AI4OPT Tutorial Lectures: Recent Results in Planted Assignment Problems

Lecture 3: Spectral graph matching

Lecturer: Jiaming Xu *Scribe: Jiaming Xu, April 20, 2023*

Spectral algorithms are powerful methods for recovering the underlying structure in planted models based on the principal eigenvectors of the observed data matrix A. Under planted models such as planted clique or planted partition models, spectral algorithms and their variants have been shown to achieve either the optimal recovery thresholds or the best possible performance within certain relaxation hierarchies. The rationale behind spectral algorithms is that the principal eigenvectors of $\mathbb{E}[A]$ contains information about underlying structures and the principal eigenvectors of A are close to those of $\mathbb{E}[A]$, provided that the spectral gap (the gap between the largest few eigenvalues and the rest of them) is much larger than the spectral norm of the perturbation $\|A - \mathbb{E}[A]\|$.

Unfortunately, in random graph matching, the observed adjancency matrices of the two Erdős-Rényi graphs have full rank and vanishing eigengaps. Therefore, for the naive spectral algorithms to suceed, it requires a very high signal-to-noise ratio, that is, the fraction of edges differed in two graphs $\delta \le n^{-C}$ for a constant C. In this lecture, we will develop a new spectral method that exactly recovers the underlying vertex correspondence with high probability when $\delta \le 1/\mathsf{polylog}(n)$, hereby achieving an exponential improvement in the noise tolerance [FMWX19a, FMWX19b].

3.1 Spectral graph matching paradigm

Write the eigenvalue decompsotions of the adjacency matrices A and B as

$$A = \sum_{i=1}^{n} \lambda_i u_i u_i^\top \text{ and } B = \sum_{j=1}^{n} \mu_j v_j v_j^\top, \tag{3.1}$$

where the eigenvalues are ordered such that

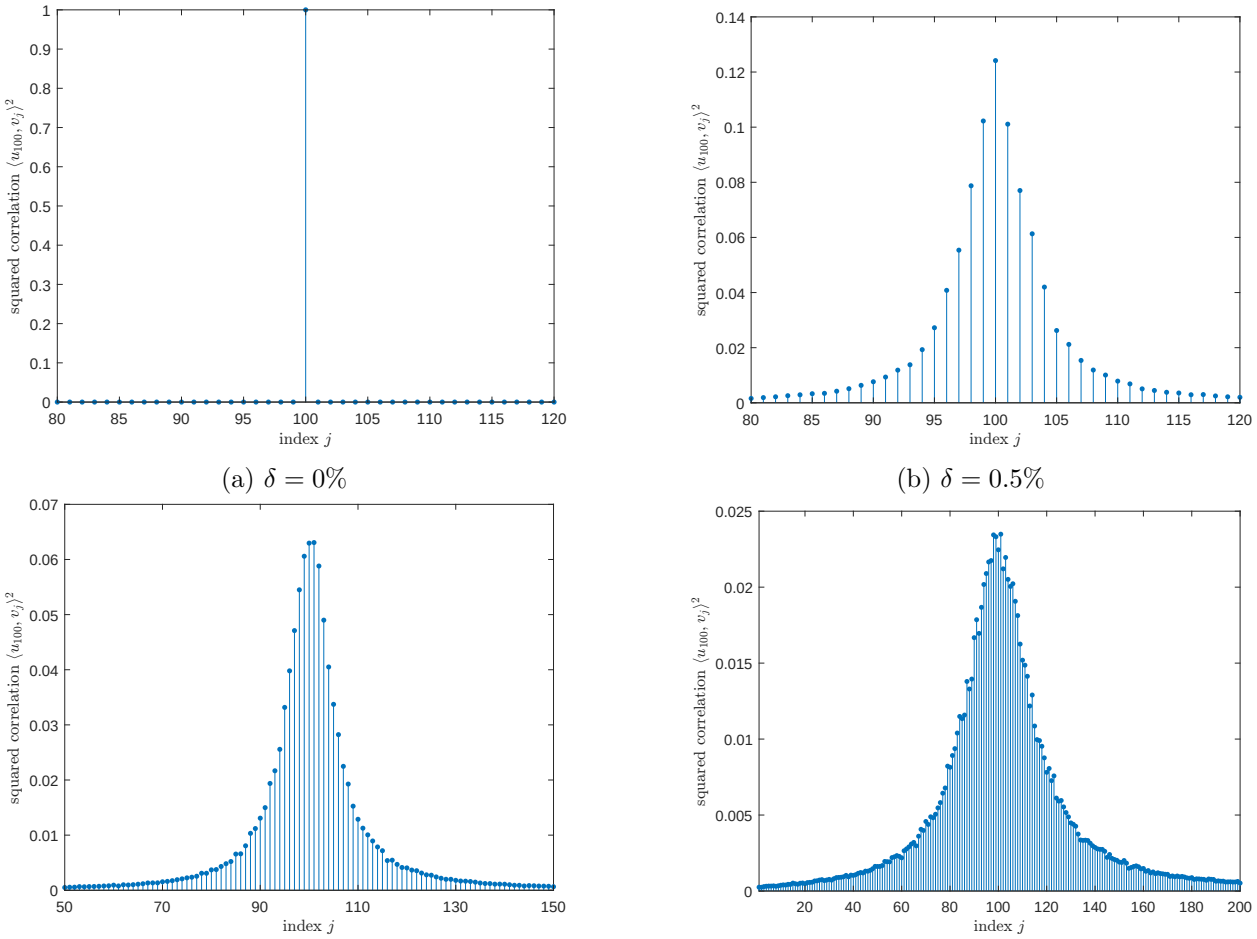
$$\lambda_1 \ge \cdots \ge \lambda_n \text{ and } \mu_1 \ge \cdots \ge \mu_n.$$

A general spectral graph matching paradigm works as follows:

1. Construct similarity matrix X based on (λ_i, u_i) and (μ_j, v_j), where $X_{k\ell}$ measures the similarity between vertex k in graph A and vertex ℓ in graph B;
2. Project X to permutation by linear assignment, that is,
$$\widehat{\Pi} \in \arg\max \langle X, \Pi \rangle . \tag{3.2}$$

Many previous spectral methods fall into the above paradigm with different constructions of the similarity matrix. These methods include:

- *Low-rank methods* that use a small number of eigenvectors of A and B. The simplest such approach uses only the leading eigenvectors, taking as the similarity matrix
$$\widehat{X} = u_1 v_1^\top. \tag{3.3}$$
Then $\widehat{\pi}$ which solves (3.2) sorts the entries of v_1 in the order of u_1.



(a) $\delta = 0\%$ (b) $\delta = 0.5\%$

(c) $\delta = 1\%$ (d) $\delta = 3\%$

Figure 3.1: Erdős-Rényi graphs: 500 vertices, edge probability $\frac{1}{2}$. The plot of $\langle u_{100}, v_j\rangle^2$, averaged across 1000 simulations.

- *Full-rank methods* that use all eigenvectors of A and B. A notable example is the popular method of Umeyama [Ume88], which sets

$$\widehat{X} = \sum_{i=1}^{n} s_i u_i v_i^\top \tag{3.4}$$

  where $s_i \in \{-1, 1\}$. The motivation is that (3.4) is the solution to the *orthogonal relaxation* of the QAP (3.6), where the feasible set is relaxed to the set of the orthogonal matrices.

All these methods turn out to perform very well with no noise, but are extremely fragile with noise. The underlying reason is that A and B have full rank and vanishing eigengaps. This leads to decorrelation of the eigenvectors u_i and v_i even when $\delta = n^{-C}$, as illustrated in Fig. 3.1.

3.2 A new spectral method: GRAMPA

As can be seen from Fig. 3.1, when the fraction of edges differed in the two graphs δ increases, the correlation between u_i and v_i decreases, and v_j becomes correlated with u_i for a wider range of j neighboring i. It suggests that to effectively leverage the correlation among the eigenvectors, we need to align all pairs of eigenvectors. This inspires our new spectral method - GRAph Matching by Pairwise eigen-Alignments, dubbed GRAMPA.

$$\boxed{X = \sum_{i,j=1}^{n} \underbrace{\frac{\eta}{(\lambda_i - \mu_j)^2 + \eta^2}}_{\text{spectral weights}} \times \underbrace{u_i^\top \mathbf{J} v_j \cdot u_i v_j^\top}_{\text{“Alignment” between } u_i \text{ and } v_j}}, \tag{3.5}$$

where η = bandwidth parameter, $\mathbf{J}$ = all-one matrix.

Note that the spectral weight penalizes pairs whose eigenvalues are far apart and the Cauchy spectral weight is inspired by the eigenvector correlation decay

$$n \cdot \mathbb{E}\left[\langle u_i, v_j \rangle^2\right] \approx \frac{\delta}{(\lambda_i - \mu_j)^2 + C\delta^2}.$$

Also, a nice feature of GRAMPA is that it is invariant to the choices of signs for u_i and v_j.

It turns out that GRAMPA is also rooted in optimization, as the similarity matrix X corresponds to the solution to a convex relaxation of the QAP, regularized by an added ridge penalty. To see this, note that graph matching can be cast as a quadratic assignment problem (QAP):

$$\arg\max_{\Pi \in S_n} \langle A, \Pi B \Pi^\top \rangle = \arg\min_{\Pi \in S_n} \|A - \Pi B \Pi^\top\|_F^2 = \arg\min_{\Pi \in S_n} \|A\Pi - \Pi B\|_F^2, \tag{3.6}$$

where the last equality holds because Π is an orthogonal matrix. Note that the objective function $\|A\Pi - \Pi B\|_F^2$ is quadratic in Π. The only non-convex part is the permutation set constraint $\Pi \in S_n$. Since the convex hull of the permutation matrices is the set of the doubly stochastic matrices, this gives rise to a natural quadratic programming relaxation:

$$\arg\min_{X \geq 0:\ X\mathbf{1}=\mathbf{1},\ X^\top \mathbf{1}=\mathbf{1}} \|AX - XB\|_F^2 \qquad \text{(QP-DS)} \tag{3.7}$$

It turns out that the GRAMPA similarity matrix X is (a multiple of)

$$\arg\min_{X:\ \mathbf{1}^\top X \mathbf{1} = n} \|AX - XB\|_F^2 + \eta^2 \|X\|_F^2, \tag{3.8}$$

which further relaxes the DS constraint to be the total-sum constraint and adds a ridge regularizer. To see this, define the Lagrangian function

$$L(X, \gamma) = \|AX - XB\|_F^2 + \eta^2 \|X\|_F^2 - 2\gamma \left(\langle X, \mathbf{J} \rangle - n\right).$$

Let $\mathsf{vec}(A)$ denote the vector formed by stacking all the columns of A. Then we have the following identity:

$$\mathsf{vec}(AXB) = (B^\top \otimes A)\mathsf{vec}(X).$$

It follows that

$$L(X, \gamma) = \mathsf{vec}(X)^\top \left(\mathbf{I} \otimes A^2 + B^2 \otimes \mathbf{I} - 2B \otimes A\right) \mathsf{vec}(X) + \eta^2 \|\mathsf{vec}(X)\|_2^2 - 2\gamma \left(\langle \mathsf{vec}(X), \mathsf{vec}(\mathbf{J}) \rangle - n\right)$$

and hence

$$\frac{\partial L(X,\gamma)}{\partial X} = 2\left(\mathbf{I}\otimes A^2 + B^2\otimes\mathbf{I} - 2B\otimes A + \eta^2\mathbf{I}\otimes\mathbf{I}\right)\mathsf{vec}(X) - 2\gamma\mathsf{vec}(\mathbf{J})$$

Setting the above partial derivative to be 0 yields that

$$\mathsf{vec}(X) = \gamma\left(\mathbf{I}\otimes A^2 + B^2\otimes\mathbf{I} - 2B\otimes A + \eta^2\mathbf{I}\otimes\mathbf{I}\right)^{-1}\mathsf{vec}(\mathbf{J}).$$

Finally, let $A = U\Lambda U^\top$ and $B = VMM^\top$ denote the eigenvalue decomposition. We get that

$$\mathbf{I}\otimes A^2 + B^2\otimes\mathbf{I} - 2B\otimes A + \eta^2\mathbf{I}\otimes\mathbf{I} = (V\otimes U)\left(\mathbf{I}\otimes\Lambda^2 + M^2\otimes\mathbf{I} - 2M\otimes\Lambda + \eta^2\mathbf{I}\otimes\mathbf{I}\right)\left(V^\top\otimes U^\top\right).$$

Therefore,

$$\begin{aligned}\mathsf{vec}(X) &= \gamma\,(V\otimes U)\left(\mathbf{I}\otimes\Lambda^2 + M^2\otimes\mathbf{I} - 2M\otimes\Lambda + \eta^2\mathbf{I}\otimes\mathbf{I}\right)^{-1}\left(V^\top\otimes U^\top\right)\mathsf{vec}(\mathbf{J})\\ &= \sum_{i,j=1}^{n}\frac{\gamma}{(\lambda_i-\mu_j)^2+\eta^2}\underbrace{(v_j\otimes u_i)\left(v_j^\top\otimes u_i^\top\right)\mathsf{vec}(\mathbf{J})}_{\mathsf{vec}\left(u_iu_i^\top\mathbf{J}v_jv_j^\top\right)}.\end{aligned}$$

3.3 Analysis of GRAMPA

In this section, we will prove the following performance guarantee of GRAMPA. Let $\delta = 1-s$ denote the fraction of differed edges in two graphs and q denote the edge probability in one of the two graphs.

Theorem 3.1. *GRAMPA achieves exact recovery with high probability if*

$$nq \gtrsim (\log n)^C \qquad \text{and} \qquad \delta \lesssim (\log n)^{-C}$$

for some absolute constant C.

Note that the performance of GRAMPA exponentially improves the performance of existing spectral matching algorithms, which require $\sigma = \frac{1}{\mathsf{poly}(n)}$ as opposed to $\sigma = O(\frac{1}{\mathsf{polylog}n})$. Also, we will show that the performance of GRAMPA holds universially across a wide range of correlated Wigner model with independent entries, e.g., Gaussian model.

3.3.1 Diagonal dominance structure

Equipped with this optimization point of view, we now explain the typical structure of solutions to the above quadratic programs including the spectral similarity matrix (3.5). It is well known that even the solution to the most stringent relaxation (3.7) is *not* the latent permutation matrix, which can be shown by proving that the KKT conditions cannot be fulfilled with high probability. In fact, a heuristic calculation explains why the solution to (3.7) is far from any permutation matrix: Let us consider the "population version" of (3.7), where the objective function is replaced by its expectation over the random instances A and B. Consider $\pi^* = id$ and the Gaussian Wigner model

$B = A + \sigma Z$, where A and Z are independent GOE matrices with $N(0, \frac{1}{n})$ off-diagonal entries and $N(0, \frac{2}{n})$ diagonal entries. Then the expectation of the objective function is

$$\begin{aligned}\mathbb{E}\|AX - XB\|_F^2 &= \mathbb{E}\|AX\|_F^2 + \mathbb{E}\|XB\|_F^2 - 2\mathbb{E}\langle AX, XA\rangle \\ &= (2+\sigma^2)\frac{n+1}{n}\|X\|_F^2 - \frac{2}{n}\mathsf{Tr}(X)^2 - \frac{2}{n}\langle X, X^\top\rangle.\end{aligned}$$

Hence the population version of the quadratic program (3.7) is

$$\min_{X \geq 0 : X\mathbf{1}=\mathbf{1}, X^\top \mathbf{1}=\mathbf{1}} (2+\sigma^2)(n+1)\|X\|_F^2 - 2\mathsf{Tr}(X)^2 - 2\langle X, X^\top\rangle, \tag{3.9}$$

whose solution[1] is

$$\overline{X} \triangleq \epsilon \mathbf{I} + (1-\epsilon)\frac{1}{n}\mathbf{J}, \qquad \epsilon = \frac{2}{2+(n+1)\sigma^2} \approx \frac{2}{n\sigma^2}. \tag{3.10}$$

This is a convex combination of the true permutation matrix and the center of the Birkhoff polytope $\frac{1}{n}\mathbf{J}$. Therefore, the population solution $\overline{X}$ is in fact a very "flat" matrix, with each entry on the order of $\frac{1}{n}$, and is close to the center of the Birkhoff polytope and far from any of its vertices.

This calculation nevertheless provides us with important structural information about the solution to such a QP relaxation: $\overline{X}$ is *diagonally dominant* for small σ, with diagonals about $2/\sigma^2$ times the off-diagonals. Although the actual solution of the relaxed program (3.7) or (3.8) is not equal to the population solution $\overline{X}$ in expectation, it is reasonable to expect that it inherits the diagonal dominance property in the sense that $\widehat{X}_{i,\pi^*(i)} > \widehat{X}_{ij}$ for all $j \neq \pi^*(i)$, which enables rounding procedures such as (3.2) to succeed.

With this intuition in mind, let us revisit the regularized quadratic program (3.8) whose solution is the spectral similarity matrix (3.5). By a similar calculation, the solution to the population version of (3.8) is given by $\alpha\mathbf{I} + \beta\mathbf{J}$, with $\alpha = \frac{2n^2}{(n(\eta^2+\sigma^2)+\sigma^2)(n(\eta^2+\sigma^2+2)+\sigma^2)} \approx \frac{2}{(\eta^2+\sigma^2)(\eta^2+\sigma^2+2)}$ and $\beta = \frac{n}{n(\eta^2+\sigma^2+2)+\sigma^2} \approx \frac{1}{\eta^2+\sigma^2+2}$, which is diagonally dominant for small σ and η. In turn, the basis of our theoretical guarantee is to establish the diagonal dominance of the actual solution $\widehat{X}$; see Fig. 3.2 for an empirical illustration.

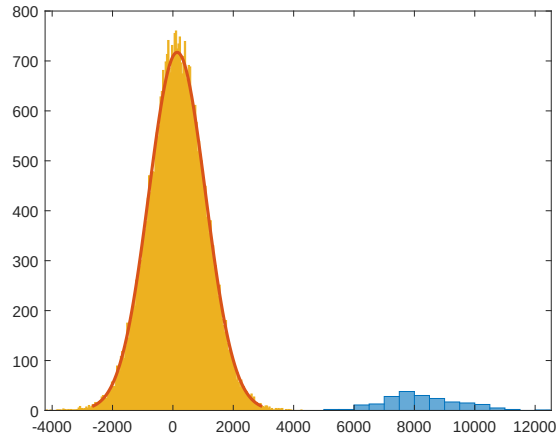
3.3.2 Gaussian Heuristic argument

To build up intuition, let us first consider the noiseless Gaussian case where $A = B$. In this case, we have

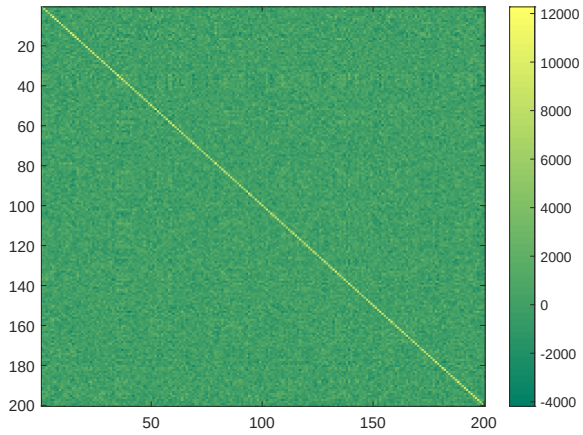
$$X = \sum_{i=1}^n \frac{1}{\eta}(u_i^\top \mathbf{J} u_i) u_i u_i^\top + \sum_{i \neq j} \frac{\eta}{\eta^2 + (\lambda_i - \lambda_j)^2}(u_i^\top \mathbf{J} u_j) u_i u_j^\top. \tag{3.11}$$

We explain why the first term is diagonally dominant, while the second term is a perturbation of smaller order. Central to our proof is the fact that $A \sim \mathsf{GOE}(n)$ is rotationally invariant in law, so that $U = (u_1, \ldots, u_n)$ is uniformly distributed on the orthogonal group and independent of $\lambda_1, \ldots, \lambda_n$. The coordinates of U are approximately independent with distribution $N(0, \frac{1}{n})$.

[1] In fact, (3.10) is the solution to (3.9) even if the constraint is relaxed to $\mathbf{1}^\top X \mathbf{1} = n$.



(a) Histogram of diagonal (blue) and off-diagonal (yellow with a normal fit) entries of $\widehat{X}$.



(b) Heat map of $\widehat{X}$.

Figure 3.2: Diagonal dominance of the GRAMPA similarity matrix X defined by (3.5) or (3.8) for the Gaussian Wigner model $B = A + \sigma Z$ with $n = 200$, $\sigma = 0.05$ and $\eta = 0.01$.

For the first term in (3.11), with high probability $u_i^\top \mathbf{J} u_i = \langle u_i, \mathbf{1}\rangle^2 \approx 1$ for every i. Applying the heuristic $(u_i)_k \overset{\text{i.i.d.}}{\sim} N(0, \frac{1}{n})$, the first term satisfies

$$\sum_{i=1}^{n} \frac{1}{\eta}(u_i^\top \mathbf{J} u_i)(u_i)_k(u_i)_\ell \approx \begin{cases} \frac{1}{\eta} & \text{if } k = \ell \\ \frac{1}{\eta\sqrt{n}} & \text{if } k \neq \ell \end{cases}. \tag{3.12}$$

For the second term in (3.11), Applying the heuristic that g, h are approximately iid $N(0, \frac{1}{n}\mathbf{I})$, we have a Hanson-Wright type bound

$$\begin{aligned} \sum_{i\neq j} \frac{\eta}{\eta^2 + (\lambda_i - \lambda_j)^2}(u_i^\top \mathbf{J} u_j)(u_i)_k(u_j)_\ell &\approx \sqrt{\frac{1}{n^2}\sum_{i\neq j}\Big(\frac{\eta}{\eta^2 + (\lambda_i - \lambda_j)^2}\Big)^2} \\ &\approx \sqrt{\iint \Big(\frac{\eta}{\eta^2 + (x-y)^2}\Big)^2 \rho(x)\rho(y)dxdy} \approx \frac{1}{\sqrt{\eta}}, \end{aligned} \tag{3.13}$$

where the second approximation uses the fact that as $n \to \infty$, the empirical spectral distribution $n^{-1}\sum_{i=1}^n \delta_{\lambda_i}$ of A converges to the Wigner semicircle law with density ρ, and the last step is an elementary computation that holds for any bounded density ρ with bounded support. Combining (3.12)–(3.13) shows that the noiseless solution X in (3.11) is indeed diagonally dominant, with diagonals approximately η^{-1} and off-diagonals at most of the order $\eta^{-1/2}$, omitting logarithmic factors.

The above heuristic argument can be made rigorous when A, B are Gaussian since the eigenvalues and eigenvectors are independent, but hard to extend to Erdős-Rényi graphs. In the next subsection, we present a rigorous proof using resolvents and local alws.

3.3.3 Proof of universality via resolvents

We will establish the diagonal dominance under a general correlated Wigner model. Consider the standardized weighted adjacency matrices A, B where (A_{ij}, B_{ij}) are independent sub-gaussian pairs satisfying

$$\mathbb{E}[A_{ij}] = \mathbb{E}[B_{ij}] = 0, \quad \mathbb{E}[A_{ij}^2] = \mathbb{E}[B_{ij}^2] = \frac{1}{n}, \quad \mathbb{E}[A_{ij}B_{ij}] = \frac{1-\delta}{n}$$

Here we omit some additional growth contions on the higher moments for simplicity.

Our key proof techniques are resolvent and local laws. For a real symmetric matrix A with spectral decomposition (3.1), its resolvent is a complex-valued matrix defined as

$$R_A(z) \triangleq (A - z\mathbf{I})^{-1} = \sum_{i=1}^{n} \frac{1}{\lambda_i - z} u_i u_i^\top, \qquad z \in \mathbb{C}\backslash\mathbb{R}$$

Denote Wigner's semicircle density and its Stieltjes transform by

$$\rho(x) = \frac{1}{2\pi}\sqrt{4-x^2} \cdot \mathbb{1}\{|x| \le 2\} \quad \text{and} \quad m(z) = \int \frac{1}{x-z}\rho(x)dx = \frac{-z+\sqrt{z^2-4}}{2}$$

Classical result in radom matri theory states that the empirical eigenvalue distribution $\frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i}$ converges to ρ. Hence, we expect that

$$\frac{1}{n}\mathsf{Tr}R_A(z) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i - z} \to m(z)$$

There exist stronger results concern the entire matrix $R_A(z)$ when $\operatorname{Im} z$ is not too small. This is called local laws [EKYY13a, EKYY13b]:

- $R_A(z) \approx m(z)\mathbf{I}$ entrywise, that is, whp
$$(R_A(z))_{ij} \approx m(z) \cdot \mathbb{1}\{i = j\} \tag{3.14}$$
- Similarly, row sum and total sum satisfy: whp
$$\sum_j (R_A(z))_{ij} \lesssim \mathsf{polylog}(n) \qquad \sum_{i,j}(R_A(z))_{ij} \approx n \cdot m(z)$$

With these local laws, we are now ready to analyze the GRAMPA similarity matrix X defined by (3.5).

Step 1: Resolvent representation

The first step is to represent X using resolvents.

Lemma 3.1. *Consider symmetric matrices A and B with spectral decompositions (3.1), and suppose that $\|A\| \le 2.5$. Then the matrix X defined in (3.5) admits the following representation*

$$X = \frac{1}{2\pi}\operatorname{Re}\oint_\Gamma R_A(z)\mathbf{J}R_B(z+\mathbf{i}\eta)dz, \tag{3.15}$$

where

$$\Gamma = \{z : |\operatorname{Re} z| = 3 \text{ and } |\operatorname{Im} z| \le \eta/2 \quad \text{or} \quad |\operatorname{Im} z| = \eta/2 \text{ and } |\operatorname{Re} z| \le 3\} \tag{3.16}$$

is the rectangular contour with vertices $\pm 3 \pm \mathbf{i}\eta/2$ (See Fig. 3.3 for an illustration).

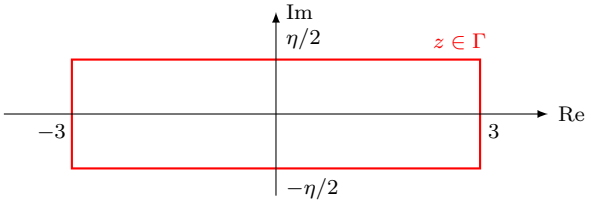


Figure 3.3: Γ encloses $\lambda_1, \ldots, \lambda_n$ but not $\mu_1 - \mathbf{i}\eta, \ldots, \mu_n - \mathbf{i}\eta$.

Proof. By definition,

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

Thus,

$$\begin{aligned}X &= \eta \sum_{i,j} u_i u_i^\top \mathbf{J} \frac{v_j v_j^\top}{(\lambda_i - \mu_j)^2 + \eta^2} \\ &= \operatorname{Im} \sum_i u_i u_i^\top \mathbf{J} R_B(\lambda_i + \mathbf{i}\eta). \end{aligned} \tag{3.17}$$

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

$$-\frac{1}{2\pi\mathbf{i}} \oint_\Gamma \frac{f(z)}{\lambda_i - z} dz = f(\lambda_i). \tag{3.18}$$

Substituting this into (3.17), we obtain

$$X = \operatorname{Im} \sum_i v_i v_i^\top \left(-\frac{1}{2\pi\mathbf{i}} \oint_\Gamma \frac{f(z)}{\lambda_i - z} dz\right) = \frac{1}{2\pi} \operatorname{Re} \oint_\Gamma R_A(z) f(z) dz, \tag{3.19}$$

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

Step 2: Leave-one-out relation

With the resolvent representation of X in Lemma 3.1, we can now apply local laws to bound X entrywise. In particular,

$$X_{k\ell} = \frac{1}{2\pi} \operatorname{Re} \oint_\Gamma \left[e_k^\top R_A(z)\mathbf{1}\right]\left[\mathbf{1}^\top R_B(z + \mathbf{i}\eta) e_\ell\right] dz$$

Note that one may attempt to directly apply (3.3.3) to bound the row sums $\mathbf{e}_k^\top R_A(z)\mathbf{1}$ and $\mathbf{e}_\ell^\top R_B(z + \mathbf{i}\eta)\mathbf{1}$. 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 X 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 and the leave-one-out analysis.

To proceed, let us recall the Schur complement identity.

Lemma 3.2 (Schur complement identity)**.** *Provided D is square and invertible,*

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \implies M^{-1} = \begin{bmatrix} S & -SBD^{-1} \\ -D^{-1}CS & D^{-1} + D^{-1}CSBD^{-1} \end{bmatrix}$$

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

The utility of this result (both theoretically and algorithmically) is to reduce matrix inverse to inverting smaller matrices. Now, let us apply Schur complement identity to study X_{11}. Write

$$A = \begin{pmatrix} a_{11} & a_1^\top \\ a_1 & A^{(1)} \end{pmatrix} \qquad R_A(z) = \begin{pmatrix} R_{A,11} & R_{A,1*} \\ R_{A,*1} & R_{A,**} \end{pmatrix}$$

By the Schur-complement formula

$$\begin{aligned} R_{A,1*}(z) &= -R_{A,11}(z) \cdot a_1^\top (A^{(1)} - z\mathbf{I})^{-1} \\ &\approx -m(z) \cdot a_1^\top R_{A^{(1)}}(z) \end{aligned},$$

where we apply the approximation (3.14). It follows that

$$\mathbf{e}_1 R_{A,1*}(z)\mathbf{1} \approx m(z) - m(z) \cdot a_1^\top R_{A^{(1)}}(z)\mathbf{1}.$$

Writing a similar expression for B, we get

$$\begin{aligned} X_{11} &\approx \frac{1}{2\pi} \operatorname{Re} \oint_\Gamma \Big(m(z) - m(z) \cdot a_1^\top R_{A^{(1)}}(z)\mathbf{1} \Big) \Big(m(z+\mathbf{i}\eta) - m(z+\mathbf{i}\eta) \cdot \mathbf{1}^\top R_{B^{(1)}}(z+\mathbf{i}\eta) b_1 \Big)\, dz \\ &\approx \frac{1}{2\pi} \operatorname{Re} a_1^\top \underbrace{\oint_\Gamma m(z)m(z+\mathbf{i}\eta) R_{A^{(1)}}(z)\mathbf{1}\mathbf{1}^\top R_{B^{(1)}}(z+\mathbf{i}\eta) dz}_{M} b_1, \end{aligned}$$

where in the last approximation we only keep the bilinear term as it captures the desired correlation. In particular, note that a_1 and b_1 are correlated, and they are independent of $A^{(1)}$ and $B^{(1)}$ and hence M. Thus, we can condition on M and apply Hanson-wright type bound.

Similarly, for off-diagonals, we get (by leaving two out and only keeping the bilinear term) that

$$X_{12} \approx \frac{1}{2\pi} \operatorname{Re} a_1^\top \underbrace{\oint_\Gamma m(z)m(z+\mathbf{i}\eta) R_{A^{(12)}}(z)\mathbf{1}\mathbf{1}^\top R_{B^{(12)}}(z+\mathbf{i}\eta) dz}_{N} b_2.$$

where $A^{(12)}$ and $B^{(12)}$ are the same as A and B by deleting the first two rows and columns. Note that here a_1 and b_2 are uncorrelated, and again they are independent of N. Thus, again we can condition on N and apply Hanson-wright type bound.

Step 3: Separating signal from noise

As aforementioned, we will apply concentration of bilinear forms. Recall that a_1, b_1 are correlated and a_1, b_2 are independent, s.t.

$$\mathbb{E}[a_1 b_1^\top] = \frac{1-\delta}{n}\mathbf{I}, \quad \mathbb{E}[a_1 b_2^\top] = 0.$$

So we expect, for deterministic matrix M,

$$a_1^\top M b_1 \approx \frac{1-\delta}{n}\mathsf{Tr}(M) \quad a_1^\top M b_2 \approx N\left(0, \frac{1}{n^2}\|M\|_{\mathrm{F}}^2\right)$$

More precisely, applying Hanson-wright type bound, we get that whp

$$\left|a_1^\top M b_1 - \frac{1-\delta}{n}\mathsf{Tr}(M)\right| \le \frac{\mathsf{polylog}(n)\,\|M\|_{\mathrm{F}}}{n}, \quad \left|a_1^\top M b_2\right| \le \frac{\mathsf{polylog}(n)\,\|M\|_{\mathrm{F}}}{n}.$$

Thus, applying the above concentration bound for the bilinear forms, we get that

$$X_{11} \approx \frac{1}{2\pi}\,\mathrm{Re}\,\frac{1-\delta}{n}\mathsf{Tr}(M), \quad X_{12} \lesssim \frac{\mathsf{polylog}(n)}{n}\,\|N\|_{\mathrm{F}}\,. \tag{3.20}$$

Step 4: Proof of diagonal dominance

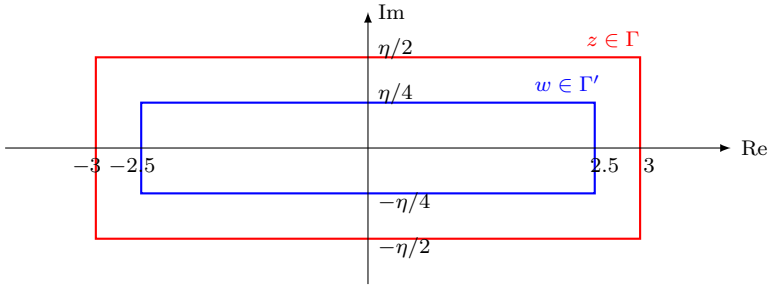
It remains to bound $\mathsf{Tr}(M)$ and $\|N\|_{\mathrm{F}}$. To bound $\mathsf{Tr}(M)$, applying the facts:

- $R_B(z+\mathbf{i}\eta)R_A(z) = \frac{1}{\mathbf{i}\eta}(R_B(z+\mathbf{i}\eta) - R_A(z) - R_B(z+\mathbf{i}\eta)(A-B)R_A(z))$
- whp, $\mathbf{1}^\top R_A(z)\mathbf{1} \approx nm(z)$, $\|A-B\| \lesssim \sqrt{\delta}$ and $\|R_A(z)\mathbf{1}\| \lesssim \sqrt{\frac{n}{\eta}}$,

we get

$$\begin{aligned}
&\frac{1}{n}\,\mathrm{Re}\,\mathsf{Tr}(M)\\
=&\frac{1}{n}\,\mathrm{Re}\oint_\Gamma dz\; m(z)m(z+\mathbf{i}\eta)\mathsf{Tr}\big[R_A(z)\mathbf{J}R_B(z+\mathbf{i}\eta)\big]\\
=&\frac{1}{n}\,\mathrm{Re}\oint_\Gamma dz\; m(z)m(z+\mathbf{i}\eta)\mathbf{1}^\top R_B(z+\mathbf{i}\eta)R_A(z)\mathbf{1}\\
=&\frac{1}{n\eta}\,\mathrm{Im}\oint_\Gamma dz\; m(z)m(z+\mathbf{i}\eta)\mathbf{1}^\top\left(R_B(z+\mathbf{i}\eta) - R_A(z) - R_B(z+\mathbf{i}\eta)(A-B)R_A(z)\right)\mathbf{1}\\
\le&\frac{1}{\eta}\underbrace{\mathrm{Im}\oint_\Gamma dz\; m(z)m(z+\mathbf{i}\eta)(m(z+\mathbf{i}\eta)-m(z))}_{2\pi+o_\eta(1)} + O\left(\frac{\sqrt{\delta}}{\eta^2}\right).
\end{aligned}$$

To bound $\|N\|_{\mathrm{F}}$, we first deform the contour as follows:



Then

$$\begin{aligned} N &= \oint_{\Gamma} m(z)m(z+\mathbf{i}\eta)R_A(z)\mathbf{J}R_B(z+\mathbf{i}\eta)dz \\ &= \oint_{\Gamma'} m(w)m(w+\mathbf{i}\eta)R_A(w)\mathbf{J}R_B(w+\mathbf{i}\eta)dw \end{aligned}$$

Applying the facts

- $\overline{m(z)} = m(\overline{z})$, $R_A(z)^* = R_A(\overline{z})$
- $R_A(z)R_A(w) = \frac{R_A(z)-R_A(w)}{z-w}$
- $|m(z)| \lesssim 1$ and $|\mathbf{1}^\top R_A(z)\mathbf{1}| \lesssim n$

we get

$$\begin{aligned} &\|M\|_{\mathrm{F}}^2 = \mathsf{Tr}(MM^*) \\ &= \oint_{\Gamma} dz \oint_{\Gamma'} dw\; m(z)m(z+\mathbf{i}\eta)m(\overline{w})m(\overline{w}-\mathbf{i}\eta)\mathsf{Tr}\big[R_A(z)\mathbf{1}\mathbf{1}^\top R_B(z+\mathbf{i}\eta)R_B(\overline{w}-\mathbf{i}\eta)\mathbf{1}\mathbf{1}^\top R_A(\overline{w})\big] \\ &= -\oint_{\Gamma} dz \oint_{\Gamma'} dw\; m(z)m(z+\mathbf{i}\eta)m(w)m(w-\mathbf{i}\eta)\mathbf{1}^\top R_A(w)R_A(z)\mathbf{1}\mathbf{1}^\top R_B(z+\mathbf{i}\eta)R_B(w-\mathbf{i}\eta)\mathbf{1} \\ &= -\oint_{\Gamma} dz \oint_{\Gamma'} dw\; m(z)m(z+\mathbf{i}\eta)m(w)m(w-\mathbf{i}\eta)\frac{\mathbf{1}^\top(R_A(z)-R_A(w))\mathbf{1}}{z-w}\frac{\mathbf{1}^\top(R_B(z+\mathbf{i}\eta)-R_B(w-\mathbf{i}\eta))\mathbf{1}}{z+\mathbf{i}\eta-(w-\mathbf{i}\eta)} \\ &\lesssim n^2 \underbrace{\oint_{\Gamma} dz \oint_{\Gamma'} dw \frac{1}{|z-w|}\frac{1}{|z-w+2\mathbf{i}\eta|}}_{\asymp 1/\eta}, \end{aligned}$$

where the last integral is bounded by considering case (i): either z or w is on the vertical strips of $\Gamma \cup \Gamma'$ and case (ii): both z and w are on the horizontal strips of $\Gamma \cup \Gamma'$.

Plugging the above bounds back to (3.20), we arrive at

$$X_{11} \approx \frac{1-\delta}{\eta} + \frac{\sqrt{\delta}}{\eta^2}, \quad X_{12} \lesssim \frac{\mathsf{polylog}(n)}{\sqrt{\eta}}.$$

Applying this together with a union bound for every $X_{k\ell}$ shows that X is diagonally dominant when

$$\sqrt{\delta} \lesssim \eta \lesssim 1/\mathsf{polylog}(n)$$

3.4 Extensions and open problems

For the dense graphs $q = \Theta(1)$, we can improve the noise tolerance to $\delta \lesssim (\log n)^{-4+\epsilon}$. For Gaussian weighted graphs, we can improve the noise teolerance to $\delta \lesssim (\log n)^{-2}$ by direct analysis following our heuristic argument above. We can also obtain a similar result for a tighter QP relaxation

$$\arg \max_{X\colon X\mathbf{1}=\mathbf{1}} \|AX - XB\|_F^2 + \eta^2 \|X\|_F^2$$

using the technique of resolvents and local laws. A fundamental open problem is to analyze the tigher relaxations

$$\arg \max_{X\colon X\mathbf{1}=\mathbf{1}, X^\top \mathbf{1}=\mathbf{1}} \|AX - XB\|_F^2 + \eta^2 \|X\|_F^2$$

and

$$\arg \max_{X\colon X\geq 0, X\mathbf{1}=\mathbf{1}, X^\top \mathbf{1}=\mathbf{1}} \|AX - XB\|_F^2$$

We have also extended the results to matching bipartite graphs. Another interesting future direction is to study other andom graphs ensembles, e.g., random geometric graphs.

References

[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.

[FMWX19a] Zhou Fan, Cheng Mao, Yihong Wu, and Jiaming Xu. Spectral graph matching and regularized quadratic relaxations I: The Gaussian model. *arxiv preprint arXiv:1907.08880*, 2019.

[FMWX19b] Zhou Fan, Cheng Mao, Yihong Wu, and Jiaming Xu. Spectral graph matching and regularized quadratic relaxations II: Erdős-Rényi graphs and universality. *arxiv preprint arXiv:1907.08883*, 2019.

[Ume88] Shinji Umeyama. An eigendecomposition approach to weighted graph matching problems. *IEEE transactions on pattern analysis and machine intelligence*, 10(5):695–703, 1988.