Foundations of Computational Mathematics
https://doi.org/10.1007/s10208-022-09575-7

Spectral Graph Matching and Regularized Quadratic Relaxations II

Erdős-Rényi Graphs and Universality

Zhou Fan[1] · Cheng Mao[2] · Yihong Wu[1] · Jiaming Xu[3]

Received: 16 August 2019 / Revised: 23 September 2021 / Accepted: 15 March 2022

Abstract

We analyze a new spectral graph matching algorithm, GRAph Matching by Pairwise eigen-Alignments (GRAMPA), for recovering the latent vertex correspondence between two unlabeled, edge-correlated weighted graphs. Extending the exact recovery guarantees established in a companion paper for Gaussian weights, in this work, we prove the universality of these guarantees for a general correlated Wigner model. In particular, for two Erdős-Rényi graphs with edge correlation coefficient $1-\sigma^2$ and average degree at least $\mathrm{polylog}(n)$, we show that GRAMPA exactly recovers the latent vertex correspondence with high probability when $\sigma \lesssim 1/\mathrm{polylog}(n)$. Moreover, we establish a similar guarantee for a variant of GRAMPA, corresponding to a tighter quadratic programming relaxation of the quadratic assignment problem. Our analysis exploits a resolvent representation of the GRAMPA similarity matrix and local laws for the resolvents of sparse Wigner matrices.

Communicated by Francis Bach. Z. Fan is supported in part by NSF Grant DMS-1916198. C. Mao is supported in part by NSF Grant DMS-2053333. Y. Wu is supported in part by the NSF Grants CCF-1527105, CCF-1900507, an NSF CAREER award CCF-1651588, and an Alfred Sloan fellowship. J. Xu is supported by the NSF Grants IIS-1838124, CCF-1850743, and CCF-1856424.

✉ Yihong Wu
yihong.wu@yale.edu

Zhou Fan
zhou.fan@yale.edu

Cheng Mao
cheng.mao@math.gatech.edu

Jiaming Xu
jx77@duke.edu

[1] Department of Statistics and Data Science, Yale University, New Haven, CT, USA

[2] School of Mathematics, Georgia Institute of Technology, Atlanta, GA, USA

[3] The Fuqua School of Business, Duke University, Durham, NC, USA

Published online: 13 June 2022

Keywords Graph matching · Quadratic assignment problem · Spectral methods · Random matrix theory · Erdős-Rényi graph

Mathematics Subject Classification Primary 90C25 · Secondary 68Q87

1 Introduction

Given two (weighted) graphs, graph matching aims at finding a bijection between the vertex sets that maximizes the total edge weight correlation between the two graphs. It reduces to the graph isomorphism problem when the two graphs can be matched perfectly. Let A and B be the (weighted) adjacency matrices of the two graphs on n vertices. Then, the graph matching problem can be formulated as solving the following *quadratic assignment problem* (QAP) [5, 18]:

$$\max_{\Pi \in \mathfrak{S}_n} \langle A, \Pi B \Pi^\top \rangle, \tag{1}$$

where $\mathfrak{S}_n$ denotes the set of permutation matrices in $\mathbb{R}^{n \times n}$ and $\langle \cdot, \cdot \rangle$ denotes the matrix inner product. The QAP is NP-hard to solve or to approximate within a growing factor [17].

In the companion paper [14], we proposed a computationally efficient spectral graph matching method, called GRAph Matching by Pairwise eigen-Alignments (GRAMPA). Let us write the spectral decompositions of A and B as

$$A = \sum_i \lambda_i v_i v_i^\top \quad \text{and} \quad B = \sum_j \mu_j w_j w_j^\top. \tag{2}$$

Given a tuning parameter $\eta > 0$, GRAMPA first constructs an $n \times n$ similarity matrix[1]

$$X = \sum_{i,j} \frac{\eta}{(\lambda_i - \mu_j)^2 + \eta^2} v_i v_i^\top \mathbf{J} w_j w_j^\top, \tag{3}$$

where $\mathbf{J}$ is the $n \times n$ all-ones matrix. Then, it outputs a permutation matrix $\widehat{\Pi}$ by "rounding" X to a permutation matrix, for example, by solving the following *linear assignment problem* (LAP)

$$\widehat{\Pi} \in \operatorname*{argmax}_{\Pi \in \mathfrak{S}_n} \langle X, \Pi \rangle. \tag{4}$$

Let $\Pi_* \in \mathfrak{S}_n$ be the latent true matching, and denote the entries of A and $\Pi_* B \Pi_*^\top$ as a_{ij} and $b_{\pi_*(i)\pi_*(j)}$. A Gaussian Wigner model is studied in [14], where $\{(a_{ij}, b_{\pi_*(i)\pi_*(j)})\}$ are i.i.d. pairs of correlated Gaussian variables such that

[1] In [14], X is defined without the factor η in the numerator. We include η here for convenience in the proof; this does not affect the algorithm as the rounded solution $\widehat{\Pi}$ is invariant to rescaling X.

$b_{\pi_*(i)\pi_*(j)} = a_{ij} + \sigma z_{ij}$ for a noise level $\sigma \geq 0$, and a_{ij} and z_{ij} are independent standard Gaussian. It is shown that GRAMPA exactly recovers the vertex correspondence Π_* with high probability when $\sigma = O(1/\log n)$. Simulation results in [14,Sect. 4.1] further show that the empirical performance of GRAMPA under the Gaussian Wigner model is very similar to that under the Erdős-Rényi model where $\{(a_{ij}, b_{\pi_*(i)\pi_*(j)})\}$ are i.i.d. pairs of correlated centered Bernoulli random variables, suggesting that the performance of GRAMPA enjoys universality.

In this paper, we prove a universal exact-recovery guarantee for GRAMPA, under a general Wigner matrix model for the weighted adjacency matrix: Let $A = (a_{ij})$ be a symmetric random matrix in $\mathbb{R}^{n\times n}$, where the entries $(a_{ij})_{i\leq j}$ are independent. Suppose that

$$\mathbb{E}\left[a_{ij}\right] = 0 \text{ for all } i, j, \qquad \mathbb{E}\left[a_{ij}^2\right] = \frac{1}{n} \text{ for all } i \neq j, \tag{5}$$

and

$$\mathbb{E}\left[\left|a_{ij}\right|^k\right] \leq \frac{C^k}{nd^{(k-2)/2}} \quad \text{for all } i, j \text{ and each } k \in \left[2, (\log n)^{10\log\log n}\right], \tag{6}$$

where $d \equiv d(n)$ is an n-dependent sparsity parameter and C is an absolute positive constant.

Of particular interest are the following special cases:

- Bounded case: The entries are bounded in magnitude by $\frac{C}{\sqrt{n}}$. Then, (6) is fulfilled for $d = n$ and all k.
- Sub-Gaussian case: The sub-Gaussian norm of each entry satisfies

$$\|a_{ij}\|_{\psi_2} \triangleq \sup_{k\geq 1} k^{-1/2}\mathbb{E}\left[\left|a_{ij}\right|^k\right]^{1/k} = O\left(1/\sqrt{n}\right). \tag{7}$$

  It is easily checked that (6) is satisfied for $d = n/(\log n)^{11\log\log n}$ and all large n.
- Erdős-Rényi graphs with edge probability $p \equiv p(n)$. We may center and scale the adjacency matrix A such that $a_{ij} \sim (\mathsf{Bern}(p) - p)/\sqrt{np(1-p)}$ for $i \neq j$, which satisfies (5) and (6) for $d = np(1-p)$ (cf. Lemma 1).

With the moment conditions (5) and (6) specified, we are ready to introduce the correlated Wigner model, which encompasses the correlated Erdős-Rényi graph model proposed in [19] as a special case.

Definition 1 (Correlated Wigner model) Let n be a positive integer, $\sigma \in [0, 1]$ an (n-dependent) noise parameter, π_* a latent permutation on $[n]$, and $\Pi_* \in \{0, 1\}^{n\times n}$ the corresponding permutation matrix such that $(\Pi_*)_{i\pi_*(i)} = 1$. Suppose that $\left\{(a_{ij}, b_{\pi_*(i)\pi_*(j)}) : i \leq j\right\}$ are independent pairs of random variables such that both $A = (a_{ij})$ and $B = (b_{ij})$ satisfy (5) and (6),

$$\mathbb{E}\left[a_{ij} b_{\pi_*(i)\pi_*(j)}\right] \geq \frac{1-\sigma^2}{n} \quad \text{for all } i \neq j, \tag{8}$$

and for a constant $C > 0$, any $D > 0$, and all $n \geq n_0(D)$,

$$\mathbb{P}\left\{\left\|A - \Pi_* B \Pi_*^\top\right\| \leq C\sigma\right\} \geq 1 - n^{-D} \tag{9}$$

where $\|\cdot\|$ denotes the spectral norm.

The parameter σ measures the effective noise level in the model. In the special case of sparse Erdős-Rényi model, A and B are the centered and normalized adjacency matrices of two Erdős-Rényi graphs, which differ by a fraction $2\sigma^2$ of edges approximately.

In this paper, we prove the following exact recovery guarantee for GRAMPA:

Theorem *(Informal statement) For the correlated Wigner model, if $d \geq \operatorname{polylog}(n)$ and $\sigma \leq c\,(\log n)^{-2\kappa}$ for any fixed constant $\kappa > 2$ and a sufficiently small constant $c > 0$, then GRAMPA with $\eta = 1/\operatorname{polylog} n$ recovers π_* exactly with high probability for large n. If furthermore a_{ij} and b_{ij} are sub-Gaussian and satisfy (7), then this holds with $\kappa = 1$.*

This theorem generalizes the exact recovery guarantee for GRAMPA proved in [14] for the Gaussian Wigner model, albeit at the expense of a slightly stronger requirement for σ than in the Gaussian case. The requirement that $d \geq \operatorname{polylog}(n)$ and $\sigma \leq 1/\operatorname{polylog}(n)$ is the state-of-the-art for polynomial time algorithms on sparse Erdős-Rényi graphs [10], although we note that the recent work of [2] provided an algorithm with super-polynomial runtime $n^{O(\log n)}$ that achieves exact recovery when $d \geq n^{o(1)}$ under the much weaker condition of $\sigma \leq 1 - (\log n)^{-o(1)}$ (see the end of Sect. 2 for more detailed discussion). Numerical experiments in the companion paper [14] suggest that the failure of GRAMPA occurs at $\sigma = C/\log n$ for some constant C, indicating that our theoretical characterization of the performance of GRAMPA here is almost tight. In [14], we further demonstrate the superior empirical performance of GRAMPA on a variety of synthetic and real datasets, in terms of both statistical accuracy and computational efficiency. In the conference version [15], GRAMPA is also shown to improve existing shape matching algorithms on 3D deformable shape data.

The analysis in [14] relies heavily on the rotational invariance of Gaussian Wigner matrices, and does not extend to non-Gaussian models. Here, instead, our universality analysis uses a *resolvent representation* of the GRAMPA similarity matrix (3) via a contour integral (cf. Proposition 1). Capitalizing on local laws for the resolvent of sparse Wigner matrices [11, 12], we show that the similarity matrix (3) is with high probability *diagonal dominant* in the sense that $\min_k X_{k\pi_*(k)} > \max_{\ell \neq \pi_*(k)} X_{k\ell}$. This enables rounding procedures as simple as thresholding to succeed.

From an optimization point of view, GRAMPA can also be interpreted as solving a *regularized quadratic programming (QP) relaxation* of the QAP. More precisely, the QAP (1) can be equivalently written as

$$\min_{\Pi \in \mathfrak{S}_n} \|A\Pi - \Pi B\|_F^2, \tag{10}$$

and the similarity matrix X in (3) is a positive scalar multiple of the solution $\widetilde{X}$ to

$$\begin{aligned} &\operatorname*{argmin}_{X \in \mathbb{R}^{n \times n}} \|AX - XB\|_F^2 + \eta^2 \|X\|_F^2 \\ &\quad \text{s.t. } \mathbf{1}^\top X \mathbf{1} = n. \end{aligned} \tag{11}$$

(See [14,Corollary 2.2].) This is a convex relaxation of the program (10) with an additional ridge regularization term. As a result, our analysis immediately yields the same exact recovery guarantees for algorithms that round the solution $\widetilde{X}$ to (11) instead of X. In Sect. 6, we study a tighter relaxation of the QAP (10) that imposes row-sum constraints, and establish the same exact recovery guarantees (up to universal constants) by employing similar technical tools.

Organization The rest of the paper is organized as follows. In Sect. 2, we state the main exact recovery guarantees for GRAMPA under the correlated Wigner model, as well as the results specialized to the (sparse) Erdős-Rényi model. We start the analysis by introducing the key resolvent representation of the GRAMPA similarity matrix in Sect. 3. As a preparation for the main proof, Sect. 4 provides the needed tools from random matrix theory. The proof of correctness for GRAMPA is then presented in Sect. 5. In Sect. 6, we extend the theoretical guarantees to a tighter QP relaxation. Finally, Sect. 7 is devoted to proving the resolvent bounds which form the main technical ingredient to our proofs.

Notation Let $[n] \triangleq \{1, \dots, n\}$. Let $\mathbf{i} = \sqrt{-1}$. In a Euclidean space $\mathbb{R}^n$ or $\mathbb{C}^n$, let $\mathbf{e}_i$ be the i-th standard basis vector, and let $\mathbf{1} = \mathbf{1}_n$ be the all-ones vector. Let $\mathbf{J} = \mathbf{J}_n$ denote the $n \times n$ all-ones matrix, and let $\mathbf{I} = \mathbf{I}_n$ denote the $n \times n$ identity matrix. The subscripts are often omitted when there is no ambiguity.

The inner product of $u, v \in \mathbb{C}^n$ is defined as $\langle u, v \rangle = u^* v$. Similarly, for matrices, $\langle A, B \rangle = \mathrm{Tr}(A^* B)$. Let $\|v\| \equiv \|v\|_2 = \langle v, v \rangle$ and $\|v\|_\infty = \sup_i |v_i|$ for vectors. Let $\|M\| \equiv \|M\|_{\mathrm{op}} = \sup_{v: \|v\|=1} \|Mv\|$, $\|M\|_F^2 = \langle M, M \rangle$, and $\|M\|_\infty = \sup_{i,j} |M_{ij}|$ for matrices.

Let $x \wedge y = \min(x, y)$ and $x \vee y = \max(x, y)$. We use $C, C', c, c', \dots$ to denote positive constants that may change at each appearance. For sequences of positive real numbers $(a_n)_{n=1}^\infty$ and $(b_n)_{n=1}^\infty$, we write $a_n \lesssim b_n$ (resp. $a_n \gtrsim b_n$) if there is a constant $C > 0$ such that $a_n \le C b_n$ (resp. $b_n \le C a_n$) for all $n \ge 1$, $a_n \asymp b_n$ if both relations $a_n \lesssim b_n$ and $a_n \gtrsim b_n$ hold, and $a_n \ll b_n$ if $a_n / b_n \to 0$ as $n \to \infty$. We write $a_n = O(b_n)$ if $|a_n| \lesssim b_n$ and $a_n = o(b_n)$ if $|a_n| \ll b_n$.

2 Exact Recovery Guarantees for GRAMPA

In this section, we state the exact recovery guarantees for GRAMPA, making the earlier informal statement precise.

Theorem 1 *Fix constants $a > 0$ and $\kappa > 2$, and let $\eta \in [1/(\log n)^a, 1]$. Consider the correlated Wigner model with $n \ge d \ge (\log n)^{c_0}$ where $c_0 > \max(32 + 4a, 4 + 7a)$. Then, there exist (a, κ)-dependent constants $C_0, n_0 > 0$ and a deterministic quantity*

$r(n) \equiv r(n, \eta, d, a)$ satisfying $r(n) \to 0$ as $n \to \infty$, such that for all $n \geq n_0$, with probability at least $1 - n^{-10}$, the matrix X in (3) satisfies

$$\max_{\ell \neq \pi_*(k)} |X_{k\ell}| \leq C_0 (\log n)^\kappa \frac{1}{\sqrt{\eta}},$$
$$\max_k \left| X_{k\pi_*(k)} - \frac{1-\sigma^2}{\eta} \right| \leq C_0 \left(\frac{r(n)}{\eta} + \frac{\sigma}{\eta^2} + (\log n)^\kappa \frac{1}{\sqrt{\eta}} \right). \tag{12}$$

If there is a universal constant K for which a_{ij} and b_{ij} are sub-Gaussian with $\|a_{ij}\|_{\psi_2}, \|b_{ij}\|_{\psi_2} \leq K/\sqrt{n}$, then the above holds also with $\kappa = 1$.

As an immediate corollary, we obtain the following exact recovery guarantee for GRAMPA.

Corollary 1 (Universal graph matching) *Under the conditions of Theorem 1, there exist constants $c, c' > 0$ such that for all $n \geq n_0$, if*

$$(\log n)^{-a} \leq \eta \leq c(\log n)^{-2\kappa} \quad \text{and} \quad \sigma \leq c'\eta, \tag{13}$$

then with probability at least $1 - n^{-10}$,

$$\min_k X_{k\pi_*(k)} > \max_{\ell \neq \pi_*(k)} X_{k\ell}, \tag{14}$$

and hence $\widehat{\Pi}$ that solves the linear assignment problem (4) equals Π_*.

Proof Let $c = 1/(64C_0^2)$ and $c' = 1/(2C_0)$, where C_0 is the constant given in Theorem 1. Then under assumption (13), we have

$$C_0 (\log n)^\kappa \sqrt{\eta} \leq C_0 (\log n)^\kappa \frac{\sqrt{c}}{(\log n)^\kappa} = C_0 \sqrt{c} \leq 1/8,$$

so $\max_{\ell \neq \pi_*(k)} |X_{k\ell}| \leq 1/(8\eta)$. We also have $C_0\sigma/\eta \leq C_0 c' = 1/2$ and $1 - \sigma^2 > 7/8$ and $C_0 r(n) < 1/8$ for all large n, so that $\max_k X_{k\pi_*(k)} > (7/8 - 1/8 - 1/2 - 1/8)/\eta = 1/(8\eta)$. This implies (14). □

An important application of the above universality result is matching two correlated sparse Erdős-Rényi graphs. Let G be an Erdős-Rényi graph with n vertices and edge probability q, denoted by $G \sim \mathsf{G}(n, q)$. Let $\mathbf{A}$ and $\mathbf{B}'$ be two copies of Erdős-Rényi graphs that are i.i.d. conditional on G, each of which is obtained from G by deleting every edge of G with probability $1 - s$ independently where $s \in [0, 1]$. Then, we have that $\mathbf{A}, \mathbf{B}' \sim \mathsf{G}(n, p)$ marginally where $p \triangleq qs$. Equivalently, we may first sample an Erdős-Rényi graph $\mathbf{A} \sim \mathsf{G}(n, p)$, and then define $\mathbf{B}'$ by

$$\mathbf{B}'_{ij} \sim \begin{cases} \mathsf{Bern}(s) & \text{if } \mathbf{A}_{ij} = 1 \\ \mathsf{Bern}\left(\frac{p(1-s)}{1-p}\right) & \text{if } \mathbf{A}_{ij} = 0. \end{cases}$$

Suppose that we observe a pair of graphs $\mathbf{A}$ and $\mathbf{B} = \Pi_*^\top \mathbf{B}' \Pi_*$, where Π_* is an unknown permutation matrix. We then wish to recover the permutation matrix Π_*.

We transform the adjacency matrices $\mathbf{A}$ and $\mathbf{B}$ so that they satisfy the moment conditions (5) and (6): Define the centered, rescaled versions of $\mathbf{A}$ and $\mathbf{B}$ by

$$A \triangleq (np(1-p))^{-1/2}(\mathbf{A} - \mathbb{E}[\mathbf{A}]) \quad \text{and} \quad B \triangleq (np(1-p))^{-1/2}(\mathbf{B} - \mathbb{E}[\mathbf{B}]). \quad (15)$$

Then, (5) clearly holds, and we check the following additional properties.

Lemma 1 *For all large n, the matrices $A = (a_{ij})$ and $B = (b_{ij})$ satisfy (6), (8), and (9) with $d = np(1-p)$ and*

$$\sigma^2 = \max\left(\frac{1-s}{1-p}, \frac{(\log n)^7}{d}\right).$$

Proof Assume without loss of generality that Π_* is the identity matrix. For any $k \geq 2$, we have

$$\begin{aligned}\mathbb{E}\left[|a_{ij}|^k\right] &= (np(1-p))^{-k/2}\left[p(1-p)^k + (1-p)p^k\right] = \frac{(1-p)^{k-1} + p^{k-1}}{nd^{(k-2)/2}} \\ &\leq \frac{1}{nd^{(k-2)/2}}.\end{aligned}$$

Thus, the moment condition (6) is satisfied. In addition, we have that for all $i < j$,

$$\mathbb{E}\left[a_{ij}b_{ij}\right] = \frac{1}{d}\mathbb{E}\left[(\mathbf{A}_{ij} - p)(\mathbf{B}_{ij} - p)\right] = \frac{1}{d}\left(ps - p^2\right) = \frac{s-p}{n(1-p)} \leq \frac{1-\sigma^2}{n},$$

where the last inequality holds by the choice of σ^2. Thus, (8) is satisfied. Moreover, let $\Delta_{ij} = \frac{1}{\sqrt{2\sigma^2}}\left(a_{ij} - b_{ij}\right)$. It follows that $\mathbb{E}\left[\Delta_{ij}\right] = 0$ and

$$\mathbb{E}\left[|\Delta_{ij}|^k\right] = \frac{2p(1-s)}{\left(2\sigma^2 d\right)^{k/2}} \leq \frac{1}{n(2\sigma^2 d)^{(k-2)/2}}$$

where the last inequality is due to $\sigma^2 \geq \frac{1-s}{1-p}$. Thus, by applying Lemma 3 and $2(\log n)^7 \leq 2\sigma^2 d \leq n$ where the upper bound follows from $p(1-s) \leq s(1-s) \leq 1/4$, there exists a constant $C > 0$ such that for any $D > 0$, with probability at least $1 - n^{-D}$ for all $n \geq n_0(D)$, we have $\|\Delta\| \leq C$ and hence $\|A - B\| \leq \sqrt{2}C\sigma$. Thus, (9) is satisfied. □

Combining Lemma 1 with Corollary 1 immediately yields a sufficient condition for GRAMPA to exactly recover Π_* in the correlated Erdős-Rényi graph model.

Corollary 2 (Erdős-Rényi graph matching) *Suppose that either* *(a) (dense case)*

$$\delta \le p \le 1-\delta, \qquad \frac{1-s}{1-p} \le (\log n)^{-c_1}$$

for constants $\delta \in (0,1)$ and $c_1 > 4$, or

(b) (sparse case)

$$np(1-p) \ge (\log n)^{c_0}, \qquad \frac{1-s}{1-p} \le (\log n)^{-c_1}$$

for constants $c_0 > 48$ and $c_1 > 8$.

There exist (δ, c_0, c_1)-dependent constants $a, n_0 > 0$ such that if $\eta = (\log n)^{-a}$ and $n \ge n_0$, then with probability at least $1 - n^{-10}$,

$$\min_k X_{k\pi_*(k)} > \max_{\ell \ne \pi_*(k)} X_{k\ell},$$

and hence the solution $\widehat{\Pi}$ to the linear assignment problem (4) coincides with Π_*.

Proof For (a), pick $\kappa = 1$ and any a such that $c_1/2 > a > 2\kappa = 2$. For (b), pick any a, κ such that $c_1/2 > a > 2\kappa > 4$ and $c_0 > 32 + 4a > 4 + 7a$. Then, all conditions of Theorem 1 and Corollary 1 are satisfied for large n, and the result follows. □

Comparison to information-theoretic limits and existing algorithmic guarantees of exact recovery For the correlated Erdős-Rényi graph model, exact recovery of the hidden vertex correspondence with high probability is shown to be information-theoretically possible if $nps - \log n \to +\infty$ and $p/s = O(\log^{-3}(n))$, and impossible if $nps^2 - \log n = O(1)$ [7, 8]. From a computational perspective, recent work [9] shows that degree matching can achieve exact recovery with high probability in polynomial time provided that $np \gg n^{4/5} \log^{7/5}(n)$ and $1 - s \ll p^4/\log^6(n)$.

This result is further improved to $np = \Omega(\log^2(n))$ and $1 - s \le O(\log^{-2}(n))$ in [10] by matching degree profiles (that is, empirical distributions of neighbors' degrees). The performance guarantee of the proposed GRAMPA method matches the state of the art of polynomial-time algorithms up to polylogarithmic factors and holds for more general models of correlated matrices. It is worth noting that a quasi-polynomial time ($n^{O(\log n)}$) algorithm is proposed in [2] which succeeds when $np \in \left[n^{o(1)}, n^{1/153}\right] \cup \left[n^{2/3}, n^{1-\epsilon}\right]$ and $s \ge (\log n)^{-o(1)}$. However, it remains open whether exact recovery is achievable in polynomial time for any constant s bounded away from 1. It is conceivable that there exists a "hard regime" where exact recovery is information-theoretically possible but computationally intractable, resembling the conjectured computational hardness for the planted clique problem [3] and the stochastic block model [6].

3 Resolvent Representation

For a real symmetric matrix A with spectral decomposition (2), its resolvent is defined by

$$R_A(z) \triangleq (A - z\mathbf{I})^{-1} = \sum_i \frac{1}{\lambda_i - z} v_i v_i^\top$$

for $z \in \mathbb{C} \setminus \mathbb{R}$. Then, we have the matrix symmetry $R_A(z)^\top = R_A(z)$, conjugate symmetry $\overline{R_A(z)} = R_A(\bar{z})$, and the following Ward identity.

Lemma 2 (Ward identity) *For any $z \in \mathbb{C} \setminus \mathbb{R}$ and any real symmetric matrix A,*

$$R_A(z)\overline{R_A(z)} = \frac{\operatorname{Im} R_A(z)}{\operatorname{Im} z}.$$

Proof By the definition of $R(z) \equiv R_A(z)$ and conjugate symmetry, it holds

$$\begin{aligned}\frac{\operatorname{Im} R(z)}{\operatorname{Im} z} &= \frac{R(z) - \overline{R(z)}}{z - \bar{z}} = \frac{(A - z\mathbf{I})^{-1} - (A - \bar{z}\mathbf{I})^{-1}}{z - \bar{z}} = (A - z\mathbf{I})^{-1}(A - \bar{z}\mathbf{I})^{-1} \\ &= R(z)\overline{R(z)}.\end{aligned}$$

$\square$

The following resolvent representation of X is central to our analysis.

Proposition 1 *Consider symmetric matrices A and B with spectral decompositions (2), and suppose that $\|A\| \leq 2.5$. Then, the matrix X defined in (3) admits the following representation*

$$X = \frac{1}{2\pi} \operatorname{Re} \oint_\Gamma R_A(z)\mathbf{J}R_B(z + \mathbf{i}\eta)dz, \tag{16}$$

where

$$\Gamma = \{z : |\operatorname{Re} z| = 3 \text{ and } |\operatorname{Im} z| \leq \eta/2 \quad \text{or} \quad |\operatorname{Im} z| = \eta/2 \text{ and } |\operatorname{Re} z| \leq 3\} \tag{17}$$

is the rectangular contour with vertices $\pm 3 \pm \mathbf{i}\eta/2$ (See Fig. 1 for an illustration).

Proof We have

$$\begin{aligned}X &= \eta \sum_{i,j} v_i v_i^\top \mathbf{J} \frac{w_j w_j^\top}{(\lambda_i - \mu_j)^2 + \eta^2} \\ &= \eta \sum_i v_i v_i^\top \mathbf{J} R_B(\lambda_i + \mathbf{i}\eta) R_B(\lambda_i - \mathbf{i}\eta)\end{aligned}$$

$$= \operatorname{Im} \sum_i v_i v_i^\top \mathbf{J} R_B(\lambda_i + \mathbf{i}\eta) \tag{18}$$

by Lemma 2. Consider the function $f : \mathbb{C} \to \mathbb{C}^{n\times n}$ defined by $f(z) = \mathbf{J}R_B(z + \mathbf{i}\eta)$. Then, each entry $f_{k\ell}$ is analytic in the region $\{z : \operatorname{Im} z > -\eta\}$. Since Γ encloses each eigenvalue λ_i of A, the Cauchy integral formula yields entrywise equality

$$-\frac{1}{2\pi\mathbf{i}} \oint_\Gamma \frac{f(z)}{\lambda_i - z} dz = f(\lambda_i). \tag{19}$$

Substituting this into (18), we obtain

$$X = \operatorname{Im} \sum_i v_i v_i^\top \left(-\frac{1}{2\pi\mathbf{i}} \oint_\Gamma \frac{f(z)}{\lambda_i - z} dz\right) = \frac{1}{2\pi} \operatorname{Re} \oint_\Gamma R_A(z) f(z) dz, \tag{20}$$

which completes the proof in view of the definition of f. □

4 Tools from Random Matrix Theory

Before proving our main results, we introduce the relevant tools from random matrix theory. In particular, the resolvent bounds in Theorem 2 constitute an important technical ingredient in our analysis.

4.1 Concentration Inequalities

We start with some known concentration inequalities in the literature.

Lemma 3 (Norm bounds) *For any constant $\varepsilon > 0$ and a universal constant $c > 0$, if $n \ge d \ge (\log n)^{6+6\varepsilon}$, then with probability at least $1 - e^{-c(\log n)^{1+\varepsilon}}$,*

$$\|A\| \le 2 + \frac{(\log n)^{1+\varepsilon}}{d^{1/4}}.$$

Proof See [12,Lemma 4.3], where we fix the parameter $\xi = 1 + \varepsilon$ in [12,Eq. (2.4)]. The notational identification is $q \equiv \sqrt{d}$. □

Lemma 4 (Concentration inequalities) *Let $\alpha, \beta \in \mathbb{R}^n$ be independent random vectors with independent entries, satisfying*

$$\begin{aligned} &\mathbb{E}[\alpha_i] = \mathbb{E}[\beta_i] = 0, \qquad \mathbb{E}[\alpha_i^2] = \mathbb{E}[\beta_i^2] = \frac{1}{n}, \\ &\max(\mathbb{E}[|\alpha_i|^k], \mathbb{E}[|\beta_i|^k]) \le \frac{1}{nd^{(k-2)/2}}, \quad \text{for each } k \in [2, (\log n)^{10\log\log n}]. \end{aligned} \tag{21}$$

For any constant $\varepsilon > 0$ and universal constants $C, c > 0$, if $n \ge d \ge (\log n)^{6+6\varepsilon}$, then:

(a) *For each $i \in [n]$, with probability at least $1 - e^{-c(\log n)^{1+\varepsilon}}$,*

$$|\alpha_i| \le \frac{C}{\sqrt{d}}. \tag{22}$$

(b) *For any deterministic vector $v \in \mathbb{C}^n$, with probability at least $1 - e^{-c(\log n)^{1+\varepsilon}}$,*

$$\left|v^\top \alpha\right| \le (\log n)^{1+\varepsilon} \left(\frac{\|v\|_\infty}{\sqrt{d}} + \frac{\|v\|_2}{\sqrt{n}}\right). \tag{23}$$

Furthermore, for any even integer $p \in [2, (\log n)^{10 \log\log n}]$,

$$\mathbb{E}\left[\left|v^\top \alpha\right|^p\right] \le (Cp)^p \left(\frac{\|v\|_\infty}{\sqrt{d}} + \frac{\|v\|_2}{\sqrt{n}}\right)^p. \tag{24}$$

(c) *For any deterministic matrix $M \in \mathbb{C}^{n\times n}$, with probability at least $1 - e^{-c(\log n)^{1+\varepsilon}}$,*

$$\left|\alpha^\top M\alpha - \frac{1}{n}\operatorname{Tr} M\right| \le (\log n)^{2+2\varepsilon} \left(\frac{2\|M\|_\infty}{\sqrt{d}} + \frac{\|M\|_F}{n}\right) \tag{25}$$

and

$$\left|\alpha^\top M\beta\right| \le (\log n)^{2+2\varepsilon} \left(\frac{2\|M\|_\infty}{\sqrt{d}} + \frac{\|M\|_F}{n}\right). \tag{26}$$

Proof See [12,Lemma 3.7, Lemma 3.8, and Lemma A.1(i)], where again we fix $\xi = 1 + \varepsilon$. □

Next, based on the above lemma, we state concentration inequalities for a bilinear form that apply to our setting directly.

Lemma 5 (Concentration of bilinear form) *Let $\alpha, \beta \in \mathbb{R}^n$ be random vectors such that the pairs (α_i, β_i) for $i \in [n]$ are independent, with*

$$\mathbb{E}[\alpha_i] = \mathbb{E}[\beta_i] = 0, \qquad \mathbb{E}\left[\alpha_i^2\right] = \mathbb{E}\left[\beta_i^2\right] = \frac{1}{n}, \qquad \mathbb{E}\left[\alpha_i\beta_i\right] \ge \frac{1-\sigma^2}{n}.$$

Let $M \in \mathbb{C}^{n\times n}$ be any deterministic matrix.

(a) *For any constant $\varepsilon > 0$, suppose (21) holds where $n \ge d \ge (\log n)^{6+6\varepsilon}$. Then, there are universal constants $C, c > 0$ such that with probability at least $1 - e^{-c(\log n)^{1+\varepsilon}}$,*

$$\left|\alpha^\top M\beta - \frac{1-\sigma^2}{n}\operatorname{Tr} M\right| \le C\,(\log n)^{2+2\varepsilon} \left(\frac{1}{n}\|M\|_F + \frac{1}{\sqrt{d}}\|M\|_\infty\right). \tag{27}$$

(b) *Suppose that α_i, β_i are sub-Gaussian with $\|\alpha_i\|_{\psi_2} = \|\beta_i\|_{\psi_2} \le \frac{K}{\sqrt{n}}$ for a constant $K > 0$. Then for any $D > 0$, there exists a constant $C \equiv C_{K,D}$ only depending on K and D such that with probability at least $1 - n^{-D}$,*

$$\left|\alpha^\top M\beta - \frac{1-\sigma^2}{n} \operatorname{Tr} M\right| \le \frac{C \log n}{n} \|M\|_F. \tag{28}$$

Proof In view of the polarization identity

$$\alpha^\top M\beta = \frac{1}{4}(\alpha+\beta)^\top M(\alpha+\beta) - \frac{1}{4}(\alpha-\beta)^\top M(\alpha-\beta),$$

it suffices to analyze the two terms separately. Note that

$$\mathbb{E}\left[(\alpha+\beta)^\top M(\alpha+\beta)\right] = \frac{4-2\sigma^2}{n} \operatorname{Tr} M,$$
$$\mathbb{E}\left[(\alpha-\beta)^\top M(\alpha-\beta)\right] = \frac{2\sigma^2}{n} \operatorname{Tr} M,$$

which yields the desired expectation $\mathbb{E}[\alpha^\top M\beta] = \frac{1-\sigma^2}{n} \operatorname{Tr} M$. Thus, it remains to study the deviation.

To prove the concentration bound (27), we obtain from (25) that, there is a universal constant $c > 0$ such that with probability at least $1 - e^{-c(\log n)^{1+\varepsilon}}$,

$$\left|(\alpha\pm\beta)^\top M(\alpha\pm\beta) - \mathbb{E}[(\alpha\pm\beta)^\top M(\alpha\pm\beta)]\right| \le (\log n)^{2+2\varepsilon}\left(\frac{1}{n}\|M\|_F + \frac{2}{\sqrt{d}}\|M\|_\infty\right),$$

from which (27) easily follows.

The sub-Gaussian concentration bound (28) follows from the Hanson–Wright inequality [16, 20]. More precisely, note that $\max\{\|\alpha+\beta\|_{\psi_2}, \|\alpha-\beta\|_{\psi_2}\} \le \|\alpha\|_{\psi_2} + \|\beta\|_{\psi_2} \le 2K/\sqrt{d}$, so taking $\delta = n^{-D}/2$ in [14,Lemma A.2] yields that with probability at least $1 - n^{-D}$,

$$\left|(\alpha\pm\beta)^\top M(\alpha\pm\beta) - \mathbb{E}\left[(\alpha\pm\beta)^\top M(\alpha\pm\beta)\right]\right| \le C_{K,D}\frac{\log n}{n}\|M\|_F,$$

which completes the proof. □

4.2 The Stieltjes Transform

Denote the semicircle density and its Stieltjes transform by

$$\rho(x) = \frac{1}{2\pi}\sqrt{4-x^2}\,\mathbf{1}_{\{|x|\le 2\}} \quad \text{and} \quad m_0(z) = \int \frac{1}{x-z}\rho(x)dx = \frac{-z+\sqrt{z^2-4}}{2}, \tag{29}$$

respectively, where $m_0(z)$ is defined for $z \notin [-2,2]$, and $\sqrt{z^2-4}$ is defined with a branch cut on $[-2,2]$ so that $\sqrt{z^2-4} \sim z$ as $|z| \to \infty$. We have the conjugate symmetry $\overline{m_0(z)} = m_0(\bar{z})$.

We record the following basic facts about the Stieltjes transform.

Proposition 2 *For each $z \in \mathbb{C} \setminus \mathbb{R}$, the Stieltjes transform $m_0(z)$ is the unique value satisfying*

$$m_0(z)^2 + zm_0(z) + 1 = 0 \quad \text{and} \quad \operatorname{Im} m_0(z) \cdot \operatorname{Im} z > 0. \tag{30}$$

Setting $\zeta(z) \triangleq \min(|\operatorname{Re} z - 2|, |\operatorname{Re} z + 2|)$, uniformly over $z \in \mathbb{C} \setminus [-2,2]$ with $|z| \le 10$,

$$\begin{aligned} &|m_0(z)| \asymp 1, \quad |\operatorname{Im} m_0(z)| \gtrsim |\operatorname{Im} z|, \quad \text{and} \quad |\operatorname{Im} m_0(z)| \asymp \\ &\begin{cases} \sqrt{\zeta(z) + |\operatorname{Im} z|} & \text{if } |\operatorname{Re} z| \le 2, \\ |\operatorname{Im} z| / \sqrt{\zeta(z) + |\operatorname{Im} z|} & \text{if } |\operatorname{Re} z| > 2. \end{cases} \end{aligned} \tag{31}$$

For $x \in [-2,2]$, the continuous extensions

$$m_0^+(x) \triangleq \lim_{z \to x:\, z \in \mathbb{C}^+} m_0(z), \quad m_0^-(x) \triangleq \lim_{z \to x:\, z \in \mathbb{C}^-} m_0(z)$$

from $\mathbb{C}^+$ and $\mathbb{C}^-$ both exist. For all $x \in [-2,2]$, these satisfy

$$\begin{aligned} &m_0^\pm(x)^2 + xm_0^\pm(x) + 1 = 0, \quad m_0^+(x) = \overline{m_0^-(x)}, \\ &\frac{1}{\pi} \operatorname{Im} m_0^+(x) = -\frac{1}{\pi} \operatorname{Im} m_0^-(x) = \rho(x), \quad |m_0^\pm(x)| = 1. \end{aligned} \tag{32}$$

Proof (30) follows from the definition of m_0. (31) follows from [11,Lemma 4.3] and continuity and conjugate symmetry of m_0. For the existence of m_0^+ (and hence also m_0^-), see, e.g., the more general statement of [4,Corollary 1]. The first claim of (32) follows from continuity and (30), the second from conjugate symmetry, the third from the Stieltjes inversion formula, and the last from the fact that the two roots of (30) at $z = x \in [-2,2]$ are $m_0^+(x)$ and $m_0^-(x) = \overline{m_0^+(x)}$, so that $1 = m_0^\pm(x)\overline{m_0^\pm(x)} = |m_0^\pm(x)|^2$. □

4.3 Resolvent Bounds

For a fixed constant $a > 0$ and all large n, we bound the resolvent $R(z) = R_A(z)$ over the spectral domain

$$\begin{aligned} &D = D_1 \cup D_2, \text{ where} \\ &D_1 = \{z \in \mathbb{C} : \operatorname{Re} z \in [-3,3],\ |\operatorname{Im} z| \in [1/(\log n)^a, 1]\}, \text{ and} \\ &D_2 = \{z \in \mathbb{C} : |\operatorname{Re} z| \in [2.6,3],\ |\operatorname{Im} z| \le 1/(\log n)^a\}. \end{aligned}$$

Here, D_1 is the union of two strips in the upper and lower half planes, and D_2 is the union of two strips in the left and right half planes.

Theorem 2 (Resolvent bounds) *Suppose $A \in \mathbb{R}^{n\times n}$ has independent entries $(a_{ij})_{i\le j}$ satisfying (5) and (6). Fix a constant $a > 0$ which defines the domain D, fix $\varepsilon > 0$, and set*

$$b = \max(16 + 3\varepsilon + 2a, 3 + 3\varepsilon + 5a/2), \qquad b' = \max(16 + 4\varepsilon + 2a, 4 + 5\varepsilon + 6a).$$

Suppose $n \ge d \ge (\log n)^{b'}$. Then for some constants $C, c, n_0 > 0$ depending on a and ε, and for all $n \ge n_0$, with probability $1 - e^{-c(\log n)(\log\log n)}$, the following hold simultaneously for every $z \in D$:

(a) *(Entrywise bound) For all $j \ne k \in [n]$,*

$$|R_{jk}(z)| \le \frac{C(\log n)^{2+2\varepsilon+a}}{\sqrt{d}}. \tag{33}$$

For all $j \in [n]$,

$$|R_{jj}(z) - m_0(z)| \le \frac{C(\log n)^{2+2\varepsilon+3a/2}}{\sqrt{d}}. \tag{34}$$

(b) *(Row sum bound) For all $j \in [n]$,*

$$\left|\mathbf{e}_j^\top R(z)\mathbf{1}\right| \le C(\log n)^{1+\varepsilon+a}. \tag{35}$$

(c) *(Total sum bound)*

$$|\mathbf{1}^\top R(z)\mathbf{1} - n \cdot m_0(z)| \le \frac{Cn(\log n)^b}{\sqrt{d}}. \tag{36}$$

The proof follows ideas of [12], and we defer this to Sect. 7. As the spectral parameter z is allowed to converge to the interval $[-2, 2]$ with increasing n, this type of result is often called a "local law" in the random matrix theory literature. The focus of the above is a bit different from the results stated in [12], as we wish to obtain explicit logarithmic bounds for $|\operatorname{Im} z| \asymp 1/\operatorname{polylog}(n)$, rather than bounds for more local spectral parameters down to the scale of $|\operatorname{Im} z| \asymp \operatorname{polylog}(n)/n$.

5 Proof of Correctness for GRAMPA

In this section, we prove Theorem 1. Note that the mapping $B \mapsto \Pi_*^\top B \Pi_*$ for any permutation Π_* induces $w_j \mapsto \Pi_*^\top w_j$ and $X \mapsto X\Pi_*$, since $\mathbf{J}\Pi_*^\top = \mathbf{J}$. By virtue of this equivariance, throughout the proof, we may assume without loss of generality that $\Pi_* = \mathbf{I}$, i.e., the underlying true permutation π_* is the identity permutation. Then, we aim to show that X is diagonally dominant, in the sense that $\min_k X_{kk} > \max_{k\ne\ell} X_{k\ell}$.

In view of Lemma 3, we have that $\|A\| \leq 2.5$ holds with probability $1 - n^{-D}$ for any $D > 0$ and all $n \geq n_0(D)$. In the following, we assume that $\|A\| \leq 2.5$ holds. On this event, by Proposition 1, we get that

$$X_{k\ell} = \frac{1}{2\pi} \operatorname{Re} \oint_{\Gamma} (\mathbf{e}_k^\top R_A(z)\mathbf{1})(\mathbf{e}_\ell^\top R_B(z + \mathbf{i}\eta)\mathbf{1}) dz \tag{37}$$

Note that one may attempt to directly apply (35) to bound the row sums $\mathbf{e}_k^\top R_A(z)\mathbf{1}$ and $\mathbf{e}_\ell^\top R_B(z + \mathbf{i}\eta)\mathbf{1}$. This would yield

$$\left|(\mathbf{e}_k^\top R_A(z)\mathbf{1})(\mathbf{e}_\ell^\top R_B(z + \mathbf{i}\eta)\mathbf{1})\right| \lesssim (\log n)^{2+2\varepsilon+2a},$$

and hence $|X_{k\ell}| \lesssim (\log n)^{2+2\varepsilon+2a}$. However, this estimate is too crude to capture the differences between the diagonal and off-diagonal entries. In fact, the row sum $\mathbf{e}_k^\top R_A(z)\mathbf{1}$ does *not* concentrate on its mean, and the deviation $\mathbf{e}_k^\top R_A(z)\mathbf{1} - m_0(z)$ and $\mathbf{e}_\ell^\top R_B(z + \mathbf{i}\eta)\mathbf{1} - m_0(z)$ is uncorrelated for $k \neq \ell$ and positively correlated for $k = \ell$. For this reason, the diagonal entries of (37) dominate the off-diagonals. Thus, it is crucial to gain a better understanding of the deviation terms. We do so by applying Schur complement decomposition.

5.1 Decomposition Via Schur Complement

We recall the classical Schur complement identity for the inverse of a block matrix.

Lemma 6 (Schur complement identity) *For any invertible matrix $M \in \mathbb{C}^{n\times n}$ and block decomposition*

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

if D is square and invertible, then

$$M^{-1} = \begin{bmatrix} S & -SBD^{-1} \\ -D^{-1}CS & D^{-1} + D^{-1}CSBD^{-1} \end{bmatrix} \tag{38}$$

where $S = (A - BD^{-1}C)^{-1}$.

We decompose $\mathbf{e}_k^\top R_A(z)\mathbf{1}$ and $\mathbf{e}_\ell^\top R_B(z + \mathbf{i}\eta)\mathbf{1}$ using this identity, focusing without loss of generality on $(k, \ell) = (1, 2)$. Let $R_{A,12} \in \mathbb{C}^{2\times 2}$ be the upper-left 2×2 submatrix of R_A, and let $R_A^{(12)} \in \mathbb{C}^{(n-2)\times(n-2)}$ be the resolvent of the $(n-2) \times (n-2)$ minor of A with the first two rows and columns removed. Let $a_1^\top$ and $a_2^\top$ be the first two rows of A with first two entries removed, and let $A_o^\top \in \mathbb{R}^{2\times(n-2)}$ be the stacking of $a_1^\top$ and $a_2^\top$.

The following deterministic lemma approximates $\mathbf{e}_1^\top R_A(z)\mathbf{1}$ based on the Schur complement.

Lemma 7 *Suppose* $|z| \leq 10$, *and*

$$\left\| R_{A,12}(z) - m_0(z)\mathbf{I} \right\| \leq \delta \tag{39}$$

where $0 \leq \delta \leq \min_{z:|z|\leq 10} |m_0(z)|/2$. *Then for a constant* $C > 0$ *and* $k = 1, 2$

$$\left| \mathbf{e}_k^\top R_A(z)\mathbf{1} - m_0(z)\left(1 - a_k^\top R_A^{(12)}(z)\mathbf{1}_{n-2}\right) \right| \leq C\delta\left(1 + \|R_A(z)\mathbf{1}\|_\infty\right). \tag{40}$$

Proof It suffices to consider $k = 1$. Applying the Schur complement identity (38), the first two rows of R_A are given by

$$\begin{bmatrix} R_{A,12} & -R_{A,12}A_o^\top R_A^{(12)} \end{bmatrix}. \tag{41}$$

Thus,

$$\begin{aligned} \mathbf{e}_1^\top R_A(z)\mathbf{1} &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} R_{A,12} & -R_{A,12}A_o^\top R_A^{(12)} \end{bmatrix} \begin{bmatrix} \mathbf{1}_2 \\ \mathbf{1}_{n-2} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \end{bmatrix} R_{A,12}\left(\mathbf{1}_2 - A_o^\top R_A^{(12)}\mathbf{1}_{n-2}\right). \end{aligned}$$

Denote $\varDelta_A \triangleq R_{A,12}(z) - m_0(z)\mathbf{I}$. Then,

$$\begin{aligned} \mathbf{e}_1^\top R_A(z)\mathbf{1} &= \begin{bmatrix} 1 & 0 \end{bmatrix} (m_0(z)\mathbf{I} + \varDelta_A)\left(\mathbf{1}_2 - A_o^\top R_A^{(12)}\mathbf{1}_{n-2}\right). \\ &= m_0(z)\left(1 - a_1^\top R_A^{(12)}\mathbf{1}_{n-2}\right) + \begin{bmatrix} 1 & 0 \end{bmatrix} \varDelta_A\left(\mathbf{1}_2 - A_o^\top R_A^{(12)}\mathbf{1}_{n-2}\right). \\ &= m_0(z)\left(1 - a_1^\top R_A^{(12)}\mathbf{1}_{n-2}\right) + O\left(\delta\left(1 + \left\|A_o^\top R_A^{(12)}\mathbf{1}_{n-2}\right\|\right)\right), \end{aligned} \tag{42}$$

where the last equality applies (39). We next upper bound $\left\|A_o^\top R_A^{(12)}\mathbf{1}_{n-2}\right\|$. In view of the fact that $C \geq |m_0(z)| \geq c$ for absolute constants c and C, the assumption (39) implies that $R_{A,12}$ is invertible with $\|R_{A,12}^{-1}\| \lesssim 1$. Using (41) again, we have

$$A_o^\top R_A^{(12)}\mathbf{1}_{n-2} = \mathbf{1}_2 - R_{A,12}^{-1}\begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix}^\top R_A\mathbf{1}_n. \tag{43}$$

It follows that

$$\left\|A_o^\top R_A^{(12)}\mathbf{1}_{n-2}\right\| \lesssim 1 + \left|\mathbf{e}_1^\top R_A\mathbf{1}_n\right| + \left|\mathbf{e}_2^\top R_A\mathbf{1}_n\right| \lesssim 1 + \|R_A\mathbf{1}_n\|_\infty. \tag{44}$$

The desired bound (40) follows by combining (42) and (44). □

5.2 Off-Diagonal Entries

Without loss of generality, we focus on the off-diagonal entry X_{12}:

$$X_{12} = \frac{1}{2\pi} \operatorname{Re} \oint_{\Gamma} \left(\mathbf{e}_1^\top R_A(z)\mathbf{1}\right) \left(\mathbf{e}_2^\top R_B(z+\mathbf{i}\eta)\mathbf{1}\right) dz.$$

For the given value $a > 0$ in Theorem 1, and for some small constant $\varepsilon > 0$, let b, b' be as defined in Theorem 2. Under the given condition for c_0 in Theorem 1, for $\varepsilon > 0$ sufficiently small, we have $c_0 > b'$ and $c_0 > 2b$—thus $d \gg (\log n)^{b'}$ so Theorem 2 applies, and also $\sqrt{d} \gg (\log n)^b$. Fix the constant κ, where $\kappa = 1$ in the sub-Gaussian case where $\|a_{ij}\|_{\psi_2}, \|b_{ij}\|_{\psi_2} \lesssim 1/\sqrt{n}$, and $\kappa > 2$ otherwise. For ease of notation, we define

$$\delta_1 = \frac{(\log n)^{2+2\varepsilon+3a/2}}{\sqrt{d}}, \quad \delta_2 = \frac{(\log n)^{1+\varepsilon+a}}{\sqrt{n}}, \quad \delta_3 = \frac{(\log n)^b}{\sqrt{d}}, \quad \delta_4 = \frac{(\log n)^{\kappa/2}}{\sqrt{n}}. \tag{45}$$

Note that we have $\delta_i = o(1)$ for each $i = 1, 2, 3, 4$, and also $\delta_1\delta_2^2 n = o(1)$.

5.2.1 Resolvent Approximation

Define an event $\mathcal{E}_1$ wherein the following hold simultaneously for all $z \in \Gamma$:

$$\left\| R_{A,12}(z) - m_0(z)\mathbf{I} \right\| \lesssim \delta_1 \tag{46}$$

$$\left\| R_{B,12}(z+\mathbf{i}\eta) - m_0(z+\mathbf{i}\eta)\mathbf{I} \right\| \lesssim \delta_1 \tag{47}$$

$$\|R_A(z)\mathbf{1}\|_\infty \lesssim \delta_2\sqrt{n} \tag{48}$$

$$\|R_B(z+\mathbf{i}\eta)\mathbf{1}\|_\infty \lesssim \delta_2\sqrt{n}. \tag{49}$$

Applying the resolvent approximations given in Theorem 2, we have that

$$\mathbb{P}\{\mathcal{E}_1\} \ge 1 - e^{-c(\log n)(\log\log n)}.$$

In the following, we assume the event $\mathcal{E}_1$ holds.

On $\mathcal{E}_1$, by Lemma 7, we get that uniformly over $z \in \Gamma$,

$$\mathbf{e}_1^\top R_A(z)\mathbf{1} = m_0(z)\left(1 - a_1^\top R_A^{(12)}\mathbf{1}_{n-2}\right) + O\left(\delta_1\delta_2\sqrt{n}\right), \tag{50}$$

$$\mathbf{e}_2^\top R_B(z+\mathbf{i}\eta)\mathbf{1} = m_0(z+\mathbf{i}\eta)\left(1 - b_2^\top R_B^{(12)}\mathbf{1}_{n-2}\right) + O\left(\delta_1\delta_2\sqrt{n}\right). \tag{51}$$

Each of (50) and (51) is itself $O(\delta_2\sqrt{n})$, by (48) and (49). Then multiplying the two, we have

$$
\begin{aligned}
&\left[\mathbf{e}_1^\top R_A(z)\mathbf{1}\right]\left[\mathbf{e}_2^\top R_B(z+\mathbf{i}\eta)\mathbf{1}\right] \\
&\quad = m_0(z)m_0(z+\mathbf{i}\eta)\left(1 - a_1^\top R_A^{(12)}\mathbf{1}_{n-2} - b_2^\top R_B^{(12)}\mathbf{1}_{n-2} + a_1^\top R_A^{(12)}\mathbf{J}_{n-2}R_B^{(12)}b_2\right) \\
&\qquad + O\left(\delta_1\delta_2^2 n\right).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\oint_\Gamma \left[\mathbf{e}_1^\top R_A(z)\mathbf{1}\right]\left[\mathbf{e}_2^\top R_B(z+\mathbf{i}\eta)\mathbf{1}\right] dz \\
&\qquad = \oint_\Gamma m_0(z)m_0(z+\mathbf{i}\eta)dz - a_1^\top g - b_2^\top h + a_1^\top M b_2 + O\left(\delta_1\delta_2^2 n\right),
\end{aligned}
\tag{52}
$$

where

$$
\begin{aligned}
g &\triangleq \oint_\Gamma m_0(z)m_0(z+\mathbf{i}\eta)R_A^{(12)}(z)\mathbf{1}_{n-2}dz, \\
h &\triangleq \oint_\Gamma m_0(z)m_0(z+\mathbf{i}\eta)R_B^{(12)}(z+\mathbf{i}\eta)\mathbf{1}_{n-2}dz, \\
M &\triangleq \oint_\Gamma m_0(z)m_0(z+\mathbf{i}\eta)R_A^{(12)}(z)\mathbf{J}_{n-2}R_B^{(12)}(z+\mathbf{i}\eta)dz.
\end{aligned}
\tag{53}
$$

5.2.2 Term-By-Term Analysis

Next, we bound the individual terms of (52). By the boundedness of $m_0(z)$, we have

$$
\oint_\Gamma m_0(z)m_0(z+\mathbf{i}\eta)dz = O(1). \tag{54}
$$

Define the event $\mathcal{E}_2$ wherein the following hold simultaneously:

$$
\left|a_1^\top g\right| + \left|b_2^\top h\right| \lesssim \delta_1\left(\|g\|_\infty + \|h\|_\infty\right) + \delta_4\left(\|g\|_2 + \|h\|_2\right) \tag{55}
$$

$$
\left|a_1^\top M b_2\right| \lesssim \delta_1\|M\|_\infty + \delta_4^2\|M\|_F. \tag{56}
$$

Note that the triple (g, h, M) is independent of the pair (a_1, b_2) and a_1 and b_2 are independent. Hence, by first conditioning on (g, h, M) and then applying (23) and (26), we get that

$$
\mathbb{P}\{\mathcal{E}_2\} \ge 1 - n^{-D}
$$

for any constant $D > 0$,[2] and all $n \ge n_0(D)$, in both the sub-Gaussian ($\kappa = 1$) and general ($\kappa > 2$) cases. Henceforth, we assume $\mathcal{E}_2$ holds. It then remains to bound the ℓ_2 and ℓ_∞ norms of g, h, and M.

[2] The constant D can be made arbitrarily large by setting the hidden constants in (55) and (56) sufficiently large.

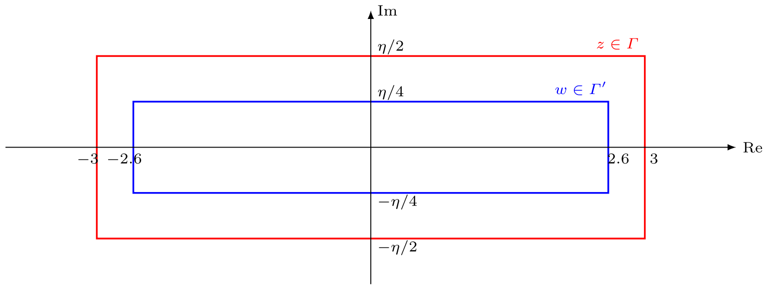

Fig. 1 Nested contours Γ and Γ'

Recall that Γ is the rectangular contour with vertices $\pm 3 \pm \mathbf{i}\frac{\eta}{2}$. Let us define another contour (to be used later) Γ' inside Γ, with vertices $\pm 2.6 \pm \mathbf{i}\frac{\eta}{4}$, cf. Fig. 1. Define the event $\mathcal{E}_3$ wherein the following hold simultaneously for all $z \in \Gamma \cup \Gamma'$:

$$
\begin{aligned}
\left\| R_A^{(12)}(z)\mathbf{1}_{n-2} \right\|_\infty &\lesssim \delta_2 \sqrt{n}, &(57)\\
\left\| R_B^{(12)}(z+\mathbf{i}\eta)\mathbf{1}_{n-2} \right\|_\infty &\lesssim \delta_2 \sqrt{n}, &(58)\\
\left| \mathbf{1}_{n-2}^\top R_A^{(12)}(z)\mathbf{1}_{n-2} - m_0(z)(n-2) \right| &\lesssim \delta_3 n, &(59)\\
\left| \mathbf{1}_{n-2}^\top R_B^{(12)}(z+\mathbf{i}\eta)\mathbf{1}_{n-2} - m_0(z+\mathbf{i}\eta)(n-2) \right| &\lesssim \delta_3 n. &(60)
\end{aligned}
$$

By Theorem 2, we have that $\mathbb{P}\{\mathcal{E}_3\} \geq 1 - e^{-c(\log n)(\log\log n)}$. In the following, we assume the event $\mathcal{E}_3$ holds.

Note that

$$
\|g\|_\infty \lesssim \sup_{z\in\Gamma} \|R_A^{(12)}(z)\mathbf{1}_{n-2}\|_\infty \lesssim \delta_2\sqrt{n}, \tag{61}
$$

where the second inequality holds in view of (57). Similarly, in view of (58), we have that $\|h\|_\infty \lesssim \delta_2\sqrt{n}$. Furthermore,

$$
\begin{aligned}
\|M\|_\infty &\lesssim \sup_{z\in\Gamma} \left\| R_A^{(12)}(z)\mathbf{J}_{n-2}R_B^{(12)}(z+\mathbf{i}\eta) \right\|_\infty \\
&\leq \sup_{z\in\Gamma} \left\| R_A^{(12)}(z)\mathbf{1}_{n-2} \right\|_\infty \left\| \mathbf{1}_{n-2}^\top R_B^{(12)}(z+\mathbf{i}\eta) \right\|_\infty \lesssim \delta_2^2 n. \qquad (62)
\end{aligned}
$$

The ℓ_2 bounds of g, h and M are deferred to Lemma 8. Applying (59), (60), and Lemma 8 with $R_A = R_A^{(12)}$ and $R_B = R_B^{(12)}$, we get $\|g\|_2^2 \lesssim n \log\frac{1}{\eta}$, $\|h\|_2^2 \lesssim n\log\frac{1}{\eta}$ and $\|M\|_F \lesssim n/\sqrt{\eta}$.

Combining the above bounds on the norms of g, h, M with (55), (56), and (54), and plugging into (52), we conclude that on the event $\{\|A\| \leq 2.5\} \cap \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$,

$$
\begin{aligned}
|X_{12}| &= 2\pi \left| \oint_{\Gamma} \left[\mathbf{e}_1^\top R_A(z)\mathbf{1}\right]\left[\mathbf{e}_2^\top R_B(z+\mathbf{i}\eta)\mathbf{1}\right] dz \right| \\
&\lesssim 1 + \delta_4\sqrt{n\log\frac{1}{\eta}} + \delta_4^2 n\frac{1}{\sqrt{\eta}} + \delta_1\delta_2^2 n \lesssim \delta_4^2 n\frac{1}{\sqrt{\eta}} = (\log n)^\kappa \frac{1}{\sqrt{\eta}}, \qquad (63)
\end{aligned}
$$

where in the third step we used $\delta_1\delta_2^2 n = o(1)$ and $\eta \le 1$ so that $\delta_4\sqrt{n} = (\log n)^{\kappa/2} \gtrsim \sqrt{\eta\log\frac{1}{\eta}} + \eta^{1/4}$.

5.2.3 Bounding the Norms of *g*, *h* and *M*

Lemma 8 *Suppose* $\|A\| \le 2.5$ *and* $\left|\mathbf{1}^\top R(z)\mathbf{1}\right| \lesssim n$ *for all* $z \in \Gamma \cup \Gamma'$ *and both* $R(z) = R_A(z)$ *and* $R(z) = R_B(z+\mathbf{i}\eta)$. *Define*

$$
\begin{aligned}
g &= \oint_{\Gamma} m_0(z)m_0(z+\mathbf{i}\eta)R_A(z)\mathbf{1}dz \\
h &= \oint_{\Gamma} m_0(z)m_0(z+\mathbf{i}\eta)R_B(z+\mathbf{i}\eta)\mathbf{1}dz \\
M &= \oint_{\Gamma} m_0(z)m_0(z+\mathbf{i}\eta)R_A(z)\mathbf{J}R_B(z+\mathbf{i}\eta)dz.
\end{aligned}
$$

Then, $\|g\|^2 \lesssim n\log\frac{1}{\eta}$, $\|h\|^2 \lesssim n\log\frac{1}{\eta}$ *and* $\|M\|_F^2 \lesssim \frac{n^2}{\eta}$.

Proof Since $\|A\| \le 2.5$, the function $m_0(z)m_0(z+\mathbf{i}\eta)R_A(z)\mathbf{1}$ is analytic in z in the region between Γ' and Γ. It follows that

$$
g = \oint_{\Gamma} m_0(z)m_0(z+\mathbf{i}\eta)R_A(z)\mathbf{1}dz = \oint_{\Gamma'} m_0(w)m_0(w+\mathbf{i}\eta)R_A(w)\mathbf{1}dw.
$$

Thus,

$$
\begin{aligned}
\|g\|^2 &\overset{(a)}{=} \oint_{\Gamma} dz \oint_{\Gamma'} dw\; m_0(z)m_0(z+\mathbf{i}\eta)m_0(\bar{w})m_0(\bar{w}-\mathbf{i}\eta)\mathbf{1}^\top R_A(\bar{w})R_A(z)\mathbf{1} \\
&\overset{(b)}{=} -\oint_{\Gamma} dz \oint_{\Gamma'} dw\; m_0(z)m_0(z+\mathbf{i}\eta)m_0(w)m_0(w-\mathbf{i}\eta)\mathbf{1}^\top R_A(w)R_A(z)\mathbf{1} \\
&\overset{(c)}{=} -\oint_{\Gamma} dz \oint_{\Gamma'} dw\; m_0(z)m_0(z+\mathbf{i}\eta)m_0(w)m_0(w-\mathbf{i}\eta)\mathbf{1}^\top \frac{R_A(z)-R_A(w)}{z-w}\mathbf{1} \\
&\overset{(d)}{\lesssim} n\oint_{\Gamma} dz \oint_{\Gamma'} \frac{1}{|z-w|} \qquad (64)
\end{aligned}
$$

where (a) applies conjugation symmetry of m_0 and R_A; (b) changes variables $w \mapsto \bar{w}$ which reverses the direction of integration along Γ'; (c) follows from the identity

$$R_A(z)R_A(w) = (A-z)^{-1}(A-w)^{-1} = \frac{1}{z-w}\left[(A-z)^{-1} - (A-w)^{-1}\right]$$
$$= \frac{1}{z-w}\left[R_A(z) - R_A(w)\right] \tag{65}$$

and (d) holds because $|m_0(z)| \asymp 1$ and $\left|\mathbf{1}^\top R_A(z)\mathbf{1}\right| \lesssim n$ for all $z \in \Gamma \cup \Gamma'$ by assumption. For either z or w in the vertical strips of $\Gamma \cup \Gamma'$ of length $O(\eta)$, we apply simply $|z-w| \gtrsim \eta$. For both z and w in the horizontal strips, i.e., $|\operatorname{Im} z| = \eta/2$ and $|\operatorname{Im} w| = \eta/4$, we apply $|z-w| \gtrsim |\operatorname{Re}(z) - \operatorname{Re}(w)| + \eta$. This gives

$$\|g\|^2 \lesssim n\left(1 + \int_{-3}^{3} dx \int_{-2.6}^{2.6} dy \frac{1}{|x-y|+\eta}\right) \lesssim n \log\frac{1}{\eta}.$$

For $\|h\|^2$, we have similarly

$$\|h\|^2 = -\oint_\Gamma dz \oint_{\Gamma'} dw\; m_0(z)m_0(z+\mathbf{i}\eta)m_0(w)m_0(w-\mathbf{i}\eta)\mathbf{1}^\top$$
$$\frac{R_B(z+\mathbf{i}\eta) - R_B(w-\mathbf{i}\eta)}{(z+\mathbf{i}\eta)-(w-\mathbf{i}\eta)}\mathbf{1} \lesssim n \oint_\Gamma dz \oint_{\Gamma'} \frac{1}{|z-w+2\mathbf{i}\eta|.}$$

We may again bound $|z-w+2\mathbf{i}\eta| \gtrsim \eta$ if either z or w belongs to a vertical strip, or $|z-w+2\mathbf{i}\eta| \gtrsim |\operatorname{Re}(z)-\operatorname{Re}(w)| + \eta$ otherwise, to obtain $\|h\|^2 \lesssim n\log(1/\eta)$.

Finally, we bound $\|M\|_F$. Since $\|A\| \le 2.5$, the function $m_0(z)m_0(z + \mathbf{i}\eta)R_A(z)\mathbf{J}R_B(z+\mathbf{i}\eta)$ is analytic in z in the region between Γ' and Γ, so

$$M = \oint_\Gamma m_0(z)m_0(z+\mathbf{i}\eta)R_A(z)\mathbf{J}R_B(z+\mathbf{i}\eta)dz = \oint_{\Gamma'} m_0(w)m_0(w+\mathbf{i}\eta)$$
$$R_A(w)\mathbf{J}R_B(w+\mathbf{i}\eta)dw.$$

Consequently, by the same arguments that leads to (64),

$$\begin{aligned}
&\|M\|_F^2\\
&= \operatorname{Tr}(M^*M)\\
&= \oint_\Gamma dz \oint_{\Gamma'} dw\; m_0(z)m_0(z+\mathbf{i}\eta)m_0(\overline{w})m_0(\overline{w}-\mathbf{i}\eta)\operatorname{Tr}\\
&\quad \times \left[R_A(z)\mathbf{1}\mathbf{1}^\top R_B(z+\mathbf{i}\eta)R_B(\overline{w}-\mathbf{i}\eta)\mathbf{1}\mathbf{1}^\top R_A(\overline{w})\right]\\
&= -\oint_\Gamma dz \oint_{\Gamma'} dw\; m_0(z)m_0(z+\mathbf{i}\eta)m_0(w)m_0(w-\mathbf{i}\eta)\mathbf{1}^\top R_A(w)R_A(z)\mathbf{1}\mathbf{1}^\top\\
&\quad R_B(z+\mathbf{i}\eta)R_B(w-\mathbf{i}\eta)\mathbf{1}\\
&= -\oint_\Gamma dz \oint_{\Gamma'} dw\; m_0(z)m_0(z+\mathbf{i}\eta)m_0(w)m_0(w-\mathbf{i}\eta)\\
&\quad \frac{\mathbf{1}^\top(R_A(z)-R_A(w))\mathbf{1}}{z-w}\frac{\mathbf{1}^\top(R_B(z+\mathbf{i}\eta)-R_B(w-\mathbf{i}\eta))\mathbf{1}}{z+\mathbf{i}\eta-(w-\mathbf{i}\eta)}
\end{aligned}$$

$$\lesssim n^2 \oint_{\Gamma} dz \oint_{\Gamma'} dw \frac{1}{|z-w|} \frac{1}{|z-w+2\mathbf{i}\eta|}.$$

If z or w belongs to a vertical strip of $\Gamma \cup \Gamma'$, of length $O(\eta)$, then $|z-w| \cdot |z-w+2\mathbf{i}\eta| \gtrsim \eta^2$; otherwise, $|z-w| \cdot |z-w+2\mathbf{i}\eta| \gtrsim (|\operatorname{Re}(z)-\operatorname{Re}(w)|+\eta)^2 \gtrsim (\operatorname{Re}(z)-\operatorname{Re}(w))^2+\eta^2$. Then

$$\|M\|_F^2 \lesssim n^2 \left(\frac{1}{\eta} + \int_{-3}^{3} dx \int_{-2.6}^{2.6} dy \frac{1}{(x-y)^2+\eta^2} \right) \lesssim \frac{n^2}{\eta}.$$

□

5.3 Diagonal Entries

Without loss of generality, we consider the diagonal entry X_{11}:

$$X_{11} = \frac{1}{2\pi} \operatorname{Re} \oint_{\Gamma} \left[\mathbf{e}_1^\top R_A(z) \mathbf{1} \right] \left[\mathbf{1}^\top R_B(z+\mathbf{i}\eta) \mathbf{e}_1 \right] dz.$$

By similar arguments as in the off-diagonal entry X_{12} that lead to (50) and (51), we obtain that for all $z \in \Gamma$,

$$\begin{aligned} \mathbf{e}_1^\top R_A(z) \mathbf{1} &= m_0(z) \left(1 - a_1^\top R_A^{(1)}(z) \mathbf{1}_{n-1} \right) + O\left(\delta_1 \delta_2 \sqrt{n} \right) \\ \mathbf{e}_1^\top R_B(z+\mathbf{i}\eta) \mathbf{1} &= m_0(z+\mathbf{i}\eta) \left(1 - b_1^\top R_B^{(1)}(z) \mathbf{1}_{n-1} \right) + O\left(\delta_1 \delta_2 \sqrt{n} \right). \end{aligned}$$

It follows that

$$\begin{aligned} &\left[\mathbf{e}_1^\top R_A(z) \mathbf{1} \right] \left[\mathbf{1}^\top R_B(z+\mathbf{i}\eta) \mathbf{e}_1 \right] \\ &\quad = m_0(z) m_0(z+\mathbf{i}\eta) \left(1 - a_1^\top R_A^{(1)} \mathbf{1}_{n-1} - \mathbf{1}_{n-1}^\top R_B^{(1)} b_1 + a_1^\top R_A^{(1)} \mathbf{J}_{n-1} R_B^{(1)} b_1 \right) \\ &\quad\quad + O\left(\delta_1 \delta_2^2 n \right), \end{aligned}$$

where, respectively, $a_1^\top$ and $b_1^\top$ are the first rows of A and B with first entries removed; and $R_A^{(1)}$ and $R_B^{(1)}$ are the resolvents of the minors of A and B with first rows and columns removed. Thus, we get that

$$\begin{aligned} &\oint_{\Gamma} \left[\mathbf{e}_1^\top R_A(z) \mathbf{1} \right] \left[\mathbf{1}^\top R_B(z+\mathbf{i}\eta) \mathbf{e}_1 \right] dz \\ &\quad = \oint_{\Gamma} m_0(z) m_0(z+\mathbf{i}\eta) dz - a_1^\top g - b_1^\top h + a_1^\top M b_1 + O\left(\delta_1 \delta_2^2 n \right), \end{aligned} \tag{66}$$

where

$$g \triangleq \oint_{\Gamma} m_0(z) m_0(z+\mathbf{i}\eta) R_A^{(1)}(z) \mathbf{1} dz,$$

$$h \triangleq \oint_{\Gamma} m_0(z)m_0(z+\mathbf{i}\eta)R_B^{(1)}(z+\mathbf{i}\eta)\mathbf{1}dz,$$
$$M \triangleq \oint_{\Gamma} m_0(z)m_0(z+\mathbf{i}\eta)R_A^{(1)}(z)\mathbf{J}R_B^{(1)}(z+\mathbf{i}\eta)dz.$$

By the same argument as in the off-diagonal entry X_{12}, we can control each term above. The only difference is that for the bilinear form, instead of using (26), applying Lemma 5 to control $a_1^\top Mb_1$ gives an extra expectation term $(1-\sigma^2)n^{-1}\operatorname{Tr} M$. Therefore, we obtain that for any fixed constant $D > 0$, with probability at least $1 - n^{-D}$, for all sufficiently large n,

$$\left|X_{11} - \frac{1-\sigma^2}{2\pi}\operatorname{Re}\frac{\operatorname{Tr} M}{n}\right| \lesssim (\log n)^\kappa \frac{1}{\sqrt{\eta}}. \tag{67}$$

Denote by $\mathcal{E}_4$ the event where the following hold simultaneously for all $z \in \Gamma$:

$$\|A - B\| \lesssim \sigma$$
$$\left|\mathbf{1}_{n-1}^\top R_A^{(1)}(z)\mathbf{1}_{n-1} - m_0(z)n\right| \lesssim \delta_3 n$$
$$\left|\mathbf{1}_{n-1}^\top R_B^{(1)}(z+\mathbf{i}\eta)\mathbf{1}_{n-1} - m_0(z+\mathbf{i}\eta)n\right| \lesssim \delta_3 n.$$

By the assumption (9) and Theorem 2, we have that $\mathbb{P}\{\mathcal{E}_4\} \geq 1-n^{-D}$ for any constant $D > 0$ and all $n \geq n_0(D)$.

We defer the analysis of $\operatorname{Tr} M$ to Lemmas 9 and 10: Assuming $\mathcal{E}_4$ holds and applying Lemmas 9 and 10 with R_A, R_B replaced by $R_A^{(1)}, R_B^{(1)}$, respectively, we get

$$\frac{1}{n}\operatorname{Re}\operatorname{Tr}(M) = \frac{2\pi + o_\eta(1)}{\eta} + O\left(\frac{\sigma}{\eta^2} + \frac{\delta_3}{\eta}\right). \tag{68}$$

Setting $r(n) = o_\eta(1) + \delta_3$, we get

$$\left|X_{11} - \frac{1-\sigma^2}{\eta}\right| \lesssim \frac{r(n)}{\eta} + \frac{\sigma}{\eta^2} + (\log n)^\kappa \frac{1}{\sqrt{\eta}}.$$

5.3.1 Analyzing the Trace of M

Lemma 9 *Suppose* $\|A\| \leq 2.5$ *and* $\|A - B\| \lesssim \sigma$ *and*

$$\left|\mathbf{1}^\top R_A(z)\mathbf{1} - m_0(z)n\right| \lesssim \delta_3 n,$$
$$\left|\mathbf{1}^\top R_B(z+\mathbf{i}\eta)\mathbf{1} - m_0(z+\mathbf{i}\eta)n\right| \lesssim \delta_3 n, \tag{69}$$

for all $z \in \Gamma$. *Define*

$$M = \oint_{\Gamma} m_0(z)m_0(z+\mathbf{i}\eta)R_A(z)\mathbf{J}R_B(z+\mathbf{i}\eta)dz.$$

Then,

$$\frac{1}{n}\operatorname{Tr} M = \frac{1}{\mathbf{i}\eta}\oint_{\Gamma} m_0(z)m_0(z+\mathbf{i}\eta)(m_0(z+\mathbf{i}\eta)-m_0(z))dz + O\left(\frac{\sigma}{\eta^2}+\frac{\delta_3}{\eta}\right).$$

Proof Applying the identity

$$R_B(z+\mathbf{i}\eta)-R_A(z) = (B-(z+\mathbf{i}\eta))^{-1}-(A-z)^{-1} = R_B(z+\mathbf{i}\eta)(A-B+\mathbf{i}\eta)R_A(z),$$

we get $R_B(z+\mathbf{i}\eta)R_A(z) = \frac{1}{\mathbf{i}\eta}(R_B(z+\mathbf{i}\eta)-R_A(z)-R_B(z+\mathbf{i}\eta)(A-B)R_A(z))$. Therefore

$$\begin{aligned}\operatorname{Tr} M &= \oint_{\Gamma} dz\; m_0(z)m_0(z+\mathbf{i}\eta)\operatorname{Tr}\left[R_A(z)\mathbf{J}R_B(z+\mathbf{i}\eta)\right]\\ &= \oint_{\Gamma} dz\; m_0(z)m_0(z+\mathbf{i}\eta)\mathbf{1}^\top R_B(z+\mathbf{i}\eta)R_A(z)\mathbf{1}\\ &= \frac{1}{\mathbf{i}\eta}\oint_{\Gamma} dz\; m_0(z)m_0(z+\mathbf{i}\eta)\mathbf{1}^\top\left(R_B(z+\mathbf{i}\eta)-R_A(z)\right.\\ &\quad \left.-R_B(z+\mathbf{i}\eta)(A-B)R_A(z)\right)\mathbf{1}. \end{aligned} \tag{70}$$

To proceed, we use the following facts. First, it holds that

$$\left|\mathbf{1}^\top R_B(z+\mathbf{i}\eta)(A-B)R_A(z)\mathbf{1}\right| \le \left\|\mathbf{1}^\top R_B(z+\mathbf{i}\eta)\right\|\|A-B\|\|R_A(z)\mathbf{1}\|.$$

For $z\in\Gamma$ with $\operatorname{Im} z=\pm\eta/2$, in view of the Ward identity given in Lemma 2 and the assumption given in (69), we get that

$$\|R_A(z)\mathbf{1}\|^2 = \mathbf{1}^\top R_A(z)\overline{R_A(z)}\mathbf{1} = \frac{2}{\eta}|\operatorname{Im}\mathbf{1}^\top R_A(z)\mathbf{1}| \lesssim \frac{n}{\eta}$$

For $z\in\Gamma$ with $\operatorname{Re} z=\pm 3$, we have that $\|R_A(z)\mathbf{1}\|^2 \le n\|R_A(z)\|^2 \lesssim n$ thanks to the assumption $\|A\|\le 2.5$. Similarly, we have $\|R_B(z+\mathbf{i}\eta)\mathbf{1}\|^2 \lesssim n/\eta$. Combining these bounds with the assumption that $\|A-B\|\lesssim\sigma$ yields that

$$\left|\mathbf{1}^\top R_B(z+\mathbf{i}\eta)(A-B)R_A(z)\mathbf{1}\right| \lesssim \frac{n\sigma}{\eta}.$$

Then applying $|m_0(z)|\asymp 1$ and (69), we obtain

$$\frac{1}{n}\operatorname{Tr} M = \frac{1}{\mathbf{i}\eta}\oint_{\Gamma} m_0(z)m_0(z+\mathbf{i}\eta)(m_0(z+\mathbf{i}\eta)-m_0(z))dz + O\left(\frac{\sigma}{\eta^2}+\frac{\delta_3}{\eta}\right).$$

□

Lemma 10 *Let Γ be the rectangular contour with vertices $\pm 3 \pm \mathbf{i}\eta/2$. Then*

$$\operatorname{Im}\left[\oint_{\Gamma} m_0(z)m_0(z+\mathbf{i}\eta)(m_0(z+\mathbf{i}\eta)-m_0(z))dz\right] = 2\pi + o_\eta(1).$$

Proof By Proposition 2, the integrand is analytic and bounded over

$$\{z \in \mathbb{C} : \ |z| \le 9, \ z \notin [-2,2], \ z+\mathbf{i}\eta \notin [-2,2]\}.$$

Hence, we may deform Γ to the contour Γ_ϵ with vertices $\pm(2+\varepsilon)\pm i\varepsilon$, and take $\varepsilon \to 0$ (for fixed η). The portion of Γ_ϵ where $|\operatorname{Re} z| > 2$ has total length $O(\varepsilon)$, so the integral over this portion vanishes as $\varepsilon \to 0$. We may apply the bounded convergence theorem for the remaining two horizontal strips of Γ_ϵ to get (recall that contour integrals are evaluated counterclockwise):

$$\begin{aligned}
&\oint_{\Gamma} m_0(z)m_0(z+i\eta)(m_0(z+i\eta)-m_0(z))dz \\
&\quad = \int_{2}^{-2} m_0^+(x)m_0(x+i\eta)(m_0(x+i\eta)-m_0^+(x))dx \\
&\qquad + \int_{-2}^{2} m_0^-(x)m_0(x+i\eta)(m_0(x+i\eta)-m_0^-(x))dx,
\end{aligned}$$

where m_0^+ and m_0^- are the limits from $\mathbb{C}^+$ and $\mathbb{C}^-$ defined in Proposition 2. Now applying the bounded convergence theorem again to take $\eta \to 0$, we have $\lim_{\eta\to 0} m_0(x+i\eta) = m_0^+(x)$ and hence

$$\begin{aligned}
&\lim_{\eta\to 0}\oint_{\Gamma} m_0(z)m_0(z+i\eta)(m_0(z+i\eta)-m_0(z))dz \\
&\quad = \int_{-2}^{2} m_0^-(x)m_0^+(x)(m_0^+(x)-m_0^-(x))dx = \int_{-2}^{2} |m_0^+(x)|^2 \cdot 2\pi\mathbf{i}\rho(x)dx = 2\pi\mathbf{i},
\end{aligned}$$

the last two steps applying (32). Thus, the imaginary part of the integral is $2\pi + o_\eta(1)$ for small η. □

6 A Tighter Regularized QP Relaxation

As discussed in the introduction, GRAMPA can be interpreted as solving the regularized QP relaxation (11) of the QAP (10). We further explore this optimization aspect in this section, and study a tighter regularized QP relaxation.

Let us begin by recalling the following QP relaxation of the QAP (10) that replaces the feasible set of permutation matrices by its convex hull, the Birkhoff polytope consisting of all doubly stochastic matrices [1, 21]:

$$\min_{X \in \mathbb{R}^{n\times n}} \|AX - XB\|_F^2$$
$$\text{s.t. } X\mathbf{1} = \mathbf{1},\ X^\top \mathbf{1} = \mathbf{1},\ X \geq 0. \tag{71}$$

This program differs from the QP relaxation (11) that underlies GRAMPA in two aspects. First, the added ridge penalty $\eta^2\|X\|_F^2$ in (11) is crucial for ensuring the desired statistical property of the solution,[3] while for (71) there is no such need for regularization. Moreover, the Birkhoff polytope constraint, being the tightest possible convex relaxation, is significantly tighter than the constraint $\mathbf{1}^\top X\mathbf{1} = n$. Although it is much slower to solve (71) than to implement GRAMPA, the doubly stochastic relaxation achieves superior performance over the weaker program (11) as demonstrated by ample empirical evidence (cf. [10, 14]); nevertheless, a rigorous theoretical understanding is still lacking.

As a further step toward understanding the relaxations, we analyze the following intermediate program between (71) and (11):

$$\min_{X \in \mathbb{R}^{n\times n}} \|AX - XB\|_F^2 + \eta^2\|X\|_F^2$$
$$\text{s.t. } X\mathbf{1} = \mathbf{1}, \tag{72}$$

where we enforce the sum of each row of X to be equal to one. The above program without the regularization term $\eta^2\|X\|_F^2$ has been studied in [1] in a small noise regime. As we are analyzing the structure of the solution rather than the value of the program, the exact recovery guarantee for GRAMPA (and hence for (11)) does not automatically carries over to the tighter program (72). Fortunately, we are able to employ similar technical tools to analyze the solution to (72), denoted henceforth by X^{c}.

The following result is the counterpart of Theorem 1 and Corollary 1:

Theorem 3 *Fix constants $a > 0$ and $\kappa > 2$, and let $\eta \in [1/(\log n)^a, 1]$.*

Consider the correlated Wigner model with $n \geq d \geq (\log n)^{c_0}$ where $c_0 > \max(34 + 11a, 8 + 12a)$. Then, there exist (α, κ)-dependent constants $C, n_0 > 0$ and a deterministic quantity $r(n) \equiv r(n, \eta, d, a)$ satisfying $r(n) \to 0$ as $n \to \infty$, such that for all $n \geq n_0$, with probability at least $1 - n^{-10}$,

$$\max_{\pi_*(k)\neq \ell} \left|n \cdot X^{\mathsf{c}}_{k\ell}\right| \leq C(\log n)^\kappa \frac{1}{\sqrt{\eta}}, \tag{73}$$

$$\max_k \left|n \cdot X^{\mathsf{c}}_{k\pi_*(k)} - \frac{4(1-\sigma^2)}{\pi\eta}\right| \leq C\left(\frac{r(n)}{\eta} + \frac{\sigma}{\eta^2} + (\log n)^\kappa \frac{1}{\sqrt{\eta}}\right). \tag{74}$$

If $\|a_{ij}\|_{\psi_2}, \|b_{ij}\|_{\psi_2} \leq K/\sqrt{n}$, then the above guarantees hold also for $\kappa = 1$, with constants possibly depending on K.

Furthermore, there exist constants $c, c' > 0$ such that for all $n \geq n_0$, if

$$(\log n)^{-a} \leq \eta \leq c(\log n)^{-2\kappa} \quad \text{and} \quad \sigma \leq c'\eta, \tag{75}$$

[3] See [14,Sect. 1.3] for a more detailed discussion in this regard.

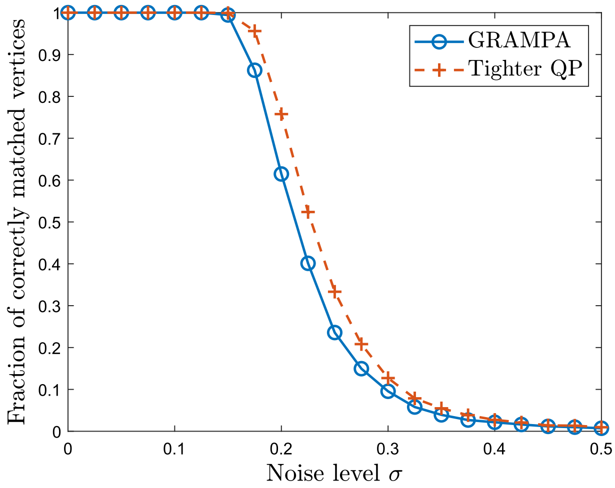

Fig. 2 Fraction of correctly matched pairs of vertices by GRAMPA and the tighter QP (72) (both followed by linear assignment rounding) on Erdős-Rényi graphs with 1000 vertices and edge density 0.5, averaged over 10 repetitions

then with probability at least $1 - n^{-10}$,

$$\min_k X_{k\pi_*(k)} > \max_{\pi_*(k)\neq\ell} X_{k\ell}. \tag{76}$$

Compared with Corollary 1, the theoretical guarantee for the tighter program (72) is similar to that for (11) and the GRAMPA method. In practice the performance of the former is slightly better (cf. Fig. 2). Furthermore, Theorem 3 applies verbatim to the solution of (72) with column-sum constraints $X^\top \mathbf{1} = \mathbf{1}$ instead. This simply follows by replacing (A, B, X, Π_*) with $(B, A, X^\top, \Pi_*^\top)$.

6.1 Structure of Solutions to QP Relaxations

Before proving Theorem 3, we first provide an overview of the structure of solutions to the QP relaxations (11), (72), and (71). Using the Karush–Kuhn–Tucker (KKT) conditions, the solution of (72) can be expressed as

$$X^{\mathsf{c}} = \sum_{i,j} \frac{\langle v_i, \mu\rangle\langle w_j, \mathbf{1}\rangle}{(\lambda_i - \mu_j)^2 + \eta^2} v_i w_j^\top, \tag{77}$$

where $\mu \in \mathbb{R}^n$ is the dual variable corresponding to the row sum constraints, chosen so that X^{c} is feasible. Since

$$X^{\mathsf{c}}\mathbf{1} = \sum_{i,j} \frac{\langle w_j, \mathbf{1}\rangle^2}{(\lambda_i - \mu_j)^2 + \eta^2} v_i v_i^\top \mu = \left\{\sum_i \tau_i v_i v_i^\top\right\} \mu,$$

where

$$\tau_i \triangleq \sum_j \frac{\langle w_j, \mathbf{1}\rangle^2}{(\lambda_i - \mu_j)^2 + \eta^2}. \tag{78}$$

Solving $X^{\mathsf{c}}\mathbf{1} = \mathbf{1}$ yields

$$\mu = \sum_i \frac{\langle v_i, \mathbf{1}\rangle}{\tau_i} v_i, \tag{79}$$

so we obtain

$$X^{\mathsf{c}} = \sum_{i,j} \frac{1}{(\lambda_i - \mu_j)^2 + \eta^2} \frac{1}{\tau_i} v_i v_i^\top \mathbf{J} w_j w_j^\top. \tag{80}$$

Let us provide some heuristics regarding the solution X^{c}. As before we can express τ_i via resolvents as follows:

$$\begin{aligned}\tau_i &= \frac{1}{\eta} \operatorname{Im} \sum_j \frac{\langle w_j, \mathbf{1}\rangle^2}{\mu_j - (\lambda_i + \mathbf{i}\eta)} = \frac{1}{\eta} \mathbf{1}^\top \left[\operatorname{Im} \sum_j \frac{1}{\mu_j - (\lambda_i + \mathbf{i}\eta)} w_j w_j^\top \right] \mathbf{1} \\ &= \frac{1}{\eta} \operatorname{Im}[\mathbf{1}^\top R_B(\lambda_i + \mathbf{i}\eta)\mathbf{1}]. \end{aligned} \tag{81}$$

Invoking the resolvent bound (36), we expect $\tau_i \approx \frac{n}{\eta} \operatorname{Im}[m_0(\lambda_i + \mathbf{i}\eta)]$, where, by properties of the Stieltjes transform (cf. Proposition 2), $\operatorname{Im}[m_0(\lambda_i + \mathbf{i}\eta)] \approx \operatorname{Im}[m_0(\lambda_i)] = \pi\rho(\lambda_i)$ as $\eta \to 0$. Thus, we have the approximation

$$X^{\mathsf{c}} \approx \frac{1}{\pi n} \sum_{i,j} \frac{\eta}{(\lambda_i - \mu_j)^2 + \eta^2} \frac{1}{\rho(\lambda_i)} v_i v_i^\top \mathbf{J} v_j w_j^\top,$$

Compared with the unconstrained solution (3), apart from normalization, the only difference is the extra spectral weight $\frac{1}{\rho(\lambda_i)}$ according to the inverse semicircle density. The effect is that eigenvalues near the edge are upweighted while eigenvalues in the bulk are downweighted, the rationale being that eigenvectors corresponding to the extreme eigenvalues are more robust to noise perturbation.

Remark 1 *(Structure of the QP solutions)* Let us point out that solution of various QP relaxations, including (71), (72), and (11), is of the following common form:

$$X = \sum_{i,j} \frac{\eta}{(\lambda_i - \mu_j)^2 + \eta^2} v_i v_i^\top S w_j w_j^\top, \tag{82}$$

where S is an $n \times n$ matrix that can depend on A and B. Specifically, from the loosest to the tightest relaxations, we have:

- For (11) with the total sum constraint, $S = \alpha\mathbf{J}$, where the dual variable $\alpha > 0$ is chosen for feasibility. Since scaling by α does not affect the subsequent rounding step, this is equivalent to (3) that we analyze.
- For (72) with the row sum constraint, $S = \mu\mathbf{1}^\top$ is rank-one with μ given in (79).
- For (71) without the positivity constraint, $S = \mu\mathbf{1}^\top + \mathbf{1}\nu^\top$ is rank-two. Unfortunately, the dual variables and the spectral structure of the optimal solution turn out to be difficult to analyze.
- For (71) with the positivity constraint, $S = \mu\mathbf{1}^\top + \mathbf{1}\nu^\top + H$, where $H \geq 0$ is the dual variable certifying the positivity of the solution and satisfies complementary slackness.

6.2 Proof of Theorem 3

We now apply the resolvent technique to analyze the behavior of the constrained solution X^{c} and establish its diagonal dominance.

6.2.1 Resolvent Representation of the Solution

We start by giving a resolvent representation of X^{c} via a contour integral.

Lemma 11 *Consider symmetric matrices A and B with the spectral decompositions (2), and suppose that $\|A\| \leq 2.5$. Then, the solution X^{c} of the program (72) admits the following representation*

$$X^{\mathsf{c}} = \frac{1}{2\pi} \operatorname{Re} \oint_\Gamma F(z) R_A(z) \mathbf{J} R_B(z + \mathbf{i}\eta), \tag{83}$$

where Γ is defined by (17) and

$$F(z) \triangleq \frac{2\mathbf{i}}{\mathbf{1}^\top R_B(z + \mathbf{i}\eta)\mathbf{1} - \mathbf{1}^\top R_B(z - \mathbf{i}\eta)\mathbf{1}}. \tag{84}$$

Proof By (81) we have $\tau_i^{-1} = \eta F(\lambda_i)$. This leads to the following contour representation of X^{c} analogous to (16) for the unconstrained solution:

$$\begin{aligned} X^{\mathsf{c}} &= \eta \sum_i F(\lambda_i) v_i v_i^\top \mathbf{J} \left\{ \sum_j \frac{1}{(\lambda_i - \mu_j)^2 + \eta^2} w_j w_j^\top \right\} \\ &\overset{\text{(a)}}{=} \operatorname{Im} \left[\sum_i F(\lambda_i) v_i v_i^\top \mathbf{J} R_B(\lambda_i + \mathbf{i}\eta) \right] \\ &\overset{\text{(b)}}{=} \operatorname{Im} \left[\frac{1}{-2\pi\mathbf{i}} \oint_\Gamma F(z) R_A(z) \mathbf{J} R_B(z + \mathbf{i}\eta) \right] \\ &= \frac{1}{2\pi} \operatorname{Re} \oint_\Gamma F(z) R_A(z) \mathbf{J} R_B(z + \mathbf{i}\eta), \end{aligned}$$

where (a) follows from the Ward identity (Lemma 2); (b) follows from Cauchy integral formula and the analyticity of F in the region enclosed by the contour Γ. $\square$

6.2.2 Entrywise Approximation

For some small constant $\varepsilon > 0$, let b, b' be as defined in Theorem 2. Under the assumptions of Theorem 3, we have $c_0 > b'$ for ε sufficiently small, so that Theorem 2 applies. Recall the notation $\delta_1, \ldots, \delta_4$ defined in (45). For sufficiently small $\varepsilon > 0$, we may also verify under the assumptions of Theorem 3 that $\delta_i = o(1)$ for each $i = 1, 2, 3, 4$, and

$$\frac{\delta_1 \delta_2^2 n}{\eta} \leq 1, \quad \frac{\delta_2^2 \delta_3 n}{\eta^2} \leq \frac{(\log n)^\kappa}{\sqrt{\eta}}, \quad \text{and} \quad \delta_3 \leq \eta^3. \tag{85}$$

We also assume throughout the proof that the high-probability event $\|A\| \leq 2.5$ holds.

Thanks to (36), we can approximate $F(z)$ by

$$\widetilde{F}(z) = \frac{1}{n} \frac{2\mathbf{i}}{m_0(z + \mathbf{i}\eta) - m_0(z - \mathbf{i}\eta)} \tag{86}$$

and approximate X^{c} by

$$\widetilde{X}^{\mathsf{c}} = \frac{1}{2\pi} \operatorname{Re} \oint_\Gamma \widetilde{F}(z) R_A(z) \mathbf{J} R_B(z + \mathbf{i}\eta) \tag{87}$$

$$= \frac{-1}{\pi n} \operatorname{Im} \oint_\Gamma \frac{1}{m_0(z + \mathbf{i}\eta) - m_0(z - \mathbf{i}\eta)} R_A(z) \mathbf{J} R_B(z + \mathbf{i}\eta). \tag{88}$$

The following lemma makes the approximation of X^{c} precise in the entrywise sense:

Lemma 12 *Suppose (85) holds. On the high-probability event where Theorem 2 holds and also $\|A\| \leq 2.5$,*

$$\|\widetilde{X}^c - X^c\|_{\ell_\infty} \lesssim \frac{\delta_2^2 \delta_3}{\eta^2} \leq \frac{(\log n)^\kappa}{n\sqrt{\eta}}, \tag{89}$$

where δ_2, δ_3 are defined in (45).

Proof For notational convenience, put $G(z) = 2\mathbf{i}/(nF(z))$ and $\widetilde{G}(z) = 2\mathbf{i}/(n\widetilde{F}(z))$. Note that $|\operatorname{Im}(z)| \leq \eta/2$ for $z \in \Gamma$, and thus $\operatorname{Im}(z + \mathbf{i}\eta)$ and $\operatorname{Im}(z - \mathbf{i}\eta)$ have different signs. Therefore,

$$|\widetilde{G}(z)| \geq |\operatorname{Im} \widetilde{G}(z)| = |\operatorname{Im} m_0(z + \mathbf{i}\eta)| + |\operatorname{Im} m_0(z - \mathbf{i}\eta)| \gtrsim \eta,$$

where the last step follows from (31). Furthermore, by (36), we have $\sup_{z \in \Gamma} |G(z) - \widetilde{G}(z)| \leq 2C\delta_3$. In view of (85), $\delta_3 \ll \eta$. Hence, we have $|G(z)| \gtrsim \eta$ and

$$\sup_{z \in \Gamma} |F(z) - \widetilde{F}(z)| \lesssim \frac{1}{n} \frac{\delta_3}{\eta^2}.$$

Finally, by (83) and (87), we have

$$|(X^{\mathsf{c}} - \widetilde{X}^{\mathsf{c}})_{k\ell}| \leq \oint_{\Gamma} dz |F(z) - \widetilde{F}(z)||e_k^\top R_A(z)\mathbf{1}||e_\ell^\top R_B(z+\mathbf{i}\eta)\mathbf{1}|.$$

By (35), for all k, ℓ, $|e_k^\top R_A(z)\mathbf{1}| \lesssim \delta_2\sqrt{n}$ and $|e_\ell^\top R_B(z+\mathbf{i}\eta)\mathbf{1}| \lesssim \delta_2\sqrt{n}$. Combining the last two displays yields the desired claim. □

In view of the entrywise approximation, we may switch our attention to the approximate solution $\widetilde{X}^{\mathsf{c}}$ and establish its diagonal dominance, assuming without loss of generality π_* is the identity permutation. The proof parallels the analysis in Sect. 5 so we focus on the differences. To make the scaling identical to the unconstrained case, define

$$Y \triangleq n\widetilde{X}^{\mathsf{c}} = \frac{1}{2\pi} \operatorname{Re} \oint_{\Gamma} f(z) R_A(z) \mathbf{J} R_B(z+\mathbf{i}\eta), \tag{90}$$

with

$$f(z) \triangleq \frac{2\mathbf{i}}{m_0(z+\mathbf{i}\eta) - m_0(z-\mathbf{i}\eta)}.$$

Compared with the unconstrained solution (16), the only difference is the weighting factor $f(z)$.

We aim to show that with probability at least $1-n^{-D}$, for any constant $D>0$, the following holds:

1. For off-diagonals, we have

$$\max_{k\neq\ell} |Y_{k\ell}| \lesssim (\log n)^{\kappa} / \sqrt{\eta}. \tag{91}$$

2. For diagonal entries, we have

$$\min_k \left| Y_{kk} - \frac{4(1-\sigma^2)}{\pi\eta} \right| \lesssim \frac{r(n)}{\eta} + \frac{\sigma}{\eta^2} + (\log n)^{\kappa} \frac{1}{\sqrt{\eta}}. \tag{92}$$

In view of Lemma 12, this implies the desired (73) and (74). Finally, analogous to Corollary 1, under the assumption (75) with constants $c = 1/(64C^2)$ and $c' = 1/(2C)$, for all sufficiently large n,

$$\frac{4(1-\sigma^2)}{\pi\eta} \geq \frac{7}{8\eta} > C\left(\frac{r(n)}{\eta} + \frac{\sigma}{\eta^2} + 2(\log n)^{\kappa}\frac{1}{\sqrt{\eta}}\right),$$

implying the diagonal dominance in (76).

6.2.3 Off-Diagonal Entries

Let us first consider Y_{12}. Recall that for $z \in \Gamma$, we have $|\operatorname{Im}(z+\mathbf{i}\eta)| \gtrsim \eta$, $|\operatorname{Im}(z-\mathbf{i}\eta)| \gtrsim \eta$, and these imaginary parts have opposite signs. Then,

$$|f(z)| \leq \frac{2}{|\operatorname{Im}[m_0(z+\mathbf{i}\eta) - m_0(z-\mathbf{i}\eta)]|} = \frac{2}{|\operatorname{Im} m_0(z+\mathbf{i}\eta)| + |\operatorname{Im} m_0(z-\mathbf{i}\eta)|} \lesssim \frac{1}{\eta}, \tag{93}$$

where the last step applies (31). Analogous to (52), we get

$$\begin{aligned} 2\pi Y_{12} &= \operatorname{Re}\left(\oint_\Gamma f(z)\left[\mathbf{e}_1^\top R_A(z)\mathbf{1}\right]\left[\mathbf{e}_2^\top R_B(z+\mathbf{i}\eta)\mathbf{1}\right]dz\right) \\ &= \operatorname{Re}\left(\alpha - a_1^\top g - b_2^\top h + a_1^\top M b_2\right) + O\left(\frac{\delta_1\delta_2^2 n}{\eta}\right), \end{aligned} \tag{94}$$

where

$$\alpha \triangleq \oint_\Gamma f(z)m_0(z)m_0(z+\mathbf{i}\eta)dz, \tag{95}$$

$$g \triangleq \oint_\Gamma f(z)m_0(z)m_0(z+\mathbf{i}\eta)R_A^{(12)}(z)\mathbf{1}_{n-2}dz, \tag{96}$$

$$h \triangleq \oint_\Gamma f(z)m_0(z)m_0(z+\mathbf{i}\eta)R_B^{(12)}(z+\mathbf{i}\eta)\mathbf{1}_{n-2}dz, \tag{97}$$

$$M \triangleq \oint_\Gamma f(z)m_0(z)m_0(z+\mathbf{i}\eta)R_A^{(12)}(z)\mathbf{J}_{n-2}R_B^{(12)}(z+\mathbf{i}\eta)dz. \tag{98}$$

Here, the constant $\operatorname{Re}\alpha$ is in fact equal to 2π, which is consistent with the row-sum constraints. Indeed, opening up $m_0(z)$ and applying the Cauchy integral formula, we have

$$\begin{aligned} \operatorname{Re}\alpha &= \operatorname{Re}\oint dz \frac{2\mathbf{i}}{m_0(z+\mathbf{i}\eta) - m_0(z-\mathbf{i}\eta)} m_0(z)m_0(z+\mathbf{i}\eta) \\ &= \int \rho(x)dx \operatorname{Re}\oint dz \frac{1}{x-z}\frac{2\mathbf{i}\, m_0(z+\mathbf{i}\eta)}{m_0(z+\mathbf{i}\eta) - m_0(z-\mathbf{i}\eta)} \\ &= \int \rho(x)dx \operatorname{Re}\left[(-2\pi\mathbf{i})\frac{2\mathbf{i}\, m_0(x+\mathbf{i}\eta)}{m_0(x+\mathbf{i}\eta) - m_0(x-\mathbf{i}\eta)}\right] \\ &= 2\pi\int \rho(x)dx \operatorname{Re}\left[\frac{2\, m_0(x+\mathbf{i}\eta)}{2\mathbf{i}\operatorname{Im} m_0(x+\mathbf{i}\eta)}\right] = 2\pi\int\rho(x)dx = 2\pi. \end{aligned} \tag{99}$$

As in Sect. 5.2.2, to bound the linear and bilinear terms, we need to bound the ℓ_∞-norms and ℓ_2-norms of g, h and M. Clearly, by (93), the ℓ_∞-norms are at most an $O(1/\eta)$ factor of those obtained in (61) and (62), i.e., $\|g\|_\infty \lesssim \delta_2\sqrt{n}/\eta$ and $\|M\|_\infty \lesssim \delta_2^2 n/\eta$. The ℓ_2-norms need to be bounded more carefully. The following result is the counterpart of Lemma 8:

Lemma 13 *Assume the same setting of Lemma 8, and define M, g, and h as in (96–98) with R_A, R_B in place of $R_A^{(12)}$, $R_B^{(12)}$. Then, $\|M\|_F^2 \lesssim n^2/\eta$, $\|g\|^2 \lesssim n\log(1/\eta)$, and $\|h\|^2 \lesssim n\log(1/\eta)$.*

Proof We start with $\|M\|_F$, as the arguments for $\|g\|$ and $\|h\|$ are analogous and simpler. Recall the contour Γ' from Fig. 1. Proceeding as in the proof of Lemma 8, we have

$$\begin{aligned}\frac{1}{n^2}\|M\|_F^2 = &-\oint_\Gamma dz \oint_{\Gamma'} dw\, m_0(z)m_0(z+\mathbf{i}\eta)m_0(w)m_0(w-\mathbf{i}\eta)f(z)f(w)\times\\ &\frac{n^{-1}\mathbf{1}^\top(R_A(z)-R_A(w))\mathbf{1}}{z-w}\frac{n^{-1}\mathbf{1}^\top(R_B(z+\mathbf{i}\eta)-R_B(w-\mathbf{i}\eta))\mathbf{1}}{z+\mathbf{i}\eta-(w-\mathbf{i}\eta)}\\ =&\underbrace{-\oint_\Gamma dz\oint_{\Gamma'} dw\, m_0(z)m_0(z+\mathbf{i}\eta)m_0(w)m_0(w-\mathbf{i}\eta)f(z)f(w)\frac{m_0(z)-m_0(w)}{z-w}\frac{m_0(z+\mathbf{i}\eta)-m_0(w-\mathbf{i}\eta)}{z+\mathbf{i}\eta-(w-\mathbf{i}\eta)}}_{\text{(I)}}\\ &+\text{(II)},\end{aligned}$$

where (II) denotes the remainder term. Applying (36), (93), and the boundedness of m_0, the residual term is bounded as

$$|\text{(II)}| \lesssim \delta_3 \oint_\Gamma dz \oint_{\Gamma'} dw|f(z)||f(w)|\frac{1}{|z-w|}\frac{1}{|z+\mathbf{i}\eta-(w-\mathbf{i}\eta)|} \lesssim \frac{\delta_3}{\eta^4} \lesssim \frac{1}{\eta}. \tag{100}$$

To control the leading term (I), let us define the auxiliary contours γ with vertices $\pm(2+2\eta)\pm(\eta/2)\mathbf{i}$ and γ' with vertices $\pm(2+\eta)\pm(\eta/4)\mathbf{i}$. By first deforming Γ' to γ' for each fixed $z\in\Gamma$, then deforming Γ to γ, and finally taking the complex modulus and applying $|m_0|\lesssim 1$, we get

$$|\text{(I)}| \lesssim \oint_\gamma dz \oint_{\gamma'} dw\, |f(z)||f(w)|\left|\frac{m_0(z)-m_0(w)}{z-w}\right|\left|\frac{m_0(z+\mathbf{i}\eta)-m_0(w-\mathbf{i}\eta)}{z+\mathbf{i}\eta-(w-\mathbf{i}\eta)}\right|.$$

The reason for performing these deformations is that for any $z\in\gamma\cup\gamma'$, since $\operatorname{Re} z\in[-2-2\eta, 2+2\eta]$, we have from (31) that $\operatorname{Im} m_0(z+\mathbf{i}\eta) \asymp \sqrt{\eta+\zeta(z)}$ and $-\operatorname{Im} m_0(z-\mathbf{i}\eta) \asymp \sqrt{\eta+\zeta(z)}$, where $\zeta(z)$ is as defined in Proposition 2. Then, we obtain from (93) the improved bound $|f(z)|\lesssim 1/\sqrt{\eta+\zeta(z)}$, and hence

$$|\text{(I)}| \lesssim \oint_\gamma dz \oint_{\gamma'} dw\, \frac{1}{\sqrt{\eta+\zeta(z)}}\frac{1}{\sqrt{\eta+\zeta(w)}}\left|\frac{m_0(z)-m_0(w)}{z-w}\right|\left|\frac{m_0(z+\mathbf{i}\eta)-m_0(w-\mathbf{i}\eta)}{z+\mathbf{i}\eta-(w-\mathbf{i}\eta)}\right|.$$

To bound the above integral, for a small constant $c_0>0$, consider the two cases where $|z-w|\geq c_0$ and $|z-w|<c_0$. For the first case $|z-w|\geq c_0$, we simply apply $|m_0|\lesssim 1$ and $\sqrt{\eta+\kappa}\geq\sqrt{\eta}$ to get that

$$\begin{aligned}\oint\oint_{|z-w|\geq c_0} dz\,dw\, \frac{1}{\sqrt{\eta+\zeta(z)}}\frac{1}{\sqrt{\eta+\zeta(w)}}\left|\frac{m_0(z)-m_0(w)}{z-w}\right|\Bigg|\\ \frac{m_0(z+\mathbf{i}\eta)-m_0(w-\mathbf{i}\eta)}{z+\mathbf{i}\eta-(w-\mathbf{i}\eta)}\Bigg| \lesssim \frac{1}{\eta}.\end{aligned} \tag{101}$$

In the second case $|z-w| < c_0$, we claim that for c_0 sufficiently small, we have

$$|m_0(z) - m_0(w)| \lesssim \sqrt{\eta + \zeta(z)} + \sqrt{\eta + \zeta(w)}, \tag{102}$$

$$|m_0(z + \mathbf{i}\eta) - m_0(w - \mathbf{i}\eta)| \lesssim \sqrt{\eta + \zeta(z)} + \sqrt{\eta + \zeta(w)}. \tag{103}$$

Indeed, if $\zeta(z) > c_0$, then (102) and (103) hold because $\sqrt{\eta + \zeta(z)} + \sqrt{\eta + \zeta(w)} \asymp 1$. If instead $\zeta(z) \le c_0$, say, $\operatorname{Re} z \ge 2 - c_0$, then from the explicit form (29) for $m_0(z)$ we get $1 + m_0(z) = \frac{2 - z + \sqrt{z^2 - 4}}{2}$ and hence

$$|1 + m_0(z)| \lesssim |z - 2| + \sqrt{|z - 2||z + 2|} \asymp \sqrt{|z - 2|} \asymp \sqrt{\eta + \zeta(z)}.$$

Furthermore, since $\operatorname{Re} w \ge \operatorname{Re} z - |z - w| \ge 2 - 2c_0$, we also have $|1 + m_0(w)| \lesssim \sqrt{\eta + \zeta(w)}$. Then, (102) follows from the triangle inequality. The case of $\operatorname{Re} z \le -2 + c_0$, and the argument for (103), are analogous.

Having established (102) and (103), we apply

$$\begin{aligned}
&\frac{\left(\sqrt{\eta + \zeta(z)} + \sqrt{\eta + \zeta(w)}\right)^2}{\sqrt{\eta + \zeta(z)}\sqrt{\eta + \zeta(w)}} \\
&\lesssim \frac{\sqrt{\eta + \max(\zeta(z), \zeta(w))}}{\sqrt{\eta + \min(\zeta(z), \zeta(w))}} \\
&\le \frac{\sqrt{\eta + \min(\zeta(z), \zeta(w))} + \sqrt{|\zeta(z) - \zeta(w)|}}{\sqrt{\eta + \min(\zeta(z), \zeta(w))}} \le 1 + \frac{\sqrt{|z - w|}}{\sqrt{\eta}}
\end{aligned}$$

to get

$$\begin{aligned}
&\oint\oint_{|z-w|<c_0} dz\, dw\, \frac{1}{\sqrt{\eta + \zeta(z)}} \frac{1}{\sqrt{\eta + \zeta(w)}} \left|\frac{m_0(z) - m_0(w)}{z - w}\right| \left|\frac{m_0(z + \mathbf{i}\eta) - m_0(w - \mathbf{i}\eta)}{z + \mathbf{i}\eta - (w - \mathbf{i}\eta)}\right| \\
&\lesssim \oint\oint_{|z-w|<c_0} dz\, dw \left(1 + \frac{\sqrt{|z - w|}}{\sqrt{\eta}}\right) \frac{1}{|z - w||z + \mathbf{i}\eta - (w - \mathbf{i}\eta)|}.
\end{aligned}$$

Then divide this into the integrals where $|z - w| < \eta$ and $|z - w| \ge \eta$, applying

$$\oint\oint_{|z-w|<\eta} dz\, dw\, \frac{1}{|z - w||z + \mathbf{i}\eta - (w - \mathbf{i}\eta)|} \lesssim \oint\oint_{|z-w|<\eta} dz\, dw\, \frac{1}{\eta^2} \lesssim \frac{1}{\eta}$$

and

$$\begin{aligned}
&\oint\oint_{\eta\le|z-w|<c_0} dz\, dw\, \frac{\sqrt{|z - w|}}{\sqrt{\eta}} \cdot \frac{1}{|z - w||z + \mathbf{i}\eta - (w - \mathbf{i}\eta)|} \\
&\lesssim \frac{1}{\sqrt{\eta}} \oint\oint_{\eta\le|z-w|<c_0} dz\, dw\, \frac{1}{|z - w|^{3/2}} \lesssim \frac{1}{\sqrt{\eta}} \frac{1}{\sqrt{\eta}} \lesssim \frac{1}{\eta}.
\end{aligned} \tag{104}$$

Combining with the first case (101), we get $|(\mathrm{I})| \lesssim 1/\eta$. Finally, combining with (100), we get $\|M\|_F^2 \lesssim n^2/\eta$ as desired.

Next we bound $\|g\|$. Proceeding as in the proof of Lemma 8 and following the same argument as above, we get

$$\frac{1}{n}\|g\|^2 \lesssim \oint_{\gamma} dz \oint_{\gamma'} dw\, |f(z)||f(w)| \frac{|m_0(z)-m_0(w)|}{|z-w|} + O\left(\frac{\delta_3}{\eta^3}\right)$$
$$\lesssim \oint_{\gamma} dz \oint_{\gamma'} dw\, \frac{1}{\sqrt{\eta+\zeta(z)}} \frac{1}{\sqrt{\eta+\zeta(w)}} \frac{|m_0(z)-m_0(w)|}{|z-w|} + O\left(\frac{\delta_3}{\eta^3}\right).$$

For $|z-w| \geq c_0$, we have

$$\oint\oint_{|z-w|\geq c_0} dz dw \frac{1}{\sqrt{\eta+\zeta(z)}} \frac{1}{\sqrt{\eta+\zeta(w)}} \frac{|m_0(z)-m_0(w)|}{|z-w|}$$
$$\lesssim \left(\oint \frac{1}{\sqrt{\eta+\zeta(z)}} dz\right)\left(\oint \frac{1}{\sqrt{\eta+\zeta(w)}} dw\right) \lesssim 1.$$

For $|z-w| < c_0$, we apply $|m_0(z)-m_0(w)| \lesssim \sqrt{\eta+\zeta(z)} + \sqrt{\eta+\zeta(w)}$ as above, so that

$$\oint\oint_{|z-w|<c_0} dz dw \frac{1}{\sqrt{\eta+\zeta(z)}} \frac{1}{\sqrt{\eta+\zeta(w)}} \frac{|m_0(z)-m_0(w)|}{|z-w|}$$
$$\lesssim \oint dz \frac{1}{\sqrt{\eta+\zeta(z)}} \oint dw \frac{1}{|z-w|} + \oint dw \frac{1}{\sqrt{\eta+\zeta(w)}} \oint dz \frac{1}{|z-w|}$$
$$\lesssim \log(1/\eta) \cdot \left(\oint dz \frac{1}{\sqrt{\eta+\zeta(z)}} + \oint dw \frac{1}{\sqrt{\eta+\zeta(w)}}\right) \lesssim \log(1/\eta).$$

Combining the above yields $\|g\|^2 \lesssim n\log(1/\eta)$. The argument for $\|h\|^2$ is the same as that for $\|g\|^2$. □

Finally, proceeding as in (55)–(56) and using the preceding norm bounds, we obtain from (94):

$$|Y_{12}| \lesssim\ 1 + \delta_4\sqrt{n\log\frac{1}{\eta}} + \frac{\delta_4^2 n}{\sqrt{\eta}} + \frac{\delta_1\delta_2^2 n}{\eta} \lesssim \frac{\delta_4^2 n}{\sqrt{\eta}} = (\log n)^{\kappa}/\sqrt{\eta},$$

with probability at least $1-n^{-D}$, for any constant D. This implies the desired (91) by the union bound.

6.2.4 Diagonal Entries

We now consider Y_{11}. Following the derivation from (66) to (67) and using Lemma 13 in place of Lemma 8, we obtain, with probability at least $1-n^{-D}$ for any constant D,

$$\left|Y_{11} - \frac{1-\sigma^2}{2\pi}\operatorname{Re}\frac{\operatorname{Tr}(M)}{n}\right| \lesssim (\log n)^{\kappa}\frac{1}{\sqrt{\eta}}, \tag{105}$$

where

$$M \triangleq \oint_\Gamma f(z)m_0(z)m_0(z+\mathbf{i}\eta)R_A^{(1)}(z)\mathbf{J}R_B^{(1)}(z+\mathbf{i}\eta)dz.$$

The trace is computed by the following result, which parallels Lemma 9 and Lemma 10:

Lemma 14 *Suppose $\delta_3 \le \eta^2$. Assume the setting of Lemma 9. Define*

$$M = \oint_\Gamma f(z)m_0(z)m_0(z+\mathbf{i}\eta)R_A(z)\mathbf{J}R_B(z+\mathbf{i}\eta)dz.$$

Then,

$$\frac{1}{n}\operatorname{Tr}(M) = \frac{8+o_\eta(1)}{\eta} + O\left(\frac{\sigma+\delta_3}{\eta^2}\right).$$

Proof Analogous to (70), we have $\frac{1}{n}\operatorname{Tr}(M) = \text{(I)} - \text{(II)}$, where

$$\text{(I)} = \frac{1}{\mathbf{i}\eta}\oint_\Gamma f(z)m_0(z)m_0(z+\mathbf{i}\eta)\frac{1}{n}\mathbf{1}^\top(R_B(z+\mathbf{i}\eta) - R_A(z))\mathbf{1}dz$$
$$\text{(II)} = \frac{1}{\mathbf{i}\eta}\oint_\Gamma f(z)m_0(z)m_0(z+\mathbf{i}\eta)\frac{1}{n}\mathbf{1}^\top R_B(z+\mathbf{i}\eta)(A-B)R_A(z)\mathbf{1}\,dz.$$

To bound (II), consider two cases:

- For $z \in \Gamma$ with $|\operatorname{Im} z| = \eta/2$, by the Ward identity and (36), we have

  $$\|R_A(z)\mathbf{1}\|^2 = \frac{2}{\eta}|\operatorname{Im}\mathbf{1}^\top R_A(z)\mathbf{1}| \lesssim \frac{n}{\eta}(|\operatorname{Im} m_0(z)| + O(\delta_3)).$$

  and similarly,

  $$\|R_B(z+\mathbf{i}\eta)\mathbf{1}\|^2 \lesssim \frac{n}{\eta}(|\operatorname{Im} m_0(z+\mathbf{i}\eta)| + O(\delta_3)).$$

  Thus, it holds that

  $$\left|\mathbf{1}^\top R_B(z+\mathbf{i}\eta)(A-B)R_A(z)\mathbf{1}\right| \lesssim \frac{n\sigma}{\eta}\left(\sqrt{|\operatorname{Im} m_0(z)\operatorname{Im} m_0(z+\mathbf{i}\eta)|} + \sqrt{\delta_3}\right).$$

  Using (31) and (93), we conclude that

  $$|f(z)|\sqrt{|\operatorname{Im} m_0(z)\operatorname{Im} m_0(z+\mathbf{i}\eta)|} \le \frac{2\sqrt{|\operatorname{Im} m_0(z)\operatorname{Im} m_0(z+\mathbf{i}\eta)|}}{|\operatorname{Im} m_0(z+\mathbf{i}\eta)| + |\operatorname{Im} m_0(z-\mathbf{i}\eta)|} \asymp 1$$

  for all $z \in \Gamma$ with $|\operatorname{Im} z| = \eta/2$.

– For $z \in \Gamma$ with $\operatorname{Re} z = \pm 3$, since $\|A\| \le 2.5$, $\left|\mathbf{1}^\top R_B(z + \mathbf{i}\eta)(A - B)R_A(z)\mathbf{1}\right| \lesssim n\sigma$.

Furthermore, by (93), $|f(z)| \lesssim \frac{1}{\eta}$ for all $z \in \Gamma$. Combining the above two cases yields

$$|(\mathrm{II})| \lesssim \frac{\sigma}{\eta^2}\left(1 + \frac{\sqrt{\delta_3}}{\eta}\right) + \frac{\sigma}{\eta} \asymp \frac{\sigma}{\eta^2},$$

since $\delta_3 \le \eta^2$ by the assumption.

For (I), applying (36) again and plugging the definition of $f(z)$ yields

$$(\mathrm{I}) = \frac{2}{\eta}\oint_\Gamma m_0(z)m_0(z+\mathbf{i}\eta)\frac{m_0(z+\mathbf{i}\eta) - m_0(z)}{m_0(z+\mathbf{i}\eta) - m_0(z-\mathbf{i}\eta)}dz + O\left(\frac{\delta_3}{\eta^2}\right).$$

We now apply an argument similar to that of Lemma 10: Note that

$$|m_0(z+\mathbf{i}\eta) - m_0(z-\mathbf{i}\eta)| \ge \operatorname{Im}(m_0(z+\mathbf{i}\eta) - m_0(z-\mathbf{i}\eta)) \gtrsim \eta$$

by (31), so the integrand is bounded for fixed η. Then deforming Γ to Γ_ϵ with vertices $\pm(2+\varepsilon) \pm \mathbf{i}\varepsilon$, taking $\varepsilon \to 0$ for fixed η, and applying the bounded convergence theorem, we have the equality

$$\begin{aligned}
&\oint_\Gamma m_0(z)m_0(z+\mathbf{i}\eta)\frac{m_0(z+\mathbf{i}\eta) - m_0(z)}{m_0(z+\mathbf{i}\eta) - m_0(z-\mathbf{i}\eta)}dz \\
&\quad = \int_2^{-2} m_0^+(x)m_0(x+\mathbf{i}\eta)\frac{m_0(x+\mathbf{i}\eta) - m_0^+(x)}{m_0(x+\mathbf{i}\eta) - m_0(x-\mathbf{i}\eta)}dx \\
&\quad\quad + \int_{-2}^{2} m_0^-(x)m_0(x+\mathbf{i}\eta)\frac{m_0(x+\mathbf{i}\eta) - m_0^-(x)}{m_0(x+\mathbf{i}\eta) - m_0(x-\mathbf{i}\eta)}dx.
\end{aligned} \tag{106}$$

We show that these integrands are uniformly bounded over small η: For any constant $\delta > 0$ and for $|x| \le 2 - \delta$, we have the lower bound

$$|m_0(x+\mathbf{i}\eta) - m_0(x-\mathbf{i}\eta)| = 2\operatorname{Im} m_0(x+\mathbf{i}\eta) \gtrsim \sqrt{\zeta(x)+\eta} \ge \sqrt{\delta}. \tag{107}$$

Then, the above integrands are bounded by $C/\sqrt{\delta}$ for $|x| \le 2-\delta$. For $|x| \in [2-\delta, 2]$, let us apply

$$|m_0(x+\mathbf{i}\eta) - m_0^+(x)| \lesssim \sqrt{\zeta(x)+\eta}$$

as follows from (102) and taking the limit $w \in \mathbb{C}^+ \to x$. We have also $|m_0^+(x) - m_0^-(x)| \asymp \sqrt{\zeta(x)} \lesssim \sqrt{\zeta(x)+\eta}$, so that

$$|m_0(x+\mathbf{i}\eta) - m_0^-(x)| \lesssim \sqrt{\zeta(x)+\eta}.$$

Combining these cases with the first inequality of (107), we see that the integrands of (106) are uniformly bounded for all small η.

Now apply the bounded convergence theorem and take the limit $\eta \to 0$, noting that $\lim_{\eta\to 0} m_0(x+\mathbf{i}\eta) = m_0^+(x)$ and $\lim_{\eta\to 0} m_0(x-\mathbf{i}\eta) = m_0^-(x)$. We get

$$\lim_{\eta\to 0}\oint_\Gamma m_0(z)m_0(z+\mathbf{i}\eta)\frac{m_0(z+\mathbf{i}\eta)-m_0(z)}{m_0(z+\mathbf{i}\eta)-m_0(z-\mathbf{i}\eta)}dz$$
$$=\int_{-2}^{2} m_0^-(x)m_0^+(x)\frac{m_0^+(x)-m_0^-(x)}{m_0^+(x)-m_0^-(x)}dx = \int_{-2}^{2}|m_0^+(x)|^2dx = 4.$$

This gives (I) $= (8+o_\eta(1))/\eta + O(\delta_3/\eta^2)$. Combining with the bound for (II) yields the lemma. □

Finally, combining (105) with Lemma 14 and $\delta_3 \ll \eta$ from (85), and applying a union bound yields the desired (92).

7 Proof of Resolvent Bounds

In this section, we prove Theorem 2. The entrywise bounds of part (a) are essentially the local semicircle law of [12,Theorem 2.8], restricted to the simpler domain $\{z : \mathrm{dist}(z,[-2,2]) \geq (\log n)^{-a}\}$ and with small modifications of the logarithmic factors. The bound in (b) follows from (a) using a straightforward Schur complement identity. The bound in (c) is more involved, and relies on the fluctuation averaging technique of [12,Sect. 5]. We provide a proof of all three statements using the tools of [12].

For each statement, it suffices to establish the claim with the stated probability for each individual point $z \in D$. The uniform statement over $z \in D$ then follows from a union bound over a sufficiently fine discretization of D (of cardinality an arbitrarily large polynomial in n) and standard Lipschitz bounds for m_0 and R_{jk} on the event of $\|A\| \leq 2.5$—we omit these details for brevity.

7.1 Notation and Matrix Identities

In this section, for $S \subset [n]$, denote by $A^{(S)} \in \mathbb{R}^{n\times n}$ the matrix A with all elements in rows and columns belonging to S replaced by 0. Denote

$$R^{(S)}(z) = (A^{(S)} - z\mathbf{I})^{-1} \in \mathbb{C}^{n\times n}.$$

Note that $R^{(S)}(z)$ is block diagonal with respect to the block decomposition $\mathbb{C}^n = \mathbb{C}^S \oplus \mathbb{C}^{[n]\setminus S}$, with $S\times S$ block equal to $(-1/z)\mathbf{I}_{|S|}$ and $([n]\setminus S)\times([n]\setminus S)$ block equal to the resolvent of the corresponding minor of A. (We will typically only access elements of $R^{(S)}$ in this $([n]\setminus S)\times([n]\setminus S)$ block, in which case $R^{(S)}$ may be understood as the resolvent of the minor of A.)

For $i \in [n]$, we write as shorthand

$$iS = \{i\}\cup S, \qquad \sum_k^{(S)} = \sum_{k\in[n]\setminus S}.$$

We usually omit the spectral argument z for brevity.

Lemma 15 (Schur complement identities) *For any $j \in [n]$,*

$$\frac{1}{R_{jj}} = a_{jj} - z - \sum_{k,\ell}^{(j)} a_{jk} R_{k\ell}^{(j)} a_{\ell j}. \tag{108}$$

For any $j \neq k \in [n]$,

$$R_{jk} = -R_{jj} \sum_{\ell}^{(j)} a_{j\ell} R_{\ell k}^{(j)} = R_{jj} R_{kk}^{(j)} \left(-a_{jk} + \sum_{\ell,m}^{(jk)} a_{j\ell} R_{\ell m}^{(jk)} a_{mk} \right), \tag{109}$$

$$\mathbf{e}_k^\top R = \mathbf{e}_k^\top R^{(j)} + \frac{R_{kj}}{R_{jj}} \cdot \mathbf{e}_j^\top R, \tag{110}$$

$$\frac{1}{R_{kk}} = \frac{1}{R_{kk}^{(j)}} - \frac{(R_{kj})^2}{R_{kk}^{(j)} R_{jj} R_{kk}}. \tag{111}$$

For any $j, k, \ell \in [n]$ with $j \notin \{k, \ell\}$,

$$R_{k\ell} = R_{k\ell}^{(j)} + \frac{R_{kj} R_{j\ell}}{R_{jj}}. \tag{112}$$

These identities hold also for any $S \subset [n]$ with R replaced by $R^{(S)}$ and with $j, k, \ell \in [n] \setminus S$.

Proof For all but (110), see [11,Lemma 4.5] and [13,Lemma 4.2]. As for (110), it is equivalent to verify that (112) holds also for $\ell = j$, which simply follows from $R_{kj}^{(j)} = 0$, due to the block diagonal structure of $R^{(j)}$. □

7.2 Entrywise Bound

We say an event occurs w.h.p. if its probability is at least $1 - e^{-c(\log n)^{1+\varepsilon}}$ for a universal constant $c > 0$. Let us show that (33) and (34) hold for $z \in D$ w.h.p.

We start with (34). Note that the jth row $\{a_{jk} : k \in [n]\}$ is independent of $A^{(j)}$ and hence $R^{(j)}$. Applying (108), (22), and (25) conditional on $A^{(j)}$, w.h.p. for all j,

$$\begin{aligned} &\left| \frac{1}{R_{jj}} + z + \frac{1}{n} \sum_{k}^{(j)} R_{kk}^{(j)} \right| \\ &= \left| a_{jj} - \sum_{k,\ell}^{(j)} a_{jk} R_{k\ell}^{(j)} a_{\ell j} + \frac{1}{n} \sum_{k}^{(j)} R_{kk}^{(j)} \right| \\ &\le (\log n)^{2+2\varepsilon} \left(\frac{1}{\sqrt{d}} + \frac{2\|R^{(j)}\|_\infty}{\sqrt{d}} + \frac{\|R^{(j)}\|_F}{n} \right). \end{aligned}$$

Note that $\|R^{(j)}\|_\infty \leq \|R^{(j)}\|$, $\|R^{(j)}\|_F \leq \sqrt{n}\|R^{(j)}\|$, and $d \leq n$. For $z \in D_1$ and any $S \subset [n]$, we have $\|R^{(S)}\| \leq 1/|\operatorname{Im} z| \leq (\log n)^a$. For $z \in D_2$, we have $\|R^{(S)}\| \leq 10$ on the event $\|A\| \leq 2.5$, which occurs w.h.p. by Lemma 3. Then in both cases, we get

$$\left|\frac{1}{R_{jj}} + z + \frac{1}{n}\sum_k^{(j)} R_{kk}^{(j)}\right| \lesssim \frac{(\log n)^{2+2\varepsilon+a}}{\sqrt{d}}. \tag{113}$$

Since $|z| \leq 10$, $|R_{kk}^{(j)}| \leq (\log n)^a$, and $d \gg (\log n)^{4+4\varepsilon}$, this implies $1/|R_{jj}| \lesssim (\log n)^a$. Let $m_n(z) = n^{-1} \operatorname{Tr} R(z)$ be the empirical Stieltjes transform. Then,

$$\begin{aligned}
&\left|m_n - \frac{1}{n}\sum_k^{(j)} R_{kk}^{(j)}\right| \\
&= \left|\frac{1}{n}R_{jj} + \frac{1}{n}\sum_k^{(j)}\left(R_{kk} - R_{kk}^{(j)}\right)\right| \\
&\overset{(112)}{=} \left|\frac{1}{n}\sum_k \frac{R_{kj}^2}{R_{jj}}\right| = \frac{\|\mathbf{e}_j^\top R\|^2}{n|R_{jj}|} \leq \frac{\|R\|^2}{n|R_{jj}|} \lesssim \frac{(\log n)^{3a}}{n}.
\end{aligned}$$

Using $d \leq n$ and combining with (113), w.h.p. for all j,

$$\left|\frac{1}{R_{jj}} + z + m_n\right| \lesssim \frac{(\log n)^{2+2\varepsilon+a}}{\sqrt{d}}. \tag{114}$$

Then by the triangle inequality, also w.h.p. for all $j \neq k$,

$$\left|\frac{1}{R_{jj}} - \frac{1}{R_{kk}}\right| \lesssim \frac{(\log n)^{2+2\varepsilon+a}}{\sqrt{d}},$$

so

$$\begin{aligned}
\left|\frac{m_n}{R_{jj}} - 1\right| &= \left|n^{-1}\sum_k \frac{R_{kk} - R_{jj}}{R_{jj}}\right| \leq \max_k \left|\frac{R_{kk} - R_{jj}}{R_{jj}}\right| \\
&= \max_k |R_{kk}| \left|\frac{1}{R_{jj}} - \frac{1}{R_{kk}}\right| \lesssim \frac{(\log n)^{2+2\varepsilon+2a}}{\sqrt{d}}.
\end{aligned}$$

For $d \gg (\log n)^{4+4\varepsilon+4a}$, this implies $\frac{3}{2}|R_{jj}| \geq |m_n| \geq |R_{jj}|/2$ w.h.p. for all j. Then also

$$\begin{aligned}
\left|\frac{1}{R_{jj}} - \frac{1}{m_n}\right| &= \frac{|R_{jj} - m_n|}{|R_{jj}||m_n|} \leq \max_k \frac{|R_{jj} - R_{kk}|}{|R_{jj}||m_n|} \leq \max_k \frac{2|R_{jj} - R_{kk}|}{|R_{jj}||R_{kk}|} \\
&= 2\max_k \left|\frac{1}{R_{jj}} - \frac{1}{R_{kk}}\right|,
\end{aligned}$$

so

$$\left| \frac{1}{R_{jj}} - \frac{1}{m_n} \right| \lesssim \frac{(\log n)^{2+2\varepsilon+a}}{\sqrt{d}}. \tag{115}$$

Combining with (114), w.h.p. we have

$$\frac{1}{m_n} + z + m_n = r_n, \qquad |r_n| \lesssim \frac{(\log n)^{2+2\varepsilon+a}}{\sqrt{d}} \ll (\log n)^{-a}.$$

Solving for m_n yields

$$m_n \in \frac{-z + r_n \pm \sqrt{z^2 - 4 - 2zr_n + r_n^2}}{2}$$

where the right side denotes the two complex square roots. Note that $|z^2-4| = |z-2| \cdot |z+2| \gtrsim (\log n)^{-a}|z|$ and $|z| \geq (\log n)^{-a}$ for all $z \in D$. Then, as $(\log n)^{-a} \gg |r_n|$, we have $|z^2 - 4| \gg |zr_n| \gg |r_n|^2$. Letting m_0 be the Stieltjes transform of the semicircle law, and letting $\widetilde{m}_0 = 1/m_0$ be the other root of the quadratic Eq. (30), we obtain by a Taylor expansion of the square root that

$$\min(|m_n - m_0|, |m_n - \widetilde{m}_0|) \lesssim |r_n| \left(1 + \frac{|z|}{\sqrt{|z^2-4|}}\right) \lesssim \frac{|r_n|}{\sqrt{\zeta(z) + |\operatorname{Im} z|}}, \tag{116}$$

where $\zeta(z)$ is as defined in Proposition 2.

To argue that this bound holds for $|m_n - m_0|$ rather than $|m_n - \widetilde{m}_0|$, consider first $z \in D_1$ with $\operatorname{Im} z > 0$. In this case $m_n \in \mathbb{C}_+$ and $\widetilde{m}_0 \in \mathbb{C}_-$. Furthermore, note that (31) implies $\operatorname{Im} m_0(z) \geq (\operatorname{Im} z)/\sqrt{\zeta(z) + \operatorname{Im} z}$, and hence $\operatorname{Im} \widetilde{m}_0 = -(\operatorname{Im} m_0)/|m_0|^2 \leq -c(\log n)^{-a}/\sqrt{\zeta(z) + \operatorname{Im} z}$. Since $\operatorname{Im} m_n > 0$ and $|r_n| \ll (\log n)^{-a}$, (116) must hold for $|m_n - m_0|$ rather than $|m_n - \widetilde{m}_0|$. The same argument applies for $z \in D_1$ with $\operatorname{Im} z < 0$. For $z \in D_2$, we have $||m_0(z)| - 1| \geq c$ and hence $|m_0(z) - \widetilde{m}_0(z)| > c$ for a constant $c > 0$. Consider the point $z' \in D_1 \cap D_2$ with $\operatorname{Re} z' = \operatorname{Re} z$ and $\operatorname{Im} z' = (\log n)^{-a}$. Note that for all $z \in D_2$, $|\frac{d}{dz} m_0(z)| \lesssim 1$ and, on the event $\|A\| \leq 2.5$, $|\frac{d}{dz} m_n(z)| \lesssim 1$ also. Thus $|m_0(z) - m_0(z')| \leq C(\log n)^{-a}$ and $|m_n(z) - m_n(z')| \leq C(\log n)^{-a}$. Since we have already shown that (116) holds for $|m_n(z') - m_0(z')|$ in the previous case, this implies also that (116) must hold for $|m_n - m_0|$ rather than for $|m_n - \widetilde{m}_0|$.

Applying $|\operatorname{Im} z| \geq (\log n)^{-a}$, (116) yields w.h.p.

$$|m_n - m_0| \lesssim (\log n)^{a/2} |r_n| \lesssim \frac{(\log n)^{2+2\varepsilon+3a/2}}{\sqrt{d}}. \tag{117}$$

Recalling (115), $|R_{jj}| \leq (\log n)^a$ and $|m_n| \leq \frac{3}{2}|R_{jj}|$, we get

$$|R_{jj} - m_n| \lesssim |R_{jj}||m_n| \cdot \frac{(\log n)^{2+2\varepsilon+a}}{\sqrt{d}} \lesssim \frac{(\log n)^{2+2\varepsilon+3a}}{\sqrt{d}}. \tag{118}$$

Combining the last two displayed equations gives the weak estimate

$$|R_{jj} - m_0| \lesssim \frac{(\log n)^{2+2\varepsilon+3a}}{\sqrt{d}}.$$

Since $d \gtrsim (\log n)^{4+4\varepsilon+6a}$ by assumption, this and $|m_0(z)| \asymp 1$ imply $|R_{jj}| \lesssim 1$ w.h.p. Then, applying the last display and (117) to the first inequality of (118) yields the desired estimate

$$|R_{jj} - m_0| \le |R_{jj} - m_n| + |m_n - m_0| \lesssim \frac{(\log n)^{2+2\varepsilon+3a/2}}{\sqrt{d}}.$$

To show (33) for the off-diagonals, we now apply (109), (22), (26) conditional on $R^{(jk)}$, $|R_{jj}| \lesssim 1$, $|R^{(j)}_{kk}| \lesssim 1$, $\|R^{(jk)}\|_\infty \le (\log n)^a$, $\|R^{(jk)}\|_F \le \sqrt{n}(\log n)^a$, and $d \le n$ to get w.h.p.

$$\begin{aligned}|R_{jk}| &= |R_{jj}||R^{(j)}_{kk}| \left| -a_{jk} + \sum_{\ell,m}^{(jk)} a_{j\ell} R^{(jk)}_{\ell m} a_{mk} \right| \\ &\lesssim (\log n)^{2+2\varepsilon} \left(\frac{1}{\sqrt{d}} + \frac{2\|R^{(jk)}\|_\infty}{\sqrt{d}} + \frac{\|R^{(jk)}\|_F}{n} \right) \lesssim \frac{(\log n)^{2+2\varepsilon+a}}{\sqrt{d}}.\end{aligned}$$

7.3 Row Sum Bound

We now show that (35) holds for $z \in D$ w.h.p. Set

$$\mathcal{Z}_i \triangleq \sum_{j,k}^{(i)} a_{ik} R^{(i)}_{kj} = \sum_{k}^{(i)} a_{ik} \left(\mathbf{e}_k^\top R^{(i)} \mathbf{1} \right) \tag{119}$$

where the last equality holds because $R^{(i)}_{ki} = 0$ for $k \ne i$. Applying (109),

$$\mathbf{e}_i^\top R \mathbf{1} = \sum_j R_{ij} = R_{ii} - R_{ii}\mathcal{Z}_i.$$

Then applying (34), w.h.p. for every $i \in [n]$,

$$\left| \mathbf{e}_i^\top R \mathbf{1} \right| \lesssim 1 + |\mathcal{Z}_i|. \tag{120}$$

Applying (23) conditional on $A^{(i)}$, w.h.p. for every $i \in [n]$,

$$|\mathcal{Z}_i| \le (\log n)^{1+\varepsilon} \left(\frac{\max_{k \ne i} |\mathbf{e}_k^\top R^{(i)} \mathbf{1}|}{\sqrt{d}} + \sqrt{\frac{\sum_k^{(i)} |\mathbf{e}_k^\top R^{(i)} \mathbf{1}|^2}{n}} \right). \tag{121}$$

For the second term above, we apply $\|R^{(i)}\| \leq (\log n)^a$ w.h.p. to get

$$\sum_{k}^{(i)} \left|\mathbf{e}_k^\top R^{(i)} \mathbf{1}\right|^2 \leq \mathbf{1}^\top \overline{R^{(i)}} R^{(i)} \mathbf{1} \leq (\log n)^{2a} n. \tag{122}$$

For the first term, we apply (110), (33), and (34) to get, w.h.p. for all $k \neq i$,

$$\left|\mathbf{e}_k^\top R^{(i)} \mathbf{1}\right| = \left|\mathbf{e}_k^\top R\mathbf{1} - \frac{R_{ki}}{R_{ii}} \cdot \mathbf{e}_i^\top R\mathbf{1}\right| \leq \left|\mathbf{e}_k^\top R\mathbf{1}\right| + \frac{C(\log n)^{2+2\varepsilon+a}}{\sqrt{d}} \left|\mathbf{e}_i^\top R\mathbf{1}\right|. \tag{123}$$

Applying $d \gg (\log n)^{4+4\varepsilon+2a}$ and substituting (122) and (123) into (121) and then into (120), we get that

$$\left|\mathbf{e}_i^\top R\mathbf{1}\right| \lesssim 1 + (\log n)^{1+\varepsilon} \left(\frac{\max_k |\mathbf{e}_k^\top R\mathbf{1}|}{\sqrt{d}} + (\log n)^a\right) \tag{124}$$

Taking the maximum over i and rearranging yields (35).

7.4 Total Sum Bound

Finally, we show that (36) holds with probability $1 - e^{-c(\log n)(\log\log n)}$ for $z \in D$. As above, we set

$$\mathcal{Z}_i = \sum_{j,k}^{(i)} a_{ik} R_{kj}^{(i)} = \sum_{k}^{(i)} a_{ik} \left(\mathbf{e}_k^\top R^{(i)} \mathbf{1}\right). \tag{125}$$

Note that if we apply (122), (123), and (35) to (121), we obtain w.h.p. that for every $i \in [n]$,

$$|\mathcal{Z}_i| \leq (\log n)^{1+\varepsilon+a}. \tag{126}$$

The main step of the proof of (36) is to use the weak dependence of $\mathcal{Z}_1, \ldots, \mathcal{Z}_n$ to obtain a bound on $n^{-1}\sum_i \mathcal{Z}_i$ that is better than $(\log n)^{1+\varepsilon+a}$. The idea is encapsulated by the following abstract lemma from [12].

Lemma 16 (Fluctuation averaging) *Let Ξ be an event defined by A, let $\mathcal{Z}_1, \ldots, \mathcal{Z}_n$ be random variables which are functions of A, let p be an (n-dependent) even integer, and let $x, y > 0$ be deterministic positive quantities. Suppose there exist random variables $\mathcal{Z}_i^{[U]}$, indexed by $U \subseteq [n]$ and $i \in [n] \setminus U$, which satisfy $\mathcal{Z}_i^{[\emptyset]} = \mathcal{Z}_i$ as well as the following conditions:*

(i) *Let a_i denote the i^{th} row of A. Then $\mathcal{Z}_i^{[U]}$ is independent of $\{a_j : j \in U\}$, and $\mathbb{E}_i\left[\mathcal{Z}_i^{[U]}\right] = 0$ where $\mathbb{E}_i$ is the partial expectation over only a_i.*

(ii) *For any* $U \subseteq S \subset [n]$ *with* $|S| \leq p$, *and for any* $i \notin S$, *denote* $u = |U| + 1$ *and*

$$\mathcal{Z}_i^{S,U} = \sum_{T: T \subseteq U} (-1)^{|T|} \mathcal{Z}_i^{[(S \setminus U) \cup T]}. \tag{127}$$

Then for a constant $C > 0$ *and any integer* $r \in [0, p]$,

$$\mathbb{E}\left[\mathbb{1}\{\Xi\} \left|\mathcal{Z}_i^{S,U}\right|^r\right] \leq \left(y(Cxu)^u\right)^r.$$

Furthermore,

$$x \leq 1/(p^5 \log n).$$

(iii) *Let* $\mathcal{A} \subset \mathbb{R}^{n \times n}$ *be the matrices satisfying* Ξ, *i.e.,* $\Xi = \{A \in \mathcal{A}\}$. *Let* $\mathcal{A}_i = \{B \in \mathbb{R}^{n \times n} : B^{(i)} = A^{(i)}$ *for some* $A \in \mathcal{A}\}$, *and define the event* $\Xi_i = \{A \in \mathcal{A}_i\}$. *For a constant* $C > 0$ *and any* U, S, i *as above,* $\mathbb{E}\left[\mathbb{1}\{\Xi_i\} \left|\mathcal{Z}_i^{S,U}\right|^2\right] \leq n^{Cp}$.

(iv) *For a constant* $C > 0$ *and any* $U \subseteq [n]$, $\mathbb{1}\{\Xi\} \left|\mathcal{Z}_i^{[U]}\right| \leq yn^C$.

(v) *For a constant* $\varepsilon > 0$, $\mathbb{P}[\Xi] \geq 1 - e^{-c(\log n)^{1+\varepsilon} p}$.

Then for constants $C', n_0 > 0$ *depending on* C, ε *above, and for all* $n \geq n_0$,

$$\mathbb{P}\left[\mathbb{1}\{\Xi\} \left|n^{-1} \sum_i \mathcal{Z}_i\right| \geq p^{12} y \left(x^2 + n^{-1}\right)\right] \leq (C'/p)^p.$$

Proof See [12,Theorem 5.6]. (The theorem is stated for $1 + \varepsilon = 3/2$ in condition (v), but the proof holds for any $\varepsilon > 0$.) □

The important condition encapsulating weak dependence above is (ii). Applying (ii) with $U = \emptyset$, the condition requires first that each $|\mathcal{Z}_i^{[S]}|$, and in particular each $|\mathcal{Z}_i| = |\mathcal{Z}_i^{[\emptyset]}|$, is of typical size Cxy. In the application of this lemma, for $S = U$ and $i \notin U$, we will define the variables $\mathcal{Z}_i^{[V]}$ for $\emptyset \subseteq V \subseteq U$ such that the quantity $\mathcal{Z}_i^{U,U}$ in (127) is the variable $\mathcal{Z}_i$ with its dependence on all $\{a_j : j \in U\}$ projected out by an inclusion–exclusion procedure. Then, condition (ii) requires that $\mathcal{Z}_i$ depends weakly on $\{a_j : j \in U\}$, in the sense that $|\mathcal{Z}_i^{U,U}|$ is of typical size $x^{|U|+1} y \cdot (C(|U|+1))^{|U|+1}$, which is roughly smaller than $|\mathcal{Z}_i|$ by a factor of $x|U|$ for each element of U. Assuming $1/\sqrt{n} \ll x \ll p^{-12}$, the above then estimates the average $|n^{-1} \sum_i \mathcal{Z}_i|$ to be of the smaller order $p^{12} yx^2 \ll xy$. We refer the reader to the discussion in [12] for additional details.

We will check that the conditions of this lemma hold for $\mathcal{Z}_i$ as defined by (125), with the appropriate construction of variables $\mathcal{Z}_i^{[U]}$. To this end, we first extend (33), (34), and (35) to $R^{(S)}$ for $|S| \leq \log n$ in the following deterministic lemma:

Lemma 17 *Suppose* (33), (34), and (35) *hold with the constant $C \equiv C_0$ for a deterministic symmetric matrix A, some $z \in D$, and all $j, k \in [n]$. Then for all $S \subset [n]$ with $|S| \leq \log n$, and all $j \neq k \in [n] \setminus S$,*

$$|R_{jj}^{(S)}(z) - m_0(z)| \leq \frac{2C_0(\log n)^{2+2\varepsilon+3a}}{\sqrt{d}}, \tag{128}$$

$$|R_{jk}^{(S)}(z)| \leq \frac{2C_0(\log n)^{2+2\varepsilon+a}}{\sqrt{d}}, \tag{129}$$

$$|\mathbf{e}_j^\top R^{(S)}(z)\mathbf{1}| \leq 2C_0(\log n)^{1+\varepsilon+a}. \tag{130}$$

Proof For integers $s \geq 0$, let

$$\Lambda_s^d = \max\left\{|R_{jj}^{(S)} - m_0| : \ |S| = s, \ j \in [n] \setminus S\right\},$$
$$\Lambda_s^o = \max\left\{|R_{jk}^{(S)}| : \ |S| = s, \ j \neq k \in [n] \setminus S\right\}.$$

When (33) and (34) hold, we have that $\Lambda_s^d \leq C_0(\log n)^{2+2\varepsilon+3a}/\sqrt{d}$ and $\Lambda_s^o \leq C_0(\log n)^{2+2\varepsilon+a}/\sqrt{d}$ for $s = 0$. By (112), we have for each $s \geq 1$ and $* \in \{d, o\}$ that

$$\Lambda_{s+1}^* \leq \Lambda_s^* + \frac{(\Lambda_s^o)^2}{|m_0| - \Lambda_s^d}. \tag{131}$$

Assume inductively that for some $s \leq \log n$,

$$\Lambda_s^d \leq \frac{C_0(\log n)^{2+2\varepsilon+3a}}{\sqrt{d}}\left(1 + \frac{4C_0(\log n)^{2+2\varepsilon+a}}{|m_0|\sqrt{d}}\right)^s,$$
$$\Lambda_s^o \leq \frac{C_0(\log n)^{2+2\varepsilon+a}}{\sqrt{d}}\left(1 + \frac{4C_0(\log n)^{2+2\varepsilon+a}}{|m_0|\sqrt{d}}\right)^s. \tag{132}$$

Applying $d \gg (\log n)^{6+4\varepsilon+2a}$, $|m_0| \geq c$, and $s \leq \log n$, this implies in particular that

$$\Lambda_s^d \leq \frac{2C_0(\log n)^{2+2\varepsilon+3a}}{\sqrt{d}}, \qquad \Lambda_s^o \leq \frac{2C_0(\log n)^{2+2\varepsilon+a}}{\sqrt{d}}.$$

We then have $|m_0| - \Lambda_s^d \geq |m_0|/2$ for $d \gg (\log n)^{4+4\varepsilon+6a}$, so (131) yields

$$\Lambda_{s+1}^* \leq \max(\Lambda_s^*, \Lambda_s^o)\left(1 + \frac{2\Lambda_s^o}{|m_0|}\right) \leq \max(\Lambda_s^*, \Lambda_s^o)\left(1 + \frac{4C_0(\log n)^{2+2\varepsilon+a}}{|m_0|\sqrt{d}}\right).$$

Thus, both bounds of (132) hold for $s + 1$, completing the induction. This establishes (128) and (129).

To show (130), set

$$\Gamma_s = \max\{|\mathbf{e}_j^\top R^{(S)}\mathbf{1}| : \ |S| = s, \ j \notin S\}.$$

When (35) holds, $\Gamma_0 \leq C_0(\log n)^{1+\varepsilon+a}$. Applying (110) and the bound $|m_0| - \Lambda_s^d \geq |m_0|/2$, we have

$$\Gamma_{s+1} \leq (1 + 2\Lambda_s^o/|m_0|)\Gamma_s \overset{(129)}{\leq} \left(1 + \frac{4C_0(\log n)^{2+2\varepsilon+3a}}{|m_0|\sqrt{d}}\right)\Gamma_s,$$

Thus, $\Gamma_s \leq 2\Gamma_0$ for all $s \leq \log n$. □

Lemma 18 *Fix $z \in D$. Let $\mathcal{Z}_i$ be defined in (125). For $U \subset [n]$ not containing i, define*

$$\mathcal{Z}_i^{[U]} = \sum_{j,k}^{(iU)} a_{ik} R_{kj}^{(iU)} = \sum_{k}^{(iU)} a_{ik} \left(\mathbf{e}_k^\top R^{(iU)} \mathbf{1}\right).$$

Let Ξ be the event where

- (33), (34), and (35) *all hold at z, for all distinct $j, k \in [n]$,*
- $|a_{ij}| \leq 1$ *for all $i, j \in [n]$, and*
- $\|A\| \leq 2.5$.

Let $p \in [2, (\log n) - 1]$ be an even integer, and set

$$x = \frac{(\log n)^{2+2\varepsilon+a}}{\sqrt{d}}, \qquad y = C'\sqrt{d}(\log n)^{-\varepsilon}$$

for a sufficiently large constant $C' > 0$. Then, all of the conditions of Lemma 16 are satisfied.

Proof Condition (i) is clear by definition, as row a_i of A is independent of $R^{(iU)}$.

To check (ii), note first that the bound $x \leq 1/(p^5 \log n)$ follows from $d \geq (\log n)^{16+4\varepsilon+2a}$. For $U \subseteq S$ and $i \notin S$ we write

$$\begin{aligned}
\mathcal{Z}_i^{S,U} &= \sum_{T:\, T \subseteq U} (-1)^{|T|} \mathcal{Z}_i^{[(S \setminus U) \cup T]} \\
&= \sum_{T:\, T \subseteq U} (-1)^{|T|} \sum_{k}^{((iS \setminus U) \cup T)} a_{ik} (\mathbf{e}_k^\top R^{((iS \setminus U) \cup T)} \mathbf{1}) \\
&= \sum_{k \in U} a_{ik} \left(\sum_{T:\, T \subseteq U \setminus \{k\}} (-1)^{|T|} (\mathbf{e}_k^\top R^{((iS \setminus U) \cup T)} \mathbf{1}) \right) \\
&\quad + \sum_{k}^{(iS)} a_{ik} \left(\sum_{T:\, T \subseteq U} (-1)^{|T|} (\mathbf{e}_k^\top R^{((iS \setminus U) \cup T)} \mathbf{1}) \right) \\
&\triangleq \sum_{k \in U} a_{ik} \alpha_k + \sum_{k}^{(iS)} a_{ik} \beta_k.
\end{aligned}$$

We claim that deterministically on the event Ξ, there is a constant $C > 0$ such that for any $W, V \subset [n]$ disjoint with $|W \cup V| \leq \log n$, and any $i \notin W \cup V$, we have

$$\left| \sum_{T:\, T \subseteq W} (-1)^{|T|} \left(\mathbf{e}_i^\top R^{(V \cup T)} \mathbf{1} \right) \right| \leq \widetilde{y}(Cxw)^w, \tag{133}$$

where $w = |W| + 1$, $x = (\log n)^{2+2\varepsilon+a}/\sqrt{d}$, and $\widetilde{y} = C\sqrt{d}(\log n)^{-1-\varepsilon}$. We will verify this claim at the end of the proof. Assuming this claim, we apply it above with $V = iS \setminus U$ and either $W = U$ or $W = U \setminus \{k\}$. Then setting $u = |U| + 1 \geq w$, we have on Ξ that

$$|\alpha_k| \leq \widetilde{y}(Cxu)^{|U|}, \qquad |\beta_k| \leq \widetilde{y}(Cxu)^{|U|+1}. \tag{134}$$

Let r be any even integer with $r \leq p \leq (\log n) - 1$. As α_k, β_k are independent of row a_i of A by definition, we have for the partial expectation $\mathbb{E}_i$ over a_i that

$$\begin{aligned}
&\mathbb{E}_i \left[\mathbb{1}\{\Xi\} \left| \mathcal{Z}_i^{S,U} \right|^r \right] \\
&\quad = \mathbb{E}_i \left[\mathbb{1}\{\Xi\} \left| \sum_{k \in U} a_{ik}\alpha_k + \sum_k^{(iS)} a_{ik}\beta_k \right|^r \right] \\
&\quad \leq \mathbb{1}\{|\alpha_k| \leq \widetilde{y}(Cxu)^{|U|} \text{ and } |\beta_k| \leq \widetilde{y}(Cxu)^{|U|+1} \text{ for all } k\} \\
&\qquad \cdot \mathbb{E}_i \left[\left| \sum_{k \in U} a_{ik}\alpha_k + \sum_k^{(iS)} a_{ik}\beta_k \right|^r \right].
\end{aligned}$$

We apply (24) for the conditional expectation $\mathbb{E}_i$, with v having entries $v_k = \alpha_k$ for $k \in U$, $v_k = \beta_k$ for $k \notin iS$, and $v_k = 0$ otherwise. Recall that $w \leq |U| \leq |S| \leq \log n$. Since $Cxw \ll 1$ and $|U|(Cxw)^{2|U|} \ll (n - |U|)(Cxw)^{2|U|+2}$ by the definition of x and $d \leq n$, the bounds (134) imply

$$\|v\|_\infty \leq \widetilde{y}(Cxu)^{|U|}, \qquad \|v\|_2 \leq \sqrt{2n} \cdot \widetilde{y}(Cxw)^{|U|+1}.$$

Then for a constant $C' > 0$, (24) gives

$$\mathbb{E}_i \left[\mathbb{1}\{\Xi\} \left| \mathcal{Z}_i^{S,U} \right|^r \right] \leq (C' r \widetilde{y}(Cxu)^u)^r.$$

Then, taking the full expectation and setting $y = C'(\log n)\widetilde{y} \geq C' r \widetilde{y}$ (since $r \leq p \leq \log n$) yields condition (ii).

For condition (iii), we have

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}\{\Xi_i\}\left|\mathcal{Z}_i^{S,U}\right|^2\right]\\
&\quad\le 2^{|U|}\sum_{T:\,T\subseteq U}\mathbb{E}[\mathbb{1}\{\Xi_i\}|\mathcal{Z}_i^{[(S\setminus U)\cup T]}|^2]\\
&\quad= 2^{|U|}\sum_{T:\,T\subseteq U}\sum_{k,k'}^{((iS\setminus U)\cup T)}\mathbb{E}[a_{ik}a_{ik'}]\mathbb{E}\left[\mathbb{1}\{\Xi_i\}(\mathbf{e}_k^\top R^{((iS\setminus U)\cup T)}\mathbf{1})(\mathbf{e}_{k'}^\top R^{((iS\setminus U)\cup T)}\mathbf{1})\right]\\
&\quad= 2^{|U|}\sum_{T:\,T\subseteq U}\sum_{k}^{((iS\setminus U)\cup T)}\mathbb{E}[a_{ik}^2]\mathbb{E}\left[\mathbb{1}\{\Xi_i\}\left|\mathbf{e}_k^\top R^{((iS\setminus U)\cup T)}\mathbf{1}\right|^2\right],
\end{aligned}$$

where the second line applies the independence of a_i and $A^{(i)}$. Note that on Ξ_i, we have $\|A^{(i)}\|\le 2.5$. Then, applying $|U|\le\log n$, the norm bound $\|R^{((iS\setminus U)\cup T)}\|\le(\log n)^a$ on Ξ_i, and $\mathbb{E}[a_{ik}^2]\le C^2/n$, we get (iii). For (iv), we apply the condition $|a_{ik}|\le 1$ by definition of Ξ, together with the bound $\|R^{(iU)}\|\le(\log n)^a$ on Ξ. Finally, (v) holds by the probability bound of $1-e^{-c(\log n)^{1+\varepsilon}}$ established for (33), (34), (35), (22), and in Lemma 3.

It remains to establish the claim (133). For $W=\emptyset$, this follows from (35). Assume then that $w\ge 1$, and write $W=\{j_1,\dots,j_{w-1}\}$ (in any order). For a function $f:\mathbb{R}^{n\times n}\to\mathbb{C}$ and any index $j\in[n]$, define $Q_jf:\mathbb{R}^{n\times n}\to\mathbb{C}$ by

$$(Q_jf)(A)=f(A)-f(A^{(j)}).$$

Note that if f is in fact a function of $A^{(S)}$, i.e., $f(A)=f(A^{(S)})$ for every matrix $A\in\mathbb{R}^{n\times n}$, then $Q_jf(A)=f(A^{(S)})-f(A^{(jS)})$. Fix i and V, and define $f(A)=\mathbf{e}_i^\top R^{(V)}\mathbf{1}$. This satisfies $f(A)=f(A^{(V)})$ for every A. Then by inclusion–exclusion, the quantity to be bounded is equivalently written as

$$\sum_{T:\,T\subseteq W}(-1)^{|T|}\left(\mathbf{e}_i^\top R^{(V\cup T)}\mathbf{1}\right)=\left(Q_{j_{w-1}}\dots Q_{j_2}Q_{j_1}f\right)(A).$$

We apply Schur complement identities to iteratively to expand $Q_{j_{w-1}}\dots Q_{j_1}f$: First applying (110), we get

$$Q_{j_1}f(A)=\mathbf{e}_i^\top R^{(V)}\mathbf{1}-\mathbf{e}_i^\top R^{(j_1V)}\mathbf{1}=R_{ij_1}^{(V)}\cdot\frac{1}{R_{j_1j_1}^{(V)}}\cdot\mathbf{e}_{j_1}^\top R^{(V)}\mathbf{1}.$$

Then applying (110), (111), and (112) to the three factors on the right side above, and using the identity

$$xyz-\widetilde{x}\widetilde{y}\widetilde{z}=xy(z-\widetilde{z})+x(\widetilde{y}-y)\widetilde{z}+(\widetilde{x}-x)\widetilde{y}\widetilde{z},$$

we get

$$Q_{j_2} Q_{j_1} f(A) = R^{(V)}_{ij_1} \cdot \frac{1}{R^{(V)}_{j_1 j_1}} \cdot \left(\frac{R^{(V)}_{j_1 j_2}}{R^{(V)}_{j_2 j_2}} \cdot \mathbf{e}^\top_{j_2} R^{(V)} \mathbf{1} \right) + R^{(V)}_{ij_1} \cdot \left(- \frac{\left(R^{(V)}_{j_1 j_2}\right)^2}{R^{(j_2 V)}_{j_1 j_1} R^{(V)}_{j_2 j_2} R^{(V)}_{j_1 j_1}} \right) \cdot \mathbf{e}^\top_{j_1} R^{(j_2 V)} \mathbf{1} + \frac{R^{(V)}_{ij_2} R^{(V)}_{j_2 j_1}}{R^{(V)}_{j_2 j_2}} \cdot \frac{1}{R^{(j_2 V)}_{j_1 j_1}} \cdot \mathbf{e}^\top_{j_1} R^{(j_2 V)} \mathbf{1}.$$

Applying (112), (111), and (110) to each factor of each summand above, and repeating iteratively, an induction argument verifies the following claims for each $t \in \{1, \ldots, w-1\}$:

- $Q_{j_t} \ldots Q_{j_1} f(A)$ is a sum of at most $\prod_{s=1}^{t-1} 4s$ summands (with the convention $\prod_{s=1}^{0} 4s = 1$), where
- Each summand is a product of at most $4t$ factors, where
- Each factor is one of the following three forms, for a set $S \subseteq V \cup W$: $R^{(S)}_{jk}$ for $j, k \notin S$ distinct, or $1/R^{(S)}_{jj}$ for $j \notin S$, or $\mathbf{e}^\top_j R^{(S)} \mathbf{1}$ for $j \notin S$. Furthermore,
- Each summand of $Q_{j_t} \ldots Q_{j_1} f(A)$ satisfies: (a) It has exactly one factor of the form $\mathbf{e}^\top_j R^{(S)} \mathbf{1}$. (b) The number of factors of the form $1/R^{(S)}_{jj}$ is less than or equal to the number of factors of the form $R^{(S)}_{jk}$ for $j \neq k$. (c) There are at least t factors of the form $R^{(S)}_{jk}$ for $j \neq k$.

Finally, we apply this with $t = w - 1$ and use the bound

$$\prod_{s=1}^{t-1} 4s \leq (4w)^w.$$

By Lemma 17, since $|W \cup V| \leq \log n$, we have $|R^{(S)}_{jk}| \leq C(\log n)^{2+2\varepsilon+a}/\sqrt{d}$, $|R^{(S)}_{jj}| \geq |m_0|/2$, and $|\mathbf{e}^\top_j R^{(S)} \mathbf{1}| \leq C(\log n)^{1+\varepsilon+a}$ on the event Ξ. Thus, we get

$$|Q_{j_{w-1}} \ldots Q_{j_1} f(A)| \leq (4w)^w \cdot \left(\frac{C(\log n)^{2+2\varepsilon+a}}{\sqrt{d}} \right)^{w-1} \cdot C(\log n)^{1+\varepsilon+a} \leq \widetilde{y}(C'xw)^w$$

for $x = (\log n)^{2+2\varepsilon+a}/\sqrt{d}$ and $\widetilde{y} = C\sqrt{d}(\log n)^{-1-\varepsilon}$, as claimed. □

We now show (36) holds for $z \in D$ with probability $1 - e^{-c(\log n)(\log \log n)}$. The diagonal bound (34) implies

$$|\operatorname{Tr} R - n \cdot m_0| \leq \frac{Cn(\log n)^{2+2\varepsilon+3a/2}}{\sqrt{d}}. \tag{135}$$

To bound the sum of off-diagonal elements of R, we apply (109) to write

$$\sum_{i \neq k} R_{ik} = -\sum_i R_{ii}\mathcal{Z}_i = -m_0\sum_i \mathcal{Z}_i - \sum_i (R_{ii} - m_0)\mathcal{Z}_i. \tag{136}$$

Applying (34) and (126) yields

$$\sum_i |(R_{ii} - m_0)\mathcal{Z}_i| \leq \frac{Cn(\log n)^{3+3\varepsilon+5a/2}}{\sqrt{d}}. \tag{137}$$

Then applying Lemma 16 with x, y, Ξ as defined in Lemma 18 and with p being the largest even integer less than $(\log n) - 1$, we have

$$\mathbb{1}\{\Xi\}\left|n^{-1}\sum_i \mathcal{Z}_i\right| \leq C(\log n)^{12} \cdot \sqrt{d}(\log n)^{-\varepsilon} \cdot (\log n)^{4+4\varepsilon+2a}/d \leq \frac{C(\log n)^{16+3\varepsilon+2a}}{\sqrt{d}} \tag{138}$$

with probability $1 - e^{-c(\log n)(\log\log n)}$. Since $\mathbf{1}^\top R\mathbf{1} = \operatorname{Tr} R + \sum_{i \neq k} R_{ik}$, multiplying (138) by $n \cdot m_0$ and combining with (135)–(137) yields (36).

References

1. Y. Aflalo, A. Bronstein, and R. Kimmel. On convex relaxation of graph isomorphism. *Proceedings of the National Academy of Sciences*, 112(10):2942–2947, 2015.
2. B. Barak, C.-N. Chou, Z. Lei, T. Schramm, and Y. Sheng. (Nearly) efficient algorithms for the graph matching problem on correlated random graphs. *arXiv preprint* arXiv:1805.02349, 2018.
3. B. Barak, S. B. Hopkins, J. A. Kelner, P. Kothari, A. Moitra, and A. Potechin. A Nearly tight sum-of-squares lower bound for the planted clique problem. In *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS*, pp 428–437, 2016.
4. P. Biane. On the free convolution with a semi-circular distribution. *Indiana University Mathematics Journal*, 46(3):705–718, 1997.
5. R. E. Burkard, E. Cela, P. M. Pardalos, and L. S. Pitsoulis. The quadratic assignment problem. In *Handbook of combinatorial optimization*, vol. 3 (D.-Z. Du, P.M. Pardalos, eds.). Kluwer Academic Publishers, 1998, pp. 241–337.
6. Y. Chen and J. Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. In *Proceedings of ICML 2014* (Also arXiv:1402.1267), Feb 2014.
7. D. Cullina and N. Kiyavash. Improved achievability and converse bounds for Erdös-Rényi graph matching. In *Proceedings of the 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science*, pp 63–72. ACM, 2016.
8. D. Cullina and N. Kiyavash. Exact alignment recovery for correlated Erdös-Rényi graphs. *arXiv preprint* arXiv:1711.06783, 2017.
9. O. E. Dai, D. Cullina, N. Kiyavash, and M. Grossglauser. On the performance of a canonical labeling for matching correlated Erdős-Rényi graphs. *arXiv preprint* arXiv:1804.09758, 2018.
10. J. Ding, Z. Ma, Y. Wu, and J. Xu. Efficient random graph matching via degree profiles. *Probability Theory and Related Fields*, pp 1–87, Sep 2020.
11. L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. The local semicircle law for a general class of random matrices. *Electron. J. Probab*, 18(59):1–58, 2013.
12. L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Spectral statistics of Erdős–Rényi graphs I: local semicircle law. *The Annals of Probability*, 41(3B):2279–2375, 2013.

13. L. Erdős, H.-T. Yau, and J. Yin. Bulk universality for generalized Wigner matrices. *Probability Theory and Related Fields*, 154(1-2):341–407, 2012.
14. Z. Fan, C. Mao, Y. Wu, and J. Xu. Spectral graph matching and regularized quadratic relaxations I: The Gaussian model. *arxiv preprint* arXiv:1907.08880, 2019.
15. Z. Fan, C. Mao, Y. Wu, and J. Xu. Spectral graph matching and regularized quadratic relaxations: Algorithm and theory. In *International Conference on Machine Learning (ICML)*, Jul 2020.
16. D. L. Hanson and F. T. Wright. A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Statist.*, 42:1079–1083, 1971.
17. K. Makarychev, R. Manokaran, and M. Sviridenko. Maximum quadratic assignment problem: Reduction from maximum label cover and LP-based approximation algorithm. *Automata, Languages and Programming*, pp 594–604, 2010.
18. P. M. Pardalos, F. Rendl, and H. Wolkowicz. The quadratic assignment problem: A survey and recent developments. In *In Proceedings of the DIMACS Workshop on Quadratic Assignment Problems, volume 16 of DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pp 1–42. American Mathematical Society, 1994.
19. P. Pedarsani and M. Grossglauser. On the privacy of anonymized networks. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp 1235–1243, 2011.
20. M. Rudelson and R. Vershynin. Hanson-Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.*, 18(82):9, 2013.
21. M. Zaslavskiy, F. Bach, and J.-P. Vert. A path following algorithm for the graph matching problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12):2227–2242, 2008.