

# Phase Transition in Convex Relaxations for Graph Alignment

**Laurent Massoulié**

*INRIA, DI/ENS, PSL Research University, Paris, France*

LAURENT.MASSOULIE@INRIA.FR

**Sushil Mahavir Varma**

*Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan, USA.*

SUSHILV@UMICH.EDU

**Louis Vassaux**

*INRIA, DI/ENS, PSL Research University, Paris, France*

LOUIS.VASSAUX@INRIA.FR

**Irène Waldspurger**

*CNRS, INRIA, Université Paris Dauphine, Paris, France*

WALDSPURGER@CEREMADE.DAUPHINE.FR

**Editors:** Steve Hanneke and Tor Lattimore

## Abstract

We study the graph alignment problem for correlated Gaussian Orthogonal Ensemble (GOE) matrices, where the goal is to recover a hidden vertex permutation given two correlated symmetric Gaussian matrices  $(A, B)$  with correlation  $1/\sqrt{1 + \sigma^2}$ . While the maximum likelihood estimator is information-theoretically optimal, its computation, which reduces to a quadratic assignment problem, is intractable. Motivated by this, we analyze convex relaxations based on minimizing  $\|AX - XB\|_F$  over the set of doubly stochastic matrices and the unit hypercube. We show that when the correlation parameter satisfies  $\sigma = o(n^{-1/2}/\log^4 n)$ , the solution of either relaxation ( $X^*$ ) concentrates around the ground-truth permutation matrix  $(\Pi^*)$ , i.e.,  $\|X^* - \Pi^*\|_F^2 = o(n)$ , implying recovery of all but a vanishing fraction of vertices after simple post-processing. Combined with existing lower bounds, our results precisely characterize that  $\|X^* - \Pi^*\|_F^2$  transitions from  $o(n)$  for  $\sigma = \tilde{o}(n^{-1/2})$  to  $\Omega(n)$  for  $\sigma = \tilde{\Omega}(n^{-1/2})$ . In doing so, our analysis significantly tightens prior results and extends them beyond doubly stochastic relaxations.

**Keywords:** GOE Alignment, Phase Transition, Convex Relaxation, Birkhoff polytope, Unit Hypercube

## 1. Introduction

The graph alignment problem is defined as finding a mapping between the vertices of two undirected graphs  $G_1$  and  $G_2$  such that the edge overlap is maximized. This problem is popular because it arises in widespread applications, e.g., network de-anonymization [Narayanan and Shmatikov \(2008\)](#), computational biology [Singh et al. \(2008\)](#), pattern recognition [Conte et al. \(2004\)](#), etc. More formally, let  $A, B \in \mathbb{R}^{n \times n}$  be the adjacency matrices (possibly weighted) of  $G_1$  and  $G_2$  respectively. The objective is to find a permutation  $\pi : [n] \rightarrow [n]$  such that the edge overlap  $\sum_{i,j \in [n]} A_{i,j} B_{\pi(i),\pi(j)}$  is maximized. We consider the problem of graph alignment, where the inputs are sampled as correlated Gaussian random matrices  $A$  and  $B$ . In particular, we first sample two independent and identically distributed (i.i.d.) Gaussian Wigner matrices  $A$  and  $Z$ . That is,  $Z$  is a symmetric matrix,  $\{Z_{i,j}\}_{i \leq j}$  are independent,  $Z_{i,i}$  are i.i.d. Gaussian with 0 mean and  $2/n$  variance for  $i \in [n]$ , and  $Z_{i,j}$  are i.i.d. Gaussian with 0 mean and  $1/n$  variance for  $i < j \in [n]$ . To define  $B$ , we first sample  $\pi^*$  uniformly at random from the set  $\mathcal{S}_n$  of permutations of  $[n]$ . We finally let

$$B' = A + \sigma Z,$$

and define

$$B_{i,j} = B'_{\pi^*(i), \pi^*(j)}, \quad i \leq j \in [n].$$

The objective is to then infer the hidden permutation  $\pi^*$ , given the correlated Gaussian matrices  $(A, B)$ . It is known that one can formulate the maximum likelihood estimation as a quadratic assignment problem (QAP) given by  $\tilde{\Pi}^* = \arg \min_{\Pi \in \mathcal{P}_n} \|A\Pi - \Pi B\|_F$ , where  $\mathcal{P}_n$  is the set of all permutation matrices defined as

$$\mathcal{P}_n = \left\{ \Pi \in \{0, 1\}^{n \times n} : \sum_{i=1}^n \Pi_{i,j} = 1 \text{ for all } j, \sum_{j=1}^n \Pi_{i,j} = 1 \text{ for all } i \right\}.$$

The papers (Ganassali, 2022; Wu et al., 2022) showed that whenever  $\sigma^2 \leq n/(4 + \varepsilon) \log n$ , we have  $\tilde{\Pi}^* = \Pi^*$  w.h.p., where  $\Pi^*$  is the permutation matrix corresponding to the hidden permutation  $\pi^*$ . While the above QAP guarantees to recover  $\pi^*$  w.h.p., QAPs are known to be NP-hard to even approximately solve (Makarychev et al., 2010). Thus, one popular approach in the literature is to consider convex relaxations of the QAP. That is, minimizing  $\|AX - XB\|_F$  over some convex set  $\mathcal{C}_n$ :

$$X^* \in \underset{X \in \mathcal{C}_n}{\text{Arg min}} \|AX - XB\|_F. \quad (1)$$

A popular choice for the convex set  $\mathcal{C}_n$  consists in letting  $\mathcal{C}_n = \mathcal{B}_n$ , that is the set of doubly stochastic matrices, also known as the Birkhoff polytope. By the Birkhoff-von Neumann theorem,  $\mathcal{B}_n$  is the convex hull of permutation matrices. Other choices for  $\mathcal{C}_n$  are possible; for instance Araya and Tyagi (2024) consider letting  $\mathcal{C}_n = \Delta_n$ , that is the simplex in  $\mathbb{R}^{n^2}$ , defined as

$$\Delta_n = \left\{ X \in \mathbb{R}_+^{n^2} : \sum_{i,j \in [n]} X_{i,j} = n \right\}.$$

Finally, we introduce a third option,  $\mathcal{C}_n = \mathcal{H}_n$ , where  $\mathcal{H}_n$  is the constrained hypercube, defined as:

$$\mathcal{H}_n := \left\{ X \in [0, 1]^{n^2} : \sum_{i,j \in [n]} X_{i,j} = n \right\}. \quad (2)$$

It clearly holds that  $\mathcal{B}_n \subset \mathcal{H}_n \subset \Delta_n$ . The objective is then to establish if the solution of the convex relaxation  $X^*$  can be used to infer the hidden permutation  $\Pi^*$ .

### 1.1. Related Work and Main Contribution

The related work on GOE alignment is summarized in Table 1. Surprisingly, only a few theoretical results are known about the performance of such convex relaxations, despite strong empirical performance Aflalo et al. (2015); Dym et al. (2017); Lyzinski et al. (2015). Indeed, Araya and Tyagi (2024) considered the noise-free case where  $\sigma = 0$ , for which they show that for both relaxations to  $\mathcal{B}_n$  and to  $\Delta_n$ , the unique minimizer  $X^*$  is the ground-truth permutation matrix  $X^* = \Pi^*$  w.h.p., for which the objective function  $\|AX - XB\|_F$  trivially equals zero. More recently, Varma et al. (2025) showed that for relaxation to  $\mathcal{B}_n$ , given any  $\delta > 0$ ,  $X^*$  verifies with high probability:

$$\begin{aligned} \sigma = n^{-1-\delta} &\Rightarrow \|X^* - \Pi^*\|_F^2 = o(n); \\ \sigma = n^{-1/2+\delta} &\Rightarrow \|X^* - \Pi^*\|_F^2 = \Omega(n). \end{aligned} \quad (3)$$

In the first case, property  $\|X^* - \Pi^*\|_F^2 = o(n)$  readily guarantees that a very crude post-processing of  $X^*$ , for instance letting

$$\forall i \in [n], \hat{\pi}(i) \in \operatorname{Arg} \max_{j \in [n]} X_{i,j}^*, \quad (4)$$

correctly recovers all but a vanishing fraction of the entries of  $\pi^*$ . In this paper, we tighten the results of Varma et al. (2025) by showing that the positive result in (3) holds for  $\sigma = \tilde{o}(n^{-1/2})$ . Combining with the negative result in (3), we now locate the threshold at which  $\|X^* - \Pi^*\|_F^2$  goes from  $o(n)$  to  $\Omega(n)$ , up to sub-polynomial factors. We note that a heuristic prediction of this  $n^{-1/2}$  threshold already appears in Section 1.4 of Fan et al. (2023); our contribution is a rigorous proof of this transition for the actual (random) minimizer of the unregularized relaxations. Lastly, our positive result also holds for the hypercube relaxation, i.e., we show that  $\|X^* - \Pi^*\|_F^2 = o(n)$  for  $\sigma = \tilde{o}(n^{-1/2})$ , for  $X^*$  defined in (1) with  $\mathcal{C}_n = \mathcal{H}_n$ . Such a tightening needed several non-trivial innovations in the proof beyond Varma et al. (2025), which we discuss in Section 3.3. Another related work is Fan et al. (2023), which considers the following regularized convex relaxation problem, coined GRAMPA:

$$\operatorname{Arg} \min_{X:1^T X \mathbf{1} = n} \|AX - XB\|_F^2 + \eta \|X\|_F^2, \quad \text{for some } \eta > 0. \quad (5)$$

Note that the optimization problem above is a further relaxation of the QAP with the addition of a quadratic regularization term. In Fan et al. (2023), the authors show that (5) recovers  $\Pi^*$  whenever  $\sigma = O(1/\log n)$ . Such a regularized relaxation is considered in Fan et al. (2023) mainly due to its theoretical tractability. Specifically, Fan et al. (2023) drop the non-negativity constraints of the Birkhoff relaxation for tractability, and the quadratic regularization  $\eta \|X\|_F^2$  is needed for good performance in this looser setting; they also consider alternative regularizations (e.g., the row-sum constraint  $X\mathbf{1} = \mathbf{1}$ ) with better empirical performance. Empirically, the regularized relaxation (5) trades off accuracy for speed: while it attains lower recovery accuracy than the unregularized simplex and Birkhoff relaxations, it is much faster to solve (Fan et al., 2023; Varma et al., 2025). Lastly, we refer the readers to Gaudio et al. (2025); Ganassali et al. (2024); Ding et al. (2021); Mao et al. (2023) and the references within for analysis beyond the GOE setting.

Paper	Algorithm Type	Algorithm Name	Noise ( $\sigma$ )
Ganassali et al. (2022)	Top Eigenvector Alignment	EIG1	$\Theta(n^{-7/6})$
Fan et al. (2023)	Regularized Relaxation	GRAMPA (5)	$O(1/\log n)$
Araya and Tyagi (2024)	Simplex Relaxation	(1) with $\mathcal{C}_n = \Delta_n$	0
Varma et al. (2025)	Birkhoff Relaxation	(1) with $\mathcal{C}_n = \mathcal{B}_n$	$\tilde{O}(n^{-1})$
Our Work	Hypercube/Birkhoff Relaxation	(1) with $\mathcal{C}_n = \mathcal{H}_n$	$\tilde{O}(n^{-1/2})$

Table 1: Comparison of the spectral and optimization-based algorithms for graph alignment on the Gaussian Wigner Model: sufficient conditions for success.

## 2. Main Result

We now state the main result of this section, a tightening of the positive result in (3):

**Theorem 1** *For correlated GOE matrices as above, if noise parameter  $\sigma$  satisfies*

$$\sigma = o\left(\frac{n^{-1/2}}{\ln(n)^4}\right), \quad (6)$$

*one then has with high probability, for both the relaxation to the Birkhoff polytope  $\mathcal{B}_n$  and to the constrained hypercube  $\mathcal{H}_n$  defined in (2),  $\|X^* - \Pi^*\|_F^2 = O(\sigma n^{3/2} \ln(n)^4) = o(n)$ .*

Before embarking on the proof, we remark that together with the negative result of [Varma et al. \(2025\)](#) recalled in (3), Theorem 1 locates the threshold at which  $\|X^* - \Pi^*\|_F^2$  goes from  $o(n)$  to  $\Omega(n)$  at  $\sigma = n^{-1/2}$ , up to sub-polynomial factors. The above result guarantees that the output of the post-processing ( $\hat{\pi}$ ) defined in (4) correctly recovers all but a vanishing fraction of the entries of  $\pi^*$  whenever  $\sigma$  satisfies (6). Formally, define overlap  $\text{ov}(\hat{\pi}, \pi^*)$  as the fraction of indices mapped identically by both permutations, i.e.,

$$\text{ov}(\hat{\pi}, \pi^*) := \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\pi^*(i) = \hat{\pi}(i)}.$$

Then,  $\text{ov}(\hat{\pi}, \pi^*) = 1 - o(1)$  w.h.p. Indeed, fix any  $i$  with  $\hat{\pi}(i) \neq \pi^*(i)$ . Since  $\hat{\pi}(i) \in \text{Arg max}_{j \in [n]} X_{i,j}^*$ , we have  $X_{i, \hat{\pi}(i)}^* \geq X_{i, \pi^*(i)}^*$ , so either  $X_{i, \pi^*(i)}^* \leq 1/2$  or  $X_{i, \hat{\pi}(i)}^* \geq X_{i, \pi^*(i)}^* > 1/2$ . In the first case the term  $(X_{i, \pi^*(i)}^* - 1)^2 \geq 1/4$ ; in the second the off-diagonal term  $(X_{i, \hat{\pi}(i)}^*)^2 > 1/4$  (recall  $\hat{\pi}(i) \neq \pi^*(i)$ ). In either case, row  $i$  contributes at least  $1/4$  to  $\|X^* - \Pi^*\|_F^2$ , so

$$\|X^* - \Pi^*\|_F^2 \geq \frac{1}{4} \sum_{i \in [n]} \mathbb{1}_{\hat{\pi}(i) \neq \pi^*(i)},$$

hence the lower bound  $1 - o(1)$  on the overlap when  $\|X^* - \Pi^*\|_F^2 = o(n)$ . We further remark that post-processing (4) is very crude, and that a more refined post-processing could be proposed. In particular one could let  $\hat{\pi}$  be chosen as the permutation in  $\mathcal{S}_n$  that solves

$$\max_{\pi \in \mathcal{S}_n} \sum_{i \in [n]} X_{i\pi(i)}^*. \quad (7)$$

This is known as the Linear Assignment Problem (LAP), or maximum weight bipartite matching problem. Due to ([Varma et al., 2025](#), Corollary 3), we are guaranteed that the overlap of the output of the LAP  $\hat{\pi}$  with  $\pi^*$  is at least  $1 - o(1)$  w.h.p. whenever  $\sigma$  satisfies (6), in light of Theorem 1. Empirically LAP appears to achieve overlap  $1 - o(1)$  for  $\sigma$  well above  $n^{-1/2}$ , but there is so far no theoretical understanding of LAP's performance in the present setup. In fact it is not possible to tighten the Frobenius guarantees since [Varma et al. \(2025\)](#) show that  $\|X^* - \Pi^*\|_F^2 = \Omega(n)$  for  $\sigma$  larger than  $n^{-1/2}$ . This gap suggests that the property governing recovery is not Frobenius proximity but a closely related one: diagonal dominance of the minimizer, i.e.,  $X_{i\pi^*(i)}^* > X_{ij}^*$  for all  $j \neq \pi^*(i)$ . Either property makes the post-processing above succeed; the difference is their reach, since the relaxations remain *empirically* diagonally dominant—and hence keep recovering  $\pi^*$ —for substantially larger  $\sigma$  than the Frobenius guarantee allows. Our analysis controls the unregularized minimizer in Frobenius norm but says nothing about its diagonal dominance, and hence does not explain this empirical success for larger  $\sigma$ . Relatedly, [Fan et al. \(2023\)](#) do establish diagonal

dominance, though only for their *regularized* minimizer (up to  $\sigma = O(1/\log n)$ ), where it enables recovery by rounding. Establishing it for the *unregularized* minimizer thus remains an important open problem.

Our contribution is to pin down a *structural* phase transition for the Birkhoff and hypercube relaxations: the point (up to sub-polynomial factors) at which the minimizer transitions from  $o(n)$  to  $\Omega(n)$  squared Frobenius distance from  $\Pi^*$ . While this threshold does not coincide with the information-theoretic (Ganassali, 2022; Wu et al., 2022) or algorithmic frontiers (Fan et al., 2023), which are shown to be much larger than  $\Theta(n^{-1/2})$ , it is an important change point in the behavior of the unregularized Birkhoff relaxation, above which different proof techniques appear to be required to establish any form of correctness guarantee. Since understanding the (most natural) Birkhoff relaxation is known to be difficult—Fan et al. (2023) describe it as “a challenging task yet to be accomplished”—we believe that precisely characterizing its different operating regimes is of independent interest. Moreover, our techniques (upcoming Lemmas 2 and 3) are not regime-specific and may be useful at larger noise levels.

### 3. Proof of Theorem 1

Thanks to invariance of Frobenius norm  $\|\cdot\|_F$  and of both  $\mathcal{B}_n$  and  $\mathcal{H}_n$  by left- or right-multiplication by a permutation matrix, we can assume without loss of generality that  $\pi^*$  is the identity. We thus aim to prove

$$\|X^* - I\|_F^2 = o(n), \quad (8)$$

where  $X^*$  minimizes

$$\|AX - X(A + \sigma Z)\|_F \quad (9)$$

over  $X \in \mathcal{C}_n$  where  $\mathcal{C}_n$  could be either  $\mathcal{H}_n$  or  $\mathcal{B}_n$ . To ensure (8), it is sufficient to show that  $\text{Tr}(X^*)$  is at least  $n(1 - o(1))$  w.h.p. In particular, as  $X^* \in \mathcal{H}_n$ , we have

$$\|X^* - I\|_F^2 = \sum_{i,j \in [n]: i \neq j} (X_{i,j}^*)^2 + \sum_{i=1}^n (X_{i,i}^* - 1)^2 = \|X^*\|_F^2 + n - 2 \text{Tr}(X^*) \leq 2(n - \text{Tr}(X^*)), \quad (10)$$

where the last inequality follows as  $X^* \in \mathcal{H}_n$ , which implies  $\|X^*\|_F \leq \sqrt{n}$ . The above inequality ascertains that  $\text{Tr}(X^*) = n(1 - o(1))$  implies  $\|X^* - I\|_F^2 = o(n)$ . So, we focus on establishing the former. Another key ingredient is the following upper bound on the objective (9):

$$\|AX^* - X^*B\|_F \leq \|A - B\|_F = \sigma \|Z\|_F \stackrel{(*)}{=} O(\sigma\sqrt{n}), \quad (11)$$

where  $(*)$  holds as  $\|Z\|_F \leq c\sqrt{n}$  w.p.  $1 - o(1)$  for  $c > 0$  large enough. We now present two different proofs for the assertion that (11) implies  $\text{Tr}(X^*)$  is close to  $n$ .

#### 3.1. Direct (Primal) Approach

In this approach, we establish the following lower bound on the objective of (9):

$$\forall X \in \mathcal{C}_n, \|AX - XB\|_F \geq \frac{1}{O(n \log^4(n))} (n - \text{Tr}(X)),$$

which then combined with (11) and (6) implies  $\text{Tr}(X^*) \geq n(1 - o(1))$ . We present the complete proof below.

**Proof** [Proof of Theorem 1] The proof is mainly divided into the following four parts.

Isolating dependence on  $B$ : The objective (9) admits the lower bounds

$$\|AX - X(A + \sigma Z)\|_F \geq \|AX - XA\|_F - \sigma\|XZ\|_F \geq \|AX - XA\|_F - \sigma\|X\|_F\|Z\|_2,$$

where  $\|Z\|_2$  denotes the operator norm of matrix  $Z$ , and we used the classical inequality  $\|XZ\|_F \leq \|Z\|_2\|X\|_F$ , valid for any two matrices  $X, Z$ . By classical results on spectra of random matrices Anderson et al. (2010), with high probability matrix  $Z$  verifies  $\|Z\|_2 \leq 2 + o(1) = O(1)$ . Also, any matrix  $X$  in  $\mathcal{C}_n$  verifies  $\|X\|_F \leq \sqrt{n}$ , so that

$$\|AX - X(A + \sigma Z)\|_F \geq \|AX - XA\|_F - 3\sigma\sqrt{n}. \quad (12)$$

*Lower bound by a linear function*: Let  $\lambda_1 \leq \dots \leq \lambda_n$  denote the spectrum of  $A$ , and  $u_1, \dots, u_n$  an associated orthonormal basis of eigenvectors of  $A$ . The matrix  $(u_1 | \dots | u_n)$  is uniformly distributed over the orthogonal group  $O_n$ . The goal of this step is to lower bound  $\|AX - XA\|_F^2$  in terms of  $n - \text{Tr}(X)$ , so that the objective is large whenever  $X$  is far from  $I$ . The natural approach—bounding by the smallest eigenvalue of  $A \otimes I - I \otimes A$ , namely  $\min_{i,j} (\lambda_i - \lambda_j)^2$ —fails, since  $\lambda_i - \lambda_j$  can be very small when  $|i - j|$  is small. We therefore project  $X - I$  onto the eigenvector pairs  $u_i u_j^\top$  with  $|i - j| \geq \epsilon n$ , so that the minimum eigenvalue gap restricted to this projected subspace can be controlled. More formally, let  $\epsilon > 0$  be chosen, possibly dependent on  $n$ . Let

$$C_\epsilon := \min_{i,j \in [n]: |i-j| \geq \epsilon n} (\lambda_i - \lambda_j)^2. \quad (13)$$

Denote by  $P_\epsilon$  the orthonormal projection defined by

$$P_\epsilon(M) := \sum_{i,j: |i-j| < \epsilon n} \langle u_i u_j^\top, M \rangle u_i u_j^\top. \quad (14)$$

Write then

$$\begin{aligned} \|AX - XA\|_F^2 &= \|A(X - I) - (X - I)A\|_F^2 \\ &= \sum_{i,j \in [n]} (\lambda_i - \lambda_j)^2 \langle u_i u_j^\top, X - I \rangle^2 \\ &\geq C_\epsilon \|(I - P_\epsilon)(X - I)\|_F^2 \\ &\stackrel{(a)}{\geq} C_\epsilon \frac{\langle (I - P_\epsilon)(J), X - I \rangle^2}{\|J\|_F^2} \\ &\stackrel{(b)}{=} C_\epsilon n^{-2} \langle (I - P_\epsilon)(J), X - I \rangle^2 \\ &\stackrel{(c)}{=} C_\epsilon n^{-2} \langle P_\epsilon(J), I - X \rangle^2, \end{aligned} \quad (15)$$

where we introduced the all-ones matrix  $J$  and used Cauchy-Schwarz inequality in (a), used  $\|J\|_F = n$  in (b), and used  $\langle J, X \rangle = \langle J, I \rangle = n$  in (c). For the lower bound (15) to reflect the trace deficiency  $n - \text{Tr}(X)$ , we need  $Y := P_\epsilon(J)$  to be close to  $I$ . This is governed by  $\epsilon$ , since  $Y$  interpolates between  $J$  for large  $\epsilon$  and  $I$  for small  $\epsilon$ . To see this, write  $J = ee^\top = \sum_{i,j \in [n]} (u_i^\top e)(u_j^\top e) u_i u_j^\top$ . At  $\epsilon = 1$ ,  $Y$  sums over all pairs and equals the matrix  $J$ ; at  $\epsilon = 1/n$ ,  $Y$  retains only the diagonal terms  $\sum_i (u_i^\top e)^2 u_i u_i^\top$ , which can be shown to be component-wise close to  $I = \sum_i u_i u_i^\top$  since the eigenvectors  $u_i$  are delocalized (being uniform on  $\mathcal{S}^{n-1}$ ). We thus want  $\epsilon$  small enough that  $Y \approx I$ ,

yet large enough that the gap  $C_\epsilon = \min_{|i-j| \geq \epsilon n} (\lambda_i - \lambda_j)^2$  does not vanish; the next step balances this trade-off.

*Analysis at extreme points of  $\mathcal{C}_n$ :* The extremal elements of  $\mathcal{C}_n$  are its elements whose entries belong to  $\{0, 1\}$ . For some  $\mathcal{C} \subset [n]^2$  of size  $n$ , we denote by  $X_{\mathcal{C}}$  the extremal element of  $\mathcal{C}_n$  whose support is  $\mathcal{C}$  (if it exists). In addition, we denote by  $\mathcal{C}_1$  the set of pairs  $(i, j)$  in  $\mathcal{C}$  with  $i \neq j$ , and by  $\mathcal{C}_2$  the set of indices  $k$  such that  $(k, k)$  is not in  $\mathcal{C}$ . We can write

$$X_{\mathcal{C}} = I + \sum_{(i,j) \in \mathcal{C}_1} e_i e_j^\top - \sum_{k \in \mathcal{C}_2} e_k e_k^\top.$$

Note that

$$|\mathcal{C}_1| = |\mathcal{C}_2| = n - \text{Tr } X_{\mathcal{C}}.$$

Write then

$$\begin{aligned} \langle P_\epsilon(J), I - X_{\mathcal{C}} \rangle &= \left\langle P_\epsilon(J), \sum_{k \in \mathcal{C}_2} e_k e_k^\top - \sum_{(i,j) \in \mathcal{C}_1} e_i e_j^\top \right\rangle \\ &= \sum_{k \in \mathcal{C}_2} \langle P_\epsilon(J), e_k e_k^\top \rangle - \sum_{(i,j) \in \mathcal{C}_1} \langle P_\epsilon(J), e_i e_j^\top \rangle \end{aligned}$$

Let now  $Y := P_\epsilon(J)$ . The previous display is then written as

$$\langle P_\epsilon(J), I - X_{\mathcal{C}} \rangle = \sum_{k \in \mathcal{C}_2} Y_{kk} - \sum_{(i,j) \in \mathcal{C}_1} Y_{ij}. \quad (16)$$

A key step to conclude our argument is the following

**Lemma 2** *Choose  $\epsilon = \kappa / \ln^4(n)$ , where  $\kappa > 0$  is a sufficiently small constant, of order  $\Omega(1)$ . Then with high probability one has:*

$$\forall i \in [n], Y_{ii} = 1 + o(1), \quad (17)$$

and

$$\forall i \neq j \in [n], |Y_{ij}| \leq \frac{1}{2}. \quad (18)$$

The proof of Lemma 2 is presented in Section 4. Intuitively, the choice  $\epsilon = \kappa / \ln^4(n)$  is essentially the largest band for which  $Y = P_\epsilon(J)$  remains component-wise close to  $I$ . In particular, recall  $Y$  is close to  $I$  for small  $\epsilon$  and close to  $J$  for large  $\epsilon$ : Lemma 2 certifies that  $\epsilon = \kappa / \ln^4(n) = o(1)$  is still on the  $Y \approx I$  side. Essentially, we take  $\epsilon$  as large as possible, while ensuring  $Y \approx I$ , because a larger band yields a larger eigenvalue gap  $C_\epsilon$ ; we show later (Lemma 3) that this choice gives  $C_\epsilon = \tilde{\Omega}(1)$ , keeping the prefactor in the resulting lower bound from vanishing. Plugging the above lemma's result in (16) we get

$$\langle P_\epsilon(J), I - X_{\mathcal{C}} \rangle \geq (1 - o(1))|\mathcal{C}_2| - \frac{1}{2}|\mathcal{C}_1| \geq \frac{1}{3}|\mathcal{C}_2| = \frac{1}{3}(n - \text{Tr}(X_{\mathcal{C}})).$$

*Concluding:* For each  $X \in \mathcal{C}_n$ , writing  $X$  as a convex combination of extremal points of  $\mathcal{C}_n$  of the form  $X_{\mathcal{C}}$  as defined above, one readily obtains

$$\forall X \in \mathcal{C}_n, \langle P_\epsilon(J), I - X \rangle \geq \frac{1}{3}(n - \text{Tr}(X)).$$

Combined with (15), this yields

$$\forall X \in \mathcal{C}_n, \|AX - XA\|_F \geq \frac{\sqrt{C_\epsilon}}{n} \frac{1}{3} (n - \text{Tr}(X)).$$

This holds in particular for the optimum  $X^*$ . Its optimality ensures that it achieves a lower objective value than the identity matrix  $I$ , for which  $\|A \times I - I(A + \sigma Z)\|_F = \sigma \|Z\|_F = O(\sigma\sqrt{n})$ , so that by (12) and the previous display:

$$O(\sigma\sqrt{n}) \geq \|AX^* - X^*(A + \sigma Z)\|_F \geq \frac{\sqrt{C_\epsilon}}{3} [1 - n^{-1} \text{Tr} X^*] - 3\sigma\sqrt{n}.$$

Thus

$$\text{Tr} X^* \geq n - O(\sigma n^{3/2} / \sqrt{C_\epsilon}).$$

The following lemma will then allow us to conclude:

**Lemma 3** *Assume that for some fixed constant  $\delta > 0$  one has*

$$\epsilon \geq n^{\delta-2/3}. \quad (19)$$

*Then with high probability*

$$C_\epsilon = \Omega(\epsilon^2), \quad (20)$$

*where  $C_\epsilon$  is defined in (13).*

Our choice  $\epsilon = \Omega(1)/\ln^4(n)$  clearly satisfies Condition (19). Lemma 3 thus ensures that  $C_\epsilon = \Omega(\epsilon^2)$ . The proof of Lemma 3 is presented in Appendix B. The intuition behind Lemma 3 is that GOE eigenvalues have typical spacing  $\lambda_{i+1} - \lambda_i \sim 1/n$ , so two eigenvalues whose indices differ by at least  $\epsilon n$  are separated by roughly  $\epsilon$ ; squaring gives  $C_\epsilon = \Omega(\epsilon^2)$ . ■

### 3.2. Dual Approach

In this section, we prove the following result:

**Proposition 4** *For correlated GOE matrices, if the noise parameter  $\sigma$  satisfies*

$$\sigma = o\left(\frac{n^{-1/2}}{\ln(n)^{5.5}}\right), \quad (21)$$

*one then has with high probability, for relaxation to the Birkhoff polytope  $\mathcal{B}_n$ , constrained hypercube  $\mathcal{H}_n$ , and simplex  $\Delta_n$ :*

$$\text{Tr}(X^*) \geq n(1 - o(1)). \quad (22)$$

In light of (10), the above proposition implies the conclusion of Theorem 1 under Condition (21), a slight strengthening of Condition (6) on  $\sigma$ . Note that the above proposition also ensures that  $\text{Tr}(X^*) \geq n(1 - o(1))$ , even for the simplex relaxation, i.e.,  $\mathcal{C}_n = \Delta_n$ . However, the reduction from  $\text{Tr}(X^*) \geq n - o(n)$  to  $\|X^* - I\|_F^2 = o(n)$  via (10) requires  $\|X^*\|_F \leq \sqrt{n}$ , which may fail for  $X^* \in \Delta_n$ . Proposition 4 still yields  $\text{Tr}(X^*) \geq n - o(n)$  for the simplex; whether this can be strengthened to a Frobenius-norm guarantee remains open.

**Proof** [Proof of Proposition 4] As before, let  $A = \sum_{i=1}^n \lambda_i u_i u_i^T$ , where  $\{\lambda_i\}_{i=1}^n$  are the eigenvalues and  $\{u_i\}_{i=1}^n$  are the corresponding orthonormal eigenvectors of  $A$ . The proof is mainly divided into the following four parts.

Trace minimization problem: Our objective is to lower bound  $\text{Tr}(X^*)$  for  $X^* \in \Delta_n$  satisfying (11), which motivates the following optimization problem:

$$T^* = \min_{X \in \mathbb{R}^{n \times n}} \text{Tr}(X) \quad \text{subject to} \quad \sum_{i,j=1}^n X_{ij} = n, X \geq 0, \|AX - XB\|_F^2 \leq c^2 \sigma^2 n. \quad (23)$$

Note that the objective is a linear function, and the constraints define a convex set. Thus, the above optimization problem is convex. The norm constraint  $\|AX - XB\|_F^2 \leq c^2 \sigma^2 n$  encodes the (near-) optimality of  $X^*$  via (11), and is essential: without it, one could trivially construct dual solutions with objective value 0.

The Dual Problem: Introduce the dual variables  $R \in \mathbb{R}^{n \times n}$  and  $\tilde{\mu}, \mu \in \mathbb{R}$  corresponding to the non-negativity, the norm constraint, and the total sum constraint, respectively. Thus, by weak duality, we have

$$\begin{aligned} T^* &\geq \max_{R \geq 0, \mu, \tilde{\mu}} \min_X \text{Tr}(X) - \langle R, X \rangle - \mu \langle J, X \rangle + \tilde{\mu} \|AX - XB\|_F^2 - \tilde{\mu} c^2 \sigma^2 n + \mu n \\ &= \max_{R \geq 0, \mu, \tilde{\mu}} \min_X \tilde{\mu} \|AX - XB\|_F^2 - \langle R + \mu J - I, X \rangle - \tilde{\mu} c^2 \sigma^2 n + \mu n \\ &\stackrel{(a)}{=} \max_{R \geq 0, \mu, \tilde{\mu}, M} \tilde{\mu} \|AM - MB\|_F^2 - \langle R + \mu J - I, M \rangle - \tilde{\mu} c^2 \sigma^2 n + \mu n \\ &\quad \text{subject to} \quad 2\tilde{\mu}(A^2 M - 2AMB + MB^2) = R + \mu J - I \\ &\stackrel{(b)}{=} \max_{R \geq 0, \mu, \tilde{\mu}, M} -\tilde{\mu} \|AM - MB\|_F^2 - \tilde{\mu} c^2 \sigma^2 n + \mu n \\ &\quad \text{subject to} \quad 2\tilde{\mu}(A^2 M - 2AMB + MB^2) = R + \mu J - I \\ &\stackrel{(c)}{=} \max_{R \geq 0, \mu, \tilde{\mu}, \tilde{M}, M} -\tilde{\mu} \|\tilde{M}\|_F^2 - \tilde{\mu} c^2 \sigma^2 n + \mu n \\ &\quad \text{subject to} \quad 2\tilde{\mu}(A\tilde{M} - \tilde{M}B) = R + \mu J - I, \tilde{M} = AM - MB. \end{aligned} \quad (24)$$

Above, (a) follows from the first order optimality condition by noting that the inner minimization is an unconstrained convex problem over  $X$ ; the constraint on  $M$  corresponds to the gradient of the minimized quantity being 0. Next, (b) holds by multiplying the constraint on both sides by  $M$ , which implies

$$\langle R + \mu J - I, M \rangle = 2\tilde{\mu} \langle A^2 M - 2AMB + MB^2, M \rangle = 2\tilde{\mu} \|AM - MB\|_F^2.$$

Lastly, (c) follows by introducing the auxiliary variable  $\tilde{M} = AM - MB$ .

Dual Certificate: Now, we provide a dual feasible solution  $(R, \mu, \tilde{\mu}, M, \tilde{M})$  which provides a lower bound on  $T^*$ . Denote  $e$  the all-ones vector. Fix  $\epsilon = \kappa / (\log n)^4$  for some large enough  $\kappa > 0$ , set  $\tilde{\mu} = n(\log n)^{10.5}$ , and

$$\tilde{M} = \frac{1}{\tilde{\mu}} \sum_{|i-j| \geq \epsilon n} \frac{(e^T u_i)(e^T u_j)}{\lambda_i - \lambda_j} u_i u_j^T, \quad Y = P_\epsilon(J) = \sum_{|i-j| < \epsilon n} (e^T u_i)(e^T u_j) u_i u_j^T,$$

Moreover, we also set  $\mu = 1 - \xi_1 - \xi_2$  with  $\xi_1 = 2\tilde{\mu}\sigma \max_{i,j \in [n]} (\tilde{M}Z)_{i,j}$  and  $\xi_2 = \max_{i,j \in [n]} (2Y - I - J)_{i,j}$ , and

$$R = (1 + \xi_1 + \xi_2) J + I - 2Y - 2\tilde{\mu}\sigma \tilde{M}Z. \quad (25)$$

Next, note that  $A \otimes I - I \otimes B$  is non-singular with probability 1 for  $\sigma > 0$  (Araya and Tyagi, 2024, Lemma 4). Thus, we set  $M$  to be the unique solution of  $\tilde{M} = AM - MB$ . The constraints are feasible by construction. The feasibility of equality constraint is verified as follows:

$$\begin{aligned} \tilde{\mu}(A\tilde{M} - \tilde{M}B) &= \tilde{\mu}(A\tilde{M} - \tilde{M}A) - \tilde{\mu}\sigma \tilde{M}Z \\ &= \sum_{|i-j| \geq \epsilon n} \frac{(e^T u_i)(e^T u_j)}{\lambda_i - \lambda_j} (Au_i u_j^T - u_i u_j^T A) - \tilde{\mu}\sigma \tilde{M}Z \\ &= \sum_{|i-j| \geq \epsilon n} (e^T u_i)(e^T u_j) u_i u_j^T - \sigma \tilde{\mu} \tilde{M}Z = J - Y - \tilde{\mu}\sigma \tilde{M}Z = (R + \mu J - I)/2. \end{aligned}$$

Also, by the definition of  $\xi_1, \xi_2$ , we have  $R = (\xi_1 J - 2\tilde{\mu}\sigma \tilde{M}Z) + (\xi_2 J - (2Y - I - J)) \geq 0$ .

*Dual Objective Value:* We first upper bound  $\|\tilde{\mu}\tilde{M}\|_F$  using eigenvalue separation established in Lemma 3. Indeed, with high probability,

$$\|\tilde{\mu}\tilde{M}\|_F^2 = \sum_{|i-j| \geq \epsilon n} \frac{(e^T u_i)^2 (e^T u_j)^2}{(\lambda_i - \lambda_j)^2} \leq \frac{1}{C_\epsilon} \sum_{|i-j| \geq \epsilon n} (e^T u_i)^2 (e^T u_j)^2 \stackrel{(*)}{\leq} O(n^2 (\log n)^{10}), \quad (26)$$

where  $(*)$  holds as, for any  $i \leq n$ , for any  $c > 1$ , it holds with probability  $1 - n^{-c}$  for large enough  $n$  that

$$e^T u_i = \frac{e^T z}{\|z\|_2} \leq \frac{2e^T z}{\sqrt{n}} \leq O(\sqrt{\log n}),$$

where  $z \sim N(0, I_n)$ . The second inequality holds from (Fan et al., 2023, Lemma 15), and the last one from (Fan et al., 2023, Lemma 13). By union bound, the same holds true simultaneously for all  $i$  with probability at least  $1 - n^{-(c-1)}$ . Moreover,  $C_\epsilon = \Omega(\epsilon^2)$  by Lemma 3, where, we set  $\epsilon = \kappa/(\log n)^4$ . Next, we upper bound  $\xi_1$ . As  $A$  and  $Z$  are independent,  $\tilde{M}$  is independent of  $Z$ . Thus, by (Fan et al., 2023, Lemma 13), for any  $i, j$ , for any  $c > 2$ , with probability at least  $1 - n^{-c}$ , we have

$$|(\tilde{M}Z)_{i,j}| = \left| \sum_{l=1}^n \tilde{M}_{i,l} Z_{l,j} \right| \leq O \left( \sqrt{\frac{\log n}{n} \sum_{l=1}^n \tilde{M}_{i,l}^2} \right) \leq O \left( \sqrt{\frac{\log n}{n}} \|\tilde{M}\|_F \right). \quad (27)$$

This inequality simultaneously holds for all pairs  $(i, j)$ , with probability  $1 - n^{-(c-2)}$ , by union bound. Thus, we get

$$\xi_1 = \max_{i,j \in [n]} |2\tilde{\mu}\sigma (\tilde{M}Z)_{i,j}| = O \left( \sigma \sqrt{\frac{\log n}{n}} \|\tilde{\mu}\tilde{M}\|_F \right) \stackrel{(26)}{=} O(\sigma \sqrt{n} (\log n)^{5.5}) = o(1), \quad (28)$$

where the last equality holds by (21). Now, we upper bound  $\xi_2$ . By Lemma 2, we readily obtain  $(2Y - I - J)_{i,j} \leq 0$  for  $i \neq j \in [n]$  and  $(2Y - I - J)_{i,i} = o(1)$  w.h.p. Thus, we have

$$\xi_2 = \max_{i,j \in [n]} (2Y - I - J)_{i,j} = o(1), \quad (29)$$

w.h.p. Combining (26), (28), and (29), we get

$$\begin{aligned} T^* &\geq -\tilde{\mu} \|\tilde{M}\|_F^2 - \tilde{\mu} c^2 \sigma^2 n + \mu n = -\tilde{\mu} \|\tilde{M}\|_F^2 - \tilde{\mu} c^2 \sigma^2 n + (1 - \xi_1 - \xi_2) n \\ &\geq -O\left(\frac{n^2 (\log n)^{10}}{\tilde{\mu}}\right) - \tilde{\mu} c^2 \sigma^2 n + n(1 - o(1)) = n - o(n), \end{aligned}$$

where the last equality holds due to our choices of  $\tilde{\mu} = n(\log n)^{10.5}$  and (21). Thus, we obtain  $T^* \geq n - o(n)$ , which completes the proof.  $\blacksquare$

### 3.3. Discussion on the two approaches

The primal and the dual approaches both provide a lower bound on  $\text{Tr}(X^*)$  by using the upper bound on the objective function given in (11). The primal approach lower bounds the trace for any feasible  $X \in \mathcal{H}_n$  of (23), while the dual approach constructs a dual feasible solution, whose objective function value lower bounds the trace via weak duality. Although the two approaches look different, they run into the *same* obstacle and resolve it with the *same* technical hammer, which is precisely why Lemmas 2 and 3 are central to both. The obstacle is that the map  $X \mapsto AX - XA$  has eigenvalues  $\{\lambda_i - \lambda_j\}_{i,j \in [n]}$ , which vanish at  $i = j$  and are tiny when  $|i - j|$  is small; the technical hammer is to restrict attention to the well-separated pairs  $|i - j| \geq \epsilon n$ , on which Lemma 3 controls the eigengap  $C_\epsilon$  and Lemma 2 controls the relevant projection of  $J$ . We make this parallel precise below. In the primal approach the obstacle appears directly: the naive bound  $\|AX - XA\|_F = \|A(X - I) - (X - I)A\|_F \geq \lambda_{\min}(A \otimes I - I \otimes A) \|X - I\|_F$  is vacuous, since the eigenvalues  $\{\lambda_i - \lambda_j\}$  vanish at  $i = j$  and can be as small as  $n^{-3/2}$  for nearby indices (Feng et al., 2019). Projecting  $X - I$  onto the band  $|i - j| \geq \epsilon n$  excises these small gaps: Lemma 3 ensures  $C_\epsilon = \tilde{\Omega}(1)$ , while Lemma 2 gives  $P_\epsilon(J) \approx I$ , so that the projected objective still captures  $n - \text{Tr}(X)$  via Cauchy–Schwarz, as in (15). The dual approach meets the identical obstacle, only inverted. Constructing a dual certificate for (23) requires the primal solution to be close to  $I$ , which by complementary slackness suggests the choice  $R = J - I$  (exactly the certificate used in Varma et al. (2025)). The dual constraint (24) then forces  $\tilde{\mu}\tilde{M}$  to satisfy  $\tilde{\mu}(A\tilde{M} - \tilde{M}A) \approx J - I$ , i.e.,  $\tilde{\mu}\tilde{M}$  is obtained by inverting the same map  $X \mapsto AX - XA$  on  $J$ :

$$\tilde{\mu}\tilde{M} \approx \sum_{i \neq j} \frac{(e^\top u_i)(e^\top u_j)}{\lambda_i - \lambda_j} u_i u_j^\top.$$

The very same small denominators  $\lambda_i - \lambda_j$  that weakened the primal bound now *blow up*  $\|\tilde{\mu}\tilde{M}\|_F$ , producing a small dual objective and hence a weak bound on  $\text{Tr}(X^*)$ ; this is what restricts Varma et al. (2025) to  $\sigma = o(n^{-1})$ . The remedy is the same as in the primal proof: we set  $R$  to approximately the projection of  $J - I$  onto  $\{u_i u_j^\top\}_{|i-j| \geq \epsilon n}$  (the remaining terms in (25) are lower order), which excises the small-gap pairs and yields  $\tilde{\mu}\tilde{M} \approx \sum_{|i-j| \geq \epsilon n} \frac{(e^\top u_i)(e^\top u_j)}{\lambda_i - \lambda_j} u_i u_j^\top$ . Now Lemma 3

bounds  $\|\tilde{\mu}\tilde{M}\|_F$  through the same eigengap  $C_\epsilon$ , while Lemma 2 controls the error incurred by projecting  $J - I$  off the band, ensuring  $R$  stays close to  $J - I$  and feasible. In short, the two proofs are mirror images: the band  $|i - j| \geq \epsilon n$  and the two lemmas play identical roles in each, one controlling the eigengap  $C_\epsilon$  (Lemma 3) and the other a projection of  $J$  (Lemma 2). Establishing these lemmas (especially Lemma 2) requires considerable technical overhead beyond Varma et al. (2025), highlighting our methodological contributions. The advantage of the dual approach compared to the direct approach is that it provides a tight lower bound on  $\text{Tr}(X^*)$  even for the simplex relaxation ( $\mathcal{C}_n = \Delta_n$ ). While this trace bound cannot directly be translated to guarantees on  $\|X^* - I\|_F$ , it still provides partial results and progress towards that goal, and could serve as a good starting point for future work on simplex relaxation.

#### 4. Proof of Lemma 2.

We shall consider the case of  $Y_{i,j}$  for  $i \neq j$  and  $i = j$  separately. The details for the case of  $i = j$  are deferred to Appendix A.3 as it is easier than the case of  $i \neq j$ . Here we focus on bounding  $Y_{i,j}$  for  $i \neq j$ . Denote by  $e$  the all-ones vector. Write, recalling that  $U = (u_1 | \dots | u_n)$ , and denoting  $V := U^\top$ :

$$\begin{aligned} Y_{i,j} &= \sum_{a,b:|b-a|<\epsilon n} \langle u_a u_b^\top, e e^\top \rangle \langle u_a u_b^\top, e_i e_j^\top \rangle = \sum_{a,b:|b-a|<\epsilon n} (u_a^\top e)(u_b^\top e)(u_b^\top e_i)(u_a^\top e_j) \\ &= \sum_{a,b:|b-a|<\epsilon n} (e_a^\top V e)(e_b^\top V e)(e_b^\top V e_i)(e_a^\top V e_j). \end{aligned}$$

Now,  $V$  is uniformly distributed over the orthogonal group  $O_n$ . Thus,  $V e_i$  is uniformly distributed over  $\mathcal{S}^{n-1}$ ,  $V e_j$  is uniformly distributed over the intersection of  $\mathcal{S}^{n-1}$  with the orthogonal of  $V e_i$ , and  $(n-2)^{-1/2}[V e - V e_i - V e_j]$  is uniformly distributed over the intersection of  $\mathcal{S}^{n-1}$  with the orthogonal of  $\text{vect}\{V e_i, V e_j\}$ . To control the distribution of  $Y_{i,j}$  we use the following construction. With a slight abuse of notation, take  $X, X', X''$  three i.i.d. Gaussian random vectors with distribution  $\mathcal{N}(0, I)$ . Let then

$$v := \frac{1}{\alpha} X, \quad v' := \frac{1}{\beta} (X' - r v), \quad v'' := \frac{1}{\gamma} (X'' - s v - t v')$$

denote the Gram-Schmidt orthonormalization of  $(X, X', X'')$ , where

$$\alpha := \|X\|, \quad r := v^\top X', \quad \beta := \|X' - r v\|, \quad s := v^\top X'', \quad t := v'^\top X'', \quad \gamma := \|X'' - s v - t v'\|.$$

Then  $Y$  has the same distribution as

$$Y := \sum_{a,b:|b-a|<\epsilon n} v_b v'_a (v_a + v'_a + \sqrt{n-2} v''_a) (v_b + v'_b + \sqrt{n-2} v''_b).$$

For two independent standard Gaussian random variables  $Z_1, Z_2$  and  $x \in \mathbb{R}$ , we have

$$\mathbb{E} \left( e^{x Z_1^2} \right) = (1 - 2x)_+^{-1/2}, \quad \mathbb{E} \left( e^{x Z_1 Z_2} \right) = (1 - x^2)_+^{-1/2}.$$

Let  $\bar{Z}_1 = (Z_1(1), \dots, Z_1(n))$ ,  $\bar{Z}_2 = (Z_2(1), \dots, Z_2(n))$  be two independent vectors both distributed according to  $\mathcal{N}(0, I)$ . By Chernoff's argument, we have the existence of two non-negative

convex functions  $g, h$ , where  $g$  achieves its minimum at 1, with  $g(1) = g'(1) = 0$  and  $g''(1) > 0$ , and likewise  $h$  achieves its minimum at 0 with  $h(0) = h'(0) = 0$ ,  $h''(0) > 0$ , such that for all  $x > 0$ ,

$$\mathbb{P}\left(\left|\sum_{i=1}^n Z_1(i)^2 - n\right| \geq xn\right) \leq 2e^{-ng(1+x)} \text{ and } \mathbb{P}\left(\left|\sum_{i=1}^n Z_1(i)Z_2(i)\right| \geq xn\right) \leq 2e^{-nh(x)}.$$

Thus for any constant  $c > 0$ , for large enough constant  $C > 0$  (for instance, we can set  $C = 3c \max(1/g''(1), 1/h''(0))$ ), taking  $x = C\sqrt{\ln(n)}/n$  we obtain

$$\mathbb{P}\left(\left|\sum_{i=1}^n Z_1(i)^2 - n\right| \geq C\sqrt{n \ln(n)}\right) \leq n^{-c} \text{ and } \mathbb{P}\left(\left|\sum_{i=1}^n Z_1(i)Z_2(i)\right| \geq C\sqrt{n \ln(n)}\right) \leq n^{-c}.$$

Thus with probability  $1 - O(n^{-c})$ , one has:

$$\begin{aligned} \alpha &= \sqrt{n}(1 + O(\sqrt{\ln(n)}/n)) = \sqrt{n} + O(\sqrt{\ln(n)}), \\ r &= \alpha^{-1}X^\top X' = O(\alpha^{-1}\sqrt{n \ln(n)}) = O(\sqrt{\ln(n)}), \\ \beta &= \sqrt{n + O(\sqrt{n \ln(n)}) + O(r/\alpha)\sqrt{n \ln(n)} + (r/\alpha)^2(n + O(\sqrt{n \ln(n)}))} \\ &= \sqrt{n} + O(\sqrt{\ln(n)}), \\ s, t &= O(\sqrt{\ln(n)}), \quad \gamma = \sqrt{n} + O(\sqrt{\ln(n)}). \end{aligned}$$

Write then, denoting  $r' := r/\alpha, s' = s/\alpha, t' = t/\beta$ :

$$\begin{aligned} \alpha\beta Y &= \sum_{|a-b| < \epsilon n} X_a(X'_b - r'X_b) \times \\ &\times \left[ \left( \frac{1}{\alpha} - \frac{r'}{\beta} - \frac{\sqrt{n-2s'}}{\gamma} + \frac{r't'\sqrt{n-2}}{\gamma} \right) X_a + \left( \frac{1}{\beta} - \frac{\sqrt{n-2t'}}{\gamma} \right) X'_a + \frac{\sqrt{n-2}}{\gamma} X''_a \right] \\ &\times \left[ \left( \frac{1}{\alpha} - \frac{r'}{\beta} - \frac{\sqrt{n-2s'}}{\gamma} + \frac{r't'\sqrt{n-2}}{\gamma} \right) X_b + \left( \frac{1}{\beta} - \frac{\sqrt{n-2t'}}{\gamma} \right) X'_b + \frac{\sqrt{n-2}}{\gamma} X''_b \right] \\ &= \sum_{|a-b| < \epsilon n} X_a(X'_b + \theta X_b) [\theta X_a + \theta X'_a + (1 + o(1))X''_a] [\theta X_b + \theta X'_b + (1 + o(1))X''_b], \end{aligned}$$

where in the last expression,  $\theta$  is a short-hand notation for  $O(\sqrt{\ln n/n})$ . This yields:

$$\begin{aligned} \alpha\beta Y &= (1 + o(1)) \sum_{|a-b| < \epsilon n} X_a X'_b X''_a X''_b + \theta^3 \sum_{|a-b| < \epsilon n} X_a X_b (X_a + X'_a)(X_b + X'_b) \\ &+ \theta \sum_{|a-b| < \epsilon n} X_a [X'_b X''_a (X_b + X'_b) + X'_b (X_a + X'_a) X''_b + X_b X''_a X''_b] \\ &+ \theta^2 \sum_{|a-b| < \epsilon n} X_a [X_b (X_a + X'_a) X''_b + X_b X''_a (X_b + X'_b) + X'_b (X_a + X'_a)(X_b + X'_b)]. \end{aligned} \tag{30}$$

Thus, we obtain a tractable expression of  $Y$  in terms of i.i.d. Gaussian random vectors  $(X, X', X'')$ . We then upper bound the above expression using concentration inequalities for Gaussian vectors and matrices. The details are deferred to Appendix A.2.

## 5. Conclusion

We studied convex relaxations of the quadratic assignment formulation for graph alignment under the correlated GOE model with correlation  $1/\sqrt{1+\sigma^2}$ . Focusing on relaxations over the Birkhoff polytope and a constrained hypercube, we established sharp recovery guarantees. In particular, we showed that when  $\sigma = o(n^{-1/2}/\log^4 n)$ , the solution of the relaxed problem concentrates around the ground-truth permutation matrix in Frobenius norm. Combined with existing lower bounds, our results precisely characterize the noise threshold at which  $\|X^* - \Pi^*\|_F^2$  transitions from  $o(n)$  to  $\Omega(n)$ . Our analysis further shows that the hypercube relaxation achieves the same guarantees as the Birkhoff relaxation, despite operating over a strictly larger feasible set. The resulting concentration bounds imply that simple post-processing methods recover all but a vanishing fraction of the true permutation. An important open direction is to develop a theoretical understanding of LAPs for  $\sigma = \Omega(n^{-1/2})$  and to extend these guarantees to broader random graph models. A promising route is to establish diagonal dominance of the unregularized minimizer, as discussed in Section 2. Like Frobenius proximity to  $\Pi^*$ , diagonal dominance guarantees that the post-processing of (4) recovers all but a vanishing fraction of  $\pi^*$ ; however, whereas Varma et al. (2025) showed that Frobenius proximity provably fails above  $\sigma = n^{-1/2}$ , diagonal dominance is expected to persist for substantially larger  $\sigma$ , in line with the empirical performance of the Birkhoff and simplex relaxations.

## References

- 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.
- Greg Anderson, Alice Guionnet, and Ofer Zeitouni. *An Introduction to Random Matrices*. Cambridge University Press, 2010.
- Ernesto Araya and Hemant Tyagi. Graph matching via convex relaxation to the simplex. *Foundations of Data Science*, pages 0–0, 2024.
- Donatello Conte, Pasquale Foggia, Carlo Sansone, and Mario Vento. Thirty years of graph matching in pattern recognition. *International journal of pattern recognition and artificial intelligence*, 18(03):265–298, 2004.
- Jian Ding, Zongming Ma, Yihong Wu, and Jiaming Xu. Efficient random graph matching via degree profiles. *Probability Theory and Related Fields*, 179:29–115, 2021.
- Nadav Dym, Haggai Maron, and Yaron Lipman. DS++ a flexible, scalable and provably tight relaxation for matching problems. *ACM Transactions on Graphics (TOG)*, 36(6):1–14, 2017.
- László Erdős, Horng-Tzer Yau, and Jun Yin. Rigidity of eigenvalues of generalized wigner matrices. *Advances in Mathematics*, 229(3):1435–1515, 2012. ISSN 0001-8708. doi: <https://doi.org/10.1016/j.aim.2011.12.010>. URL <https://www.sciencedirect.com/science/article/pii/S0001870811004063>.
- Zhou Fan, Cheng Mao, Yihong Wu, and Jiaming Xu. Spectral graph matching and regularized quadratic relaxations I: algorithm and gaussian analysis. *Foundations of Computational Mathematics*, 23(5):1511–1565, 2023.
- Renjie Feng, Gang Tian, and Dongyi Wei. Small gaps of GOE. *Geometric and Functional Analysis*, 29(6):1794–1827, 2019.
- Luca Ganassali. Sharp threshold for alignment of graph databases with gaussian weights. In *Mathematical and Scientific Machine Learning*, pages 314–335. PMLR, 2022.
- Luca Ganassali, Marc Lelarge, and Laurent Massoulié. Spectral alignment of correlated gaussian matrices. *Advances in Applied Probability*, 54(1):279–310, 2022.
- Luca Ganassali, Marc Lelarge, and Laurent Massoulié. Correlation detection in trees for planted graph alignment. *The Annals of Applied Probability*, 34(3):2799–2843, 2024.
- Julia Gaudio, Miklós Z Rácz, and Anirudh Sridhar. Average-case and smoothed analysis of graph isomorphism. *The Annals of Applied Probability*, 35(2):1373–1406, 2025.
- Friedrich Götze, Holger Sambale, and Arthur Sinulis. Concentration inequalities for polynomials in  $\alpha$ -sub-exponential random variables. *Electronic Journal of Probability*, 26:1 – 22, 2021. doi: [10.1214/21-EJP606](https://doi.org/10.1214/21-EJP606). URL <https://doi.org/10.1214/21-EJP606>.

Vince Lyzinski, Donniell E Fishkind, Marcelo Fiori, Joshua T Vogelstein, Carey E Priebe, and Guillermo Sapiro. Graph matching: Relax at your own risk. *IEEE transactions on pattern analysis and machine intelligence*, 38(1):60–73, 2015.

Konstantin Makarychev, Rajsekar Manokaran, and Maxim Sviridenko. Maximum quadratic assignment problem: Reduction from maximum label cover and lp-based approximation algorithm. In *International Colloquium on Automata, Languages, and Programming*, pages 594–604. Springer, 2010.

Cheng Mao, Yihong Wu, Jiaming Xu, and Sophie H Yu. Random graph matching at otter’s threshold via counting chandeliers. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 1345–1356, 2023.

Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*, pages 111–125. IEEE, 2008.

Rohit Singh, Jinbo Xu, and Bonnie Berger. Global alignment of multiple protein interaction networks with application to functional orthology detection. *Proceedings of the National Academy of Sciences*, 105(35):12763–12768, 2008.

Sushil Mahavir Varma, Irène Waldspurger, and Laurent Massoulié. Graph alignment via birkhoff relaxation. *CoRR*, 2025. URL <https://openreview.net/forum?id=YdbJKHz4TV>.

Yihong Wu, Jiaming Xu, and H Yu Sophie. Settling the sharp reconstruction thresholds of random graph matching. *IEEE Transactions on Information Theory*, 68(8):5391–5417, 2022.

## Appendix A. Proof Details for Lemma 2

### A.1. A concentration lemma

We first need the following technical lemma concerning concentration of sums of i.i.d. random variables with zero mean and moment bounds.

**Lemma 5** *Let  $U \in \mathbb{R}^n$  be an i.i.d. random vector with  $\mathbb{E}(U) = 0$  and  $\mathbb{E}(U_i^{2k}) \leq r_k$  for some constants  $k, r_k > 0$ . Then there exists a constant  $c_k$  depending only on  $r_k$  and  $k$  such that, with probability at least  $1 - c_k n^{-k/3}$ ,*

$$\sum_{a=1}^n U_a \leq n^{2/3}.$$

**Proof** We have

$$\mathbb{E} \left( \sum_{a=1}^n U_a \right)^{2k} = \mathbb{E} \left( \sum_{i_1, \dots, i_{2k}=1}^n \prod_{j=1}^{2k} U_{i_j} \right). \tag{31}$$

Observe that, for any  $(i_1, \dots, i_{2k}) \in \{1, \dots, n\}^{2k}$ , if  $\{i_1, \dots, i_{2k}\}$  has cardinality strictly greater than  $k$ , then it must be that an index appears exactly once in the  $2k$ -uplet, hence, since  $\mathbb{E}(U) = 0$ ,

$$\mathbb{E} \left( \prod_{j=1}^{2k} U_{i_j} \right) = 0.$$

Therefore, if we denote

$$\mathcal{E}_{n,k} = \{E \subset \{1, \dots, n\}, \text{Card}(E) \leq k\},$$

we have

$$\mathbb{E} \left( \sum_{a=1}^n U_a \right)^{2k} \leq \sum_{E \in \mathcal{E}_{n,k}} \sum_{i_1, \dots, i_{2k} \in E} \mathbb{E} \left( \prod_{j=1}^{2k} U_{i_j} \right) \stackrel{(*)}{\leq} \sum_{E \in \mathcal{E}_{n,k}} \text{Card}(E)^{2k} \mathbb{E}(U^{2k}) \stackrel{(**)}{\leq} n^k k^{2k} r_k,$$

where  $(*)$  follows as, by Hölder's inequality, for any set of iid random variables  $Y_1, \dots, Y_s$  and any set of nonnegative integers  $n_1, \dots, n_s$ , it holds that

$$\mathbb{E} \left| \prod_{j=1}^s Y_j^{n_j} \right| \leq \mathbb{E} |Y_1^{n_1 + \dots + n_s}|,$$

and  $(**)$  is true because  $\text{Card}(E) \leq k$  for any  $E \in \mathcal{E}_{n,k}$  and  $\text{Card}(\mathcal{E}_{n,k}) \leq n^k$ . Now, the Markov's inequality gives us

$$\mathbb{P} \left( \left| \sum_{a=1}^n U_a \right| > n^{2/3} \right) \leq n^{-4k/3} n^k k^{2k} r_k = r_k k^{2k} n^{-k/3}.$$

This completes the proof. ■

### A.2. The case of $Y_{i,j}$ for $i \neq j$

Each of the sums in (30) can be split according to whether  $a = b$  or  $a \neq b$ . The resulting sums when  $a = b$  give rise to contributions of the following types:

$$\begin{aligned} S_0 &:= \sum_{a=1}^n X_a X'_a (X''_a)^2, \quad S_1 := \theta \sum_{a=1}^n X_a^2 X'_a X''_a, \quad S'_1 := \theta \sum_{a=1}^n X_a^2 (X''_a)^2, \\ S_2 &:= \theta^2 \sum_{a=1}^n X_a^3 X''_a, \quad S_3 := \theta^3 \sum_{a=1}^n X_a^4, \end{aligned}$$

as well as sums of lesser order such as  $\theta^3 \sum_{a=1}^n X_a^3 X'_a$  and  $\theta^3 \sum_{a=1}^n X_a^2 X'_a{}^2$ . We start by upper bounding  $S_0$  below:

$$S_0 = \sum_{a=1}^n X_a X'_a (X''_a)^2 \leq O \left( \sqrt{\log n \sum_{a=1}^n (X'_a)^2 (X''_a)^4} \right) \leq O(\sqrt{n}(\log n)^2),$$

where the first inequality holds w.h.p.  $(1 - n^{-c})$  by (Fan et al., 2023, Lemma 13) (conditioning on  $X'$  and  $X''$ ) and the second one by bounding each component of  $X'$  and  $X''$  by  $O(\sqrt{\log n})$  w.h.p.  $(1 - n^{-c})$ . The same argument gives w.h.p.  $(1 - n^{-c})$

$$S_1 = O((\log n)^{2.5}), \quad S_2 = O((\log n)^3/\sqrt{n}).$$

Next, with probability at least  $1 - n^{-c}$ , we have

$$\begin{aligned} \sum_{a=1}^n X_a^2 (X_a'')^2 &= n + \sum_{a=1}^n (X_a^2 - 1) + \sum_{a=1}^n ((X_a'')^2 - 1) + \sum_{a=1}^n (X_a^2 - 1)((X_a'')^2 - 1) \\ &\leq n + O(n^{2/3}) \quad \text{by Lemma 5,} \end{aligned}$$

which ensures  $S_1' = O(\sqrt{n \log n})$ . Similar argument gives  $S_3 = O((\log n)^{3/2}/\sqrt{n})$ . So, letting  $Y = Y_{=} + Y_{\neq}$  where the subscript corresponds to summations with  $a = b$  or  $a \neq b$ , recalling that  $\alpha\beta = \Omega(n)$ , we find

$$Y_{=} = O((\log n)^2/\sqrt{n}).$$

It remains to control  $Y_{\neq}$ . Taking into account symmetries between  $X, X', X''$  and between  $a$  and  $b$ , denoting by  $S$  the summation over  $a, b$  in  $[n]$  such that  $a \neq b$  and  $|a - b| < \epsilon n$ , we find (omitting the sums of lesser order as we did for  $Y_{=}$ ) that

$$Y_{\neq} = O\left(\frac{1}{n} \max \left[ S(X_a X_b X_a' X_b''), \theta S(X_a^2 X_b' X_b''), \theta S(X_a X_b X_a' X_b'), \right. \right. \\ \left. \left. \theta^2 S(X_a^2 X_b X_b'), \theta^2 S(X_a^2 (X_b')^2), \theta^3 S(X_a^2 X_b^2) \right] \right).$$

**Control of  $S(X_a^2 X_b^2)$ :** Rewrite it as

$$S(1) + S(X_a^2 - 1) + S(X_b^2 - 1) + S((X_a^2 - 1)(X_b^2 - 1)).$$

The first term  $S(1)$  is  $(1 + o(1))n^2\epsilon$ . The second and third term are weighted sums of i.i.d. centered random variables with weights  $(1 + o(1))(2n\epsilon)$ , and Chernoff bounds entail that these are with probability  $1 - O(n^{-c})$  equal to  $O(n\epsilon\sqrt{n \ln n})$ . The third term is a quadratic form of a centered i.i.d. random vector with sub-exponential tails. Thus by Proposition 1.1 in Götze et al. (2021) we have

$$\mathbb{P}(|S(X_a^2 - 1)S(X_b^2 - 1)| \geq t) \leq 2 \exp\left(-\frac{1}{C} \min\left(\frac{t^2}{\|M\|_F^2}, \sqrt{\frac{t}{\|M\|_{op}}}\right)\right),$$

where  $C$  is a universal constant and  $M_{ab} = \mathbf{1}_{a \neq b, |b-a| < \epsilon n}$ . Here we have  $\|M\|_{op} \leq 2n\epsilon$ ,  $\|M\|_F^2 \leq 2n^2\epsilon$ , and we find that by taking  $t = \Omega(n\epsilon \ln^2 n)$ , we obtain a right-hand side in  $O(n^{-\Omega(1)})$ . Altogether, with probability  $1 - O(n^{-c})$  one has  $S(X_a^2 X_b^2) = O(n^2\epsilon)$ , and thus its contribution to the upper-bound on  $Y_{\neq}$  is  $O(n^{-1}\theta^3 n^2\epsilon) = O(\ln^{3/2} n/n^{1/2})$ .

**Control of  $S(X_a^2 X_b'^2)$ :** Rewrite it as

$$S(1) + S(X_a^2 - 1) + S(X_b'^2 - 1) + S((X_a^2 - 1)(X_b'^2 - 1)).$$

We can apply the same bounds as in the previous paragraph, and we obtain that the contribution to the upper-bound on  $Y_{\neq}$  is with probability  $1 - O(n^{-c})$  in  $O(n^{-1}\theta^2 n^2\epsilon) = O(\epsilon \ln n)$ .

**Control of  $S(X_a X'_a X_b X'_b)$ :** Write  $\tilde{X}_a = X_a X'_a$ . Thus we are again evaluating  $S(\tilde{X}_a \tilde{X}_b)$ , where here  $\tilde{X}_a$  are i.i.d., zero-mean, and sub-exponential. Relying again on Proposition 1.1 in [Götze et al. \(2021\)](#) we have

$$\mathbb{P}(|S(\tilde{X}_a \tilde{X}_b)| \geq t) \leq 2 \exp \left( -\frac{1}{C} \min \left( \frac{t^2}{\|M\|_F^2}, \sqrt{\frac{t}{\|M\|_{op}}} \right) \right),$$

and for  $t = \Omega(n\epsilon \ln^2 n)$  we find that with probability  $1 - O(n^{-c})$ ,  $S(Y_a Y_b) = O(n\epsilon \ln^2 n)$ . Thus the contribution to the upper bound on  $Y_{\neq}$  is  $O(n^{-1}\theta n\epsilon \ln^2 n) = O(\ln^{3/2} n/n^{1/2})$ .

**Control of  $S(X_a^2 X_b X'_b)$ :** This admits, conditionally on  $X$ , a Gaussian distribution with variance  $\sum_b \left( X_b \sum_{a \neq b, |b-a| < n\epsilon} X_a^2 \right)^2$ . Thus this is, with probability  $1 - O(n^{-c})$ , no larger than

$$\sqrt{\ln(n)} \sqrt{\sum_b \left( X_b \sum_{a \neq b, |b-a| < n\epsilon} X_a^2 \right)^2}.$$

We can bound with high  $(1 - O(n^{-c}))$  probability each  $X_a$  by  $O(\sqrt{\ln n})$ , giving an upper bound in

$$O(\sqrt{\ln(n)} \sqrt{n\epsilon^2 \ln^3 n n^2}) = O(\epsilon n^{3/2} \ln^2 n).$$

Thus we get an overall contribution to the upper-bound on  $Y_{\neq}$  bounded by  $O(n^{-1}\theta^2 n^{3/2} \ln^2 n) = O(\ln^3 n/n^{1/2})$ .

**Control of  $S(X_a^2 X'_b X''_b)$ :** Conditionally on  $X, X'$  this is a Gaussian random variable with variance  $\sum_b \left( X'_b \sum_{a \neq b, |b-a| < n\epsilon} X_a^2 \right)^2$ . Again with high probability  $(1 - o(n^{-c}))$  we obtain the upper bound

$$O \left( \sqrt{\ln n} \sqrt{\sum_b \left( X'_b \sum_{a \neq b, |b-a| < n\epsilon} X_a^2 \right)^2} \right)$$

Upper-bounding the entries of  $X$  and  $X'$  by  $O(\sqrt{\ln n})$  this is also

$$O \left( (\ln^2 n) \sqrt{n(n\epsilon)^2} \right) = O((\ln^2 n) \epsilon n^{3/2}).$$

The contribution to the upper-bound of  $Y_{\neq}$  is then  $O(n^{-1}\theta(\ln^2 n) \epsilon n^{3/2}) = O(\epsilon \ln^{5/2} n)$ .

**Control of  $S(X_a X_b X'_a X''_b)$ :** Conditionally on  $X, X'$  this is Gaussian with variance

$$\sum_b \left( X_b \sum_{a \neq b, |b-a| < n\epsilon} X_a X'_a \right)^2.$$

The term  $X_b \sum_{a \neq b, |b-a| < n\epsilon} X_a X'_a$  is Gaussian with variance  $V_b := X_b^2 \sum_{a \neq b, |b-a| < n\epsilon} X_a^2$  conditionally on  $X$ . This variance term  $V_b$  is, with high probability  $(1 - o(n^{-c}))$ ,  $O(\epsilon n \ln^2 n)$ , so that

$(X_b \sum_{a \neq b: |b-a| < n\epsilon} X_a X'_a)^2$  is, with high probability  $(1 - o(n^{-c}))$ ,  $O(\epsilon n \ln^3 n)$ . Thus with high probability  $(1 - o(n^{-c}))$ ,  $S(X_a X_b X'_a X'_b)$  is

$$O\left(\sqrt{\ln n} \sqrt{\epsilon n^2 \ln^3 n}\right) = O(\sqrt{\epsilon} n \ln^2 n).$$

Thus the contribution to the upper-bound on  $Y_{\neq}$  is  $O(n^{-1} \sqrt{\epsilon} n \ln^2 n) = O(\sqrt{\epsilon} \ln^2 n)$ . Of all these 6 terms, the last one is the largest. By a union bound, assuming we took  $c > 2$  in our control in  $(1 - o(n^{-c}))$  of the probability of desired events, we thus have  $|Y_{ij}| \leq O(\sqrt{\epsilon} \ln^2 n)$  for all  $i, j \in [n], i \neq j$ . Thus it follows that with high probability, for all  $i \neq j \in [n]$ ,  $|Y_{ij}| \leq 1/2$  provided  $\epsilon \leq \kappa / \ln^4 n$  for sufficiently small (but still  $\Omega(1)$ ) constant  $\kappa > 0$ .

### A.3. The case of $Y_{i,j}$ for $i = j$

**Bounding  $Y_{kk}$ :** This will follow the same path as for bounding  $Y_{ij}$ , while being slightly simpler. Write

$$Y_{kk} = \sum_{a,b: |b-a| < n\epsilon} (e_a^\top V e)(e_b^\top V e)(e_a^\top V e_k)(e_b^\top V e_k).$$

We use two Gaussian vectors  $X, X'$  as before to form  $V e_k = \alpha^{-1} X$ , and

$$V e = \alpha^{-1} X + \beta^{-1} \sqrt{n-1} (X' - r \alpha^{-1} X),$$

so that

$$\begin{aligned} \alpha^2 Y_{kk} &= \sum_{a,b: |b-a| < n\epsilon} X_a X_b \times \\ &\quad \left( \left[ \frac{1}{\alpha} - \frac{\sqrt{n-1}r}{\alpha\beta} \right] X_a + \frac{\sqrt{n-1}}{\beta} X'_a \right) \left( \left[ \frac{1}{\alpha} - \frac{\sqrt{n-1}r}{\alpha\beta} \right] X_b + \frac{\sqrt{n-1}}{\beta} X'_b \right) \\ &= \sum_{a,b: |b-a| < n\epsilon} X_a X_b (\theta X_a + (1 + o(1)) X'_a) (\theta X_b + (1 + o(1)) X'_b) \\ &= (1 + o(1)) \sum_{a,b: |b-a| < n\epsilon} X_a X_b [X'_a X'_b + \theta (X_a X'_b + X'_a X_b) + \theta^2 X_a X_b], \end{aligned}$$

where as before  $\theta = O(\sqrt{\ln n/n})$ . We first evaluate the sums over  $a = b$ , which read

$$S_0 = \sum_a X_a^2 (X'_a)^2, \quad S_1 = \theta \sum_a X_a^3 X'_a, \quad S_2 = \theta^2 \sum_a X_a^4.$$

By Lemma 5, the first sum can be written

$$S_0 = n + \sum_a X_a^2 (X_a'^2 - 1) + \sum_a (X_a^2 - 1) = n + O(n^{2/3}).$$

Thus, with high probability  $1 - o(n^{-c})$ ,  $S_0 = n + O(n^{2/3})$ . The same argument entails that with high probability  $(1 - o(n^{-c}))$ ,  $S_1 = O(\theta n^{2/3}) = O(\sqrt{\ln n} n^{1/6})$ ,  $S_2 = O(\theta^2 n) = O(\ln n)$ . We decompose again  $Y_{kk}$  as  $Y_{=} + Y_{\neq}$ . The previous evaluations together with  $\alpha^2 = n(1 + o(1))$  give us

$$Y_{=} = 1 + o(1).$$

Next we evaluate  $Y_{\neq}$ . We readily have:

$$Y_{\neq} = O\left(\frac{1}{n} \max [S(X_a X_b X'_a X'_b), \theta S(X_a^2 X_b X'_b), \theta^2 S(X_a^2 X_b^2)]\right).$$

As previously established,  $S(X_a^2 X_b^2) = O(n^2 \epsilon)$ , and the corresponding contribution to the upper-bound on  $Y_{\neq}$  is  $O(n^{-1} \theta^2 n^2 \epsilon) = O(\epsilon \ln n)$ . We also established  $S(X_a^2 X_b X'_b)$  is  $O(\epsilon n^{3/2} \ln^2 n)$ . Thus its contribution to the upper-bound on  $Y_{\neq}$  is  $O(n^{-1} \theta \epsilon n^{3/2} \ln^2 n) = O(\epsilon \ln^{5/2} n)$ . Finally, we have established that  $S(X_a X_b X'_a X'_b) = O(n \epsilon \ln^2 n)$ . This term thus gives a contribution in  $O(\epsilon \ln^2 n)$  to the upper bound. Summarizing, we find that  $Y_{\neq} = O(\epsilon \ln^{5/2} n)$ . Under the previously identified condition that  $\epsilon \leq \kappa / \ln^4 n$ , this is  $o(1)$ . For this regime, we thus have that

$$Y_{kk} = 1 + o(1),$$

which concludes the proof of Lemma 2.

### Appendix B. Proof of Lemma 3.

Introduce the semi-circle density

$$\forall x \in \mathbb{R}, \quad \rho_{sc}(x) := \frac{1}{2\pi} \sqrt{(4 - x^2)_+},$$

and the so-called *classical location*  $\nu_j$  of eigenvalue  $\lambda_j$  of the Gaussian Wigner matrix  $A_2$ , defined through

$$\int_{-\infty}^{\nu_j} \rho_{sc}(x) dx = \frac{j}{n}.$$

The main result in Erdős, Yau and Yin [Erdős et al. \(2012\)](#) implies the existence of some constant  $C = \Theta(1)$  such that with high probability,

$$\forall j \in [n], \quad |\lambda_j - \nu_j| \leq (\ln n)^{C \ln(\ln n)} n^{-2/3}. \quad (32)$$

In fact the results of [Erdős et al. \(2012\)](#) give a stronger guarantee, a precision on the probability of the “rigidity property” to fail, and apply beyond the case of Gaussian Wigner matrices, but their implication (32) is sufficient for our purpose. Note that  $\rho_{sc}(x)$  is upper-bounded by  $1/\pi$  on  $\mathbb{R}$ , so that

$$\forall i, j \in [n], \quad i < j, \quad \int_{\nu_i}^{\nu_j} \rho_{sc}(x) dx \leq \frac{1}{\pi} (\nu_j - \nu_i).$$

By definition of the classical locations  $\nu_j$ , this entails

$$\forall i, j \in [n], \quad i < j, \quad \nu_j - \nu_i \geq \pi \frac{j - i}{n}.$$

Recall Lemma 3’s assumption that  $\epsilon \geq n^{\delta-2/3}$  for some  $\delta = \Omega(1)$ . The term  $(\ln n)^{C \ln(\ln n)}$  in (32) is less than  $n^{\delta/2}$  for large enough  $n$ . It then follows from (32) and the previous display that

$$\begin{aligned} \forall i, j \in [n], \quad i < j, \quad \lambda_j - \lambda_i &\geq \nu_j - \nu_i - 2n^{\delta/2-2/3} \\ &\geq \pi \frac{j-i}{n} - 2n^{\delta/2-2/3}. \end{aligned}$$

This yields

$$\begin{aligned}
 C_\epsilon &= \min_{j-i \geq \epsilon n} (\lambda_j - \lambda_i)^2 \\
 &\geq \min_i (\lambda_{i+\epsilon n} - \lambda_i)^2 \\
 &\geq (\pi\epsilon - 2n^{\delta/2-2/3})^2 \\
 &\geq ((\pi - 1)\epsilon + \Omega(n^{\delta-2/3}) - 2n^{\delta/2-2/3})^2 \\
 &\geq (\pi - 1)^2 \epsilon^2
 \end{aligned}$$

for sufficiently large  $n$ , which concludes the proof.