

---

# Graph Alignment via Birkhoff Relaxation

---

**Sushil Mahavir Varma**  
Industrial and Systems Engineering  
University of Michigan Ann Arbor  
Michigan, USA  
sushilv@umich.edu

**Irène Waldspurger**  
CNRS, INRIA  
Université Paris Dauphine  
Paris, France  
waldspurger@ceremade.dauphine.fr

**Laurent Massoulié**  
INRIA, DI/ENS  
PSL Research University  
Paris, France  
laurent.massoulie@inria.fr

## Abstract

We consider the graph alignment problem, wherein the objective is to find a vertex correspondence between two graphs that maximizes the edge overlap. The graph alignment problem is an instance of the quadratic assignment problem (QAP), known to be NP-hard in the worst case even to approximately solve. In this paper, we analyze Birkhoff relaxation, a tight convex relaxation of QAP, and present theoretical guarantees on its performance when the inputs follow the Gaussian Wigner Model. More specifically, the weighted adjacency matrices are correlated Gaussian Orthogonal Ensemble with correlation  $1/\sqrt{1+\sigma^2}$ . Denote the optimal solutions of the QAP and Birkhoff relaxation by  $\Pi^*$  and  $X^*$  respectively. We show that  $\|X^* - \Pi^*\|_F^2 = o(n)$  when  $\sigma = o(n^{-1})$  and  $\|X^* - \Pi^*\|_F^2 = \Omega(n)$  when  $\sigma = \Omega(n^{-0.5})$ . Thus, the optimal solution  $X^*$  transitions from a small perturbation of  $\Pi^*$  for small  $\sigma$  to being well separated from  $\Pi^*$  as  $\sigma$  becomes larger than  $n^{-0.5}$ . This result allows us to guarantee that simple rounding procedures on  $X^*$  align  $1 - o(1)$  fraction of vertices correctly whenever  $\sigma = o(n^{-1})$ . This condition on  $\sigma$  to ensure the success of the Birkhoff relaxation is state-of-the-art.

## 1 Introduction

Consider two undirected graphs  $G_1$  and  $G_2$  with vertices  $[n]$ . The *graph matching/alignment* problem is defined as finding a mapping between the vertices of  $G_1$  and  $G_2$  such that the edge overlap is maximized. The graph isomorphism problem is a special case when  $G_1 = G_2$  up to the permutation of the vertices. The graph matching problem has widespread applications in network de-anonymization [28], computational biology [30], pattern recognition [6], etc., underlining the need to design and study efficient algorithms for this problem.

More formally, let  $A$  and  $B$  be the adjacency matrices (possibly weighted) of  $G_1$  and  $G_2$  respectively, then, the vertex correspondence that maximizes the edge overlap is given by the optimal solution of the following quadratic assignment problem:

$$\Pi^* = \arg \min_{X \in \mathcal{P}_n} \|AX - XB\|_F^2, \quad (1)$$

where  $\mathcal{P}_n$  is the set of all  $n \times n$  permutation matrices. Quadratic assignment problems are NP-hard in the worst case and are known to be difficult to even solve approximately [24]. However, for typical graphs, one could expect to efficiently solve the graph matching problem. Indeed,

polynomial time algorithms are known for the graph isomorphism problem when  $G_1 = G_2$  is an Erdős-Rényi random graph [5, 7]. More generally, the setting with noise, where  $G_1, G_2$  are not exactly equal but are *correlated* Erdős-Rényi graphs, has been extensively studied in the last decade [8, 9, 11, 12, 18, 17, 26, 25, 27, 2].

A popular class of algorithms in practice with good empirical and computational performance is the convex relaxations of (1). We are interested in the Birkhoff relaxation, which is a tight convex relaxation, wherein the set of permutation matrices  $\mathcal{P}_n$  is replaced by its convex hull to get

$$X^* = \arg \min_{X \in \mathcal{B}_n} \|AX - XB\|_F^2, \quad (2)$$

where  $\mathcal{B}_n$  is the Birkhoff polytope, i.e., the set of all doubly stochastic matrices. Such a relaxation has been an attractive approach in practice, with strong empirical performance for shape matching [1] and ordering images in a grid [10]. It is also observed in [23] that Birkhoff relaxation, when combined with indefinite relaxation, yields excellent empirical performance on real datasets. More generally, similar relaxations are proposed for gene sequencing [14], and computer vision [4].

While these results highlight the usefulness of the Birkhoff relaxation to practice due to its attractive computational and empirical performance, there is limited understanding of its theoretical performance. Theoretical guarantees for alternate convex relaxations of the QAP (1) are available in the literature [11, 12, 2], however, they lack in their empirical performance compared to the Birkhoff relaxation (see Section 5). Motivated by this gap in the literature, we provide the first theoretical guarantees on the performance of the Birkhoff relaxation (2) when  $A, B$  are sampled from the Gaussian Wigner Model with correlation  $1/\sqrt{1 + \sigma^2}$ . Imposing such distributional assumptions on the input graphs allows for mathematical analysis, as otherwise, graph matching is an NP-hard problem in the worst case.

## 1.1 Gaussian Wigner Model

We say that a matrix  $A := \{A_{ij}\}_{i,j \in [n]}$  is a (GOE) matrix if  $A_{ii} \sim N(0, 2/n)$  for all  $i \in [n]$ ,  $A_{ij} = A_{ji} \sim N(0, 1/n)$  for all  $i \neq j$ , and  $\{A_{ij} : i \leq j\}$  are mutually independent. We say that  $A, B \in \mathbb{R}^{n \times n}$  follows the Gaussian Wigner Model when for all  $i, j \in [n]$ , we have  $B_{\pi^*(i), \pi^*(j)} = A_{ij} + \sigma Z_{ij}$ , where  $A, Z$  are i.i.d. GOE matrices. Alternatively, we can write  $B = (\tilde{\Pi}^*)^T (A + \sigma Z) (\tilde{\Pi}^*)$ , where  $\tilde{\Pi}_{ij}^* = \mathbb{1}\{j = \pi^*(i)\}$ . Given an observation of  $A$  and  $B$ , the goal is to infer the ground truth  $\pi^*$  that aligns  $A$  and  $B$ , which corresponds to the optimal solution of QAP (1) for  $\sigma^2 \leq O(n/\log n)$  by [15, 32]. In particular, [15, 32] shows that  $\tilde{\Pi}^* = \Pi^*$  with probability  $1 - o(1)$  for  $\sigma^2 \leq O(n/\log n)$  and  $n$  large enough. We fix  $\sigma \leq 1$  for simplicity.

## 1.2 Main Contributions

In this paper, we establish bounds on the distance of the optimal solution of the Birkhoff relaxation  $X^*$  from the optimal solution of the QAP  $\Pi^*$ , which is equal to the true permutation  $\tilde{\Pi}^*$  with high probability. Our contributions are two-fold.

Firstly, we show that  $\|X^* - \Pi^*\|_F^2 = o(n)$  whenever  $\sigma = o(n^{-1})$ , asserting that  $X^*$  is close to the true permutation  $\Pi^*$ . This bound in turn implies that one can use simple rounding procedures on  $X^*$  to correctly align  $1 - o(1)$  fraction of vertices. Our proof technique is motivated by the literature on the sensitivity analysis of convex problems. In particular, we formulate the dual of a version of (2) for  $\sigma = 0$ . The novelty of the proof lies in the careful construction of a feasible dual solution, which in turn, provides us useful bounds on  $X^*$ . Secondly, we show that if  $\sigma = \Omega(n^{-0.5})$ , then  $\|X^* - \Pi^*\|_F^2 = \Omega(n)$ , and so,  $X^*$  is far away from the true permutation  $\Pi^*$ .

The above two results establish a phase transition in the behavior of (2): the optimal solution  $X^*$  transitions from being a small perturbation of  $\Pi^*$  for slowly growing  $\sigma$  to moving far away from  $\Pi^*$  as  $\sigma$  becomes greater than  $1/\sqrt{n}$ . In the next section, we discuss that the population version of (2) exhibits such a phase transition at  $1/\sqrt{n}$  and so we believe our sufficient condition on  $\sigma$  is loose, i.e., one can possibly show  $\|X^* - \Pi^*\|_F^2 = o(n)$  for  $\sigma = o(n^{-0.5})$ . Nonetheless, to the best of our knowledge, the sufficient condition  $\sigma = o(n^{-1})$  is the state-of-the-art to ensure that (2) succeeds in recovering the true permutation.

Note that  $\sigma = 0$  is the special case of the graph isomorphism problem. As the value of  $\sigma$  increases, the correlation between the two graphs reduces, which makes it harder to align them. In other words, noise increases, making it harder to extract the correct signal. Thus,  $\sigma$  serves as a tuning parameter to increase the hardness of the problem. The takeaway is then to establish that the Birkhoff relaxation succeeds for large values of  $\sigma$ , establishing the robustness of the relaxation. We make progress in this direction by improving upon the previously known result [2] that shows the simplex relaxation succeeds for  $\sigma = 0$ . These results contribute to explaining the strong empirical performance of the Birkhoff relaxation.

### 1.3 Related Work

Regarding practical algorithms for graph alignment, there is a rich literature on optimization-based methods [21, 10, 22, 11, 2] and simple spectral-based methods [31, 13, 16]. The works [11, 2] focus on the Gaussian Wigner model and are closest to our paper. In particular, [11] proposes the GRAMPA algorithm which solves the following convex relaxation of (1):

$$\arg \min_{X: \mathbf{1}^T X \mathbf{1} = n} \|AX - XB\|_F^2 + \eta \|X\|_F^2, \quad \text{for some } \eta > 0. \quad (3)$$

Note that the optimization problem above further relaxes (2) and adds a quadratic regularizer. The authors in [11] establish that GRAMPA (3) succeeds in recovering  $\Pi^*$  whenever  $\sigma = O(1/\log n)$ . While [11] asserts that GRAMPA succeeds with a weaker condition on  $\sigma$  than ours, its empirical performance is observed to be worse than the Birkhoff relaxation (2). In addition, GRAMPA requires one to tune the regularization parameter  $\eta$  which is completely avoided for the Birkhoff relaxation (2). In addition, [2] analyzes the Simplex Relaxation defined as follows:

$$\arg \min_{X: \mathbf{1}^T X \mathbf{1} = n, X \geq 0} \|AX - XB\|_F^2. \quad (4)$$

They prove that the simplex relaxation (4) succeeds for  $\sigma = 0$ . On the other hand, we consider a tighter convex relaxation and allow  $\sigma$  to be non-zero. Comparison to these methods is summarized in Table 1. We refer the reader to [19, 17, 9, 27] and the references within for analysis beyond the GOE setting.

| Paper    | Algorithm Type                | Algorithm Name | Noise Threshold             |
|----------|-------------------------------|----------------|-----------------------------|
| [16]     | Top Eigenvector Alignment     | EIG1           | $\sigma = \Theta(n^{-7/6})$ |
| [11]     | Regularized Convex Relaxation | GRAMPA (3)     | $\sigma = O(1/\log n)$      |
| [2]      | Simplex Relaxation            | Simplex (4)    | $\sigma = 0$                |
| Our work | Birkhoff Relaxation           | Birkhoff (2)   | $\sigma = O(n^{-1})$        |

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

## 2 Main Result

We now present the main theorem of the paper that characterizes when  $X^*$  is well separated from  $\Pi^*$  and when  $X^*$  is a small perturbation of  $\Pi^*$ .

**Theorem 1.** *Let  $A, B \in \mathbb{R}^{n \times n}$  follow the Gaussian Wigner model with some  $\sigma \leq 1$  (possibly dependent on  $n$ ) and let  $\Pi^*$  and  $X^*$  be defined as in (1) and (2) respectively. Then, for any fixed  $\epsilon > 0$  and  $n$  large enough, the following statements hold with probability  $1 - o(1)$ .*

- **Well-Separation:** *When  $\sigma \geq n^{-0.5+\epsilon}$  then  $\|X^* - \Pi^*\|_F^2 \geq \delta n$  for some  $\delta > 0$ .*
- **Small-Perturbation:** *When  $\sigma \leq n^{-1-\epsilon}$  then  $\|X^* - \Pi^*\|_F^2 \leq 10n^{1-\epsilon/32}$ .*

Using the small-perturbation result above, one can implement a simple rounding procedure on  $X^*$  to recover a  $1 - o(1)$  fraction of the permutation  $\pi^*$  correctly.

**Corollary 2.** *Under the hypothesis of Theorem 1, let  $\hat{\pi}(i) = \arg \max_j X_{ij}^*$  for all  $i \in [n]$ . If  $\sigma \leq n^{-1-\epsilon}$  for some  $\epsilon > 0$ , then  $\sum_{i=1}^n \mathbb{1} \{\hat{\pi}(i) \neq \pi^*(i)\} = o(n)$  with probability  $1 - o(1)$  for  $n$  large enough.*

In practice, one may want to implement more sophisticated rounding procedure such as Hungarian projection. Theorem 1 ensures that Hungarian projection also succeeds to recover a  $1 - o(1)$  fraction of the permutation  $\pi^*$  correctly.

**Corollary 3.** *Under the hypothesis of Theorem 1, let  $\tilde{\Pi} \in \arg \max_{\Pi \in \mathcal{P}_n} \langle X^*, \Pi \rangle$ . If  $\sigma \leq n^{-1-\epsilon}$  for some  $\epsilon > 0$ , then  $\langle \tilde{\Pi}^*, \tilde{\Pi} \rangle \geq n - o(n)$  with probability  $1 - o(1)$  for  $n$  large enough.*

We refer the reader to Appendix A for proof details. Note that a similar negative result for the well-separated case may not hold. In particular, while  $X^*$  is far away from  $\Pi^*$  for  $\sigma = \Omega(n^{-0.5})$ , it does not imply that the Birkhoff relaxation fails to recover the permutation  $\pi^*$  correctly, as rounding procedures like above can be used to post-process  $X^*$  which could recover  $\Pi^*$ . As our proof technique is based on sensitivity analysis of convex problems, we do not explore this direction in this paper, except in Section 5, where we conduct numerical experiments to provide insights.

To gain intuition, consider the population version of (2), where we replace the objective with its expected value to get the following optimization problem for  $\tilde{\Pi}^* = I$ :

$$\min_{X \in \mathcal{B}_n} (2 + \sigma^2)(n + 1) \|X\|_F^2 - 2\text{Tr}(X)^2 - 2\langle X, X^T \rangle,$$

whose optimal value is given by

$$\bar{X}^* = \epsilon I + \frac{1 - \epsilon}{n} J, \quad \text{where} \quad \epsilon = \frac{2}{2 + \sigma^2(n + 1)},$$

where we denote the all-ones  $n \times n$  matrix by  $J$ . The above can be verified using the KKT conditions. Now, using the above characterization, we get

$$\|I - \bar{X}^*\|_F = (1 - \epsilon) \left\| I - \frac{J}{n} \right\| = (1 - \epsilon) \sqrt{n - 1} \approx \frac{\sigma^2 n \sqrt{n}}{2 + \sigma^2 n} = \begin{cases} \Theta(\sqrt{n}) & \text{if } \sigma \sqrt{n} = \Omega(1) \\ o(\sqrt{n}) & \text{if } \sigma \sqrt{n} = o(1). \end{cases}$$

The first case above asserts that whenever  $\sigma \gg n^{-0.5}$ , we have  $\|I - \bar{X}^*\| = \Theta(\sqrt{n})$ , i.e.,  $\bar{X}^*$  is far from  $I$ , so, we should not expect  $X^*$  to be close to  $I$  as well. The first assertion of Theorem 1 formalizes this intuition. The above equation also shows that  $\bar{X}^*$  is close to  $I$  whenever  $\sigma \ll n^{-0.5}$ , so we expect  $X^*$  to be a small perturbation of  $I$ . Theorem 1 formalizes this only for  $\sigma \ll n^{-1}$ , which only partially resolves this case. However, analyzing the Birkhoff convex relaxation (2) is known to be a challenging task, and this paper provides the first results in understanding this relaxation. In particular, one of the main difficulties is in handling the non-negativity constraints. The paper [11] circumvents such a difficulty by relaxing the non-negativity constraints and compensating for it with a quadratic regularizer in the objective function. Such a modification allows them to obtain a closed-form expression of  $X^*$ , which is then shown to satisfy certain desirable properties (diagonal dominance). More recently, [2] preserves the non-negativity constraints, but the result is restricted to  $\sigma = 0$  and the proof exploits the structural properties of (2) that hold only when  $B = A$ . We, on the other hand, tackle the challenges head-on presented by non-negativity constraints and also allow  $\sigma > 0$ . Our proof technique is to carefully construct a suitable feasible solution of the dual of (2), which provides required bounds on  $X^*$ . The main difficulty in constructing such a feasible dual certificate is its high dimension as the non-negativity constraints result in  $n^2$  dual variables.

### 3 Proof of Theorem 1

Without loss of generality, we assume that  $\tilde{\Pi}^* = I$ . Indeed, if  $\tilde{\Pi}^* \neq I$ , then, we have

$$\|AX - XB\|_F = \|AX - X(\tilde{\Pi}^*)^T(A + \sigma Z)\tilde{\Pi}^*\|_F = \|AX(\tilde{\Pi}^*)^T - X(\tilde{\Pi}^*)^T(A + \sigma Z)\|_F.$$

Thus, defining

$$\tilde{X} = \arg \min_{X \in \mathcal{B}_n} \|AX - X(A + \sigma Z)\|_F^2,$$

we conclude that  $\tilde{X} = X^*(\tilde{\Pi}^*)^T$  and so  $\|X^* - \Pi^*\|_F = \|\tilde{X} - \Pi^*(\tilde{\Pi}^*)^T\|_F = \|\tilde{X} - I\|_F$ , where the last equality holds as  $\tilde{\Pi}^* = \Pi^*$  with probability  $1 - o(1)$  by [15, 32]. So, in summary, we have  $\tilde{\Pi}^* = \Pi^* = I$  without loss of generality.

### 3.1 Part I: Well-Separation

*Proof of Theorem 1 (Part I):*. We start by upper bounding  $\|AX^* - X^*B\|$  by the objective function value of (2) for  $J/n$ .

**Claim 4.** For  $n$  large enough, with probability at least  $1 - 4ne^{-n^\epsilon/4}$ , we have

$$\|AX^* - X^*B\|_F^2 \leq \frac{1}{n^2} \|AJ - JB\|_F^2 \leq 9n^\epsilon$$

The proof details of the above claim is deferred to Appendix B.1. We now construct a lower bound on  $\|AX^* - X^*B\|_F^2$  as a function of  $\|X^* - I\|_F$ . We have

$$\begin{aligned} \|AX^* - X^*B\|_F^2 &= \|A(I - X^*) - (I - X^*)B + \sigma Z\|_F^2 \\ &= \|A(I - X^*) - (I - X^*)B\|_F^2 + \sigma^2 \|Z\|_F^2 + 2\sigma \langle A(I - X^*) - (I - X^*)B, Z \rangle \\ &= \|A(I - X^*) - (I - X^*)B\|_F^2 + \sigma^2 \|Z\|_F^2 + 2\sigma \langle AZ - ZB, I - X^* \rangle \\ &= \|A(I - X^*) - (I - X^*)B\|_F^2 + \sigma^2 \|Z\|_F^2 - 2\sigma \langle AZ - ZA, X^* \rangle - 2\sigma^2 \langle Z^2, I - X^* \rangle \\ &\geq \frac{\sigma^2 n}{2} - 2\sigma \max_{i \neq j} |(AZ - ZA)_{ij}| \sum_{i \neq j} X_{ij}^* - 2\sigma^2 \|Z^2\|_F \|I - X^*\|_F, \end{aligned} \quad (5)$$

where the last inequality is true because  $\|Z\|_F^2 = \frac{2}{n} \|\tilde{z}\|_2^2$ , where  $\tilde{z} \sim N(0, I_{n(n+1)/2})$ . Thus, by [11, Lemma 15], with probability at least  $1 - e^{-\sqrt{n}}$ , we have  $\|Z\|_F^2 \geq n - \tilde{c}\sqrt{n} \geq n/2$ , for a sufficiently large  $n$  since  $\tilde{c} > 0$  is a constant independent of  $n$ . We now bound  $\max_{i \neq j} |(AZ - ZA)_{ij}|$  in the following claim.

**Claim 5.** For  $n$  large enough, w.p.  $1 - 2n^2e^{-n^\epsilon}$ , we have  $\max_{i \neq j} |(AZ - ZA)_{ij}| \leq 8n^{\epsilon/2-0.5}$ .

The proof details of the claim is deferred to Appendix B.2. Substituting the above bound back in (5) and noting that  $\sum_{i \neq j} X_{ij}^* \leq n$  as  $X^* \in \mathcal{B}_n$ , we get

$$\begin{aligned} \|AX^* - X^*B\|_F^2 &\geq \frac{\sigma^2 n}{2} - 16\sigma n^{1/2+\epsilon/2} - 2\sigma^2 \|Z^2\|_F \|I - X^*\|_F \\ &\geq \frac{\sigma^2 n}{4} - 2c\sigma^2 \sqrt{n} \|I - X^*\|_F. \end{aligned} \quad (6)$$

where the last inequality holds for  $n$  large enough as  $\sigma^2 n = \Theta(n^{2\epsilon})$  and  $\sigma n^{1/2+\epsilon/2} = \Theta(n^{3\epsilon/2})$ . In addition, we use the inequality  $\|Z^2\|_F \leq \sqrt{n} \|Z^2\|_2 \leq \sqrt{n} \|Z\|_2^2 \leq c\sqrt{n}$  for some  $c > 0$  sufficiently large, which holds with probability at least  $1 - e^{-n}$  by [3, Lemma 6.3]. Now combining the upper bound in Claim 4 with the lower bound in (6), we get

$$2c\sigma^2 \sqrt{n} \|I - X^*\|_F \geq \frac{\sigma^2 n}{4} - 9n^\epsilon \stackrel{*}{\geq} \frac{\sigma^2 n}{8} \implies \|I - X^*\|_F \geq \frac{\sqrt{n}}{16c}.$$

where  $(*)$  holds for  $n$  large enough. The above assertion holds with probability at least  $1 - 5n^2e^{-n^\epsilon/4} \geq 1 - e^{-n^{\epsilon/2}}$  for  $n$  large enough by the union bound over all the high probability bounds in the proof. This completes the proof with  $\delta = 1/(16c)$ , a universal constant.  $\square$

Note that the above proof does not rely on specific properties of a GOE matrices. Thus, such a conclusion would hold for more general symmetric matrices with i.i.d. entries that concentrate sufficiently, e.g., subgaussian concentration.

### 3.2 Part II: Small-Perturbation

The main idea of the proof is to construct a suitable dual feasible solution that provides an upper bound on the off-diagonal entries of  $X^*$ . As the primal problem (2) is non-linear, resulting in an involved dual problem, we start up considering a simpler optimization problem corresponding to  $\sigma = 0$  in (2). As we will see below, this analysis, in turn, allows us to conclude meaningful bounds on  $X^*$  for  $\sigma = o(n^{-1})$ . For  $\sigma = 0$ , we have  $\min_{X \in \mathcal{B}_n} \|AX - XA\|_F^2$ . Note that  $X = I$  is an

optimal solution as it results in 0 objective function value. Thus, any optimal solution must satisfy  $AX - XA = 0$ , and so, we can further simplify to write the following optimization problem.

$$\min_{X \in \mathcal{B}_n} 0 \quad \text{subject to } AX - XA = 0.$$

We introduce dual variables  $R \in \mathbb{R}_+^{n \times n}$ ,  $\mu, \tilde{\mu} \in \mathbb{R}^n$  for the non-negativity, row sum, and column sum constraints, and  $M \in \mathbb{R}^{n \times n}$  for the constraint  $AX - XA = 0$ . Then, the Lagrangian is given by

$$\min_X \max_{R \geq 0, \mu, \tilde{\mu}, M} \langle AX - XA, M \rangle - \langle R, X \rangle + \mu^T X \mathbf{1} + \mathbf{1}^T X \tilde{\mu} - \mu^T \mathbf{1} - \tilde{\mu}^T \mathbf{1}.$$

Now, we swap the order of the min and the max to obtain the dual problem.

$$\begin{aligned} & \max_{R \geq 0, \mu, \tilde{\mu}, M} \min_X \langle AX - XA, M \rangle - \langle R, X \rangle + \mu^T X \mathbf{1} + \mathbf{1}^T X \tilde{\mu} - \mu^T \mathbf{1} - \tilde{\mu}^T \mathbf{1} \\ &= \max_{R \geq 0, \mu, \tilde{\mu}, M} \min_X \langle AM - MA - R + \mu \mathbf{1}^T + \mathbf{1} \tilde{\mu}^T, X \rangle - \mu^T \mathbf{1} - \tilde{\mu}^T \mathbf{1} \\ &= \max_{R \geq 0, \mu, \tilde{\mu}, M} -\mu^T \mathbf{1} - \tilde{\mu}^T \mathbf{1} \quad \text{subject to, } AM - MA - R + \mu \mathbf{1}^T + \mathbf{1} \tilde{\mu}^T = 0. \end{aligned}$$

We are now looking for a dual feasible solution  $(R, \mu, \tilde{\mu}, M)$  that also satisfies strong duality. We construct an approximately feasible dual solution, i.e., we have

$$\mu^T \mathbf{1} + \tilde{\mu}^T \mathbf{1} = 0, \quad R \geq 0, \quad AM - MA - R + \mu \mathbf{1}^T + \mathbf{1} \tilde{\mu}^T \approx 0 \quad (7)$$

which implies for any  $X \in \mathcal{B}_n$  with  $AX - XA = 0$

$$\begin{aligned} \langle R, X \rangle &= \langle R, X \rangle - \langle AX - XA, M \rangle - \mu^T \mathbf{1} - \tilde{\mu}^T \mathbf{1} \\ &= \langle R, X \rangle - \langle AM - MA, X \rangle - \langle \mu \mathbf{1}^T + \mathbf{1} \tilde{\mu}^T, X \rangle \\ &= \langle R - AM + MA - \mu \mathbf{1}^T - \mathbf{1} \tilde{\mu}^T, X \rangle \approx 0. \end{aligned}$$

To get a meaningful bound, we set  $R = J - I$  and appropriately construct  $(M, \mu, \tilde{\mu})$  which implies

$$\sum_{j \neq i} X_{ij} \approx 0.$$

We formalize this argument in the lemma below:

**Lemma 6.** *Under the hypothesis of Theorem 1, for any  $X \in \mathcal{B}_n$  and any  $\epsilon > 0$ , for  $n$  large enough, we have*

$$\sum_{j \neq i} X_{ij} \leq 2n^{3/2+7\epsilon/8} \|AX - XA\|_F + 4n^{1-\epsilon/32} \quad \text{w.p. } 1 - o_n(1).$$

Now to show that the bound obtained in the above lemma is small enough, we upper bound  $\|AX - XA\|_F$  for the optimal solution  $X = X^*$  in the following lemma.

**Lemma 7.** *Under the setting of Theorem 1, there exists a constant  $c > 0$  such that  $\|AX^* - X^*A\|_F \leq c\sigma\sqrt{n}$  with probability at least  $1 - e^{-n}$  for  $n$  large enough.*

The proof strategy of the above lemma is as follows: As  $I \in \mathcal{B}_n$  is a feasible solution to (2), we immediately obtain  $\|AX^* - X^*B\|_F \leq \|A - B\|_F = \sigma\|Z\|_F = O(\sigma\sqrt{n})$ , where the last inequality follows as  $\|Z\|_F = O(\sqrt{n})$  for a GOE matrix. The rest of the argument is to show that  $\|AX^* - X^*A\|_F \lesssim \|AX^* - X^*B\|_F + O(\sigma\sqrt{n})$  by writing  $B = A + \sigma Z$ , expanding  $\|AX^* - X^*B\|_F$ , and appropriately bounding terms involving  $\sigma$ . The proof details are deferred to Appendix C.1.

*Proof of Theorem 1 (Part II):* By Lemma 6 and Lemma 7, for  $\sigma = n^{-1-\epsilon}$  we get

$$\sum_{i,j \in [n]: i \neq j} X_{ij}^* \leq 2c\sigma n^{2+7\epsilon/8} + 4n^{1-\epsilon/32} = 2cn^{1-\epsilon/8} + 4n^{1-\epsilon/32} \leq 5n^{1-\epsilon/32}, \quad (8)$$

where the first inequality holds with probability at least  $1 - e^{-n} - o_n(1) = 1 - o_n(1)$  by the union bound. The last inequality holds for all  $n$  large enough (depending on  $\epsilon$ ). We are now ready to obtain a bound on  $\|X^* - I\|_F$ .

$$\begin{aligned}\|X^* - I\|_F^2 &= \sum_{i,j \in [n]: i \neq j} (X_{ij}^*)^2 + \sum_{i=1}^n (1 - X_{ii}^*)^2 = \|X^*\|_F^2 + n - 2 \sum_{i=1}^n X_{ii}^* \stackrel{*}{\leq} 2n - 2 \sum_{i=1}^n X_{ii}^* \\ &= 2 \sum_{i,j \in [n]: i \neq j} X_{ij}^* \stackrel{(8)}{\leq} 10n^{1-\epsilon/32},\end{aligned}$$

where  $(*)$  follows as  $\|X\|_F^2 \leq n \max_{j \in [n]} \sum_{i=1}^n X_{ij}^2 \leq n \max_{j \in [n]} \sum_{i=1}^n |X_{ij}| = n$  for any  $X \in \mathcal{B}_n$ . This completes the second part of the proof of Theorem 1.

**Remark 8** (Generalization beyond GOE Matrices). *The main difficulties are to get a handle on the eigenvalues and eigenvectors of  $A$ . In particular, we explicitly use the properties of a GOE matrix in two places in the proof. (1) Eigenvalue separation [Claim 8]: We use tail bounds on the eigenvalues of a random matrix from [5]. The results of [5] are applicable for all Wigner matrices (i.i.d. subgaussian entries), and so Claim 8 can be extended to Wigner matrices. (2) Eigenvector Concentration [Claim 7]: To prove Claim 7, we rely on the fact that the orthonormal eigenvectors of a GOE matrix is uniformly distributed on  $\mathcal{S}_{n-1}$  to obtain upper and lower concentrations on  $\langle \mathbf{1}, u_i \rangle$ . This step is the bottleneck, as we require such concentration results for general Wigner matrices.*

## 4 Proof of Lemma 6

*Proof of Lemma 6.* For the GOE matrix  $A$ , let  $\{\lambda_i\}_{i=1}^n$  be the set of eigenvalues and  $\{u_i\}_{i=1}^n$  be the corresponding set of orthonormal eigenvectors. Now, we construct a dual feasible solution  $(R, M, \mu, \tilde{\mu})$  using eigenvectors of  $A$  as the basis vectors. First, we set  $\tilde{\mu} = \mathbf{0}$  and

$$R = \sum_{i,j=1}^n w_{ij} u_i u_j^T = J - I, \quad \text{with } w_{ij} = \langle u_i, \mathbf{1} \rangle \langle u_j, \mathbf{1} \rangle - \mathbb{1}\{i = j\}.$$

Now, to ensure approximate dual feasibility and strong duality as in (7), we carefully set  $(M, \mu)$  as follows. We have

$$\mu = \sum_{i=1}^n w_i u_i \quad \text{with } w_i = \langle u_i, \mathbf{1} \rangle + \frac{C-1}{\langle u_i, \mathbf{1} \rangle} \mathbb{1}\left\{|\langle u_i, \mathbf{1} \rangle| \geq n^{-\epsilon/16}\right\}$$

where  $C = 1 - \frac{n}{\#\{|\langle u_i, \mathbf{1} \rangle| \geq n^{-\epsilon/16}\}}$ , with  $\#\{|\langle u_i, \mathbf{1} \rangle| \geq n^{-\epsilon/16}\} = \sum_{i=1}^n \mathbb{1}\{|\langle u_i, \mathbf{1} \rangle| \geq n^{-\epsilon/16}\}$ .

Note that  $\langle \mu, \mathbf{1} \rangle = 0$  which ensures strong duality. One can quickly verify it as follows:

$$\langle \mu, \mathbf{1} \rangle = \sum_{i=1}^n w_i \langle u_i, \mathbf{1} \rangle = \sum_{i=1}^n \langle u_i, \mathbf{1} \rangle^2 + (C-1) \#\{|\langle u_i, \mathbf{1} \rangle| \geq n^{-\epsilon/16}\} = \sum_{i=1}^n \langle u_i, \mathbf{1} \rangle^2 - n = 0.$$

Next, we set

$$M = \sum_{i,j \in [n]: i \neq j} \frac{\tilde{w}_{ij}}{\lambda_i - \lambda_j} u_i u_j^T \quad \text{with } \tilde{w}_{ij} = (\langle u_i, \mathbf{1} \rangle \langle u_j, \mathbf{1} \rangle - \langle u_i, \mathbf{1} \rangle w_j) \mathbb{1}\{i \neq j\}.$$

Now, we show that  $(R, \mu, \tilde{\mu}, M)$  is approximately dual feasible as in (7). By construction, we have

$$\begin{aligned}R - (AM - MA) - \mathbf{1}\mu^T &= \sum_{i=1}^n w_{ii} u_i u_i^T + \sum_{i,j \in [n]: i \neq j} \langle u_i, \mathbf{1} \rangle w_j u_i u_j^T - \mathbf{1}\mu^T \\ &= \sum_{i=1}^n w_{ii} u_i u_i^T + \sum_{i,j \in [n]: i \neq j} \langle u_i, \mathbf{1} \rangle w_j u_i u_j^T - \sum_{i=1}^n \langle u_i, \mathbf{1} \rangle u_i \mu^T \\ &= \sum_{i=1}^n w_{ii} u_i u_i^T - \sum_{i=1}^n \langle u_i, \mathbf{1} \rangle w_i u_i u_i^T \\ &= - \sum_{i \in [n]: |\langle u_i, \mathbf{1} \rangle| \leq n^{-\epsilon/16}} u_i u_i^T - C \sum_{i \in [n]: |\langle u_i, \mathbf{1} \rangle| \geq n^{-\epsilon/16}} u_i u_i^T \triangleq D.\end{aligned}$$

Now, we show that  $\|D\|_F^2 = o(n)$  asserting that  $R - (AM - MA) - \mathbf{1}\mu^T$  is small. As the Frobenius norm is unitary invariant, we have

$$\begin{aligned}\|D\|_F &= \sqrt{\#\{|\langle u_i, \mathbf{1} \rangle| \leq n^{-\epsilon/16}\} + C^2 \#\{|\langle u_i, \mathbf{1} \rangle| \geq n^{-\epsilon/16}\}} \\ &\leq \sqrt{\#\{|\langle u_i, \mathbf{1} \rangle| \leq n^{-\epsilon/16}\} + C^2 n}.\end{aligned}$$

Next, we get a handle on  $C$  using the following claim, proved at the end of this section.

**Claim 9.** *There exists a constant  $c > 0$  such that, for large enough  $n > 0$ , we have, with a probability of at least  $1 - e^{-cn^{1-\epsilon/16}}$ ,  $\#\{|\langle u_i, \mathbf{1} \rangle| \leq n^{-\epsilon/16}\} \leq 3n^{1-\epsilon/16}$ .*

To prove the above claim, we use the fact that the orthonormal eigenvector basis of a GOE matrix is uniformly distributed on  $\mathcal{S}_{n-1}$  and the details are provided in Appendix C.2. Thus,  $\langle u_i, \mathbf{1} \rangle$  is approximately a standard normal and so  $\#\{|\langle u_i, \mathbf{1} \rangle| \leq n^{-\epsilon/16}\} \approx n^{1-\epsilon/16}$ . By the above claim

$$0 \geq C = 1 - \frac{n}{n - \#\{|\langle u_i, \mathbf{1} \rangle| \leq n^{-\epsilon/16}\}} \geq 1 - \frac{n}{n - 3n^{1-\epsilon/16}} \geq -5n^{-\epsilon/16}, \quad (9)$$

where the last inequality follows for  $n > 0$  large enough (depending on  $\epsilon$ ). The bound on  $C$  implies

$$\|D\|_F \leq \sqrt{\#\{|\langle u_i, \mathbf{1} \rangle| \leq n^{-\epsilon/16}\} + C^2 n} \leq \sqrt{3n^{1-\epsilon/16} + 25n^{1-\epsilon/8}} \leq 2n^{1/2-\epsilon/32}, \quad (10)$$

where the last inequality holds for  $n$  large enough (depending on  $\epsilon$ ). Now, for any  $X \in \mathcal{B}_n$ , we have

$$\begin{aligned}\langle R - (AM - MA) - \mathbf{1}\mu^T, X \rangle &= \langle R, X \rangle - \langle AM - MA, X \rangle - \langle \mathbf{1}\mu^T, X \rangle \\ &= \sum_{i,j \in [n]: i \neq j} X_{ij} - \langle M, AX - XA \rangle - \langle \mu, \mathbf{1} \rangle = \sum_{i,j \in [n]: i \neq j} X_{ij} - \langle M, AX - XA \rangle,\end{aligned}$$

where the last inequality holds as  $\langle \mu, \mathbf{1} \rangle = 0$  by construction. Thus, we have

$$\sum_{i,j \in [n]: i \neq j} X_{ij} = \langle M, AX - XA \rangle + \langle X, D \rangle \stackrel{(10)}{\leq} \langle M, AX - XA \rangle + 2n^{1-\epsilon/32},$$

where the last inequality holds as  $\langle X, D \rangle \leq \|X\|_F \|D\|_F$  and  $\|X\|_F^2 \leq n \max_{j \in [n]} \sum_{i=1}^n X_{ij}^2 \leq n \max_{j \in [n]} \sum_{i=1}^n |X_{ij}| = n$  for any  $X \in \mathcal{B}_n$ . Now, to complete the proof, we upper bound  $\langle M, AX - XA \rangle$  below. Let  $X = \sum_{i,j=1}^n x_{ij} u_i u_j^T$  for some  $x_{ij} \in \mathbb{R}$  for all  $i, j \in [n]$ . Then,

$$\begin{aligned}\langle M, AX - XA \rangle &= \left\langle \sum_{i,j \in [n]: i \neq j} \frac{\tilde{w}_{ij}}{\lambda_i - \lambda_j} u_i u_j^T, \sum_{i,j \in [n]: i \neq j} x_{ij} (\lambda_i - \lambda_j) u_i u_j^T \right\rangle \\ &= \left\langle \sum_{i,j \in [n]: i \neq j} \frac{\tilde{w}_{ij}}{|\lambda_i - \lambda_j| + n^{-1-\epsilon}} u_i u_j^T, \sum_{i,j \in [n]: i \neq j} x_{ij} (|\lambda_i - \lambda_j| + n^{-1-\epsilon}) u_i u_j^T \right\rangle \\ &\stackrel{*}{\leq} \sqrt{\sum_{i,j \in [n]: i \neq j} \frac{\tilde{w}_{ij}^2}{(|\lambda_i - \lambda_j| + n^{-1-\epsilon})^2}} \cdot (\|AX - XA\|_F + n^{-0.5-\epsilon}) \quad (11)\end{aligned}$$

where  $(*)$  follows from the Cauchy-Schwarz inequality. In addition, we upper bound the second term by using the triangle inequality and noting that  $n \geq \|X\|_F^2 = \sum_{i,j=1}^n x_{ij}^2 \geq \sum_{i,j \in [n]: i \neq j} x_{ij}^2$  for all  $X \in \mathcal{B}_n$ . In addition, we also note that  $\sum_{i,j \in [n]: i \neq j} x_{ij}^2 (\lambda_i - \lambda_j)^2 = \|AX - XA\|_F^2$ . Now, we focus on getting a handle on the term involving the eigenvalue separation  $(\lambda_i - \lambda_j)$ . We have

$$\begin{aligned}&\sqrt{\sum_{i,j \in [n]: i \neq j} \frac{\tilde{w}_{ij}^2}{(|\lambda_i - \lambda_j| + n^{-1-\epsilon})^2}} \\ &= (1 - C) \sqrt{\sum_{i,j \in [n]: i \neq j} \frac{\langle u_i, \mathbf{1} \rangle^2}{\langle u_j, \mathbf{1} \rangle^2 (|\lambda_i - \lambda_j| + n^{-1-\epsilon})^2} \mathbb{1}\{|\langle u_j, \mathbf{1} \rangle| \geq n^{-\epsilon/16}\}} \\ &\stackrel{*}{\leq} 2n^{\epsilon/16} \sqrt{\sum_{i,j \in [n]: i \neq j} \frac{\langle u_i, \mathbf{1} \rangle^2}{(|\lambda_i - \lambda_j| + n^{-1-\epsilon})^2}} \stackrel{**}{\leq} 2n^{\epsilon/8} \sqrt{\sum_{i,j \in [n]: i \neq j} \frac{1}{(|\lambda_i - \lambda_j| + n^{-1-\epsilon})^2}}, \quad (12)\end{aligned}$$



where (\*) follows for  $n$  large enough by (9). In addition, (\*\*) follows with probability  $1 - ne^{-n^{\epsilon/8}/8} - ne^{-n^{1/4}}$ . Indeed, one can combine the union bound over  $i \in [n]$  with  $|\langle u_i, \mathbf{1} \rangle| = \frac{|\langle z, \mathbf{1} \rangle|}{\|z\|_2} \leq \frac{2|\langle z, \mathbf{1} \rangle|}{\sqrt{n}} \leq n^{\epsilon/16}$ , where  $z \sim N(0, I_n)$ . The second inequality holds with probability  $1 - e^{-n^{1/4}}$  for  $n$  large enough (e.g. see: [11, Lemma 15]). In addition, the last inequality holds with probability  $1 - e^{-n^{\epsilon/8}/8}$  (e.g. see: [11, Lemma 13]). Now, we get a handle on the eigenvalue separation in the following claim:

**Claim 10.** For  $n$  large enough, w.p.  $1 - o_n(1)$ , we have  $\sum_{i,j \in [n]: i \neq j} \frac{1}{(|\lambda_i - \lambda_j| + n^{-1-\epsilon})^2} \leq n^{3+3\epsilon/2}$ .

Note that naively bounding  $(|\lambda_i - \lambda_j| + n^{-1-\epsilon})^{-2} \leq n^{2+2\epsilon}$  results in a weaker upper bound of  $n^{4+2\epsilon}$ . We perform a much tighter analysis by carefully using the tail bounds for the separation of the eigenvalue from [29]. In particular, depending on whether  $|i - j|$  is small or large, we perform a separate analysis to lower bound  $|\lambda_i - \lambda_j|$ . The analysis is especially delicate when  $|i - j|$  is small: directly using the tail bounds of [29] combined with a union bound over  $i \in [n]$  is not sufficient.

We instead use the tail bounds to first compute  $\mathbb{E} \left[ \sum_{|i-j|=O(1)} \frac{1}{(|\lambda_i - \lambda_j| + n^{-1-\epsilon})^2} \right]$  and then use the Markov's inequality to get a tighter upper bound. Note that the trick to introduce  $n^{-1-\epsilon}$  in (11) is crucial here to ensure the finiteness of the expectation. The proof details of this claim are deferred to Appendix C.3. Now, we continue with the proof of Lemma 6 below.

Using the above claim along with (11) and (12), for  $n$  large enough, we get

$$\langle M, AX - XA \rangle \leq 2n^{3/2+7\epsilon/8} (\|AX - XA\|_F + n^{-0.5-\epsilon})$$

which further implies

$$\begin{aligned} \sum_{i \neq j} X_{ij} &\leq 2n^{3/2+7\epsilon/8} \|AX - XA\|_F + 2n^{1-\epsilon/8} + 2n^{1-\epsilon/32} \\ &\leq 2n^{3/2+7\epsilon/8} \|AX - XA\|_F + 4n^{1-\epsilon/32}, \end{aligned}$$

where the last inequality holds for  $n$  large enough (depending on  $\epsilon$ ). This completes the proof with probability  $1 - o_n(1)$  for  $n$  large enough by the union bound over all the bounds in the proof.  $\square$

## 5 Simulations

We conduct simulations on the Gaussian Wigner Model to verify our results and provide further insights (The code is publicly available at [https://github.com/smv30/convex\\_rel\\_for\\_graph\\_alignment](https://github.com/smv30/convex_rel_for_graph_alignment)). We set  $n = 400$  (unless otherwise specified) and consider  $\sigma \in \{0, 0.1, 0.2, \dots, 1\}$ . For these parameters, we test the performance of three convex relaxations: GRAMPA [11], simplex [2], and Birkhoff. We set the regularization parameter to 0.2 in GRAMPA as suggested by the authors. The convex relaxations are solved using the `cvxpy` library in Python using SCS (Splitting Conic Solver) and we set `use_indirect` to `True`, recommended for large instances for better memory management. We then project the solution of the convex relaxation to the set of permutation matrices using the Hungarian algorithm. For each  $n$  and  $\sigma$ , we repeat the simulation 10 times and report the average fraction of correctly matched vertices in Figure 1. We run the simulations on a CPU with 50GB of memory and impose a maximum run time of 3 hours for each instance.

On the left of Figure 1, we observe that the Birkhoff relaxation outperforms both Simplex and GRAMPA by exactly aligning all vertices for  $\sigma$  up to 0.5. On the other hand, the performance of Simplex and GRAMPA starts degrading for  $\sigma = 0.4$  and  $\sigma = 0.2$  respectively. Such a strong empirical performance motivates the theoretical analysis of the Birkhoff relaxation.

In the center plot of Figure 1, we plot the fraction of matched vertices and  $\|X^* - \Pi^*\|_F / \sqrt{n}$  as a function of  $\sigma$  for Birkhoff relaxation. In conjunction with Theorem 1 (case I),  $\|X^* - \Pi^*\|_F / \sqrt{n}$  increases rapidly and converges to one as a function of  $\sigma$ . However, even when  $\|X^* - \Pi^*\|_F / \sqrt{n}$  is close to one for  $\sigma \in [0.3, 0.5]$ , the Birkhoff relaxation still manages to align all the vertices. In particular, while the optimal solution  $X^*$  is not close to  $\Pi^*$ , it has a slight bias toward  $\Pi^*$  as opposed to other permutation matrices. So,  $X^*$  is projected to  $\Pi^*$  in the post-processing step of projection onto the set of permutation matrices. This result suggests that Birkhoff relaxation combined with post-processing could succeed beyond  $\sigma \sim n^{-0.5}$ .

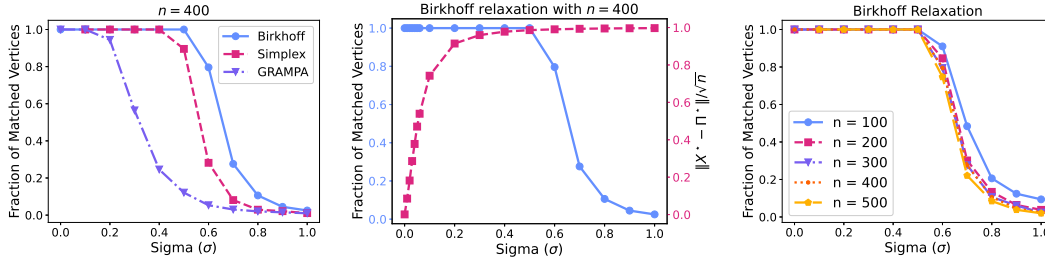


Figure 1: Fraction of Matched Vertices and  $\|X^* - \Pi^*\|_F / \sqrt{n}$  as a function of  $\sigma$  for the Gaussian Wigner Model: Performance of GRAMPA [11], Simplex [2], and Birkhoff Relaxations

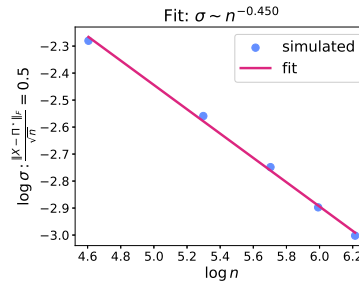


Figure 2: Log-log plot for  $\sigma$  such that  $\frac{\|X^* - \Pi^*\|_F}{\sqrt{n}} = 0.5$  as a function of  $n$ .

In the right plot of Figure 1, we test the performance of Birkhoff relaxation for  $n \in \{100, 200, 300, 400, 500\}$ . Although the fraction of correctly matched vertices as a function of  $\sigma$  worsens as  $n$  increases, the performance degradation is gradual. Such an empirical observation hints that the Birkhoff relaxation combined with post-processing could succeed for nearly constant  $\sigma$ . Although we take a first step in this direction, non-trivial ideas are needed for stronger guarantees.

In Figure 2, we empirically compute the value of  $\sigma$  at which  $\|X^* - \Pi^*\|_F$  transitions from  $o(\sqrt{n})$  to  $\Omega(\sqrt{n})$ . In particular, for each  $n \in \{100, \dots, 500\}$ , we implement the Birkhoff relaxation for  $\sigma \in \{0, 0.01, \dots, 1\}$  to infer the value of  $\sigma$  at which  $\|X^* - \Pi^*\|_F / \sqrt{n} = 0.5$ . A linear regression between these thresholds ( $\log \sigma$ ) and  $\log n$  outputs a slope of  $-0.45$ , supporting the phase transition at  $\sigma = \Theta(1/\sqrt{n})$  as in Theorem 1.

## 6 Conclusion and Future Work

In this paper, we study Birkhoff relaxation, a tight convex relaxation of the Quadratic Assignment Problem (QAP). The input is sampled from a Gaussian Wigner model with correlation  $1/\sqrt{1 + \sigma^2}$ . We show that  $X^*$  (optimal solution of Birkhoff relaxation) is a small perturbation of  $\Pi^*$  (optimal solution of QAP) when  $\sigma \ll n^{-1}$ , and  $X^*$  is far away from  $\Pi^*$  when  $\sigma \gg n^{-0.5}$ . More specifically, we show that  $\|X^* - \Pi^*\|_F^2 = o(n)$  when  $\sigma = o(n^{-1})$  and  $\|X^* - \Pi^*\|_F^2 = \Omega(n)$  when  $\sigma = \Omega(n^{-0.5})$ . This result allows us to align  $1 - o(1)$  fraction of vertices whenever  $\sigma = o(n^{-1})$ .

Based on heuristic calculations using the population version of the Birkhoff relaxation, we believe the condition to ensure  $\|X^* - \Pi^*\|_F^2 = o(n)$  can be improved to  $\sigma = o(n^{-0.5})$ , which is an immediate future work. In addition, while we show that  $\|X^* - \Pi^*\|_F^2 = \Omega(n)$  when  $\sigma = \Omega(n^{-0.5})$ , it does not imply the failure of Birkhoff relaxation, i.e., one can still post-process  $X^*$  appropriately to recover  $\Pi^*$ . Hence, another future direction is to establish the success of post-processing procedures combined with the Birkhoff relaxation for  $\sigma$  greater than  $n^{-0.5}$ .

## References

- [1] 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.
- [2] Ernesto Araya and Hemant Tyagi. Graph matching via convex relaxation to the simplex. *Foundations of Data Science*, pages 0–0, 2024.
- [3] G Ben Arous, Amir Dembo, and Alice Guionnet. Aging of spherical spin glasses. *Probability theory and related fields*, 120:1–67, 2001.
- [4] Tolga Birdal and Umut Simsekli. Probabilistic permutation synchronization using the riemannian structure of the birkhoff polytope. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11105–11116, 2019.
- [5] Béla Bollobás. Distinguishing vertices of random graphs. In *North-Holland Mathematics Studies*, volume 62, pages 33–49. Elsevier, 1982.
- [6] 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.
- [7] Tomek Czajka and Gopal Pandurangan. Improved random graph isomorphism. *Journal of Discrete Algorithms*, 6(1):85–92, 2008.
- [8] Osman Emre Dai, Daniel Cullina, Negar Kiyavash, and Matthias Grossglauser. Analysis of a canonical labeling algorithm for the alignment of correlated erdos-rényi graphs. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 3(2):1–25, 2019.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] Zhou Fan, Cheng Mao, Yihong Wu, and Jiaming Xu. Spectral graph matching and regularized quadratic relaxations II: Erdős-rényi graphs and universality. *Foundations of Computational Mathematics*, 23(5):1567–1617, 2023.
- [13] Soheil Feizi, Gerald Quon, Mariana Recamonde-Mendoza, Muriel Medard, Manolis Kellis, and Ali Jadbabaie. Spectral alignment of graphs. *IEEE Transactions on Network Science and Engineering*, 7(3):1182–1197, 2019.
- [14] Fajwel Fogel, Rodolphe Jenatton, Francis Bach, and Alexandre d’Aspremont. Convex relaxations for permutation problems. *Advances in neural information processing systems*, 26, 2013.
- [15] Luca Ganassali. Sharp threshold for alignment of graph databases with gaussian weights. In *Mathematical and Scientific Machine Learning*, pages 314–335. PMLR, 2022.
- [16] Luca Ganassali, Marc Lelarge, and Laurent Massoulié. Spectral alignment of correlated gaussian matrices. *Advances in Applied Probability*, 54(1):279–310, 2022.
- [17] 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.
- [18] Luca Ganassali and Laurent Massoulié. From tree matching to sparse graph alignment. In *Conference on Learning Theory*, pages 1633–1665. PMLR, 2020.
- [19] 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.

- [20] Zhao-Lin Jiang and Tin-Yau Tam. On the spectral norm of a doubly stochastic matrix and level-k circulant matrix. *Special Matrices*, 12(1):20230106, 2024.
- [21] Itay Kezurer, Shahar Z Kovalsky, Ronen Basri, and Yaron Lipman. Tight relaxation of quadratic matching. In *Computer graphics forum*, volume 34(5), pages 115–128. Wiley Online Library, 2015.
- [22] Shuyang Ling. On the exactness of SDP relaxation for quadratic assignment problem. arXiv:2408.05942, 2024.
- [23] 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.
- [24] 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.
- [25] Cheng Mao, Mark Rudelson, and Konstantin Tikhomirov. Random graph matching with improved noise robustness. In *Conference on Learning Theory*, pages 3296–3329. PMLR, 2021.
- [26] Cheng Mao, Mark Rudelson, and Konstantin Tikhomirov. Exact matching of random graphs with constant correlation. *Probability Theory and Related Fields*, 186(1):327–389, 2023.
- [27] 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.
- [28] 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.
- [29] Hoi Nguyen, Terence Tao, and Van Vu. Random matrices: tail bounds for gaps between eigenvalues. *Probability Theory and Related Fields*, 167:777–816, 2017.
- [30] 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.
- [31] Shinji Umeyama. An eigendecomposition approach to weighted graph matching problems. *IEEE transactions on pattern analysis and machine intelligence*, 10(5):695–703, 1988.
- [32] 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.

## NeurIPS Paper Checklist

### 1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [Yes]

Justification: Theorem 1 and Corollary 2 formalize the claims in the abstract and introduction. To the best of our knowledge, these results are the state-of-the-art for Graph Alignment via Birkhoff Relaxation.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

### 2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: In Section 2 (Main Result) and Section 6 (Conclusion and Future Work), we discuss that our theoretical results can be improved on and how they pave the way for further research on Birkhoff Relaxation.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best

judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

### 3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: The proof of Corollary 2 is provided in Section 2. In addition, Section 3 is dedicated to the proof of Theorem 1, while Section 4 and the Appendix include all the necessary proofs of Lemmas and Claims.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

### 4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: In Section 5 (Simulations), we mention the Python functions (including the arguments) used to solve the convex problems discussed in the paper. Values of all hyperparameters used are also indicated.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.

- (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
- (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
- (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

## 5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [\[Yes\]](#)

Justification: The code can be found on this link: [https://github.com/smv30/convex\\_rel\\_for\\_graph\\_alignment](https://github.com/smv30/convex_rel_for_graph_alignment).

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

## 6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [\[Yes\]](#)

Justification: In Section 5 (Simulations), we mention the Python functions (including the arguments) used to solve the convex problems discussed in the paper. Values of all hyperparameters used are also indicated.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.

- The full details can be provided either with the code, in appendix, or as supplemental material.

## 7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: For each set of parameters, we repeat the simulation 10 times and report the complete distribution of the objective in Figure 1.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

## 8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: We have included the CPU specifications and run time in the simulation section.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

## 9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines?>



Answer: [Yes]

Justification:

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

#### 10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: The paper provides theoretical guarantees for a well-known algorithm that has already been used in practice. A better theoretical understanding allows for more confidence while deploying such an algorithm in practice.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

#### 11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The results (simulations and otherwise) do not pose any such risk.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring

that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.

- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

## 12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [NA]

Justification: The paper does not use existing assets.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, [paperswithcode.com/datasets](https://paperswithcode.com/datasets) has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

## 13. New assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: The paper does not release any new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

## 14. Crowdsourcing and research with human subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

**15. Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

**16. Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The paper does not involve any LLMs.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.

## A Rounding Procedures

*Proof of Corollary 2.* We bound the number of incorrectly matched indices as follows:

$$\begin{aligned} \sum_{i=1}^n \mathbb{1}\{\hat{\pi}(i) \neq \pi^*(i)\} &\leq \sum_{i=1}^n \mathbb{1}\left\{X_{i\pi^*(i)}^* \leq \frac{1}{2}\right\} \leq 4 \sum_{i=1}^n \left(1 - X_{i\pi^*(i)}^*\right)^2 \\ &\leq 4\|X^* - \tilde{\Pi}^*\|_F^2 \stackrel{*}{=} 4\|X^* - \Pi^*\|_F^2 \leq 40n^{1-\epsilon/32}, \end{aligned}$$

where  $(*)$  holds as  $\tilde{\Pi} = \Pi^*$  with probability  $1 - o(1)$  by [15, 32]. The last inequality is followed by the second part of Theorem 1 with probability  $1 - o(1)$ .  $\square$

*Proof of Corollary 3.* Similar to the proof of Theorem 1, let  $\tilde{\Pi}^* = \Pi^* = I$  without loss of generality. Now, let  $\tilde{\Pi} \in \arg \max_{\Pi \in \mathcal{P}_n} \langle X^*, \Pi \rangle$ . Then, we have

$$\langle X^*, \tilde{\Pi} \rangle \geq \langle X^*, I \rangle \stackrel{(8)}{\geq} n - o(n).$$

Also, we have

$$\langle X^*, \tilde{\Pi} \rangle = \langle I, \tilde{\Pi} \rangle + \langle X^* - I, \tilde{\Pi} \rangle \leq \langle I, \tilde{\Pi} \rangle + \|X^* - I\|_F \|\tilde{\Pi}\|_F \leq \langle I, \tilde{\Pi} \rangle + o(n),$$

where the last inequality follows by Theorem 1 for  $\sigma \geq n^{-1-\epsilon}$ . Combining the above two inequalities, we get  $\langle I, \tilde{\Pi} \rangle \geq n - o(n)$  showing that the Hungarian solution  $\tilde{\Pi}$  overlaps with the correct permutation  $I$  for at least  $1 - o(1)$  fraction of the vertices, concluding the proof.  $\square$

## B Technical Details for Part I: Well-Separation

### B.1 Proof of Claim 4

We have

$$\begin{aligned} \|AX^* - X^*B\|_F^2 &\leq \frac{1}{n^2} \|AJ - JB\|_F^2 = \frac{1}{n^2} \|AJ - JA - \sigma JZ\|_F^2 \\ &\stackrel{*}{\leq} \frac{3}{n^2} (\|AJ\|_F^2 + \|JA\|_F^2 + \sigma^2 \|JZ\|_F^2) \\ &\stackrel{**}{\leq} \frac{6}{n} \sum_{i=1}^n \left( \sum_{k=1}^n A_{ik} \right)^2 + \frac{3}{n} \sum_{i=1}^n \left( \sum_{k=1}^n Z_{ik} \right)^2 \leq 9n^\epsilon, \end{aligned} \quad (13)$$

where  $(*)$  uses the inequality  $(a + b + c)^2 \leq 3(a^2 + b^2 + c^2)$ , and  $(**)$  holds as  $\sigma \leq 1$ . Moreover, the last inequality follows with probability at least  $1 - 4ne^{-n^\epsilon/4}$  for  $n$  large enough. Indeed,

$\sum_{k=1}^n A_{ik} \sim N(0, 1 + 1/n)$  and so  $\left| \sum_{k=1}^n A_{ik} \right| \leq n^{\epsilon/2}$  for all  $i \in [n]$  with probability  $1 - 2e^{-n^\epsilon/4}$

for  $n$  large enough. Thus, since  $A, Z$  are i.i.d., by the union bound,  $\sum_{i=1}^n (\sum_{k=1}^n A_{ik})^2 \leq n^{1+\epsilon}$  and  $\sum_{i=1}^n (\sum_{k=1}^n Z_{ik})^2 \leq n^{1+\epsilon}$  with probability at least  $1 - 4ne^{-n^\epsilon/4}$ .

### B.2 Proof of Claim 5

For any  $i, j \in [n]$ , we have

$$\begin{aligned} (AZ)_{ij} &= \sum_{k=1}^n A_{ik} Z_{kj} \stackrel{*}{\leq} \frac{2n^{\epsilon/2}}{\sqrt{n}} \|A_{i,:}\|_2 \quad \text{w.p. at least } 1 - e^{-n^\epsilon} \\ &\stackrel{**}{\leq} \frac{4n^{\epsilon/2}}{\sqrt{n}}, \end{aligned}$$

where  $(*)$  is true because  $\{Z_{kj} : k \in [n]\}$  are independent from each other, and independent from  $A$ . Thus,  $(*)$  follows by conditioning on  $A$  and using [11, Lemma 13]. Next,  $(**)$  holds with probability at least  $1 - e^{-\sqrt{n}}$  as  $\|A_{i,:}\|_2 \leq \frac{\sqrt{2}}{\sqrt{n}} \|z\|_2$ , where  $z \sim N(0, I_n)$  and by [11, Lemma 15],

$\|z\|_2^2 \leq n + \tilde{c}\sqrt{n} \leq 2n$  for sufficiently large  $n$ . Thus, by the union bound,  $(AZ)_{ij} \leq 4n^{\epsilon/2-1/2}$  with probability at least  $1 - 2e^{-n^\epsilon}$ . By further employing union bound on  $i, j \in [n]$ , with probability at least  $1 - 2n^2e^{-n^\epsilon}$ , we have

$$\max_{i \neq j} |(AZ - ZA)_{ij}| \leq 2 \max_{i \neq j} |(AZ)_{ij}| \leq \frac{8n^{\epsilon/2}}{\sqrt{n}}.$$

## C Technical Details for Part II: Small-Perturbation

### C.1 Proof of Lemma 7

We prove the desired inequality in the event where

$$\|Z\|_F^2 \leq \tilde{c}^2 n,$$

for some constant  $\tilde{c} > 0$ , which occurs with probability at least  $1 - e^{-n}$  if  $\tilde{c}$  is large enough, from [11, Lemma 15], because  $\|Z\|_F^2$  is a sum of  $\frac{n(n+1)}{2}$  independent squared Gaussian variables with variance  $\frac{2}{n}$ .

First, we show that  $\|AX^* - X^*B\|_F = O(\sigma\sqrt{n})$ . As  $X^*$  is the minimizer of (2) and  $I \in \mathcal{B}_n$ ,

$$\|AX^* - X^*B\|_F^2 \leq \|A - B\|_F^2 = \sigma^2 \|Z\|_F^2 \leq \tilde{c}^2 \sigma^2 n.$$

We now obtain an upper bound on  $\|AX^* - X^*A\|_F$  using the upper bound on  $\|AX^* - X^*B\|_F$ :

$$\begin{aligned} \|AX^* - X^*A\|_F &\leq \|AX^* - X^*B\|_F + \|X^*(A - B)\|_F \\ &= \|AX^* - X^*B\|_F + \sigma \|X^*Z\|_F \\ &\leq \tilde{c}\sigma\sqrt{n} + \sigma \|X^*\|_2 \|Z\|_F \\ &\leq \tilde{c}\sigma\sqrt{n} + \tilde{c}\sigma\sqrt{n} \|X^*\|_2 \\ &\leq 2\tilde{c}\sigma\sqrt{n}. \end{aligned}$$

The last inequality follows as  $X^* \in \mathcal{B}_n$  and the spectral norm of a doubly stochastic matrix is at most 1, e.g., see [20]. It completes the proof with  $c = 2\tilde{c}$ .

### C.2 Proof of Claim 9

Let  $U \in \mathbb{R}^{n \times n}$  be the matrix whose lines are  $u_1, \dots, u_n$ . It is orthogonal, and its distribution is uniform on the set of orthogonal matrices. Therefore,  $U\mathbf{1}$  is uniformly distributed in  $\sqrt{n}\mathcal{S}_{n-1}$ , so that it is equal in distribution to  $\sqrt{n} \frac{z}{\|z\|_2}$ , for  $z \sim N(0, I_n)$ . In particular,

$$\begin{aligned} &\mathbb{P}\left(\#\left\{|\langle u_i, \mathbf{1} \rangle| \leq n^{-\epsilon/16}\right\} > 3n^{1-\epsilon/16}\right) \\ &= \mathbb{P}\left(\#\left\{|z_i| \leq n^{-1/2-\epsilon/16}\|z\|_2\right\} > 3n^{1-\epsilon/16}\right) \\ &\leq \mathbb{P}\left(\#\left\{|z_i| \leq 2n^{-\epsilon/16}\right\} > 3n^{1-\epsilon/16}\right) + \mathbb{P}(\|z\|_2 > 2\sqrt{n}) \\ &\stackrel{*}{\leq} \mathbb{P}\left(\sum_{i=1}^n \mathbb{1}\left\{|z_i| \leq 2n^{-\epsilon/16}\right\} > 3n^{1-\epsilon/16}\right) + e^{-c_1 n} \quad (\text{for some constant } c_1 > 0) \\ &\stackrel{**}{\leq} \mathbb{P}\left(\sum_{i=1}^n \left(\mathbb{1}\left\{|z_i| \leq 2n^{-\epsilon/16}\right\} - \mathbb{E}\mathbb{1}\left\{|z_i| \leq 2n^{-\epsilon/16}\right\}\right) > n^{1-\epsilon/16}\right) + e^{-c_1 n} \\ &\stackrel{***}{\leq} e^{-c_2 n^{1-\epsilon/16}} + e^{-c_1 n} \quad (\text{for some constant } c_2 > 0). \end{aligned}$$

Inequality (\*) is true, again, from [11, Lemma 15]. Inequality (\*\*) is true because

$$\mathbb{E}\left(\sum_{i=1}^n \mathbb{1}\left\{|z_i| \leq 2n^{-\epsilon/16}\right\}\right) = \frac{n}{\sqrt{2\pi}} \int_{-2n^{-\epsilon/16}}^{2n^{-\epsilon/16}} e^{-\frac{t^2}{2}} dt \leq 2n\sqrt{\frac{2}{\pi}} n^{-\epsilon/16} \leq 2n^{1-\epsilon/16}.$$

Inequality (\*\*\*) is due to Bernstein's inequality. This completes the proof of claim with  $c = c_2/2$  as  $e^{-c_2 n^{1-\epsilon/16}} + e^{-c_1 n} \leq e^{-c_2 n^{1-\epsilon/16}/2}$  for  $n$  large enough.

### C.3 Proof of Claim 10

We pick  $L > 0$  (constant depending on  $\epsilon$ ) such that  $\min_{|i-j|>L} |\lambda_i - \lambda_j| \geq n^{-1-\epsilon/2}$  with probability at least  $1 - o(1)$  by [29, Corollary 2.5]. So we have

$$\begin{aligned} \sum_{i \neq j} \frac{1}{(|\lambda_i - \lambda_j| + n^{-1-\epsilon})^2} &= \sum_{k=0}^{\lceil \frac{n}{L} \rceil} \sum_{|i-j| \in [kL, (k+1)L]} \frac{1}{(|\lambda_i - \lambda_j| + n^{-1-\epsilon})^2} \\ &\leq \sum_{|i-j| \leq L} \frac{1}{(|\lambda_i - \lambda_j| + n^{-1-\epsilon})^2} + \sum_{k=1}^{\lceil \frac{n}{L} \rceil} \sum_{|i-j| \in [kL, (k+1)L]} \frac{1}{(|\lambda_i - \lambda_j|)^2} \\ &\leq \sum_{|i-j| \leq L} \frac{1}{(|\lambda_i - \lambda_j| + n^{-1-\epsilon})^2} + \sum_{k=1}^{\lceil \frac{n}{L} \rceil} \sum_{|i-j| \in [kL, (k+1)L]} \frac{n^{2+\epsilon}}{k^2} \\ &\leq \sum_{|i-j| \leq L} \frac{1}{(|\lambda_i - \lambda_j| + n^{-1-\epsilon})^2} + n^{3+\epsilon} L \sum_{k=1}^{\lceil \frac{n}{L} \rceil} \frac{1}{k^2} \\ &\leq \sum_{|i-j| \leq L} \frac{1}{(|\lambda_i - \lambda_j| + n^{-1-\epsilon})^2} + 10Ln^{3+\epsilon}. \end{aligned}$$

Now, we bound the remaining terms below. We have

$$\begin{aligned} &\mathbb{E} \left[ \frac{1}{(|\lambda_j - \lambda_i| + n^{-1-\epsilon})^2} \right] \\ &= \int_0^\infty \mathbb{P} \left( \frac{1}{(|\lambda_j - \lambda_i| + n^{-1-\epsilon})^2} > x \right) dx \\ &= \int_0^\infty \mathbb{P} \left( |\lambda_j - \lambda_i| < \frac{1}{\sqrt{x}} - n^{-1-\epsilon} \right) dx \\ &= \int_0^{n^{2+2\epsilon}} \mathbb{P} \left( |\lambda_j - \lambda_i| < \frac{1}{\sqrt{x}} - n^{-1-\epsilon} \right) dx + \int_{n^{2+2\epsilon}}^\infty \mathbb{P} \left( |\lambda_j - \lambda_i| < \frac{1}{\sqrt{x}} - n^{-1-\epsilon} \right) dx \\ &\leq \int_0^{n^{2+2\epsilon}} \mathbb{P} \left( |\lambda_j - \lambda_i| < \frac{1}{\sqrt{x}} \right) dx \leq 2c_0 n^{2+\epsilon}, \end{aligned}$$

where the last inequality holds by [29, Corollary 2.2] for some constant  $c_0 > 0$ . Using the above inequality, we get

$$\mathbb{E} \left[ \sum_{|i-j| \leq L} \frac{1}{(|\lambda_j - \lambda_i| + n^{-1-\epsilon})^2} \right] \leq 2c_0 Ln^{3+\epsilon}.$$

Now, by the Markov's inequality, with probability  $1 - n^{-\epsilon/4}$ , we get

$$\sum_{|i-j| \leq L} \frac{1}{(|\lambda_j - \lambda_i| + n^{-1-\epsilon})^2} \leq 2c_0 Ln^{3+5\epsilon/4}.$$

Thus, we get

$$\sum_{i \neq j} \frac{1}{(|\lambda_i - \lambda_j| + n^{-1-\epsilon})^2} \leq 2c_0 Ln^{3+5\epsilon/4} + 10Ln^{3+\epsilon} \leq n^{3+3\epsilon/2},$$

where the last inequality holds for  $n$  large enough (depending on  $\epsilon$ ). The proof is now complete.