Supplement to "Spectral Graph Matching and Regularized Quadratic Relaxations"

Zhou Fan, Cheng Mao, Yihong Wu, and Jiaming Xu

We present the proofs of our main results and various claims in this supplement.

A Equivalence of GRAMPA and regularized quadratic relaxations

We first establish the equivalence of similarity matrices defined by (3), (11) and (12), as claimed in Section 1.3.

Lemma A.1. *The similarity matrix $\widehat{X}$ defined by* (3) *is the minimizer of the unconstrained program (11), and $\alpha\widehat{X}$ is the minimizer of the constrained program (12) for some (random) scalar multiplier $\alpha > 0$.*

Proof. To show that $\widehat{X}$ solves (11), note that the objective function in (11) is quadratic, with first order optimality condition

$$A^2X + XB^2 - 2AXB + \eta^2 X = \mathbf{J}.$$

Setting $\mathbf{x} = \mathsf{vec}(X)$ and writing this in vectorized form with Kronecker products

$$\big[(\mathbf{I}_n \otimes A - B \otimes \mathbf{I}_n)^2 + \eta^2 \mathbf{I}_{n^2}\big]\mathbf{x} = \mathbf{1}_{n^2},$$

we see that the vectorized solution to (11) is

$$\widehat{\mathbf{x}} = \big[(\mathbf{I}_n \otimes A - B \otimes \mathbf{I}_n)^2 + \eta^2 \mathbf{I}_{n^2}\big]^{-1}\mathbf{1}_{n^2} \in \mathbb{R}^{n^2}.$$

Applying the spectral decomposition (2), we get

$$\widehat{\mathbf{x}} = \sum_{ij} \frac{1}{(\lambda_i - \mu_j)^2 + \eta^2}(v_j \otimes u_i)(v_j \otimes u_i)^\top \mathbf{1}_{n^2} = \sum_{ij} \frac{u_i^\top \mathbf{J}_n v_j}{(\lambda_i - \mu_j)^2 + \eta^2}\mathsf{vec}(u_i v_j^\top), \tag{A.1}$$

which is exactly the vectorization of $\widehat{X}$ in (3).

Recall that $\widetilde{X}$ denotes the minimizer of (12). Introducing a Lagrange multiplier $2\alpha \in \mathbb{R}$ for the constraint, the first-order stationarity condition is

$$A^2X + XB^2 - 2AXB + \eta^2 X = \alpha\mathbf{J},$$

and hence $\widetilde{X} = \alpha\widehat{X}$. To find α, note that $\mathbf{1}^\top\widetilde{X}\mathbf{1} = \alpha\mathbf{1}^\top\widehat{X}\mathbf{1} = n$. Furthermore, from (3) we have

$$\mathbf{1}^\top\widehat{X}\mathbf{1} = \sum_{ij} \frac{\langle u_i, \mathbf{1}\rangle^2\langle v_j, \mathbf{1}\rangle^2}{(\lambda_i - \mu_j)^2 + \eta^2} > 0.$$

Hence $\alpha > 0$. These claims together establish the lemma. □

B Resolvent representation

As noted in the Section 2.3, the proof of Theorem 2.2 hinges on a resolvent representation of the similarity matrix $\widehat{X}$, which we prove in this section. To ease the notation, we let $X \triangleq \eta\widehat{X}$ throughout the proofs.

Recall that for a real symmetric matrix A with spectral decomposition (2), its resolvent is

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

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 B.1 (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

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

□

Proposition B.2. *Consider symmetric matrices A and B with spectral decompositions* (2)*, and suppose that $\|A\| \le 2.5$. Then the matrix $X \triangleq \eta\widehat{X}$, where $\widehat{X}$ is defined in (3), admits the following representation*

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

where

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

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

Proof. We have

$$\begin{aligned} X &= \eta \sum_{i,j} u_i u_i^\top \mathbf{J} \frac{v_j v_j^\top}{(\lambda_i - \mu_j)^2 + \eta^2} \\ &= \eta \sum_i u_i u_i^\top \mathbf{J} R_B(\lambda_i + \mathbf{i}\eta) R_B(\lambda_i - \mathbf{i}\eta) \\ &= \operatorname{Im} \sum_i u_i u_i^\top \mathbf{J} R_B(\lambda_i + \mathbf{i}\eta) \end{aligned} \tag{B.3}$$

by Lemma B.1. 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

$$-\frac{1}{2\pi\mathbf{i}} \oint_\Gamma \frac{f(z)}{\lambda_i - z} dz = f(\lambda_i). \tag{B.4}$$

Substituting this into (B.3), we obtain

$$X = \operatorname{Im} \sum_i u_i u_i^\top \left(-\frac{1}{2\pi \mathbf{i}} \oint_\Gamma \frac{f(z)}{\lambda_i - z} dz \right) = \frac{1}{2\pi} \operatorname{Re} \oint_\Gamma R_A(z) f(z) dz, \tag{B.5}$$

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

C Tools from random matrix theory

Before proving our main results, we introduce some useful tools from random matrix theory. In particular, the resolvent bounds in Theorem C.6 constitute an important technical ingredient in our analysis.

C.1 Concentration inequalities

We start with some known concentration inequalities in the literature.

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

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

Proof. See [EKYY13b, Lemma 4.3], where we fix the parameter $\xi = 1+\varepsilon$ in [EKYY13b, Eq. (2.4)]. The notational identification is $q \equiv \sqrt{d}$. □

Lemma C.2 (Hanson-Wright inequality)**.** *Let z be a sub-Gaussian vector in $\mathbb{R}^n$, and let M be a fixed matrix in $\mathbb{C}^{n\times n}$. Then we have with probability at least $1-\delta$ that*

$$|z^\top M z - \operatorname{Tr} M| \leq 2C\|z\|_{\psi_2}^2 \|M\|_F \log(1/\delta),$$

where C is a universal constant and $\|z\|_{\psi_2}$ is the sub-Gaussian norm of z.

Proof. See [RV13, Section 3.1] for this complex-valued version of the Hanson-Wright inequality. □

Lemma C.3 (Concentration inequalities)**.** *Let $\alpha, \beta \in \mathbb{R}^n$ be independent random vectors with independent entries, satisfying*

$$\mathbb{E}[\alpha_i] = \mathbb{E}[\beta_i] = 0, \qquad \mathbb{E}[\alpha_i^2] = \mathbb{E}[\beta_i^2] = \frac{1}{n},$$

$$\max(\mathbb{E}[|\alpha_i|^k], \mathbb{E}[|\beta_i|^k]) \leq \frac{1}{nd^{(k-2)/2}}, \qquad \textit{for each } k \in [2, (\log n)^{10 \log\log n}]. \tag{C.1}$$

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

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

$$|\alpha_i| \leq \frac{C}{\sqrt{d}}. \tag{C.2}$$

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

$$\left| v^\top \alpha \right| \le (\log n)^{1+\varepsilon} \left(\frac{\|v\|_\infty}{\sqrt{d}} + \frac{\|v\|_2}{\sqrt{n}} \right). \tag{C.3}$$

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

$$\mathbb{E}\left[\left| v^\top \alpha \right|^p \right] \le (Cp)^p \left(\frac{\|v\|_\infty}{\sqrt{d}} + \frac{\|v\|_2}{\sqrt{n}} \right)^p. \tag{C.4}$$

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

$$\left| \alpha^\top M \alpha - \frac{1}{n} \operatorname{Tr} M \right| \le (\log n)^{2+2\varepsilon} \left(\frac{2\|M\|_\infty}{\sqrt{d}} + \frac{\|M\|_F}{n} \right) \tag{C.5}$$

and

$$\left| \alpha^\top M \beta \right| \le (\log n)^{2+2\varepsilon} \left(\frac{2\|M\|_\infty}{\sqrt{d}} + \frac{\|M\|_F}{n} \right). \tag{C.6}$$

Proof. See [EKYY13b, Lemma 3.7, Lemma 3.8, and Lemma A.1(i)], where we fix $\xi = 1 + \varepsilon$. □

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

Lemma C.4 (Concentration of bilinear forms)**.** *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*

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

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

(a) *For any constant* $\varepsilon > 0$*, suppose (C.1) 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* $1 - e^{-c(\log n)^{1+\varepsilon}}$,

$$\left| \alpha^\top M \beta - \frac{1-\sigma^2}{n} \operatorname{Tr} M \right| \le C\, (\log n)^{2+2\varepsilon} \left(\frac{1}{n}\|M\|_F + \frac{1}{\sqrt{d}}\|M\|_\infty \right). \tag{C.7}$$

(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* $1 - n^{-D}$,

$$\left| \alpha^\top M \beta - \frac{1-\sigma^2}{n} \operatorname{Tr} M \right| \le \frac{C \log n}{n} \|M\|_F. \tag{C.8}$$

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

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

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 (C.7), we obtain from (C.5) 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 (C.7) easily follows.

The sub-Gaussian concentration bound (C.8) follows from the Hanson-Wright inequality [HW71, RV13]. 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 Lemma C.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. □

C.2 The Stieltjes transform

Denote the semicircle density and its Stieltjes transform by

$$\rho(x) = \frac{1}{2\pi}\sqrt{4-x^2}\,\mathbf{1}_{\{|x|\le 2\}} \quad \text{and} \quad m_0(z) = \int \frac{1}{x-z}\rho(x)dx = \frac{-z+\sqrt{z^2-4}}{2} \tag{C.9}$$

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 C.5. *For each $z \in \mathbb{C}\setminus\mathbb{R}$, the Stieltjes transform $m_0(z)$ is the unique value satisfying*

$$m_0(z)^2 + zm_0(z) + 1 = 0 \quad \text{and} \quad \operatorname{Im} m_0(z)\cdot\operatorname{Im} z > 0. \tag{C.10}$$

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

$$|m_0(z)| \asymp 1, \ \ |\operatorname{Im} m_0(z)| \gtrsim |\operatorname{Im} z|, \quad \text{and} \quad |\operatorname{Im} m_0(z)| \asymp \begin{cases} \sqrt{\zeta(z) + |\operatorname{Im} z|} & \text{if } |\operatorname{Re} z| \le 2, \\ |\operatorname{Im} z|/\sqrt{\zeta(z)+|\operatorname{Im} z|} & \text{if } |\operatorname{Re} z| > 2. \end{cases} \tag{C.11}$$

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

$$m_0^\pm(x)^2 + xm_0^\pm(x) + 1 = 0, \ \ m_0^+(x) = \overline{m_0^-(x)}, \ \ \frac{1}{\pi}\operatorname{Im} m_0^+(x) = -\frac{1}{\pi}\operatorname{Im} m_0^-(x) = \rho(x), \ \ |m_0^\pm(x)| = 1. \tag{C.12}$$

Proof. (C.10) follows from the definition of m_0. (C.11) follows from [EKYY13a, 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 [Bia97, Corollary 1]. The first claim of (C.12) follows from continuity and (C.10), the second from conjugate symmetry, the third from the Stieltjes inversion formula, and the last from the fact that the two roots of (C.10) at $z = x \in [-2, 2]$ are $m_0^+(x)$ and $m_0^-(x) = \overline{m_0^+(x)}$, so that $1 = m_0^\pm(x)\overline{m_0^\pm(x)} = |m_0^\pm(x)|^2$. □

C.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

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

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 C.6 (Resolvent bounds)**.** *Suppose $A \in \mathbb{R}^{n\times n}$ has independent entries $(a_{ij})_{i\leq j}$ satisfying (13) and (14). 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 $n \geq d \geq (\log n)^{b'}$. Then for some constants $C, c, n_0 > 0$ depending on a and ε, and for all $n \geq n_0$, with probability $1 - e^{-c(\log n)(\log\log n)}$, the following hold simultaneously for every $z \in D$:

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

$$|R_{jk}(z)| \leq \frac{C(\log n)^{2+2\varepsilon+a}}{\sqrt{d}}. \tag{C.13}$$

For all $j \in [n]$,

$$|R_{jj}(z) - m_0(z)| \leq \frac{C(\log n)^{2+2\varepsilon+3a/2}}{\sqrt{d}}. \tag{C.14}$$

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

$$\left|\mathbf{e}_j^\top R(z)\mathbf{1}\right| \leq C(\log n)^{1+\varepsilon+a}. \tag{C.15}$$

(c) (Total sum bound)

$$|\mathbf{1}^\top R(z)\mathbf{1} - n \cdot m_0(z)| \leq \frac{Cn(\log n)^b}{\sqrt{d}}. \tag{C.16}$$

The proof follows ideas of [EKYY13b], and we defer this to Section E. 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 [EKYY13b], 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 $|\operatorname{Im} z| \asymp \operatorname{polylog}(n)/n$.

D Proofs of guarantees for the correlated Wigner model

Our main result, Theorem 2.2, is an immediate consequence of the following theorem.

Theorem D.1. *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 $c_0 > \max(32 + 4a, 4 + 7a)$. Then there exist (a, 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 the following holds: For all $n \geq n_0$, with probability at least $1 - n^{-10}$, the matrix $X \triangleq \eta \widehat{X}$, where $\widehat{X}$ is defined in (3), satisfies*

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

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 $\kappa = 1$.

Proof of Theorem 2.2. Let $c = 1/(64C^2)$ and $c' = 1/(2C)$, where C is the constant given in Theorem D.1. Then under assumption (17), we have

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

so $\max_{\ell \neq \pi_*(k)} |X_{k\ell}| \leq 1/(8\eta)$. We also have $C\sigma/\eta \leq Cc' = 1/2$ and $1 - \sigma^2 > 7/8$ and $Cr(n) < 1/8$ 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 (18). □

In the rest of this section, we prove Theorem D.1, following the outline presented in Section 2.3.

Note that the mapping $B \mapsto \Pi_*^\top B \Pi_*$ for any permutation Π_* induces $v_j \mapsto \Pi_*^\top v_j$ 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 $\min_k X_{kk} > \max_{k \neq \ell} X_{k\ell}$.

In view of Lemma C.1, 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 B.2, we get that

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

Note that one may attempt to directly apply (C.15) 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 (D.2) 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.

D.1 Decomposition via Schur complement

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

Lemma D.2 (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

$$M^{-1} = \begin{bmatrix} S & -SBD^{-1} \\ -D^{-1}CS & D^{-1} + D^{-1}CSBD^{-1} \end{bmatrix} \tag{D.3}$$

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 sub-matrix 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 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 D.3. *Suppose $|z| \le 10$, and*

$$\|R_{A,12}(z) - m_0(z)\mathbf{I}\| \le \delta \tag{D.4}$$

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

$$\left|\mathbf{e}_k^\top R_A(z)\mathbf{1} - m_0(z)\left(1 - a_k^\top R_A^{(12)}(z)\mathbf{1}_{n-2}\right)\right| \le C\delta\left(1 + \|R_A(z)\mathbf{1}\|_\infty\right). \tag{D.5}$$

Proof. It suffices to consider $k = 1$. Applying the Schur complement identity (D.3), the first two rows of R_A are given by

$$\begin{bmatrix} R_{A,12} & -R_{A,12}A_o^\top R_A^{(12)} \end{bmatrix}. \tag{D.6}$$

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

$$\begin{aligned} \mathbf{e}_1^\top R_A(z)\mathbf{1} &= \begin{bmatrix} 1 & 0 \end{bmatrix}(m_0(z)\mathbf{I} + \Delta_A)\left(\mathbf{1}_2 - A_o^\top R_A^{(12)}\mathbf{1}_{n-2}\right). \\ &= m_0(z)\left(1 - a_1^\top R_A^{(12)}\mathbf{1}_{n-2}\right) + \begin{bmatrix} 1 & 0 \end{bmatrix}\Delta_A\left(\mathbf{1}_2 - A_o^\top R_A^{(12)}\mathbf{1}_{n-2}\right). \\ &= m_0(z)\left(1 - a_1^\top R_A^{(12)}\mathbf{1}_{n-2}\right) + O\left(\delta\left(1 + \left\|A_o^\top R_A^{(12)}\mathbf{1}_{n-2}\right\|\right)\right), \end{aligned} \tag{D.7}$$

where the last equality applies (D.4). 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 (D.4) implies that $R_{A,12}$ is invertible with $\|R_{A,12}^{-1}\| \lesssim 1$. Using (D.6) again, we have

$$A_o^\top R_A^{(12)} \mathbf{1}_{n-2} = \mathbf{1}_2 - R_{A,12}^{-1} \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix}^\top R_A \mathbf{1}_n. \tag{D.8}$$

It follows that

$$\left\| A_o^\top R_A^{(12)} \mathbf{1}_{n-2} \right\| \lesssim 1 + \left| \mathbf{e}_1^\top R_A \mathbf{1}_n \right| + \left| \mathbf{e}_2^\top R_A \mathbf{1}_n \right| \lesssim 1 + \|R_A \mathbf{1}_n\|_\infty. \tag{D.9}$$

The desired bound (D.5) follows by combining (D.7) and (D.9). □

D.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 D.1, and for some small constant $\varepsilon > 0$, let b, b' be as defined in Theorem C.6. Under the given condition for c_0 in Theorem D.1, for $\varepsilon > 0$ sufficiently small, we have $c_0 > b'$ and $c_0 > 2b$—thus $d \gg (\log n)^{b'}$ so Theorem C.6 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

$$\delta_1 = \frac{(\log n)^{2+2\varepsilon+3a/2}}{\sqrt{d}}, \quad \delta_2 = \frac{(\log n)^{1+\varepsilon+a}}{\sqrt{n}}, \quad \delta_3 = \frac{(\log n)^b}{\sqrt{d}}, \quad \delta_4 = \frac{(\log n)^{\kappa/2}}{\sqrt{n}}. \tag{D.10}$$

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

D.2.1 Resolvent approximation

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

$$\|R_{A,12}(z) - m_0(z)\mathbf{I}\| \lesssim \delta_1 \tag{D.11}$$

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

$$\|R_A(z)\mathbf{1}\|_\infty \lesssim \delta_2 \sqrt{n} \tag{D.13}$$

$$\|R_B(z + \mathbf{i}\eta)\mathbf{1}\|_\infty \lesssim \delta_2 \sqrt{n}. \tag{D.14}$$

Applying the resolvent approximations given in Theorem C.6, we have that

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

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

On $\mathcal{E}_1$, by Lemma D.3, we get that uniformly over $z \in \Gamma$,

$$\mathbf{e}_1^\top R_A(z) \mathbf{1} = m_0(z) \left(1 - a_1^\top R_A^{(12)} \mathbf{1}_{n-2} \right) + O\left(\delta_1 \delta_2 \sqrt{n} \right), \tag{D.15}$$

$$\mathbf{e}_2^\top R_B(z + \mathbf{i}\eta) \mathbf{1} = m_0(z + \mathbf{i}\eta) \left(1 - b_2^\top R_B^{(12)} \mathbf{1}_{n-2} \right) + O\left(\delta_1 \delta_2 \sqrt{n} \right). \tag{D.16}$$

Each of (D.15) and (D.16) is itself $O(\delta_2\sqrt{n})$, by (D.13) and (D.14). Then multiplying the two, we have

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

It follows that

$$\begin{aligned}&\oint_\Gamma \left[\mathbf{e}_1^\top R_A(z)\mathbf{1}\right]\left[\mathbf{e}_2^\top R_B(z+\mathbf{i}\eta)\mathbf{1}\right]dz\\&=\oint_\Gamma m_0(z)m_0(z+\mathbf{i}\eta)dz - a_1^\top g - b_2^\top h + a_1^\top M b_2 + O\left(\delta_1\delta_2^2 n\right),\end{aligned}\tag{D.17}$$

where

$$\begin{aligned}g &\triangleq \oint_\Gamma m_0(z)m_0(z+\mathbf{i}\eta)R_A^{(12)}(z)\mathbf{1}_{n-2}dz,\\h &\triangleq \oint_\Gamma m_0(z)m_0(z+\mathbf{i}\eta)R_B^{(12)}(z+\mathbf{i}\eta)\mathbf{1}_{n-2}dz,\\M &\triangleq \oint_\Gamma m_0(z)m_0(z+\mathbf{i}\eta)R_A^{(12)}(z)\mathbf{J}_{n-2}R_B^{(12)}(z+\mathbf{i}\eta)dz.\end{aligned}\tag{D.18}$$

D.2.2 Term-by-term analysis

Next, we bound the individual terms of (D.17). By the boundedness of $m_0(z)$, we have

$$\oint_\Gamma m_0(z)m_0(z+\mathbf{i}\eta)dz = O(1).\tag{D.19}$$

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

$$\left|a_1^\top g\right| + \left|b_2^\top h\right| \lesssim \delta_1\left(\|g\|_\infty + \|h\|_\infty\right) + \delta_4\left(\|g\|_2+\|h\|_2\right)\tag{D.20}$$

$$\left|a_1^\top M b_2\right| \lesssim \delta_1\|M\|_\infty + \delta_4^2\|M\|_F.\tag{D.21}$$

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 (C.3) and (C.6), we get that

$$\mathbb{P}\left\{\mathcal{E}_2\right\} \geq 1-n^{-D}$$

for any constant $D>0$,[1] and all $n\geq n_0(D)$, 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.

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. D.1. Define the event $\mathcal{E}_3$ wherein the

[1] The constant D can be made arbitrarily large by setting the hidden constants in (D.20) and (D.21) sufficiently large.

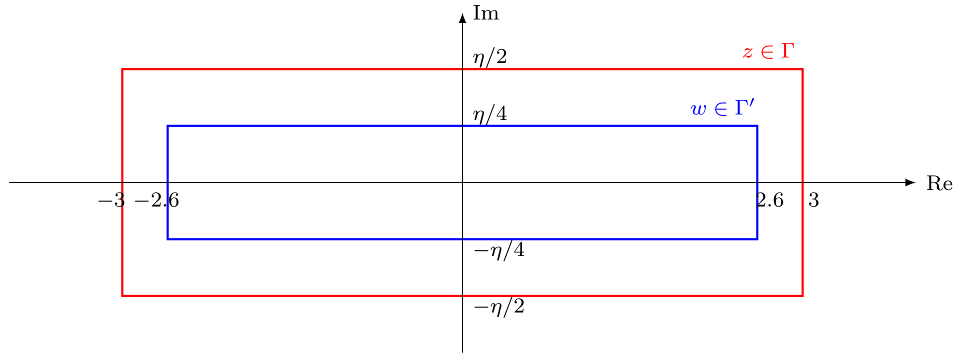

Figure D.1: Nested contours Γ and Γ'.

following hold simultaneously for all $z \in \Gamma \cup \Gamma'$:

$$\left\| R_A^{(12)}(z)\mathbf{1}_{n-2} \right\|_\infty \lesssim \delta_2\sqrt{n}, \tag{D.22}$$

$$\left\| R_B^{(12)}(z+\mathbf{i}\eta)\mathbf{1}_{n-2} \right\|_\infty \lesssim \delta_2\sqrt{n}, \tag{D.23}$$

$$\left| \mathbf{1}_{n-2}^\top R_A^{(12)}(z)\mathbf{1}_{n-2} - m_0(z)(n-2) \right| \lesssim \delta_3 n, \tag{D.24}$$

$$\left| \mathbf{1}_{n-2}^\top R_B^{(12)}(z+\mathbf{i}\eta)\mathbf{1}_{n-2} - m_0(z+\mathbf{i}\eta)(n-2) \right| \lesssim \delta_3 n. \tag{D.25}$$

By Theorem C.6, 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

$$\|g\|_\infty \lesssim \sup_{z\in\Gamma} \|R_A^{(12)}(z)\mathbf{1}_{n-2}\|_\infty \lesssim \delta_2\sqrt{n}, \tag{D.26}$$

where the second inequality holds in view of (D.22). Similarly, in view of (D.23), we have that $\|h\|_\infty \lesssim \delta_2\sqrt{n}$. Furthermore,

$$\|M\|_\infty \lesssim \sup_{z\in\Gamma} \left\| R_A^{(12)}(z)\mathbf{J}_{n-2}R_B^{(12)}(z+\mathbf{i}\eta) \right\|_\infty \leq \sup_{z\in\Gamma} \left\| R_A^{(12)}(z)\mathbf{1}_{n-2} \right\|_\infty \left\| \mathbf{1}_{n-2}^\top R_B^{(12)}(z+\mathbf{i}\eta) \right\|_\infty \lesssim \delta_2^2 n. \tag{D.27}$$

The ℓ_2 bounds of g, h and M are deferred to Lemma D.4 below. Applying (D.24), (D.25), and Lemma D.4 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 (D.20), (D.21), and (D.19), and plugging into (D.17), we conclude that on the event $\{\|A\| \leq 2.5\} \cap \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$,

$$\begin{aligned} |X_{12}| &= 2\pi \left| \oint_\Gamma [\mathbf{e}_1^\top R_A(z)\mathbf{1}][\mathbf{e}_2^\top R_B(z+\mathbf{i}\eta)\mathbf{1}]dz \right| \\ &\lesssim 1 + \delta_4\sqrt{n\log\frac{1}{\eta}} + \delta_4^2 n\frac{1}{\sqrt{\eta}} + \delta_1\delta_2^2 n \lesssim \delta_4^2 n \frac{1}{\sqrt{\eta}} = (\log n)^\kappa \frac{1}{\sqrt{\eta}}, \end{aligned} \tag{D.28}$$

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

D.2.3 Bounding the norms of g, h and M

Lemma D.4. *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

$$\begin{aligned} \|g\|^2 &\overset{(a)}{=} \oint_\Gamma dz \oint_{\Gamma'} dw \; m_0(z)m_0(z+\mathbf{i}\eta)m_0(\bar w)m_0(\bar w - \mathbf{i}\eta)\mathbf{1}^\top R_A(\bar w) R_A(z)\mathbf{1} \\ &\overset{(b)}{=} -\oint_\Gamma dz \oint_{\Gamma'} dw \; m_0(z)m_0(z+\mathbf{i}\eta)m_0(w)m_0(w - \mathbf{i}\eta)\mathbf{1}^\top R_A(w) R_A(z)\mathbf{1} \\ &\overset{(c)}{=} -\oint_\Gamma dz \oint_{\Gamma'} dw \; m_0(z)m_0(z+\mathbf{i}\eta)m_0(w)m_0(w - \mathbf{i}\eta)\mathbf{1}^\top \frac{R_A(z) - R_A(w)}{z-w}\mathbf{1} \\ &\overset{(d)}{\lesssim} n \oint_\Gamma dz \oint_{\Gamma'} \frac{1}{|z-w|} \end{aligned} \tag{D.29}$$

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

$$R_A(z)R_A(w) = (A-z)^{-1}(A-w)^{-1} = \frac{1}{z-w}[(A-z)^{-1} - (A-w)^{-1}] = \frac{1}{z-w}[R_A(z) - R_A(w)] \tag{D.30}$$

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

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

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 (D.29),

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

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

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

□

D.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 (D.15) and (D.16), 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

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

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

$$\begin{aligned} &\oint_\Gamma \left[\mathbf{e}_1^\top R_A(z)\mathbf{1}\right]\left[\mathbf{1}^\top R_B(z+\mathbf{i}\eta)\mathbf{e}_1\right] dz \\ &= \oint_\Gamma m_0(z)m_0(z+\mathbf{i}\eta)dz - a_1^\top g - b_1^\top h + a_1^\top M b_1 + O\left(\delta_1\delta_2^2 n\right), \end{aligned} \tag{D.31}$$

where

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

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 (C.6), applying Lemma C.4 to control $a_1^\top M b_1$ 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,

$$\left|X_{11} - \frac{1-\sigma^2}{2\pi}\operatorname{Re}\frac{\operatorname{Tr} M}{n}\right| \lesssim (\log n)^\kappa \frac{1}{\sqrt{\eta}}. \tag{D.32}$$

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

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

By the assumption (16) and Theorem C.6, we have that $\mathbb{P}\{\mathcal{E}_4\}\ge 1-n^{-D}$ for any constant $D>0$ and all $n\ge n_0(D)$.

We defer the analysis of $\operatorname{Tr} M$ to Lemma D.5 and Lemma D.6 below: Assuming $\mathcal{E}_4$ holds and applying Lemma D.5 and Lemma D.6 with R_A, R_B replaced by $R_A^{(1)}, R_B^{(1)}$, respectively, we get

$$\frac{1}{n}\operatorname{Re}\operatorname{Tr}(M) = \frac{2\pi + o_\eta(1)}{\eta} + O\left(\frac{\sigma}{\eta^2} + \frac{\delta_3}{\eta}\right). \tag{D.33}$$

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

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

D.3.1 Analyzing the trace of M

Lemma D.5. *Suppose $\|A\| \le 2.5$ and $\|A-B\| \lesssim \sigma$ and*

$$\left|\mathbf{1}^\top R_A(z)\mathbf{1} - m_0(z)n\right| \lesssim \delta_3 n,$$
$$\left|\mathbf{1}^\top R_B(z+\mathbf{i}\eta)\mathbf{1} - m_0(z+\mathbf{i}\eta)n\right| \lesssim \delta_3 n, \tag{D.34}$$

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

$$\begin{aligned}\operatorname{Tr} M &= \oint_\Gamma dz\ m_0(z)m_0(z+\mathbf{i}\eta)\operatorname{Tr}\left[R_A(z)\mathbf{J}R_B(z+\mathbf{i}\eta)\right]\\ &= \oint_\Gamma dz\ m_0(z)m_0(z+\mathbf{i}\eta)\mathbf{1}^\top R_B(z+\mathbf{i}\eta)R_A(z)\mathbf{1}\\ &= \frac{1}{\mathbf{i}\eta}\oint_\Gamma dz\ m_0(z)m_0(z+\mathbf{i}\eta)\mathbf{1}^\top\left(R_B(z+\mathbf{i}\eta) - R_A(z) - R_B(z+\mathbf{i}\eta)(A-B)R_A(z)\right)\mathbf{1}.\end{aligned} \tag{D.35}$$

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 B.1 and the assumption given in (D.34), 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 (D.34), 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 D.6. *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 C.5, 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):

$$\begin{aligned}
&\oint_\Gamma m_0(z)m_0(z+i\eta)(m_0(z+i\eta)-m_0(z))dz \\
&= \int_2^{-2} m_0^+(x)m_0(x+i\eta)(m_0(x+i\eta)-m_0^+(x))dx + \int_{-2}^{2} m_0^-(x)m_0(x+i\eta)(m_0(x+i\eta)-m_0^-(x))dx,
\end{aligned}$$

where m_0^+ and m_0^- are the limits from $\mathbb{C}^+$ and $\mathbb{C}^-$ defined in Proposition C.5. 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

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

the last two steps applying (C.12). Thus the imaginary part of the integral is $2\pi + o_\eta(1)$ for small η. □

E Proof of resolvent bounds

In this section, we prove Theorem C.6. The entrywise bounds of part (a) are essentially the local semicircle law of [EKYY13b, Theorem 2.8], restricted to the simpler domain $\{z : \operatorname{dist}(z, [-2,2]) \ge (\log n)^{-a}\}$ 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 [EKYY13b, Section 5]. We provide a proof of all three statements using the tools of [EKYY13b].

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 $\|A\| \le 2.5$—we omit these details for brevity.

E.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 E.1 (Schur complement identities)**.** *For any $j \in [n]$,*

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

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

$$R_{jk} = -R_{jj} \sum_{\ell}^{(j)} a_{j\ell} R_{\ell k}^{(j)} = R_{jj} R_{kk}^{(j)} \left(-a_{jk} + \sum_{\ell,m}^{(jk)} a_{j\ell} R_{\ell m}^{(jk)} a_{mk} \right), \tag{E.2}$$

$$\mathbf{e}_k^\top R = \mathbf{e}_k^\top R^{(j)} + \frac{R_{kj}}{R_{jj}} \cdot \mathbf{e}_j^\top R, \tag{E.3}$$

$$\frac{1}{R_{kk}} = \frac{1}{R_{kk}^{(j)}} - \frac{(R_{kj})^2}{R_{kk}^{(j)} R_{jj} R_{kk}}. \tag{E.4}$$

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

$$R_{k\ell} = R_{k\ell}^{(j)} + \frac{R_{kj} R_{j\ell}}{R_{jj}}. \tag{E.5}$$

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 (E.3), see [EKYY13a, Lemma 4.5] and [EYY12, Lemma 4.2]. As for (E.3), it is equivalent to verify that (E.5) holds also for $\ell = j$, which simply follows from $R_{kj}^{(j)} = 0$, due to the block diagonal structure of $R^{(j)}$. □

E.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 (C.13) and (C.14) hold for $z \in D$ w.h.p.

We start with (C.14). Note that the jth row $\{a_{jk} : k \in [n]\}$ is independent of $A^{(j)}$ and hence $R^{(j)}$. Applying (E.1), (C.2), and (C.5) conditional on $A^{(j)}$, w.h.p. for all j,

$$\Big| \frac{1}{R_{jj}} + z + \frac{1}{n} \sum_k^{(j)} R_{kk}^{(j)} \Big| = \Big| a_{jj} - \sum_{k,\ell}^{(j)} a_{jk} R_{k\ell}^{(j)} a_{\ell j} + \frac{1}{n} \sum_k^{(j)} R_{kk}^{(j)} \Big| \leq (\log n)^{2+2\varepsilon} \Big(\frac{1}{\sqrt{d}} + \frac{2\|R^{(j)}\|_\infty}{\sqrt{d}} + \frac{\|R^{(j)}\|_F}{n} \Big).$$

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

$$\left|\frac{1}{R_{jj}} + z + \frac{1}{n}\sum_k^{(j)} R_{kk}^{(j)}\right| \lesssim \frac{(\log n)^{2+2\varepsilon+a}}{\sqrt{d}}. \tag{E.6}$$

Since $|z| \le 10$, $|R_{kk}^{(j)}| \le (\log n)^a$, 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

$$\left|m_n - \frac{1}{n}\sum_k^{(j)} R_{kk}^{(j)}\right| = \left|\frac{1}{n}R_{jj} + \frac{1}{n}\sum_k^{(j)}(R_{kk} - R_{kk}^{(j)})\right| \overset{(E.5)}{=} \left|\frac{1}{n}\sum_k \frac{R_{kj}^2}{R_{jj}}\right| = \frac{\|\mathbf{e}_j^\top R\|^2}{n|R_{jj}|} \le \frac{\|R\|^2}{n|R_{jj}|} \lesssim \frac{(\log n)^{3a}}{n}.$$

Using $d \le n$ and combining with (E.6), w.h.p. for all j,

$$\left|\frac{1}{R_{jj}} + z + m_n\right| \lesssim \frac{(\log n)^{2+2\varepsilon+a}}{\sqrt{d}}. \tag{E.7}$$

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

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

so

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

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

$$\left|\frac{1}{R_{jj}} - \frac{1}{m_n}\right| = \frac{|R_{jj} - m_n|}{|R_{jj}||m_n|} \le \max_k \frac{|R_{jj}-R_{kk}|}{|R_{jj}||m_n|} \le \max_k \frac{2|R_{jj}-R_{kk}|}{|R_{jj}||R_{kk}|} = 2\max_k\left|\frac{1}{R_{jj}} - \frac{1}{R_{kk}}\right|,$$

so

$$\left|\frac{1}{R_{jj}} - \frac{1}{m_n}\right| \lesssim \frac{(\log n)^{2+2\varepsilon+a}}{\sqrt{d}}. \tag{E.8}$$

Combining with (E.7), 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 $|z| \ge (\log n)^{-a}$ 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 equation (C.10), we obtain by a Taylor expansion of the square-root that

$$\min(|m_n - m_0|, |m_n - \widetilde{m}_0|) \lesssim |r_n| \left(1 + \frac{|z|}{\sqrt{|z^2-4|}}\right) \lesssim \frac{|r_n|}{\sqrt{\zeta(z) + |\operatorname{Im} z|}}, \tag{E.9}$$

where $\zeta(z)$ is as defined in Proposition C.5.

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 (C.11) 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 $|r_n| \ll (\log n)^{-a}$, (E.9) 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 (E.9) holds for $|m_n(z') - m_0(z')|$ in the previous case, this implies also that (E.9) must hold for $|m_n - m_0|$ rather than for $|m_n - \widetilde{m}_0|$.

Applying $|\operatorname{Im} z| \geq (\log n)^{-a}$, (E.9) yields w.h.p.

$$|m_n - m_0| \lesssim (\log n)^{a/2} |r_n| \lesssim \frac{(\log n)^{2+2\varepsilon+3a/2}}{\sqrt{d}}. \tag{E.10}$$

Recalling (E.8), $|R_{jj}| \leq (\log n)^a$ and $|m_n| \leq \frac{3}{2}|R_{jj}|$, we get

$$|R_{jj} - m_n| \lesssim |R_{jj}||m_n| \cdot \frac{(\log n)^{2+2\varepsilon+a}}{\sqrt{d}} \lesssim \frac{(\log n)^{2+2\varepsilon+3a}}{\sqrt{d}}. \tag{E.11}$$

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 (E.10) to the first inequality of (E.11) yields the desired estimate

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

To show (C.13) for the off-diagonals, we now apply (E.2), (C.2), (C.6) conditional on $R^{(jk)}$, $|R_{jj}| \lesssim 1$, $|R_{kk}^{(j)}| \lesssim 1$, $\|R^{(jk)}\|_\infty \leq (\log n)^a$, $\|R^{(jk)}\|_F \leq \sqrt{n}(\log n)^a$, and $d \leq n$ to get w.h.p.

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

E.3 Row sum bound

We now show that (C.15) holds for $z \in D$ w.h.p. Set

$$\mathcal{Z}_i \triangleq \sum_{j,k}^{(i)} a_{ik} R_{kj}^{(i)} = \sum_{k}^{(i)} a_{ik} \left(\mathbf{e}_k^\top R^{(i)} \mathbf{1} \right) \tag{E.12}$$

where the last equality holds because $R_{ki}^{(i)} = 0$ for $k \neq i$. Applying (E.2),

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

Then applying (C.14), w.h.p. for every $i \in [n]$,

$$\left| \mathbf{e}_i^\top R \mathbf{1} \right| \lesssim 1 + |\mathcal{Z}_i|. \tag{E.13}$$

Applying (C.3) conditional on $A^{(i)}$, w.h.p. for every $i \in [n]$,

$$|\mathcal{Z}_i| \leq (\log n)^{1+\varepsilon} \left(\frac{\max_{k \neq i} |\mathbf{e}_k^\top R^{(i)} \mathbf{1}|}{\sqrt{d}} + \sqrt{\frac{\sum_k^{(i)} |\mathbf{e}_k^\top R^{(i)} \mathbf{1}|^2}{n}} \right). \tag{E.14}$$

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

$$\sum_k^{(i)} \left| \mathbf{e}_k^\top R^{(i)} \mathbf{1} \right|^2 \leq \mathbf{1}^\top \overline{R^{(i)}} R^{(i)} \mathbf{1} \leq (\log n)^{2a} n. \tag{E.15}$$

For the first term, we apply (E.3), (C.13), and (C.14) to get, w.h.p. for all $k \neq i$,

$$\left| \mathbf{e}_k^\top R^{(i)} \mathbf{1} \right| = \left| \mathbf{e}_k^\top R \mathbf{1} - \frac{R_{ki}}{R_{ii}} \cdot \mathbf{e}_i^\top R \mathbf{1} \right| \leq \left| \mathbf{e}_k^\top R \mathbf{1} \right| + \frac{C(\log n)^{2+2\varepsilon+a}}{\sqrt{d}} \left| \mathbf{e}_i^\top R \mathbf{1} \right|. \tag{E.16}$$

Applying $d \gg (\log n)^{4+4\varepsilon+2a}$ and substituting (E.15) and (E.16) into (E.14) and then into (E.13), we get that

$$\left| \mathbf{e}_i^\top R \mathbf{1} \right| \lesssim 1 + (\log n)^{1+\varepsilon} \left(\frac{\max_k |\mathbf{e}_k^\top R \mathbf{1}|}{\sqrt{d}} + (\log n)^a \right) \tag{E.17}$$

Taking the maximum over i and rearranging yields (C.15).

E.4 Total sum bound

Finally, we show that (C.16) holds with probability $1 - e^{-c(\log n)(\log \log n)}$ for $z \in D$. As above, we set

$$\mathcal{Z}_i = \sum_{j,k}^{(i)} a_{ik} R_{kj}^{(i)} = \sum_{k}^{(i)} a_{ik} \left(\mathbf{e}_k^\top R^{(i)} \mathbf{1} \right). \tag{E.18}$$

Note that if we apply (E.15), (E.16), and (C.15) to (E.14), we obtain w.h.p. that for every $i \in [n]$,

$$|\mathcal{Z}_i| \leq (\log n)^{1+\varepsilon+a}. \tag{E.19}$$

The main step of the proof of (C.16) 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 [EKYY13b].

Lemma E.2 (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

$$\mathcal{Z}_i^{S,U} = \sum_{T:\, T \subseteq U} (-1)^{|T|} \mathcal{Z}_i^{[(S \setminus U) \cup T]}. \tag{E.20}$$

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

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

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 $n \geq n_0$,

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

Proof. See [EKYY13b, 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 (E.20) 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 [EKYY13b] for additional details.

We will check that the conditions of this lemma hold for $\mathcal{Z}_i$ as defined by (E.18), with the appropriate construction of variables $\mathcal{Z}_i^{[U]}$. To this end, we first extend (C.13), (C.14), and (C.15) to $R^{(S)}$ for $|S| \leq \log n$ in the following deterministic lemma:

Lemma E.3. *Suppose* (C.13), (C.14), *and* (C.15) *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* $|S| \le \log n$*, and all* $j \neq k \in [n] \setminus S$*,*

$$|R_{jj}^{(S)}(z) - m_0(z)| \le \frac{2C_0(\log n)^{2+2\varepsilon+3a}}{\sqrt{d}}, \tag{E.21}$$

$$|R_{jk}^{(S)}(z)| \le \frac{2C_0(\log n)^{2+2\varepsilon+a}}{\sqrt{d}}, \tag{E.22}$$

$$|\mathbf{e}_j^\top R^{(S)}(z)\mathbf{1}| \le 2C_0(\log n)^{1+\varepsilon+a}. \tag{E.23}$$

Proof. For integers $s \ge 0$, let

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

When (C.13) and (C.14) hold, we have that $\Lambda_s^d \le C_0(\log n)^{2+2\varepsilon+3a}/\sqrt{d}$ and $\Lambda_s^o \le C_0(\log n)^{2+2\varepsilon+a}/\sqrt{d}$ for $s = 0$. By (E.5), we have for each $s \ge 1$ and $* \in \{d, o\}$ that

$$\Lambda_{s+1}^* \le \Lambda_s^* + \frac{(\Lambda_s^o)^2}{|m_0| - \Lambda_s^d}. \tag{E.24}$$

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

$$\Lambda_s^d \le \frac{C_0(\log n)^{2+2\varepsilon+3a}}{\sqrt{d}}\left(1 + \frac{4C_0(\log n)^{2+2\varepsilon+a}}{|m_0|\sqrt{d}}\right)^s,\ \Lambda_s^o \le \frac{C_0(\log n)^{2+2\varepsilon+a}}{\sqrt{d}}\left(1 + \frac{4C_0(\log n)^{2+2\varepsilon+a}}{|m_0|\sqrt{d}}\right)^s. \tag{E.25}$$

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

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

We then have $|m_0| - \Lambda_s^d \ge |m_0|/2$ for $d \gg (\log n)^{4+4\varepsilon+6a}$, so (E.24) yields

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

Thus both bounds of (E.25) hold for $s+1$, completing the induction. This establishes (E.21) and (E.22).

To show (E.23), set

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

When (C.15) holds, $\Gamma_0 \le C_0(\log n)^{1+\varepsilon+a}$. Applying (E.3) and the bound $|m_0| - \Lambda_s^d \ge |m_0|/2$, we have

$$\Gamma_{s+1} \le (1 + 2\Lambda_s^o/|m_0|)\Gamma_s \overset{(E.22)}{\le} \left(1 + \frac{4C_0(\log n)^{2+2\varepsilon+3a}}{|m_0|\sqrt{d}}\right)\Gamma_s,$$

Thus $\Gamma_s \le 2\Gamma_0$ for all $s \le \log n$. $\square$

Lemma E.4. *Fix $z \in D$. Let $\mathcal{Z}_i$ be defined in* (E.18)*. For $U \subset [n]$ not containing i, define*

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

Let Ξ be the event where

- (C.13), (C.14)*, and* (C.15) *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 E.2 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

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

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

$$\left| \sum_{T\colon T \subseteq W} (-1)^{|T|} \left(\mathbf{e}_i^\top R^{(V \cup T)} \mathbf{1} \right) \right| \leq \widetilde{y} (Cxw)^w, \tag{E.26}$$

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

$$|\alpha_k| \leq \widetilde{y} (Cxu)^{|U|}, \qquad |\beta_k| \leq \widetilde{y} (Cxu)^{|U|+1}. \tag{E.27}$$

Let r be any even integer with $r \le p \le (\log n) - 1$. 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

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

We apply (C.4) 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 $w \le |U| \le |S| \le \log n$. Since $Cxw \ll 1$ and $|U|(Cxw)^{2|U|} \ll (n - |U|)(Cxw)^{2|U|+2}$ by the definition of x and $d \le n$, the bounds (E.27) imply

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

Then for a constant $C' > 0$, (C.4) gives

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

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

For condition (iii), we have

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

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 (C.13), (C.14), (C.15), (C.2), and in Lemma C.1.

It remains to establish the claim (E.26). For $W = \emptyset$, this follows from (C.15). Assume then that $w \ge 1$, and write $W = \{j_1, \ldots, j_{w-1}\}$ (in any order). For a function $f : \mathbb{R}^{n \times n} \to \mathbb{C}$ and any index $j \in [n]$, define $Q_j f : \mathbb{R}^{n \times n} \to \mathbb{C}$ by

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

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 $Q_j f(A) = f(A^{(S)}) - f(A^{(jS)})$. 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

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

We apply Schur complement identities to iteratively to expand $Q_{j_{w-1}} \dots Q_{j_1} f$: First applying (E.3), we get

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

Then applying (E.3), (E.4), and (E.5) 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

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

Applying (E.5), (E.4), and (E.3) to each factor of each summand above, and repeating iteratively, an induction argument verifies the following claims for each $t \in \{1, \dots, w-1\}$:

- $Q_{j_t} \dots Q_{j_1} f(A)$ 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
- jach factor is one of the following three forms, for a set $S \subseteq V \cup W$: $R_{jk}^{(S)}$ for $j, k \notin S$ distinct, or $1/R_{jj}^{(S)}$ for $j \notin S$, or $\mathbf{e}_j^\top R^{(S)} \mathbf{1}$ for $j \notin S$. Furthermore,
- Each summand of $Q_{j_t} \dots Q_{j_1} f(A)$ satisfies: (a) It has exactly one factor of the form $\mathbf{e}_j^\top R^{(S)} \mathbf{1}$. (b) The number of factors of the form $1/R_{jj}^{(S)}$ is less than or equal to the number of factors of the form $R_{jk}^{(S)}$ for $j \neq k$. (c) There are at least t factors of the form $R_{jk}^{(S)}$ 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 E.3, since $|W \cup V| \leq \log n$, we have $|R_{jk}^{(S)}| \leq C(\log n)^{2+2\varepsilon+a}/\sqrt{d}$, $|R_{jj}^{(S)}| \geq |m_0|/2$, and $|\mathbf{e}_j^\top R^{(S)} \mathbf{1}| \leq C(\log n)^{1+\varepsilon+a}$ on the event Ξ. Thus we get

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

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

We now show (C.16) holds for $z \in D$ with probability $1 - e^{-c(\log n)(\log\log n)}$. The diagonal bound (C.14) implies

$$|\operatorname{Tr} R - n \cdot m_0| \le \frac{Cn(\log n)^{2+2\varepsilon+3a/2}}{\sqrt{d}}. \tag{E.28}$$

To bound the sum of off-diagonal elements of R, we apply (E.2) to write

$$\sum_{i \neq k} R_{ik} = -\sum_i R_{ii}\mathcal{Z}_i = -m_0 \sum_i \mathcal{Z}_i - \sum_i (R_{ii} - m_0)\mathcal{Z}_i. \tag{E.29}$$

Applying (C.14) and (E.19) yields

$$\sum_i |(R_{ii} - m_0)\mathcal{Z}_i| \le \frac{Cn(\log n)^{3+3\varepsilon+5a/2}}{\sqrt{d}}. \tag{E.30}$$

Then applying Lemma E.2 with x, y, Ξ as defined in Lemma E.4 and with p being the largest even integer less than $(\log n) - 1$, we have

$$\mathbb{1}\{\Xi\} \left| n^{-1} \sum_i \mathcal{Z}_i \right| \le C(\log n)^{12} \cdot \sqrt{d}(\log n)^{-\varepsilon} \cdot (\log n)^{4+4\varepsilon+2a}/d \le \frac{C(\log n)^{16+3\varepsilon+2a}}{\sqrt{d}} \tag{E.31}$$

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 (E.31) by $n \cdot m_0$ and combining with (E.28)–(E.30) yields (C.16).

F Proof of Lemma 2.3

We now prove Lemma 2.3. Assume without loss of generality that Π_* is the identity matrix. For any $k \ge 2$ we have

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

Thus, the moment conditions (13) and (14) are satisfied. In addition, we have that for all $i < j$,

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

where the last equality holds by the choice of σ^2. Thus, (15) 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

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

where the last inequality is due to $\sigma^2 \ge \frac{1-s}{1-p}$. Thus, by applying Lemma C.1 and $2(\log n)^7 \le 2\sigma^2 d \le n$ where the upper bound follows from $p(1-s) \le s(1-s) \le 1/4$, there exists a constant $C > 0$ such that for any $D > 0$, with probability at least $1 - n^{-D}$ for all $n \ge n_0(D)$, we have $\|\Delta\| \le C$ and hence $\|A - B\| \le \sqrt{2}C\sigma$. Thus (16) is satisfied.

G A tighter regularized QP relaxation

As discussed in Section 1.3, GRAMPA can be interpreted as solving the regularized QP relaxation (12) of the QAP. We further explore this optimization aspect in this section. As a further step toward understanding convex relaxations of the QAP, we analyze the following intermediate program between (10) and (12):

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

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 [ABK15] 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 (12)) does not automatically carries over to the tighter program (G.1). Fortunately, we are able to employ similar technical tools to analyze the solution to (G.1), denoted henceforth by X^{c}.

The following result is the counterpart of Theorem D.1 and Theorem 2.2:

Theorem G.1. *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 $1 - n^{-10}$,*

$$\max_{\pi_*(k)\neq\ell} |n \cdot X^{\mathsf{c}}_{k\ell}| \leq C(\log n)^\kappa \frac{1}{\sqrt{\eta}}, \tag{G.2}$$

$$\max_k \left| n \cdot X^{\mathsf{c}}_{k\pi_*(k)} - \frac{4(1-\sigma^2)}{\pi\eta} \right| \leq C\left(\frac{r(n)}{\eta} + \frac{\sigma}{\eta^2} + (\log n)^\kappa \frac{1}{\sqrt{\eta}} \right). \tag{G.3}$$

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

$$(\log n)^{-a} \leq \eta \leq c(\log n)^{-2\kappa} \quad \text{and} \quad \sigma \leq c'\eta, \tag{G.4}$$

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

$$\min_k X_{k\pi_*(k)} > \max_{\pi_*(k)\neq\ell} X_{k\ell}. \tag{G.5}$$

Compared with Theorem 2.2, the theoretical guarantee for the tighter program (G.1) is similar to that for (12) and the GRAMPA method. In practice the performance of the former is slightly better (cf. Fig. G.1). Furthermore, Theorem G.1 applies verbatim to the solution of (G.1) with column-sum constraints $X^\top\mathbf{1} = \mathbf{1}$ instead. This simply follows by replacing (A, B, X, Π_*) with $(B, A, X^\top, \Pi_*^\top)$.

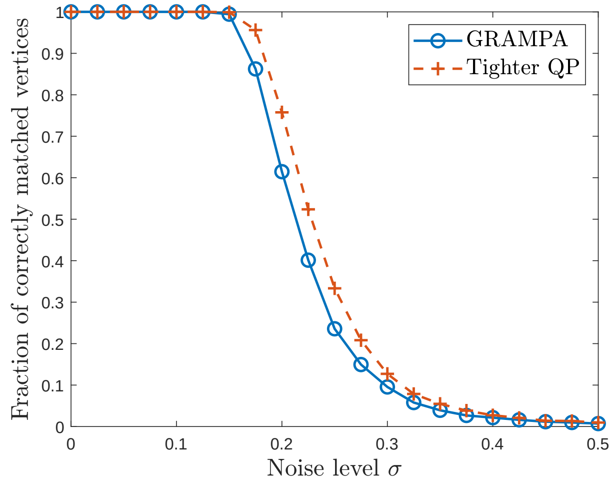

Figure G.1: Fraction of correctly matched pairs of vertices by GRAMPA and the tighter QP (G.1) (both followed by linear assignment rounding) on Erdős-Rényi graphs with 1000 vertices and edge density 0.5, averaged over 10 repetitions.

G.1 Structure of solutions to QP relaxations

Before proving Theorem G.1, we first provide an overview of the structure of solutions to the QP relaxations (12), (G.1) and (10). Using the Karush–Kuhn–Tucker (KKT) conditions, the solution of (G.1) can be expressed as

$$X^{\mathsf{c}} = \sum_{i,j} \frac{\langle u_i, \mu \rangle \langle v_j, \mathbf{1} \rangle}{(\lambda_i - \mu_j)^2 + \eta^2} u_i v_j^\top, \tag{G.6}$$

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

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

where

$$\tau_i \triangleq \sum_j \frac{\langle v_j, \mathbf{1} \rangle^2}{(\lambda_i - \mu_j)^2 + \eta^2}. \tag{G.7}$$

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

$$\mu = \sum_i \frac{\langle u_i, \mathbf{1} \rangle}{\tau_i} u_i, \tag{G.8}$$

so we obtain

$$X^{\mathsf{c}} = \sum_{i,j} \frac{1}{(\lambda_i - \mu_j)^2 + \eta^2} \frac{1}{\tau_i} u_i u_i^\top \mathbf{J} v_j v_j^\top. \tag{G.9}$$

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

$$\tau_i = \frac{1}{\eta}\,\mathrm{Im}\sum_j \frac{\langle v_j, \mathbf{1}\rangle^2}{\mu_j - (\lambda_i + \mathbf{i}\eta)} = \frac{1}{\eta}\mathbf{1}^\top\left[\mathrm{Im}\sum_j \frac{1}{\mu_j - (\lambda_i + \mathbf{i}\eta)} v_j v_j^\top\right]\mathbf{1}$$
$$= \frac{1}{\eta}\,\mathrm{Im}[\mathbf{1}^\top R_B(\lambda_i + \mathbf{i}\eta)\mathbf{1}]. \tag{G.10}$$

Invoking the resolvent bound (C.16), we expect $\tau_i \approx \frac{n}{\eta}\,\mathrm{Im}[m_0(\lambda_i + \mathbf{i}\eta)]$, where, by properties of the Stieltjes transform (cf. Proposition C.5), $\mathrm{Im}[m_0(\lambda_i + \mathbf{i}\eta)] \approx \mathrm{Im}[m_0(\lambda_i)] = \pi\rho(\lambda_i)$ as $\eta \to 0$. Thus we have the approximation

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

Compared with the unconstrained solution $\widehat{X}$ defined in (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 G.2 (Structure of the QP solutions)**.** *Let us point out that solution of various QP relaxations, including (10), (G.1), and (12), are of the following common form:*

$$X = \sum_{i,j}\frac{\eta}{(\lambda_i - \mu_j)^2 + \eta^2} u_i u_i^\top S v_j v_j^\top, \tag{G.11}$$

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 (12) 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 effect the subsequent rounding step, this is equivalent to $\eta\widehat{X}$ that we have analyzed.*
- *For (G.1) with the row sum constraint, $S = \mu\mathbf{1}^\top$ is rank-one with μ given in (G.8).*
- *For (10) 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 (10) 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.*

G.2 Proof of Theorem G.1

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

G.2.1 Resolvent representation of the solution

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

Lemma G.3. *Consider symmetric matrices A and B with the spectral decompositions* (2)*, and suppose that $\|A\| \le 2.5$. Then the solution X^{c} of the program* (G.1) *admits the following representation*

$$X^{\mathsf{c}} = \frac{1}{2\pi} \operatorname{Re} \oint_{\Gamma} F(z) R_A(z) \mathbf{J} R_B(z + \mathbf{i}\eta), \tag{G.12}$$

where Γ is defined by (B.2) *and*

$$F(z) \triangleq \frac{2\mathbf{i}}{\mathbf{1}^\top R_B(z+\mathbf{i}\eta)\mathbf{1} - \mathbf{1}^\top R_B(z-\mathbf{i}\eta)\mathbf{1}}. \tag{G.13}$$

Proof. By (G.10) we have $\tau_i^{-1} = \eta F(\lambda_i)$. This leads to the following contour representation of X^{c} analogous to (B.1) for the unconstrained solution:

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

where (a) follows from the Ward identity (Lemma B.1); (b) follows from Cauchy integral formula and the analyticity of F in the region enclosed by the contour Γ. $\square$

G.2.2 Entrywise approximation

For some small constant $\varepsilon > 0$, let b, b' be as defined in Theorem C.6. Under the assumptions of Theorem G.1, we have $c_0 > b'$ for ε sufficiently small, so that Theorem C.6 applies. Recall the notation $\delta_1, \dots, \delta_4$ defined in (D.10). For sufficiently small $\varepsilon > 0$, we may also verify under the assumptions of Theorem G.1 that $\delta_i = o(1)$ for each $i = 1, 2, 3, 4$, and

$$\frac{\delta_1 \delta_2^2 n}{\eta} \le 1, \quad \frac{\delta_2^2 \delta_3 n}{\eta^2} \le \frac{(\log n)^\kappa}{\sqrt{\eta}}, \quad \text{and} \quad \delta_3 \le \eta^3. \tag{G.14}$$

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

Thanks to (C.16), we can approximate $F(z)$ by

$$\widetilde{F}(z) = \frac{1}{n} \frac{2\mathbf{i}}{m_0(z+\mathbf{i}\eta) - m_0(z-\mathbf{i}\eta)} \tag{G.15}$$

and approximate X^{c} by

$$\widetilde{X}^{\mathsf{c}} = \frac{1}{2\pi} \operatorname{Re} \oint_{\Gamma} \widetilde{F}(z) R_A(z) \mathbf{J} R_B(z+\mathbf{i}\eta) \tag{G.16}$$

$$= \frac{-1}{\pi n} \operatorname{Im} \oint_{\Gamma} \frac{1}{m_0(z+\mathbf{i}\eta) - m_0(z-\mathbf{i}\eta)} R_A(z) \mathbf{J} R_B(z+\mathbf{i}\eta). \tag{G.17}$$

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

Lemma G.4. *Suppose (G.14) holds. On the high-probability event where Theorem C.6 holds and also* $\|A\| \le 2.5$*,*

$$\|\widetilde{X}^{\mathsf{c}} - X^{\mathsf{c}}\|_{\ell_\infty} \lesssim \frac{\delta_2^2 \delta_3}{\eta^2} \le \frac{(\log n)^\kappa}{n\sqrt{\eta}}, \tag{G.18}$$

where δ_2, δ_3 *are defined in (D.10).*

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 $|\operatorname{Im}(z)| \le \eta/2$ for $z \in \Gamma$, and thus $\operatorname{Im}(z+\mathbf{i}\eta)$ and $\operatorname{Im}(z-\mathbf{i}\eta)$ have different signs. Therefore

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

where the last step follows from (C.11). Furthermore, by (C.16), we have $\sup_{z\in\Gamma} |G(z) - \widetilde{G}(z)| \le 2C\delta_3$. In view of (G.14), $\delta_3 \ll \eta$. 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 (G.12) and (G.16), we have

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

By (C.15), 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 Section D so we focus on the differences. To make the scaling identical to the unconstrained case, define

$$Y \triangleq n\widetilde{X}^{\mathsf{c}} = \frac{1}{2\pi} \operatorname{Re} \oint_{\Gamma} f(z) R_A(z) \mathbf{J} R_B(z+\mathbf{i}\eta), \tag{G.19}$$

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 (B.1), 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

$$\max_{k\neq\ell}|Y_{k\ell}| \lesssim (\log n)^{\kappa}/\sqrt{\eta}. \tag{G.20}$$

2. For diagonal entries, we have

$$\min_k \left|Y_{kk} - \frac{4(1-\sigma^2)}{\pi\eta}\right| \lesssim \frac{r(n)}{\eta} + \frac{\sigma}{\eta^2} + (\log n)^{\kappa}\frac{1}{\sqrt{\eta}}. \tag{G.21}$$

In view of Lemma G.4, this implies the desired (G.2) and (G.3). Finally, analogous to Theorem 2.2, under the assumption (G.4) 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 (G.5).

G.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

$$|f(z)| \leq \frac{2}{|\operatorname{Im}[m_0(z+\mathbf{i}\eta) - m_0(z-\mathbf{i}\eta)]|} = \frac{2}{|\operatorname{Im} m_0(z+\mathbf{i}\eta)| + |\operatorname{Im} m_0(z-\mathbf{i}\eta)|} \lesssim \frac{1}{\eta}, \tag{G.22}$$

where the last step applies (C.11). Analogous to (D.17), we get

$$\begin{aligned} 2\pi Y_{12} &= \operatorname{Re}\left(\oint_\Gamma f(z)\left[\mathbf{e}_1^\top R_A(z)\mathbf{1}\right]\left[\mathbf{e}_2^\top R_B(z+\mathbf{i}\eta)\mathbf{1}\right]dz\right) \\ &= \operatorname{Re}\left(\alpha - a_1^\top g - b_2^\top h + a_1^\top M b_2\right) + O\left(\frac{\delta_1\delta_2^2 n}{\eta}\right), \end{aligned} \tag{G.23}$$

where

$$\alpha \triangleq \oint_\Gamma f(z)m_0(z)m_0(z+\mathbf{i}\eta)dz, \tag{G.24}$$

$$g \triangleq \oint_\Gamma f(z)m_0(z)m_0(z+\mathbf{i}\eta)R_A^{(12)}(z)\mathbf{1}_{n-2}dz, \tag{G.25}$$

$$h \triangleq \oint_\Gamma f(z)m_0(z)m_0(z+\mathbf{i}\eta)R_B^{(12)}(z+\mathbf{i}\eta)\mathbf{1}_{n-2}dz, \tag{G.26}$$

$$M \triangleq \oint_\Gamma f(z)m_0(z)m_0(z+\mathbf{i}\eta)R_A^{(12)}(z)\mathbf{J}_{n-2}R_B^{(12)}(z+\mathbf{i}\eta)dz. \tag{G.27}$$

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

$$\begin{aligned}
\operatorname{Re}\alpha =& \operatorname{Re}\oint dz \frac{2\mathbf{i}}{m_0(z+\mathbf{i}\eta)-m_0(z-\mathbf{i}\eta)} m_0(z)m_0(z+\mathbf{i}\eta) \\
=& \int \rho(x)dx \operatorname{Re}\oint dz \frac{1}{x-z}\frac{2\mathbf{i}\ m_0(z+\mathbf{i}\eta)}{m_0(z+\mathbf{i}\eta)-m_0(z-\mathbf{i}\eta)} \\
=& \int \rho(x)dx \operatorname{Re}\left[(-2\pi\mathbf{i})\frac{2\mathbf{i}\ m_0(x+\mathbf{i}\eta)}{m_0(x+\mathbf{i}\eta)-m_0(x-\mathbf{i}\eta)}\right] \\
=& 2\pi \int \rho(x)dx \operatorname{Re}\left[\frac{2\,m_0(x+\mathbf{i}\eta)}{2\mathbf{i}\operatorname{Im} m_0(x+\mathbf{i}\eta)}\right] = 2\pi\int\rho(x)dx = 2\pi.
\end{aligned} \tag{G.28}$$

As in Section D.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 (G.22), the ℓ_∞-norms are at most an $O(1/\eta)$ factor of those obtained in (D.26) and (D.27), 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 D.4:

Lemma G.5. *Assume the same setting of Lemma D.4, and define M, g, and h as in (G.25–G.27) 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. D.1. Proceeding as in the proof of Lemma D.4, we have

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

where (II) denotes the remainder term. Applying (C.16), (G.22), and the boundedness of m_0, the residual term is bounded as

$$|\text{(II)}| \lesssim \delta_3 \oint_\Gamma dz \oint_{\Gamma'} dw |f(z)||f(w)| \frac{1}{|z-w|}\frac{1}{|z+\mathbf{i}\eta-(w-\mathbf{i}\eta)|} \lesssim \frac{\delta_3}{\eta^4} \lesssim \frac{1}{\eta}. \tag{G.29}$$

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 (C.11) 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 C.5. Then we obtain from (G.22) the improved bound $|f(z)| \lesssim 1/\sqrt{\eta+\zeta(z)}$, and hence

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

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

$$\oint \oint_{|z-w|\geq c_0} dz\, dw \, \frac{1}{\sqrt{\eta+\zeta(z)}} \frac{1}{\sqrt{\eta+\zeta(w)}} \left| \frac{m_0(z)-m_0(w)}{z-w} \right| \left| \frac{m_0(z+\mathbf{i}\eta)-m_0(w-\mathbf{i}\eta)}{z+\mathbf{i}\eta-(w-\mathbf{i}\eta)} \right| \lesssim \frac{1}{\eta}. \tag{G.30}$$

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

$$|m_0(z)-m_0(w)| \lesssim \sqrt{\eta+\zeta(z)} + \sqrt{\eta+\zeta(w)}, \tag{G.31}$$

$$|m_0(z+\mathbf{i}\eta)-m_0(w-\mathbf{i}\eta)| \lesssim \sqrt{\eta+\zeta(z)} + \sqrt{\eta+\zeta(w)}. \tag{G.32}$$

Indeed, if $\zeta(z)>c_0$, then (G.31) and (G.32) hold because $\sqrt{\eta+\zeta(z)}+\sqrt{\eta+\zeta(w)} \asymp 1$. If instead $\zeta(z) \leq c_0$, say, $\operatorname{Re} z \geq 2-c_0$, then from the explicit form (C.9) 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 $\operatorname{Re} w \geq \operatorname{Re} z - |z-w| \geq 2-2c_0$, we also have $|1+m_0(w)| \lesssim \sqrt{\eta+\zeta(w)}$. Then (G.31) follows from the triangle inequality. The case of $\operatorname{Re} z \leq -2+c_0$, and the argument for (G.32), are analogous.

Having established (G.31) and (G.32), we apply

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

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 $|z-w| \geq \eta$, 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

$$\oint\oint_{\eta\le|z-w|<c_0} dz\,dw\,\frac{\sqrt{|z-w|}}{\sqrt{\eta}}\cdot\frac{1}{|z-w||z+\mathbf{i}\eta-(w-\mathbf{i}\eta)|}$$
$$\lesssim\frac{1}{\sqrt{\eta}}\oint\oint_{\eta\le|z-w|<c_0} dz\,dw\,\frac{1}{|z-w|^{3/2}}\lesssim\frac{1}{\sqrt{\eta}}\frac{1}{\sqrt{\eta}}\lesssim\frac{1}{\eta}. \tag{G.33}$$

Combining with the first case (G.30), we get $|(\mathrm{I})|\lesssim 1/\eta$. Finally, combining with (G.29), we get $\|M\|_F^2\lesssim n^2/\eta$ as desired.

Next we bound $\|g\|$. Proceeding as in the proof of Lemma D.4 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 $|z-w|\ge c_0$, we have

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

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

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

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 (D.20)–(D.21) and using the preceeding norm bounds, we obtain from (G.23):

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

with probability at least $1-n^{-D}$, for any constant D. This implies the desired (G.20) by the union bound.

G.2.4 Diagonal entries

We now consider Y_{11}. Following the derivation from (D.31) to (D.32) and using Lemma G.5 in place of Lemma D.4, we obtain, with probability at least $1-n^{-D}$ for any constant D,

$$\left|Y_{11} - \frac{1-\sigma^2}{2\pi}\operatorname{Re}\frac{\operatorname{Tr}(M)}{n}\right| \lesssim (\log n)^{\kappa}\frac{1}{\sqrt{\eta}}, \tag{G.34}$$

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 D.5 and Lemma D.6:

Lemma G.6. *Suppose $\delta_3 \le \eta^2$. Assume the setting of Lemma D.5. 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 (D.35), we have $\frac{1}{n}\operatorname{Tr}(M) = \text{(I)} - \text{(II)}$, where

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

To bound (II), consider two cases:

- For $z\in\Gamma$ with $|\operatorname{Im} z| = \eta/2$, by the Ward identity and (C.16), 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 (C.11) and (G.22), 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 $\|A\| \leq 2.5$, $\left|\mathbf{1}^\top R_B(z+\mathbf{i}\eta)(A-B)R_A(z)\mathbf{1}\right| \lesssim n\sigma$.

Furthermore, by (G.22), $|f(z)| \lesssim \frac{1}{\eta}$ 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 $\delta_3 \leq \eta^2$ by the assumption.

For (I), applying (C.16) 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 D.6: Note that

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

by (C.11), 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

$$\begin{aligned}&\oint_\Gamma m_0(z)m_0(z+\mathbf{i}\eta)\frac{m_0(z+\mathbf{i}\eta) - m_0(z)}{m_0(z+\mathbf{i}\eta) - m_0(z-\mathbf{i}\eta)} dz \\ &= \int_2^{-2} m_0^+(x)m_0(x+\mathbf{i}\eta)\frac{m_0(x+\mathbf{i}\eta) - m_0^+(x)}{m_0(x+\mathbf{i}\eta) - m_0(x-\mathbf{i}\eta)} dx + \int_{-2}^{2} m_0^-(x)m_0(x+\mathbf{i}\eta)\frac{m_0(x+\mathbf{i}\eta) - m_0^-(x)}{m_0(x+\mathbf{i}\eta) - m_0(x-\mathbf{i}\eta)} dx.\end{aligned} \tag{G.35}$$

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

$$|m_0(x+\mathbf{i}\eta) - m_0(x-\mathbf{i}\eta)| = 2\operatorname{Im} m_0(x+\mathbf{i}\eta) \gtrsim \sqrt{\zeta(x)+\eta} \geq \sqrt{\delta}. \tag{G.36}$$

Then the above integrands are bounded by $C/\sqrt{\delta}$ for $|x| \leq 2-\delta$. 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 (G.31) 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 (G.36), we see that the integrands of (G.35) are uniformly bounded for all small η.

Now we 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

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

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

Finally, combining (G.34) with Lemma G.6 and $\delta_3 \ll \eta$ from (G.14), and applying a union bound yields the desired (G.21).

H Signal-to-noise heuristics

We justify the choice of the Cauchy weight kernel in (4) by a heuristic signal-to-noise calculation for $\widehat{X}$. For simplicity, consider the Gaussian Wigner model $B = A + \sigma Z$, where A and Z are independent GOE matrices. We assume without loss of generality that π^* is the identity, so that diagonal entries of $\widehat{X}$ indicate similarity between matching vertices of A and B. Then for the rounding procedure in (5), we may interpret $n^{-1} \operatorname{Tr} \widehat{X}$ and $(n^{-2} \sum_{i,j:\, i \neq j} \widehat{X}_{ij}^2)^{1/2} \approx n^{-1} \|\widehat{X}\|_F$ as the average signal strength and noise level in $\widehat{X}$. Let us define a corresponding signal-to-noise ratio as

$$\mathrm{SNR} = \frac{\mathbb{E}[\operatorname{Tr} \widehat{X}]}{\mathbb{E}[\|\widehat{X}\|_F^2]^{1/2}}$$

and compute this quantity in the Gaussian Wigner model.

We abbreviate the spectral weights $w(\lambda_i, \mu_j)$ as w_{ij}. For $\widehat{X}$ defined by (3) with any weight kernel $w(x, y)$, we have

$$\operatorname{Tr} \widehat{X} = \sum_{ij} w_{ij} \cdot u_i^\top \mathbf{J} v_j \cdot u_i^\top v_j.$$

Applying that (A, B) is equal in law to $(OAO^\top, OBO^\top)$ for a rotation O such that $O\mathbf{1} = \sqrt{n}\mathbf{e}_k$, we obtain for every k that

$$\mathbb{E}[\operatorname{Tr} \widehat{X}] = \sum_{ij} n \cdot \mathbb{E}[w_{ij} \cdot u_i^\top (\mathbf{e}_k \mathbf{e}_k^\top) v_j \cdot u_i^\top v_j].$$

Then averaging over $k = 1, \ldots, n$ and applying $\sum_k \mathbf{e}_k \mathbf{e}_k^\top = \mathbf{I}$ yield that

$$\mathbb{E}[\operatorname{Tr} \widehat{X}] = \sum_{ij} \mathbb{E}[w_{ij} (u_i^\top v_j)^2].$$

For the noise, we have

$$\|\widehat{X}\|_F^2 = \operatorname{Tr} \widehat{X} \widehat{X}^\top = \sum_{i,j,k,l} w_{ij} w_{kl} (u_i^\top \mathbf{J} v_j)(u_k^\top \mathbf{J} v_\ell) \operatorname{Tr}(u_i v_j^\top \cdot v_\ell u_k^\top) = \sum_{ij} w_{ij}^2 (u_i^\top \mathbf{J} v_j)^2.$$

Applying the equality in law of (A, B) and $(OAO^\top, OBO^\top)$ for a uniform random orthogonal matrix O, and writing $r = O\mathbf{1}/\sqrt{n}$, we get

$$\mathbb{E}[\|\widehat{X}\|_F^2] = \sum_{ij} n^2 \cdot \mathbb{E}[w_{ij}^2 (u_i^\top r)^2 (v_j^\top r)^2].$$

Here, $r = (r_1, \ldots, r_n)$ is a uniform random vector on the unit sphere, independent of (A, B). For any deterministic unit vectors u, v with $u^\top v = \alpha$, we may rotate to $u = \mathbf{e}_1$ and $v = \alpha \mathbf{e}_1 + \sqrt{1 - \alpha^2} \mathbf{e}_2$ to get

$$\mathbb{E}[(u^\top r)^2 (v^\top r)^2] = \mathbb{E}[r_1^2 \cdot (\alpha r_1 + \sqrt{1 - \alpha^2} r_2)^2] = \alpha^2 \mathbb{E}[r_1^4] + (1 - \alpha^2) \mathbb{E}[r_1^2 r_2^2] = \frac{1 + 2\alpha^2}{n(n+2)},$$

where the last equality applies an elementary computation. Bounding $1 + 2\alpha^2 \in [1, 3]$ and applying this conditional on (A, B) above, we obtain

$$\mathbb{E}[\|\widehat{X}\|_F^2] = \frac{cn}{n+2} \sum_{ij} \mathbb{E}[w_{ij}^2]$$

for some value $c \in [1,3]$.

To summarize,

$$\text{SNR} \asymp \frac{\sum_{ij} \mathbb{E}[w(\lambda_i, \mu_j)(u_i^\top v_j)^2]}{\sqrt{\sum_{ij} \mathbb{E}[w(\lambda_i, \mu_j)^2]}}.$$

The choice of weights which maximizes this SNR would satisfy $w(\lambda_i, \mu_j) \propto (u_i^\top v_j)^2$. Recall that for $n^{-1+\varepsilon} \ll \sigma^2 \ll n^{-\varepsilon}$ and i, j in the bulk of the spectrum, we have the approximation (9). Thus this optimal choice of weights takes a Cauchy form, which motivates our choice in (4).

We note that this discussion is only heuristic, and maximizing this definition of SNR does not automatically imply any rigorous guarantee for exact recovery of π^*. Our proposal in (4) is a bit simpler than the optimal choice suggested by (9): The constant C in (9) depends on the semicircle density near λ_i, but we do not incorporate this dependence in our definition. Also, while (9) depends on the noise level σ, our main result in Theorem 2.2 shows that η need not be set based on σ, which is usually unknown in practice.

References

[ABK15] Yonathan Aflalo, Alexander Bronstein, and Ron Kimmel. On convex relaxation of graph isomorphism. *Proceedings of the National Academy of Sciences*, 112(10):2942–2947, 2015.

[Bia97] Philippe Biane. On the free convolution with a semi-circular distribution. *Indiana University Mathematics Journal*, pages 705–718, 1997.

[EKYY13a] László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. The local semicircle law for a general class of random matrices. *Electron. J. Probab*, 18(59):1–58, 2013.

[EKYY13b] László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Spectral statistics of Erdős–Rényi graphs I: local semicircle law. *The Annals of Probability*, 41(3B):2279–2375, 2013.

[EYY12] László Erdős, Horng-Tzer Yau, and Jun Yin. Bulk universality for generalized Wigner matrices. *Probability Theory and Related Fields*, 154(1-2):341–407, 2012.

[HW71] 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.

[RV13] Mark Rudelson and Roman Vershynin. Hanson-Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.*, 18:no. 82, 9, 2013.