

# Ranking and synchronization from pairwise measurements via SVD

Alexandre d’Aspremont\*, Mihai Cucuringu<sup>†‡</sup>, Hemant Tyagi<sup>‡</sup>

June 20, 2019

## Abstract

Given a measurement graph  $G = (V, E)$  and an unknown signal  $r \in \mathbb{R}^n$ , we investigate algorithms for recovering  $r$  from pairwise measurements of the form  $r_i - r_j$ ;  $\{i, j\} \in E$ . This problem arises in a variety of applications, such as ranking teams in sports data and time synchronization of distributed networks. Framed in the context of ranking, the task is to recover the ranking of  $n$  teams (induced by  $r$ ) given a small subset of noisy pairwise rank offsets. We propose a simple SVD-based algorithmic pipeline for both the problem of time synchronization and ranking. We provide a detailed theoretical analysis in terms of robustness against both sampling sparsity and noise perturbations with outliers, using results from matrix perturbation and random matrix theory. Our theoretical findings are complemented by a detailed set of numerical experiments on both synthetic and real data, showcasing the competitiveness of our proposed algorithms with other state-of-the-art methods.

**Keywords:** ranking, group synchronization, spectral algorithms, matrix perturbation theory, singular value decomposition, low-rank matrix completion.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Related work</b>	<b>3</b>
2.1	Synchronization literature . . . . .	4
2.2	Ranking literature . . . . .	5
<b>3</b>	<b>Problem setup and main results</b>	<b>7</b>
3.1	Main idea: SVD-based spectral algorithm . . . . .	7
3.2	Summary of main theoretical results . . . . .	8
<b>4</b>	<b>Theoretical results for SVD-RS and SVD-NRS</b>	<b>9</b>
4.1	Analysis of Algorithm 1 (SVD-RS) . . . . .	10
4.2	Analysis of Algorithm 2 (SVD-NRS) . . . . .	13
<b>5</b>	<b>Proofs</b>	<b>14</b>
5.1	Proof of Theorem 1 . . . . .	14
5.2	Proof of Theorem 2 . . . . .	16
5.3	Proof of Theorem 3 . . . . .	17
5.4	Proof of Theorem 4 . . . . .	18
5.5	Proof of Theorem 5 . . . . .	19
5.6	Proof of Theorem 6 . . . . .	20
<b>6</b>	<b>Matrix completion as a preprocessing step for ranking</b>	<b>22</b>

\*CNRS & Ecole Normale Supérieure, Paris, France. E-mail address: aspremon@ens.fr

<sup>†</sup>Department of Statistics and Mathematical Institute, University of Oxford, Oxford, UK. Email: mi-hai.cucuringu@stats.ox.ac.uk

<sup>‡</sup>INRIA Lille-Nord Europe, Lille, France; MODAL project-team. Email: hemant.tyagi@inria.fr

<sup>§</sup>Alan Turing Institute, London, UK. This work was supported by EPSRC grant EP/N510129/1.

<b>7</b>	<b>Numerical Experiments</b>	<b>23</b>
7.1	Synthetic data . . . . .	23
7.2	Real data . . . . .	24
<b>8</b>	<b>Conclusion and future directions</b>	<b>28</b>
<b>A</b>	<b>Matrix perturbation analysis</b>	<b>32</b>
<b>B</b>	<b>Useful concentration inequalities</b>	<b>33</b>
B.1	Spectral norm of random symmetric matrices . . . . .	33
B.2	Bernstein inequality . . . . .	33
B.3	Product of a random matrix (raised to a power) and a fixed vector . . . . .	33
<b>C</b>	<b>Proof of Theorem 1</b>	<b>33</b>
C.1	Proof of Lemma 1 . . . . .	33
C.2	Proof of Lemma 2 . . . . .	34
C.3	Proof of Lemma 3 . . . . .	34
C.4	Proof of Lemma 4 . . . . .	35
<b>D</b>	<b>Proof of Lemmas from Section 5.2</b>	<b>36</b>
D.1	Proof of Lemma 5 . . . . .	36
D.2	Proof of Lemma 6 . . . . .	38
D.3	Proof of Lemma 7 . . . . .	40
D.4	Proof of Lemma 8 . . . . .	41
<b>E</b>	<b>Proofs of Lemmas from Section 5.5</b>	<b>41</b>
E.1	Proof of Lemma 9 . . . . .	41

# 1 Introduction

Let  $r = (r_1, \dots, r_n)^T \in \mathbb{R}^n$  be an unknown signal and  $G = ([n], E)$  be an undirected measurement graph. Given a subset of noisy pairwise measurements of the form  $r_i - r_j$  for each  $\{i, j\} \in E$ , the goal is to estimate the original vector  $r$ . Clearly, this is only possible only up to a global shift. Moreover, when measurements are exact without any measurement noise, one can recover the strength vector  $r$  if and only if the graph  $G$  is connected, by simply considering a spanning tree, fixing the value of the root node, and traversing the tree while propagating the information by summing the given offsets. For simplicity, we assume the graph is connected, otherwise it is not possible to estimate the offset values between nodes belonging to different connected components of the graph.

Instantiations of the above problem are ubiquitous in engineering, machine learning and computer vision, and have received a great deal of attention in the recent literature. There are two main classes of applications where this problem arises.

- *Time synchronization of wireless networks.* A popular application arises in engineering, and is known as time synchronization of distributed networks [26, 33], where clocks measure noisy time offsets  $r_i - r_j$ , and the goal is to recover  $r_1, \dots, r_n \in \mathbb{R}$ .
- *Ranking.* A fundamental problem in information retrieval is that of recovering the ordering induced by the latent strengths or scores  $r_1, \dots, r_n \in \mathbb{R}$  of a set of  $n$  players, that is best reflected by the given set of pairwise comparisons  $r_i - r_j$ . We refer the reader to [16] and references therein for a detailed overview.

A naive approach by sequentially propagating the information along spanning trees is doomed to fail in the presence of noise, due to accumulation of the errors. To this end, in order to increase the robustness to noise, one typically aims to simultaneously integrate all the pairwise measurements in a globally consistent framework. This line of thought appears in the literature in the context of the *group synchronization* problem [45], for recovering group elements from noisy pairwise measurements of their ratios. We briefly review the

current literature in group synchronization in Section 2.1, which almost exclusively pertains to synchronization over compact groups. The problem we study in this paper can be construed as synchronization over the real line, hence a non-compact group. There exists a very rich literature on ranking, and it is beyond the scope of our work to provide an extensive review of it. Instead, in Section 2.2, we give a brief overview of the literature on ranking and time synchronization, and point out some of the methods that relate to our work. In particular, we focus on spectral methods and approaches that leverage low-rank matrix completion as a pre-processing step.

**Contributions.** We propose SVD-RS, a simple spectral algorithm for ranking and synchronization from pairwise comparisons, along with a normalized version denoted SVD-NRS<sup>1</sup>. We provide a detailed theoretical consistency analysis for both algorithms for a random measurement model (see Section 3) in terms of robustness against sampling sparsity of the measurement graph and noise level. Additionally, we provide extensive numerical experiments on both synthetic and real data, showing that in certain noise and sparsity regimes, our proposed algorithms perform comparable or better than state-of-the-art methods.

On the theoretical side, our specific contributions can be summarized as follows.

- For SVD-RS, we provide  $\ell_2$  and  $\ell_\infty$  recovery guarantees for the **score** vector  $r$  (see Theorem 4). For instance, in the setting when  $r_i = i$ ,  $\Omega(n \log n)$  measurements suffice for  $\ell_2$  recovery, while  $\Omega(n^{4/3}(\log n)^{2/3})$  measurements suffice for  $\ell_\infty$  recovery.
- The  $\ell_\infty$  analysis of SVD-RS leads to guarantees for **rank** recovery in terms of the maximum displacement error between the recovered ranking and the ground truth, as in Theorem 3. For e.g., when  $r_i = i$ ,  $\Omega(n^{4/3}(\log n)^{2/3})$  measurements suffice.
- For SVD-NRS, we provide in Theorem 6,  $\ell_2$  recovery guarantees for the **score** vector  $r$ , and leave the  $\ell_\infty$  guarantees for future work, essentially by following a similar, though more intricate, pipeline. Similar to SVD-RS, for  $r_i = i$ ,  $\Omega(n \log n)$  measurements suffice for  $\ell_2$  recovery.

We remark that all the above recovery results hold with high probability.

**Outline.** The remainder of this paper is organized as follows. Section 2 is a brief survey of the relevant literature, with a focus on group synchronization and ranking. Section 3 starts with the formal setup of the problem, presents the gist of our SVD-based approach along with the two algorithms SVD-RS, SVD-NRS, and summarizes our main theoretical results. Section 4 details and interprets our theoretical results, with the main steps of the proofs outlined in Section 5. Section 6 discusses the low-rank matrix completion problem and its applicability in the setting of this paper. Section 7 contains numerical experiments on various synthetic and real data sets. Finally, Section 8 summarizes our findings along with future research directions. The Appendix contains additional technical details for the theoretical results, as well as further numerical experiments.

**Notation.** Vectors and matrices are denoted by lower case and upper case letters respectively. For a matrix  $A \in \mathbb{R}^{m \times n}$  where  $m \geq n$ , we denote its singular values by  $\sigma_1 \geq \dots \geq \sigma_n$  and the corresponding left (resp. right) singular vectors by  $u_i$  (resp.  $v_i$ ).  $\|A\|_2$  denotes the spectral norm (largest singular value),  $\|A\|_*$  denotes the nuclear norm (sum of the singular values), and  $\|A\|_{\max} := \max_{i,j} |A_{ij}|$  denotes the max-norm of  $A$ .  $\mathcal{R}(A)$  denotes the range space of its columns. We denote  $e$  to be the all ones column vector. The symbol  $\mathcal{U}$  is used to denote the uniform distribution. For positive numbers  $a, b$ , we denote  $a \lesssim b$  (resp.  $a \gtrsim b$ ) to mean that there exists an absolute constant  $C > 0$  such that  $a \leq Cb$  (resp.  $a \geq Cb$ ).

## 2 Related work

This section is a brief survey of the relevant literature for the tasks we consider out in this paper. The first part focuses on the group synchronization problem, a relevant instance of it being that of synchronization of clocks arising often in engineering, while the second part surveys the very rich ranking literature, with an emphasis on spectral methods and approaches that leverage low-rank matrix completion.

<sup>1</sup>By normalized, we mean that the measurement matrix is normalized by the degree matrix of the graph, see Section 4.

## 2.1 Synchronization literature

Much of the engineering literature has focused on least-squares approaches for solving the time synchronization problem. In its simplest terms, the approach can be formulated as follows. Let  $m = |E|$  denote the number of edges in  $G$ , and  $B$  denote the edge-vertex incidence matrix of size  $m \times n$ . For  $l = \{i, j\} \in E$ , with  $i < j$ , the entries  $B_{li}$  and  $B_{lj}$  are given by  $B_{li} = 1$ ,  $B_{lj} = -1$ . If  $i' \notin \{i, j\}$  then  $B_{li'} = 0$ . Let  $w \in \mathbb{R}^m$  encode the pairwise rank measurements  $C_{ij}$ , for all edges  $\{i, j\} \in E$ . The least-squares solution to the ranking problem can be obtained by solving

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \|Bx - w\|_2^2, \tag{2.1}$$

where  $B$  denotes the design matrix with two non-zero entries per row corresponding to the edge  $\{i, j\}$  indexed by  $l$ . In a related line of work, Hirani et al. [28] show that the problem of least-squares ranking on graphs has far-reaching rich connections with seemingly unrelated areas, such as spectral graph theory and multilevel methods for graph Laplacian systems, Hodge decomposition theory and random clique complexes in topology.

The problem we consider in this paper can also be framed in the context of the *group synchronization* problem, of finding group elements from noisy measurements of their ratios. For example, synchronization over the special orthogonal group  $SO(d)$  consists of estimating a set of  $n$  unknown  $d \times d$  matrices  $R_1, \dots, R_n \in SO(d)$  from noisy measurements of a subset of the pairwise ratios  $R_i R_j^{-1}$  via

$$\underset{R_1, \dots, R_n \in SO(d)}{\text{minimize}} \quad \sum_{\{i, j\} \in E} w_{ij} \|R_i^{-1} R_j - R_{ij}\|_F^2, \tag{2.2}$$

where  $\|\cdot\|_F$  denotes the Frobenius norm, and  $w_{ij}$  are non-negative weights representing the confidence in the noisy pairwise measurements  $R_{ij}$ . Spectral and semidefinite programming (SDP) relaxations for solving an instance of the above synchronization problem were originally introduced and analyzed by Singer [45], in the context of angular synchronization over the group  $SO(2)$  of planar rotations. There, one is asked to estimate  $n$  unknown angles  $\theta_1, \dots, \theta_n \in [0, 2\pi)$  given  $m$  noisy measurements of their offsets  $\theta_i - \theta_j \pmod{2\pi}$ . The difficulty of the problem is amplified, on one hand, by the amount of **noise** in the offset measurements, and on the other hand by **sparsity** - the fact that  $m \ll \binom{n}{2}$ , i.e., only a very small subset of all possible pairwise offsets are measured. In general, one may consider other groups  $\mathcal{G}$  (such as  $SO(d)$ ,  $O(d)$ ) for which there are available noisy measurements  $g_{ij}$  of ratios between the group elements

$$g_{ij} = g_i g_j^{-1}, g_i, g_j \in \mathcal{G}. \tag{2.3}$$

The set  $E$  of pairs  $\{i, j\}$  for which a ratio of group elements is available can be realized as the edge set of a graph  $G = (V, E)$ ,  $|V| = n$ ,  $|E| = m$ , with vertices corresponding to the group elements  $g_1, \dots, g_n$ , and edges to the available pairwise measurements  $g_{ij} = g_i g_j^{-1}$ . As long as the group  $\mathcal{G}$  is compact and has a real or complex representation, one may construct a real or Hermitian matrix (which may also be a matrix of matrices) where the element in the position  $\{i, j\}$  is the matrix representation of the measurement  $g_{ij}$  (possibly a matrix of size  $1 \times 1$ , as is the case for  $\mathbb{Z}_2$ ), or the zero matrix if there is no direct measurement for the ratio of  $g_i$  and  $g_j$ . For example, the rotation group  $SO(3)$  has a real representation using  $3 \times 3$  rotation matrices, and the group  $SO(2)$  of planar rotations has a complex representation as points on the unit circle.

The setting we consider in this paper, namely that of recovering points on the real line from a subset of noisy pairwise differences, is essentially synchronization over the non-compact group  $\mathbb{R}$ . In recent work in the context of ranking from pairwise cardinal and ordinal measurements [16], the real line was compactified by wrapping it over the upper half of the unit circle, making the problem amenable to standard synchronization over the compact group  $SO(2)$ . The estimated solution allowed for the recovery of the player rankings, after a post-processing step of modding out the best circular permutation. Note that the proposed approach only focused on recovering the individual rankings, and not the magnitude (i.e., strength) of each player, as we propose to do in this paper.

Very recently, Ozyesil et al. [42] proposed an approach that allows one to consider the synchronization problem in the setting of non-compact groups. The authors leverage a compactification process, that relies on a mapping from a non-compact domain into a compact one, for solving the synchronization problem. The contraction mapping enables one to transform measurements from a Cartan motion group to an associated compact group, whose synchronization solution provides a solution for the initial synchronization problem over the original domain.

## 2.2 Ranking literature

Fogel et al. introduced SERIAL-RANK [24, 25], a ranking algorithm that explicitly leverages the connection with seriation, a classical ordering problem that considers the setting where the user has available a similarity matrix (with  $\pm 1$  entries) between a set of  $n$  items, and assumes that there exists an underlying one-dimensional ordering such that the similarity between items decreases with their distance. The authors propose a spectral algorithm SERIAL-RANK, and show that under an Erdős Renyi random graph model and a given noise model, it recovers the underlying ranking with  $\Omega(n^{3/2} \log^4 n)$  comparisons where the estimation error is bounded in the  $\ell_\infty$  norm.

A popular theme in the literature is the rank aggregation setting (where players meet in multiple matches), where there exists a latent probability matrix  $P \in [0, 1]^{n \times n}$ , wherein  $P_{ij}$  denotes the probability that player  $i$  defeats player  $j$ , with  $P_{ij} + P_{ji} = 1$ . For each pair  $(i, j)$ , one observes  $Y_{ij} \in \{0, 1\} \sim \text{Bern}(P_{ij})$ . A number of so-called Random Utility models have been considered in the literature, starting with the seminal work of Bradley-Terry-Luce (BTL) [7], which is by far the most popular model considered. In the most basic version of the BTL model, the probability that player  $i$  beats player  $j$  is given by  $P_{ij} = \frac{w_i}{w_i + w_j}$ , where the vector  $w \in \mathbb{R}_+^n$  is the parameter vector to be inferred from the data, with  $w_i$  being a proxy for the score or strength associated to player  $i$ . In the Thurstone model [48],  $P_{ij} = \Phi(s_i - s_j)$ , where  $\Phi$  denotes the standard normal cumulative distribution function (CDF), and  $s \in \mathbb{R}^n$  is the score vector. Negahban et al. proposed RANK-CENTRALITY in the context of the rank aggregation problem from multiple ranking systems, under a Bradley-Terry-Luce (BTL) model [39, 40]. Rank-Centrality is an iterative algorithm that estimates the ranking scores from the stationary distribution of a certain random walk on the graph of players. Edges encode the outcome of pairwise comparisons, and are directed  $i \rightarrow j$  towards the winner, where the weight captures the proportion of times  $j$  defeated  $i$  in case of multiple direct matches. The authors show that under some assumptions on the connectivity of the underlying graph, their algorithm estimates the underlying score vector of the BTL model with  $\Omega(n \text{poly}(\log n))$  comparisons. Very recently, Agarwal et al. [1] propose Accelerated Spectral Ranking, a provably faster spectral ranking algorithm in the setting of the multinomial logit (MNL) and BTL models. The authors considered a random walk that has a faster mixing time than the random walks associated with previous algorithms (including Rank-Centrality [40]), along with improved sample complexity bounds (of the order  $n \text{poly}(\log n)$ ) for recovery of the MNL and BTL parameters.

Cucuringu [16] introduced SYNC-RANK, formulating the problem of ranking with incomplete noisy information as an instance of the group synchronization problem over the group  $\text{SO}(2)$  of planar rotations. SYNC-RANK starts by making the ansatz that items are embedded on the upper-half of the unit circle in the complex plane, and transforms the pairwise input measurements into pairwise angle offsets. It then solves the spectral or SDP relaxation of synchronization, and finally chooses amongst the circular permutations of the ranking induced by the synchronized solution by minimizing the number of upsets. The method was shown to compare favorably with a number of state-of-the-art methods, including RANK-CENTRALITY [39], SERIAL-RANK [25], and a SVD-based ranking algorithm similar to the one we consider in this paper, but without the projection step and its normalized extension, and without any theoretical guarantees (see Remark 2 for details). Along the same lines, and inspired by the group synchronization framework, Fanuël and Suykens [23] propose a certain deformation of the combinatorial Laplacian, in particular, the so-called dilation Laplacian whose spectrum is leveraged for ranking in directed networks of pairwise comparisons. The method is reported to perform well on both synthetic and real data, and enjoys the property that it can place an emphasis of the top- $k$  items in the ranking.

More recently, De Bacco et al. [20] proposed SPRING-RANK, an efficient physically-inspired algorithm, based on solving a linear system of equations, for the task of inferring hierarchical rankings in directed networks, which also comes with a statistical significance test for the inferred hierarchy. The model considered also incorporates the assumption that interactions are more likely to occur between individuals with similar ranks, which is sometimes the case in real applications, such as sport competitions. The authors compare SPRING-RANK to a wealth of algorithms, including the above-mentioned SYNC-RANK and SERIAL-RANK, in terms of an accuracy defined as the fraction of edges whose direction is consistent with the inferred ranking, and conclude that, on average across various synthetic and real data sets, SPRING-RANK and SYNC-RANK have the highest accuracy.

Volkovs and Zemel [51] consider two instances (in the unsupervised and supervised settings) of the

preference aggregation problem where the task is to combine multiple preferences (either in the form of binary comparisons or as score differences) over objects into a single consensus ranking. Both settings rely on a newly introduced Multinomial Preference model (MPM) which uses a multinomial generative process to model the observed preferences. Dalal et al. [18] consider a multi-objective rank aggregation problem (where pairwise information is in the form of score differences), and rely on the framework of the combinatorial Hodge decomposition [30]. In particular, they first formulate the problem of reducing global inconsistencies and then propose techniques for identifying local observations which can reduce such global inconsistencies.

**Low-rank matrix completion approaches for ranking.** Gleich and Lim [27] proposed a new method for ranking a set of items from pairwise observations in the form  $r_i - r_j$ . They first use matrix completion to fill in a partial skew-symmetric matrix, followed by a simple row-sum to recover the rankings. Using standard recovery results from matrix completion literature, it follows that in the absence of noise, and under some coherence assumptions on the score vector  $r$ , the true ranking can be recovered exactly from a random subset of pairs of size  $\Omega(n \text{poly}(\log n))$ . In case the observations are given as a rating matrix (users and items), they show how to transform it into a pairwise comparison matrix involving score differences so that their method can be applied. Ye et al. [55] proposed a rank-minimization approach to aggregate the predicted confidence scores of multiple models. The authors cast the score fusion problem as that of finding a shared rank-2 pairwise relationship matrix that renders each of the original score matrices from the different models to be decomposed into the common rank-2 matrix and the sparse error component.

The recent work of Levy et al [36] relies on matrix completion in the context of the rank aggregation problem from a small subset of noisy pairwise comparisons, wherein the user observes  $L$  repeated comparisons between the same set of players. The authors work in the setting of the Bradley-Terry-Luce model (BTL) that assumes that a set of latent scores (or strengths) underlies all players, and each individual pairwise comparison is a probabilistic outcome, as a function of the underlying scores. The resulting pipeline is reported to improve over state-of-the-art in both simulated scenarios and real data. Prior to that, Kang et al. [32] also relied on matrix completion and proposed an algorithm for top- $N$  recommender systems. The authors filled in the user-item matrix based on a low-rank assumption, by considering a nonconvex relaxation, as opposed to the usual nuclear norm, and argued that it provides a better rank approximation and empirically leads to an accuracy superior to that of any state-of-the-art algorithm for the top- $N$  recommender problem, on a comprehensive set of experiments on real data sets.

Another relevant line of work is that of Massimino and Davenport [38], who consider an adaptation of the one-bit matrix completion framework [19] to the setting of pairwise comparison matrices. One observes measurements  $Y_{ij} = \pm 1$  with probability  $f(M_{ij})$ , where  $M_{ij} = r_i - r_j$ , and  $f(M_{ij}) = \mathbb{P}(M_{ij} > 0)$ ; for example, the authors consider  $f(x) = (1 + e^{-x})^{-1}$ . On a related note, Yang and Wakin [54] considered the rank aggregation setting, extending the work of Gleich and Lim [27] to the setup of non-transitive matrices, for which it does not necessarily hold true that  $Y_{i,j} = Y_{i,k} + Y_{k,j}, \forall i, j, k$ , as is the case in the special setting when  $Y_{i,j} = r_i - r_j$  (and thus  $Y = re^T - er^T$ ), for a score vector  $r$ . In particular, Yang and Wakin were interested in modeling and recovering  $Y$  itself as opposed to the one-dimensional ranking, and introduced a model for non-transitive pairwise comparisons  $Y_{ij} = r_i a_j - r_j a_i$ , where, for example,  $r_i$  could denote offensive strength of player  $i$ , and  $a_j$  defensive strength of player  $j$ , thus giving  $Y_{ij}$  the interpretation of the anticipated margin of victory for player  $i$  over player  $j$ . The authors then propose a low-rank matrix completion approach based on alternating minimization, along with recovery guarantees for the estimate  $\hat{Y}$  of the form  $\|\hat{Y} - Y\|_F \leq \epsilon$ , after  $\log 1/\epsilon$  iterations.

The recent seminal work of Rajkumar and Agarwal [44] considered the question of recovering rankings from  $O(n \log n)$  comparisons between the items. Their proposed pipeline starts by mapping the input binary entries of the comparison matrix to a low-rank matrix. More specifically, a *link function* is applied entry-wise to the input comparison matrix, which renders the resulting incomplete matrix to be of low rank. Standard low-rank matrix completion is applied to complete the matrix, followed by the inverse link function applied entry wise, which effectively maps the data back to binary pairwise measurements. Finally, the Copeland ranking procedure [15] is applied that ranks items by their Copeland scores, which is effectively given by the number of wins in the comparison matrix after thresholding the entries with respect to 0.5. This approach can be seen as a generalization of the pipeline proposed by Gleich and Lim [27], in the sense that both lines of work rely on matrix completion followed by row-sum, the main difference being that [44] uses the link function that maps the binary input matrix to a low-rank one. Borrowing tools from matrix completion,

restricted strong convexity and prior work on Random Utility Models, Negahban et al. [41] propose a nuclear norm regularized optimization problem for learning the parameters of the MultiNomial Logit (MNL) model that best explains the data, typically given as user preference in the form of choices and comparisons. The authors show that the convex relaxation for learning the MNL model is minimax optimal up to a logarithmic factor, by comparing its performance to a fundamental lower bound.

### 3 Problem setup and main results

Our formal setup is as follows. Consider an undirected graph  $G = ([n], E)$  and an unknown vector  $r \in \mathbb{R}^n$ , where  $r_i$  is the value associated with node  $i$ . In particular,  $G$  is assumed to be a  $G(n, p)$ , i.e., the popular Erdős-Rényi random graph model, where edges between vertices are present independently with probability  $p$ . Moreover, we assume  $r_i$  to be bounded uniformly, i.e.,  $r_i \in [0, M]$  for each  $i$ . Hence  $r_i - r_j \in [-M, M]$  for all  $i, j$ . Requiring  $r_i \geq 0$  is only for convenience and w.l.o.g. Importantly,  $M$  is *not assumed* to be known to the algorithm.

For each  $\{i, j\} \in E$ , we are given noisy, independent measurements  $R_{ij}$  where

$$R_{ij} = \begin{cases} r_i - r_j; & \text{w.p } \eta \\ \sim \mathcal{U}[-M, M]; & \text{w.p } (1 - \eta). \end{cases} \quad (3.1)$$

The parameter  $\eta \in [0, 1]$  controls the level of noise. We will refer to this model as the **Erdős-Rényi Outliers** model, or in short,  $\text{ERO}(n, p, \eta)$ . This noise model was also considered in previous work in ranking [16] and angular synchronization [45]. Our goal is two fold - we would like to recover<sup>2</sup> the score vector  $r$ , and also the ranking  $\pi$  induced by  $r$ .

#### 3.1 Main idea: SVD-based spectral algorithm

We start by forming the measurement matrix  $H \in \mathbb{R}^{n \times n}$ , where

- $H_{ii} = 0, \forall i = 1, \dots, n$ ,
- $H_{ij} = R_{ij}$  and  $H_{ji} = -R_{ij}$ , if  $\{i, j\} \in E$ , and
- $H_{ij} = 0$ , if  $\{i, j\} \notin E$ .

If  $G$  is the complete graph and the measurements are noise free, it holds true that  $H = re^T - er^T$ , which is a rank 2 skew-symmetric matrix. Denoting  $\alpha = \frac{r^T e}{n}$ , one can verify that the two non-zero left singular vectors are  $u_1 = e/\sqrt{n}, u_2 = \frac{r - \alpha e}{\|r - \alpha e\|_2}$  with equal non-zero singular values  $\sigma_1 = \sigma_2 = \|r - \alpha e\|_2 \sqrt{n}$  (see Lemma 1). Therefore, given any orthonormal basis for  $\text{span}\{u_1, u_2\}$ , we can simply find a vector orthonormal to  $e/\sqrt{n}$  and which lies in  $\text{span}\{u_1, u_2\}$ ; this will give us candidate solutions  $\pm \frac{r - \alpha e}{\|r - \alpha e\|_2}$ . Multiplying these candidates by  $\sigma_1/\sqrt{n}$  recovers the scale information of  $r$ , while the sign ambiguity is resolved by selecting the candidate which is most consistent with the measurements. If one is interested only in ranking the items, then there is of course no need to estimate the scale above.

If  $G$  is not complete (thus there are missing edges) and the measurements are noisy, then  $H$  will typically not be rank 2, but as will see shortly, it can be thought of as a perturbation of  $\eta p(re^T - er^T)$ . Proceeding similarly to the noiseless case, we can find the top two singular vectors of  $H$  (denoted by  $\hat{u}_1, \hat{u}_2$ ), project  $u_1 = e/\sqrt{n}$  on to  $\text{span}\{\hat{u}_1, \hat{u}_2\}$  to obtain  $\bar{u}_1$ , and then find a unit vector in  $\text{span}\{\hat{u}_1, \hat{u}_2\}$  which is orthonormal to  $\bar{u}_1$  (call this  $\tilde{u}_2$ ). If the noise level is not too large and we have sufficiently many edges in  $G$ , then one can imagine that  $\tilde{u}_2$  will be close (up to a sign flip) to  $u_2$ , and  $\hat{\sigma}_i \approx \eta p \sigma_i$ . This implies that the centered version of  $\frac{\hat{\sigma}_1 \tilde{u}_2}{\eta p \sqrt{n}}$  will be close (up to a sign flip) to  $r - \alpha e$ , which we can resolve as described above. This approach is outlined formally as Algorithm 1, namely SVD-RS (SVD-Ranking and Synchronization) which is a spectral method for recovering the ranks and scores (up to a global shift) of a collection of  $n$  items. Additionally, we also consider a “normalized” version of SVD-RS, wherein  $H$  is replaced by  $H_{\text{ss}} = D^{-1/2} H D^{-1/2}$  with  $D$  being a diagonal matrix and  $D_{ii} = \sum_{j=1}^n |H_{ij}|$ . Clearly,  $H_{\text{ss}}$  is also skew symmetric. Such a normalization step is particularly useful when the degree distribution is skewed, and is commonly employed in other problem domains involving spectral methods, such as clustering (see for eg. [34]). The ensuing algorithm, namely SVD-NRS (SVD-Normalized Ranking and Synchronization), is outlined as Algorithm 2.

<sup>2</sup>Clearly, this is possible only up to a global shift.

---

**Algorithm 1** SVD-RS

---

- 1: **Input:** Measurement graph  $G = ([n], E)$  and pairwise measurements  $R_{ij}$  for  $\{i, j\} \in E$ . Also need  $\eta, p$  for recovering  $r$ .
  - 2: **Output:** Rank estimates:  $\hat{\pi}$  and score estimates  $\hat{r} \in \mathbb{R}^n$ .
  - 3: Form measurement matrix  $H \in \mathbb{R}^{n \times n}$  using  $R_{ij}$  as outlined in Section 3.1.
  - 4: Find the top two left singular vectors (resp. singular values) of  $H$ , namely  $\hat{u}_1, \hat{u}_2$  (resp.  $\hat{\sigma}_1, \hat{\sigma}_2$ ).
  - 5: Obtain vector  $\bar{u}_1$  as the orthogonal projection of  $u_1 = e/\sqrt{n}$  on to  $\text{span}\{\hat{u}_1, \hat{u}_2\}$ .
  - 6: Obtain a unit vector  $\tilde{u}_2 \in \text{span}\{\hat{u}_1, \hat{u}_2\}$  such that  $\tilde{u}_2 \perp \bar{u}_1$ .
  - 7: **Rank recovery:** Obtain ranking  $\tilde{\pi}$  induced by  $\tilde{u}_2$  (up to global sign ambiguity) and reconcile its global sign by minimizing the number of upsets. Output ranking estimate  $\hat{\pi}$ .
  - 8: **Recovering  $r$ :** Obtain  $\tilde{w} = \frac{\hat{\sigma}_1}{\eta p \sqrt{n}} \tilde{u}_2$  and  $\hat{r} = \tilde{w} - \frac{e^T \tilde{w}}{n} e$ . Reconcile global sign of  $\hat{r}$  by forming  $\hat{R} \in \mathbb{R}^{n \times n}$ , where  $\hat{R}_{i,j} = \hat{r}_i - \hat{r}_j$ , and choosing  $\hat{r}$  which minimizes  $\sum_{i,j} \left| \hat{R}_{i,j} - H_{i,j} \right|^2$ .
- 

---

**Algorithm 2** SVD-NRS

---

- 1: **Input:** Measurement graph  $G = ([n], E)$  and pairwise measurements  $R_{ij}$  for  $\{i, j\} \in E$ . Also need  $\eta, p$  for recovering  $r$ .
  - 2: **Output:** Rank estimates:  $\hat{\pi}$  and score estimates  $\hat{r} \in \mathbb{R}^n$ .
  - 3: Form  $H_{ss} = \bar{D}^{-1/2} H \bar{D}^{-1/2}$  with  $H$  formed using  $R_{ij}$  as outlined in Section 3.1.  $\bar{D}$  is a diagonal matrix with  $\bar{D}_{ii} = \sum_{j=1}^n |H_{ij}|$ .
  - 4: Find the top two left singular vectors (resp. singular values) of  $H_{ss}$ , namely  $\hat{u}_1, \hat{u}_2$  (resp.  $\hat{\sigma}_1, \hat{\sigma}_2$ ).
  - 5: Obtain vector  $\bar{u}_1$  as the orthogonal projection of  $u_1 = \frac{\bar{D}^{-1/2} e}{\|\bar{D}^{-1/2} e\|_2}$  on to  $\text{span}\{\hat{u}_1, \hat{u}_2\}$ .
  - 6: Obtain a unit vector  $\tilde{u}_2 \in \text{span}\{\hat{u}_1, \hat{u}_2\}$  such that  $\tilde{u}_2 \perp \bar{u}_1$ .
  - 7: **Rank recovery:** Obtain ranking  $\tilde{\pi}$  induced by  $\bar{D}^{1/2} \tilde{u}_2$  (up to global sign ambiguity) and reconcile its global sign by minimizing the number of upsets. Output ranking estimate  $\hat{\pi}$ .
  - 8: **Recovering  $r$ :** Obtain  $\tilde{w} = \frac{\hat{\sigma}_1}{\eta p \|\bar{D}^{-1/2} e\|_2} \bar{D}^{1/2} \tilde{u}_2$  and  $\hat{r} = \tilde{w} - \frac{e^T \tilde{w}}{n} e$ . Reconcile global sign of  $\hat{r}$  by forming  $\hat{R} \in \mathbb{R}^{n \times n}$ , where  $\hat{R}_{i,j} = \hat{r}_i - \hat{r}_j$ , and choosing  $\hat{r}$  which minimizes  $\sum_{i,j} \left| \hat{R}_{i,j} - H_{i,j} \right|^2$ .
- 

**Remark 1.** Note that in order to recover the ranks, Algorithms 1 and 2 do not need to know  $\eta, p$ . However, in order to recover the scale of (the zero mean version of)  $r$ , they do require to know  $\eta, p$ .

**Remark 2.** Note that the SVD-based algorithm introduced in [16] considers the four possible rankings induced by the singular vectors  $\{\pm \hat{u}_1, \pm \hat{u}_2\}$ , and chooses the one that minimizes the number of upsets. In contrast, SVD-RS first computes  $\bar{u}_1$  as the orthogonal projection of the all ones vector  $e$  on to  $\text{span}\{\hat{u}_1, \hat{u}_2\}$ , finds a vector  $\tilde{u}_2 \in \text{span}\{\hat{u}_1, \hat{u}_2\}$  perpendicular to  $\bar{u}_1$ , and finally extracts the ranking induced by  $\tilde{u}_2$ . Furthermore, unlike [16], SVD-NRS introduces an additional normalization step of the measurement matrix  $H$  to alleviate potential issues arising from skewed degree distributions. Such a normalization step is fairly common for spectral methods in general. For example, this operator was considered in [17] in the context of angular synchronization [45] and the graph realization problem, and [46] who introduced Vector Diffusion Maps for nonlinear dimensionality reduction and explored the interplay with the connection-Laplacian operator for vector fields over manifolds.

### 3.2 Summary of main theoretical results

1. Theorem 3 gives guarantees for rank recovery (up to global order reversal) for SVD-RS in terms of the maximum displacement error (defined in (4.6)) between the ranking  $\tilde{\pi}$  obtained in Step 7 of Algorithm 1, and the ground truth ranking  $\pi$ . Denoting  $\alpha = \frac{r^T e}{n}$  to be the average score, it states that if

$$\|r - \alpha e\|_2 \gtrsim \frac{M}{\eta \sqrt{p}}, \quad p \gtrsim \frac{\log n}{n}, \quad (3.2)$$



and if  $n$  is large enough, then  $\|\tilde{\pi} - \pi\|_\infty \lesssim \frac{\|r - \alpha e\|_2}{\rho} \gamma(n, M, \eta, p, \varepsilon, r)$  w.h.p. Here  $\gamma(n, M, \eta, p, \varepsilon, r)$  is the bound on the  $\ell_\infty$  error between  $\tilde{u}_2$  and  $u_2$  (up to global sign) in Theorem 2. For concreteness, if  $r_i = i$ , we show in Example 3 that for a fixed  $\delta \in (0, 1)$ , if  $p \gtrsim \frac{1}{\eta^2 \delta^{2/3}} \left( \frac{(\log n)^{2/3}}{n^{2/3}} \right)$  and  $n$  is large enough, then  $\|\tilde{\pi} - \pi\|_\infty \lesssim \delta \frac{n}{\log n}$  w.h.p. Note that this implies that  $\Omega(n^{4/3}(\log n)^{2/3})$  measurements suffice.

2. Theorem 4 provides  $\ell_2$  and  $\ell_\infty$  recovery guarantees for SVD-RS for estimating the score vector  $r$  by  $\hat{r}$  obtained in Step 8 of Algorithm 1 (up to a global shift and sign), and essentially states the following.

- (a) If the conditions in (3.2) hold, then  $\exists \beta \in \{-1, 1\}$  such that

$$\|\hat{r} - \beta(r - \alpha e)\|_2 \leq \frac{M}{\eta\sqrt{p}} + \sqrt{\frac{M\|r - \alpha e\|_2}{\eta p^{1/2}}}$$

holds w.h.p. If  $r_i = i$ , we show in Example 4 that if  $p \gtrsim \frac{\log n}{\eta^2 \delta^2 n}$  for a fixed  $\delta \in (0, 1)$ , then  $\|\hat{r} - \beta(r - \alpha e)\|_2 \lesssim \frac{n^{3/2}}{(\log n)^{1/4}} \sqrt{\delta}$  w.h.p. Hence  $\Omega(n \log n)$  measurements suffice for  $\ell_2$  recovery.

- (b) If the conditions in (3.2) hold, and  $n$  is large enough, then there exists  $\beta \in \{-1, 1\}$  such that

$$\|\hat{r} - \beta(r - \alpha e)\|_\infty \lesssim \|r - \alpha e\|_2 [C(n, M, \eta, p, r) + \sqrt{n} C^2(n, M, \eta, p, r)] + \frac{M(M - \alpha)}{\eta\sqrt{p}\|r - \alpha e\|_2}$$

holds, with  $C(n, M, \eta, p, r)$  as defined in (4.3). When  $r_i = i$ , we show in Example 4 that if  $p \gtrsim \frac{1}{\eta^2 \delta^{2/3}} \left( \frac{(\log n)^{2/3}}{n^{2/3}} \right)$  for a fixed  $\delta \in (0, 1)$ , and  $n$  is large enough, then  $\|\hat{r} - \beta(r - \alpha e)\|_\infty \lesssim \frac{n}{\log n} \delta$  holds w.h.p.

3. Theorem 6 provides a  $\ell_2$  error bound for SVD-NRS for recovering  $r$  by  $\hat{r}$  obtained in Step 8 of Algorithm 2 (up to a global shift and sign), in the same spirit as Theorem 4. While the conditions on  $p$  are more convoluted than those in Theorem 4 – in large part due to the normalization – the essence is the same. In the setting where  $r_i = i$ , we can derive a simplified version of the theorem as shown in Example 6. Let us now denote  $\alpha = \frac{r^T (\mathbb{E}[D])^{-1} e}{e^T (\mathbb{E}[D])^{-1} e}$  and  $\alpha' = \frac{e^T (r - \alpha e)}{n}$ . Theorem 6 states that if  $p \gtrsim \frac{\log n}{\eta^2 \delta^2 n}$  for fixed  $\delta \in (0, 1)$ , then there exists  $\beta \in \{-1, 1\}$  such that w.h.p,

$$\|\hat{r} - \beta(r - (\alpha + \alpha')e)\|_2 \lesssim n^{3/2} \eta^{1/4} \delta^{1/2} + \delta^{1/2} \frac{n^{3/2}}{(\log n)^{1/4}}.$$

The closest work to our setting are those of [27] and [54] which rely on matrix completion followed by a row sum to compute rankings. In our experiments, we also consider matrix completion as a preprocessing step before applying our SVD based algorithms, and find that in general it does improve the performance of our methods. However, this is at the cost of an additional computational overhead; our SVD based methods rely on computing the top two singular vectors of the input matrix which in practice can be done very efficiently. In general, the leading singular values and singular vectors can be computed using iterative techniques at a typical cost of  $O(pn^2)$ , using for example, a simple power method (since all iterations only require a matrix-vector product at a cost of  $pn^2$ , where  $p$  is the sampling probability). In the sparse setting, the computational cost is essentially linear in the number of nonzero entries in the matrix, thus making the approach scalable to large measurement graphs. Moreover the theoretical guarantees we provide for our problem setting are to our knowledge new and have been lacking in the literature currently.

## 4 Theoretical results for SVD-RS and SVD-NRS

We begin by detailing the theoretical results for SVD-RS in Section 4.1, followed by those for SVD-NRS in Section 4.2. Throughout, we instantiate the main theorems for the special setting where  $r_i = i$ , for ease of interpretability.

## 4.1 Analysis of Algorithm 1 (SVD-RS)

We will assume throughout this section that for  $\{i, j\} \in E$ , the measurement  $R_{ij}$  corresponds to  $i < j$ . This is clearly without loss of generality. From (3.1), we can write the entry  $H_{ij}$  (for  $i < j$ ) as the following mixture model

$$H_{ij} = \begin{cases} r_i - r_j & \text{w.p } \eta p \\ N_{ij} \sim \mathcal{U}[-M, M] & \text{w.p } (1 - \eta)p \\ 0 & \text{w.p } 1 - p, \end{cases}$$

with  $(H_{ij})_{i < j}$  being independent random variables. Since  $H_{ii} = 0$  and  $H_{ij} = -H_{ji}$  (by construction of  $H$ ), therefore for all  $i, j \in [n]$ , we obtain  $\mathbb{E}[H_{ij}] = \eta p(r_i - r_j)$ . In particular, we have that

$$\mathbb{E}[H] = \eta p C, \quad (4.1)$$

where  $C = re^T - er^T$  is a skew-symmetric matrix of rank 2. We can decompose  $H$  as

$$H = \mathbb{E}[H] + Z = \eta p C + Z,$$

where  $Z$  is a random noise matrix with  $Z_{ii} = 0, \forall i = 1, \dots, n$ . For  $1 \leq i < j \leq n$  the entries of  $Z$  are defined by the following mixture model

$$Z_{ij} = \begin{cases} (r_i - r_j) - \eta p(r_i - r_j) & ; \text{ w.p } \eta p \\ N_{ij} - \eta p(r_i - r_j) & ; \text{ w.p } (1 - \eta)p \\ -\eta p(r_i - r_j) & ; \text{ w.p } (1 - p). \end{cases} \quad (4.2)$$

Note that  $(Z_{ij})_{i < j}$  are independent<sup>3</sup> random variables and  $Z_{ij} = -Z_{ji}$ . Hence the matrices  $H, C, Z$  are all skew-symmetric. We now proceed to show in Theorem 1 that, provided  $p$  is large enough, it holds true that  $\tilde{u}_2$  (obtained in Step 6 of SVD-RS) is close to  $u_2 = \frac{r - \alpha e}{\|r - \alpha e\|_2}$  in the  $\ell_2$  norm (up to a global sign) with high probability. The proof is outlined in Section 5.1.

**Theorem 1** ( $\ell_2$  recovery, ERO model). *Denoting  $\alpha = \frac{r^T e}{n}$ , for a given  $0 < \varepsilon \leq 1/2$ , let  $\|r - \alpha e\|_2 \geq \frac{24M}{\eta} \sqrt{\frac{5}{3p}}(2 + \varepsilon)$ . Let  $\tilde{u}_2 \in \mathbb{R}^n$  be the vector obtained in Step 6 of SVD-RS (with  $\|\tilde{u}_2\|_2 = 1$ ) and  $u_2 = \frac{r - \alpha e}{\|r - \alpha e\|_2}$ . Then, there exists  $\beta \in \{-1, 1\}$  and a constant  $c_\varepsilon > 0$  depending only on  $\varepsilon$  such that*

$$\|\tilde{u}_2 - \beta u_2\|_2^2 \leq \frac{120M}{\eta} \sqrt{\frac{5}{3p}} \frac{(2 + \varepsilon)}{\|r - \alpha e\|_2},$$

with probability at least  $1 - 2n \exp\left(-\frac{20pn}{3c_\varepsilon}\right)$ .

Theorem 1 says that for  $\delta \in (0, 1)$ , if  $p \gtrsim \max\left\{\frac{M^2 \log n}{\eta^2 \delta^2 \|r - \alpha e\|_2^2}, \frac{\log n}{n}\right\}$ , then with high probability, we have

$$\|\tilde{u}_2 - \beta u_2\|_2^2 \lesssim \frac{\delta}{\sqrt{\log n}}.$$

Note that the bounds involved are invariant to the scaling of  $r$ , and essentially depend on the variance of the normalized entries of  $r$ , wherein each  $\frac{r_i}{M} \in [0, 1]$ . We can see that as  $\frac{\|r - \alpha e\|_2^2}{M^2}$  becomes small, then the corresponding condition on  $p$  becomes more stringent.

**Example 1.** *Consider the case where  $r_i = i$  for  $i = 1, \dots, n$ . Then  $M = n$ ,  $\alpha = \frac{n+1}{2}$  and  $\|r - \alpha e\|_2 = \Theta(n^{3/2})$ . Hence Theorem 1 now says that if  $p \gtrsim \frac{\log n}{\eta^2 \delta^2 n}$ , then with high probability, we have  $\|\tilde{u}_2 - \beta u_2\|_2^2 \lesssim \frac{\delta}{\sqrt{\log n}}$ .*

If in addition,  $n$  is also large enough, then we show in Theorem 2 that  $\tilde{u}_2$  is close to  $u_2$  (up to a global sign) in the  $\ell_\infty$  norm as well. The proof is outlined in Section 5.2.

<sup>3</sup>Since  $Z_{ii} = 0$ , clearly  $(Z_{ij})_{i \leq j}$  are independent as well.

**Theorem 2** ( $\ell_\infty$  recovery, ERO model). *With the same notation as in Theorem 1, for a given  $0 < \varepsilon \leq 1/2$ , let  $\|r - \alpha e\|_2 \geq \frac{24M}{\eta} \sqrt{\frac{5}{3p}}(2 + \varepsilon)$ . Assume  $p \geq \max\left\{\frac{1}{2n}, \frac{2 \log n}{15n}\right\}$ . Choose  $\xi > 1$ ,  $0 < \kappa < 1$  and define  $\mu = \frac{2}{\kappa+1}$ . Let  $n$  satisfy  $\frac{16}{\kappa} \leq (\log n)^\xi$ . Then, there exists  $\beta \in \{-1, 1\}$  and constants  $C_\varepsilon, c_\varepsilon > 0$  depending only on  $\varepsilon$  such that with probability at least*

$$1 - 2n \exp\left(-\frac{20pn}{3c_\varepsilon}\right) - \frac{4}{n} - 2n^{1-\frac{1}{4}(\log_\mu n)^{\xi-1}(\log_\mu e)^{-\xi}}$$

we have that

$$\|\tilde{u}_2 - \beta u_2\|_\infty \leq 4(2 + \sqrt{2})C(n, M, \eta, p, \varepsilon, r) + 4\sqrt{n}C^2(n, M, \eta, p, \varepsilon, r),$$

where

$$C(n, M, \eta, p, \varepsilon, r) = C_\varepsilon \left[ \left( \frac{M\sqrt{\log n}}{\eta\sqrt{p}\|r - \alpha e\|_2} + \frac{M^2(\log n)^{2\xi}}{\eta^2 p \|r - \alpha e\|_2^2} \right) \left( \frac{1}{\sqrt{n}} + \frac{M - \alpha}{\|r - \alpha e\|_2} \right) + \frac{M^3}{\eta^3 p^{3/2} \|r - \alpha e\|_2^3} \right]. \quad (4.3)$$

Let us look at bounding the admittedly complicated looking term  $C(n, M, \eta, p, \varepsilon, r)$ . For  $\delta \in (0, 1)$ , say  $p$  satisfies

$$p \gtrsim \frac{M^2}{\|r - \alpha e\|_2^2} \max \left\{ \frac{n \log^3 n}{\eta^2 \delta^2} \left( \frac{1}{\sqrt{n}} + \frac{M - \alpha}{\|r - \alpha e\|_2} \right)^2, \frac{\sqrt{n}(\log n)^{2\xi+1}}{\eta^2 \delta} \left( \frac{1}{\sqrt{n}} + \frac{M - \alpha}{\|r - \alpha e\|_2} \right), \frac{(\sqrt{n} \log n)^{2/3}}{\eta^2 \delta^{2/3}} \right\}. \quad (4.4)$$

Then we have that  $C(\cdot) \lesssim \frac{\delta}{\sqrt{n} \log n}$ . Hence if additionally  $p \gtrsim \max\left\{\frac{M^2}{\eta^2 \|r - \alpha e\|_2^2}, \frac{\log n}{n}\right\}$  and  $n = \Omega(1)$  hold, then Theorem 2 says that

$$\|\tilde{u}_2 - \beta u_2\|_\infty \lesssim \frac{\delta}{\sqrt{n} \log n}$$

holds w.h.p. Note that the condition on  $p$  is stricter than that in Theorem 1.

**Example 2.** *Let us revisit the case where  $r_i = i$  for all  $i$ . The condition (4.4) then simplifies to*

$$p \gtrsim \max \left\{ \frac{1}{\eta^2 \delta^2} \left( \frac{(\log n)^3}{n} \right), \frac{1}{\eta^2 \delta} \left( \frac{(\log n)^{2\xi+1}}{n} \right), \frac{1}{\eta^2 \delta^{2/3}} \left( \frac{(\log n)^{2/3}}{n^{2/3}} \right) \right\}. \quad (4.5)$$

Since  $\frac{M^2}{\|r - \alpha e\|_2^2} = \frac{1}{n}$ , we conclude that for a fixed  $\delta \in (0, 1)$ , if  $p \gtrsim \frac{1}{\eta^2 \delta^{2/3}} \left( \frac{(\log n)^{2/3}}{n^{2/3}} \right)$  and  $n = \Omega(1)$ , then  $\|\tilde{u}_2 - \beta u_2\|_\infty \lesssim \frac{\delta}{\sqrt{n} \log n}$  holds with high probability.

**Recovering ranks in the  $\ell_\infty$  norm.** Let us assume that  $\tilde{u}_2$  is aligned with  $u_2$ , i.e.,  $\|\tilde{u}_2 - u_2\|_\infty \leq \gamma$  where  $\gamma$  is the bound in Theorem 2. For any  $s \in \mathbb{R}^n$ , we say that the permutation  $\pi : [n] \rightarrow [n]$  is *consistent* with  $s$  if for all pairs  $(i, j)$ ,  $\pi(i) < \pi(j)$  implies  $s_i \geq s_j$ . For the purpose of recovering the rankings, we will consider the entries of  $r$  to be pairwise distinct, i.e.  $r_i \neq r_j$  for all  $i \neq j$ . Hence there is a unique permutation  $\pi$  which is consistent with  $r$ . Let  $\tilde{\pi}$  be a permutation that is consistent with  $\tilde{u}_2$ , note that this is not unique since some entries of  $\tilde{u}_2$  could have the same value. Our goal is to bound the maximum displacement error between  $\pi$  and  $\tilde{\pi}$  defined as

$$\|\tilde{\pi} - \pi\|_\infty := \max_i \left( \sum_{\pi(j) > \pi(i)} \mathbf{1}_{\tilde{\pi}(j) < \tilde{\pi}(i)} + \sum_{\pi(j) < \pi(i)} \mathbf{1}_{\tilde{\pi}(j) > \tilde{\pi}(i)} \right). \quad (4.6)$$

To this end, we have the following theorem the proof of which is provided in Section 5.3. Notice that bounding  $\|\pi - \tilde{\pi}\|_\infty$  requires a bound on  $\|\tilde{u}_2 - u_2\|_\infty$  which is obtained from Theorem 2. The proof technique is essentially the same as that of [25, Theorem 24].

**Theorem 3.** Assuming  $r_i \neq r_j$  for all  $i \neq j$ , let  $\pi$  denote the (unique) ranking consistent with  $r$  and also define  $\rho := \min_{i \neq j} |r_i - r_j|$ . Assuming  $\beta = 1$ , denote  $\gamma(n, M, \eta, p, \varepsilon, r)$  to be the bound on  $\|\tilde{u}_2 - u_2\|_\infty$  in Theorem 2. Then under the notation and assumptions in Theorem 2, we have with probability at least  $1 - 2n \exp\left(-\frac{20pn}{3c_\varepsilon}\right) - \frac{4}{n} - 2n^{1-\frac{1}{4}(\log_\mu n)^{\xi-1}(\log_\mu e)^{-\xi}}$  that

$$\|\tilde{\pi} - \pi\|_\infty \leq \frac{4\|r - \alpha e\|_2}{\rho} \gamma(n, M, \eta, p, \varepsilon, r) \quad (4.7)$$

holds for all rankings  $\tilde{\pi}$  which are consistent with  $\tilde{u}_2$ .

**Example 3.** Let us examine the bound in (4.7) for the case where  $r_i = i$  for  $i = 1, \dots, n$ . We have seen that  $\|r - \alpha e\|_2 = \Theta(n^{3/2})$ . Theorem 2 tells us that for a fixed  $\delta \in (0, 1)$ , if  $p \gtrsim \frac{1}{\eta^2 \delta^{2/3}} \left(\frac{(\log n)^{2/3}}{n^{2/3}}\right)$  and  $n = \Omega(1)$ , then  $\|\tilde{u}_2 - u_2\|_\infty \leq \gamma(\cdot) \lesssim \frac{\delta}{\sqrt{n} \log n}$  w.h.p. Since  $\rho = 1$ , we obtain from (4.7) that

$$\|\tilde{\pi} - \pi\|_\infty \lesssim \delta \frac{n}{\log n}.$$

**Recovering scores in the  $\ell_2$  and  $\ell_\infty$  norms.** We now bound the error between the score vector estimate  $\hat{r}$  (obtained in Step 8 of Algorithm 1) and  $r$  up to a global shift and sign. This is shown in the following theorem for the  $\ell_2$  and  $\ell_\infty$  norms. The proof is outlined in Section 5.4. Note that both  $\hat{r}$  and  $r - \alpha e$  are centered vectors.

**Theorem 4.** Recall  $\hat{r} \in \mathbb{R}^n$  as defined in Step 8 of Algorithm 1.

1. Under the notation and assumptions of Theorem 1, there exists  $\beta \in \{-1, 1\}$  such that

$$\|\hat{r} - \beta(r - \alpha e)\|_2 \leq \frac{8M}{\eta} \sqrt{\frac{5}{3p}} (2 + \varepsilon) + \sqrt{\frac{120M(2 + \varepsilon)\|r - \alpha e\|_2}{\eta p^{1/2}}} \left(\frac{5}{3}\right)^{1/4} \quad (4.8)$$

with probability at least  $1 - 2n \exp\left(-\frac{20pn}{3c_\varepsilon}\right)$ .

2. Under the notation and assumptions of Theorem 2, there exists  $\beta \in \{-1, 1\}$  such that

$$\begin{aligned} \|\hat{r} - \beta(r - \alpha e)\|_\infty &\leq \frac{16}{3} \|r - \alpha e\|_2 [(2 + \sqrt{2})C(n, M, \eta, p, \varepsilon, r) + \sqrt{n}C^2(n, M, \eta, p, \varepsilon, r)] \\ &\quad + 8\sqrt{\frac{5}{3}} (2 + \varepsilon) \frac{M(M - \alpha)}{\eta\sqrt{p}\|r - \alpha e\|_2} \end{aligned} \quad (4.9)$$

with probability at least

$$1 - 2n \exp\left(-\frac{20pn}{3c_\varepsilon}\right) - \frac{4}{n} - 2n^{1-\frac{1}{4}(\log_\mu n)^{\xi-1}(\log_\mu e)^{-\xi}}.$$

**Example 4.** When  $r_i = i$  for all  $i$ , the bound in (4.8) is of the form

$$\|\hat{r} - \beta(r - \alpha e)\|_2 \lesssim \frac{n}{\eta\sqrt{p}} + \frac{n^{5/4}}{\sqrt{\eta}p^{1/4}}. \quad (4.10)$$

Recall from Theorem 1 that  $p$  is required to satisfy  $p \gtrsim \frac{\log n}{\eta^2 \delta^2 n}$  for  $\delta \in (0, 1)$ . Then, (4.10) simplifies to

$$\|\hat{r} - \beta(r - \alpha e)\|_2 \lesssim \frac{n^{3/2}}{(\log n)^{1/4}} \sqrt{\delta}.$$

Turning our attention to (4.9), recall from Theorem 2 that we require  $p \gtrsim \frac{1}{\eta^2 \delta^{2/3}} \left(\frac{(\log n)^{2/3}}{n^{2/3}}\right)$  and  $n = \Omega(1)$ . This leads to

$$\|\hat{r} - \beta(r - \alpha e)\|_\infty \lesssim \frac{n}{\log n} \delta + \frac{\sqrt{n}}{\eta\sqrt{p}} \lesssim \frac{n}{\log n} \delta + \frac{n^{5/6}}{(\log n)^{1/3}} \delta^{1/3} \lesssim \frac{n}{\log n} \delta.$$

## 4.2 Analysis of Algorithm 2 (SVD-NRS)

We now analyze Algorithm 2, namely SVD-NRS, under the ERO model. The random matrix  $\bar{D}^{-1/2}H\bar{D}^{-1/2}$  will concentrate around  $(\mathbb{E}[\bar{D}])^{-1/2}\mathbb{E}[H](\mathbb{E}[\bar{D}])^{-1/2}$ . Recall from (4.1) that  $\mathbb{E}[H] = \eta p(re^T - er^T)$ , hence  $(\mathbb{E}[\bar{D}])^{-1/2}\mathbb{E}[H](\mathbb{E}[\bar{D}])^{-1/2}$  is a rank-2 matrix. It is easy to verify that

$$\mathbb{E}[\bar{D}_{ii}] = \eta p \sum_{j=1}^n |r_i - r_j| + (1 - \eta)p \frac{M}{2}.$$

It will be useful to denote

$$\max_i \mathbb{E}[\bar{D}_{ii}] = p \left[ \eta \max_i \left( \sum_{j=1}^n |r_i - r_j| \right) + (1 - \eta) \frac{M}{2} \right] = p\lambda_{\max}, \quad (4.11)$$

$$\min_i \mathbb{E}[\bar{D}_{ii}] = p \left[ \eta \min_i \left( \sum_{j=1}^n |r_i - r_j| \right) + (1 - \eta) \frac{M}{2} \right] = p\lambda_{\min}. \quad (4.12)$$

Recall  $\tilde{u}_2 \in \mathbb{R}^n$  obtained in Step 6 of SVD-NRS, and denote  $u_2 = \frac{(\mathbb{E}[\bar{D}])^{-1/2}(r - \alpha e)}{\|(\mathbb{E}[\bar{D}])^{-1/2}(r - \alpha e)\|_2}$ . The following theorem bounds the  $\ell_2$  error between  $\tilde{u}_2$  and  $u_2$  up to a global sign, its proof is given in Section 5.5.

**Theorem 5** ( $\ell_2$  recovery, ERO model). *Denote*

$$\alpha = \frac{r^T(\mathbb{E}[\bar{D}])^{-1}e}{e^T(\mathbb{E}[\bar{D}])^{-1}e}, \quad A(\eta, M) = \eta M^2 + (1 - \eta) \frac{M}{2} \quad \text{and} \quad C_1(\eta, M) = 4A^{1/4}(\eta, M).$$

For  $\varepsilon > 0$ , denote

$$\begin{aligned} \sigma_{\min} &= \frac{\eta \|r - \alpha e\|_2 \sqrt{n}}{\lambda_{\max}}, \quad \sigma_{\max} = \frac{\eta \|r - \alpha e\|_2 \sqrt{n}}{\lambda_{\min}}, \\ \tilde{\Delta} &= 16M \sqrt{\frac{5}{3} \frac{pn}{pn}} \frac{(2 + \varepsilon)}{p\lambda_{\min}} + \frac{C_1(\eta, M)(np \log n)^{1/4} \sigma_{\max}}{p^{3/2} \lambda_{\min}^{3/2}} \left( \frac{C_1(\eta, M)(np \log n)^{1/4}}{\sqrt{p\lambda_{\min}}} + 2\sqrt{2} \right). \end{aligned}$$

Let  $\tilde{u}_2 \in \mathbb{R}^n$  be the vector obtained in Step 6 of SVD-NRS (with  $\|\tilde{u}_2\|_2 = 1$ ) and denote  $u_2 = \frac{(\mathbb{E}[\bar{D}])^{-1/2}(r - \alpha e)}{\|(\mathbb{E}[\bar{D}])^{-1/2}(r - \alpha e)\|_2}$ . If

$$p \geq \frac{M^2}{9A(\eta, M)} \frac{\log n}{n}, \quad p \geq \frac{16(\sqrt{2} + 1)^2 A(\eta, M) n \log n}{\lambda_{\min}^2} \quad \text{and} \quad \tilde{\Delta} \leq \frac{\sigma_{\min}}{3} \quad (4.13)$$

hold, then there exists  $\beta \in \{-1, 1\}$  and a constant  $c_\varepsilon > 0$  depending only on  $\varepsilon$  such that

$$\|\tilde{u}_2 - \beta u_2\|_2^2 \leq \frac{15\tilde{\Delta}}{\sigma_{\min}}$$

holds with probability at least  $1 - \frac{2}{n} - 2n \exp\left(-\frac{20pn}{3c_\varepsilon}\right)$ .

The quantities  $\sigma_{\min}, \sigma_{\max}$  are lower and upper bounds on the non-zero singular values of the rank 2 matrix  $(\mathbb{E}[\bar{D}])^{-1/2}\mathbb{E}[H](\mathbb{E}[\bar{D}])^{-1/2}$ , while  $\tilde{\Delta}$  is an upper bound on  $\|\bar{D}^{-1/2}H\bar{D}^{-1/2} - (\mathbb{E}[\bar{D}])^{-1/2}\mathbb{E}[H](\mathbb{E}[\bar{D}])^{-1/2}\|_2$ . The theorem essentially states that if  $p$  is sufficiently large, then  $\tilde{u}_2$  is close to  $u_2$  up to a sign flip. In order to get a sense of the precise scaling of the quantities involved, we consider the special case  $r_i = i$ .

**Example 5.** Consider  $r_i = i$  for all  $i$  with  $\eta \in (0, 1)$  fixed, hence  $M = n$ . Denoting  $S_i = \sum_{j=1}^n |r_i - r_j|$ , one can verify that  $S_i = i^2 - i(n + 1) + \frac{n^2 + n}{2}$ . This implies that  $\lambda_{\max}, \lambda_{\min} = \Theta(\eta n^2) = A(\eta, M)$ , and so,  $C(\eta, M) = \Theta(\eta^{1/4} n^{1/2})$ . Now let us find the scaling of  $\alpha$ . Since  $\mathbb{E}[\bar{D}_{ii}] = \Theta(p\eta n^2)$ , we obtain

$$\alpha = \frac{\sum_{i=1}^n i(\mathbb{E}[\bar{D}_{ii}])^{-1}}{\sum_{i=1}^n (\mathbb{E}[\bar{D}_{ii}])^{-1}} = \Theta\left(\frac{\sum_{i=1}^n i}{n}\right) = \Theta(n).$$

This in particular implies that  $\|r - \alpha e\|_2^2 = \sum_{j=1}^n (j - \alpha)^2 = \Theta(n^3)$ . We then obtain  $\sigma_{\min}, \sigma_{\max} = \Theta(1)$ ; plugging the aforementioned scalings in the expression for  $\tilde{\Delta}$  leads to

$$\tilde{\Delta} = O\left(\frac{1}{\eta\sqrt{np}} + \frac{(\log n)^{1/4}}{(\eta p)^{5/4}n^{9/4}} \left(\frac{(\log n)^{1/4}}{(\eta p)^{1/4}} + 2\sqrt{2}\right)\right).$$

The first two conditions on  $p$  in (4.13) translate to  $p \gtrsim \frac{1}{\eta} \frac{\log n}{n}$ . Since  $\sigma_{\min} = \Theta(1)$ , hence the third condition in (4.13) holds provided  $\tilde{\Delta} \lesssim 1$ . In fact, for any  $\delta \in (0, 1)$ , we have  $\tilde{\Delta} \lesssim \frac{\delta}{\sqrt{\log n}}$  if  $p$  additionally satisfies

$$\frac{1}{\eta\sqrt{np}} \lesssim \frac{\delta}{\sqrt{\log n}} \Leftrightarrow p \gtrsim \frac{\log n}{\eta^2 \delta^2 n} \quad \text{and} \quad \frac{(\log n)^{1/4}}{(\eta p)^{5/4}n^{9/4}} \lesssim \frac{\delta}{\sqrt{\log n}} \Leftrightarrow p \gtrsim \frac{(\log n)^{3/5}}{\eta n^{9/5} \delta^{4/5}}.$$

Clearly, the first condition on  $p$  dominates since  $\frac{\log n}{\eta^2 \delta^2 n} \gtrsim \frac{(\log n)^{3/5}}{\eta n^{9/5} \delta^{4/5}}$ . To summarize, Theorem 5 states that if  $p$  satisfies

$$p \gtrsim \frac{\log n}{\eta^2 \delta^2 n}, \tag{4.14}$$

then  $\exists \beta \in \{-1, 1\}$  such that  $\|\tilde{u}_2 - \beta u_2\|_2^2 \lesssim \delta$ .

**Recovering scores in the  $\ell_2$  norm.** The following theorem bounds the  $\ell_2$  error between the score vector estimate  $\hat{r}$  (obtained in Step 8 of Algorithm 2) and  $r$  up to a global sign and shift. Its proof is outlined in Section 5.6. Note that both  $\hat{r}$  and  $r - (\alpha + \alpha')e$  are centered vectors.

**Theorem 6.** Recall  $\hat{r} \in \mathbb{R}^n$  as defined in Step 8 of Algorithm 2. Under the notation and assumptions of Theorem 5, and denoting  $\alpha' = \frac{e^{\mathbf{T}(r - \alpha e)}}{n}$ , there exists  $\beta \in \{-1, 1\}$  such that

$$\begin{aligned} \|\hat{r} - \beta(r - (\alpha + \alpha')e)\|_2 &\leq \frac{2}{\eta p} \left[ \sqrt{\frac{3}{n}} (\sigma_{\max} + \tilde{\Delta}) (2\sqrt{2} + 1)^{1/2} (A(\eta, M) \eta p \log n)^{1/4} \left( \sqrt{p \lambda_{\max}} + \frac{\lambda_{\max}}{\lambda_{\min}} \right) \right. \\ &\quad \left. + \frac{\tilde{\Delta} p \lambda_{\max}}{\sqrt{n}} + \sqrt{\frac{15 \tilde{\Delta}}{\sigma_{\min} n}} \sigma_{\max} p \lambda_{\max} \right] \end{aligned}$$

holds with probability at least  $1 - \frac{2}{n} - 2n \exp\left(-\frac{20pn}{3c_\varepsilon}\right)$ .

**Example 6.** Consider the scenario where  $r_i = i$  for each  $i$ . Then as discussed above for Theorem 5, if  $p$  satisfies (4.14), we obtain the bound

$$\begin{aligned} \|\hat{r} - \beta(r - (\alpha + \alpha')e)\|_2 &\lesssim \frac{1}{\eta p} \left[ \frac{(\eta n^3 p \log n)^{1/4}}{\sqrt{n}} (\sqrt{p \eta n^2} + 1) + \frac{p \eta n^2 \delta}{\sqrt{n} \log n} + \sqrt{\frac{\delta}{n \sqrt{\log n}}} p \eta n^2 \right] \\ &\lesssim \frac{1}{\eta p} \left[ \eta^{3/4} n^{5/4} (\log n)^{1/4} p^{3/4} + \eta p \delta^{1/2} \frac{n^{3/2}}{(\log n)^{1/4}} \right] \\ &= \frac{n^{5/4} (\log n)^{1/4}}{(\eta p)^{1/4}} + \delta^{1/2} \frac{n^{3/2}}{(\log n)^{1/4}} \\ &\lesssim n^{3/2} \eta^{1/4} \delta^{1/2} + \delta^{1/2} \frac{n^{3/2}}{(\log n)^{1/4}} \quad (\text{using (4.14)}). \end{aligned}$$

## 5 Proofs

### 5.1 Proof of Theorem 1

We now outline the proof of Theorem 1 by highlighting the steps involved, along with the intermediate Lemmas (the proofs of which are in Appendix C). The proof is broken into the following steps.

**Step 1: Singular values and singular vectors of  $C$ .** We begin by finding the singular values and singular vectors of  $C$ .

**Lemma 1.** For  $C = re^T - er^T$ , it holds true that  $C = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T$ , with  $\sigma_1 = \sigma_2 = \|r - \alpha e\|_2 \sqrt{n}$ , and

$$u_1 = v_2 = \frac{e}{\sqrt{n}}; \quad v_1 = -\frac{r - \alpha e}{\|r - \alpha e\|_2}; \quad u_2 = \frac{r - \alpha e}{\|r - \alpha e\|_2},$$

where  $\alpha = \frac{r^T e}{n}$ .

**Step 2: Bounding the spectral norm of  $Z$ .** Recall that  $(Z_{ij})_{i \leq j}$  are independent zero-mean random variables with  $Z_{ii} = 0$  and moreover,  $Z_{ji} = -Z_{ij}$ . Denote  $Z = S - S^T$ , where  $S$  is an upper-triangular matrix with  $S_{ij} = Z_{ij}$  for  $i \leq j$ . Note that  $\|Z\|_2 \leq \|S\|_2 + \underbrace{\|S^T\|_2}_{=\|S\|_2} \leq 2\|S\|_2$ . We define a symmetric matrix

$$\tilde{S} := \begin{bmatrix} 0 & S \\ S^T & 0 \end{bmatrix} \in \mathbb{R}^{2n \times 2n}.$$

It is easy to check that  $\|\tilde{S}\|_2 = \|S\|_2$ , and thus  $\|Z\|_2 \leq 2\|\tilde{S}\|_2$ . Since  $\tilde{S}$  is a random symmetric matrix with the entries  $(\tilde{S}_{ij})_{i \leq j}$  being centered, independent and bounded random variables, we can bound  $\|\tilde{S}\|_2$  using a recent result of Bandeira and van Handel [2, Corollary 3.12, Remark 3.13] that gives non-asymptotic bounds on the spectral norm of such random symmetric matrices. This altogether leads to the following Lemma.

**Lemma 2.** For  $Z \in \mathbb{R}^{n \times n}$  as defined in (4.2), we have that

$$\|Z\|_2 \leq 8M \sqrt{\frac{5}{3}pn}(2 + \varepsilon) \quad (5.1)$$

holds with probability at least  $1 - 2n \exp\left(-\frac{20pn}{3c_\varepsilon}\right)$ .

**Step 3: Using Wedin's bound.** Let us denote  $U = [u_1 \ u_2] \in \mathbb{R}^{n \times 2}$  (recall  $u_1, u_2$  from Lemma 1). Denote  $\hat{U} = [\hat{u}_1 \ \hat{u}_2] \in \mathbb{R}^{n \times 2}$  where  $\hat{u}_1, \hat{u}_2$  are the left singular vectors corresponding to the top two singular values of  $H$ . The following Lemma states that  $\mathcal{R}(U)$  is close to  $\mathcal{R}(\hat{U})$  if  $\|Z\|_2$  is small. The proof uses Wedin's bound [52] for perturbation of singular subspaces.

**Lemma 3.** Given  $H = \eta p C + Z$ , where  $\|Z\|_2 \leq \Delta$ , if  $\Delta < \eta p \|r - \alpha e\|_2 \sqrt{n}$ , then

$$\|(I - \hat{U} \hat{U}^T)U\|_2 \leq \frac{\Delta}{\eta p \|r - \alpha e\|_2 \sqrt{n} - \Delta} \quad (=:\delta). \quad (5.2)$$

**Step 4: Analyzing the projection step.** Next, we project  $u_1$  on  $\text{span}(\hat{U})$  to obtain  $\bar{u}_1$ , find a unit vector orthogonal to  $\bar{u}_1$  lying in  $\mathcal{R}(\hat{U})$  (denote this by  $\tilde{u}_2$ ), and show that  $\tilde{u}_2$  is close to  $u_2$  up to a sign flip.

**Lemma 4.** With  $\delta$  as defined in Lemma 3 let  $\delta \leq 1/2$  hold. Then, there exists  $\beta \in \{-1, 1\}$  such that

$$\|\tilde{u}_2 - \beta u_2\|_2^2 \leq 10\delta.$$

**Step 5: Putting it together.** We now use the above Lemma's to readily obtain Theorem 1 (see appendix for details). Using  $\|Z\|_2 \leq \Delta = 8M \sqrt{\frac{5}{3}pn}(2 + \varepsilon)$  in the expression for  $\delta$ , one can verify that

$$\delta \leq 1/2 \Leftrightarrow 3\Delta \leq \eta p \|r - \alpha e\|_2 \sqrt{n} \quad (5.3)$$

holds if  $\|r - \alpha e\|_2 \geq \frac{24M}{\eta} \sqrt{\frac{5}{3p}}(2 + \varepsilon)$ . By using the bound on  $\Delta$  in (5.3) in the expression for  $\delta$ , we finally obtain

$$\delta \leq \frac{12M}{\eta} \sqrt{\frac{5}{3p}} \frac{(2 + \varepsilon)}{\|r - \alpha e\|_2}.$$

Plugging this in Lemma 4 yields the stated  $\ell_2$  bound and completes the proof.

## 5.2 Proof of Theorem 2

Recall that  $U = [u_1 \ u_2]$ ,  $\hat{U} = [\hat{u}_1 \ \hat{u}_2] \in \mathbb{R}^{n \times 2}$  correspond to the two largest singular vectors of  $\mathbb{E}[H]$ , and  $H$  respectively. The proof is broken down into three main steps. The first step involves bounding  $\|\tilde{u}_2 - \beta u_2\|_\infty$  in terms of  $\|\hat{U} - UO^*\|_{\max}$  – where  $O^*$  is a  $2 \times 2$  orthogonal matrix that “aligns”  $U$  with  $\hat{U}$  – provided  $\|Z\|_2$  is small. The second step involves bounding  $\|\hat{U} - UO^*\|_{\max}$ , and in the third step we combine the results from the previous steps. The proofs of all lemmas are provided in Appendix D.

**Step 1: Bounding  $\|\tilde{u}_2 - \beta u_2\|_\infty$  in terms of  $\|\hat{U} - UO^*\|_{\max}$ .** The following Lemma states this formally.

**Lemma 5.** *For  $\Delta$  as defined in Lemma 3, let  $\Delta \leq \frac{\eta p \|r - \alpha e\|_2 \sqrt{n}}{3}$  hold. Then for any orthogonal matrix  $O \in \mathbb{R}^{2 \times 2}$ , there exists  $\beta \in \{-1, 1\}$  depending on  $O$  such that*

$$\|\tilde{u}_2 - \beta u_2\|_\infty \leq 4\|\hat{U} - UO\|_{\max} \left( 2 + \frac{\sqrt{n}(M - \alpha)}{\|r - \alpha e\|_2} \right) + 4\sqrt{n}\|\hat{U} - UO\|_{\max}^2. \quad (5.4)$$

Moreover, there exists an orthogonal matrix  $O^* \in \mathbb{R}^{2 \times 2}$  such that

$$\|\hat{U} - UO^*\|_2 \leq \frac{3\Delta}{\eta p \sqrt{n} \|r - \alpha e\|_2}. \quad (5.5)$$

**Step 2: Bounding  $\|\hat{U} - UO^*\|_{\max}$ .** Since  $H^T = -H$ , therefore  $-H^2 = HH^T$ . Using the SVD  $H = \hat{U}\hat{\Sigma}\hat{V}$ , we arrive at

$$\begin{aligned} HH^T\hat{U} &= -H^2\hat{U} = \hat{U}\hat{\Sigma}^2 \\ \Rightarrow \hat{U} &= -H^2\hat{U}\hat{\Sigma}^{-2} = -(\mathbb{E}[H]^2 + \mathbb{E}[H]Z + Z\mathbb{E}[H] + Z^2)\hat{U}\hat{\Sigma}^{-2}. \end{aligned} \quad (5.6)$$

For convenience of notation, denote  $\sigma = \sigma_i(\mathbb{E}[H]) = \eta p \sqrt{n} \|r - \alpha e\|_2$ ;  $i = 1, 2$ . Since  $\mathbb{E}[H]^T = -\mathbb{E}[H]$ , we get

$$\begin{aligned} -\mathbb{E}[H]^2 &= \mathbb{E}[H]\mathbb{E}[H]^T = U\Sigma^2U^T = \sigma^2(UO^*)(UO^*)^T \\ UO^* &= -\frac{1}{\sigma^2}(\mathbb{E}[H])^2(UO^*). \end{aligned} \quad (5.7)$$

Subtracting (5.7) from (5.6) leads to

$$\begin{aligned} \hat{U} - UO^* &= \underbrace{-\mathbb{E}[H]^2(\hat{U} - UO^*)\hat{\Sigma}^{-2}}_{E_1} + \underbrace{\mathbb{E}[H]^2UO^*(\sigma^{-2}I - \hat{\Sigma}^{-2})}_{E_2} - \underbrace{\mathbb{E}[H]Z\hat{U}\hat{\Sigma}^{-2}}_{E_3} - \underbrace{Z\mathbb{E}[H]\hat{U}\hat{\Sigma}^{-2}}_{E_4} - \underbrace{Z^2\hat{U}\hat{\Sigma}^{-2}}_{E_5} \\ &= E_1 + E_2 - E_3 - E_4 - E_5 \\ \Rightarrow \|\hat{U} - UO^*\|_{\max} &\leq \sum_{i=1}^5 \|E_i\|_{\max}. \end{aligned} \quad (5.8)$$

We now proceed to bound each term in the RHS of (5.8). This is stated precisely in the following Lemma.

**Lemma 6.** *If  $\Delta \leq \frac{\sigma}{3} = \frac{\eta p \sqrt{n} \|r - \alpha e\|_2}{3}$ , then the following holds true.*

1.  $\|E_1\|_{\max} \leq \frac{27\Delta}{4\sigma} \left( \frac{1}{\sqrt{n}} + \frac{M - \alpha}{\|r - \alpha e\|_2} \right)$ .
2.  $\|E_2\|_{\max} \leq \frac{21\Delta}{4\sigma} \left( \frac{1}{\sqrt{n}} + \frac{M - \alpha}{\|r - \alpha e\|_2} \right)$ .
3.  $\|E_3\|_{\max} \leq \frac{9\Delta}{4\sigma} \left( \frac{1}{\sqrt{n}} + \frac{M - \alpha}{\|r - \alpha e\|_2} \right)$ .
4.  $\|E_4\|_{\max} \leq \frac{9}{4\sigma} (\|Zu_1\|_\infty + \|Zu_2\|_\infty)$ .
5.  $\|E_5\|_{\max} \leq \frac{27\Delta^3}{4\sigma^3} + \frac{9}{4\sigma^2} (\|Z^2u_1\|_\infty + \|Z^2u_2\|_\infty)$ .



Next, we will bound  $\|Zu_i\|_\infty$  and  $\|Z^2u_i\|_\infty$  for  $i = 1, 2$ . Bounding  $\|Zu_i\|_\infty$  is a relatively straightforward consequence of Bernstein's inequality [6, Corollary 2.11], we state this in the following Lemma.

**Lemma 7.** *If  $p \geq \frac{2 \log n}{15n}$  then the following holds.*

1.  $\mathbb{P}(\|Zu_1\|_\infty \geq \frac{2\sqrt{2}+4}{3}M\sqrt{15p \log n}) \leq \frac{2}{n}$ .
2.  $\mathbb{P}(\|Zu_2\|_\infty \geq \frac{2\sqrt{2}+4}{3}M\sqrt{15pBn \log n}) \leq \frac{2}{n}$  where  $B = \frac{(M-\alpha)^2}{\|r-\alpha e\|_2^2}$ .

Bounding  $\|Z^2u_i\|_\infty$  is a considerably more challenging task since the entries within a row of  $Z^2$  will not be independent now. However, leveraging a recent entry-wise concentration result for the product of a random matrix (raised to a power) and a fixed vector [21, Theorem 15] (see Theorem 11), we are able to do so and arrive at the following Lemma.

**Lemma 8.** *Let  $p \geq \frac{1}{2n}$  hold. Choose  $\xi > 1, 0 < \kappa < 1$  and define  $\mu = \frac{2}{\kappa+1}$ . Then if  $\frac{16}{\kappa} \leq (\log n)^\xi$ , the following holds true.*

1.  $\mathbb{P}(\|Z^2u_1\|_\infty \geq 8\sqrt{np}M^2(\log n)^{2\xi}) \leq n^{1-\frac{1}{4}(\log_\mu n)^{\xi-1}(\log_\mu e)^{-\xi}}$ .
2.  $\mathbb{P}(\|Z^2u_2\|_\infty \geq 8npM^2 \frac{M-\alpha}{\|r-\alpha e\|_2} (\log n)^{2\xi}) \leq n^{1-\frac{1}{4}(\log_\mu n)^{\xi-1}(\log_\mu e)^{-\xi}}$ .

**Step 3: Putting it together.** We now combine the results of the preceding Lemmas to derive the final approximation bound. Recall from Lemma 2 that  $\Delta = 8M\sqrt{\frac{5}{3}pn}(2 + \varepsilon)$ . Also,  $\sigma = \eta p \sqrt{n} \|r - \alpha e\|_2$  and so

$$\frac{\Delta}{\sigma} = \frac{8M}{\eta} \sqrt{\frac{5}{3p}} \frac{(2 + \varepsilon)}{\|r - \alpha e\|_2}. \quad (5.9)$$

Using Lemmas 6,7,8 and (5.9) in (5.8), we get

$$\begin{aligned} \|\widehat{U} - UO^*\|_{\max} &\leq \left( \frac{114M}{\eta} \sqrt{\frac{5}{3p}} \frac{(2 + \varepsilon)}{\|r - \alpha e\|_2} \right) \left( \frac{1}{\sqrt{n}} + \frac{M - \alpha}{\|r - \alpha e\|_2} \right) \\ &\quad + \left( \frac{3\sqrt{15}(2\sqrt{2} + 4)}{4} \right) \left( \frac{M\sqrt{\log n}}{\eta\sqrt{p}\|r - \alpha e\|_2} \right) \left( \frac{1}{\sqrt{n}} + \frac{M - \alpha}{\|r - \alpha e\|_2} \right) \\ &\quad + \frac{18M^2(\log n)^{2\xi}}{\eta^2 p \|r - \alpha e\|_2^2} \left( \frac{1}{\sqrt{n}} + \frac{M - \alpha}{\|r - \alpha e\|_2} \right) + \frac{27}{4} \left( \frac{8\sqrt{5}}{3} \right)^3 \frac{M^3}{\eta^3 p^{3/2} \|r - \alpha e\|_2^3} \\ &\leq C_\varepsilon \left[ \left( \frac{M\sqrt{\log n}}{\eta\sqrt{p}\|r - \alpha e\|_2} + \frac{M^2(\log n)^{2\xi}}{\eta^2 p \|r - \alpha e\|_2^2} \right) \left( \frac{1}{\sqrt{n}} + \frac{M - \alpha}{\|r - \alpha e\|_2} \right) + \frac{M^3}{\eta^3 p^{3/2} \|r - \alpha e\|_2^3} \right] \\ &= C(n, M, \eta, p, \varepsilon, r) \end{aligned}$$

where  $C_\varepsilon > 0$  is a universal constant depending only on  $\varepsilon$ . Plugging this in (5.4), we arrive at the stated  $\ell_\infty$  bound in the Theorem. The lower bound on the success probability follows readily via a union bound on the events stated in Lemmas 2, 7, 8.

### 5.3 Proof of Theorem 3

We assume w.l.o.g that  $\pi$  is identity so that  $\pi(i) = i$ . For any  $i \in [n]$ , we have for all rankings  $\tilde{\pi}$  which are consistent with  $\tilde{u}_2$  that

$$|i - \tilde{\pi}(i)| = \sum_{j>i} \mathbf{1}_{\{\tilde{\pi}(j) < \tilde{\pi}(i)\}} + \sum_{j<i} \mathbf{1}_{\{\tilde{\pi}(j) > \tilde{\pi}(i)\}} \leq \sum_{j>i} \mathbf{1}_{\{\tilde{u}_{2,j} \geq \tilde{u}_{2,i}\}} + \sum_{j<i} \mathbf{1}_{\{\tilde{u}_{2,j} \leq \tilde{u}_{2,i}\}},$$

leading to the bound

$$\|\pi - \tilde{\pi}\|_\infty = \max_i |i - \tilde{\pi}(i)| \leq \sum_{j>i} \mathbf{1}_{\{\tilde{u}_{2,j} \geq \tilde{u}_{2,i}\}} + \sum_{j<i} \mathbf{1}_{\{\tilde{u}_{2,j} \leq \tilde{u}_{2,i}\}}.$$

Now for a given  $i' \in [n]$ , we can decompose  $\tilde{u}_{2,i'} - \tilde{u}_{2,j}$  as

$$\tilde{u}_{2,i'} - \tilde{u}_{2,j} = (\tilde{u}_{2,i'} - u_{2,i'}) + (u_{2,i'} - u_{2,j}) + (u_{2,j} - \tilde{u}_{2,j}),$$

which in turn implies  $\tilde{u}_{2,i'} - \tilde{u}_{2,j} \leq (u_{2,i'} - u_{2,j}) + 2\|\tilde{u}_2 - u_2\|_\infty$ . Hence if  $j$  is such that  $i' > j$ , then  $\|\tilde{u}_2 - u_2\|_\infty \leq \frac{|u_{2,i'} - u_{2,j}|}{2}$  implies

$$\tilde{u}_{2,i'} - \tilde{u}_{2,j} \leq \underbrace{(u_{2,i'} - u_{2,j})}_{\leq 0} + |u_{2,i'} - u_{2,j}| = 0.$$

Using this, we obtain

$$\begin{aligned} \sum_{j < i'} \mathbf{1}_{\{\tilde{u}_{2,j} < \tilde{u}_{2,i'}\}} &\leq \sum_{j < i'} \mathbf{1}_{\left\{\|\tilde{u}_2 - u_2\|_\infty > \frac{|u_{2,i'} - u_{2,j}|}{2}\right\}} \\ &\leq \sum_{j < i'} \mathbf{1}_{\left\{\gamma(n, M, \eta, p, \varepsilon, r) > \frac{|u_{2,i'} - u_{2,j}|}{2}\right\}} \quad (\text{since } \|\tilde{u}_2 - u_2\|_\infty \leq \gamma(n, M, \eta, p, \varepsilon, r)) \\ &= \sum_{j < i'} \mathbf{1}_{\{2\gamma(n, M, \eta, p, \varepsilon, r)\|r - \alpha e\|_2 > |r_{i'} - r_j|\}} \\ &= \left\lfloor \frac{2\gamma(n, M, \eta, p, \varepsilon, r)\|r - \alpha e\|_2}{\rho} \right\rfloor \leq \frac{2\gamma(n, M, \eta, p, \varepsilon, r)\|r - \alpha e\|_2}{\rho}. \end{aligned}$$

The same bound holds for  $\sum_{j > i'} \mathbf{1}_{\{\tilde{u}_{2,j} > \tilde{u}_{2,i'}\}}$  and hence the statement of the Theorem follows.

#### 5.4 Proof of Theorem 4

Recall that  $\tilde{w} = \frac{\hat{\sigma}_1}{\eta p \sqrt{n}} \tilde{u}$ . Let us denote

$$w = \frac{\sigma}{\eta p \sqrt{n}} \beta u_2 = \|r - \alpha e\|_2 \beta u_2 = (r - \alpha e) \beta.$$

For any norm  $\|\cdot\|$  on  $\mathbb{R}^n$ , we have by triangle inequality that

$$\|\hat{r} - w\| \leq \|\tilde{w} - w\| + \left| \frac{e^T \tilde{w}}{n} \right| \|e\|.$$

Since  $e^T w = 0$ , therefore we obtain via Cauchy-Schwarz that

$$\left| \frac{e^T \tilde{w}}{n} \right| = \left| \frac{e^T (\tilde{w} - w)}{n} \right| \leq \frac{\|\tilde{w} - w\|_2}{\sqrt{n}}.$$

Hence we have that

$$\|\hat{r} - w\|_2 \leq 2\|\tilde{w} - w\|_2 \quad \text{and} \quad \|\hat{r} - w\|_\infty \leq \|\tilde{w} - w\|_\infty + \frac{\|\tilde{w} - w\|_2}{\sqrt{n}} \leq 2\|\tilde{w} - w\|_\infty.$$

It remains to bound  $\|\tilde{w} - w\|_2, \|\tilde{w} - w\|_\infty$ . Our starting point will be

$$\begin{aligned} \|\tilde{w} - w\| &= \frac{1}{\eta p \sqrt{n}} \|\hat{\sigma}_1 \tilde{u}_2 - \sigma_1 \beta u_2\| \\ &= \frac{1}{\eta p \sqrt{n}} \|(\hat{\sigma}_1 - \sigma_1) \tilde{u}_2 + \sigma_1 (\tilde{u}_2 - \beta u_2)\| \\ \Rightarrow \|\tilde{w} - w\| &\leq \frac{1}{\eta p \sqrt{n}} (|\hat{\sigma}_1 - \sigma_1| \|\tilde{u}_2\| + \sigma_1 \|\tilde{u}_2 - \beta u_2\|) \\ &\leq \frac{1}{\eta p \sqrt{n}} (\Delta \|\tilde{u}_2\| + \eta p \|r - \alpha e\|_2 \sqrt{n} \|\tilde{u}_2 - \beta u_2\|). \end{aligned} \tag{5.10}$$

1. **Bounding**  $\|\tilde{w} - w\|_2$ . From (5.10), we have that

$$\|\tilde{w} - w\|_2 \leq \frac{1}{\eta p \sqrt{n}} (\Delta + \eta p \|r - \alpha e\|_2 \sqrt{n}) \|\tilde{u}_2 - \beta u_2\|_2$$

Plugging the expression for  $\Delta$  from Lemma 2, and the bound on  $\|\tilde{u}_2 - \beta u_2\|_2$  from Theorem 1, we readily obtain the bound in part (i) of the Theorem.

2. **Bounding**  $\|\tilde{w} - w\|_\infty$ . From (5.10), we have that

$$\begin{aligned} \|\tilde{w} - w\|_\infty &\leq \frac{1}{\eta p \sqrt{n}} [\Delta \|\tilde{u}_2\|_\infty + \sigma_1 \|\tilde{u}_2 - \beta u_2\|_\infty] \\ &\leq \frac{1}{\eta p \sqrt{n}} [\Delta (\|\tilde{u}_2 - \beta u_2\|_\infty + \|u_2\|_\infty) + \sigma_1 \|\tilde{u}_2 - \beta u_2\|_\infty] \\ &= \frac{1}{\eta p \sqrt{n}} \left[ (\sigma_1 + \Delta) \|\tilde{u}_2 - \beta u_2\|_\infty + \Delta \frac{(M - \alpha)}{\|r - \alpha e\|_2} \right]. \end{aligned} \quad (5.11)$$

Plugging the expression for  $\sigma_1$  (from Lemma 1) and  $\Delta$  (from Lemma 2), followed by some simplification, we arrive at the bound in part (ii) of the Theorem.

## 5.5 Proof of Theorem 5

The outline is similar to the proof of Theorem 1 with some technical changes. We first note that the SVD of  $(\mathbb{E}[\bar{D}])^{-1/2} \mathbb{E}[H] (\mathbb{E}[\bar{D}])^{-1/2}$  is given by  $\sigma(u_1 v_1^T + u_2 v_2^T)$  where

$$\begin{aligned} \sigma &= \eta p \|(\mathbb{E}[\bar{D}])^{-1/2} (r - \alpha e)\|_2 \|(\mathbb{E}[\bar{D}])^{-1/2} e\|_2, \quad v_2 = u_1 = \frac{(\mathbb{E}[\bar{D}])^{-1/2} e}{\|(\mathbb{E}[\bar{D}])^{-1/2} e\|_2}, \\ v_1 = -u_2 &= -\frac{(\mathbb{E}[\bar{D}])^{-1/2} (r - \alpha e)}{\|(\mathbb{E}[\bar{D}])^{-1/2} (r - \alpha e)\|_2}, \end{aligned} \quad (5.12)$$

and  $\alpha = \frac{r^T (\mathbb{E}[\bar{D}])^{-1} e}{e^T (\mathbb{E}[\bar{D}])^{-1} e}$ . This is verified easily by proceeding as in the proof of Lemma 1. One can also readily see that  $\sigma_{\min} \leq \sigma \leq \sigma_{\max}$ . Now let us write

$$\bar{D}^{-1/2} H \bar{D}^{-1/2} = (\mathbb{E}[\bar{D}])^{-1/2} \mathbb{E}[H] (\mathbb{E}[\bar{D}])^{-1/2} + \tilde{Z}$$

where  $\tilde{Z} = \bar{D}^{-1/2} (H - \mathbb{E}[H]) \bar{D}^{-1/2} + \bar{D}^{-1/2} \mathbb{E}[H] \bar{D}^{-1/2} - (\mathbb{E}[\bar{D}])^{-1/2} \mathbb{E}[H] (\mathbb{E}[\bar{D}])^{-1/2}$ . In order to bound  $\|\tilde{Z}\|_2$ , we will first need to establish the concentration of  $\bar{D}$  around  $\mathbb{E}[\bar{D}]$ .

**Lemma 9.** Denote  $A(\eta, M) = \eta M^2 + (1 - \eta) \frac{M}{2}$ . If  $p \geq \frac{M^2}{9A(\eta, M)} \frac{\log n}{n}$  then,

$$\mathbb{P}(\|\bar{D} - \mathbb{E}[\bar{D}]\|_2 \geq 2(\sqrt{2} + 1) \sqrt{A(\eta, M) n p \log n}) \leq 2/n.$$

The proof is deferred to Appendix E. Conditioned on the event in Lemma 9, we have that

$$\bar{D}_{ii} \in [\mathbb{E}[\bar{D}_{ii}] \pm 2(\sqrt{2} + 1) \sqrt{A(\eta, M) n p \log n}], \quad \forall i = 1, \dots, n.$$

In particular, if  $2(\sqrt{2} + 1) \sqrt{A(\eta, M) n p \log n} \leq \frac{1}{2} \min_i \mathbb{E}[\bar{D}_{ii}] = \frac{p \lambda_{\min}}{2}$ , then

$$\min_i \bar{D}_{ii} \geq p \lambda_{\min} / 2. \quad (5.13)$$

Using triangle inequality, we obtain the bound

$$\|\tilde{Z}\|_2 \leq \underbrace{\|\bar{D}^{-1/2} (H - \mathbb{E}[H]) \bar{D}^{-1/2}\|_2}_{\tilde{Z}_1} + \underbrace{\|\bar{D}^{-1/2} \mathbb{E}[H] \bar{D}^{-1/2} - (\mathbb{E}[\bar{D}])^{-1/2} \mathbb{E}[H] (\mathbb{E}[\bar{D}])^{-1/2}\|_2}_{\tilde{Z}_2}.$$

From sub-multiplicativity of the spectral norm and also (5.13), we have that

$$\|\tilde{Z}_1\|_2 \leq \frac{\|H - \mathbb{E}[H]\|_2}{\min_i \bar{D}_{ii}} \leq \frac{2\Delta}{p\lambda_{\min}}. \quad (5.14)$$

Here, we used the bound  $\|H - \mathbb{E}[H]\|_2 \leq \Delta$  from Lemma 2. In order to bound  $\|\tilde{Z}_2\|_2$ , we add and subtract  $\mathbb{E}[\bar{D}]$  from  $\bar{D}$  and apply triangle inequality. This yields

$$\begin{aligned} \|\tilde{Z}_2\|_2 &\leq \|(\bar{D}^{-1/2} - (\mathbb{E}[\bar{D}])^{-1/2})\mathbb{E}[H](\bar{D}^{-1/2} - (\mathbb{E}[\bar{D}])^{-1/2})\|_2 \\ &\quad + \|(\bar{D}^{-1/2} - (\mathbb{E}[\bar{D}])^{-1/2})\mathbb{E}[H]\bar{D}^{-1/2}\|_2 + \|\bar{D}^{-1/2}\mathbb{E}[H](\bar{D}^{-1/2} - (\mathbb{E}[\bar{D}])^{-1/2})\|_2 \\ &\leq \|\bar{D}^{-1/2} - (\mathbb{E}[\bar{D}])^{-1/2}\|_2^2 \|\mathbb{E}[H]\|_2 + 2\|\bar{D}^{-1/2} - (\mathbb{E}[\bar{D}])^{-1/2}\|_2 \|\mathbb{E}[H]\|_2 \|\bar{D}^{-1/2}\|_2. \end{aligned} \quad (5.15)$$

Note that

$$\begin{aligned} \|(\bar{D}^{-1/2} - (\mathbb{E}[\bar{D}])^{-1/2})\|_2 &= \|\bar{D}^{-1/2}(\bar{D}^{1/2} - (\mathbb{E}[\bar{D}])^{1/2})(\mathbb{E}[\bar{D}])^{-1/2}\|_2 \\ &\leq \|\bar{D}^{-1/2}\|_2 \|(\bar{D}^{1/2} - (\mathbb{E}[\bar{D}])^{1/2})\|_2 \|(\mathbb{E}[\bar{D}])^{-1/2}\|_2. \end{aligned}$$

Moreover, we can bound  $\|(\bar{D}^{1/2} - (\mathbb{E}[\bar{D}])^{1/2})\|_2 \leq \|\bar{D} - \mathbb{E}[\bar{D}]\|_2^{1/2}$  since  $\bar{D}, \mathbb{E}[\bar{D}] \succ 0$  and  $(\cdot)^{1/2}$  is operator monotone (see [5, Theorem X.1.1]). Using  $\|\bar{D}^{-1/2}\|_2 \leq \sqrt{\frac{2}{p\lambda_{\min}}}$ ,  $\|(\mathbb{E}[\bar{D}])^{-1/2}\|_2 \leq \sqrt{\frac{1}{p\lambda_{\min}}}$  and the bound from Lemma 9 in (5.15), we get

$$\|\tilde{Z}_2\|_2 \leq C_1^2(\eta, M) \frac{(np \log n)^{1/4}}{p\lambda_{\min}} \|\mathbb{E}[H]\|_2 + \frac{2\sqrt{2}C_1(\eta, M) \|\mathbb{E}[H]\|_2 (np \log n)^{1/4}}{(p\lambda_{\min})^{3/2}}, \quad (5.16)$$

where  $C_1(\eta, M) = 4A^{1/4}(\eta, M)$ . From (5.14), (5.16), we get

$$\begin{aligned} \|\tilde{Z}\|_2 &\leq \frac{2\Delta}{p\lambda_{\min}} + \frac{C_1(\eta, M) \|\mathbb{E}[H]\|_2 (np \log n)^{1/4}}{(p\lambda_{\min})^{3/2}} \left[ \frac{C_1(\eta, M) (np \log n)^{1/4}}{\sqrt{p\lambda_{\min}}} + 2\sqrt{2} \right] \\ &\leq 16M \sqrt{\frac{5}{3} pn} \frac{(2 + \varepsilon)}{p\lambda_{\min}} + \frac{C_1(\eta, M) (np \log n)^{1/4} \sigma}{(p\lambda_{\min})^{3/2}} \left( \frac{C_1(\eta, M) (np \log n)^{1/4}}{\sqrt{p\lambda_{\min}}} + 2\sqrt{2} \right) \quad (\text{using Lemma 2, (5.12)}) \\ &= \tilde{\Delta}. \end{aligned}$$

Let  $\hat{U} = [\hat{u}_1 \ \hat{u}_2]$  denote the top two left singular vectors of  $\bar{D}^{-1/2} H \bar{D}^{-1/2}$ . Then from Wedin's bound, we know that

$$\tilde{\Delta} < \sigma \Rightarrow \|(I - \hat{U}\hat{U}^T)U\|_2 \leq \frac{\tilde{\Delta}}{\sigma - \tilde{\Delta}} = \delta. \quad (5.17)$$

Finally, Lemma 4 can be used here unchanged. Hence if  $\delta \leq 1/2 \Leftrightarrow \tilde{\Delta} \leq \sigma/3$ , then there exists  $\beta \in \{-1, 1\}$  such that  $\|\tilde{u}_2 - \beta u_2\|_2^2 \leq 10\delta$ . Since  $\tilde{\Delta} \leq \sigma_{\min}/3 \Leftrightarrow \delta \leq \frac{3\tilde{\Delta}}{2\sigma_{\min}} \Rightarrow \tilde{\Delta} \leq \sigma/3$ , we obtain the stated bound on  $\|\tilde{u}_2 - \beta u_2\|_2^2$ . The lower bound on the success probability follows by applying the union bound to the events in Lemmas 2, 9.

## 5.6 Proof of Theorem 6

Recall that  $\tilde{w} = \frac{\hat{\sigma}_1}{\eta p \|\bar{D}^{-1/2} e\|_2} \bar{D}^{1/2} \tilde{u}_2$  and  $\hat{r} = \tilde{w} - \frac{e^T \tilde{w}}{n} e$ . Let us denote

$$\begin{aligned} w &= \frac{\sigma}{\eta p \|\mathbb{E}[\bar{D}]\|_2} (\mathbb{E}[\bar{D}])^{1/2} \beta u_2 = (r - \alpha e) \beta, \\ w' &= w - \frac{e^T w}{n} e = \beta(r - (\alpha + \alpha') e), \end{aligned}$$

where  $\alpha' = \frac{e^T (r - \alpha e)}{n}$ . Then using triangle inequality, we can bound  $\|\hat{r} - w'\|_2$  as

$$\|\hat{r} - w'\|_2 \leq \|w - \tilde{w}\|_2 + \left| \frac{e^T (\tilde{w} - w)}{n} \right| \|e\|_2 \leq 2\|w - \tilde{w}\|_2.$$

In order to bound  $\|w - \tilde{w}\|_2$ , note that we can write  $\tilde{w} - w = \frac{1}{np}(a_1 + a_2 + a_3 + a_4)$  where

$$\begin{aligned} a_1 &= \frac{\widehat{\sigma}_1}{\|\bar{D}^{-1/2}e\|_2} \bar{D}^{1/2}\tilde{u}_2 - \frac{\widehat{\sigma}_1}{\|\bar{D}^{-1/2}e\|_2} (\mathbb{E}[\bar{D}])^{1/2}\tilde{u}_2, \\ a_2 &= \frac{\widehat{\sigma}_1}{\|\bar{D}^{-1/2}e\|_2} (\mathbb{E}[\bar{D}])^{1/2}\tilde{u}_2 - \frac{\widehat{\sigma}_1}{\|(\mathbb{E}[\bar{D}])^{-1/2}e\|_2} (\mathbb{E}[\bar{D}])^{1/2}\tilde{u}_2, \\ a_3 &= \frac{\widehat{\sigma}_1}{\|(\mathbb{E}[\bar{D}])^{-1/2}e\|_2} (\mathbb{E}[\bar{D}])^{1/2}\tilde{u}_2 - \frac{\sigma}{\|(\mathbb{E}[\bar{D}])^{-1/2}e\|_2} (\mathbb{E}[\bar{D}])^{1/2}\tilde{u}_2, \\ a_4 &= \frac{\sigma}{\|(\mathbb{E}[\bar{D}])^{-1/2}e\|_2} (\mathbb{E}[\bar{D}])^{1/2}\tilde{u}_2 - \frac{\sigma}{\|(\mathbb{E}[\bar{D}])^{-1/2}e\|_2} (\mathbb{E}[\bar{D}])^{1/2}\beta u_2. \end{aligned}$$

Since  $\|\tilde{w} - w\|_2 \leq \frac{1}{np} \sum_{i=1}^4 \|a_i\|_2$  we will now bound  $\|a_i\|_2$  for each  $i$ . Before proceeding, recall from the proof of Theorem 5.5 that  $\min_i \bar{D}_{ii} \geq \frac{p\lambda_{\min}}{2}$ . We also have

$$\max_i \bar{D}_{ii} \leq p\lambda_{\max} + \frac{p\lambda_{\min}}{2} \leq \frac{3p\lambda_{\max}}{2},$$

and  $\widehat{\sigma}_1 \leq \sigma + \tilde{\Delta}$  where the latter is due to Weyl's inequality.

**Bounding  $\|a_1\|_2$ .** Since  $\sigma \leq \sigma_{\max}$ , we have that

$$\|a_1\|_2 \leq \frac{\tilde{\Delta} + \sigma_{\max}}{\|\bar{D}^{-1/2}e\|_2} \|\bar{D}^{1/2} - (\mathbb{E}[\bar{D}])^{1/2}\|_2 \leq \frac{\tilde{\Delta} + \sigma_{\max}}{\|\bar{D}^{-1/2}e\|_2} \|\bar{D} - \mathbb{E}[\bar{D}]\|_2^{1/2} \quad (5.18)$$

where the second inequality is due to  $\bar{D}, \mathbb{E}[\bar{D}] \succ 0$  and since  $(\cdot)^{1/2}$  is operator monotone (see [5, Theorem X.1.1]). Using the bound  $\|\bar{D}^{-1/2}e\|_2 \geq \sqrt{\frac{2n}{3p\lambda_{\max}}}$ , along with the bound for  $\|\bar{D} - \mathbb{E}[\bar{D}]\|_2$  (from Lemma 9) in (5.18), we arrive at

$$\|a_1\|_2 \leq \sqrt{\frac{3}{2n}} p\lambda_{\max} (\sigma_{\max} + \tilde{\Delta}) (2\sqrt{2} + 1)^{1/2} (A(\eta, M) np \log n)^{1/4}.$$

**Bounding  $\|a_2\|_2$ .** We begin by noting that

$$\|a_2\|_2 = \widehat{\sigma}_1 \left| \frac{1}{\|\bar{D}^{-1/2}e\|_2} - \frac{1}{\|(\mathbb{E}[\bar{D}])^{-1/2}e\|_2} \right| \|(\mathbb{E}[\bar{D}])^{1/2}\tilde{u}_2\|_2 \leq (\sigma_{\max} + \tilde{\Delta}) \left| \frac{\|(\mathbb{E}[\bar{D}])^{-1/2}e\|_2 - \|\bar{D}^{-1/2}e\|_2}{\|\bar{D}^{-1/2}e\|_2 \|(\mathbb{E}[\bar{D}])^{-1/2}e\|_2} \right|.$$

Since  $\|\bar{D}^{-1/2}e\|_2 \geq \sqrt{\frac{2n}{3p\lambda_{\max}}}$ ,  $\|(\mathbb{E}[\bar{D}])^{-1/2}e\|_2 \geq \frac{\sqrt{n}}{\sqrt{p\lambda_{\max}}}$ , and

$$\begin{aligned} \left| \|(\mathbb{E}[\bar{D}])^{-1/2}e\|_2 - \|\bar{D}^{-1/2}e\|_2 \right| &\leq \|(\bar{D}^{-1/2} - (\mathbb{E}[\bar{D}])^{-1/2})e\|_2 \\ &\leq \|\bar{D}^{-1/2}\|_2 \|\bar{D}^{1/2} - (\mathbb{E}[\bar{D}])^{1/2}\|_2 \|(\mathbb{E}[\bar{D}])^{-1/2}\|_2 \sqrt{n} \\ &\leq \frac{\sqrt{2}}{p\lambda_{\min}} (2\sqrt{2} + 1)^{1/2} (A(\eta, M) np \log n)^{1/4}, \end{aligned}$$

we can bound  $\|a_2\|_2$  as

$$\|a_2\|_2 \leq \sqrt{\frac{3}{n}} (\sigma_{\max} + \tilde{\Delta}) \frac{\lambda_{\max}}{\lambda_{\min}} (2\sqrt{2} + 1)^{1/2} (A(\eta, M) np \log n)^{1/4}.$$

**Bounding  $\|a_3\|_2$ .** This is easily achieved by noting that

$$\|a_3\|_2 = \frac{|\widehat{\sigma}_1 - \sigma|}{\|(\mathbb{E}[\bar{D}])^{-1/2}e\|_2} \|(\mathbb{E}[\bar{D}])^{1/2}\tilde{u}_2\|_2 \leq \frac{\tilde{\Delta} \sqrt{p\lambda_{\max}}}{\sqrt{n} / \sqrt{p\lambda_{\max}}} = \frac{\tilde{\Delta} p \lambda_{\max}}{\sqrt{n}}.$$

**Bounding  $\|a_4\|_2$ .** This is also easily achieved by noting that

$$\begin{aligned} \|a_4\|_2 &\leq \frac{\sigma}{\|(\mathbb{E}[\bar{D}])^{-1/2}e\|_2} \|(\mathbb{E}[\bar{D}])^{1/2}\|_2 \|\tilde{u}_2 - \beta u_2\|_2 \\ &\leq \frac{\sigma_{\max} \sqrt{p\lambda_{\max}}}{\sqrt{n}/\sqrt{p\lambda_{\max}}} \left( \sqrt{15} \sqrt{\frac{\tilde{\Delta}}{\sigma_{\min}}} \right) \\ &\leq \sqrt{\frac{15\tilde{\Delta}}{\sigma_{\min}n}} \sigma_{\max} p \lambda_{\max}. \end{aligned}$$

The stated bound now follows from  $\|\hat{r} - \beta(r - (\alpha + \alpha')e)\|_2 \leq \frac{2}{\eta p} \sum_{i=1}^4 \|a_i\|_2$ .

## 6 Matrix completion as a preprocessing step for ranking

*Low-rank matrix completion* is the problem of recovering the missing entries of a low-rank matrix given a subset of its entries. This line of research started with the results in [12] and [13] which showed that given a rank- $r$  matrix  $C$  of size  $n_1 \times n_2$  (with  $n = \max\{n_1, n_2\}$ ), one can recover it by observing only  $O(nr \text{polylog}(n))$  randomly selected entries (under some assumptions on  $C$ ) via a simple convex optimization algorithm. This was partly inspired by similar approaches used previously in the compressed sensing literature [10, 11]. This problem has received tremendous attention in the last decade, due to a number of applications such as in phase retrieval [8], computer vision [49, 43] and sensor network localization [17] to name a few.

More formally, let us assume for simplicity that  $r = O(1)$  and the SVD of  $C = \sum_{i \in [r]} \sigma_i u_i v_i^T$  satisfies

$$\|u_i\|_{\infty} \leq \sqrt{\frac{\mu}{n_1}}, \quad \|v_i\|_{\infty} \leq \sqrt{\frac{\mu}{n_2}}; \quad \forall i, j, \quad (6.1)$$

for some  $\mu \geq 1$ . Here  $\mu$  is a measure of how spread out the entries of  $u_i, v_i$  are - the smaller the value of  $\mu$ , the better. Say we observe  $m$  entries of  $C$  on a subset  $\Omega \subset [n_1] \times [n_2]$ , sampled uniformly at random. Denoting  $P_{\Omega} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^{n_1 \times n_2}$  to be the projection operator on  $\Omega$ , it was shown in [13] that the solution  $\hat{C}$  of

$$\min \|X\|_* \quad \text{s.t.} \quad P_{\Omega}(X) = P_{\Omega}(C) \quad (6.2)$$

equals  $C$  with high probability, provided  $m = \Omega(\mu^4 n \log^2 n)$ . The nuclear norm minimization problem (6.2) is a SDP and hence can be solved in polynomial time using, for eg., interior point methods. In fact, the result holds under a different Bernoulli sampling model too wherein each entry of  $M$  is observed independently with a certain probability [12, Section 4.1]. Moreover, say that the observations are noisy, i.e, we observe

$$Y_{i,j} = C_{i,j} + Z_{i,j}; \quad (i, j) \in \Omega \quad \Leftrightarrow \quad P_{\Omega}(Y) = P_{\Omega}(C) + P_{\Omega}(Z)$$

where  $Z$  is the noise matrix. Say  $\|P_{\Omega}(Z)\|_F \leq \delta$ . Then, it was shown in [9] (under certain additional conditions) that the solution of

$$\min \|X\|_* \quad \text{s.t.} \quad \|P_{\Omega}(X) - P_{\Omega}(Y)\|_F \leq \delta \quad (6.3)$$

is stable, i.e., the estimation error  $\|\hat{C} - C\|_F$  is bounded by a term proportional to  $\delta$ .

In our setting, we observe (noisy versions of) a subset of the entries of the  $n \times n$  matrix  $C = r e^T - e r^T$  where each off-diagonal entry of  $C$  is revealed with probability  $p$ . Since  $C$  has rank 2, it is natural to consider estimating  $C$  via matrix completion as a preprocessing step, and then subsequently applying SVD-RS or SVD-NRS on the obtained estimate of  $C$  for recovering the underlying ranks and scores. In order to understand the sample complexity for successful matrix completion, we need to express  $\mu$  defined in (6.1) in terms of the score vector  $r$ . To this end, we see from Lemma 1 that

$$\|u_2\|_{\infty}, \|v_1\|_{\infty} = \frac{\|r - \alpha e\|_{\infty}}{\|r - \alpha e\|_2} \leq \frac{M - \alpha}{\|r - \alpha e\|_2}$$

and  $\|u_1\|_\infty, \|v_2\|_\infty = 1/\sqrt{n}$ . Hence it follows that

$$\mu = \max \left\{ \frac{(M - \alpha)\sqrt{n}}{\|r - \alpha e\|_2}, 1 \right\}.$$

In order to get a sense of the scaling of  $\mu$ , consider the setting  $r_i = i$ . As seen before, we then have  $\|r - \alpha e\|_2 = \Theta(n^{3/2})$ ,  $M = n$ , and  $\alpha = \Theta(n)$ . This in turn implies that  $\mu = \Theta(1)$  which is the ideal scenario for matrix completion. In our numerical experiments, we rely on the TFOCS software library [3], that allows for construction of first-order methods for a variety of convex optimization problems [4]. In our implementation, we also set the diagonal entries of  $X$  to be equal to zero since the same is true for  $C$ . We do not enforce the skew-symmetry constraints  $X_{ij} = -X_{ji}$  for  $(i, j) \in \Omega$  so the solution  $\widehat{C}$  to  $C$  is not guaranteed to be skew symmetric. Instead, we output  $\frac{\widehat{C} - \widehat{C}^T}{2}$  as the final (skew-symmetric) estimate obtained from the preprocessing step. Note that since  $C^T = -C$ , we have

$$\left\| \frac{\widehat{C} - \widehat{C}^T}{2} - C \right\|_F = \left\| \frac{\widehat{C} - \widehat{C}^T - C + C^T}{2} \right\|_F \leq \|C - \widehat{C}\|_F.$$

Finally, we end with a motivating discussion for the applicability of the matrix completion approach in the setting of ranking from pairwise comparisons. The recent work [50] of Udell and Townsend provides a compelling argument on why big data matrices are approximately low rank. The authors consider a simple generative model for matrices, assuming that each row or column is associated to a (possibly high dimensional) bounded latent variable, with the individual matrix entries being generated by applying a piecewise analytic function to these latent variables. While the initial resulting matrices are typically full rank, the authors show that one can approximate every entry of an  $m \times n$  matrix drawn from the above model to within a fixed absolute error by a low rank matrix whose rank grows as  $O(\log(m + n))$ . In other words, any sufficiently large matrix from such a latent variable model can be approximated by a low rank matrix (up to a small entrywise error). The paradigm that “*nice latent variables models are of log-rank*” ([50]) is also applicable in the ranking setting, where one assumes that the final ranking (or skill) of each player varies smoothly as a function of covariate information, which is typically available in many real word applications. For example, in sport or clinical data, covariates may provide additional information about both the ranked players and the rankers/judges that can overall lead to better aggregated results.

## 7 Numerical Experiments

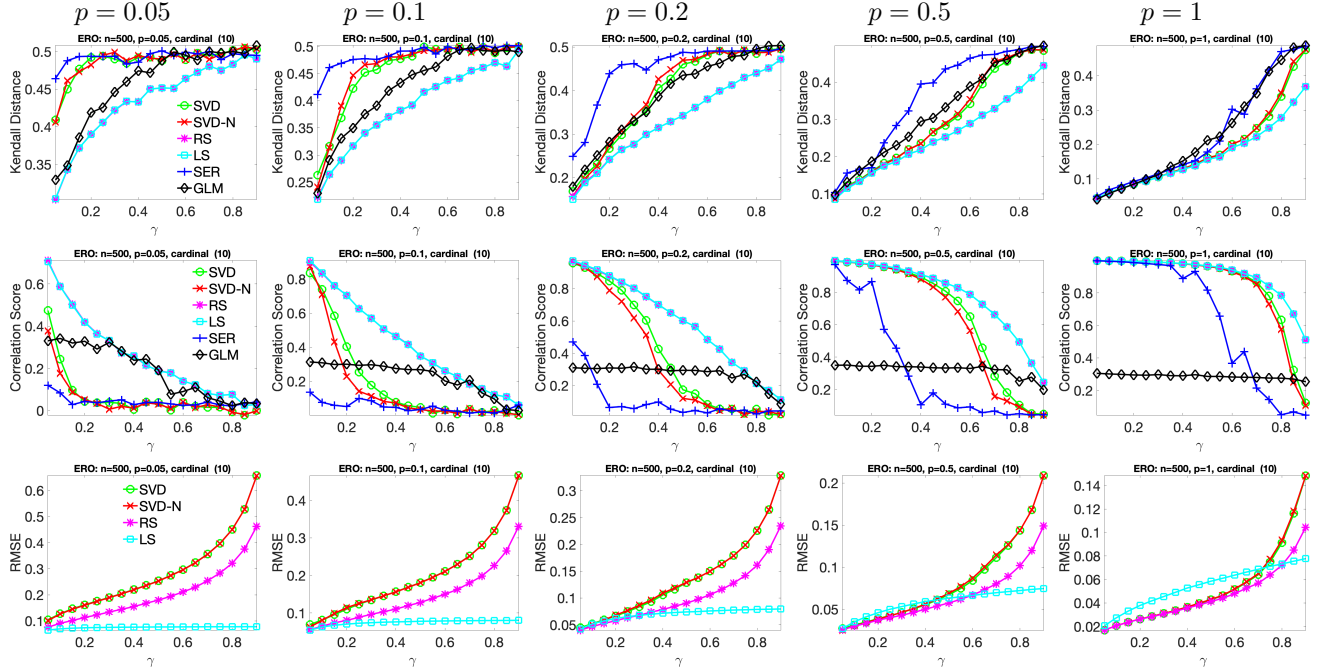
We compare the performance of Algorithm 1 SVD-RS (SVD in the figure legends for brevity) and Algorithm 2 SVD-NRS (SVD-N in the figure legends) with that of a number of other algorithms from the literature, including SERIAL-RANK (SER) along with its extension to the generalized linear model setting (GLM) [24], as well as Row-Sum (RS) and a least-squared based formulation (LS) as considered in [16]. This is done for synthetic data, under the ERO noise model in Section 7.1, and real data in Section 7.2.

### 7.1 Synthetic data

This section details the outcomes of synthetic numerical experiments, under the ERO model, where the measurement graph is Erdős-Rényi and the noise comes in the form of outliers from a uniform distribution.

We consider two sets of experiments, a first one where the strengths  $r_i$  are uniformly distributed in  $[0, 1]$ , and a second one where they are Gamma distributed with shape parameter  $a = 0.5$  and scale parameter  $b = 1$ . This choice of parameters for the Gamma distribution leads to a skewed distribution of player strengths, and subsequently a skewed distribution of node degrees, which is a setting realistic in practice but known to be challenging for spectral methods. Figure 1, respectively Figure 2, pertain to the experiment where the strengths are Gamma distributed without, respectively with, the matrix completion pre-processing step. Similarly, Figure 3, respectively Figure 4, concern the scenario where the strengths are drawn from a uniform distribution, without, respectively with, the matrix completion step.

In the synthetic experiments, we fix the number of nodes ( $n$ ), and vary the edge density ( $p$ ), and noise level ( $\gamma = 1 - \eta$ ). Whenever ground truth is available, we plot in the top row of Figures 1, 2, 3, 4 the



**Figure 1:** Performance statistics in terms of Kendall distance (top row), correlation score (middle row), and RMSE (bottom row), for synthetic data with scores generated by a **Gamma** distribution, **without** the matrix-completion step.

*Kendall Distance* between the recovered strength vector  $\hat{r}$  and the ground truth  $r$ , for different noise levels  $\gamma$ . The Kendall distance counts the number of pairs of candidates that are ranked in different order (flips), in the original ranking versus the recovered one. The middle row of each Figure plots the *Correlation Score*, computed as the Pearson correlation between the ground truth  $r$  and the recovered  $\hat{r}$ . Finally, the bottom row in each Figure plots the *RMSE* error, defined as  $\sqrt{\frac{1}{n} \|r - \hat{r}\|_2}$ , after having having centered  $r$  and  $\hat{r}$ .

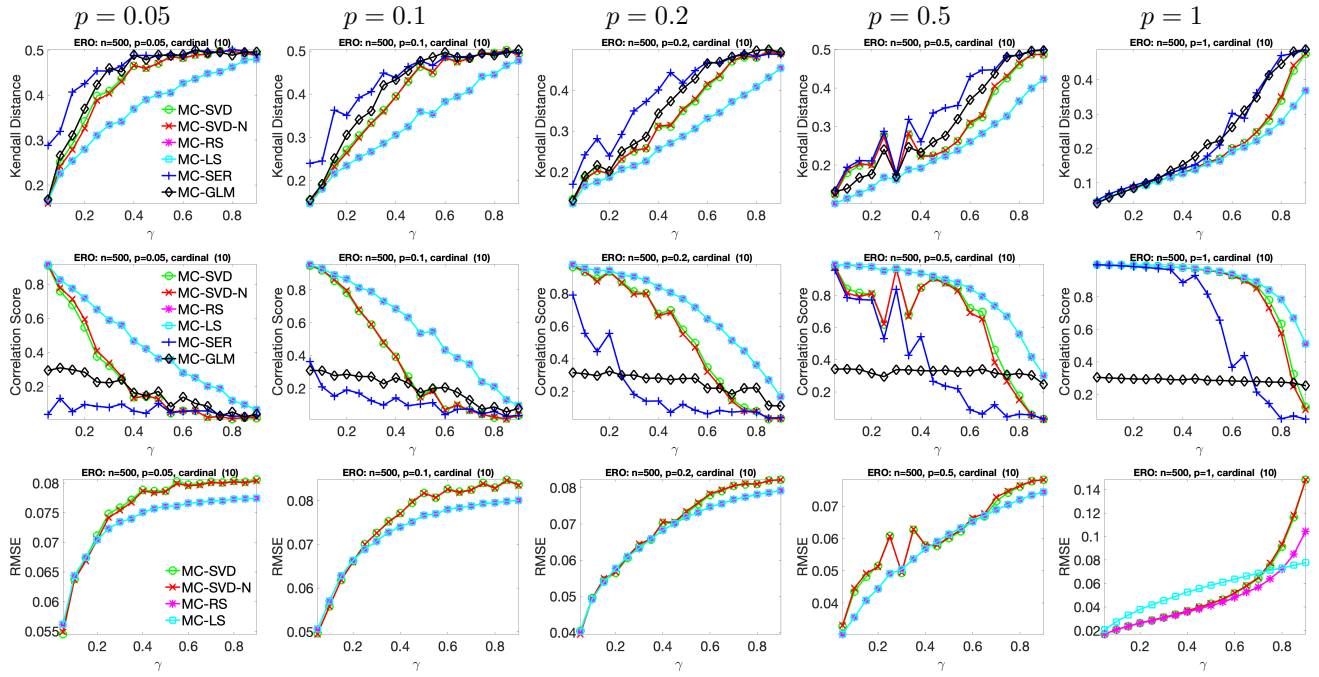
Without matrix completion, the results from Figure 1 show that the two SVD-based approaches on average outperform SER and GLM, and are surpassed in performance by RS and LS. For denser graphs with  $p = \{0.5, 1\}$ , the performance of the SVD-based methods is on par with that of RS and LS up until the very high noise levels, across the three performance metrics. Figure 2 shows the results after the matrix-completion step, which improves the performance of all methods, and maintains the above mentioned relative performance of the methods.

Figure 3 shows the results without the matrix completion step, when the scores are uniformly distributed. For very sparse graphs, RS and LS are the best performers, but the ordering of the next best-performing methods depends on the evaluation metric. In terms of Kendall distance, GLM comes next (though not a scalable approach for large  $n$ ), followed by the SVD methods, and SER. However, in terms of correlation, the SVD methods consistently outperform SER and GLM across all levels of noise. For denser graphs with  $p = \{0.5, 1\}$ , the performance of the SVD-based methods is on par with that of RS and significantly better than LS in terms of RMSE, while when comparing on the Kendall distance, GLM is the best performer followed by all other methods which return almost identical performance, except SER which is a clear underperformer. Similar relative performances are reported in Figure 4 after the matrix completion step, with the difference that, in terms of RMSE, the two SVD methods are the best performers overall, and surpass RS and LS by a small margin.

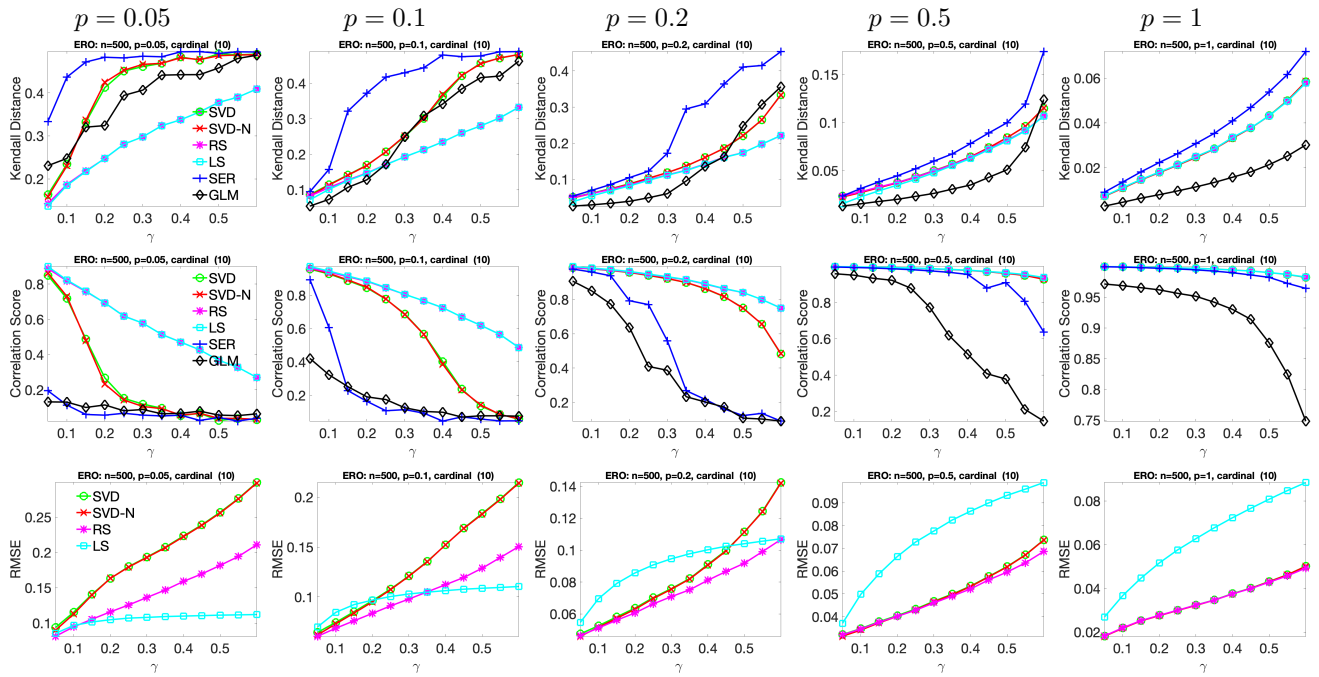
## 7.2 Real data

We apply all ranking algorithms to a variety of networks, and measure outcome by the number of upsets in the final ranking, computed with respect to the given pairwise measurements. For each barplot, we color in red, blue and magenta, respectively, the top three best performing methods (achieving the lowest number

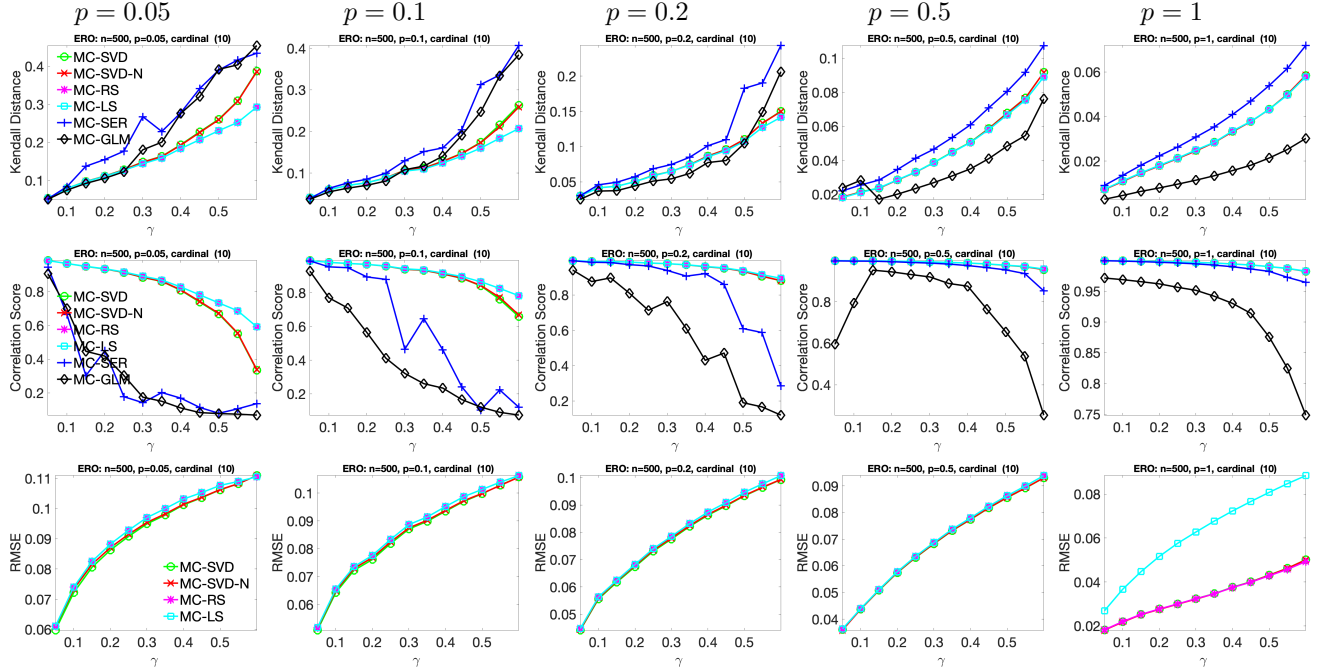




**Figure 2:** Performance statistics in terms of Kendall distance (top row), correlation score (middle row), and RMSE (bottom row), for synthetic data with scores generated by a **Gamma** distribution, **with** the low-rank matrix completion pre-processing step.



**Figure 3:** Performance statistics in terms of Kendall distance (top row), correlation score (middle row), and RMSE (bottom row), for synthetic data with scores generated from a **Uniform** distribution, **without** the matrix-completion step.



**Figure 4:** Performance statistics in terms of Kendall distance (top row), correlation score (middle row), and RMSE (bottom row), for synthetic data with scores generated from a **Uniform** distribution, **with** the low-rank matrix completion pre-processing step.

of upsets). For brevity, in each data sets, we primarily comment on the performance of the two SVD-based methods relative to the other four methods.

**Animal dominance network.** Our first example is a set of two networks of animal dominance among captive monk parakeets [29]. The input matrix  $A$  is such that  $A_{ij}$  denotes the number of aggressive wins by animal  $i$  toward animal  $j$ . The study covers two groups (of 21, respectively, 19 individuals), and spans across four quarters. We follow the same approach as in [20], and only tested the algorithms on Quarters 3 and 4, based on guidance from the behavioral ecologist who collected and analyzed the data. Since the animals did not know each other well at the start of the experiment, the hierarchies were still in the transient process of forming. Figure 5 compares performance in terms of the number of upsets, with and without the matrix completion step. For two out of the four data sets, SVD-N is the best performer, without matrix completion. After matrix completion, SVD ranks second for two of the data sets, while SVD-N ranks first on one data set and comes third on two of the other data sets.

**Faculty hiring networks.** Our next example covers three North American academic hiring networks, that track the flow of academics between universities [14]. The flow is captured in a directed graph with adjacency matrix  $A$ , such that  $A_{ij}$  is the number of faculty at university  $j$  who received their doctorate from university  $i$ . Figure 6 shows the number of upsets attained by each algorithm, for three different disciplines: (a) Computer Science ( $n = 206$ ), (b) Business ( $n = 113$ ) and (c) History ( $n = 145$ ), before and after the matrix completion step. Without matrix completion, SVD-N ranks first in History, comes second in the remaining two data sets. After matrix completion, SVD-N ranks second in Business and third in History, while SVD ranks third in Computer Science and Business.

**Microsoft Halo 2 Tournament.** Our third experiment was performed on a real data set of game outcomes collected during the Beta testing period for the Xbox game<sup>4</sup> Halo 2. The graph has a total of  $n = 606$  players, and 6227 head-to-head games. After removing the low-degree nodes, i.e., players who have played less than

<sup>4</sup>Credits for using the Halo 2 Beta data set are given to Microsoft Research Ltd. and Bungie.

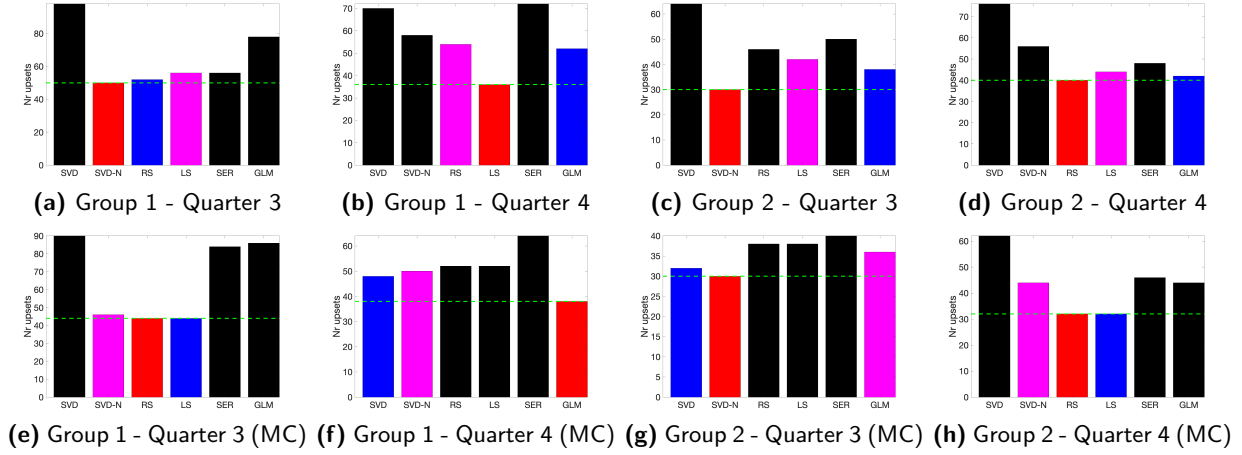


Figure 5: Number of upsets in the animal dominance networks, for two groups in Quarters 3-4.

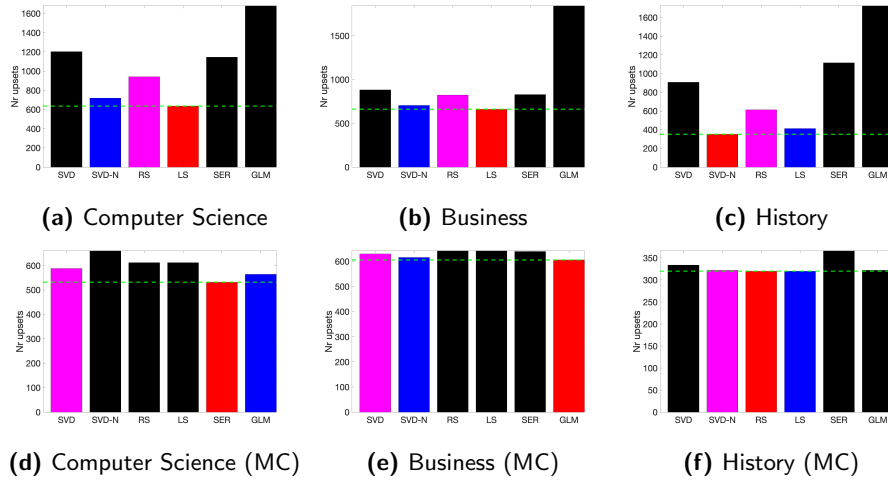


Figure 6: Number of upsets in the faculty hiring networks. Top row: without matrix completion. Bottom row: with the low-rank matrix completion.

3 games, we arrive at a graph with  $n = 535$  nodes and 6109 edges, with an average degree of roughly 23. In this example illustrated in Figure 7, LS achieves the smallest number of upsets, followed by RS and SVD-N. After matrix completion, RS, LS and SER are the top three best performers.

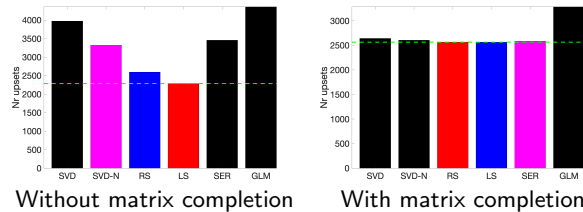


Figure 7: Recovery results for the Microsoft Halo Tournament data set, before and after the low-rank matrix completion pre-processing step.

**Premier League.** For the Premier League data set, we consider the four seasons during 2009-2013 in Figure 8. Without matrix completion, SVD-N ranks first for two seasons and comes second for one season,

while SVD ranks first in one of the seasons. After matrix completion, SVD-N ranks second for two seasons, and SVD ranks first in one of the seasons.

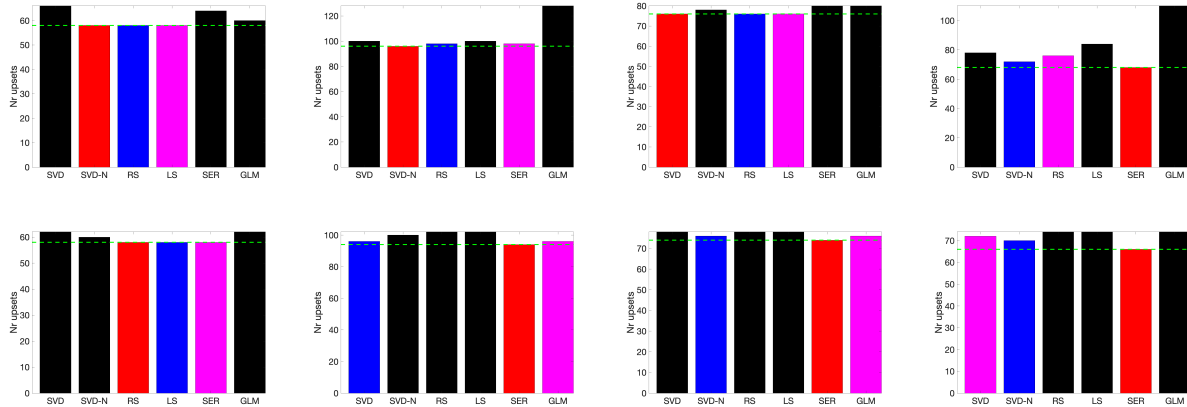
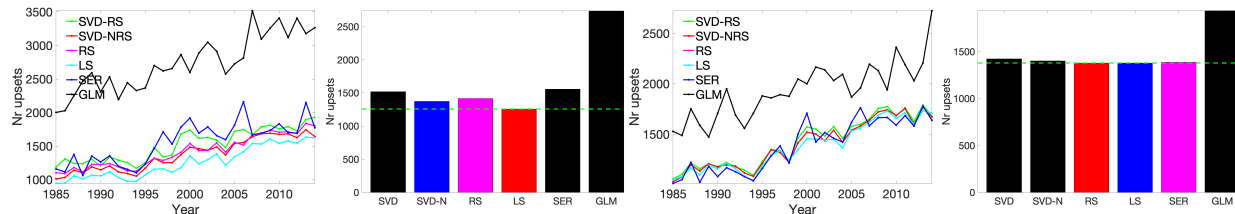


Figure 8: Premier League

**NCAA College Basketball.** Our final real data set comes from sports, and contains the outcomes of NCAA College Basketball matches during the regular season, for the years 1985 - 2014. Figure 9 (a) shows the number of upsets across time (without matrix completion), while plot (b) shows the average over the entire period, for each algorithm. On average, it is typically the case that a pair of teams play each other at most once. We remark that the number of upsets shown in Figure 9 (a) and (c) is ascending since the number of teams grew over the years. For example, during the 1985 season, the league contained 282 teams and 3737 games took place, while in 2014, there were 351 teams and 5362 games. We remark that, on average, the best performer was LS, followed by SVD-N, and that RS, while GLM was by far the worst performer. After matrix completion, the top three best performing methods are RS, LS and SER.



(a) Yearly number of upsets (b) Average over 1985 - 2014 (c) Yearly number of upsets (d) Average over 1985 - 2014 (without Matrix Completion). (without Matrix Completion). (with Matrix Completion). (with Matrix Completion).

Figure 9: Comparison of the algorithms for the NCAA College Basketball data set (1984 - 2014), before and after the matrix completion step.

## 8 Conclusion and future directions

We considered the problems of ranking and time synchronization given a subset of noisy pairwise comparisons, and proposed an SVD-based algorithmic pipeline to solve both tasks.

We analyzed the robustness of SVD-RS in the form of  $\ell_2$  and  $\ell_\infty$  recovery guarantees for the score vector  $r$ , against sampling sparsity and noise perturbation, using tools from matrix perturbation and random matrix theory. The  $\ell_\infty$  analysis of SVD-RS leads to guarantees for rank recovery in terms of the maximum displacement error with respect to the ground truth. We also introduced SVD-NRS, a normalized version of SVD-RS, and provided  $\ell_2$  recovery guarantees for the score vector  $r$ .

We have augmented our theoretical analysis with a comprehensive set of numerical experiments on both synthetic and real data (five different real data sets, which altogether contained 42 distinct problem instances/comparison graphs), showcasing the competitiveness of our approach compared to other algorithms. In particular, SVD-NRS was shown to perform particularly well on many of the real data sets, often returning the best performance in terms of the number of upsets induced by the recovered rankings.

There are several avenues for future work. An interesting direction pertains to extending our analysis to the setting of very sparse graphs, with  $p$  on the order of  $\frac{1}{n}$ , by leveraging recent regularization techniques [31, 35]. Analysis of other, perhaps more realistic, noise models would also be of interest, as well as obtaining theoretical guarantees for the pipeline with low-rank matrix completion as a pre-processing step.

## Acknowledgements

A.A. is at the département d’informatique de IENS, Ecole normale supérieure, UMR CNRS 8548, PSL Research University, 75005 Paris, France, and INRIA, and would like to acknowledge support from the ML and Optimisation joint research initiative with the fonds AXA pour la recherche and Kamet Ventures as well as a Google focused award. A part of this work was done while H.T. was affiliated to The Alan Turing Institute, London, and the School of Mathematics, University of Edinburgh. M.C. and H.T. also acknowledge support from the EPSRC grant EP/N510129/1 at The Alan Turing Institute.

## References

- [1] Arpit Agarwal, Prathamesh Patil, and Shivani Agarwal. Accelerated spectral ranking. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pages 70–79, 2018.
- [2] Afonso S. Bandeira and Ramon van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Ann. Probab.*, 44(4):2479–2506, 07 2016.
- [3] Stephen R Becker, Emmanuel J Candès, and Michael C Grant. TFOCS: Templates for First-Order Conic Solvers.
- [4] Stephen R Becker, Emmanuel J Candès, and Michael C Grant. Templates for convex cone problems with applications to sparse signal recovery. *Mathematical programming computation*, 3(3):165, 2011.
- [5] R. Bhatia. *Matrix Analysis*. Springer New York, 1996.
- [6] Stéphane Boucheron, Gabor Lugosi, and Pascal Massart. *Concentration inequalities : a non asymptotic theory of independence*. Oxford University Press, 2013.
- [7] R. A. Bradley and M. E. Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, page 324345, 1952.
- [8] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski. Phase retrieval via matrix completion. *SIAM Journal on Imaging Sciences*, 6(1):199–225, 2013.
- [9] E. J. Candes and Y. Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98(6):925–936, June 2010.
- [10] E. J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, Feb 2006.
- [11] E.J. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406 – 5425, 2006.
- [12] Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717, Apr 2009.

- [13] Emmanuel J. Candès and Terence Tao. The power of convex relaxation: Near-optimal matrix completion. *CoRR*, abs/0903.1476, 2009.
- [14] Aaron Clauset, Samuel Arbesman, and Daniel B Larremore. Systematic inequality and hierarchy in faculty hiring networks. *Science Advances*, 1(1):e1400005, 2015.
- [15] A. H. Copeland. A "reasonable" social welfare function. In *In Seminar on Mathematics in Social Sciences, University of Michigan*, 1951.
- [16] M. Cucuringu. Sync-Rank: Robust Ranking, Constrained Ranking and Rank Aggregation via Eigenvector and Semidefinite Programming Synchronization. *IEEE Transactions on Network Science and Engineering*, 3(1):58–79, 2016.
- [17] Mihai Cucuringu, Yaron Lipman, and Amit Singer. Sensor network localization by eigenvector synchronization over the Euclidean group. *ACM Trans. Sen. Netw.*, 8(3):19:1–19:42, August 2012.
- [18] Onkar Dalal, Srinivasan H Sengemedu, and Subhajit Sanyal. Multi-objective ranking of comments on web. In *Proceedings of the 21st international conference on World Wide Web*, pages 419–428. ACM, 2012.
- [19] Mark A Davenport, Yaniv Plan, Ewout Van Den Berg, and Mary Wootters. 1-bit matrix completion. *Information and Inference: A Journal of the IMA*, 3(3):189–223, 2014.
- [20] Caterina De Bacco, Daniel B. Larremore, and Cristopher Moore. A physical model for efficient ranking in networks. *Science Advances*, 4(7), 2018.
- [21] Justin Eldridge, Mikhail Belkin, and Yusu Wang. Unperturbed: spectral analysis beyond Davis-Kahan. In *ALT*, 2018.
- [22] László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Spectral statistics of Erdős-Rényi graphs I: Local semicircle law. *Ann. Probab.*, 41(3B):2279–2375, 05 2013.
- [23] M. Fanuel and J.A.K. Suykens. Deformed Laplacians and spectral ranking in directed networks. *Applied and Computational Harmonic Analysis*, 2017.
- [24] F. Fogel, A. d’Aspremont, and M. Vojnovic. Serialrank: Spectral ranking using seriation. In *Advances in Neural Information Processing Systems 27*, pages 900–908, 2014.
- [25] Fajwel Fogel, Alexandre d’Aspremont, and Milan Vojnovic. Spectral ranking using seriation. *Journal of Machine Learning Research*, 17(88):1–45, 2016.
- [26] A. Giridhar and P. R. Kumar. Distributed clock synchronization over wireless networks: Algorithms and analysis. In *45th IEEE Conference on Decision and Control*, pages 4915–4920, 2006.
- [27] David F Gleich and Lek-heng Lim. Rank aggregation via nuclear norm minimization. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 60–68. ACM, 2011.
- [28] Anil N Hirani, Kaushik Kalyanaraman, and Seth Watts. Least squares ranking on graphs. *arXiv preprint arXiv:1011.1716*, 2010.
- [29] Elizabeth A Hobson and Simon DeDeo. Social feedback and the emergence of rank in animal society. *PLoS computational biology*, 11(9):e1004411, 2015.
- [30] Xiaoye Jiang, Lek-Heng Lim, Yuan Yao, and Yinyu Ye. Statistical ranking and combinatorial Hodge theory. *Mathematical Programming*, 127(1):203–244, 2011.
- [31] Antony Joseph and Bin Yu. Impact of regularization on spectral clustering. *Ann. Statist.*, 44(4):1765–1791, 2016.

- [32] Zhao Kang, Chong Peng, and Qiang Cheng. Top-n recommender system via matrix completion. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.
- [33] R. Karp, J. Elson, D. Estrin, and S. Shenker. Optimal and global time synchronization in sensor networks. Technical report, Center for Embedded Networked Sensing, University of California, Los Angeles, 2003.
- [34] Jérôme Kunegis, Stephan Schmidt, Andreas Lommatzsch, Jürgen Lerner, Ernesto William De Luca, and Sahin Albayrak. Spectral analysis of signed graphs for clustering, prediction and visualization. *SDM*, 10:559–570, 2010.
- [35] Can M Le, Elizaveta Levina, and Roman Vershynin. Sparse random graphs: regularization and concentration of the Laplacian. *arXiv preprint arXiv:1502.03049*, 2015.
- [36] Tal Levy, Alireza Vahid, and Raja Giryes. Ranking recovery from limited comparisons using low-rank matrix completion. *arXiv preprint arXiv:1806.05419*, 2018.
- [37] Ren-Cang Li. On perturbations of matrix pencils with real spectra. *Mathematics of Computation*, 62(205):231–265, 1994.
- [38] Andrew K Massimino and Mark A Davenport. One-bit matrix completion for pairwise comparison matrices. 2013.
- [39] Sahand Negahban, Sewoong Oh, and Devavrat Shah. Iterative ranking from pair-wise comparisons. In *Advances in Neural Information Processing Systems 25*, pages 2474–2482, 2012.
- [40] Sahand Negahban, Sewoong Oh, and Devavrat Shah. Rank centrality: Ranking from pairwise comparisons. *Operations Research*, 65(1):266–287, 2017.
- [41] Sahand Negahban, Sewoong Oh, Kiran K Thekumparampil, and Jiaming Xu. Learning from comparisons and choices. *Journal of Machine Learning Research*, 19:1–95, 2018.
- [42] Onur Ozyesil, Nir Sharon, and Amit Singer. Synchronization over Cartan motion groups via contraction. *SIAM Journal on Applied Algebra and Geometry*, 2(2):207–241, 2018.
- [43] Onur Özyeşil, Vladislav Voroninski, Ronen Basri, and Amit Singer. A survey of structure from motion. *Acta Numerica*, 26:305–364, 2017.
- [44] Arun Rajkumar and Shivani Agarwal. When can we rank well from comparisons of  $O(n \log(n))$  non-actively chosen pairs? In *Conference on Learning Theory*, pages 1376–1401, 2016.
- [45] A. Singer. Angular synchronization by eigenvectors and semidefinite programming. *Appl. Comput. Harmon. Anal.*, 30(1):20–36, 2011.
- [46] A. Singer and H.-T. Wu. Vector diffusion maps and the Connection Laplacian. *Communications on Pure and Applied Mathematics*, 65(8):1067–1144, 2012.
- [47] G.W. Stewart and Ji guang Sun. *Matrix Perturbation Theory*. Academic Press, 1990.
- [48] L.L. Thurstone. A law of comparative judgement. *Psychological Review*, 34:278–286, 1927.
- [49] Carlo Tomasi and Takeo Kanade. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9(2):137–154, 1992.
- [50] M. Udell and A. Townsend. Why are big data matrices approximately low rank? *SIAM Journal on Mathematics of Data Science*, 1(1):144–160, 2019.
- [51] Maksims N. Volkovs and Richard S. Zemel. New learning methods for supervised and unsupervised preference aggregation. *J. Mach. Learn. Res.*, 15(1):1135–1176, January 2014.
- [52] Per-Åke Wedin. Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics*, 12(1):99–111, Mar 1972.

- [53] Hermann Weyl. Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung). *Mathematische Annalen*, 71(4):441–479, Dec 1912.
- [54] D. Yang and M. B. Wakin. Modeling and recovering non-transitive pairwise comparison matrices. In *2015 International Conference on Sampling Theory and Applications (SampTA)*, pages 39–43, May 2015.
- [55] Guangnan Ye, Dong Liu, I-Hong Jhuo, and Shih-Fu Chang. Robust late fusion with rank minimization. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3021–3028. IEEE, 2012.

## A Matrix perturbation analysis

Let  $A, \tilde{A} \in \mathbb{C}^{m \times n}$  where we assume  $m \geq n$  w.l.o.g. Let

$$\sigma_1(A) \geq \cdots \geq \sigma_n(A), \quad \sigma_1(\tilde{A}) \geq \cdots \geq \sigma_n(\tilde{A})$$

denote the singular values of  $A$  and  $\tilde{A}$  respectively, and denote  $E = \tilde{A} - A$ .

To begin with, we would like to quantify the perturbation of the singular values of  $\tilde{A}$  with respect to those of  $A$ . Weyl’s inequality [53] is a very useful result in this regard.

**Theorem 7** (Weyl’s inequality [53]). *It holds that*

$$\left| \sigma_i(\tilde{A}) - \sigma_i(A) \right| \leq \|E\|_2, \quad i = 1, \dots, n.$$

Next, let us write the singular value decomposition of  $A$  as

$$A = [U_1 \ U_2 \ U_3] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \\ 0 & 0 \end{bmatrix} [V_1 \ V_2]^*$$

and the same for  $\tilde{A}$  (conformal partition) so that

$$\tilde{A} = [\tilde{U}_1 \ \tilde{U}_2 \ \tilde{U}_3] \begin{bmatrix} \tilde{\Sigma}_1 & 0 \\ 0 & \tilde{\Sigma}_2 \\ 0 & 0 \end{bmatrix} [\tilde{V}_1 \ \tilde{V}_2]^*.$$

We will like to quantify the perturbation of  $\mathcal{R}(U_1), \mathcal{R}(V_1)$ , this is given precisely by Wedin’s theorem [52]. Before introducing the theorem, we need some definitions. Let  $U, \tilde{U} \in \mathbb{C}^{n \times k}$  (for  $k \leq n$ ) have orthonormal columns respectively and let  $\sigma_1 \geq \cdots \geq \sigma_k$  denote the singular values of  $U^* \tilde{U}$ . Then the  $k$  canonical angles between  $\mathcal{R}(U), \mathcal{R}(\tilde{U})$  are defined as  $\theta_i := \cos^{-1}(\sigma_i)$  for  $1 \leq i \leq k$ , with each  $\theta_i \in [0, \pi/2]$ . It is usual to define  $k \times k$  diagonal matrices  $\Theta(\mathcal{R}(U), \mathcal{R}(\tilde{U})) := \text{diag}(\theta_1, \dots, \theta_k)$  and  $\sin \Theta(\mathcal{R}(U), \mathcal{R}(\tilde{U})) := \text{diag}(\sin \theta_1, \dots, \sin \theta_k)$ . Denoting  $\|\cdot\|$  to be any unitarily invariant norm (Frobenius, spectral, etc.), it is useful to know that the following relation holds (see for eg., [37, Lemma 2.1], [47, Corollary I.5.4]).

$$\|\|\sin \Theta(\mathcal{R}(U), \mathcal{R}(\tilde{U}))\|\| = \|\|(I - \tilde{U} \tilde{U}^*)U\|\|.$$

Let  $\Phi$  be the matrix of canonical angles between  $\mathcal{R}(U_1), \mathcal{R}(\tilde{U}_1)$  and let  $\Theta$  be the matrix of canonical angles between  $\mathcal{R}(V_1), \mathcal{R}(\tilde{V}_1)$ .

**Theorem 8** (Wedin [52]). *Suppose that there are numbers  $\alpha, \delta > 0$  such that*

$$\min \sigma(\tilde{\Sigma}_1) \geq \alpha + \delta \quad \text{and} \quad \max \sigma(\Sigma_2) \leq \alpha. \tag{A.1}$$

Then,

$$\max \{ \|\sin \Theta\|_2, \|\sin \Phi\|_2 \} \leq \frac{\|E\|_2}{\delta}.$$

Note that due to (A.1) the bounds are restricted to subspaces corresponding to the largest singular values. But one can derive bounds on  $\max \{ \|\sin \Theta\|_F, \|\sin \Phi\|_F \}$  without this restriction (see for eg. [47, Theorem 4.1]).



## B Useful concentration inequalities

We outline here some useful concentration inequalities that are used in our proofs.

### B.1 Spectral norm of random symmetric matrices

We will make use of the following result for bounding the spectral norm of random symmetric matrices with independent, centered and bounded entries.

**Theorem 9** ([2, Corollary 3.12, Remark 3.13]). *Let  $X$  be an  $n \times n$  symmetric matrix whose entries  $X_{ij}$  ( $i \leq j$ ) are independent, centered random variables. Then there exists for any  $0 < \varepsilon \leq 1/2$  a universal constant  $c_\varepsilon$  such that for every  $t \geq 0$ ,*

$$\mathbb{P}(\|X\|_2 \geq (1 + \varepsilon)2\sqrt{2}\tilde{\sigma} + t) \leq n \exp\left(-\frac{t^2}{c_\varepsilon \tilde{\sigma}_*^2}\right) \quad (\text{B.1})$$

where

$$\tilde{\sigma} := \max_i \sqrt{\sum_j \mathbb{E}[X_{ij}^2]}, \quad \tilde{\sigma}_* := \max_{i,j} \|X_{ij}\|_\infty.$$

Note that it suffices to employ upper bound estimates on  $\tilde{\sigma}, \tilde{\sigma}_*$  in (B.1).

### B.2 Bernstein inequality

**Theorem 10** ([6, Corollary 2.11]). *Let  $X_1, \dots, X_n$  be independent random variables with  $|X_i| \leq b$  for all  $i$ , and  $v = \sum_{i=1}^n \mathbb{E}[X_i^2]$ . Then for any  $t \geq 0$ ,*

$$\mathbb{P}\left(\left|\sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right| \geq t\right) \leq 2 \exp\left(-\frac{t^2}{2(v + \frac{bt}{3})}\right).$$

### B.3 Product of a random matrix (raised to a power) and a fixed vector

Given a random symmetric matrix  $X$  with independent entries on and above the diagonal, and a fixed vector  $u$ , Eldridge et al. [21, Theorem 15] provided an upper tail bound for  $|(X^k u)_i|$ . The proof technique therein followed that of Erdős et al. [22]. In fact, the proof goes through even if  $X$  is a random skew-symmetric matrix, the version stated below is for such matrices.

**Theorem 11** ([21, Theorem 15]). *Let  $X$  be a  $n \times n$  skew-symmetric and centered random matrix. Let  $u \in \mathbb{R}^n$  be a fixed vector with  $\|u\|_\infty = 1$ . Choose  $\xi > 1$ ,  $0 < \kappa < 1$  and define  $\mu = \frac{2}{\kappa+1}$ . If  $\mathbb{E}[|X_{ij}|^m] \leq 1/n$  for all  $m \geq 2$ , and  $k \leq \frac{\kappa}{8}(\log n)^{2\xi}$ , then we have for any given  $\alpha \in [n]$  that*

$$\mathbb{P}(|(X^k u)_\alpha| \geq (\log n)^{k\xi}) \leq n^{-\frac{1}{4}(\log_\mu n)^{\xi-1}(\log_\mu e)^{-\xi}}.$$

## C Proof of Theorem 1

### C.1 Proof of Lemma 1

We begin by noting that for  $\alpha = \frac{r^T e}{n}$ ,  $\langle e, r - \alpha e \rangle = 0$  and so  $u_1, u_2$  (similarly  $v_1, v_2$ ) are orthogonal. Moreover,  $u_1, u_2, v_1, v_2$  also have unit  $\ell_2$  norm. Finally, we note that

$$u_1 v_1^T + u_2 v_2^T = \frac{1}{\sqrt{n}\|r - \alpha e\|_2} [-e(r - \alpha e)^T + (r - \alpha e)e^T] = \frac{1}{\sqrt{n}\|r - \alpha e\|_2} \underbrace{[re^T - er^T]}_C,$$

which completes the proof.

## C.2 Proof of Lemma 2

We begin by considering the second moment of  $S_{ij}$  for  $i < j$ . Note that  $\mathbb{E}[N_{ij}] = 0$  and  $\mathbb{E}[N_{ij}^2] = M^2/3$ . We thus obtain

$$\begin{aligned}\mathbb{E}[(S)_{ij}^2] &= \eta p(1 - \eta p)^2(r_i - r_j)^2 + (1 - \eta)p\mathbb{E}[N_{ij} - \eta p(r_i - r_j)]^2 + (1 - p)(r_i - r_j)^2\eta^2 p^2 \\ &= \eta p(1 - \eta p)^2(r_i - r_j)^2 + (1 - \eta)p\left[\frac{M^2}{3} + \eta^2 p^2(r_i - r_j)^2\right] + (1 - p)(r_i - r_j)^2\eta^2 p^2\end{aligned}$$

which together with  $(r_i - r_j)^2 \leq M^2$  yields

$$\begin{aligned}\mathbb{E}[(S)_{ij}^2] &\leq \eta p(1 - \eta p)^2 M^2 + (1 - \eta)p\left[\frac{M^2}{3} + \eta^2 p^2 M^2\right] + (1 - p)\eta^2 p^2 M^2 \\ &= [\eta p(1 - \eta p)^2 + (1 - \eta)p\left(\frac{1}{3} + \eta^2 p^2\right) + (1 - p)\eta^2 p^2]M^2 \\ &\leq \left(\eta p + \frac{4}{3}p + \eta^2 p^2 \leq \frac{10}{3}p\right)M^2 = \frac{10p}{3}M^2.\end{aligned}\tag{C.1}$$

We thus have that  $\mathbb{E}[(S)_{ij}^2] \leq \frac{10p}{3}M^2$  for  $i < j$ ; the same bound holds for  $i > j$  as well. Hence we can bound the quantities  $\tilde{\sigma}, \tilde{\sigma}_*$  defined in Theorem 9 as follows.

$$\begin{aligned}\tilde{\sigma} &:= \max_i \sqrt{\sum_j \mathbb{E}[(\tilde{S})_{ij}^2]} \leq \sqrt{\frac{10p}{3}nM^2} = M\sqrt{\frac{10p}{3}n}, \\ \tilde{\sigma}_* &:= \max_{i,j} \|(\tilde{S})_{ij}\|_\infty \leq 2M.\end{aligned}$$

Then by invoking Theorem 9, we obtain for any given  $t \geq 0$ ,  $0 < \varepsilon \leq 1/2$  that

$$\begin{aligned}\mathbb{P}(\|\tilde{S}\|_2 \geq (1 + \varepsilon)2\sqrt{2}M\sqrt{\frac{10p}{3}n} + t) &\leq \mathbb{P}(\|\tilde{S}\|_2 \geq (1 + \varepsilon)2\sqrt{2}\tilde{\sigma} + t) \\ &\leq 2n \exp\left(\frac{-t^2}{c_\varepsilon \sigma_*^2}\right) \leq 2n \exp\left(\frac{-t^2}{4c_\varepsilon M^2}\right),\end{aligned}$$

where  $c_\varepsilon > 0$  depends only on  $\varepsilon$ . Plugging  $t = 2\sqrt{2}M\sqrt{\frac{10p}{3}n}$  we obtain the stated bound.

## C.3 Proof of Lemma 3

Note that Weyl's inequality [53] (see Theorem 7) for singular value perturbation readily yields

$$\sigma_i(H) \in [\sigma_i(\eta p C) \pm \Delta], \quad \forall i = 1, \dots, n.$$

In particular, using the expressions for  $\sigma_1(C), \sigma_2(C)$  from Lemma 1, we have that

$$\begin{aligned}\sigma_1(H), \sigma_2(H) &\geq \eta p\sqrt{n}\|r - \alpha e\|_2 - \Delta, \\ \sigma_3(H), \dots, \sigma_n(H) &\leq \Delta.\end{aligned}$$

Then using Wedin's bound for perturbation of singular subspaces [52] (see Theorem 8), we obtain

$$\|\sin \Theta(\mathcal{R}(\hat{U}), \mathcal{R}(U))\|_2 = \|(I - \hat{U}\hat{U}^T)U\|_2 \leq \frac{\Delta}{\eta p\sqrt{n}\|r - \alpha e\|_2 - \Delta} (=:\delta)$$

provided  $\Delta < \eta p\sqrt{n}\|r - \alpha e\|_2$  holds.

## C.4 Proof of Lemma 4

To begin with, note that  $\bar{u}_1 = \hat{U}\hat{U}^T u_1$  is the orthogonal projection of  $u_1$  on  $\mathcal{R}(\hat{U})$ . A unit vector orthogonal to  $\bar{u}_1$  and lying in  $\mathcal{R}(\hat{U})$  is given by

$$\tilde{u}_2 = \frac{1}{\|\bar{u}_1\|_2} \hat{U} \begin{bmatrix} \mp \langle \hat{u}_2, \bar{u}_1 \rangle \\ \pm \langle \hat{u}_1, \bar{u}_1 \rangle \end{bmatrix}.$$

Denoting  $\bar{u}_2$  to be the orthogonal projection of  $u_2$  on  $\mathcal{R}(\hat{U})$ , we can decompose  $\langle \tilde{u}_2, u_2 \rangle$  as

$$\langle \tilde{u}_2, u_2 \rangle = \langle \tilde{u}_2, u_2 - \frac{\bar{u}_2}{\|\bar{u}_2\|_2} \rangle + \langle \tilde{u}_2, \frac{\bar{u}_2}{\|\bar{u}_2\|_2} \rangle,$$

which leads to

$$|\langle \tilde{u}_2, u_2 \rangle| \geq |\langle \tilde{u}_2, \frac{\bar{u}_2}{\|\bar{u}_2\|_2} \rangle| - |\langle \tilde{u}_2, u_2 - \frac{\bar{u}_2}{\|\bar{u}_2\|_2} \rangle|. \quad (\text{C.2})$$

We will now bound the two terms in the RHS of (C.2) starting with the first term.

**Lower bounding the first term in RHS of (C.2).** Since  $\langle \hat{u}_2, \bar{u}_1 \rangle = \langle \hat{u}_2, u_1 \rangle$ ,  $\langle \hat{u}_1, \bar{u}_1 \rangle = \langle \hat{u}_1, u_1 \rangle$ , we obtain

$$\begin{aligned} \langle \tilde{u}_2, \frac{\bar{u}_2}{\|\bar{u}_2\|_2} \rangle &= \frac{u_2^T}{\|\bar{u}_1\|_2 \|\bar{u}_2\|_2} \hat{U} \underbrace{\hat{U}^T \hat{U}}_I \begin{bmatrix} \mp \langle \hat{u}_2, \bar{u}_1 \rangle \\ \pm \langle \hat{u}_1, \bar{u}_1 \rangle \end{bmatrix} \\ &= \frac{1}{\|\bar{u}_1\|_2 \|\bar{u}_2\|_2} [\mp \langle u_2, \hat{u}_1 \rangle \langle \hat{u}_2, u_1 \rangle \pm \langle u_2, \hat{u}_2 \rangle \langle \hat{u}_1, u_1 \rangle] \\ &= \frac{1}{\|\bar{u}_1\|_2 \|\bar{u}_2\|_2} [\pm \det(U^T \hat{U})]. \end{aligned}$$

Recall the definition of  $\delta$  in (5.2). The following simple claims will be useful for us.

**Claim 1.** *With  $\delta$  as defined in (5.2) we have  $|\det(U^T \hat{U})| \geq 1 - \delta^2$ .*

*Proof.* Since  $U^T \hat{U}$  is a  $2 \times 2$  matrix, we have that

$$\left| \det(U^T \hat{U}) \right| \geq \sigma_{\min}^2(U^T \hat{U}) = 1 - \|\sin \Theta(\mathcal{R}(\hat{U}), \mathcal{R}(U))\|_2^2 \geq 1 - \delta^2.$$

□

**Claim 2.** *It holds that  $\|\bar{u}_1 - u_1\|_2, \|\bar{u}_2 - u_2\|_2 \leq \delta$ .*

*Proof.* We will only show the bound for  $\|\bar{u}_1 - u_1\|_2$  as an identical argument holds for bounding  $\|\bar{u}_2 - u_2\|_2$ . Using the facts  $\bar{u}_1 = \hat{U}\hat{U}^T u_1$  and  $u_1 = U U^T u_1$ , we obtain

$$\|u_1 - \bar{u}_1\|_2 = \|(I - \hat{U}\hat{U}^T)u_1\|_2 = \|(I - \hat{U}\hat{U}^T)U U^T u_1\|_2 \leq \|(I - \hat{U}\hat{U}^T)U U^T\|_2 = \|(I - \hat{U}\hat{U}^T)U\|_2 \leq \delta$$

where the last inequality follows from (5.2). □

Using Claims 1, 2 we can lower bound the first term in RHS of (C.2) as follows.

$$\left| \langle \tilde{u}_2, \frac{\bar{u}_2}{\|\bar{u}_2\|_2} \rangle \right| \geq \frac{1 - \delta^2}{(1 + \delta)^2} = \frac{1 - \delta}{1 + \delta} = 1 - \frac{2\delta}{1 + \delta} \geq 1 - 2\delta. \quad (\text{C.3})$$

Upper bounding the second term in RHS of (C.2). We can do this as follows.

$$\begin{aligned}
\left| \left\langle \tilde{u}_2, u_2 - \frac{\bar{u}_2}{\|\bar{u}_2\|_2} \right\rangle \right| &\leq \|u_2 - \frac{\bar{u}_2}{\|\bar{u}_2\|_2}\|_2 \quad (\text{Cauchy-Schwartz and since } \|\tilde{u}_2\|_2 = 1) \\
&= \|u_2 - \bar{u}_2 + \bar{u}_2 - \frac{\bar{u}_2}{\|\bar{u}_2\|_2}\|_2 \\
&\leq \|u_2 - \bar{u}_2\|_2 + \|\bar{u}_2 - \frac{\bar{u}_2}{\|\bar{u}_2\|_2}\|_2 \\
&\leq \delta + \left| 1 - \frac{1}{\|\bar{u}_2\|_2} \right| \quad (\text{Using Claim 2}) \\
&= \delta + \frac{|\|\bar{u}_2\|_2 - 1|}{\|\bar{u}_2\|_2} \\
&\leq \delta + \frac{\delta}{1 - \delta} \quad (\text{since } \|\bar{u}_2\|_2 \in [1 - \delta, 1 + \delta] \text{ from Claim 2}) \\
&\leq 3\delta \quad (\text{whenever } \delta \leq 1/2).
\end{aligned} \tag{C.4}$$

Altogether, we conclude that if  $\delta \leq 1/2$  then using (C.3), (C.4) in (C.2), we obtain

$$|\langle \tilde{u}_2, u_2 \rangle| \geq 1 - 5\delta,$$

and consequently, there exists  $\beta \in \{-1, 1\}$  such that

$$\|\tilde{u}_2 - \beta u_2\|_2^2 = 2 - 2\langle \tilde{u}_2, \beta u_2 \rangle \leq 10\delta. \tag{C.5}$$

## D Proof of Lemmas from Section 5.2

### D.1 Proof of Lemma 5

*Proof.* Recall from the proof of Lemma 4 that

$$\tilde{u}_2 = \frac{1}{\|\bar{u}_1\|_2} \hat{U} \begin{bmatrix} \mp \langle \hat{u}_2, \bar{u}_1 \rangle \\ \pm \langle \hat{u}_1, \bar{u}_1 \rangle \end{bmatrix} = \frac{1}{\|\bar{u}_1\|_2} \hat{U} \begin{bmatrix} 0 & \mp 1 \\ \pm 1 & 0 \end{bmatrix} \hat{U}^T \bar{u}_1 = \frac{1}{\|\bar{u}_1\|_2} \hat{U} \begin{bmatrix} 0 & \mp 1 \\ \pm 1 & 0 \end{bmatrix} \hat{U}^T u_1, \tag{D.1}$$

where the last equality follows from the fact  $\langle \hat{u}_2, \bar{u}_1 \rangle = \langle \hat{u}_2, u_1 \rangle$ ,  $\langle \hat{u}_1, \bar{u}_1 \rangle = \langle \hat{u}_1, u_1 \rangle$ . Let  $O$  be any  $2 \times 2$  orthogonal matrix, clearly,  $UO$  also corresponds to the two largest left singular vectors of  $\mathbb{E}[H]$  (since  $\sigma_1(\mathbb{E}[H]) = \sigma_2(\mathbb{E}[H])$ ). Denote  $P = \hat{U} - UO$ . Plugging in (D.1), we obtain

$$\begin{aligned}
\tilde{u}_2 &= \frac{1}{\|\bar{u}_1\|_2} \hat{U} \underbrace{\begin{bmatrix} 0 & \mp 1 \\ \pm 1 & 0 \end{bmatrix}}_D \hat{U}^T \bar{u}_1 [O^T U^T + P] u_1 \\
&= \pm \det(O) u_2 + \underbrace{\frac{1}{\|\bar{u}_1\|_2} U O D P^T u_1}_{w_1} + \underbrace{\frac{1}{\|\bar{u}_1\|_2} P D O^T U^T u_1}_{w_2} + \underbrace{\frac{1}{\|\bar{u}_1\|_2} P D P^T u_1}_{w_3} \\
\Rightarrow \|\tilde{u}_2 \mp \det(O) u_2\|_\infty &\leq \|w_1\|_\infty + \|w_2\|_\infty + \|w_3\|_\infty.
\end{aligned} \tag{D.2}$$

Denoting

$$P = [p_1 \ p_2], \quad O = \begin{bmatrix} o_{11} & o_{12} \\ o_{21} & o_{22} \end{bmatrix}$$

we will now bound  $\|w_1\|_\infty, \|w_2\|_\infty, \|w_3\|_\infty$ .

To begin with,

$$\begin{aligned}
w_1 &= \frac{1}{\|\bar{u}_1\|_2} [u_1 \ u_2] O \begin{bmatrix} 0 & \mp 1 \\ \pm 1 & 0 \end{bmatrix} \begin{bmatrix} \langle p_1, u_1 \rangle \\ \langle p_2, u_1 \rangle \end{bmatrix} \\
&= \frac{1}{\|\bar{u}_1\|_2} [u_1 \ u_2] O \begin{bmatrix} \mp \langle p_2, u_1 \rangle \\ \pm \langle p_1, u_1 \rangle \end{bmatrix} \\
&= \frac{1}{\|\bar{u}_1\|_2} [u_1 \ u_2] \begin{bmatrix} \mp o_{11} \langle p_2, u_1 \rangle \pm o_{12} \langle p_1, u_1 \rangle \\ \mp o_{21} \langle p_2, u_1 \rangle \pm o_{22} \langle p_1, u_1 \rangle \end{bmatrix} \\
\Rightarrow \|w_1\|_\infty &\leq \frac{1}{\|\bar{u}_1\|_2} (|\langle p_2, u_1 \rangle| + |\langle p_1, u_1 \rangle|) (\|u_1\|_\infty + \|u_2\|_\infty). \tag{D.3}
\end{aligned}$$

Clearly,  $\|u_1\|_\infty = \frac{1}{\sqrt{n}}$  and  $\|u_2\|_\infty \leq \frac{M-\alpha}{\|r-\alpha e\|_2}$ . Using Hölder's inequality,

$$|\langle p_i, u_1 \rangle| \leq \|P\|_{\max} \|u\|_1 = \|P\|_{\max} \sqrt{n}; \quad i = 1, 2.$$

Moreover, from Claim 2, we know that  $\|\bar{u}_1\|_2 \geq 1 - \delta$  where we recall  $\delta$  defined in (5.2). We saw in (5.3) that  $\Delta \leq \frac{\eta p \|r-\alpha e\|_2 \sqrt{n}}{3} \Leftrightarrow \delta \leq 1/2$ , which implies  $\|\bar{u}_1\|_2 \geq 1/2$ . Plugging these bounds in (D.3), we obtain

$$\|w_1\|_\infty \leq 4\|P\|_{\max} \left( 1 + \frac{\sqrt{n}(M-\alpha)}{\|r-\alpha e\|_2} \right). \tag{D.4}$$

Next, we can bound  $\|w_2\|_\infty$  as follows.

$$\begin{aligned}
w_2 &= \frac{1}{\|\bar{u}_1\|_2} [p_1 \ p_2] \begin{bmatrix} 0 & \mp 1 \\ \pm 1 & 0 \end{bmatrix} O^T \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
&= \frac{1}{\|\bar{u}_1\|_2} [p_1 \ p_2] \begin{bmatrix} \mp o_{12} \\ \pm o_{11} \end{bmatrix} \\
\Rightarrow \|w_2\|_\infty &\leq \frac{2\|P\|_{\max}}{\|\bar{u}_1\|_2} \leq 4\|P\|_{\max} \tag{D.5}
\end{aligned}$$

since  $\Delta \leq \frac{\eta p \|r-\alpha e\|_2 \sqrt{n}}{3}$ .

We now bound  $\|w_3\|_\infty$ .

$$\begin{aligned}
w_3 &= \frac{1}{\|\bar{u}_1\|_2} [p_1 \ p_2] \begin{bmatrix} 0 & \mp 1 \\ \pm 1 & 0 \end{bmatrix} \begin{bmatrix} \langle p_1, u_1 \rangle \\ \langle p_2, u_1 \rangle \end{bmatrix} \\
&= \frac{1}{\|\bar{u}_1\|_2} [p_1 \ p_2] \begin{bmatrix} \mp \langle p_2, u_1 \rangle \\ \pm \langle p_1, u_1 \rangle \end{bmatrix} \\
\Rightarrow \|w_3\|_\infty &\leq \frac{1}{\|\bar{u}_1\|_2} (|\langle p_2, u_1 \rangle| + |\langle p_1, u_1 \rangle|) \|P\|_{\max} \\
&\leq 4\|P\|_{\max}^2 \|u_1\|_1 = 4\sqrt{n} \|P\|_{\max}^2 \quad (\text{since } \Delta \leq \frac{\eta p \|r-\alpha e\|_2 \sqrt{n}}{3} \text{ and using Hölder's inequality}). \tag{D.6}
\end{aligned}$$

Plugging (D.4), (D.5), (D.6) in (D.2), we obtain

$$\begin{aligned}
\|\tilde{u}_2 \mp \det(O)u_2\|_\infty &\leq 4\|P\|_{\max} \left( 2 + \frac{\sqrt{n}(M-\alpha)}{\|r-\alpha e\|_2} \right) + 4\sqrt{n} \|P\|_{\max}^2 \\
&= 4\|\widehat{U} - UO\|_{\max} \left( 2 + \frac{\sqrt{n}(M-\alpha)}{\|r-\alpha e\|_2} \right) + 4\sqrt{n} \|\widehat{U} - UO\|_{\max}^2.
\end{aligned}$$

The above bound is true for any orthogonal matrix  $O$ , it is not difficult to see that there exists an orthogonal matrix  $O = O^*$  such that  $\widehat{U}$  is ‘‘aligned’’ with  $UO^*$ . Indeed, for any orthogonal  $O$ , we first obtain via triangle inequality that

$$\|\widehat{U} - UO\|_2 \leq \|(I - UU^T)\widehat{U}\|_2 + \|UU^T\widehat{U} - UO\|_2 = \|\sin \Theta(\mathcal{R}(U), \mathcal{R}(\widehat{U}))\|_2 + \|U^T\widehat{U} - O\|_2. \tag{D.7}$$

Denoting  $\tilde{U}\tilde{D}\tilde{V}^T$  to be the SVD of  $U^T\hat{U}$ , we choose  $O = O^* = \tilde{U}\tilde{V}^T$  (so  $O^*$  is orthogonal). Denoting  $\theta_p$  to be the principal angle between  $\mathcal{R}(U), \mathcal{R}(\hat{U})$ , we obtain

$$\|U^T\hat{U} - O^*\|_2 = \|I - \tilde{D}\|_2 = 1 - \cos\theta_p \leq \sin\theta_p = \|\sin\Theta(\mathcal{R}(U), \mathcal{R}(\hat{U}))\|_2.$$

Plugging this in (D.7) leads to

$$\begin{aligned} \|\hat{U} - UO^*\|_2 &\leq 2\|\sin\Theta(\mathcal{R}(U), \mathcal{R}(\hat{U}))\|_2 \\ &\leq \frac{2\Delta}{\eta p\sqrt{n}\|r - \alpha e\|_2 - \Delta} \quad (\text{see proof of Lemma 3}) \\ &\leq \frac{3\Delta}{\eta p\sqrt{n}\|r - \alpha e\|_2}. \quad (\text{since } \Delta \leq \frac{\eta p\|r - \alpha e\|_2\sqrt{n}}{3}) \end{aligned}$$

□

## D.2 Proof of Lemma 6

*Proof.* 1. (**Bounding**  $\|E_1\|_{\max}$ ) Denoting  $(\hat{U} - UO^*)_i$  to be the  $i^{\text{th}}$  column of  $\hat{U} - UO^*$ , we note that

$$\begin{aligned} E_1 &= -\mathbb{E}[H]^2(\hat{U} - UO^*)\hat{\Sigma}^{-2} \\ &= \sigma^2 UU^T(\hat{U}^T - UO^{*T})\hat{\Sigma}^{-2} \\ &= (u_1 u_1^T + u_2 u_2^T) \left[ (\hat{U} - UO^*)_1 \frac{\sigma^2}{\hat{\sigma}_1^2} \quad (\hat{U} - UO^*)_2 \frac{\sigma^2}{\hat{\sigma}_2^2} \right] \\ &= \left[ u_1 \langle u_1, (\hat{U} - UO^*)_1 \rangle \frac{\sigma^2}{\hat{\sigma}_1^2} + u_2 \langle u_2, (\hat{U} - UO^*)_1 \rangle \frac{\sigma^2}{\hat{\sigma}_1^2} \quad u_1 \langle u_1, (\hat{U} - UO^*)_2 \rangle \frac{\sigma^2}{\hat{\sigma}_2^2} + u_2 \langle u_2, (\hat{U} - UO^*)_2 \rangle \frac{\sigma^2}{\hat{\sigma}_2^2} \right]. \end{aligned}$$

For the  $i^{\text{th}}$  entry in the first column ( $i = 1, \dots, n$ ), we have that

$$\begin{aligned} &\left| u_{1,i} \langle u_1, (\hat{U} - UO^*)_1 \rangle \frac{\sigma^2}{\hat{\sigma}_1^2} + u_{2,i} \langle u_2, (\hat{U} - UO^*)_1 \rangle \frac{\sigma^2}{\hat{\sigma}_1^2} \right| \\ &\leq \|u_1\|_\infty \underbrace{\|(\hat{U} - UO^*)_1\|_2}_{\leq \|\hat{U} - UO^*\|_2} \frac{\sigma^2}{\hat{\sigma}_1^2} + \|u_2\|_\infty \underbrace{\|(\hat{U} - UO^*)_1\|_2}_{\leq \|\hat{U} - UO^*\|_2} \frac{\sigma^2}{\hat{\sigma}_1^2}. \end{aligned}$$

Since  $\|u_1\|_\infty = \frac{1}{\sqrt{n}}$ ,  $\|u_2\|_\infty \leq \frac{M-\alpha}{\|r-\alpha e\|_2}$ , and  $\hat{\sigma}_1 \geq \sigma - \Delta \geq \frac{2\sigma}{3}$ , we obtain

$$\begin{aligned} &\left| u_{1,i} \langle u_1, (\hat{U} - UO^*)_1 \rangle \frac{\sigma^2}{\hat{\sigma}_1^2} + u_{2,i} \langle u_2, (\hat{U} - UO^*)_1 \rangle \frac{\sigma^2}{\hat{\sigma}_1^2} \right| \\ &\leq \frac{9}{4} \|\hat{U} - UO^*\|_2 \left( \frac{1}{\sqrt{n}} + \frac{M-\alpha}{\|r-\alpha e\|_2} \right) \\ &\leq \frac{27\Delta}{4\sigma} \left( \frac{1}{\sqrt{n}} + \frac{M-\alpha}{\|r-\alpha e\|_2} \right) \end{aligned}$$

where in the last inequality, we used  $\|\hat{U} - UO^*\|_2 \leq \frac{3\Delta}{\sigma}$  (from Lemma 5). The same entry-wise bound holds for the second column, leading to the stated bound on  $\|E_1\|_{\max}$ .

2. (**Bounding**  $\|E_2\|_{\max}$ ) Denoting the  $(i, j)^{\text{th}}$  entry of  $O^*$  by  $o_{ij}^*$ , we have that

$$\begin{aligned} E_2 &= \mathbb{E}[H]^2 UO^*(\sigma^{-2}I - \hat{\Sigma}^{-2}) \\ &= -UU^T \sigma^2 (UO^*) \begin{bmatrix} \sigma^{-2} - \hat{\sigma}_1^{-2} & 0 \\ 0 & \sigma^{-2} - \hat{\sigma}_2^{-2} \end{bmatrix} \\ &= -\sigma^2 [(\sigma^{-2} - \hat{\sigma}_1^{-2})(u_1 o_{11}^* + u_2 o_{21}^*) \quad (\sigma^{-2} - \hat{\sigma}_2^{-2})(u_1 o_{12}^* + u_2 o_{22}^*)]. \end{aligned}$$

Hence the  $i^{\text{th}}$  entry of the first column can be bounded as follows.

$$\begin{aligned} \sigma^2 |\sigma^{-2} - \hat{\sigma}_1^{-2}| |u_{1,i} o_{11}^* + u_{2,i} o_{12}^*| &\leq \frac{\Delta(2\sigma + \Delta)}{(\sigma - \Delta)^2} (\|u_1\|_\infty + \|u_2\|_\infty) \quad (\text{using } |\hat{\sigma}_1 - \sigma| \leq \Delta) \\ &\leq \frac{21\Delta}{4\sigma} \left( \frac{1}{\sqrt{n}} + \frac{M - \alpha}{\|r - \alpha e\|_2} \right) \quad (\text{using } \Delta \leq \frac{\sigma}{3}). \end{aligned}$$

The same entry-wise bound holds for the second column, leading to the stated bound on  $\|E_2\|_{\max}$ .

3. **(Bounding  $\|E_3\|_{\max}$ )** We have that

$$\begin{aligned} E_3 &= -\mathbb{E}[H] Z \hat{U} \hat{\Sigma}^{-2} \\ &= -\sigma (u_1 v_1^T + u_2 v_2^T) Z [\hat{\sigma}_1^{-2} \hat{u}_1 \quad \hat{\sigma}_2^{-2} \hat{u}_2] \\ &= -\sigma [u_1 \langle v_1, Z \hat{u}_1 \rangle \hat{\sigma}_1^{-2} + u_2 \langle v_2, Z \hat{u}_1 \rangle \hat{\sigma}_1^{-2} \quad u_1 \langle v_1, Z \hat{u}_2 \rangle \hat{\sigma}_2^{-2} + u_2 \langle v_2, Z \hat{u}_2 \rangle \hat{\sigma}_2^{-2}]. \end{aligned}$$

The  $i^{\text{th}}$  entry of the first column can be bounded as follows.

$$\begin{aligned} \sigma |u_{1,i} \langle v_1, Z \hat{u}_1 \rangle \hat{\sigma}_1^{-2} + u_{2,i} \langle v_2, Z \hat{u}_1 \rangle \hat{\sigma}_1^{-2}| &\leq \frac{\sigma}{\hat{\sigma}_1^2} (\|u_1\|_\infty |\langle v_1, Z \hat{u}_1 \rangle| + \|u_2\|_\infty |\langle v_2, Z \hat{u}_1 \rangle|) \\ &\leq \frac{\sigma}{\hat{\sigma}_1^2} \|Z\|_2 \left( \frac{1}{\sqrt{n}} + \frac{M - \alpha}{\|r - \alpha e\|_2} \right) \quad (\text{using Cauchy-Schwarz}) \\ &\leq \frac{9\Delta}{4\sigma} \left( \frac{1}{\sqrt{n}} + \frac{M - \alpha}{\|r - \alpha e\|_2} \right) \quad (\text{using } \hat{\sigma}_1 \geq 2\sigma/3). \end{aligned}$$

The same entry-wise bound holds for the second column, leading to the stated bound on  $\|E_3\|_{\max}$ .

4. **(Bounding  $\|E_4\|_{\max}$ )** We have that

$$\begin{aligned} E_4 &= Z \mathbb{E}[H] \hat{U} \hat{\Sigma}^{-2} = \sigma Z (u_1 v_1^T + u_2 v_2^T) [\hat{\sigma}_1^{-2} \hat{u}_1 \quad \hat{\sigma}_2^{-2} \hat{u}_2] \\ &= \sigma \left[ Z u_1 \frac{v_1^T \hat{u}_1}{\hat{\sigma}_1^2} + Z u_2 \frac{v_2^T \hat{u}_1}{\hat{\sigma}_1^2} \quad Z u_1 \frac{v_1^T \hat{u}_2}{\hat{\sigma}_2^2} + Z u_2 \frac{v_2^T \hat{u}_2}{\hat{\sigma}_2^2} \right]. \end{aligned}$$

The magnitude of the  $i^{\text{th}}$  entry of the first column can be bounded as

$$\begin{aligned} \frac{\sigma}{\hat{\sigma}_1^2} |(v_1^T \hat{u}_1)(Z u_1)_i + (v_2^T \hat{u}_1)(Z u_2)_i| &\leq \frac{\sigma}{\hat{\sigma}_1^2} (\|Z u_1\|_\infty + \|Z u_2\|_\infty) \quad (\text{Using Cauchy-Schwarz}) \\ &\leq \frac{9}{4\sigma} (\|Z u_1\|_\infty + \|Z u_2\|_\infty) \quad (\text{using } \hat{\sigma}_1 \geq 2\sigma/3). \end{aligned}$$

The same entry-wise bound holds for the second column, leading to the stated bound on  $\|E_4\|_{\max}$ .

5. **(Bounding  $\|E_5\|_{\max}$ )** We can write

$$E_5 = Z^2 \hat{U} \hat{\Sigma}^{-2} = Z^2 (\hat{U} - UO^*) \hat{\Sigma}^{-2} + Z^2 (UO^*) \hat{\Sigma}^{-2}$$

which in turn implies  $\|E_5\|_{\max} \leq \|Z^2 (\hat{U} - UO^*) \hat{\Sigma}^{-2}\|_{\max} + \|Z^2 (UO^*) \hat{\Sigma}^{-2}\|_{\max}$ . We will now bound these terms individually, below. To begin with,

$$\begin{aligned} \|Z^2 (\hat{U} - UO^*) \hat{\Sigma}^{-2}\|_{\max} &\leq \|Z^2 (\hat{U} - UO^*) \hat{\Sigma}^{-2}\|_2 \\ &\leq \Delta^2 \|\hat{U} - UO^*\|_2 \|\hat{\Sigma}^{-2}\|_2 \quad (\text{sub-multiplicativity of } \|\cdot\|_2) \\ &\leq \frac{3\Delta^3}{\sigma} \|\hat{\Sigma}^{-2}\|_2 \quad (\text{from Lemma 5}) \\ &\leq \frac{27\Delta^3}{4\sigma^3} \quad (\text{using } \hat{\sigma}_i \geq 2\sigma/3). \end{aligned}$$

Additionally, denoting  $O^* = [o_1^* \ o_2^*]$ , we can write

$$Z^2(UO^*)\widehat{\Sigma}^{-2} = \underbrace{[Z^2u_1 \ Z^2u_2]}_{\tilde{A}}[\widehat{\sigma}_1^{-2}o_1^* \ \widehat{\sigma}_2^{-2}o_2^*] = [\widehat{\sigma}_1^{-2}\tilde{A}o_1^* \ \widehat{\sigma}_2^{-2}\tilde{A}o_2^*]$$

which in turn implies

$$\begin{aligned} \|Z^2(UO^*)\widehat{\Sigma}^{-2}\|_{\max} &\leq \max \left\{ \frac{\|\tilde{A}o_1^*\|_{\infty}}{\widehat{\sigma}_1^2}, \frac{\|\tilde{A}o_2^*\|_{\infty}}{\widehat{\sigma}_2^2} \right\} \\ &\leq \frac{9}{4\sigma^2} \max \left\{ \|\tilde{A}o_1^*\|_{\infty}, \|\tilde{A}o_2^*\|_{\infty} \right\} \quad (\text{using } \widehat{\sigma}_i \geq 2\sigma/3) \\ &\leq \frac{9}{4\sigma^2} (\|Z^2u_1\|_{\infty} + \|Z^2u_2\|_{\infty}). \end{aligned}$$

□

### D.3 Proof of Lemma 7

*Proof.* 1. For a fixed  $i$ , consider  $\sum_{j=1}^n Z_{ij}u_{1j}$  which is a sum of independent random variables. From the definition of  $Z_{ij}$ , we can see that  $|Z_{ij}u_{1j}| \leq \frac{2M}{\sqrt{n}}$  for each  $j$ . Moreover,  $\mathbb{E}[(Z_{ij}u_{1j})^2] = \frac{1}{n}\mathbb{E}[Z_{ij}^2] \leq \frac{10pM^2}{3n}$  where the last inequality was shown in the proof of Lemma 2 (see (C.1)). Hence  $\sum_{i=1}^n \mathbb{E}[(Z_{ij}u_{1j})^2] \leq \frac{10pM^2}{3}$  and it follows from Bernstein's inequality (see Theorem B.2) and the union bound that

$$\mathbb{P}(\|Zu_1\|_{\infty} \geq t) \leq 2n \exp \left( -\frac{3t^2}{20pM^2 + \frac{4tM}{\sqrt{n}}} \right). \quad (\text{D.8})$$

We want to choose  $t$  such that

$$3t^2 \geq 2 \log n \left( 20pM^2 + \frac{4tM}{\sqrt{n}} \right) \Leftrightarrow 3t^2 - 40pM^2 \log n - \frac{8tM \log n}{\sqrt{n}} \geq 0. \quad (\text{D.9})$$

Since  $t \geq 0$ , the above inequality is achieved iff

$$t \geq \frac{\frac{8M \log n}{\sqrt{n}} + \sqrt{\frac{64M^2 \log^2 n}{n} + 480pM^2 \log n}}{6}. \quad (\text{D.10})$$

Hence if

$$\frac{64M^2 \log^2 n}{n} \leq 480pM^2 \log n \Leftrightarrow p \geq \frac{2 \log n}{15n}$$

holds, then the RHS of (D.10) is upper bounded by

$$\frac{\frac{8M \log n}{\sqrt{n}} + \sqrt{960pM^2 \log n}}{6} \leq \frac{\frac{8M \log n}{\sqrt{2 \log n}} \sqrt{15p} + 8M \sqrt{15p \log n}}{6} = \frac{2\sqrt{2} + 4}{3} M \sqrt{15p \log n}.$$

Therefore if  $t \geq \frac{2\sqrt{2}+4}{3} M \sqrt{15p \log n}$  is satisfied, then it implies that (D.9) holds. Hence plugging  $t = \frac{2\sqrt{2}+4}{3} M \sqrt{15p \log n}$  in (D.8) leads to the statement of the Lemma.

2. For a fixed  $i$ , consider now  $\sum_{j=1}^n Z_{ij}u_{2j}$  which is also a sum of independent random variables. One can verify that  $|Z_{ij}u_{2j}| \leq 2M\sqrt{B}$  for each  $j$ , with  $B$  as defined in the Lemma. Moreover,  $\mathbb{E}[(Z_{ij}u_{2j})^2] \leq B\mathbb{E}[Z_{ij}^2] \leq \frac{10pM^2B}{3}$  (see (C.1)) and hence,  $\sum_{i=1}^n \mathbb{E}[(Z_{ij}u_{2j})^2] \leq \frac{10pM^2Bn}{3}$  and it follows from Bernstein's inequality (see Theorem B.2) and the union bound that

$$\mathbb{P}(\|Zu_2\|_{\infty} \geq t) \leq 2n \exp \left( -\frac{3t^2}{20pM^2Bn + 4t\sqrt{BM}} \right). \quad (\text{D.11})$$

Proceeding identically as before, the reader is invited to verify that if  $p \geq \frac{2 \log n}{15n}$ , then the stated bound on  $\|Zu_2\|_{\infty}$  is obtained by plugging  $t = \frac{2\sqrt{2}+4}{3} M \sqrt{15pBn \log n}$  in (D.11).

□



## D.4 Proof of Lemma 8

*Proof.* 1. For a given  $i$ , we can write

$$\begin{aligned}
|(Z^2 u_1)_i| &= \|u_1\|_\infty \left| (Z^2 \underbrace{\frac{u_1}{\|u_1\|_\infty}}_{\tilde{u}_1})_i \right| \\
&= \frac{1}{\sqrt{n}} |(Z^2 \tilde{u}_1)_i| \quad (\text{with } \|\tilde{u}_1\|_\infty = 1) \\
&= \frac{\gamma^2}{\sqrt{n}} |((Z/\gamma)^2 \tilde{u}_1)_i|
\end{aligned} \tag{D.12}$$

for any  $\gamma > 0$ . For any integer  $m \geq 2$ , we have that

$$\begin{aligned}
\mathbb{E} \left[ \left( \frac{|Z_{ij}|}{\gamma} \right)^m \right] &= \frac{1}{\gamma^m} [\eta p |r_i - r_j|^m (1 - \eta p)^m + (1 - \eta) p \mathbb{E}[|N_{ij} - \eta p(r_i - r_j)|^m] + (1 - p) |\eta p(r_i - r_j)|^m] \\
&\leq \frac{1}{\gamma^m} [\eta p M^m + p(2M)^m + (\eta p)^m M^m] \\
&\leq \frac{1}{\gamma^m} [2^{m+1} p M^m].
\end{aligned}$$

Setting  $\gamma = (2)^{3/2} \sqrt{pn} M$  leads to

$$\mathbb{E} \left[ \left( \frac{|Z_{ij}|}{\gamma} \right)^m \right] \leq \frac{2^{m+1} p M^m}{(2)^{3m/2} p^{m/2} M^m n^{m/2}} = \frac{1}{2^{\frac{m}{2}-1}} \frac{1}{p^{\frac{m}{2}-1}} \frac{1}{n^{\frac{m}{2}}}.$$

It is easy to check that  $p \geq \frac{1}{2n}$  implies  $\mathbb{E} \left[ \left( \frac{|Z_{ij}|}{\gamma} \right)^m \right] \leq 1/n$ . Therefore invoking Theorem 11 for  $k = 2$ , we obtain for the stated choices of  $\xi, \mu, \kappa$  that  $|((Z/\gamma)^2 \tilde{u}_1)_i| \leq (\log n)^{2\xi}$  holds with probability at least  $1 - n^{-\frac{1}{4}(\log_\mu n)^{\xi-1}(\log_\mu e)^{-\xi}}$ . Plugging this in (D.12) with the expression for  $\gamma$ , and using the union bound, we obtain the statement of the Lemma.

2. The steps are identical to those above, with the only difference being  $\|u_2\|_\infty = \frac{M-\alpha}{\|r-\alpha e\|_2}$ .  $\square$

## E Proofs of Lemmas from Section 5.5

### E.1 Proof of Lemma 9

Note that  $\bar{D}_{ii} = \sum_{j=1}^n |H_{ij}|$  is the sum of independent random variables, where  $|H_{ij}| \leq M \forall i, j$ . Moreover,

$$\sum_{j=1}^n \mathbb{E}[|H_{ij}|^2] = \eta p \sum_{j=1}^n (r_i - r_j)^2 + (1 - \eta) \frac{npM}{2} \leq \eta pn M^2 + (1 - \eta) np \frac{M}{2} = npA(\eta, M).$$

Invoking Bernstein's inequality, it follows for any given  $i$  and  $t \geq 0$  that

$$\mathbb{P}(|\bar{D}_{ii} - \mathbb{E}[\bar{D}_{ii}]| \geq t) \leq 2 \exp \left( - \frac{t^2}{2(npA(\eta, M) + \frac{Mt}{3})} \right).$$

In order to get a high probability bound, we require  $t^2 \geq 4 \log n (npA(\eta, M) + \frac{Mt}{3})$ . Since  $t \geq 0$ , this is equivalent to saying that

$$t \geq \frac{\frac{4M}{3} \log n + \sqrt{(\frac{4M}{3} \log n)^2 + 16pn \log n A(\eta, M)}}{2}. \tag{E.1}$$

If  $(\frac{4M}{3} \log n)^2 \leq 6pn \log n A(\eta, M)$  or equivalently  $p \geq \frac{M^2}{9A(\eta, M)} \frac{\log n}{n}$ , then the RHS of (E.1) is bounded by

$$\sqrt{16pn \log n A(\eta, M)} \frac{\sqrt{2} + 1}{2} = 2(\sqrt{2} + 1) \sqrt{pn \log n} \sqrt{A(\eta, M)}.$$

Hence taking  $t = 2(\sqrt{2} + 1) \sqrt{pn \log n} \sqrt{A(\eta, M)}$  and applying the union bound, we obtain the statement of the Lemma.