JMLR: Workshop and Conference Proceedings vol 40:1–40, 2015 28th Annual Conference on Learning Theory

Improved Sum-of-Squares Lower Bounds for Hidden Clique and Hidden Submatrix Problems

Yash Deshpande YASH.DESHPANDE@STANFORD.EDU

Andrea Montanari MONTANARI@STANFORD.EDU

Stanford University.

Editors: Elad Hazan and Peter Grünwald

Abstract

Given a large data matrix $A \in \mathbb{R}^{n\times n}$, we consider the problem of determining whether its entries are i.i.d. from some known marginal distribution $A_{ij} \sim P_0$, or instead A contains a principal submatrix $A_{\mathsf{Q},\mathsf{Q}}$ whose entries have marginal distribution $A_{ij} \sim P_1 \neq P_0$. As a special case, the hidden (or planted) clique problem is finding a planted clique in an otherwise uniformly random graph.

Assuming unbounded computational resources, this hypothesis testing problem is statistically solvable provided $|\mathsf{Q}| \geq C \log n$ for a suitable constant C. However, despite substantial effort, no polynomial time algorithm is known that succeeds with high probability when $|\mathsf{Q}| = o(\sqrt{n})$. Recently, Meka and Wigderson (2013) proposed a method to establish lower bounds for the hidden clique problem within the Sum of Squares (SOS) semidefinite hierarchy.

Here we consider the degree-4 SOS relaxation, and study the construction of Meka and Wigderson (2013) to prove that SOS fails unless $k \geq C\, n^{1/3}/\log n$. An argument presented by Barak (2014) implies that this lower bound cannot be substantially improved unless the witness construction is changed in the proof. Our proof uses the moment method to bound the spectrum of a certain random association scheme, i.e. a symmetric random matrix whose rows and columns are indexed by the edges of an Erdös-Renyi random graph.

1. Introduction

Characterizing the computational complexity of statistical estimation and statistical learning problems is an outstanding challenge. On one hand, a large part of research in this area focuses on the analysis of specific polynomial-time algorithms, thereby providing upper bounds on the problem complexity. On the other hand, information-theoretic techniques are used to derive fundamental limits beyond which no algorithm can solve the statistical problem under study. While in some cases algorithmic and information-theoretic bounds match, in many other examples a large gap remains in which the problem is solvable assuming unbounded computational resources but simple algorithms fail. The hidden clique and hidden submatrix problems are prototypical examples of this category.

In the hidden submatrix problem, we are given a symmetric data matrix $A \in \mathbb{R}^{n\times n}$ and two probability distributions P_0 and P_1 on the real line, with $\mathbb{E}_{P_0}\{X\} = 0$ and $\mathbb{E}_{P_1}\{X\} = \mu > 0$. We want to distinguish between two hypotheses (we set by convention $A_{ii} = 0$ for all $i \in [n] = \{1, 2, \dots, n\}$):

Hypothesis H_0: The entries of A above the diagonal $(A_{ij})_{i<j}$ are independent and identically distributed (i.i.d.) random variables with the same marginal law $A_{ij} \sim P_0$.

Hypothesis H_1: Given a (hidden) subset $\mathsf{Q} \subseteq [n]$ the entries $(A_{ij})_{i<j}$ are independent with

$$A_{ij} \sim \begin{cases} P_1 & \text{if } \{i,j\} \subset \mathsf{Q}, \\ P_0 & \text{otherwise.} \end{cases} \tag{1}$$

Further, Q is a uniformly random subset conditional on its size, that is fixed $|\mathsf{Q}| = k$.

The estimation version of this problem is also of interest, wherein the special subset Q is known to exist, and an algorithm is sought that identifies Q with high probability.

This model encapsulates the basic computational challenges underlying a number of problems in which we need to estimate a matrix that is both sparse and low-rank. Such problems arise in various fields such as genomics, signal processing, social network analysis, and machine learning (Shabalin et al. (2009); Johnstone and Lu (2009); Oymak et al. (2012)).

The hidden clique (or 'planted clique') problem (Jerrum (1992)) is a special case of the above setting, and has attracted considerable interest within theoretical computer science. Let δ_x denote the Dirac delta distribution at the point $x \in \mathbb{R}$. The hidden clique problem corresponds to the distributions

$$P_1 = \delta_{+1}\,, \quad P_0 = \frac{1}{2}\,\delta_{+1} + \frac{1}{2}\,\delta_{-1}\,. \tag{2}$$

In this case, the data matrix A can be interpreted as the adjacency matrix of a graph G over n vertices (wherein $A_{ij} = +1$ encodes presence of edge $\{i,j\}$ in G, and $A_{ij} = -1$ its absence). Under hypothesis H_1, the set Q induces a clique in the (otherwise) random graph G. For the rest of this introduction, we shall focus on the hidden clique problem, referring to Section 2 for a formal statement of our general results.

The largest clique in a uniformly random graph has size $2\log_2 n + o(\log n)$, with high probability (Grimmett and McDiarmid (1975)). Thus, allowing for exhaustive search, the hidden clique problem can be solved when $k \geq (2+\varepsilon)\log_2 n$. On the other hand, despite significant efforts (Alon et al. (1998); Ames and Vavasis (2011); Dekel et al. (2011); Feige and Ron (2010); Deshpande and Montanari (2014)), no polynomial time algorithm is known to work when $k = o(\sqrt{n})$. As mentioned above, this is a prototypical case for which a large gap exists between performances of well-understood polynomial-time algorithms, and the ultimate information-theoretic (or statistical) limits. This has motivated an ongoing quest for computational lower bounds.

At first sight this appears to be straightforward: finding the maximum clique in a graph is a classical NP-hard problem (Karp (1972)). Even a very rough approximation to its size is hard to find (Hastad (1996); Khot (2001)). In particular, it is hard to detect the presence of a clique of size $n^{1-\varepsilon}$ in a graph with n vertices.

However, as is well-known, these worst-case hardness results do not imply computational lower bounds when problem instances are distributed according to a natural statistical model. Over the last two years there have been fascinating advances in crafting careful reductions that preserve the instance distribution in specific cases (Berthet and Rigollet (2013); Ma and Wu (2013); Chen and Xu (2014); Hajek et al. (2014); Cai et al. (2015)). This line of work typically establishes that several detection problems (sparse PCA, hidden submatrix, hidden community) are at least as hard as the hidden clique problem with $k = o(\sqrt{n})$. This approach has two limitations:

(i) It yields conditional statements relying on the unproven assumption that the hidden clique problem is hard. In absence of any 'completeness' result, this is a strong assumption that calls for further scrutiny.

(ii) Reductions among instance distributions are somewhat fragile with respect changes in the distribution. For instance, it is not known whether the hidden submatrix problem with Gaussian distributions $P_0 = \mathsf{N}(0,1)$ and $P_1 = \mathsf{N}(\mu,1)$ is at least as hard as the hidden clique problem, although a superficial look might suggest that they are very similar[1].

A complementary line of attack consists in proving unconditional lower bounds for broad classes of algorithms. In an early contribution, Jerrum (1992) established such a lower bound for a class of Markov Chain Monte Carlo methods. Feldman et al. (2012) considered a query-based formulation of the problem and proved a similar result for 'statistical algorithms.' Closer to the present paper is the work of Feige and Krauthgamer (2000), who analyzed the Lovász-Schrijver semidefinite programming (SDP) hierarchy. Remarkably, these authors proved that r rounds of this hierarchy (with complexity $n^{O(r)}$) fail to detect the hidden clique unless $k \gtrsim \sqrt{n}/2^r$. (Here and below we write $f(n,r,\dots) \gtrsim g(n,r,\dots)$ if there exists a constant C such that $f(n,r,\dots) \geq C\, g(n,r,\dots)$.)

While this failure of the Lovász-Schrijver hierarchy provides insightful evidence towards the hardness of the hidden-clique problem, an even stronger indication could be obtained by establishing an analogous result for the Sum of Squares (SOS) hierarchy (Shor (1987); Lasserre (2001); Parrilo (2003)). This SDP hierarchy unifies most convex relaxations developed for a variety of combinatorial optimization problems. Its close connection with the unique games conjecture has led to the idea that SOS might indeed be an 'optimal' algorithm for a broad class of problems (Barak and Steurer (2014)). Furthermore, many of the low-rank estimation problems mentioned above include naturally quadratic constraints, that are most naturally expressed within the SOS hierarchy.

The SOS hierarchy is formulated in terms of a sequence of polynomial optimization problems. The level of a relaxation in the hierarchy corresponds to the largest degree d of any monomial whose value is explicitly treated as a decision variable. Meka and Wigderson (2013) proposed a construction of a sequence of feasible solutions, or witnesses (one for each degree d), that can be used to prove lower bounds for the hidden clique problem within the SOS hierarchy. The key technical step consisted in proving that a certain moment matrix is positive semidefinite: unfortunately this part of their proof contained a fatal flaw.

In the present paper we undertake the more modest task of analyzing the Meka-Wigderson witness for the level $d = 4$ of the SOS hierarchy. This is the first level at which the SOS hierarchy differs substantially from the baseline spectral algorithm of Alon et al. (1998), or from the Lovász-Schrijver hierarchy. We prove that this relaxation fails unless

$$k \gtrsim \frac{n^{1/3}}{\log n}. \tag{3}$$

Notice that the natural guess would be that the SOS hierarchy fails (for any bounded d) whenever $k = o(\sqrt{n})$. While our result falls short of establishing this, an argument presented in Barak (2014) shows that this is a limitation of the Meka-Wigderson construction. In other words, the bound of Eq. (3) can be improved (by more than a logarithmic factor) only by changing the witness construction of Meka-Wigderson that we and Meka et al. (2015) have used.

1. We note here that Ma and Wu (2013) establish this for $\mu = o(\sqrt{\log n})$, whereas one would expect the reduction to hold also for $\mu = O(1)$.

Apart from the lower bound on the hidden clique problem, our analysis provides two additional sets of results:

- We apply a similar witness construction to the hidden submatrix problem with entries distributions $P_0 = \mathsf{N}(0,1)$, $P_1 = \mathsf{N}(\mu,1)$. We define a polynomial-time computable statistical test that is based on a degree-4 SOS relaxation of a nearly optimal combinatorial test. We show that this fails unless $k \gtrsim \mu^{-1} n^{1/3}/\log n$.

- As mentioned above, the main technical contribution consists in proving that a certain random matrix is (with high probability) positive semidefinite. Abstractly, the random matrix in question is function of an underlying (Erdös-Renyi) random graph G over n vertices. The matrix has rows/columns indexed by subsets of size at most $d/2 = 2$, and elements depending by the subgraphs of G induced by those subsets. We shall loosely refer to this type of random matrix as to a *random association scheme*.

  In order to prove that this witness is positive semidefinite, we decompose the linear space on which it acts into the irreducible subrepresentations of the group of permutations over n objects. We then use the moment method to characterize each submatrix defined by this decomposition, and combine the results to obtain our final condition for positivity.

  We believe that both the matrix definition and the proof technique are so natural that they are likely to be useful in related problems.

- As an illustration of the last point, our analysis covers the case of Erdös-Renyi graphs with sublinear average degree (namely, with average degree of order n^{1-a}, $a < 1/12$). In particular, it is easy to derive sum-of-squares lower bounds for finding cliques in such graphs from our main theorem.

The rest of the paper is organized as follows. In Section 2 we state our main technical result, which concerns the spectrum of random association schemes. We then show that it implies lower bounds for the hidden clique and hidden submatrix problem. Section 3 presents a brief outline of the proof. Finally, Section A presents the proof of our main technical result.

While this paper was being written, we became aware through Barak (2014) that Meka et al. (2015) proved that the degree-d SOS relaxation is unsuccessful unless $k \gtrsim n^{1/d}$ for arbitrary, constant d. Their work follows the path of Meka and Wigderson (2013), wherein the proof of positivity of the moment matrix is achieved by a trace method calculation similar to ours. However, as explained in the proof strategy in Section 3 below, their analysis does not account for certain spectral properties of the moment matrix. Ultimately, this results in establishing a lower bound of $n^{1/4}$ for $d = 4$, which is slightly weaker than our result. However, of course, Meka et al. (2015) are able to handle $d > 4$, which we do not consider.

2. Main results

In this section we present our results. Subsection 2.1 introduces a feasible random association scheme that is a slight generalization of the witness developed in Meka and Wigderson (2013) (for the degree $d = 4$ SOS). We state conditions implying that this matrix is positive semidefinite with high probability. These conditions are in fact obtained by specializing a more general result stated in Proposition 6. We then derive implications for hidden cliques and hidden submatrices.

2.1. Positivity of the Meka-Wigderson witness

We will denote by $\mathbb{G}(n,p)$ the undirected Erdös-Renyi random graph model on n vertices, with edge probability p. A graph $G=(V,E)\sim\mathbb{G}(n,p)$ has vertex set $V=[n]\equiv\{1,2,\dots,n\}$, and edges set E defined by letting, for each $i<j\in[n]$, $\{i,j\}\in E$ independently with probability p.

The random association scheme $M=M(G,\underline{\alpha})$ can be thought as a parametric generalization of the adjacency matrix of G, depending on the graph G and parameters $\underline{\alpha}=(\alpha_1,\alpha_2,\alpha_3,\alpha_4)\in\mathbb{R}^4$. In order to define the matrix M we first need to set up some notation. For an integer r, we let $\binom{[n]}{r}$ denote the set of all subsets of $[n]$ of size *exactly* r, and $\binom{[n]}{\le r}$ denote the set of all subsets of size *at most* r. We also let $\emptyset$ denote the empty set.

We shall often identify the collections of subsets of size one, $\binom{[n]}{1}=\{\{i\}:i\in[n]\}$ with $[n]$. Also, we identify $\binom{[n]}{2}$ with the set of ordered pairs $\{(i,j):i,j\in[n],i<j\}$. If $A=\{i,j\}$ with $i<j$ we call i (j) the head (respectively, tail) of A denoted by $h(A)$ (respectively, $t(A)$).

Given the graph G and a set $A\subseteq[n]$, we let G_A denote the subgraph of G induced by A. We define the indicator $\mathcal{G}_A$

$$\mathcal{G}_A=\begin{cases}1 & \text{if } G_A \text{ is a clique,}\\ 0 & \text{otherwise.}\end{cases} \tag{4}$$

For convenience of notation we let $\mathcal{G}_{ij}\equiv\mathcal{G}_{\{i,j\}}$ and $g_A=\mathcal{G}_A-\mathbb{E}\{\mathcal{G}_A\}$ be the centered versions of the variables $\mathcal{G}_{ij}$. We also set $g_{ii}\equiv 0$.

We can now define the matrix $M=M(G,\underline{\alpha})\in\mathbb{R}^{\binom{[n]}{\le 2}\times\binom{[n]}{\le 2}}$ as follows. For any pair of sets $A,B\in\binom{[n]}{\le 2}$ we have:

$$M_{A,B}=\alpha_{|A\cup B|}\mathcal{G}_{A\cup B}\,, \tag{5}$$

with $\alpha_0\equiv 1$.

Theorem 1 *Suppose $\underline{\alpha}$, p satisfy:*

$$\alpha_1=\kappa\,,\quad \alpha_2=2\frac{\kappa^2}{p}\,,\quad \alpha_3=\frac{\kappa^3}{p^3}\,,\quad \alpha_4=8\frac{\kappa^4}{p^6}\,,\quad c(\kappa\log n)^{1/4}n^{1/6}\le p\le 1., \tag{6}$$

for some $\kappa\in[\log n/n, n^{-2/3}/\log n]$ and c a large enough absolute constant. If $G\sim\mathbb{G}(n,p)$ is a random graph with edge probability p then, for every n large enough,

$$\mathbb{P}\big\{M(G,\underline{\alpha})\succeq 0\big\}\ge 1-\frac{1}{n}\,. \tag{7}$$

The proof of this theorem can be found in Section A. As mentioned above, a more general set of conditions that imply $M(G,\underline{\alpha})\succeq 0$ with high probability is given in Proposition 6. The proof of Theorem 1 consists in checking that the conditions of Proposition 6 hold and deriving the consequences.

2.2. A Sum of Squares lower bound for Hidden Clique

We denote by $\mathbb{G}(n,p,k)$ hidden clique model, i.e. the distribution over graphs $G=(V,E)$, with vertex set $V=[n]$, a subset $\mathsf{Q}\subseteq[n]$ of k uniformly random vertices forming a clique, and every other edge present independently with probability p.

The SOS relaxation of degree $d = 4$ for the maximum clique problem (Tulsiani (2009); Barak (2014)) is a semidefinite program, whose decision variable is a matrix $X \in \mathbb{R}^{\binom{[n]}{\le 2}\times\binom{[n]}{\le 2}}$:

$$\begin{aligned} \text{maximize} \quad & \sum_{i\in[n]} X_{\{i\},\{i\}}\,, \\ \text{subject to:} \quad & X \succeq 0, \quad X_{S_1,S_2} \in [0,1]\,, \\ & X_{S_1,S_2} = 0 \quad \text{when } S_1\cup S_2 \text{ is not a clique in } G\,, \\ & X_{S_1,S_2} = X_{S_3,S_4} \quad \text{for all } S_1\cup S_2 = S_3\cup S_4\,, \\ & X_{\emptyset,\emptyset} = 1. \end{aligned} \tag{8}$$

Denote by $\mathsf{Val}(G; d = 4)$ the value of this optimization problem for graph G (which is obviously an upper bound on the size of the maximum clique in G). We can then try to detect the clique (i.e. distinguish hypothesis H_1 and H_0 defined in the introduction), by using the test statistics

$$T(G) = \begin{cases} 0 & \text{if } \mathsf{Val}(G;4) \le c_* k, \\ 1 & \text{if } \mathsf{Val}(G;4) > c_* k. \end{cases} \tag{9}$$

with c_* a numerical constant. The rationale for this test is as follows: if we replace $\mathsf{Val}(G;4)$ by the size of the largest clique, then the above test is essentially optimal, i.e. detects the clique with high probability as soon as $k \gtrsim \log n$ (with $c_* = 1$).

We then have the following immediate consequence of Theorem 1.

Corollary 2 *Suppose $G \sim \mathbb{G}(n, 1/2)$. Then, with probability at least $1 - n^{-1}$, the degree-4 SOS relaxation has value*

$$\mathsf{Val}(G;4) \gtrsim \frac{n^{1/3}}{\log n}\,. \tag{10}$$

Proof Consider $M(\underline{\alpha}, G)$ from Theorem 1 (with $p = 1/2$). For $M(\underline{\alpha}, G)$ to be positive semidefinite with high probability, we set $\kappa = c_0\, n^{-2/3}/\log n$ for some absolute constant c_0. It is easy to check that $M(\underline{\alpha}, G)$ is a feasible point for the optimization problem (8). Recalling that $M_{\{i\},\{i\}} = \alpha_1 = \kappa$, we conclude that the objective function at this point is $n\kappa = c_0 n^{1/3}/\log n$, and the claim follows. ■

We are now in position to derive a formal lower bound on the test (9).

Theorem 3 *The degree-4 Sum-of-Squares test for the maximum clique problem, defined in Eq. (9), fails to distinguish between $G \sim \mathbb{G}(n, k, 1/2)$ and $G \sim \mathbb{G}(n, 1/2)$ with high probability if $k \lesssim n^{1/3}/\log n$.*

In particular, $T(G) = 1$ with high probability both for $G \sim \mathbb{G}(n, k, 1/2)$, and for $G \sim \mathbb{G}(n, 1/2)$.

Proof [Proof of Theorem 3] Assume $k \le c_1 n^{1/3}/\log n$ for c_1 a sufficiently small constant. For $G \sim \mathbb{G}(n, 1/2)$, Corollary 2 immediately implies that $\mathsf{Val}(G;4) \ge c_* k$, with high probability.

For $G \sim \mathbb{G}(n, k, 1/2)$, we obviously have $\mathsf{Val}(G;4) \ge k$ (because SOS gives a relaxation). To obtain a larger lower bound, recall that $\mathsf{Q} \subseteq [n]$ indicates the vertices in the clique. The subgraph

G_{Q^c} induced by the set of vertices $\mathsf{Q}^c = [n]\backslash\mathsf{Q}$ is distributed as $\mathbb{G}(n-k, 1/2)$. Further, we obviously have

$$\mathsf{Val}(G; 4) \geq \mathsf{Val}(G_{\mathsf{Q}^c}; 4)\,. \tag{11}$$

Indeed we can always set to 0 variables indexed by sets $A \subseteq [n]$ with $A \not\subseteq \mathsf{Q}^c$. Hence, applying again Corollary 2, we deduce that, with probability $1-(n-k)^{-1}$, $\mathsf{Val}(G; 4) \geq C(n-k)^{1/3}/\log(n-k)$, which is larger than $c_* k$. Hence $T(G) = 1$ with high probability. ■

2.3. A Sum of Squares lower bound for Hidden Submatrix

As mentioned in the introduction, in the hidden submatrix problem we are given a matrix $A \in \mathbb{R}^{n\times n}$, which is generated according with either hypothesis H_0 or hypothesis H_1 defined there. To avoid unnecessary technical complications, we shall consider distributions $P_0 = \mathsf{N}(0, 1)$ (for all the entries in A under H_0) and $P_1 = \mathsf{N}(\mu, 1)$ (for the entries A_{ij}, $i, j \in \mathsf{Q}$ under H_1) .

In order to motivate our definition of an SOS-based statistical test, we begin by introducing a nearly-optimal combinatorial test, call it T_{comb}. This test essentially look for a principal submatrix of A of dimension k, with average value larger than $\mu/2$. Formally

$$T_{\mathrm{comb}}(A) \equiv \begin{cases} 1 & \text{if } \exists x \in \{0,1\}^n \text{ such that } \sum_{i\in[n]} x_i \leq k, \text{ and} \\ & \quad \text{and } \sum_{i,j\in[n],i<j} A_{ij}x_i x_j \geq \frac{1}{2}\binom{k}{2}\mu, \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

A straightforward union-bound calculation shows that $T_{\mathrm{comb}}(\,\cdot\,)$ succeeds with high probability provided $k \gtrsim \mu^{-2}\log n$.

As in the previous section, the degree-4 SOS relaxation of the set of binary vectors $x \in \{0,1\}^n$ consists in the following convex set of matrices

$$\mathcal{C}_4(n) \equiv \Big\{ X \in \mathbb{R}^{\binom{[n]}{\leq 2}\times\binom{[n]}{\leq 2}} : \quad X \succeq 0, \quad X_{S_1,S_2} \in [0,1]\,, \quad X_{\emptyset,\emptyset} = 1, \\ X_{S_1,S_2} = X_{S_3,S_4} \quad \text{for all } S_1 \cup S_2 = S_3 \cup S_4 \Big\}\,. \tag{13}$$

This suggests the following relaxation of the test $T_{\mathrm{comb}}(\,\cdot\,)$:

$$T(A) = \begin{cases} 1 & \text{if there exists } X \in \mathcal{C}_4(n) \text{ such that } \sum_{i\in[n]} X_{\{i\},\{i\}} \leq k, \text{ and} \\ & \quad \sum_{i,j\in[n],i<j} A_{ij}X_{\{i\},\{j\}} \geq c_*\mu k^2, \\ 0 & \text{otherwise}\,. \end{cases} \tag{14}$$

We begin by stating a corollary of Theorem 1.

Corollary 4 *Assume A is distributed according to hypothesis H_0, i.e. $A_{ij} \sim \mathsf{N}(0,1)$ for all $i, j \in [n]$. Then, with probability at least $1 - 2n^{-1}$, there exists $X \in \mathcal{C}_4(n)$ such that*

$$\sum_{i\in[n]} X_{\{i\},\{i\}} \lesssim \frac{n^{1/3}}{\log n}\,, \qquad \sum_{i,j\in[n],i<j} A_{ij}X_{\{i\},\{j\}} \gtrsim \frac{n^{2/3}}{(\log n)^2}\,. \tag{15}$$

Proof Fix λ a sufficiently large constant and let G be graph with adjacency matrix $\mathcal{G}$ given by $\mathcal{G}_{ij} = \mathbb{I}(A_{ij} \geq \lambda)$. Note that this is an Erdös-Renyi random graph $G \sim \mathbb{G}(n,p)$ with edge probability $p = \Phi(-\lambda)$. (Throughout this proof, we let $\phi(z) \equiv e^{-z^2/2}/\sqrt{2\pi}$ denote the Gaussian density, and $\Phi(z) \equiv \int_{-\infty}^{z} \phi(t)\,\mathrm{d}t$ the Gaussian distribution function.)

We choose $X = M(G, \underline{\alpha})$ a random association scheme, where $\underline{\alpha}$ is set according to Theorem 1, with

$$\kappa = \frac{c_2}{n^{2/3}\log n}\,, \tag{16}$$

with c a suitably small constant. This ensures that the conditions of Theorem 1 are satisfied, whence $X \in \mathcal{C}_4(n)$ with high probability. Further, by definition

$$\sum_{i\in[n]} X_{\{i\},\{i\}} = n\kappa = \frac{c_2\, n^{1/3}}{\log n}\,. \tag{17}$$

It remains to check that the second inequality in (15) hold. We have

$$\sum_{i,j\in[n],i<j} A_{ij} X_{\{i\},\{j\}} = \frac{2\kappa^2}{p} \sum_{i,j\in[n],i<j} A_{ij}\mathcal{G}_{ij}\,. \tag{18}$$

Note that

$$\mathbb{E}\Big\{ \sum_{i,j\in[n],i<j} A_{ij}\mathcal{G}_{ij}\Big\} = \binom{n}{2}\mathbb{E}\big\{A_{12}\,\mathbb{I}(A_{12}\geq\lambda)\big\} = \binom{n}{2}\phi(\lambda)\,. \tag{19}$$

Note that the random variables $(A_{ij}\mathcal{G}_{ij})_{i<j}$ are independent and subgaussian. By a standard concentration-of-measure argument we have, with probability at least $1-n^{-2}$, for a suitably small constant c', $\sum_{i<j} A_{ij}\mathcal{G}_{ij} \geq c' n^2 \phi(\lambda)$ and hence

$$\sum_{i,j\in[n],i<j} A_{ij} X_{\{i\},\{j\}} \gtrsim \kappa^2 n^2 \gtrsim \frac{n^{2/3}}{(\log n)^2}\,. \tag{20}$$

■

Theorem 5 *Consider the Hidden Submatrix problem with entries' distributions $P_0 = \mathsf{N}(0,1)$, and $P_1 = \mathsf{N}(\mu,1)$.*

Then, the degree-4 Sum-of-Squares, defined in Eq. (14), fails to distinguish between hypotheses H_0 and H_1 if $k \lesssim \mu^{-1} n^{1/3}/\log n$. In particular, $T(A) = 1$ with high probability both under H_0 and under H_1.

Proof First consider A distributed according to hypothesis H_0. Note that, if $X_0 \in \mathcal{C}_4(n)$ and $s \in [0,1]$ is a scaling factor, then $s\,X_0 \in \mathcal{C}_4$. Therefore (by choosing $s = ckn^{-1/3}\log n$ for a suitable constant c) Corollary 4 implies that with high probability there exists $X \in \mathcal{C}_4(n)$ such that

$$\sum_{i\in[n]} X_{\{i\},\{i\}} \leq k\,, \qquad \sum_{i,j\in[n],i<j} A_{ij} X_{\{i\},\{j\}} \gtrsim \frac{k\, n^{1/3}}{\log n}\,. \tag{21}$$

Therefore, for $\mu\, k \le c\, n^{1/3}/\log n$ with c a sufficiently small constant, we have $\sum_{i<j} A_{ij} X_{\{i\},\{j\}} \ge c_* \mu\, k^2$ and therefore $T(A) = 1$ with high probability.

Consider next A distributed according to hypothesis H_1. Note that $A = \mu\, 1_{\mathsf{Q}} 1_{\mathsf{Q}}^{\mathsf{T}} + \widetilde{A}$, where 1_{Q} is the indicator vector of set Q, and $\widetilde{A}$ is distributed according to H_0. Since $\sum_{i<j} A_{ij} X_{\{i\},\{j\}}$ is increasing in A, we also have that $T(\widetilde{A}) = 1$ implies $T(A) = 1$. As shown above, for $\mu\, k \le c\, n^{1/3}/\log n$, we have $T(\widetilde{A}) = 1$ with high probability, and hence $T(A) = 1$. ∎

3. Further definitions and proof strategy

In order to prove $M(G, \underline{\alpha}) \succeq 0$, we will actually study a new matrix $N(G, \underline{\alpha}) \in \mathbb{R}^{\binom{[n]}{\le 2} \times \binom{[n]}{\le 2}}$ defined as follows:

$$N_{A,B} = \alpha_{|A \cup B|} \prod_{i \in A \setminus B, j \in B \setminus A} \mathcal{G}_{ij}. \tag{22}$$

Notice that $M_{A,B} = N_{A,B} \mathcal{G}_A \mathcal{G}_B$, i.e. M is obtained from N by setting to zero columns (rows) indexed by sets A, B that do not induce cliques in G. Thus, $N \succcurlyeq 0$ implies $M \succcurlyeq 0$.

We also define the matrix $H \in \mathbb{R}^{\left(\binom{[n]}{1} \cup \binom{[n]}{2}\right) \times \left(\binom{[n]}{1} \cup \binom{[n]}{2}\right)}$ that is the Schur complement of N with respect to entry $N_{\emptyset,\emptyset} = 1$. Formally:

$$H_{A,B} = N_{A,B} - \alpha_{|A|} \alpha_{|B|}\,, \tag{23}$$

where, as before, we define $\alpha_0 = 1$. Furthermore we denote by $H_{a,b}$, for $a, b \in \{1, 2\}$, the restriction of H to rows indexed by $\binom{[n]}{a}$ and columns indexed by $\binom{[n]}{b}$. (This abuse of notation will not be a source of confusion in what follows, since we will always use explicit values in $\{1, 2\}$ for the subscripts a, b.)

Since H is the Schur complement of N, $H \succeq 0$ implies $N \succeq 0$ and hence $M \succeq 0$. The next section is devoted to prove $H \succeq 0$: here we sketch the main ingredients.

Technically, we control the spectrum of H by first computing eigenvalues and eigenspaces of its expectation $\mathbb{E}H$ and then controlling the random part $H - \mathbb{E}H$ by the moment method, i.e. computing moments of the form $\mathbb{E}\mathsf{Tr}\{(H - \mathbb{E}H)^{2m}\}$. The key challenge is that the simple Weyl inequality $\lambda_{\min}(H) \ge \lambda_{\min}(\mathbb{E}H) - \|H - \mathbb{E}H\|_2$ is too weak for proving the desired result. We instead decompose H in its blocks $H_{1,1}$, $H_{1,2}$, $H_{2,2}$ and prove the inequalities stated in Proposition 6, cf. Eqs. (55) to (57). Briefly, these allow us to conclude that:

$$H_{1,1} \succcurlyeq 0. \tag{24}$$

$$H_{2,2} \succcurlyeq H_{1,2}^{\mathsf{T}} H_{1,1}^{-1} H_{1,2}, \tag{25}$$

which are the Schur complement conditions guaranteeing $H \succcurlyeq 0$. While characterizing $H_{1,1}$ is relatively easy (indeed this block is essentially the adjacency matrix of G), the most challenging part of the proof consists in showing a sufficient condition for Eq. (25) (see Eq. (57) below). In order to prove this bound, we need to decompose $H_{2,2}$ and $H_{1,2}$ along the eigenspaces of $\mathbb{E}H_{2,2}$, and carefully control each of the corresponding sub-blocks.

In the rest of this section we demonstrate the essentials of our strategy to show the weaker assertion $H_{2,2} \succcurlyeq 0$. We will assume that p is order one, for concreteness $p = 1/2$ which corresponds to the hidden clique problem. It suffices to show that

$$\mathbb{E}H_{2,2} \succcurlyeq \mathbb{E}H_{2,2} - H_{2,2}. \tag{26}$$

The expected value $\mathbb{E}H_{2,2}$ has 3 distinct eigenspaces $\mathbb{V}_0, \mathbb{V}_1, \mathbb{V}_2$ that form an orthogonal decomposition of $\mathbb{R}^{\binom{[n]}{2}}$. Crucially, these spaces admit a simple description as follows:

$$\mathbb{V}_0 \equiv \{v \in \mathbb{R}^{\binom{[n]}{2}} : \exists u \in \mathbb{R} \text{ s.t. } v_{\{i,j\}} = u \text{ for all } i < j\}, \tag{27}$$

$$\mathbb{V}_1 \equiv \{v \in \mathbb{R}^{\binom{[n]}{2}} : \exists u \in \mathbb{R}^n, \text{ s.t. } \langle 1_n, u\rangle = 0 \text{ and } \quad v_{\{i,j\}} = u_i + u_j \text{ for all } i < j\}, \tag{28}$$

$$\mathbb{V}_2 \equiv (\mathbb{V}_0 \oplus \mathbb{V}_1)^{\perp}. \tag{29}$$

If $\mathcal{P}_a$ is the orthogonal projector onto $\mathbb{V}_a$ we have that $\mathbb{E}H_{2,2} = \lambda_0\mathcal{P}_0 + \lambda_1\mathcal{P}_1 + \lambda_2\mathcal{P}_2$ where $\lambda_0 \approx n^2\kappa^4, \lambda_1 \approx n\kappa^3$ and $\lambda_2 \approx \kappa^2$ (see Proposition 21 for a formal statement).

Now, consider the entry indexed by $\{i,j\}, \{k,\ell\} \in \binom{[n]}{2}$:

$$\begin{aligned}
(H_{2,2})_{\{i,j\},\{k,\ell\}} &= -\alpha_2^2 + \alpha_4 \mathcal{G}_{ik}\mathcal{G}_{i\ell}\mathcal{G}_{jk}\mathcal{G}_{j\ell} && (30)\\
&= -\alpha_2^2 + \alpha_4(p + g_{ik})(p + g_{i\ell})(p + g_{jk})(p + g_{j\ell}) && (31)\\
&= -\alpha_2^2 + \alpha_4 p^4 + \alpha_4 p^3(g_{ik} + g_{i\ell} + g_{jk} + g_{j\ell})\\
&\quad + \alpha_4 p^2(g_{ik}g_{i\ell} + g_{ik}g_{jk}g_{jk}g_{j\ell} + g_{i\ell}g_{j\ell} + g_{ik}g_{j\ell} + g_{i\ell}g_{jk})\\
&\quad + \alpha_4 p(g_{ik}g_{i\ell}g_{jk} + g_{ik}g_{jk}g_{j\ell} + g_{ik}g_{i\ell}g_{j\ell} + g_{i\ell}g_{jk}g_{j\ell}) + \alpha_4 g_{ij}g_{i\ell}g_{jk}g_{j\ell}. && (32)
\end{aligned}$$

The decomposition Eq. (32) holds only when $\{i,j\}$ and $\{k,\ell\}$ are disjoint. Since the number of pairs $\{i,j\}, \{k,\ell\}$ that intersect are at most $n^3 \ll n^4$, it is natural to conjecture that these pairs are negligible, and in this outline we shall indeed assume that this is true (the complete proof deals with these pairs as well). The random portion $\mathbb{E}H_{2,2} - H_{2,2}$ involves the last 15 terms of the above decomposition. Each term is indexed by a pair (η, ν) where $1 \le \eta \le 4$ denotes the number of g_{ij} variables in the term and $1 \le \nu \le \binom{4}{\eta}$ the exact choice of η (out of 4) variables used. In accordance with notation used in the proof, we let $\widetilde{J}_{\eta,\nu}$ denote the matrix with $\{i,j\}, \{k,\ell\}$ entry is the (η,ν) entry in the decomposition Eq. (32). See Table 1 and Eq. (178) for a formal definition of the matrices $\widetilde{J}_{\eta,\nu}$. Hence we obtain (the $\approx$ below is due to the intersecting pairs, which we have ignored):

$$H_{2,2} - \mathbb{E}H_{2,2} \approx \sum_{\eta \le 4} \sum_{\nu \le \binom{4}{\eta}} \widetilde{J}_{\eta,\nu}\,. \tag{33}$$

We are therefore left with the task of proving

$$\mathbb{E}H_{2,2} \succcurlyeq Q \equiv -\sum_{\eta} \sum_{\nu} \widetilde{J}_{\eta,\nu}. \tag{34}$$

Viewed in the decomposition given by $\mathbb{V}_0, \mathbb{V}_1, \mathbb{V}_2$, Eq. (34) is satisfied if:

$$\begin{pmatrix} \lambda_0 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_2 \end{pmatrix} \succcurlyeq \begin{pmatrix} \|\mathcal{P}_0 Q \mathcal{P}_0\|_2 & \|\mathcal{P}_0 Q \mathcal{P}_1\|_2 & \|\mathcal{P}_0 Q \mathcal{P}_2\|_2 \\ \|\mathcal{P}_1 Q \mathcal{P}_0\|_2 & \|\mathcal{P}_1 Q \mathcal{P}_1\|_2 & \|\mathcal{P}_1 Q \mathcal{P}_2\|_2 \\ \|\mathcal{P}_2 Q \mathcal{P}_0\|_2 & \|\mathcal{P}_2 Q \mathcal{P}_1\|_2 & \|\mathcal{P}_2 Q \mathcal{P}_2\|_2 \end{pmatrix} \tag{35}$$

The bulk of the proof is devoted to developing operator norm bounds for the matrices $\mathcal{P}_a\widetilde{J}_{\eta,\nu}\mathcal{P}_b$ that hold with high probability. We then bound $\mathcal{P}_aQ\mathcal{P}_b$ using triangle inequality

$$\|\mathcal{P}_aQ\mathcal{P}_b\|_2 \le \sum_{\eta,\nu} \left\|\mathcal{P}_a\widetilde{J}_{\eta,\nu}\mathcal{P}_b\right\|_2. \tag{36}$$

The matrices $\widetilde{J}_{4,1}, \widetilde{J}_{3,\nu}, \widetilde{J}_{2,1}, \widetilde{J}_{2,6}$ turn out to have an approximate "Wigner"-like behavior, in the following sense. Note that these are symmetric matrices of size $\binom{n}{2} \approx n^2/2$ with random zero-mean entries bounded by α_4. If their entries were *independent*, they would have operator norms of order $\alpha_4\sqrt{n^2/2} \approx \kappa^4 n$ (Füredi and Komlós (1981)). Although the entries are actually not independent, the conclusion still holds for $\widetilde{J}_{4,1}, \widetilde{J}_{3,\nu}, \widetilde{J}_{2,1}, \widetilde{J}_{2,6}$ and they have operator norms of order $\kappa^4 n$. Hence $\|\mathcal{P}_a\widetilde{J}_{\eta,\nu}\mathcal{P}_b\|_2 \le \|\widetilde{J}_{\eta,\nu}\|_2 \approx \kappa^4 n$ for these cases.

We are now left with the cases $(\widetilde{J}_{1,\nu})_{1\le\nu\le 4}$ and $(\widetilde{J}_{2,\nu})_{2\le\nu\le 5}$. These require more care, since their typical norms are significantly larger than n. For instance consider $\widetilde{J}_{1,\nu}$ where

$$(\widetilde{J}_{1,\nu})_{\{i,j\},\{k,\ell\}} = g_{ik}. \tag{37}$$

Viewed as a matrix in $\mathbb{R}^{n^2\times n^2}$, $\widetilde{J}_{1,\nu}$ corresponds to the matrix $\alpha_4 g \otimes (1_n 1_n)^{\mathsf{T}}$ where $\otimes$ denotes the standard Kronecker product and $g \in \mathbb{R}^{n\times n}$ is the matrix with (i,j) entry being g_{ij}. By standard results on Wigner random matrices (Füredi and Komlós (1981)), $\|g\|_2 \lesssim \sqrt{n}$ with high probability. Hence:

$$\left\|g \otimes 1_n 1_n^{\mathsf{T}}\right\|_2 = \|g\|_2 \left\|1_n 1_n^{\mathsf{T}}\right\|_2 \lesssim n^{3/2}, \tag{38}$$

with high probability. This suggests that $\|\widetilde{J}_{1,\nu}\|_2 \lesssim \alpha_4 n^{3/2} \approx \kappa^4 n^{3/2}$ with high probability. This turns out to be the correct order for all the matrices $\widetilde{J}_{1,\nu}$ and $\widetilde{J}_{2,\nu}$ under consideration.

This heuristic calculation shows the need to be careful with these terms. Indeed, a naive application of this results yields that $\|\mathcal{P}_aQ\mathcal{P}_b\|_2 \lesssim \kappa^4 n^{3/2}$. Recalling Eq. (35), this imposes that $\lambda_2 \gg \kappa^4 n^{3/2}$. Since we have $\lambda_2 \approx \kappa^2$, we obtain the condition $\kappa \ll n^{-3/4}$. The parameter κ turns out to be related to the size of the planted clique through $k \approx n\kappa$. Hence this argument can only prove that the SOS hierarchy fails to detect hidden cliques of size $k \ll n^{1/4}$. Indeed, the result of Meka et al. (2015) specialized to $d = 4$ amounts to such a consideration.

In order to improve over this, and establish Theorem 1 we prove that matrices $\widetilde{J}_{1,\nu}$ and $\widetilde{J}_{2,\nu}$ satisfy certain spectral properties with respect to the subspaces $\mathbb{V}_0, \mathbb{V}_1, \mathbb{V}_2$. For instance consider the sum $\widetilde{J}_{2,3} + \widetilde{J}_{2,5}$. For any $v \in \mathbb{R}^{\binom{[n]}{2}}$

$$(\widetilde{J}_{2,3}v + \widetilde{J}_{2,5}v)_{\{i,j\}} = \sum_{k<\ell} p^2(g_{ik}g_{i\ell} + g_{jk}g_{j\ell})v_{\{k,\ell\}} \tag{39}$$

$$= u_i + u_j, \tag{40}$$

where we let $u_i \equiv= \sum_{k<\ell} p^2(g_{ik}g_{i\ell})v_{\{k,\ell\}}$. It follows that $(\widetilde{J}_{2,3}v + \widetilde{J}_{2,5})v \in \mathbb{V}_0 \oplus \mathbb{V}_1$ hence $\mathcal{P}_2(\widetilde{J}_{2,3} + \widetilde{J}_{2,5}) = 0$. By taking transposes we obtain that $(\widetilde{J}_{2,2} + \widetilde{J}_{2,4})\mathcal{P}_2 = 0$. In a similar fashion we obtain that $\mathcal{P}_2(\sum_\nu \widetilde{J}_{1,\nu}) = (\sum_\nu \widetilde{J}_{1,\nu})\mathcal{P}_2 = 0$. See Lemmas 28, 29 for formal statements and proofs.

Using these observations and Eq. (36) we obtain that $\|\mathcal{P}_2 Q \mathcal{P}_2\| \lesssim \kappa^4 n$, while for any other pair $(a,b) \in \{0,1,2\}^2$ we have that $\|\mathcal{P}_a Q \mathcal{P}_b\| \lesssim \kappa^4 n^{3/2}$. As noted before, since $\lambda_0 \approx n^2\kappa^4$, $\lambda_1 \approx n\kappa^3$ and $\lambda_1 \approx \kappa^2$ whence the condition in Eq. (35) reduces to:

$$\begin{pmatrix} n^2\kappa^4 & 0 & 0 \\ 0 & n\kappa^3 & 0 \\ 0 & 0 & \kappa^2 \end{pmatrix} - \kappa^4 \begin{pmatrix} n^{3/2} & n^{3/2} & n^{3/2} \\ n^{3/2} & n^{3/2} & n^{3/2} \\ n^{3/2} & n^{3/2} & n \end{pmatrix} \succcurlyeq 0. \tag{41}$$

The $2,2$ entry of this matrix inequality yields that $\kappa^2 - \kappa^4 n \gg 0$ or $\kappa \ll n^{-1/2}$. Considering the $(1,1)$ entry yields a similar condition. The key condition is that corresponding to the minor indexed by rows (and columns) $1,2$:

$$\begin{pmatrix} n\kappa^3 & -n^{3/2}\kappa^4 \\ -n^{3/2}\kappa^4 & \kappa^2 \end{pmatrix} \succcurlyeq 0. \tag{42}$$

This requires that $n\kappa^5 \gg n^3\kappa^8$ or, equivalently $\kappa \ll n^{-2/3}$. Translating this to clique size $k = n\kappa$, we obtain the condition $k \ll n^{1/3}$. This calculation thus demonstrates the origin of the threshold of $n^{1/3}$ beyond which the Meka-Wigderson witness fails to be positive semidefinite. The counterexample of Barak (2014) shows that our estimates are fairly tight (indeed, up to a logarithmic factor).

References

Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. In *Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms*, pages 594–598. Society for Industrial and Applied Mathematics, 1998.

Brendan P.W. Ames and Stephen A. Vavasis. Nuclear norm minimization for the planted clique and biclique problems. *Mathematical programming*, 129(1):69–89, 2011.

Boaz Barak. Sums of Squares upper bounds, lower bounds, and open questions (Lecture notes, Fall 2014). http://www.boazbarak.org/sos/, 2014.

Boaz Barak and David Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. arXiv:1404.5236, 2014.

Quentin Berthet and Philippe Rigollet. Complexity theoretic lower bounds for sparse principal component detection. In *Conference on Learning Theory*, pages 1046–1066, 2013.

T Tony Cai, Tengyuan Liang, and Alexander Rakhlin. Computational and statistical boundaries for submatrix localization in a large noisy matrix. arXiv:1502.01988, 2015.

Yudong Chen and Jiaming Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. arXiv:1402.1267, 2014.

Yael Dekel, Ori Gurel-Gurevich, and Yuval Peres. Finding hidden cliques in linear time with high probability. In *ANALCO*, pages 67–75. SIAM, 2011.

Yash Deshpande and Andrea Montanari. Finding hidden cliques of size $\sqrt{N/e}$ in nearly linear time. *Foundations of Computational Mathematics*, pages 1–60, 2014. ISSN 1615-3375. doi: 10.1007/s10208-014-9215-y.

Uriel Feige and Robert Krauthgamer. Finding and certifying a large hidden clique in a semirandom graph. *Random Structures and Algorithms*, 16(2):195–208, 2000.

Uriel Feige and Dorit Ron. Finding hidden cliques in linear time. *DMTCS Proceedings*, (01): 189–204, 2010.

Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh Vempala, and Ying Xiao. Statistical algorithms and a lower bound for planted clique. arXiv:1201.1214, 2012.

Zoltán Füredi and János Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.

Geoffrey R Grimmett and Colin JH McDiarmid. On colouring random graphs. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 77, pages 313–324. Cambridge Univ Press, 1975.

Bruce Hajek, Yihong Wu, and Jiaming Xu. Computational lower bounds for community detection on random graphs. *arXiv preprint arXiv:1406.6625*, 2014.

Johan Hastad. Clique is hard to approximate within $n^{1-\epsilon}$. In *Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on*, pages 627–636. IEEE, 1996.

Mark Jerrum. Large cliques elude the Metropolis process. *Random Structures & Algorithms*, 3(4): 347–359, 1992.

Iain M Johnstone and Arthur Yu Lu. On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association*, 104(486), 2009.

Richard M. Karp. Reducibility among combinatorial problems. In R. E. Miller and J. W. Thatcher, editors, *Complexity of Computer Computations*. Plenum, 1972.

Subhash Khot. Improved inapproximability results for maxclique, chromatic number and approximate graph coloring. In *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, pages 600–609. IEEE, 2001.

Jean B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

Zongming Ma and Yihong Wu. Computational barriers in minimax submatrix detection. arXiv:1309.5914, 2013.

Raghu Meka and Avi Wigderson. Association schemes, non-commutative polynomial concentration, and sum-of-squares lower bounds for planted clique. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 20, page 105, 2013.

Raghu Meka, Aaron Potechin, and Avi Wigderson. Sum-of-squares lower bounds for the planted clique problem. In *ACM Symposium on Theory of Computing*, 2015.

Samet Oymak, Amin Jalali, Maryam Fazel, Yonina C Eldar, and Babak Hassibi. Simultaneously structured models with application to sparse and low-rank matrices. arXiv:1212.3753, 2012.

Pablo A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical programming*, 96(2):293–320, 2003.

Jean-Pierre Serre. Linear representations of finite groups. *Graduate Texts in Mathematics*, 42, 1977.

Andrey A Shabalin, Victor J Weigman, Charles M Perou, and Andrew B Nobel. Finding large average submatrices in high dimensional data. *The Annals of Applied Statistics*, pages 985–1012, 2009.

NZ Shor. Class of global minimum bounds of polynomial functions. *Cybernetics and Systems Analysis*, 23(6):731–734, 1987.

Madhur Tulsiani. CSP gaps and reductions in the Lasserre hierarchy. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 303–312. ACM, 2009.

Appendix A. Proofs

A.1. Definitions and notations

Throughout the proof we denote the identity matrix in m dimensions by I_m, and the all-ones vector by 1_m. We let $\mathcal{Q}_n = 1_n 1_n^{\mathsf{T}}/n$ be the projector onto the all ones vector 1_n, and $\mathcal{Q}_n^{\perp} = \mathrm{I}_n - \mathcal{Q}_n$ its orthogonal complement.

The indicator function of property A is denoted by $\mathbb{I}(A)$. The set of first m integers is denoted by $[m] = \{1, 2, \dots, m\}$.

As mentioned above, we write $f(n, r, \dots) \gtrsim g(n, r, \dots)$ if *there exists a constant* C such that $f(n, r, \dots) \geq C\, g(n, r, \dots)$. Similarly we write $f(n, r, \dots) \gg g(n, r, \dots)$ if, *for any constant* C, we have $f(n, r, \dots) \geq C\, g(n, r, \dots)$ for all n large enough. These conditions are always understood to hold uniformly with respect to the extra arguments $r, \dots$, provided these belong to a range depending on n, that will be clear from the context.

We finally use the shorthand $\bar{n} \equiv n \log n$.

A.2. Main technical result and proof of Theorem 1

The key proposition is the following which controls the matrices $H_{a,b}$. A set of conditions for the parameters $\underline{\alpha}$ is stated in terms of two matrices $\overline{W}, W \in \mathbb{R}^{3\times 3}$. Below we will develop approximations to these matrices, under the parameter values of Theorem 1. This facilitates checking the conditions of Proposition 6.

Proposition 6 *Consider the symmetric matrices $\overline{W}, W \in \mathbb{R}^{3\times 3}$, where $\overline{W}$ is diagonal, and given by:*

$$\overline{W}_{00} = \alpha_2 + 2(n-2)\alpha_3 p + \frac{(n-2)(n-3)}{2}\alpha_4 p^4 - \frac{n(n-1)}{2}\alpha_2^2\,, \tag{43}$$

$$\overline{W}_{11} = \alpha_2 + (n-4)\alpha_3 p - (n-3)\alpha_4 p^4\,, \tag{44}$$

$$\overline{W}_{22} = \alpha_2 - 2\alpha_3 p + \alpha_4 p^4\,, \tag{45}$$

and W is defined by:

$$W_{00} = C\alpha_3\bar{n}^{1/2} + C\alpha_4\bar{n}^{3/2} + \frac{C(\alpha_3\bar{n})^2}{\alpha_1} + \frac{\left(n^{3/2}\alpha_3 p^2 + 2\sqrt{n}\alpha_2 + C\alpha_3\bar{n}\right)^2}{n(\alpha_2 p - \alpha_1^2)}\,, \tag{46}$$

$$W_{01} = C\alpha_3\bar{n}^{1/2} + C\alpha_4\bar{n}^{3/2} + \frac{C}{\alpha_1}(\alpha_3\bar{n})(C\alpha_3\bar{n} + \sqrt{n}\alpha_2) + \frac{1}{n(\alpha_2 p - \alpha_1^2)}(n^{3/2}\alpha_3 p^2 + 2\sqrt{n}\alpha_2 + C\alpha_3\bar{n})(3\alpha_3\bar{n})\,, \tag{47}$$

$$W_{02} = C\alpha_3\bar{n}^{1/2} + C\alpha_4\bar{n}^{3/2} + \frac{C(\alpha_3\bar{n})^2}{\alpha_1} + \frac{C}{n(\alpha_2 p - \alpha_1^2)}\left(n^{3/2}\alpha_3 p^2 + 2\sqrt{n}\alpha_2 + C\alpha_3\bar{n}\right)(\alpha_3\bar{n})\,, \tag{48}$$

$$W_{11} = C\alpha_3\bar{n}^{1/2} + C\alpha_4\bar{n}^{3/2} + \frac{2}{\alpha_1}\left(C\alpha_3\bar{n} + \sqrt{n}\alpha_2\right)^2 + \frac{C(\alpha_3\bar{n})^2}{n(\alpha_2 p - \alpha_1^2)}\,, \tag{49}$$

$$W_{12} = C\alpha_3\bar{n}^{1/2} + C\alpha_4\bar{n}^{3/2} + \frac{C}{\alpha_1}(\alpha_3\bar{n})(C\alpha_3\bar{n} + \sqrt{n}\alpha_2) + \frac{C(\alpha_3\bar{n})^2}{n(\alpha_2 p - \alpha_1^2)}\,, \tag{50}$$

$$W_{22} = C\alpha_3\bar{n}^{1/2} + C\alpha_4\bar{n} + \frac{C(\alpha_3\bar{n})^2}{\alpha_1} + \frac{C(\alpha_3\bar{n})^2}{n(\alpha_2 p - \alpha_1^2)}\,. \tag{51}$$

Assume the following conditions hold for a suitable constant C:

$$\alpha_1 \geq 2\alpha_2 p + 2\alpha_2\bar{n}^{1/2}, \tag{52}$$

$$\alpha_2 p^2 \geq \alpha_1^2\,, \tag{53}$$

$$\overline{W} \succcurlyeq W\,. \tag{54}$$

Then with probability exceeding $1 - n^{-1}$ all of the following are true:

$$H_{11} \succcurlyeq 0\,, \tag{55}$$

$$H_{11}^{-1} \preccurlyeq \frac{1}{n(\alpha_2 p - \alpha_1^2)}\mathcal{Q}_n + \frac{2}{\alpha_1}\mathcal{Q}_n^{\perp}\,, \tag{56}$$

$$H_{22} \succcurlyeq \frac{2}{\alpha_1}H_{12}^{\mathsf{T}}\mathcal{Q}_n^{\perp}H_{12} + \frac{1}{n(\alpha_2 p - \alpha_1^2)}H_{12}^{\mathsf{T}}\mathcal{Q}_n H_{12}. \tag{57}$$

The next two lemmas develop simplified expressions for matrices $\overline{W}$, W under the parameter choices of Theorem 1.

Lemma 7 *Setting $(\underline{\alpha}, p)$ as in Theorem 1, there exists $\delta_n = \delta_n(\kappa, p)$ with $\delta_n(\kappa, p) \to 0$ as $n \to \infty$, such that*

$$\left|\overline{W}_{00} - \frac{2n^2\kappa^4}{p^2}\right| \leq \delta_n\overline{W}_{00}\,, \tag{58}$$

$$\left|\overline{W}_{11} - \frac{n\kappa^3}{p^2}\right| \leq \delta_n\overline{W}_{11}\,, \tag{59}$$

$$\left|\overline{W}_{22} - \frac{2\kappa^2}{p}\right| \leq \delta_n\overline{W}_{22}\,. \tag{60}$$

Lemma 8 *Setting $(\underline{\alpha}, p)$ as in Theorem 1, there exists $\delta_n = \delta_n(\kappa, p)$ with $\delta_n(\kappa, p) \to 0$ as $n \to \infty$, such that, for some absolute constant C,*

$$\left| W_{00} - \frac{n^2\kappa^4}{p^2} \right| \le \delta_n W_{00}\,, \tag{61}$$

$$\left| W_{11} - C\frac{\kappa^4 \bar{n}^{3/2}}{p^6} \right| \le \delta_n W_{11}\,, \tag{62}$$

$$\left| W_{22} - C\frac{\kappa^3\sqrt{\bar{n}}}{p^3} - C\frac{\kappa^5\bar{n}^2}{p^6} \right| \le \delta_n W_{22}\,, \tag{63}$$

and, for every $a \neq b \in \{0,1,2\}$,

$$\left| W_{ab} - C\frac{\kappa^4 \bar{n}^{3/2}}{p^6} \right| \le \delta_n W_{ab}, \tag{64}$$

With Proposition 6 and the auxiliary Lemmas 8, 7 in hand, the proof of Theorem 1 is straightforward.

Proof [Proof of Theorem 1] As noted in Section 3 it suffices to prove that $H \succeq 0$. By taking the Schur complement with respect to H_{11}, we obtain that $H \succeq 0$ if and only if

$$H_{11} \succcurlyeq 0 \quad \text{and} \quad H_{22} \succcurlyeq H_{12}^{\mathsf{T}} H_{11}^{-1} H_{12}\,. \tag{65}$$

Suppose that the conditions of Proposition 6 are verified under the values of $\underline{\alpha}, p$ specified as in Theorem 1. Then we have $H_{11} \succeq 0$ by Eq. (55). Further by Eqs. (56) and (57), we have

$$H_{22} \succcurlyeq H_{12}^{\mathsf{T}} \left(\frac{2}{\alpha_1}\mathcal{Q}_n^{\perp} + \frac{1}{n(\alpha_2 p - \alpha_1^2)}\mathcal{Q}_n \right) H_{12} \tag{66}$$

$$\succcurlyeq H_{12}^{\mathsf{T}} H_{11}^{-1} H_{12}\,, \tag{67}$$

which yields the desired (65).

We are now left to verify the conditions of Proposition 6. To begin, we verify that $\alpha_1 \gtrsim 2\alpha_2 p + 2\alpha_2 \bar{n}^{1/2}$. This condition is satisfied if:

$$p \gg \kappa \bar{n}^{1/2}\,. \tag{68}$$

For this, it suffices that

$$(\kappa \log n)^{1/4} n^{1/6} \gg \kappa \bar{n}^{1/2}\,. \tag{69}$$

$$\text{or } \kappa \lesssim n^{-4/9} (\log n)^{-1/3}. \tag{70}$$

Since $\kappa \le n^{-2/3}$, this is true.

The condition $\alpha_2 p - \alpha_1^2 \ge 0$ holds since $\alpha_2 p - \alpha_1^2 = 2\kappa^2 - \kappa^2 = \kappa^2 > 0$.

It remains to check that $\overline{W} \succcurlyeq W$. By Sylvester's criterion, we need to verify that:

$$\overline{W}_{00} - W_{00} > 0\,, \tag{71}$$

$$\begin{vmatrix} \overline{W}_{00} - W_{00} & -W_{01} \\ -W_{01} & \overline{W}_{11} - W_{11} \end{vmatrix} > 0\,, \tag{72}$$

$$\begin{vmatrix} \overline{W}_{00} - W_{00} & -W_{01} & -W_{02} \\ -W_{01} & \overline{W}_{11} - W_{11} & -W_{12} \\ -W_{02} & -W_{12} & \overline{W}_{22} - W_{22} \end{vmatrix} > 0\,. \tag{73}$$

It suffices to check the above values using the simplifications provided by Lemmas 7 and 8 respectively as follows. Throughout, we will assume that n is large enough, and write δ_n for a generic sequence such that $\delta_n \to 0$ uniformly over $\kappa \in [\log n/n, c^{-4}n^{-2/3}/\log n]$, $p \in [c(\kappa \log n)^{1/4}n^{1/6}, 1]$.

For Eq. (71), using Lemmas 7 and 8 we have that:

$$\overline{W}_{00} - W_{00} \geq \frac{n^2\kappa^4}{2p^2}, \tag{74}$$

Hence , $\overline{W}_{00} - W_{00} \geq n^2\kappa^4/2p^2 > 0$ for large enough n.

For Eq. (72) to hold we need:

$$(\overline{W}_{00} - W_{00})(\overline{W}_{11} - W_{11}) - W_{01}^2 > 0. \tag{75}$$

By Lemmas 7 and 8 we have:

$$\overline{W}_{11} - W_{11} \geq \frac{n\kappa^3}{p^2}(1-\delta_n) - \frac{C\kappa^4\bar{n}^{3/2}}{p^6}(1+\delta_n). \tag{76}$$

The ratio of the two terms above is (up to a constant) given by $p^4/(\kappa n^{1/2}(\log n)^{3/2}) \to \infty$, hence for n large enough we have $\overline{W}_{11} - W_{11} \geq n\kappa^3/2p^2$. Thus Eq. (72) holds if

$$\left(\frac{n^2\kappa^4}{p^2}\right)\left(\frac{n\kappa^3}{p^2}\right) \gg \left(\frac{\kappa^4\bar{n}^{3/2}}{p^6}\right)^2 \tag{77}$$

$$\text{or } p^8 \gg \kappa(\log n)^3. \tag{78}$$

However as we set $p \gtrsim (\kappa \log n)^{1/4}n^{1/6}$, this is satisfied for n large. Indeed this implies that:

$$\begin{vmatrix} \overline{W}_{00} - W_{00} & -W_{01} \\ -W_{01} & \overline{W}_{11} - W_{11} \end{vmatrix} \geq \frac{n^3\kappa^7}{2p^4}\,. \tag{79}$$

Consider now Eq. (73). Expanding the determinant along the third column

$$(\overline{W}_{22} - W_{22})\begin{vmatrix} \overline{W}_{00} - W_{00} & -W_{01} \\ -W_{01} & \overline{W}_{11} - W_{11} \end{vmatrix} + W_{12}\begin{vmatrix} \overline{W}_{00} - W_{00} & -W_{01} \\ -W_{02} & -W_{12} \end{vmatrix} - W_{02}\begin{vmatrix} -W_{01} & \overline{W}_{11} - W_{11} \\ -W_{02} & -W_{12} \end{vmatrix} > 0\,. \tag{80}$$

We start by noting that, for all n large enough,

$$\overline{W}_{22} - W_{22} \geq \frac{3\kappa^2}{2p}. \tag{81}$$

Indeed, by Lemma 7 and 8, to prove this claim it is sufficient to show that

$$\frac{\kappa^2}{p} \geq C\left(\frac{\kappa^5\bar{n}^2}{p^6} + \frac{\kappa^3\bar{n}^{1/2}}{p^3}\right), \tag{82}$$

for a large enough constant C or:

$$p \geq C\max\left(n^{2/5}\kappa^{3/5}(\log n)^{2/5}, \kappa^{1/2}(n\log n)^{1/4}\right) \tag{83}$$

This is satisfied when we choose $p \geq c(\kappa \log n)^{1/4} n^{1/6}$ when we choose c a large enough constant. Along with the argument for the second condition above, this implies that:

$$(\overline{W}_{22} - W_{22}) \begin{vmatrix} \overline{W}_{00} - W_{00} & -W_{01} \\ -W_{01} & \overline{W}_{11} - W_{11} \end{vmatrix} \geq \frac{n^3 \kappa^9}{2p^5}, \tag{84}$$

for large enough n.

We now consider the second term. Let $w \equiv C\kappa^4 \bar{n}^{3/2}/p^6$. Then by Lemmas 7 and 8, for all n large enough:

$$0 \leq -W_{12} \begin{vmatrix} \overline{W}_{00} - W_{00} & -W_{01} \\ -W_{02} & -W_{12} \end{vmatrix} \leq \frac{3}{2} w^2 \left(\frac{n^2 \kappa^4}{p^2} + w \right) \tag{85}$$

$$\leq \frac{2n^2 \kappa^4 w^2}{p^2}, \tag{86}$$

as $n^2\kappa^4/p^2 > 2w$ whenever $p \geq (\log n)^{3/8} n^{-1/8}$. As we have $p \geq n^{-1/12}$ this is satisfied.

Similarly, for the third term

$$0 \leq W_{02} \begin{vmatrix} -W_{01} & \overline{W}_{11} - W_{11} \\ -W_{02} & -W_{12} \end{vmatrix} \leq \frac{3w^2}{2} \left(w + \frac{n\kappa^3}{p^2} \right). \tag{87}$$

The second term in the parentheses above dominates when $p \geq \kappa^{1/4} (\log n)^{3/8} n^{1/8}$ which holds as we keep $p \geq c(\kappa \log n)^{1/4} n^{1/6}$. Hence:

$$W_{02} \begin{vmatrix} -W_{01} & \overline{W}_{11} - W_{11} \\ -W_{02} & -W_{12} \end{vmatrix} \leq \frac{2n\kappa^3 w^2}{p^2}. \tag{88}$$

Thus, using Eqs. (84), (86), (88), we conclude that Eq. (73) holds if

$$\frac{n^3 \kappa^9}{2p^5} \geq \frac{2n^2 \kappa^4 w^2}{p^2} + \frac{2n\kappa^3 w^2}{p^2} \tag{89}$$

$$= \frac{2(1 + n\kappa) n \kappa^3 w^2}{p^2}. \tag{90}$$

For this, it suffices that:

$$\frac{n^3 \kappa^9}{p^5} \gtrsim \frac{n^2 \kappa^4 w^2}{p^2}, \tag{91}$$

or, equivalently, $p^9 \geq c_1 n^2 \kappa^3 (\log n)^3$ for an appropriate c_1 large enough. This holds under the stated condition $p \geq c(\kappa \log n)^{1/4} n^{1/6}$ provided c is large enough. This completes the proof of Theorem 1. ■

The proofs of Lemma 8 and 7 follow by a simple calculation and are given in Section A.3.

Our key technical result is Proposition 6. Its proof is organized as follows. We analyze the expectation matrices $\mathbb{E}\{H_{22}\}$, $\mathbb{E}\{H_{12}\}$ in Section A.5. We then control the random components $H_{11} - \mathbb{E}\{H_{11}\}$ in Section A.6, $H_{12} - \mathbb{E}\{H_{12}\}$ in Section A.8, and $H_{22} - \mathbb{E}\{H_{22}\}$ in Section A.7. The application of the moment method to these deviations requires the definition of various specific graph primitives, which we isolate in Section A.4 for easy reference. Finally, we combine the results to establish Proposition 6 in Section A.9.

A.3. Proofs of Lemmas 8 and 7

Proof [Proof of Lemma 8] Recall that W_{00} is defined as:

$$W_{00} = \alpha_3 \bar{n}^{1/2} + C\alpha_4 \bar{n}^{3/2} + \frac{C(\alpha_3 \bar{n})^2}{\alpha_1} + \frac{\left(n\sqrt{n}\alpha_3 p^2 + 2\sqrt{n}\alpha_2 + 3\alpha_3 \bar{n}\right)^2}{n(\alpha_2 p - \alpha_1^2)}. \tag{92}$$

Firstly, since $p \geq c(\kappa \log n)^{1/4} n^{1/6}$, and $n\kappa \geq \log n$, we have that $p \geq n^{-1/12}$ asymptotically. Hence:

$$\frac{n\sqrt{n}\alpha_3 p^2}{\alpha_3 \bar{n}} = \frac{p^2\sqrt{n}}{\log n} \to \infty. \tag{93}$$

Similarly:

$$\frac{n\sqrt{n}\alpha_3 p^2}{\sqrt{n}\alpha_2} = \frac{n\kappa}{2} \to \infty. \tag{94}$$

Also:

$$\frac{\alpha_4 \bar{n}^{3/2}}{n^3\alpha_3^2 p^4 / n(\alpha_2 p - \alpha_1^2)} \lesssim \frac{\kappa^4 \bar{n}^2}{(n^3\kappa^6 / n\kappa^2 p^2)} \tag{95}$$

$$= \frac{\log^2 n}{np^4} \leq \frac{\log^2 n}{n^{7/8}} \to 0. \tag{96}$$

$$\frac{(\alpha_3 \bar{n})^2/\alpha_1)}{n^3\alpha_3^2 p^2 / n(\alpha_2 p - \alpha_1^2)} \lesssim \frac{\kappa \log^2 n}{p^4} \leq \frac{\kappa \log^2 n}{\sqrt{n}} \to 0. \tag{97}$$

$$\frac{\alpha_3 \bar{n}^{1/2}}{\alpha_4 \bar{n}^{3/2}} \leq \frac{\kappa^3/p^3}{\kappa^4 \bar{n}/p^3} = \frac{p^3}{\kappa \bar{n}} \to 0. \tag{98}$$

Hence the term $(n\sqrt{n}\alpha_3 p^2)^2 / n(\alpha_2 p - \alpha_1)$ is dominant in W_{00} and the first claim of the lemma follows.

For W_{01} we have the equation:

$$W_{01} = \alpha_3 \bar{n}^{1/2} + C\alpha_4 \bar{n}^{3/2} + \frac{C}{\alpha_1}(\alpha_3 \bar{n})(\alpha_3(\bar{n} + \sqrt{n}\alpha_2)) + \frac{1}{n(\alpha_2 p - \alpha_1^2)}(n\sqrt{n}\alpha_3 p^2 + 2\sqrt{n}\alpha_2 + \alpha_3 \bar{n})(\alpha_3 \bar{n}). \tag{99}$$

It suffices to check that $C\alpha_4 \bar{n}^{3/2}$ is the dominant term. By the argument in W_{00} we already have that the first term is negligible. Further since, $\alpha_3 \bar{n}/\sqrt{n}\alpha_2 = \kappa\sqrt{n}\log n / p^2 = (\kappa \log n)^{1/2} n^{1/6} \to 0$, to prove that the third term is negligible, it suffices that

$$\frac{(\alpha_3 \bar{n})(\sqrt{n}\alpha_2)}{\alpha_1 \alpha_4 \bar{n}^{3/2}} \leq \frac{\kappa^5 p^{-3}}{\kappa^5 p^{-6}\sqrt{\log n}} = \frac{p^2}{\sqrt{\log n}} \to 0. \tag{100}$$

By the estimates in W_{00} the fourth term is negligible if:

$$\frac{(n\sqrt{n}\alpha_3 p^2)(\alpha_3 \bar{n})}{n(\alpha_2 p - \alpha_1^2)\alpha_4 \bar{n}^{3/2}} \to 0 \tag{101}$$

$$\text{i.e. } \frac{n^{5/2}\log n p^- 4\kappa^6}{n^{5/2}\log n^{3/2} p^{-6}\kappa^6} = \frac{p^2}{\sqrt{\log n}} \to 0. \tag{102}$$

This implies the claim for W_{01}. The calculation for W_{02} and W_{12} is similar.

We now consider W_{11} given by:

$$W_{11} = \alpha_3 \bar{n}^{1/2} + C\alpha_4 \bar{n}^{3/2} + \frac{C}{\alpha_1}\left(C\alpha_3\bar{n} + \sqrt{n}\alpha_3 p^2 + 2\alpha_2\right)^2 + \frac{C(\alpha_3\bar{n})^2}{n(\alpha_2 p - \alpha_1^2)}. \tag{103}$$

As in W_{00}, the first term is negligible. For the third term, first we note that $\alpha_3\bar{n}/\alpha_2 = (\kappa \log n)n/p^2 \geq \log^2 n \to \infty$. Hence to prove that the third term is negligible, it suffices that:

$$\frac{(\alpha_3\bar{n})^2}{\alpha_1\alpha_4\bar{n}^{3/2}} \leq \kappa\sqrt{\bar{n}} \to 0. \tag{104}$$

The final term in W_{11} is negligible by the same argument, since $n(\alpha_2 p - \alpha_1^2) = n\kappa^2 \geq \alpha_1$.

W_{22} is given by:

$$W_{22} = \alpha_3 \bar{n}^{1/2} + C\alpha_4\bar{n} + \frac{C(\alpha_3\bar{n})^2}{\alpha_1} + \frac{C(\alpha_3\bar{n})^2}{n(\alpha_2 p - \alpha_1^2)}. \tag{105}$$

Since $n(\alpha_2 p - \alpha_1^2) = n\kappa^2 \geq \alpha_1$ it is easy to see that the third term dominates the fourth above. To see that the first dominates the second, it suffices that their ratio diverge i.e.

$$\frac{\alpha_3\bar{n}^{1/2}}{\alpha_4\bar{n}} = \frac{p^3}{\kappa\sqrt{\bar{n}}} \tag{106}$$

$$\geq \frac{p^3}{\kappa \log n\sqrt{n}} \tag{107}$$

$$= c^3(\kappa \log n)^{1/8} n^{1/4} \to \infty, \tag{108}$$

as $\kappa \geq 1/n$. Thus we have that the first and third terms dominate the contribution for W_{22}. This completes the proof of the lemma. ■

Proof [Proof of Lemma 7] $\overline{W}_{00}$ is given by:

$$\overline{W}_{00} = \alpha_2 + 2(n-2)\alpha_3 p + \frac{(n-2)(n-3)}{2}\alpha_4 p^4 - \frac{n(n-1)}{2}\alpha_2^2. \tag{109}$$

It is straightforward to check that the third and fourth terms dominates the sum above i.e.:

$$\frac{\overline{W}_{00}}{\frac{(n-2)(n-3)}{2}\alpha_4 p^4 - \frac{n(n-1)}{2}\alpha_2^2} \to 1. \tag{110}$$

Further we have:

$$\frac{(n-2)(n-3)}{2}\alpha_4 p^4 - \frac{n(n-1)}{2}\alpha_2^2 = (1+\delta_n)\frac{2n^2\kappa^4}{p^2}, \tag{111}$$

for some $\delta_n \to 0$. The claim for $\overline{W}_{00}$ then follows.

The claims for $\overline{W}_{11}$ and $\overline{W}_{22}$ follow in the same fashion as above where we instead use the following, adjusting δ_n appropriately:

$$\frac{\overline{W}_{11}}{n\alpha_3 p} = \frac{\alpha_2 + (n-4)\alpha_3 p - (n-3)\alpha_4 p^4}{n\alpha_3 p} \to 1 \tag{112}$$

$$\frac{\overline{W}_{22}}{\alpha_2} = \frac{\alpha_2 - 2\alpha_3 p + \alpha_4 p^4}{\alpha_2} \to 1. \tag{113}$$

■

A.4. Graph definitions and moment method

In this section we define some family of graphs that will be useful in the moment calculations of Sections A.6, A.7 and A.8. We then state and prove a moment method lemma, that will be our basic tool for controlling the norm of random matrices.

Definition 9 *A* cycle *of length m is a graph $D = (V, E)$ with vertices $V = \{v_1, \dots v_m\}$ and edges $E = \{\{v_i, v_{i+1}\} : i \in [m]\}$ where addition is taken modulo m.*

Definition 10 *A* couple *is an ordered pair of vertices (u, v) where we refer to the first vertex in the couple as the* head *and the second as the* tail*.*

Definition 11 *A* bridge *of length $2m$ is a graph $B = (V, E)$ with vertex set $V = \{u_i, v_i, w_i : i \in [m]\}$, and edges $E = \{\{u_i, v_i\}, \{u_i, w_i\}, \{u_{i+1}, v_i\}, \{u_{i+1}, w_i\} : i \in [m]\}$ where addition above is modulo m. We regard (v_i, w_i) for $i \in [m]$ as couples in the bridge.*

Definition 12 *A* ribbon *of length m is a graph $R = (V, E)$ with vertex set $V = \{u_1 \dots u_m, v_1 \dots v_m\}$ and edge set $E = \{\{u_i, u_{i+1}\}, \{u_i, v_{i+1}\}, \{v_i, u_{i+1}\}, \{v_i, v_{i+1}\} : i \in [m]\}$ where addition is modulo m. Further we call the subgraph induced by the 4-tuple $(u_i, v_i, u_{i+1}, v_{i+1})$ a* face *of the ribbon and we call the ordered pairs (u_i, v_i), $i \in [m]$ couples of ribbon.*

Each face of the ribbon has 4 edges, hence there are $\binom{4}{\eta}$ ways to remove $4 - \eta$ edges from the face. We define a ribbons of class η, type ν and length $2m$ as follows.

Definition 13 *For $1 \le \eta \le 4$ and $1 \le \nu \le \binom{4}{\eta}$, we define a ribbon of length $2m$, class η and type ν to be the graph obtained from a ribbon of length $2m$ by keeping η edges in each face of the ribbon, so that the following happens. The subgraphs induced by the tuples $(u_{2i-1}, v_{2i-1}, u_{2i}, v_{2i})$ and $(u_{2i+1}, v_{2i+1}, u_{2i}, v_{2i})$ for $i \ge 1$ are faces of class η and type ν as shown in Table 1.*

For brevity, we write (η, ν)-ribbon to denote a ribbon of class η and type ν.

Definition 14 *A (η, ν)-star ribbon $S = (V, E)$ of length $2m$ is a graph formed from a (η, ν)-ribbon $R(V', E')$ of length $2m$ by the following process. For each face $(u_i, v_i, u_{i+1}, v_{i+1})$ we identify* either *the vertex pair (u_i, u_{i+1}) or the pair (v_i, v_{i+1}) and delete the self loop formed, if any, from the edge set. Note here that the choice of the pair identified can differ across faces of R.*

We let $\mathcal{S}^m_{\eta,\nu}$ denote this collection of (η, ν)-star ribbons.

Definition 15 *A labeled graph is a pair $(F = (V, E), \ell)$ where F is a graph and $\ell : V \to [n]$ maps the vertices of the graph to labels in $[n]$. We define a* valid labeling *to be one that satisfies the following conditions:*

1. *Every couple of vertices (u, v) in the graph satisfies $\ell(u) < \ell(v)$.*
2. *For every edge $e = \{v_1, v_2\} \in E$, $\ell(v_1) \neq \ell(v_2)$.*

A labeling of F is called contributing *if, in addition to being valid, the following happens. For every edge $e = \{u, v\} \in E$, there exists an edge $e' = \{u', v'\} \neq e$ such that $\{\ell(u), \ell(v)\} = \{\ell(u'), \ell(v')\}$. In other words, a labeling is contributing if it is valid and has the property that every labeled edge occurs* at least twice *in F.*

Remark 16 *Suppose F is one of the graphs defined above and C is a face of F. We write, with slight abuse of notation, $C \subseteq F$ to denote "a face C of the graph F". Furthermore, to lighten notation, we will often write $e \in F$ for an edge e in the graph F.*

Definition 17 *Let $\mathfrak{L}(F)$ denote the set of valid labelings of a graph $F = (V, E)$ and $\mathfrak{L}_2(F)$ denote the set of contributing labelings. Further, we define*

$$v_*(F) = \max_{\ell \in \mathfrak{L}_2(F)} \mathsf{range}(\ell) \tag{114}$$

where $\mathsf{range}(\ell) = \{i \in [n] : i = \ell(u), u \text{ is a vertex in } F\}$.

The following is a simple and general moment method lemma.

Lemma 18 *Given a matrix $X \in \mathbb{R}^{m' \times n'}$, suppose that there exist constants c_1, $c_2, c_3, c_4, c_5 \geq 0$ satisfying $c_2 \geq c_4$ and for any integer $r > 0$:*

$$\mathbb{E}\mathsf{Tr}\{(X^{\mathsf{T}}X)^r\} \leq \binom{n}{c_1 r + c_2} (c_5)^{2r} (c_1 r + c_2)^{c_3 r + c_4}. \tag{115}$$

Then, for every n large enough, with probability exceeding $1 - n^{-(\Gamma - c_2)/2}$ we have that

$$\|X\|_2 \leq c_4 \sqrt{\exp(c_1 \Gamma) n^{c_1} (\log n)^{c_3 - c_1}}. \tag{116}$$

Proof By rescaling X we can assume that $c_5 = 1$. Since $\mathsf{Tr}\{(X^{\mathsf{T}}X)^{2r}\} = \sum_i (\sigma_i(X))^{2r}$ where $\sigma_i(X)$ are the singular values of X ordered $\sigma_1(X) \geq \sigma_2(X) \dots \sigma_N(X)$, we have that:

$$\|X\|_2^{2r} = \sigma_1(X)^{2r} \leq \mathsf{Tr}\{(X^{\mathsf{T}}X)^r\}. \tag{117}$$

Then, by Markov inequality and the given assumption:

$$\mathbb{P}\{\|X\|_2 \geq t\} \leq \mathbb{P}\left\{\mathsf{Tr}\{(X^{\mathsf{T}}X)^{2r}\} \geq t^{2r}\right\} \tag{118}$$

$$\leq t^{-2r} \mathbb{E}\mathsf{Tr}\left\{(X^{\mathsf{T}}X)^{2r}\right\} \tag{119}$$

$$\leq \binom{n}{c_1 r + c_2} (c_1 r + c_2)^{c_3 r + c_4}. \tag{120}$$

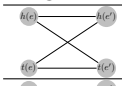
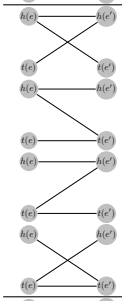
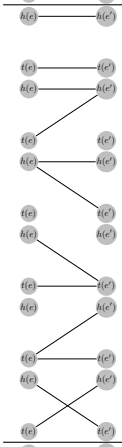
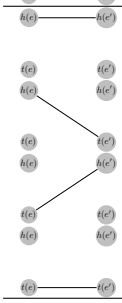
Figure	Ribbon class(η)	Ribbon type(ν)	Typical norm
	4	1	$\bar{n}$
	3	1	$\bar{n}$
	3	2	$\bar{n}$
	3	3	$\bar{n}$
	3	4	$\bar{n}$
	2	1	$\bar{n}$
	2	2	$\bar{n}^{3/2}$
	2	3	$\bar{n}^{3/2}$
	2	4	$\bar{n}^{3/2}$
	2	5	$\bar{n}^{3/2}$
	2	6	$\bar{n}$
	1	1	$\bar{n}^{3/2}$
	1	2	$\bar{n}^{3/2}$
	1	3	$\bar{n}^{3/2}$
	1	4	$\bar{n}^{3/2}$

Table 1: Definition of the different ribbon classes and types.

Using $\binom{n}{k} \leq (ne/k)^k$ we have:

$$\mathbb{P}\left\{\|X\|_2 \geq t\right\} \leq t^{-2r}(ne)^{c_1 r + c_2}(c_1 r + c_2)^{(c_3 - c_1)r + c_4 - c_2} \tag{121}$$

$$= \exp\left\{(c_1 r + c_2)(\log n + 1) + ((c_3 - c_1)r + c_4 - c_2)\log(c_1 r + c_2) - 2r \log t\right\}. \tag{122}$$

Setting $r = \lceil (\log n - c_2)/c_1 \rceil$ and using $c_2 \geq c_4$ we obtain the bound:

$$\mathbb{P}\left\{\|X\|_2 \geq t\right\} \leq \exp\left\{\log n(\log n + 1) + (c_3/c_1 - 1)(\log n)\log\log n - (\log n - c_2)\log(t^{2/c_1})\right\} \tag{123}$$

$$\leq \exp\left\{\log n \log\left(ne(\log n)^{c_3/c_1 - 1}\right) - (\log n - c_2)\log(t^{2/c_1})\right\}. \tag{124}$$

We can now set $t = \left\{\exp(\Gamma) n (\log n)^{c_3/c_1 - 1}\right\}^{c_1/2}$ whereupon the bound on the right hand side is at most $n^{-(\Gamma - c_2)/2}$ for every n large enough. This yields the claim of the lemma. ■

The next lemma specialized the previous one to the type of random matrices we will be interested in.

Lemma 19 *For a matrix $X \in \mathbb{R}^{m' \times n'}$, suppose there exists a sequence of graphs $G_X(r)$ with vertex, edge sets V_r, E_r respectively, a set $\mathfrak{L}(G_X(r))$ of labelings $\ell : V_r \to [n]$ and a constant $\beta > 0$ such that:*

$$\mathsf{Tr}\left\{(X^{\mathsf{T}}X)^r\right\} = \beta^{2r} \sum_{\ell \in \mathfrak{L}(G_X(r))} \prod_{e \in G_X(r)} g_{\ell(e)}, \tag{125}$$

where, for $e = \{u, v\}$, $\ell(e) = \{\ell(u), \ell(v)\}$. Let $\mathfrak{L}_2(G_X(r)) \subseteq \mathfrak{L}(G_X(r))$ denote the subset of contributing labelings (i.e. the set of labelings $\ell \in \mathfrak{L}(G_X(r))$ such that every labeled edge in $G_X(r)$ is repeated at least twice*). Further define $v(r)$ and $v^*(r)$ by:*

$$v(r) \equiv |V_r| \,, \tag{126}$$

$$v_*(r) \equiv v_*(G_X(r)). \tag{127}$$

Then

$$\mathbb{E}\mathsf{Tr}\left\{(X^{\mathsf{T}}X)^r\right\} \le \beta^{2r} \left|\mathfrak{L}_2(G_X(r))\right| \tag{128}$$

$$\le \binom{n}{v_*(r)} \beta^{2r} v_*(r)^{v(r)}. \tag{129}$$

Proof By rescaling X it suffices to show the case $\beta = 1$. Taking expectations on either side of Eq. (125) we have that:

$$\mathbb{E}\mathsf{Tr}\left\{(X^{\mathsf{T}}X)^r\right\} = \sum_{\ell \in \mathfrak{L}(G_X(r))} \mathbb{E}\left\{\prod_{e \in G_X(r)} g_{\ell(e)}\right\}. \tag{130}$$

The variables $g_{\ell(e)}$ are centered and independent and bounded by 1. Hence the only terms that do not vanish in the summation above correspond to labelings ℓ wherein every labeled edge occurs at least twice, i.e. precisely when $\ell \in \mathfrak{L}_2(G_X(r))$. By the boundedness of $g_{\ell(e)}$, the contribution of each non-vanishing term is at most 1, hence

$$\mathbb{E}\mathsf{Tr}\left\{(X^{\mathsf{T}}X)^r\right\} \le \left|\mathfrak{L}_2(G_X(r))\right|. \tag{131}$$

It now remains to prove that $|\mathfrak{L}_2(G_X(r)| \le \binom{n}{v_*(r)} v(r)^{v(r)}$. By definition, ℓ can map the vertices in V_r to at most $v_*(r)$ distinct labels. There are at most $\binom{n}{v_*(r)}$ distinct ways to pick these labels in $[n]$, and at most $v_*(r)^{v(r)}$ ways to assign the $v_*(r)$ labels to $v(r)$ vertices, yielding the required bound. ■

Lemma 20 *Consider the setting of Lemma 19. If we additionally have*

$$v_*(r) \leq c_1 r + c_2 \tag{132}$$

$$v(r) = c_3 r + c_4, \tag{133}$$

where $c_3 \leq 2c_1$ *then* $\|X\|_2 \lesssim \beta \bar{n}^{c_1/2}$ *with probability at least* $1 - n^{-5}$.

Proof The proof follows by combining Lemmas 19 and 18. ■

A.5. The expected values $\mathbb{E}\{H_{22}\}, \mathbb{E}\{H_{12}\}$

In this section we characterize the eigenstructure of the expectations $\mathbb{E}\{H_{22}\}$, $\mathbb{E}\{H_{12}\}$. These can be viewed as linear operators on $\mathbb{R}^{\binom{[n]}{2}}$ that are invariant under the action of permutations[2] on $\mathbb{R}^{\binom{[n]}{2}}$. By Schur's Lemma Serre (1977), their eigenspace decomposition corresponds to the decomposition of $\mathbb{R}^{\binom{[n]}{2}}$ into irreducible subrepresentations of the group of permutations. This is given by $\mathbb{R}^{\binom{[n]}{2}} = \mathbb{V}_0 \oplus \mathbb{V}_1 \oplus \mathbb{V}_2$, where

$$\mathbb{V}_0 \equiv \{v \in \mathbb{R}^{\binom{[n]}{2}} : \exists u \in \mathbb{R} \text{s.t.} v_{\{i,j\}} = u \text{ for all } i < j\} \tag{134}$$

$$\mathbb{V}_1 \equiv \{v \in \mathbb{R}^{\binom{[n]}{2}} : \exists u \in \mathbb{R}^n, \text{s.t.} \langle 1_n, u\rangle = 0 \text{ and } \quad v_{\{i,j\}} = u_i + u_j \text{ for all } i < j\} \tag{135}$$

$$\mathbb{V}_2 \equiv (\mathbb{V}_0 \oplus \mathbb{V}_1)^{\perp}. \tag{136}$$

An alternative approach to defining the spaces $\mathbb{V}_a$ is to let $\mathbb{V}_0 = \mathsf{span}(v_0), \mathbb{V}_1 = \mathsf{span}(v_1^i, i = 1 \dots n), \mathbb{V}_2 = \mathsf{span}(v_2^{ij}, 1 \leq i < j \leq n)$, where

$$(v_0)_A = \sqrt{\frac{2}{n(n-1)}} \tag{137}$$

$$(v_1^i)_A = \begin{cases} \sqrt{\frac{n-2}{n(n-1)}} & \text{if } A = \{i, \cdot\} \\ -\frac{2}{\sqrt{n(n-1)(n-2)}} & \text{otherwise.} \end{cases} \tag{138}$$

$$(v_2^{ij})_A = \begin{cases} \sqrt{\frac{n-3}{n-1}} & \text{if } A = \{i,j\} \\ -\frac{1}{n-2}\sqrt{\frac{n-3}{n-1}} & \text{if } A = \{i,\cdot\} \text{ or } \{j,\cdot\} \\ \frac{1}{\binom{n-2}{2}}\sqrt{\frac{n-3}{n-1}} & \text{otherwise.} \end{cases} \tag{139}$$

Notice that $\dim(\mathbb{V}_0) = 1, \dim(\mathbb{V}_1) = n-1, \dim(\mathbb{V}_2) = n(n-3)/2$, and that $\{v_1^i\}_{i\in[n]}, \{v_1^{i,j}\}_{i,j\in[n]}$ are overcomplete sets. For $a \in \{0,1,2\}$, we denote by V_a the matrix whose rows are given by this overcomplete basis of $\mathbb{V}_a$

It is straightforward to check that the two definitions of the orthogonal decomposition $\mathbb{R}^{\binom{[n]}{2}} = \mathbb{V}_0 \oplus \mathbb{V}_1 \oplus \mathbb{V}_2$ given above coincide. We let $\mathcal{P}_a \in \mathbb{R}^{\binom{[n]}{2}\times\binom{[n]}{2}}$ denote the orthogonal projector on the space $\mathbb{V}_a$.

The following proposition gives the eigenstructure of $\mathbb{E}\{H_{22}\}$.

2. A permutation $\sigma : [n] \to [n]$ acts on $\mathbb{R}^{\binom{[n]}{2}}$ by permuting the indices in $\binom{[n]}{2}$ in the obvious way, namely $\sigma(\{i,j\}) = \{\sigma(i), \sigma(j)\}$.

Proposition 21 *The matrix $\mathbb{E}\{H_{22}\}$ has the following spectral decomposition*

$$\mathbb{E}\{H_{22}\} = \lambda_0 \mathcal{P}_0 + \lambda_1 \mathcal{P}_1 + \lambda \mathcal{P}_2 \,, \tag{140}$$

where

$$\lambda_0 = \alpha_2 + 2(n-2)\alpha_3 p + \frac{(n-2)(n-3)}{2}\alpha_4 p^4 - \frac{n(n-1)}{2}\alpha_2^2 \,, \tag{141}$$

$$\lambda_1 = \alpha_2 + (n-4)\alpha_3 p - (n-3)\alpha_4 p^4 \,, \tag{142}$$

$$\lambda_2 = \alpha_2 - 2\alpha_3 p + \alpha_4 p^4 . \tag{143}$$

Proof It is straightforward to verify that the vectors v_ℓ^A defined above are eigenvectors of $\mathbb{E}\{H_{22}\}$. The eigenvalues are then given by $\lambda_\ell = \langle v_\ell^A, \mathbb{E}\{H_{22}\} v_\ell^A \rangle$ for an arbitrary choice of $A = \{i\}$ or $\{i,j\}$. ■

Remark 22 *The above eigenvalues can also be computed using Meka and Wigderson (2013) which relies on the theory of association schemes. We preferred to present a direct and self-contained derivation.*

We now have a similar proposition for $\mathbb{E}\{H_{12}\} \in \mathbb{R}^{\binom{[n]}{1} \times \binom{[n]}{2}}$. More precisely, we decompose $\mathbb{R}^{\binom{[n]}{1}}$ in $\mathsf{span}(1_m)$ and its orthogonal complement, and $\mathbb{R}^{\binom{[n]}{2}} = \mathbb{V}_0 \oplus \mathbb{V}_1 \oplus \mathbb{V}_2$ as above.

Proposition 23 *The following hold for all n large enough:*

$$\mathcal{Q}_n^\perp \mathbb{E}\{H_{12}\} \mathcal{P}_0 = 0 \tag{144}$$

$$\left\| \mathcal{Q}_n^\perp \mathbb{E}\{H_{12}\} \mathcal{P}_1 \right\|_2 \le \sqrt{n}\alpha_2 \tag{145}$$

$$\mathcal{Q}_n^\perp \mathbb{E}\{H_{12}\} \mathcal{P}_2 = 0 \tag{146}$$

$$\| \mathcal{Q}_n \mathbb{E}\{H_{12}\} \mathcal{P}_0 \|_2 \le n^{3/2}\alpha_3 p^2 + 2\sqrt{n}\alpha_2 \tag{147}$$

$$\mathcal{Q}_n \mathbb{E}\{H_{12}\} \mathcal{P}_1 = 0 \tag{148}$$

$$\mathcal{Q}_n \mathbb{E}\{H_{12}\} \mathcal{P}_2 = 0. \tag{149}$$

Proof For $A \in \binom{[n]}{1}$ and $B \in \binom{[n]}{2}$:

$$(\mathbb{E}\{H_{12}\})_{A,B} = \begin{cases} \alpha_3 p^2 - \alpha_1 \alpha_2 & \text{if } |A \cap B| = 0 \\ \alpha_2 - \alpha_1 \alpha_2 & \text{if } |A \cap B| = 1. \end{cases} \tag{150}$$

Recall from the definition of the space $\mathbb{V}_1 = \mathsf{span}(\{v_1^A\}_{A \in \binom{[n]}{1}})$. We can write $\mathbb{E}\{H_{12}\}$ as:

$$\mathbb{E}\{H_{12}\} = \frac{\binom{n-1}{2}(\alpha_3 p^2 - \alpha_1 \alpha_2) + (n-1)(\alpha_2 - \alpha_3 p^2)}{\sqrt{\binom{n}{2}}} 1_n v_0^{\mathsf{T}} + \sqrt{\frac{(n-1)(n-2)}{n}}(\alpha_2 - \alpha_3 p^2) V_1 . \tag{151}$$

This implies all but the second and the fourth claims immediately as $V_1\mathcal{P}_0 = V_1\mathcal{P}_2 = 0$, $\mathcal{Q}_n V_1 = 0$ and $\mathcal{Q}_n^{\perp} 1_n = 0$. For the second claim, the above decomposition yields:

$$\left\|\mathcal{Q}_n^{\perp}\mathbb{E}\{H_{12}\}\mathcal{P}_1\right\|_2 = \max_{x\in\mathbb{V}_1:\|x\|_2\le 1}\left\|\sqrt{\frac{(n-1)(n-2)}{n}}(\alpha_2-\alpha_3 p^2)V_1 x\right\|_2 \tag{152}$$

$$= \sqrt{\frac{(n-1)(n-2)}{n}}(\alpha_2-\alpha_3 p^2)\sqrt{\lambda_{\max}(V_1 V_1^{\mathsf{T}})}. \tag{153}$$

Since $\langle v_1^A, v_1^{A'}\rangle = -1/(n-1)$ when $A \neq A'$ and 1 otherwise, we have that:

$$V_1 V_1^{\mathsf{T}} = \frac{n}{n-1}\,\mathrm{I}_n - \frac{1}{n-1}\,1_n(1_n)^{\mathsf{T}}, \tag{154}$$

hence $\lambda_{\max}(V_1 V_1^{\mathsf{T}}) = n/(n-1)$. This implies that:

$$\left\|\mathcal{Q}_n^{\perp}\mathbb{E}\{H_{12}\}\mathcal{P}_1\right\|_2 = \sqrt{n-2}(\alpha_2-\alpha_3 p^2) \le \sqrt{n}\alpha_2. \tag{155}$$

For the fourth claim, the expression for $\mathbb{E}\{H_{12}\}$ above yields that:

$$\|\mathcal{Q}_n\mathbb{E}\{H_{12}\}\mathcal{P}_0\|_2 = \frac{\binom{n-1}{2}(\alpha_3 p^2-\alpha_1\alpha_2)+(n-1)(\alpha_2-\alpha_3 p^2)}{\sqrt{\binom{n}{2}}}\sqrt{n} \tag{156}$$

$$\le \frac{\binom{n-1}{2}\alpha_3 p^2}{\sqrt{\frac{n-1}{2}}} + \frac{(n-1)\alpha_2}{\sqrt{\frac{n-1}{2}}} \tag{157}$$

$$\le n\sqrt{n}\alpha_3 p^2 + 2\sqrt{n}\alpha_2. \tag{158}$$

■

A.6. Controlling $H_{11} - \mathbb{E}\{H_{11}\}$

The block H_{11} is a linear combination of the identity and the adjacency matrix of G. Hence, its spectral properties are well understood, since the seminal work of Füredi-Komlós Füredi and Komlós (1981). While the nest proposition could be proved using these results, we present an self-contained proof for pedagogical reasons, as the same argument will be repeated several times later for more complex examples.

Proposition 24 *Suppose that $\underline{\alpha}$ satisfies:*

$$\frac{\alpha_1}{2} - \alpha_2 p \gtrsim \alpha_2 \bar{n}^{1/2}\,, \tag{159}$$

$$\alpha_2 p - \alpha_1^2 \ge 0\,, \qquad \alpha_1 \ge 0. \tag{160}$$

Then with probability at least $1-n^{-5}$:

$$H_{11} \succcurlyeq 0\,, \tag{161}$$

$$H_{11}^{-1} \preccurlyeq \frac{1}{n(\alpha_1 p-\alpha_1^2)}\mathcal{Q}_n + \frac{2}{\alpha_1}\mathcal{Q}_n^{\perp} \tag{162}$$

Proof First, note that:

$$\mathbb{E}\{H_{11}\} = (\alpha_1 - \alpha_2 p)\mathrm{I}_n + (\alpha_2 p - \alpha_1^2)\, n\mathcal{Q}_n\,. \tag{163}$$

Furthermore, for $A, B \in \binom{[n]}{1}$, $A \neq B$, $(H_{11} - \mathbb{E}\{H_{11}\})_{A,B} = \alpha_2 g_{AB}$. Here, we identify elements of $\binom{[n]}{1}$ with elements of $[n]$ in the natural way. Thus, expanding $\mathsf{Tr}\left\{((H_{11} - \mathbb{E}H_{11})^{\mathsf{T}}(H_{11} - \mathbb{E}H_{11}))^m\right\}$ we obtain:

$$\mathsf{Tr}\left\{\left((H_{11} - \mathbb{E}\{H_{11}\})^{\mathsf{T}}(H_{11} - \mathbb{E}\{H_{11}\})\right)^m\right\} = \alpha_2^{2m} \sum_{A_1 \dots A_m, A_1' \dots A_m'} \prod_{\ell=1}^{m} g_{A_\ell A_\ell'} g_{A_{\ell+1} A_\ell'}, \tag{164}$$

where we set $A_{m+1} \equiv A_1$. Let $D(m)$ be a cycle of length $2m$, V_D, E_D be its vertex and edge sets respectively, and ℓ be a labeling that assigns to the vertices labels $A_1, A_1', A_2, A_2' \dots A_m, A_m'$ in order. Then the summation over indices $A_1 \dots A_m, A_1' \dots A_m'$ can be expressed as a sum over such labelings of the cycle $D(m)$, i.e.:

$$\mathsf{Tr}\left\{\left((H_{11} - \mathbb{E}\{H_{11}\})^{\mathsf{T}}(H_{11} - \mathbb{E}\{H_{11}\})\right)^m\right\} = \alpha_2^{2m} \sum_{\ell \in \mathfrak{L}(D)} \prod_{e=\{u,v\} \in E_D} g_{\ell(u)\ell(v)}. \tag{165}$$

Let $\mathfrak{L}_2(D(m))$ denote the set of contributing labelings of $D(m)$. By Lemma 20, it suffices to show that $\max_{\ell \in \mathfrak{L}_2(D(m))} |\mathsf{range}(\ell)| \leq m + 1$. Since for a contributing labeling ℓ of $D(m)$, every edge must occur at least twice, there are at most m unique labelings of the edges of $D(m)$. If we consider the graph obtained from (D, ℓ) by identifying in D the vertices with the same label, we obtain a connected graph with at most m edges, hence at most $m+1$ unique vertices. This implies that there are at most $m + 1$ unique labels in the range of a contributing labeling ℓ. Hence with probability at least $1 - n^{-5}$:

$$\|H_{11} - \mathbb{E}\{H_{11}\}\|_2 \lesssim \alpha_2 \bar{n}^{1/2}, \tag{166}$$

Hence with the same probability:

$$H_{11} \succcurlyeq (\alpha_1 - \alpha_2 p - C\alpha_2 \bar{n}^{1/2})\mathrm{I}_n + (\alpha_2 p - \alpha_1^2)\, n\, \mathcal{Q}_n, \tag{167}$$

for some constant C. Under the condition $\alpha_1/2 - \alpha_2 p \gtrsim \alpha_2 \bar{n}^{1/2}$ (with a sufficiently large constant which we suppress) we have that:

$$H_{11} \succcurlyeq \frac{\alpha_1}{2}\mathrm{I}_n + (\alpha_2 p - \alpha_1^2)\, n\, \mathcal{Q}_n\,, \tag{168}$$

or, equivalently,

$$H_{11} \succcurlyeq \frac{\alpha_1}{2}\mathcal{Q}_n^{\perp} + (\alpha_2 p - \alpha_1^2)\, n\, \mathcal{Q}_n. \tag{169}$$

Inverting this inequality yields the claim for H_{11}^{-1}. This completes the proof of the proposition. ■

A.7. Controlling $H_{22} - \mathbb{E}\{H_{22}\}$

The following proposition is the key result of this subsection.

Proposition 25 *With probability at least* $1 - 25n^{-5}$ *the following hold:*

$$\text{For } a \in \{0,1\} \qquad \|\mathcal{P}_a(H_{22} - \mathbb{E}\{H_{22}\})\mathcal{P}_a\|_2 \lesssim \alpha_3 \bar{n}^{1/2} + \alpha_4 \bar{n}^{3/2}\,, \tag{170}$$

$$\|\mathcal{P}_2(H_{22} - \mathbb{E}\{H_{22}\})\mathcal{P}_2\|_2 \lesssim \alpha_3 \bar{n}^{1/2} + \alpha_4 \bar{n}\,, \tag{171}$$

$$\text{For } a \neq b \in \{0,1,2\} \qquad \|\mathcal{P}_a(H_{22} - \mathbb{E}\{H_{22}\})\mathcal{P}_b\|_2 \lesssim \alpha_3 \bar{n}^{1/2} + \alpha_4 \bar{n}^{3/2}\,. \tag{172}$$

Recall that:

$$(H_{22})_{A,B} = \begin{cases} -\alpha_2^2 + \alpha_2 & \text{if } A = B \\ -\alpha_2^2 + \alpha_3(p + g_{t(A)t(B)}) & \text{if } h(A) = h(B), A \neq B \\ -\alpha_2^2 + \alpha_3(p + g_{h(A)t(B)}) & \text{if } t(A) = h(B), A \neq B \\ -\alpha_2^2 + \alpha_3(p + g_{t(A)h(B)}) & \text{if } h(A) = t(B), A \neq B \\ -\alpha_2^2 + \alpha_3(p + g_{h(A)h(B)}) & \text{if } t(A) = t(B), A \neq B \\ -\alpha_2^2 + \alpha_4(p + g_{h(A)h(B)})(p + g_{h(A)t(B)})(p + g_{t(A)h(B)})(p + g_{t(A)t(B)}) & \text{if } |A \cap B| = 0. \end{cases} \tag{173}$$

When $|A \cap B| = 0$ (last case above) we can expand $H_{A,B}$ as a sum of sixteen terms:

$$\begin{aligned} H_{A,B} &= \alpha_4(p + g_{h(A)h(B)})(p + g_{h(A)t(B)})(p + g_{t(A)h(B)})(p + g_{t(A)t(B)}) - \alpha_2^2 \qquad (174) \\ &= (\alpha_4 p^4 - \alpha_2^2) + \alpha_4 p^3(g_{h(A)h(B)} + g_{h(A)t(B)} + g_{t(A)h(B)} + g_{t(A)t(B)}) \\ &\quad + \alpha_4 p^2(g_{h(A)h(B)}g_{h(A)t(B)} + g_{h(A)h(B)}g_{t(A)h(B)} + g_{h(A)h(B)}g_{t(A)t(B)} \\ &\qquad + g_{h(A)t(B)}g_{t(A)h(B)} + g_{h(A)t(B)}g_{t(A)t(B)} + g_{t(A)h(B)}g_{t(A)t(B)}) \\ &\quad + \alpha_4 p(g_{h(A)h(B)}g_{h(A)t(B)}g_{t(A)h(B)} + g_{h(A)h(B)}g_{h(A)t(B)}g_{t(A)t(B)} \\ &\qquad + g_{h(A)h(B)}g_{t(A)h(B)}g_{t(A)t(B)} + g_{h(A)t(B)}g_{t(A)h(B)}g_{t(A)t(B)}) \\ &\quad + \alpha_4 g_{h(A)h(B)}g_{h(A)t(B)}g_{t(A)h(B)}g_{t(A)t(B)}. \qquad (175) \end{aligned}$$

Compactly, we can represent the above summation as follows. Each term above is indexed by a pair (η, ν) where $0 \leq \eta \leq 4$ denotes the number of variables $g_{\cdot,\cdot}$ occurring in the product, and $\nu \leq \binom{4}{\eta}$ determines exactly which η-tuple of g variables occur. For instance, when $\eta = 1$, we have $\binom{4}{1}$ terms $\alpha_4 p^3 g_{h(A)h(B)}, \alpha_4 p^3 g_{h(A)t(B)}, \alpha_4 p^3 g_{t(A)h(B)}, \alpha_4 p^3 g_{t(A)t(B)}$. Equivalently, if $R_{A,B}(\eta, \nu)$ is a labeled (η, ν)-ribbon with exactly one face and vertices labeled $h(A), t(A), h(B), t(B)$ in order, each term corresponds to one specific class and type of ribbon, i.e.

$$H_{A,B} = \sum_{\eta,\nu} \alpha_4 p^{4-\eta} \prod_{e=\{i,j\} \in R_{A,B}(\eta,\nu)} g_{ij}.$$

The exact mapping of the pair (η, ν) to the choice of edges in $R_{A,B}(\eta, \nu)$ is given in Table 1. With a slight abuse of terminology, we refer to η as the *class* and ν the *type* of the term. We define the

matrices $J_{\eta,\nu}$ (for $\eta = 1, 2, 3, 4$ and $\nu = \binom{4}{\eta}$) and K as follows.

$$(J_{\eta,\nu})_{A,B} \equiv \begin{cases} \alpha_4 p^{4-\eta} \prod_{\{i,j\}\in R_{A,B}(\eta,\nu)} g_{ij} & \text{if } |A \cap B| = 0\,, \\ 0 & \text{otherwise.} \end{cases} \tag{176}$$

$$K_{A,B} \equiv \begin{cases} \alpha_3 g_{t(A)t(B)} & \text{if } h(A) = h(B), A \neq B\,, \\ \alpha_3 g_{h(A)t(B)} & \text{if } t(A) = h(B), A \neq B\,, \\ \alpha_3 g_{t(A)h(B)} & \text{if } h(A) = t(B), A \neq B\,, \\ \alpha_3 g_{h(A)h(B)} & \text{if } t(A) = t(B), A \neq B\,, \\ 0 & \text{otherwise.} \end{cases} \tag{177}$$

The matrices $J_{\eta,\nu}$ vanish on the set of entries A, B where A and B have non-zero intersection. This causes the failure of certain useful spectral properties with respect to the spaces $\mathbb{V}_0, \mathbb{V}_1, \mathbb{V}_2$. Consequently, for our proof, it is useful to define the matrices $\widetilde{J}_{\eta,\nu}$ that do not have this constraint.

$$(\widetilde{J}_{\eta,\nu})_{A,B} \equiv \alpha_4 p^{4-\eta} \prod_{\{i,j\}\in R_{A,B}(\eta,\nu)} g_{ij}. \tag{178}$$

Here we ignore the constraint that A, B do not intersect, and follow the convention that $g_{ii} = 0$ for every $i \in [n]$.

Thus, with Eq. (173) we arrive at the following expansion:

$$H_{22} - \mathbb{E}\{H_{22}\} = K + \sum_{\eta=1}^{4} \sum_{\nu=1}^{\binom{4}{\eta}} J_{\eta,\nu} \tag{179}$$

$$= K + J_{2,1} + J_{2,6} + J_{4,1} + \sum_{\nu=1}^{4} J_{3,\nu} + \sum_{\nu=1}^{4} (J_{1,\nu} - \widetilde{J}_{1,\nu}) + \sum_{\nu=2}^{5} (J_{2,\nu} - \widetilde{J}_{2,\nu})$$
$$+ \sum_{\nu=1}^{4} \widetilde{J}_{1,\nu} + \sum_{\nu=2}^{5} \widetilde{J}_{2,\nu}. \tag{180}$$

We now prove a sequence of lemmas regarding the spectral properties of the matrices $K, J_{\eta,\nu}$. The first one concerns the case $\eta = 2$, $\nu = 1, 6$ and $\eta = 4$, $\nu = 1$.

Lemma 26 *With probability at least $1 - 3n^{-5}$, we have that:*

$$\|J_{2,1} + J_{2,6} + J_{4,1}\|_2 \lesssim \alpha_4 \bar{n} \tag{181}$$

Proof By the triangle inequality:

$$\|J_{2,1} + J_{2,6} + J_{4,1}\|_2 \leq \|J_{2,1}\|_2 + \|J_{2,6}\|_2 + \|J_{4,1}\|_2\,. \tag{182}$$

We prove that with probability at least $1 - n^{-5}$

$$\|J_{\eta,\nu}\|_2 \lesssim \alpha_4 \bar{n}, \tag{183}$$

for $(\eta, \nu) = (2,1), (2,6), (4,1)$. The claim then follows by a union bound.

Let $R(\eta, \nu, m)$ denote a (η, ν)-ribbon of length $2m$. Then, by expanding the product we have:

$$\mathsf{Tr}\left\{(J_{\eta,\nu}^{\mathsf{T}} J_{\eta,\nu})^r\right\} = \sum_{\ell \in \mathfrak{L}(R(\eta,\nu,m))} (\alpha_4 p^{4-\eta})^{2m} \left\{\prod_{e \in R(\eta,\nu,m)} g_{\ell(e)}\right\}. \tag{184}$$

Here we write $\ell(e)$ in place of the pair $\ell(u), \ell(v)$ when u, v are the end vertices of e. Since $R(\eta, \nu, m)$ has $4m+2$ vertices, by Lemma 20 it suffices to prove that $\max_{\ell \in \mathfrak{L}_2(R(\eta,\nu,m))} \mathsf{range}(\ell) = 2m + 2$.

We first prove this for the case $\eta = 2$ and $\nu = 1, 6$. Let ℓ be a contributing labeling of the ribbon $R(\eta, \nu, m)$ of length $2m$. Let $\mathbf{G}(\eta, \nu)$ denote the graph obtained by identifying in $R(\eta, \nu, m)$ every vertex with the same label according to ℓ. We have:

$$\#\text{ connected components in } \mathbf{G}(\eta,\nu) \leq \#\text{ connected components in } R(\eta,\nu,m) = 2 \tag{185}$$

$$\#\text{ edges in } \mathbf{G}_{\eta,\nu} \leq \frac{\#\text{ edges in } R(\eta,\nu,m)}{2} = 2m. \tag{186}$$

It follows that there are at most $2m + 2$ unique vertices in $\mathbf{G}(\eta, \nu, m)$ and hence, at most $2m + 2$ unique labels in $\mathsf{range}(\ell)$.

We now prove the condition $\max_{\ell \in \mathfrak{L}_2(R(\eta,\nu,m))} \mathsf{range}(\ell) = 2m + 2$ for $\eta = 4, \nu = 1$, induction on m. The base case is $m = 1$ (or a ribbon of length 2), wherein it is obvious that a contributing labeling ℓ can have at most $4 = 2m + 2$ unique labels. Now, assume the claim is true for ribbons of length at most $2m > 1$ and we will prove it for $R(4, 1, m + 1)$ of length $2m + 2$. Consider any contributing labeling ℓ of $R(4, 1, m + 1)$. We now have the following cases

1. For every vertex $u \in R(4, 1, m + 1)$, there exists $u' \neq u$ such that $\ell(u') = \ell(u)$.
2. There exists vertex $u \in R(4, 1, m + 1)$ with a unique label $i = \ell(u)$ and the degree of u is 4.

For case 1, if every label in the range of ℓ occurs at least twice in $R(4, 1, m)$, the number of unique labels is bounded by $2(m+1)$, since $R(4, 1, m)$ has only $4(m+1)$ vertices, hence the claim follows.

For case 2, let (u_1, v_1) and (u_2, v_2) be the neighboring couples of u. If u is connected to all of u_1, v_1, u_2, v_2, since the edges connected to u must occur twice, it must hold that $\ell(u_1) = \ell(u_2)$ and $\ell(v_1) = \ell(v_2)$ (recall indeed that $\ell(u_1) < \ell(v_1)$, $\ell(u_2) < \ell(v_2)$ by definition of a valid labeling). Hence, we can contract the ribbon removing the couple containing u and all edges and identifying the couples (u_1, v_1) with (u_2, v_2). We obtain now a ribbon $\tilde{R}4, 1, m)$ of length $2m$ and an induced labeling $\tilde{\ell}$ thereof which is contributing. By induction hypothesis, $\mathsf{range}(\tilde{\ell}) \leq 2m + 2$, hence $\mathsf{range}(\ell) = \mathsf{range}(\tilde{\ell}) + 2 \leq 2(m + 1) + 2$. This completes the proof. ∎

Lemma 27 *With probability at least* $1 - 8n^{-5}$*, we have*

$$\left\|\sum_{\nu=1}^{4} J_{3,\nu}\right\|_2 \lesssim \alpha_4 p\bar{n}\,. \tag{187}$$

Proof By the triangle inequality, it suffices to show that for $\nu \in \{1, \dots, 4\}$, with probability $1 - n^{-(\Gamma-2)/2}$:

$$\|J_{3,\nu}\|_2 \le \alpha_4 p \bar{n} \,. \tag{188}$$

We prove the above for the case $\nu = 2$. The other case follow from analogous arguments. Firstly, define the matrices $\widetilde{J}_{3,2} \in \mathbb{R}^{\binom{[n]}{2} \times \binom{[n]}{2}}$ and $Q \in \mathbb{R}^{n^2 \times n^2}$ as follows:

$$(\widetilde{J}_{3,2})_{\{i,j\},\{k,l\}} = \alpha_4 p g_{ik} g_{il} g_{jl}, \tag{189}$$

$$Q_{(i,j),(k,l)} = g_{ik} g_{il} g_{jl}. \tag{190}$$

Note also that $\widetilde{J}_{3,2}$ differs from $J_{3,2}$ only in the entries $\{i,j\}, \{k,\ell\}$ where $j = k$. The rows (columns) of Q above are indexed by *ordered* pairs $(i,j) \in [n] \times [n]$. Now we define the projector $\mathcal{P}_{\binom{[n]}{2}} : \mathbb{R}^{n^2} \to \mathbb{R}^{\binom{[n]}{2}}$ by letting, for all $i, j \in [n]$,

$$(\mathcal{P}_{\binom{[n]}{2}}(x))_{\{i,j\}} = x_{(i,j)} \,. \tag{191}$$

Then we have $\widetilde{J}_{3,2} = \alpha_4 p \mathcal{P}_{\binom{[n]}{2}} Q \mathcal{P}^{\mathsf{T}}_{\binom{[n]}{2}}$ and, consequently, $\|\widetilde{J}_{3,2}\|_2 \le \alpha_4 p \, \|Q\|_2$. Therefore it suffices to bound the latter, which we do again by the moment method. Firstly we define:

$$U_{(i,j),(k,l)} = \sum_{q \in [n]} g_{iq} g_{qk} g_{ij} \mathbb{I}(j = l) \,, \tag{192}$$

$$D_{(i,j),(k,l)} = \sum_{q \in [n]} g_{jq} g_{ql} g_{ij} \mathbb{I}(i = k) \,. \tag{193}$$

Then we have, for any integer $m \ge 1$,

$$\begin{aligned}
\mathsf{Tr}((Q^{\mathsf{T}}Q)^m) &= \sum_{i_1,i_2,\dots,i_{2m} \in [n]} \sum_{j_1,j_2,\dots,j_m \in [n]} Q^{\mathsf{T}}_{(i_1,j_1),(i_2,j_2)} Q_{(i_2,j_2),(i_3,j_3)} Q^{\mathsf{T}}_{(i_3,j_3),(i_4,j_4)} \cdots Q_{(i_{2m},j_{2m}),(i_1,j_1)} \\
&= \sum_{i_1,i_2,\dots,i_{2m} \in [n]} \sum_{j_1,j_2,\dots,j_m \in [n]} \big(g_{i_1 i_2} g_{j_1 j_2} g_{i_1 j_2}\big) \cdot \big(g_{i_2 i_3} g_{j_2 j_3} g_{j_2 i_3}\big) \cdot \big(g_{i_3 i_4} g_{j_3 j_4} g_{i_3 j_4}\big) \cdots \big(g_{i_{2m} i_1} g_{j_{2m} j_1} g_{j_{2m} i_1}\big) \\
&= \sum_{i_1,i_2,\dots,i_{2m} \in [n]} \sum_{j_1,j_2,\dots,j_{2m} \in [n]} \big(g_{i_1 i_2} g_{i_2 i_3} g_{i_3 i_4} \cdots g_{i_{2m} i_1}\big) \big(g_{j_1 j_2} g_{j_2 j_3} g_{j_3 j_4} \cdots g_{j_{2m} j_1}\big) \big(g_{i_1 j_2} g_{j_2 i_3} g_{i_3 j_4} \cdots g_{j_{2m} i_1}\big) \\
&= \sum_{i_1,i_2,\dots,i_{2m} \in [n]} \sum_{j_1,j_2,\dots,j_{2m} \in [n]} \big(g_{i_1 i_2} g_{i_2 i_3} g_{i_1 j_2}\big) \big(g_{j_2 j_3} g_{j_3 j_4} g_{j_2 i_3}\big) \big(g_{i_3 i_4} g_{i_4 i_5} g_{i_3 j_4}\big) \cdots \big(g_{j_{2m} j_1} g_{j_1 j_2} g_{j_{2m} i_1}\big) \,.
\end{aligned}$$

Then we have

$$\mathsf{Tr}((Q^{\mathsf{T}}Q)^m) = \mathsf{Tr}((UD)^m) \,. \tag{194}$$

Hence

$$\|Q\|_2 \le \mathsf{Tr}((Q^{\mathsf{T}}Q)^m)^{1/2m} \le \mathsf{Tr}((UD)^m)^{1/2m} \le \big(n^2 \|U\|_2^m \|D\|_2^m\big)^{1/2m} \le n^{1/m} \|U\|_2 \,, \tag{195}$$

where in the last step we used the fact that $\|U\|_2 = \|D\|_2$ by symmetry. Since m can be taken arbitrarily large, we conclude that $\|Q\|_2 \le \|U\|_2$ and we proceed to bound the latter.

Now let $T \in \mathbb{R}^{n^2 \times n^2}$ be the element-wise multiplication by g, i.e.

$$T_{(i,j),(k,l)} = g_{ij}\mathbb{I}(i=k)\mathbb{I}(j=l)\,. \tag{196}$$

Then we have

$$U = T \cdot \left(g^2 \otimes \mathrm{I}_n\right) \tag{197}$$

Here $g \in \mathbb{R}^{n\times n}$ is the matrix with i,j entry being g_{ij}. Since $|g_{ij}| \le 1$, we have $\|T\|_2 \le 1$ and therefore

$$\|Q\|_2 \le \|U\|_2 \le \|T\|_2 \|g^2 \otimes \mathrm{I}\|_2 \le \|g^2 \otimes \mathrm{I}\|_2 \le \|g^2\|_2 \le \|g\|_2^2\,. \tag{198}$$

Finally, similar to Proposition 24 we have that $\|g\| \lesssim \bar{n}^{1/2}$ with probability at least $1 - n^{-5}$, hence with the same probability:

$$\left\|\widetilde{J}_{3,2}\right\|_2 \lesssim \alpha_4 p \bar{n}\,. \tag{199}$$

By triangle inequality $\|J_{3,2}\|_2 \le \|\widetilde{J}_{3,2}\|_2 + \|\widetilde{J}_{3,2} - J_{3,2}\|_2$, hence to complete the proof we now bound $\|\widetilde{J}_{3,2} - J_{3,2}\|_2$ using the moment method. Recall that $\widetilde{J}_{3,2}$ and $J_{3,2}$ differ in the entry $\{i,j\},\{k,\ell\}$ only if $j=k$. Hence:

$$\begin{aligned}\mathsf{Tr}\left\{\left((\widetilde{J}_{3,2}-J_{3,2})^{\mathsf{T}}(\widetilde{J}_{3,2}-J_{3,2})\right)^m\right\} &= (\alpha_4 p)^{2m} \sum_{i_1\ldots i_{2m}, j_1\ldots j_{2m}, \forall q\, i_q<j_q} \prod_{q=1}^{m}\Big(g_{i_q i_{q+1}} g_{i_q j_{q+1}} g_{j_q j_{q+1}} \\ &\qquad g_{i_{q+1}i_{q+2}} g_{j_{q+1}i_{q+2}} g_{j_{q+1}j_{q+2}} \mathbb{I}(j_1 = i_2 = j_3 = i_4 = \cdots = i_{2m})\Big) \qquad (200)\\ &= (\alpha_4 p)^{2m} \sum_{\ell \in \widetilde{\mathfrak{L}}(R(3,2,m))} \prod_{e=\{u,v\}\in R(3,2,m)} g_{\ell(u)\ell(v)}\,. \qquad (201)\end{aligned}$$

Here, $R(3,2,m)$ is a $(3,2)$-ribbon of length $2m$ and $\widetilde{\mathfrak{L}}(R(3,2,m)$ is a collection of labelings of $R(3,2,m)$ satisfying the following criteria

1. For every couple $(u,v) \in R(3,2,m)$, $\ell(u) < \ell(v)$.
2. Let $(u_1,v_1),(u_2,v_2)\ldots(u_{2m},v_{2m})$ denote the couples in $R(3,2,m)$. Then $\ell(v_1) = \ell(u_2) = \ell(v_3) = \ell(u_4)\ldots$.

Let $\widetilde{\mathfrak{L}}_2(R(3,2,m))$ denote the subset of contributing labelings, i.e. those that satisfy the additional criterion that every labeled edge is repeated twice. By Lemma 20 it suffices to show that $v_*(R(3,2,m)) = \max_{\ell\in\widetilde{\mathfrak{L}}(R(3,2,m))} |\mathsf{range}(\ell)| \le m+2$. We prove this by induction. For the base case of $m=1$, since every edge is repeated twice under a contributing labeling, it is easy to see that there are at most 3 unique labels. Assume the induction hypothesis that $v_* R(3,2,m-1) \le m+1$. Let ℓ be a contributing labeling of $R(3,2,m)$. Then one of the following must happen:

1. No vertex in $R(3,2,m)$ has a unique label under ℓ.

2. There exists a vertex w of degree 4 with a unique label under ℓ.

The second condition follows because the vertices of degree smaller than 4 already have non-unique labels due to condition 2 of the labeling set $\widetilde{\mathfrak{L}}(R(3,2,m))$.

In case 1, $R(3,2,m)$ can have at most $2m/2+1 = m+1 < m+2$ unique labels under ℓ. In case 2, since w has a unique label and degree 4 the neighboring $(u,v),(u',v')$ have the same labels under ℓ i.e. $\ell(u)=\ell(u')$ and $\ell(v)=\ell(v')$. Hence we can identify the couples $(u,v),(u',v')$, delete w and its incident edges to obtain a ribbon $\tilde{R}(3,2,m-1)$ of length $2m-2$ and an induced labeling $\tilde{\ell}$ thereof. By the induction hypothesis $\mathsf{range}(\tilde{\ell}) \le m+1$ hence $\mathsf{range}(\tilde{\ell}) = \mathsf{range}(\tilde{\ell})+1 \le m+2$, as required. By Lemma 20 we obtain that $\left\|\widetilde{J}_{3,2} - J_{3,2}\right\|_2 \lesssim \alpha_4 p\bar{n}$ with probability at least $1-n^{-5}$.

By Eq. (199), it follows that with probability at least $1-2n^{-5}$, $\|J_{3,2}\|_2 \lesssim \alpha_4 p\bar{n} \lesssim \alpha_4\bar{n}$. This completes the proof of the lemma. ∎

For the case $\eta = 1$ we prove the following

Lemma 28 *Recall that $\mathcal{P}_2 : \mathbb{R}^{\binom{[n]}{2}} \to \mathbb{R}^{\binom{[n]}{2}}$ is the orthogonal projector onto the space $\mathbb{V}_2 \subseteq \mathbb{R}^{\binom{[n]}{2}}$ (defined in Section A.5). Firstly, we have that $\mathcal{P}_2(\sum_{\nu=1}^4 \widetilde{J}_{1,\nu})\mathcal{P}_2 = 0$ Further, with probability at least $1-4n^{-5}$, we have that:*

$$\left\|\sum_{\nu=1}^4 \widetilde{J}_{1,\nu}\right\|_2 \lesssim \alpha_4 \bar{n}^{3/2} \tag{202}$$

Proof Recall from the definition of $\widetilde{J}_{1,\nu}$ that

$$\sum_{\nu=1}^4 (\widetilde{J}_{1,\nu})_{\{i,j\},\{k,\ell\}} = p^3(g_{ik}+g_{i\ell}+g_{jk}+g_{j\ell}). \tag{203}$$

Now, for any $v \in \mathbb{R}^{\binom{[n]}{2}}$:

$$\begin{aligned}\left(\sum_{\nu=1}^4 \widetilde{J}_{1,\nu} v\right)_{\{i,j\}} &= \sum_{k<\ell} p^3(g_{ik}+g_{i\ell}+g_{jk}+g_{j\ell})v_{\{k,\ell\}} \\ &= u_i + u_j,\end{aligned}$$

where we define $u_i \equiv \sum_{k\le\ell} p^3(g_{ik}+g_{i\ell})v_{\{k,\ell\}}$. It follows that $\sum_{\nu=1}^4 \widetilde{J}_{\eta,\nu} v \in \mathbb{V}_2^{\perp} = \mathbb{V}_0 \oplus \mathbb{V}_1$, and hence $\mathcal{P}_2 \sum_{\nu=1}^4 \widetilde{J}_{\eta,\nu} = 0$. Since $\sum_{\nu=1}^4 \widetilde{J}_{1,\nu}$ is symmetric we obtain the first claim.

We prove the second claim –cf. Eq. (202)– by the moment method, similar to Lemma 26. Let $R(1,\nu,m)$ be a $(1,\nu)$-ribbon of length $2m$. Then:

$$\mathsf{Tr}\left\{(\widetilde{J}_{1,\nu}^{\mathsf{T}}\widetilde{J}_{1,\nu})^r\right\} = (\alpha_4 p^3) \sum_{\ell\in\mathfrak{L}(R(1,\nu,m))} \left\{\prod_{e=\{u,v\}\in R(1,\nu,m)} g_{\ell(u)\ell(v)}\right\}. \tag{204}$$

By Lemma 20 it suffices to prove that $v_*(R(1,\nu,m)) = 3m+2$. The claim then follows, using Lemma 20 and the union bound.

Let $\ell \in \mathfrak{L}_2(R(1,\nu,m))$ be a contributing labeling of a ribbon $R(1,\nu,m)$ of length $2m$. Let $\mathbf{G}(1,\nu,m)$ be the graph obtained by identifying vertices in $R(1,\nu,m)$ with the same label. Notice that $R(1,\nu,m)$ is a union of a cycle $D(m)$ of length $2m$ and $2m+1$ isolated vertices. The isolated vertices can have arbitrary labels, hence $v_*(R(1,\nu,m)) = 2m+1+v_*(D(2m)) = 3m+2$ as proved in Proposition 24. ■

In a similar fashion, we bound the norm of the terms $\widetilde{J}_{2,2}$, $\widetilde{J}_{2,3}$, $\widetilde{J}_{2,4}$, $\widetilde{J}_{2,5}$:

Lemma 29 *We have that:*

$$(\widetilde{J}_{2,2}+\widetilde{J}_{2,4})\mathcal{P}_2 = 0, \tag{205}$$

$$\mathcal{P}_2(\widetilde{J}_{2,3}+\widetilde{J}_{2,5}) = 0. \tag{206}$$

Further with probability at least $1-2n^{-4}$

$$\left\|\widetilde{J}_{2,2}\right\|_2 \lesssim (\alpha_4 p^2)\bar{n}^{3/2}, \tag{207}$$

$$\left\|\widetilde{J}_{2,4}\right\|_2 \lesssim (\alpha_4 p^2)\bar{n}^{3/2}. \tag{208}$$

Proof It is easy to check that $\widetilde{J}_{2,2} = \widetilde{J}_{2,3}^{\mathsf{T}}$ and $\widetilde{J}_{2,4} = \widetilde{J}_{2,5}^{\mathsf{T}}$. We prove Eq. (206), from which Eq. (205) follows by taking transposes of each side. From the definition of $\widetilde{J}_{2,\nu}$ we have for any $v \in \mathbb{R}^{\binom{[n]}{2}}$

$$(\widetilde{J}_{2,3}v+\widetilde{J}_{2,5}v)_{\{i,j\}} = \sum_{k<\ell} p^2(g_{ik}g_{i\ell}+g_{jk}g_{j\ell})v_{\{k,\ell\}} \tag{209}$$

$$= u_i + u_j, \tag{210}$$

where we let $u_i \equiv= \sum_{k<\ell} p^2(g_{ik}g_{i\ell})v_{\{k,\ell\}}$. It follows that $(\widetilde{J}_{2,3}v+\widetilde{J}_{2,5})v \in \mathbb{V}_0 \oplus \mathbb{V}_1$ hence $\mathcal{P}_2(\widetilde{J}_{2,3}+\widetilde{J}_{2,5}) = 0$.

We prove the claim on the spectral norm for $\widetilde{J}_{2,2}$. The claim for $\widetilde{J}_{2,4}$ holds in an analogous fashion. Let $R(2,2,m)$ be a $(2,2)$-ribbon of length m. Then:

$$\mathsf{Tr}\left\{(\widetilde{J}_{2,2}^{\mathsf{T}}\widetilde{J}_{2,2})^m\right\} = \sum_{\ell\in\mathfrak{L}(R(2,2,m))} (\alpha_4 p^2)^{2m} \prod_{e=\{u,v\}\in R(2,2,m))} g_{\ell(u)\ell(v)}. \tag{211}$$

By Lemma 20, it suffices to show that $v_*(R(2,2,m)) = 3m+2$. i.e a contributing labeling ℓ maps to at most $3m+2$ unique labels. Notice that $R(2,2,m)$ is the union of $m+1$ isolated vertices and a bridge $B(m)$ of length $2m$. The isolated vertices are unconstrained and hence contribute at most $m+1$ new labels. It suffices, hence, to prove that $B(m)$ has at most $2m+1$ unique labels under its labeling $\ell_{B(m)}$ induced by ℓ. Since, $\ell_{B(m)}$ is contributing for $B(m)$, it suffices that $v_*(B(m)) = 2m+1$. We prove this by induction on m. In the base case of $m=1$, this implies it has at most $3 = (2\cdot 1+1)$ unique labels. Assuming that the claim is true for bridges of length at most $2m$ for $m>1$, we show that it holds for a bridge $B(m+1)$ of length $2m+2$. $B(m+1)$ contains $3m+4$ vertices hence there are 3 cases:

1. For every vertex $u \in B$ there exists a different vertex $u' \in B$ such that $\ell_B(u) = \ell_B(u')$.

2. There exists a vertex $u \in B$ which has a unique label under ℓ_B and u has degree 4.

3. There exists a vertex $u \in B$ which has a unique label under ℓ_B with degree 2.

In the first case, $|\mathsf{range}(\ell)| \le (3m+4))/2 \le 2(m+1)+1$ hence the claim holds.

In the second case, we have that the neighboring couples are (u_1, v_1), (u_2, v_2) then $\ell_{B(m+1)}(u_1) = \ell_{B(m+1)}(u_2)$ and $\ell_{B(m+1)}(v_1) = \ell_{B(m+1)}(v_2)$. We can then contract the neighbors of u and delete u and incident edges to obtain a bridge $\tilde{B}(m)$ (and induced labeling $\ell_{\tilde{B}(m)}$ of length $2m$). By induction $\ell_{\tilde{B}(m)}$ maps to at most $2m+1$ labels, hence $\ell_{B(m)}$ to at most $2m+1+1 \le 2(m+1)+1$ labels.

In the third case, if u has neighbors u_1, u_2 then $\ell_{B(m+1)}(u_1) = \ell_{B(m+1)}(u_2)$. If we now identify the neighbors of u with the same label, and delete u and the edges incident on it, we obtain a bridge $\tilde{B}(m)$ of length $2m$, and an induced labeling $\ell_{\tilde{B}(m)}$ which is contributing. By induction, $\tilde{B}(m)$ has at most $2m+1$ unique labels, hence $B(m+1)$ has at most $2m+1+2 = 2(m+1)+1$ unique labels. This completes the induction. ■

Finally, we have to deal with the remainder terms (recall that matrix K is defined in Eq. (177)).

Lemma 30 *We have with probability at least* $1 - n^{-5}$ *that:*

$$\|K\|_2 \lesssim \alpha_3 \bar{n}^{1/2} \tag{212}$$

Proof We compute $\mathsf{Tr}\left\{(K^{\mathsf{T}}K)^m\right\}$. Note that:

$$\mathsf{Tr}\left\{(K^{\mathsf{T}}K)^m\right\} = \sum_{A_1,B_1\dots A_m B_m} \prod_{l=1}^{r} (K_{A_l B_l} K_{A_{l+1} B_l}) \tag{213}$$

$$= \sum_{A_1,B_1\dots A_m B_m} \prod_{l=1}^{r} K_{A_l B_l} K_{A_{l+1} B_l} \mathbb{I}(|A_l \cap B_l| = 1)\mathbb{I}(|A_{l+1} \cap B_l| = 1). \tag{214}$$

Here we set $A_{m+1} \equiv A_1$. The second equality follows since K is supported on entries A, B such that A, B share exactly one vertex. Recalling the definition of star ribbons, each term that does not vanish in the summation above corresponds a labeling of a star ribbon $S(2,1,m) \in \mathcal{S}^m_{2,1}$ formed from a $(2,1)$-ribbon of length $2m$, i.e. we have:

$$\mathsf{Tr}\left\{(K^{\mathsf{T}}K)^m\right\} = \alpha_3^{2m} \sum_{S(2,1,m)\in\mathcal{S}^m_{2,1}} \sum_{\ell\in\mathfrak{L}(S(2,1,m))} \prod_{e=\{u,v\}\in S(2,1,m)} g_{\ell(u),\ell(v)}. \tag{215}$$

Since there are at most $2^{2m} = 4^m$ star ribbons of length $2m$, it suffices by a simple extension of Lemma 20, to show that $v_*(S(2,1,m)) = m+2$. Note that every $S(2,1,m)$ is a union of 2 paths, one of length m' and the other of length $2m-m'$ for some $m' \in [2m]$, hence has at most 2 connected components. Let ℓ be a contributing labeling of $S(2,1,m)$ and $\mathbf{G}_{S(2,1,m)}$ be the graph obtained by identifying vertices in $S(2,1,m)$ with the same label. Since $S(2,1,m)$ is a union of two paths, $\mathbf{G}_{S(2,1,m)}$ has at most 2 connected components. Furthermore, since ℓ is a contributing labeling, every labeled edge in $S(2,1,m)$ repeats at least twice, hence $\mathbf{G}_{2,1}(m)$ has at most $2m/2 = m$ edges. Consequently, it has at most $m+2$ vertices, implying that $v_*(S(2,1,m)) \le m+2$. ■

Finally, we deal with the differences $J_{\eta,\nu} - \widetilde{J}_{\eta,\nu}$. (Recall that $J_{\eta,\nu}$ and $\widetilde{J}_{\eta,\nu}$ are defined in Eqs. (176) and (178).)

Lemma 31 *With probability at least* $1 - 6n^{-5}$, *for each* $\eta \leq 2$ *and* $\nu \leq \binom{4}{\eta}$*:*

$$\left\| J_{\eta,\nu} - \widetilde{J}_{\eta,\nu} \right\|_2 \lesssim \alpha_4 \bar{n} \tag{216}$$

Proof We first consider $\mathsf{Tr}\left\{((\widetilde{J}_{\eta,\nu} - J_{\eta,\nu})^{\mathsf{T}}(\widetilde{J}_{\eta,\nu} - J_{\eta,\nu}))^m\right\}$. Let $R(\eta,\nu,m)$ be a (η,ν)-ribbon of length $2m$. As in the previous lemmas, we can write $\mathsf{Tr}\left\{((\widetilde{J}_{\eta,\nu} - J_{\eta,\nu})^{\mathsf{T}}(\widetilde{J}_{\eta,\nu} - J_{\eta,\nu}))^m\right\}$ as a sum over labelings of $R(\eta,\nu,m)$ as follows:

$$\mathsf{Tr}\left\{((\widetilde{J}_{\eta,\nu} - J_{\eta,\nu})^{\mathsf{T}}(\widetilde{J}_{\eta,\nu} - J_{\eta,\nu}))^m\right\} = (\alpha_4 p^{4-\eta})^{2m} \sum_{\ell \in \widetilde{\mathfrak{L}}(R(\eta,\nu,m))} \prod_{e=\{u,v\} \in R(\eta,\nu,m)} g_{\ell(u),\ell(v)}. \tag{217}$$

Here we restrict the labelings ℓ to the subset $\widetilde{\mathfrak{L}}(R(\eta,\nu,m)$ that satisfy the criteria:

1. For every couple (u,v), $\ell(u) < \ell(v)$.
2. Consider any adjacent pair of couples $(u_1,v_1),(u_2,v_2)$ in $R(\eta,\nu,m)$, at least one of u_1,v_1,u_2,v_2 has degree 0. Assume this is u_1 (without loss of generality), then either $\ell(u_1) = \ell(u_2)$ or $\ell(u_1) = \ell(v_2)$.

On taking expectations the only labelings that do not vanish satisfy the additional criterion that every labeled edge is repeated at least twice in $R(\eta,\nu,m)$. We call this set of labelings $\widetilde{\mathfrak{L}}_2(R(\eta,\nu,m))$. As in Lemma 29 it suffices to show that $\left|\widetilde{\mathfrak{L}}_2(R(\eta,\nu,m))\right| \leq \binom{n}{2m+2}(2^{2m}(2m+2)^{3m+2})$. This follows from the same arguments as in Lemmas 29, 28 (for $\eta = 1,2$ respectively), with the additional caveat that the isolated vertices in $R(\eta,\nu,m)$ are not unconstrained as before. Indeed, once the labels of the connected component of $R(\eta,\nu,m)$ are decided, there are only 2^m possible ways of choosing the labels for the isolated vertices. Consequently, we have the bound:

$$\mathbb{E}\mathsf{Tr}\left\{((\widetilde{J}_{\eta,\nu} - J_{\eta,\nu})^{\mathsf{T}}(\widetilde{J}_{\eta,\nu} - J_{\eta,\nu}))^m\right\} \leq (\alpha_4 p^{4-\eta})^{2m} \left|\widetilde{\mathfrak{L}}_2(R(\eta,\nu,m))\right| \tag{218}$$

$$\leq \binom{n}{2m+2}(2\alpha_4 p^{4-\eta})^{2m}(2m+2)^{3m+2}. \tag{219}$$

Applying Lemma 18, union bound and the triangle inequality yields the final result. ■

We can now prove Proposition 25.

Proof [Proof of Proposition 25] The intersection of high probability events of Lemmas 26, 27, 28, 29, 30 and 31 holds with probability at least $1 - 25n^{-5}$. We will condition on this event for the proof of the proposition.

We bound each of the projections $\mathcal{P}_a(H_{22} - \mathbb{E}\{H_{22}\})\mathcal{P}_b$ for $a,b \in \{0,1,2\}$ using the decomposition (180).

- Let us first consider $a = b$, $a, b \in \{0, 1\}$, cf. Eq. (170). By application of above lemmas, triangle inequality, the fact that $\|\mathcal{P}_a X \mathcal{P}_b\|_2 \le \|\mathcal{P}_a\|_2 \|X\|_2 \|\mathcal{P}_b\|_2 \le \|X\|_2$ for any $X \in \mathbb{R}^{\binom{[n]}{2}}$ in the decomposition Eq. (180), we get

$$\|\mathcal{P}_a(H_{22} - \mathbb{E}\{H_{22}\})\mathcal{P}_a\|_2 \lesssim \alpha_3 \bar{n}^{1/2} + \alpha_4 \bar{+} \alpha_4 \bar{n}^{3/2}) \tag{220}$$

$$\lesssim \alpha_3 \bar{n}^{1/2} + \alpha_4 \bar{n}^{3/2}, \tag{221}$$

  This proves Eq. (170).

- The case $a = b = 2$ is treated in the same manner, with the only difference that, when bounding $\|\mathcal{P}_2(H_{22} - \mathbb{E}\{H_{22}\})\mathcal{P}_2\|$, the terms of the type $\alpha_4 \bar{n}^{3/2}$ do not appear (see Lemmas 28, 29). Hence:

$$\|\mathcal{P}_2(H_{22} - \mathbb{E}\{H_{22}\})\mathcal{P}_2\|_2 \lesssim \alpha_3 \bar{n}^{1/2} + \alpha_4 \bar{n} \tag{222}$$

$$\le \alpha_3 \bar{n}^{1/2} + \alpha_4 \bar{n}. \tag{223}$$

  This proves Eq. (171).

- The bound for the cross terms $\|\mathcal{P}_a(H_{22} - \mathbb{E}\{H_{22}\})\mathcal{P}_b\|_2$ for $a \neq b$ is identical to that for the case $a = b = 0$ above.

  This proves Eq. (172) and hence finishes our proof of Proposition 25.

■

A.8. Controlling $H_{12} - \mathbb{E}\{H_{12}\}$

We prove the following proposition for the deviation $H_{12} - \mathbb{E}\{H_{12}\}$

Proposition 32 *With probability at least* $1 - 5n^{-5}$ *the following are true.*

$$\|H_{12} - \mathbb{E}\{H_{12}\}\|_2 \lesssim \alpha_3 \bar{n}. \tag{224}$$

Recall that an entry of $H_{12} \in \mathbb{R}^{\binom{[n]}{1} \times \binom{[n]}{2}}$ can be written as:

$$(H_{12})_{A,B} = \begin{cases} \alpha_2 - \alpha_1 \alpha_2 & \text{if } |A \cap B| = 1 \\ \alpha_3 (p + g_{A,h(B)})(p + g_{A,t(B)}) - \alpha_1 \alpha_2 & \text{otherwise.} \end{cases} \tag{225}$$

Define the matrices $L_{\eta,\nu} \in \mathbb{R}^{\binom{[n]}{1} \times \binom{[n]}{2}}$ for $\eta = 1, 2$, $\nu \le \binom{\eta}{\nu}$ and $\widetilde{L}_{1,\nu}$ for $\nu = 1, 2$:

$$(L_{2,1})_{A,B} \equiv \begin{cases} \alpha_3 g_{A,h(B)} g_{A,t(B)} & \text{if } |A \cap B| = 0 \\ 0 & \text{otherwise.} \end{cases} \tag{226}$$

$$(L_{1,1})_{A,B} \equiv \begin{cases} \alpha_3 p g_{A,h(B)} & \text{if } |A \cap B| = 0 \\ 0 & \text{otherwise.} \end{cases} \tag{227}$$

$$(L_{1,2})_{A,B} \equiv \begin{cases} \alpha_3 p g_{A,t(B)} & \text{if } |A \cap B| = 0 \\ 0 & \text{otherwise.} \end{cases} \tag{228}$$

It thus follows that:

$$H_{12} - \mathbb{E}\{H_{12}\} = L_{1,1} + L_{1,2} + L_{2,2}. \tag{229}$$

We first prove two Lemmas on the spectral properties of the matrices $L_{\eta,\nu}$

Lemma 33 *With probability at least* $1 - n^{-5}$*, we have that*

$$\|L_{2,1}\|_2 \lesssim \alpha_3 \bar{n}. \tag{230}$$

Proof Note that:

$$\mathsf{Tr}\left\{(L_{2,1}L_{2,1}^{\mathsf{T}})\right\} = \sum_{A_1 \dots A_{m+1}, B_1 \dots B_m} \prod_{l=1}^{r} g_{A_l h(B_l)} g_{A_l t(B_l)} g_{A_{l+1} h(B_l)} g_{A_{l+1} t(B_l)} \tag{231}$$

Equivalently, letting $B(m)$ be a bridge of length $2m$ we have:

$$\mathsf{Tr}\left\{(L_{2,1}L_{2,1}^{\mathsf{T}})^m\right\} = \sum_{\ell \in \mathfrak{L}(B)} \prod_{e=\{u,v\} \in B} g_{\ell(u)\ell(v)}. \tag{232}$$

By Lemma 20 it suffices to show that $v_*(B(m)) \leq 2m + 1$. This argument is already covered in Lemma 29 and the claim hence follows. ■

Lemma 34 *With probability exceeding* $1 - 2n^{-5}$ *the following holds:*

$$\max_{\nu=1,2} \|L_{1,\nu}\|_2 \lesssim \alpha_3 \bar{n}. \tag{233}$$

Proof We prove the claim for $L_{1,1}$. The same argument applies for $L_{1,2}$ with minor modifications.

$$\mathsf{Tr}\left\{(L_{1,2}L_{1,2}^{\mathsf{T}})^m\right\} = \sum_{A_1 \dots A_{m+1}, B_1 \dots B_m} (\alpha_3 p)^{2m} \prod_{l=1}^{m} g_{A_l h(B_\ell)} g_{A_{l+1} h(B_l)}\,. \tag{234}$$

The above a sum over labelings of a bridge $B(m)$ of type 1 and class 1, of length $2m$. This is union of a cycle $D(m)$ of length $2m$, and m isolated vertices. The lemma follows from Lemma 20 if $v_*(B(m)) \leq 2m + 1$. But by the above decomposition $v_*(B(m)) \leq v_*(D(m)) + m = m + 1 + m = 2m + 1$, as in Proposition 24. This completes the proof. ■

We can now prove Proposition 32.

Proof [Proof of Proposition 32] The intersection of favorable events of lemmas 33, 34 probability at least $1 - 5n^{(\Gamma-4)/2}$. The required claim then follows from Lemmas 33, 34 and triangle inequality. ■

A.9. Proof of Proposition 6

The intersection of high probability favorable events of Propositions 24, 25 and 32 holds with probability at least $1 - 30n^{-5} \geq 1 - n^{-4}$ for large enough n. By Proposition 24 we already have the required bounds on H_{11} and H_{11}^{-1}, cf. Eqs. (55) and (56). It remains to show that on the same event:

$$H_{22} \succcurlyeq \frac{2}{\alpha_1} H_{12}^{\mathsf{T}} \mathcal{Q}_n H_{12} + \frac{1}{n(\alpha_2 p - \alpha_1^2)} H_{12}^{\mathsf{T}} \mathcal{Q}_n^{\perp} H_{12}\,, \tag{235}$$

or, equivalently,

$$\text{Or} \quad \mathbb{E}\{H_{22}\} \succcurlyeq \mathbb{E}\{H_{22}\} - H_{22} + \frac{2}{\alpha_1} H_{12}^{\mathsf{T}} \mathcal{Q}_n H_{12} + \frac{1}{n(\alpha_2 p - \alpha_1^2)} H_{12}^{\mathsf{T}} \mathcal{Q}_n^{\perp} H_{12}. \tag{236}$$

Let $\overline{W}, W \in \mathbb{R}^{3\times 3}$ be two matrices that satisfy, for $a, b \in \{0,1,2\}$:

$$\overline{W}_{ab} = \|\mathcal{P}_a \mathbb{E}\{H_{22}\} \mathcal{P}_b\|_2 \tag{237}$$

$$\begin{aligned} W_{ab} \geq \|\mathcal{P}_a (H_{22} - \mathbb{E}\{H_{22}\}) \mathcal{P}_b\|_2 + \frac{2}{\alpha_1} \left\|\mathcal{Q}_n^{\perp} H_{12} \mathcal{P}_a\right\|_2 \left\|\mathcal{Q}_n^{\perp} H_{12} \mathcal{P}_b\right\|_2 \\ + \frac{1}{n(\alpha_2 p - \alpha_1^2)} \|\mathcal{Q}_n H_{12} \mathcal{P}_a\|_2 \|\mathcal{Q}_n H_{12} \mathcal{P}_b\|_2\,. \end{aligned} \tag{238}$$

By expanding the Rayleigh quotient of each term in Eq. (236), and noting that $\overline{W}_{ab} = 0$ for $a \neq b$, it is straightforward to see that Eq. (236) holds if

$$\alpha_2 p - \alpha_1^2 \geq 0\,, \tag{239}$$

$$\overline{W} \succcurlyeq W\,. \tag{240}$$

The first condition correspond to assumption (53). For the second one, we develop explicit expressions of $\overline{W}$, W as follows. For $\overline{W}$, we use Proposition 21, that yields immediately $\overline{W}_{a,b} = 0$ for $a \neq b$ as claimed, and $\overline{W}_{0,0}$, $\overline{W}_{1,1}$, $\overline{W}_{2,2}$ as in Eqs. (43), (44), (45).

In order to develop expressions for W we note that it is sufficient to guarantee

$$\begin{aligned} W_{ab} \geq \|\mathcal{P}_a (H_{22} - \mathbb{E}\{H_{22}\}) \mathcal{P}_b\|_2 \\ + \frac{2}{\alpha_1} \left(\left\|\mathcal{Q}_n^{\perp} \mathbb{E}\{H_{12}\} \mathcal{P}_a\right\|_2 + \|H_{12} - \mathbb{E}\{H_{12}\}\|_2\right) \left(\left\|\mathcal{Q}_n^{\perp} \mathbb{E}\{H_{12}\} \mathcal{P}_b\right\|_2 + \|H_{12} - \mathbb{E}\{H_{12}\}\|_2\right) \\ + \frac{1}{n(\alpha_2 p - \alpha_1^2)} \left(\|\mathcal{Q}_n \mathbb{E}\{H_{12}\} \mathcal{P}_a\|_2 + \|H_{12} - \mathbb{E}\{H_{12}\}\|_2\right) \left(\|\mathcal{Q}_n \mathbb{E}\{H_{12}\} \mathcal{P}_b\|_2 + \|H_{12} - \mathbb{E}\{H_{12}\}\|_2\right). \end{aligned} \tag{241}$$

Using the upper bounds in Propositions 23, 25, 32 we obtain the expressions in Eqs. (46) to (51). This completes the proof.