘‘Mathematica’’, we find that the four solutions for the quartic equation above for $\alpha^2 = 3/2$ are complex numbers all with nonzero imaginary parts. Since the equation above does not have real solutions, the optimization problem (66) has maximum at $\alpha = \sqrt{3}/\sqrt{2}$. On the other hand, for any $\alpha > \sqrt{3}/\sqrt{2}$, $f(\alpha, 1, 0) = 3 - 2\alpha^2$ is always negative. Hence we have $\alpha_1^* = \sqrt{3}/\sqrt{2}$.

We also claim that for any $\alpha \in (0, \sqrt{3}/\sqrt{2})$, $\mathcal{R}(\alpha) = [0, 1]^2$. It suffices to show that for any $y \in [0, 1]$,

$$y^4 + 2y^2 - 2\alpha^2 y + 1 > 0.$$

Suppose for $\alpha \in (0, \sqrt{3}/\sqrt{2})$ there is a $\hat{y} \in [0, 1]$ such that $\hat{y}^4 + 2\hat{y}^2 - 2\alpha^2 \hat{y} + 1 \leq 0$. Then by $\alpha^2 < 3/2$ and $\hat{y} \neq 0$, we have

$$\hat{y}^4 + 2\hat{y}^2 - 3\hat{y} + 1 < 0.$$

Since $y^4 + 2y^2 - 3y + 1$ is positive at $y = 0$ and negative at \hat{y} , the continuity of $y^4 + 2y^2 - 3y + 1$ implies that there is a $y_1 \in [0, 1]$ such that (67) holds for $\alpha^2 = 3/2$, which is a contradiction. The claim follows.

Next we introduce α_2^* . Increasing α beyond α_1^* , we are interested in the first point α_2^* at which the function $f(\alpha_2^*, y_1, y_2)$ has at least one real stationary point and the value of $f(\alpha_2^*, y_1, y_2)$ at this point is zero. Observe that at the stationary points $y_1 = y_2$ and y_1 satisfies (67). Then α_2^* is determined by solving

$$\begin{aligned} y_1^4 + 2y_1^2 - 2\alpha^2 y_1 + 1 &= 0, \\ 4 - 2y_1 - \frac{2}{1 + y_1^2} \alpha^2 &= 0, \quad y_1 \in [0, 1], \alpha \in (\sqrt{3}/\sqrt{2}, \sqrt{2}). \end{aligned}$$

Using ‘‘Mathematica’’ to solve the equations above, we obtain only one real solution $y_1 = 1/3, \alpha = 5\sqrt{2}/(3\sqrt{3})$. Then we have $\alpha_2^* = 5\sqrt{2}/(3\sqrt{3})$ and $f(\alpha_2^*, 1/3, 1/3) = 0$. We verify that $f(\alpha_2^*, 1/3, y_2) < 0$ for $y_2 \in [0, 1] \setminus \{1/3\}$ and $f(\alpha_2^*, y_1, 1/3) < 0$ for $y_1 \in [0, 1] \setminus \{1/3\}$. By plotting $f(\alpha_2^*, y_1, y_2)$ in Figure 6, we see that the set $\mathcal{R}(\alpha_2^*)$ is connected through a single point $(1/3, 1/3)$.

REFERENCES

- [1] ACHLIOPTAS, D. and COJA-OGHLAN, A. (2008). Algorithmic barriers from phase transitions. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science* 793–802. IEEE, New York.
- [2] ACHLIOPTAS, D., COJA-OGHLAN, A. and RICCI-TERSENGHI, F. (2011). On the solution-space geometry of random constraint satisfaction problems. *Random Structures Algorithms* **38** 251–268. [MR2663730](#)
- [3] ALON, N., KRIVELEVICH, M. and SUDAKOV, B. (1998). Finding a large hidden clique in a random graph. *Random Structures Algorithms* **13** 457–466.
- [4] BERTHET, Q. and RIGOLLET, P. (2013). Complexity theoretic lower bounds for sparse principal component detection. In *Conference on Learning Theory* 1046–1066.
- [5] BERTHET, Q. and RIGOLLET, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.* **41** 1780–1815. [MR3127849](#)
- [6] BHAMIDI, S., DEY, P. S. and NOBEL, A. B. (2012). Energy landscape for large average submatrix detection problems in Gaussian random matrices. Preprint. Available at [arXiv:1211.2284](#).
- [7] COJA-OGHLAN, A. and EFTHYMIU, C. (2011). On independent sets in random graphs. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms* 136–144. SIAM, Philadelphia.
- [8] FORTUNATO, S. (2010). Community detection in graphs. *Phys. Rep.* **486** 75–174.
- [9] GAMARNIK, D. and SUDAN, M. (2014). Limits of local algorithms over sparse random graphs. In *Proceedings of the 5th Conference on Innovations in Theoretical Computer Science* 369–376. ACM, New York.
- [10] GAMARNIK, D. and SUDAN, M. (2014). Performance of the survey propagation-guided decimation algorithm for the random NAE-K-SAT problem. Preprint. Available at [arXiv:1402.0052](#).
- [11] GAMARNIK, D. and ZADIK, I. (2017). High-dimensional regression with binary coefficients. Estimating squared error and a phase transition. Preprint. Available at [arXiv:1701.04455](#).
- [12] KARP, R. M. (1976). The probabilistic analysis of some combinatorial search algorithms. In *Algorithms and complexity: New directions and recent results* 1–19. [MR0445898](#)
- [13] LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York. [MR0691492](#)
- [14] MADEIRA, S. C. and OLIVEIRA, A. L. (2004). Biclustering algorithms for biological data analysis: A survey. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **1** 24–45.
- [15] MONTANARI, A. (2015). Finding one community in a sparse graph. *J. Stat. Phys.* **161** 273–299. [MR3401018](#)
- [16] RAHMAN, M. and VIRÁG, B. (2014). Local algorithms for independent sets are half-optimal. Preprint. Available at [arXiv:1402.0485](#).
- [17] SHABALIN, A. A., WEIGMAN, V. J., PEROU, C. M. and NOBEL, A. B. (2009). Finding large average submatrices in high dimensional data. *Ann. Appl. Stat.* 985–1012. [MR2750383](#)
- [18] SUN, X. and NOBEL, A. B. (2013). On the maximal size of large-average and ANOVA-fit submatrices in a Gaussian random matrix. *Bernoulli* **19** 275–294. [MR3019495](#)

SLOAN SCHOOL OF MANAGEMENT, E62-563
 MASSACHUSETTS INSTITUTE OF TECHNOLOGY
 100 MAIN STREET
 CAMBRIDGE, MASSACHUSETTS 02140
 USA
 E-MAIL: gamarnik@mit.edu

DEPARTMENT OF ELECTRICAL ENGINEERING
 AND COMPUTER SCIENCE
 MASSACHUSETTS INSTITUTE OF TECHNOLOGY
 77 MASSACHUSETTS AVENUE
 CAMBRIDGE, MASSACHUSETTS 02139
 USA
 E-MAIL: quanli@mit.edu